Informatica® Cloud Data Integration
July 2024

Tasks

Informatica Cloud Data Integration Tasks
July 2024

© Copyright Informatica LLC 2006, 2024

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, Informatica Cloud, Informatica Intelligent Cloud Services, PowerCenter, PowerExchange, and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at https://www.informatica.com/trademarks.html. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2024-07-12

Table of Contents

Preface 10

Informatica Resources 10

- Informatica Documentation 10
- Informatica Intelligent Cloud Services web site 10
- Informatica Intelligent Cloud Services Communities 10
- Informatica Intelligent Cloud Services Marketplace 10
- Data Integration connector documentation 11
- Informatica Knowledge Base 11
- Informatica Intelligent Cloud Services Trust Center 11
- Informatica Global Customer Support 11

Chapter 1: Data integration tasks 12

Data filters 12

- Simple data filters 13
- Advanced data filters 13
- Data filter operators 14
- Data filter variables 15
- Rules and guidelines for data filters 15

Field metadata 16

Field expressions 17

- Creating a field expression 17
- Transformation language components for expressions 18
- Expression syntax 18
- String and numeric literals 18
- Rules and guidelines for expressions 19
- Adding comments to expressions 19
- Reserved words 20

Advanced session properties 21

Advanced session properties in advanced mode 27

Parameter files 28

Serverless usage properties 29

Schedules 29

- Repeat frequency 30
- Time zones and schedules 31
- Daylight Savings Time changes and schedules 31
- Creating a schedule 32
- Running a task on a schedule 32

Email notification 33

Preprocessing and postprocessing commands 33

- Preprocessing and postprocessing SQL commands 34

Preprocessing and postprocessing operating system commands. . . . 34
Monitoring a job. . . . 34
Data catalog discovery for sources. . . . 34
Catalog search. . . . 35
Discovering and selecting a catalog object. . . . 37
Stopping a job. . . . 38
Guidelines for sources and targets in data integration tasks. . . . 38
Rules and guidelines for flat file sources and targets. . . . 38
Rules and guidelines for database sources and targets. . . . 39

Chapter 2: Mapping tasks. . . . 40

Mapping task templates. . . . 40
Related objects. . . . 41
Advanced relationships. . . . 42
Spark session properties. . . . 42
SQL ELT optimization. . . . 44
SQL ELT optimization in advanced mode. . . . 44
Runtime strategies. . . . 45
CDC runtime properties in mapping tasks. . . . 46
Simultaneous task runs. . . . 47
Schema change handling. . . . 48
Dynamic schema handling options. . . . 48
Dynamic schema change handling rules and guidelines. . . . 49
Schema mismatch handling. . . . 50
Mapping task configuration. . . . 50
Creating a mapping task. . . . 50
Configuring sources. . . . 52
Configuring targets 54
Configuring CDC runtime properties. . . . 55
Configuring input parameters. . . . 56
Configuring persistent values. . . . 56
Configuring runtime options. . . . 57
Configuring the runtime strategy. . . . 62
CLAIRE Tuning. . . . 63
Guidelines to get an accurate recommendation. . . . 63
Configuring tuning. . . . 63
Initial tuning. . . . 64
Initial tuning results. . . . 64
Continuous tuning. . . . 65
Viewing and editing mapping task details. . . . 65
Sequence values. . . . 66
Running a mapping task. . . . 66
Reprocessing incrementally-loaded source files 66

Chapter 3: Dynamic mapping tasks. 69

Parameters in dynamic mapping tasks. 69
- Parameter scope. 71
- Parameter settings. 71

Jobs and job groups. 71
- Job settings. 72

Schema change handling. 73
- Schema change handling options. 73

Configuring a dynamic mapping task. 74
- Defining a dynamic mapping task. 74
- Configuring default parameters. 75
- Configuring jobs. 75
- Configuring groups. 76
- Configuring runtime options. 76

Chapter 4: Synchronization tasks. 77

Task operations. 77

Synchronization task sources. 78
- Rules and guidelines for multiple-object databases. 78

Synchronization task targets. 78
- Flat file target creation. 78
- Database target truncation 79
- Salesforce targets and IDs for related objects. 79
- Update columns. 79

Column names in flat files. 80

Field mappings. 80
- Field data types. 80
- Mapplets in field mappings. 81
- Lookup conditions. 81
- Lookup return values. 81
- Rules and guidelines for lookups. 82

Configuring a synchronization task. 83
- Synchronization prerequisite tasks. 83
- Defining a synchronization task. 83
- Configuring the source. 84
- Configuring the target. 86
- Configuring the data filters. 87
- Configuring the field mapping. 88
- Configuring a schedule and advanced options. 90

Viewing synchronization task details. 91

Running a synchronization task. 91
- Rules and guidelines for running a synchronization task. 92

Chapter 5: Data transfer tasks. 93

Task operations. 93
Data transfer task sources. 94
Source filters. 94
Sort conditions. 94
Custom queries. 94
Second sources. 95
Lookup condition. 95
Second source filters. 96
Data transfer task targets. 96
Database target truncation. 97
Update columns. 97
Field mapping. 97
Field data types. 98
Configuring a data transfer task. 98
Defining the data transfer task. 98
Configuring the source. 99
Configuring a second source. 99
Configuring the target. 100
Configuring the field mapping. 101
Configuring runtime options. 101
Running a data transfer task. 101

Chapter 6: Replication tasks. 102

Load types. 102
Full load. 102
Replication task sources. 103
Replication task targets. 103
Replicate data to a database target. 103
Replicate data to a flat file target. 103
Reset a database target. 104
Resetting a target table. 104
Rules and guidelines for resetting a target table. 104
Table and column names in a database target. 104
Table name truncation. 105
Duplicate tables names from same replication task. 105
Duplicate table names from different replication tasks. 105
Column name truncation. 105
Target prefixes. 106
Creating target tables. 106
Replication task schedules. 106
Configuring a replication task. 107

Rules and guidelines for configuring replication tasks. . . . 107
Replication prerequisite tasks. . . . 108
Defining a replication task. . . . 108
Configuring the source. . . . 108
Configuring the target. . . . 109
Configuring the field exclusions. . . . 110
Configuring the data filters. . . . 111
Configuring a schedule and advanced options. . . . 111
Viewing replication task details. . . . 112
Running a replication task. . . . 112
Rules and guidelines for running a replication task. . . . 112

Chapter 7: Masking tasks. . . . 113

Rules and guidelines for masking tasks. . . . 113
Masking task options. . . . 114
Source objects. . . . 114
Schema graph. . . . 114
Target task operations. . . . 115
Inplace masking. . . . 116
Update partial sandbox. . . . 116
Refresh fields. . . . 116
Validation reports. . . . 116
Staging database. . . . 117
Start the staging connection. . . . 117
H2 database configuration requirements. . . . 118
Installing and configuring H2 database manually on Windows. . . . 118
Installing H2 database manually on Linux. . . . 119
Data subset. . . . 120
Data subset options. . . . 121
Automatic task recovery. . . . 121
Parameter files in data filters. . . . 122
Configure relationship behavior. . . . 123
Data subset use cases for two objects. . . . 124
Case 1. Select the default path with filter on Account. . . . 124
Case 2. Select the configured path with filter on Account. . . . 125
Case 3. Select the default path with filter on Contact. . . . 127
Case 4. Select the configured path with filter on Contact. . . . 128
Data subset use cases for three objects. . . . 129
Case 1. Default path. . . . 130
Case 2. Configured path. . . . 131
Data subset rows. . . . 131
Data subset rows example. . . . 131
Refresh metadata. . . . 133

Reset task. . . . 134
Apply masking rules. . . . 134
Masking rule assignments. . . . 135
Add mapplets. . . . 135
Target fields. . . . 136
Default masking rules package. . . . 136
Schedule options. . . . 139
Email notification options. . . . 139
Advanced options. . . . 139
Configuring a masking task. . . . 140
Prerequisites. . . . 140
Step 1. Define the masking task. . . . 140
Step 2. Configure the source 141
Step 3. Configure the target. . . . 142
Step 4. Configure the data subset. . . . 142
Step 5. Define data masking rules. . . . 143
Step 6. Schedule the masking task. . . . 144
Masking task maintenance. . . . 144
Editing a masking task. . . . 145
Running a masking task manually. . . . 145
Refreshing the metadata in a masking task. . . . 145
Stopping a masking task. . . . 146
Resetting a masking task. . . . 147
Configuring masking task permissions. . . . 147
Copying a masking task. . . . 147
Renaming a masking task. . . . 147
Deleting a masking task. . . . 148
Exporting a masking task. . . . 148
Downloading mapping XML. . . . 148
Downloading validation reports. . . . 148
Dictionary files for data masking. . . . 149
Consistent masked output. . . . 150
Rules and guidelines. . . . 150
Example. . . . 151

Chapter 8: Masking rules. . . . 152

Masking rules. . . . 152
Repeatable output. . . . 153
Seed value. . . . 153
Optimize dictionary usage. . . . 154
Unique Substitution. . . . 154
Preprocessing and postprocessing expressions. . . . 154
Credit card masking. . . . 155

Credit card parameters. . . . 155
Email masking. . . . 156
Advanced email masking. . . . 156
IP address masking. . . . 157
Key masking. . . . 157
Key string masking. . . . 157
Key numeric masking. . . . 159
Key date masking. . . . 159
Nullification masking. . . . 160
Phone number masking. . . . 160
Random masking. . . . 160
Random string masking. . . . 161
Random numeric masking. . . . 161
Random date masking. . . . 162
SIN masking. . . . 163
SSN masking. . . . 164
Substitution masking. . . . 164
Substitution masking parameters. . . . 165
Custom substitution masking. . . . 165
Custom substitution masking parameters. . . . 166
Custom substitution lookup example. . . . 167
Custom substitution dictionary lookup use cases. . . . 168
Dependent masking. . . . 168
Dependent masking parameters. . . . 168
URL masking. . . . 169
Custom masking. . . . 169
Mapplet masking. . . . 169

Chapter 9: PowerCenter tasks. . . . 172

PowerCenter workflows. . . . 172
Supported transformations and mapping objects. . . . 173
Exception handling in stored procedures. . . . 174
Pre-session and post-session commands. . . . 174
Rules and guidelines for PowerCenter sources and targets. . . . 174
FTP/SFTP connections for PowerCenter tasks. . . . 175
Web Service connections for PowerCenter tasks. . . . 175
Parameters in PowerCenter tasks. . . . 175
PowerCenter task configuration. . . . 175
Configuring a PowerCenter task. . . . 176
Running a PowerCenter task. . . . 177

Index. . . . 179

Preface

Use *Tasks* to learn how to set up and run Data Integration tasks manually or on a schedule.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit https://docs.informatica.com.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at http://www.informatica.com/cloud. This site contains information about Informatica Cloud integration services.

Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

https://network.informatica.com/community/informatica-network/products/cloud-integration

Developers can learn more and share tips at the Cloud Developer community:

https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers

Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and mapplets:

https://marketplace.informatica.com/

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit https://docs.informatica.com.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit https://search.informatica.com. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at https://www.informatica.com/trust-center.html.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The Informatica Intelligent Cloud Services Status page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, on the Informatica Intelligent Cloud Services Status page, click **SUBSCRIBE TO UPDATES**. You can choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Global Customer Support

You can contact a Global Support Center through the Informatica Network or by telephone.

To find online support resources on the Informatica Network, click **Contact Support** in the Informatica Intelligent Cloud Services Help menu to go to the **Cloud Support** page. The **Cloud Support** page includes system status information and community discussions. Log in to Informatica Network and click **Need Help** to find additional resources and to contact Informatica Global Customer Support through email.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at https://www.informatica.com/services-and-training/support-services/contact-us.html.

CHAPTER 1

Data integration tasks

A data integration task is a process that you configure to analyze, extract, transform, and load data. You can run individual tasks manually or set tasks to run on a schedule.

You can use the following tasks to integrate data:

- Mapping. Use to process data based on the data flow logic defined in a mapping.
- Dynamic mapping. Use to run multiple jobs with different parameters based on the data flow logic defined in the same mapping.
- Synchronization. Use to load data and integrate applications, databases, and files. Includes add-on functionality such as mapplets.
- Data transfer. Use to move data from a source to a target. Optionally, sort and filter data before loading it to the target.
- Replication. Use to replicate data from Salesforce or database sources to database or file targets. You might replicate data to archive the data, perform offline reporting, or consolidate and manage data.
- Masking. Use to replace source data in sensitive columns with realistic test data for non-production environments. Masking rules define the logic to replace the sensitive data. Assign masking rules to the columns you need to mask.
- PowerCenter. Use to import a PowerCenter workflow and run it as a Data Integration PowerCenter task.

When you create a task, Data Integration walks you through the required steps. The options and properties that display depend on the task type, the options that you select, and the licenses enabled for the organization. For example, for a synchronization task, advanced Salesforce target options display on the **Schedule** page of the task wizard if you select a Salesforce target connection for the task on the **Target** page and your organization has the DSS Advanced Options license.

You can create a workflow of multiple tasks by linking the tasks in taskflows. For more information, see *Taskflows*.

Data filters

You can create the following type of data filters for any type of task:

- Simple
- Advanced

You can create a set of data filters for each object included in a replication task or synchronization task. Each set of data filters acts independently of the other sets.

Simple data filters

You can create one or more simple data filters.

When you create multiple simple data filters, the associated task creates an AND operator between the filters and loads rows that apply to all simple data filters.

For example, you load rows from the Account Salesforce object to a database table. However, you want to load only accounts that have greater than or equal to $100,000 in annual revenue and that have more than 500 employees. You configure the following simple data filters:

Field	Operator	Field Value
AnnualRevenue	>=	100000
NumberOfEmployees	>	500

Configuring simple data filters

You configure simple data filters in the task wizard.

1. To create a simple data filter, perform one of the following actions:
   - On the **Data Filters** page, click **Simple**, and then click **New**.
     The **Data Filter** dialog box appears.
   - In the **Filters** area on the **Sources** page, select **Simple** and then click **New**.
   - In a data transfer task, in the **Filters** area on the **Source** or **Second Source** page, select **Simple** and then click **New**.
2. Specify the object on which to create the data filter.

   You create separate data filters for each source object included in the task.
3. Enter the filter condition based on the field, operator, and field value.
4. Click **OK**.
5. Create additional simple data filters as needed.

   To delete a data filter, click the **Delete** icon next to the data filter.
6. Click **Next**.

Advanced data filters

Create an advanced data filter to create complex expressions that use AND, OR, or nested conditions.

When you create an advanced data filter, you enter one expression that contains all filters. The expression that you enter becomes the WHERE clause in the query used to retrieve records from the source.

For mappings in SQL ELT mode, use expression syntax that is valid in your cloud data warehouse. For more information, see the documentation for your cloud data warehouse. For other types of mappings, use Informatica transformation language expression syntax.

For example, you create a mapping to load rows from the Account Salesforce object to a database table. However, you want to load records where the billing state is California or New York and the annual revenue is greater than or equal to $100,000. You configure the following advanced filter expression:

```
(BillingState = 'CA' OR BillingState = 'NY') AND (AnnualRevenue >= 100000)
```

When you create a data filter on a Salesforce object, the corresponding task generates a SOQL query with a WHERE clause. The WHERE clause represents the data filter. The SOQL query must be less than 20,000 characters. If the query exceeds the character limit, the following error appears:

```
Salesforce SOQL limit of 5000 characters has been exceeded for the object: <Salesforce object>. Please exclude more fields or decrease the filters.
```

Note: Filter conditions are not validated until runtime.

Configuring advanced data filters

Configure advanced data filters in the task wizard. You can convert simple data filters to an advanced data filter, but you cannot convert an advanced data filter to simple data filters.

1. To create an advanced data filter, perform one of the following actions:
   - On the **Data Filters** page, click **New** > **Advanced**.
     To convert all simple data filters to one advanced data filter, on the **Data Filters** page, select a simple data filter and then click **Advanced**.
   - In the **Filters** area on the **Sources** page, select **Simple** and then click **New**.
   - In a data transfer task, in the **Filters** area on the **Source** or **Second Source** page, select **Advanced**.
     To convert all simple data filters to one advanced data filter, select **Advanced**.
2. If necessary, specify the object on which to create the data filter.

   You create separate data filters for each source object included in the task.
3. Enter the filter expression.

   Click the field name to add the field to the expression.
4. Click **OK**.

   To delete a data filter, click the **Delete** icon next to the data filter.
5. Click **Next**.

Data filter operators

You can use specific operators with each field type.

For mappings in SQL ELT mode, use operators that are valid in your cloud data warehouse's expression language. For more information, see the documentation for your cloud data warehouse. For other types of mappings, use the Informatica transformation language operators that are listed in the following table.

The following table shows the operators you can use for each field type:

Field type	Operators
Boolean	=, !=, Is Null, Is Not Null
Currency	=, !=, <, <=, >, >=, Is Null, Is Not Null
Date	=, !=, <, <=, >, >=, Is Null, Is Not Null
Datetime	=, !=, <, <=, >, >=, Is Null, Is Not Null
Double	=, !=, <, <=, >, >=, Is Null, Is Not Null
ID	=, !=, Is Null, Is Not Null

Field type	Operators
Int	=, !=, <, <=, >, >=, Is Null, Is Not Null
Reference	=, !=, Is Null, Is Not Null
String	=, !=, LIKE'_%', LIKE'%_', LIKE'%_%', Is Null, Is Not Null, <, <=, >, >=
Textarea	=, !=, LIKE'_%', LIKE'%_', LIKE'%_%', Is Null, Is Not Null, <, <=, >, >=
All other field types	=, !=, Is Null, Is Not Null

Data filter variables

Data filter variables represent the date or time that a task previously ran. Use data filter variables to help capture the source data that changed since the last task run. You can use data filter variables in simple and advanced data filter conditions.

You can use the following data filter variables:

Variable	Description
$LastRunDate	The start date in GMT time zone of the last task run that was successful or ended with a warning. Does not include time. For example, 2018-09-24. Can be used as a value for filter where the field type is DATE.
$LastRunTime	The start date and time in GMT time zone of the last task run that was successful or ended with a warning. For example, 2018-09-24 15:23:23. Can be used as a value for filter where the field type is DATETIME. You can't use the $LastRunTime variable with DATE fields .

You can't use data filter variables in expressions in mappings SQL ELT mode.

For example, you can include the following simple filter condition:

```
LastModifiedDate > $LastRunTime
```

Note: Consider time zone differences when comparing dates across time zones. The date and time of the $LastRunDate and $LastRunTime variables are based on the time zone set in Informatica Intelligent Cloud Services. The date and time of the actual job is based on the GMT time zone for Salesforce sources and the database server for database sources. The difference in the time zones may yield unexpected results.

Rules and guidelines for data filters

Use the following rules and guidelines for data filters:

- When you create a data filter in a mapping in SQL ELT mode, use expression syntax that is valid in your cloud data warehouse. For more information, see the documentation for your cloud data warehouse. For other types of mappings, use Informatica transformation language expression syntax.
- Data filters must contain valid SQL or SOQL operators.
- You cannot include simple and advanced data filters in the same task.
- When you convert a simple data filter to an advanced data filter, you cannot convert the advanced data filter back to a simple data filter.

- A task fails if the fields included in the data filter no longer exist or if the data types of the fields change. If a data type changes, edit the task.
- You can select Equals, Not Equals, Is Null, or Is Not Null operators on fields of the Other data type.
- Applications do not apply filters with Equals, Starts With, or Ends With operators and string fields that contain data that starts or ends with a single quotation mark. To filter these records, use the Contains operator.
- You can only use IS NULL and LIKE operators in data filters for fields of the Text, Ntext, and Image data types.
- If you specify a date and no time for a date/time filter, Data Integration uses 00:00:00 (12:00:00 a.m.) as the time.
- You cannot create a simple data filter in a synchronization task that includes a flat file source. You can create an advanced data filter.
- The list of available operators in a simple data filter depends on the data type of the field included in the data filter. Some operators do not apply to all fields included in data filters.
- When you enter more than one simple data filter, applications filter rows that meet the requirements of all data filters.
- When you use a parameter in a data filter, start the data filter with the parameter. For example, use $ $Sales=100000 instead of 100000=$$Sales.

Field metadata

You can view and edit field metadata such as the type, precision, and scale for parameterized source, target, and lookup objects with certain connection types. You can view and edit field metadata in mapping tasks and dynamic mapping tasks.

To view and edit field metadata in a mapping task, expand the **Field Data Types** area on the appropriate page of the mapping task editor. You configure source and lookup field metadata on the **Sources** page and target field metadata on the **Targets** page.

To view and edit field metadata in a dynamic mapping task, in the row that contains the object you want to view metadata for, click **Settings** and expand the **Field Data Types** area.

If you edit field metadata in the task and the field metadata changes after the task is saved, Data Integration uses the updated metadata. Typically, the is the desired behavior. However, if the task uses a flat file connection and you want to retain the metadata used at design time, enable the **Retain existing fields at runtime option**.

You can't update field metadata in mapping tasks that are based on mappings in SQL ELT mode.

To see if a connector supports field metadata configuration, see the help for the appropriate connector.

Field expressions

For some task types, you can configure the field mapping. The field mapping defines how source fields are mapped to target fields. You can specify an expression for each field mapping.

You can map multiple source fields to the same target field. For example, you can map SourceFieldA and SourceFieldB to TargetFieldC.

Data Integration might suggest operations when you map multiple source fields to a single target field. For example, if you map multiple text fields to a target text field, Data Integration concatenates the source text fields by default. You can change the default expression.

Data Integration provides a transformation language that includes SQL-like functions to transform source data. Use these functions to write expressions, which modify data or test whether data matches the conditions that you specify.

For more information about functions and the Data Integration transformation language, see *Function Reference*.

Creating a field expression

Create a field expression in a task wizard.

1. On the **Field Mappings** or **Input Parameters** page, select the target field for which you want to add an expression.
2. Perform one of the following actions based on your task type:
   - Click **Add or Edit Expression**.
   - Click **Actions** > **Edit Expression**, or click the field name.

   By default, the **Field Expression** dialog box shows the source field as the expression, which indicates that the target contains the same value as the source.
3. Enter the new field expression.

   To include source fields and system variables in the expression, you can select them from the **Source Fields** and **System Variables** tabs to insert them into the expression or you can add them to the expression manually.
4. Click **Validate Mapping** to validate the field mappings.
5. Click **Save**.

Validating expressions in field mappings

Use the following rules and guidelines when you validate an expression in a field mapping:

- When you validate mappings, Data Integration performs the following validations:
  - Verifies that the source and target fields in the task exist in the source or target. If the field does not exist, an error appears.
  - Verifies that all column data types are string and all field expressions contain string operations when the source and target are flat files.
  - Verifies that the correct parameters are used for each function and that the function is valid.
- The expression validator does not perform case-sensitive checks on field names.

- The expression validator verifies that the data type of a field in an expression matches the data type expected by the containing function. However, the expression validator does not check for incompatible data types between the following sets of objects:
  - Source and target fields of tasks.
  - Source field in a lookup condition and the lookup field.
  - Output of an expression or lookup and the target field.
    The expression or lookup with these incompatible data types may validate successfully, but, at runtime, the task fails and an error appears.
- If you map a string source field to a number target field, the validation succeeds. Data Integration tries to convert the string to a number using the atoi (ASCII to Integer) C function.
- The expression validator does not validate lookups.

Transformation language components for expressions

The transformation language includes the following components to create simple or complex expressions:

- Fields. Use the name of a source field to refer to the value of the field.
- Literals. Use numeric or string literals to refer to specific values.
- Functions. Use these SQL-like functions to change data in a task.
- Operators. Use transformation operators to create expressions to perform mathematical computations, combine data, or compare data.
- Constants. Use the predefined constants to reference values that remain constant, such as TRUE.

In mappings in SQL ELT mode, you use your cloud data warehouse's native expression components to create expressions. You don't use the Informatica transformation language to create expressions.

Expression syntax

You can create a simple expression that only contains a field, such as ORDERS, or a numeric literal, such as 10. You can also write complex expressions that include functions nested within functions, or combine different fields using the transformation language operators.

Note: Although the transformation language is based on standard SQL, there are differences between the two languages.

String and numeric literals

You can include numeric or string literals.

Enclose string literals within single quotation marks. For example:

```
'Alice Davis'
```

String literals are case sensitive and can contain any character except a single quotation mark. For example, the following string is not allowed:

```
'Joan's car'
```

To return a string containing a single quotation mark, use the CHR function:

```
'Joan' || CHR(39) || 's car'
```

Do not use single quotation marks with numeric literals. Just enter the number you want to include. For example:

```
.05
```

or

```
$$Sales_Tax
```

Rules and guidelines for expressions

Use the following rules and guidelines when you write expressions:

- For each source field, you can perform a lookup or create an expression. You cannot do both.
- You cannot use strings in numeric expressions.
  For example, the expression `1 + '1'` is not valid because you can only perform addition on numeric data types. You cannot add an integer and a string.
- You cannot use strings as numeric parameters.
  For example, the expression `SUBSTR(TEXT_VAL, '1', 10)` is not valid because the SUBSTR function requires an integer value, not a string, as the start position.
- You cannot mix data types when using comparison operators.
  For example, the expression `123.4 = '123.4'` is not valid because it compares a decimal value with a string.
- You can pass a value from a field, literal string or number, or the results of another expression.
- Separate each argument in a function with a comma.
- Except for literals, the transformation language is not case sensitive.
- The colon (:), comma (,), and period (.) have special meaning and should be used only to specify syntax.
- Data integration tasks treat a dash (-) as a minus operator.
- If you pass a literal value to a function, enclose literal strings within single quotation marks. Do not use quotation marks for literal numbers. Data integration tasks treat any string value enclosed in single quotation marks as a character string.
- Do not use quotation marks to designate fields.
- You can nest multiple functions within an expression. Data integration tasks evaluate the expression starting with the innermost function.
- When you use a parameter in an expression, use the appropriate function to convert the value to the necessary data type. For example, you might use the following expression to define a quarterly bonus for employees:

  ```
  IIF((EMP_SALES < TO_INTEGER($$SalesQuota), 200, 0)
  ```

Adding comments to expressions

You can use the following comment specifiers to insert comments in expressions:

- Two dashes:

  ```
  -- These are comments
  ```

- Two forward slashes:

  ```
  // These are comments
  ```

Data integration tasks ignore all text on a line preceded by comment specifiers. For example, to concatenate two strings, enter the following expression with comments in the middle of the expression:

```
 -- This expression concatenates first and last names for customers:
FIRST_NAME -- First names from the CUST table
|| // Concat symbol
LAST_NAME // Last names from the CUST table
// Joe Smith Aug 18 1998
```

Data integration tasks ignore the comments and evaluates the expression as follows:

```
FIRST_NAME || LAST_NAME
```

You cannot continue a comment to a new line:

```
-- This expression concatenates first and last names for customers:
FIRST_NAME -- First names from the CUST table
|| // Concat symbol
LAST_NAME // Last names from the CUST table
Joe Smith Aug 18 1998
```

In this case, data integration tasks do not validate the expression because the last line is not a valid expression.

Reserved words

Some keywords, such as constants, operators, and system variables, are reserved for specific functions. These include:

- :EXT
- :INFA
- :LKP
- :MCR
- :SD
- :SEQ
- :SP
- :TD
- AND
- DD_DELETE
- DD_INSERT
- DD_REJECT
- DD_UPDATE
- FALSE
- NOT
- NULL
- OR
- PROC_RESULT
- SPOUTPUT
- TRUE
- WORKFLOWSTARTTIME

The following words are reserved for Informatica Intelligent Cloud Services:

- ABORTED
- DISABLED
- FAILED
- NOTSTARTED
- STARTED

- STOPPED
- SUCCEEDED

Note: You cannot use a reserved word to name a field. Reserved words have predefined meanings in expressions.

Advanced session properties

Advanced session properties are optional properties that you can configure in mapping tasks, dynamic mapping tasks, and Visio templates. Use caution when you configure advanced session properties. The properties are based on PowerCenter advanced session properties and might not be appropriate for use with all tasks.

You can configure the following types of advanced session properties:

- General
- Performance
- Advanced
- Error handling

Note: Advanced mode uses a different set of advanced session properties. Mappings in SQL ELT mode don't use advanced session properties.

General options

The following table describes the general options:

General options	Description
Write Backward Compatible Session Log File	Writes the session log to a file.
Session Log File Name	Name for the session log. Use any valid file name. You can customize the session log file name in one of the following ways: - Using a static name. A static log file name is a simple static string with or without a file extension. If you use a static name, the log file name is appended with a sequence number each time the task runs, for example samplelog.1, samplelog.2. When the maximum number of log files is reached, the numbering sequence begins a new cycle. - Using a dynamic name. A log file name is dynamic when it includes a parameter defined in a parameter file or a system variable. You can include any of the following system variables: - $CurrentTaskName. Replaced with the task name. - $CurrentTime. Replaced with the current time. - $CurrentRunId. Replaced with the run ID for the current job. If you use a dynamic name, the file name is unique for every task run. The Maximum Number of Log Files property is not applied. To purge old log files, delete the files manually.
Session Log File Directory	Directory where the session log is saved. Use a directory local to the Secure Agent to run the task. By default, the session log is saved to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/logs`

General options	Description
$Source Connection Value	Source connection name for Visio templates .
$Target Connection Value	Target connection name for Visio templates .
Source File Directory	Source file directory path. Use for flat file connections only.
Target File Directory	Target file directory path. Use for flat file connections only.
Treat Source Rows as	When the task reads source data, it marks each row with an indicator that specifies the target operation to perform when the row reaches the target. Use one of the following options: - Insert. All rows are marked for insert into the target. - Update. All rows are marked for update in the target. - Delete. All rows are marked for delete from the target. - Data Driven. The task uses the Update Strategy object in the data flow to mark the operation for each source row.
Commit Type	Commit type to use. Use one of the following options. - Source. The task performs commits based on the number of source rows. - Target. The task performs commits based on the number of target rows. - User Defined. The task performs commits based on the commit logic defined in the Visio template. When you do not configure a commit type, the task performs a target commit.
Commit Interval	Interval in rows between commits. When you do not configure a commit interval, the task commits every 10,000 rows.
Commit on End of File	Commits data at the end of the file.
Rollback Transactions on Errors	Rolls back the transaction at the next commit point when the task encounters a non-fatal error. When the task encounters a transformation error, it rolls back the transaction if the error occurs after the effective transaction generator for the target.
Java Classpath	Java classpath to use. The Java classpath is added to the beginning of the system classpath when the task runs. Use this option when you use third-party Java packages, built-in Java packages, or custom Java packages in a Java transformation.

Performance settings

The following table describes the performance settings:

Performance settings	Description
DTM Buffer Size	Amount of memory allocated to the task from the DTM process. By default, a minimum of 12 MB is allocated to the buffer at run time. Use one of the following options: - Auto. Enter Auto to use automatic memory settings. When you use Auto, configure **Maximum Memory Allowed for Auto Memory Attributes**. - A numeric value. Enter the numeric value that you want to use. The default unit of measure is bytes. Append KB, MB, or GB to the value to specify a different unit of measure. For example, 512MB. You might increase the DTM buffer size in the following circumstances: - When a task contains large amounts of character data, increase the DTM buffer size to 24 MB. - When a task contains n partitions, increase the DTM buffer size to at least n times the value for the task with one partition. - When a source contains a large binary object with a precision larger than the allocated DTM buffer size, increase the DTM buffer size so that the task does not fail.
Incremental Aggregation	Performs incremental aggregation for tasks based on Visio templates .
Reinitialize Aggregate Cache	Overwrites existing aggregate files for a task that performs incremental aggregation.
Enable High Precision	Processes the Decimal data type to a precision of 28.
Session Retry on Deadlock	The task retries a write on the target when a deadlock occurs.
SQL ELT Optimization	Type of SQL ELT optimization. Use one of the following options: - None. The task processes all transformation logic for the task. - To Source. The task pushes as much of the transformation logic to the source database as possible. - To Target. The task pushes as much of the transformation logic to the target database as possible. - Full. The task pushes as much of the transformation logic to the source and target databases as possible. The task processes any transformation logic that it cannot push to a database. - $$PushdownConfig. The task uses the SQL ELT optimization type specified in the user-defined parameter file for the task. When you use $$PushdownConfig, ensure that the user-defined parameter is configured in the parameter file. When you use SQL ELT optimization, do not use the **Error Log Type** property. For more information, see the help for the appropriate connector. The SQL ELT optimization functionality varies depending on the support available for the connector. For more information, see the help for the appropriate connector.

Performance settings	Description
Create Temporary View	Allows the task to create temporary view objects in the database when it pushes the task to the database. Use when the task includes an SQL override in the Source Qualifier transformation or Lookup transformation. You can also use for a task based on a Visio template that includes a lookup with a lookup source filter.
Create Temporary Sequence	Allows the task to create temporary sequence objects in the database. Use when the task is based on a Visio template that includes a Sequence Generator transformation.
Enable cross-schema SQL ELT optimization	Enables SQL ELT optimization for tasks that use source or target objects associated with different schemas within the same database. To see if cross-schema SQL ELT optimization is applicable to the connector you use, see the help for the relevant connector. This property is enabled by default.
Allow SQL ELT Optimization for User Incompatible Connections	Indicates that the database user of the active database has read permission on idle databases. If you indicate that the database user of the active database has read permission on idle databases, and it does not, the task fails. If you do not indicate that the database user of the active database has read permission on idle databases, the task does not push transformation logic to the idle databases.
Session Sort Order	Order to use to sort character data for the task.

Advanced options

The following table describes the advanced options:

Advanced options	Description
Constraint Based Load Ordering	Currently not used in Informatica Intelligent Cloud Services.
Cache Lookup() Function	Caches lookup functions in Visio templates with unconnected lookups. Overrides lookup configuration in the template. By default, the task performs lookups on a row-by-row basis, unless otherwise specified in the template.

Advanced options	Description
Default Buffer Block Size	Size of buffer blocks used to move data and index caches from sources to targets. By default, the task determines this value at run time. Use one of the following options: - Auto. Enter Auto to use automatic memory settings. When you use Auto, configure Maximum Memory Allowed for Auto Memory Attributes. - A numeric value. Enter the numeric value that you want to use. The default unit of measure is bytes. Append KB, MB, or GB to the value to specify a different unit of measure. For example, 512MB. The task must have enough buffer blocks to initialize. The minimum number of buffer blocks must be greater than the total number of Source Qualifiers, Normalizers for COBOL sources, and targets. The number of buffer blocks in a task = DTM Buffer Size / Buffer Block Size. Default settings create enough buffer blocks for 83 sources and targets. If the task contains more than 83, you might need to increase DTM Buffer Size or decrease Default Buffer Block Size.
Line Sequential Buffer Length	Number of bytes that the task reads for each row. Data Integration dynamically increases the maximum line sequential buffer length from the default of 1024 bytes.
Maximum Memory Allowed for Auto Memory Attributes	Maximum memory allocated for automatic cache when you configure the task to determine the cache size at run time. You enable automatic memory settings by configuring a value for this attribute. Enter a numeric value. The default unit is bytes. Append KB, MB, or GB to the value to specify a different unit of measure. For example, 512MB. If the value is set to zero, the task uses default values for memory attributes that you set to auto.
Maximum Percentage of Total Memory Allowed for Auto Memory Attributes	Maximum percentage of memory allocated for automatic cache when you configure the task to determine the cache size at run time. If the value is set to zero, the task uses default values for memory attributes that you set to auto.
Additional Concurrent Pipelines for Lookup Cache Creation	Restricts the number of pipelines that the task can create concurrently to pre-build lookup caches. You can configure this property when the **Pre-build Lookup Cache** property is enabled for a task or transformation. When the **Pre-build Lookup Cache** property is enabled, the task creates a lookup cache before the Lookup receives the data. If the task has multiple Lookups, the task creates an additional pipeline for each lookup cache that it builds. To configure the number of pipelines that the task can create concurrently, select one of the following options: - Auto. The task determines the number of pipelines it can create at run time. - Numeric value. The task can create the specified number of pipelines to create lookup caches.
Custom Properties	Configure custom properties for the task. You can override the custom properties that the task uses after the job has started. The task also writes the override value of the property to the session log.

Advanced options	Description
Pre-build Lookup Cache	Allows the task to build the lookup cache before the Lookup receives the data. The task can build multiple lookup cache files at the same time to improve performance. You can configure this option in a Visio template or in a task. The task uses the task-level setting if you configure the Lookup option as Auto for a Visio template. Configure one of the following options: - Always allowed. The task can build the lookup cache before the Lookup receives the first source row. The task creates an additional pipeline to build the cache. - Always disallowed. The task cannot build the lookup cache before the Lookup receives the first row. When you use this option, configure the **Configure the Additional Concurrent Pipelines for Lookup Cache Creation** property. The task can pre-build the lookup cache if this property is greater than zero.
DateTime Format String	Date time format for the task. You can specify seconds, milliseconds, or nanoseconds. To specify seconds, enter `MM/DD/YYYY HH24:MI:SS`. To specify milliseconds, enter `MM/DD/YYYY HH24:MI:SS.MS`. To specify microseconds, enter `MM/DD/YYYY HH24:MI:SS.US`. To specify nanoseconds, enter `MM/DD/YYYY HH24:MI:SS.NS`. By default, the format specifies microseconds, as follows: MM/DD/YYYY HH24:MI:SS.US.
Pre 85 Timestamp Compatibility	Do not use with Data Integration.

Error handling

The following table describes the error handling options:

Error handling options	Description
Stop on Errors	Indicates how many non-fatal errors the task can encounter before it stops the session. Non-fatal errors include reader, writer, and DTM errors. Enter the number of non-fatal errors you want to allow before stopping the session. The task maintains an independent error count for each source, target, and transformation. If you specify 0, non-fatal errors do not cause the session to stop.
Override Tracing	Overrides tracing levels set on an object level.
On Stored Procedure Error	Determines the behavior when a task based on a Visio template encounters pre-session or post-session stored procedure errors. Use one of the following options: - Stop Session. The task stops when errors occur while executing a pre-session or post-session stored procedure. - Continue Session. The task continues regardless of errors. By default, the task stops.
On Pre-Session Command Task Error	Determines the behavior when a task that includes pre-session shell commands encounters errors. Use one of the following options: - Stop Session. The task stops when errors occur while executing pre-session shell commands. - Continue Session. The task continues regardless of errors. By default, the task stops.

Error handling options	Description
On Pre-Post SQL Error	Determines the behavior when a task that includes pre-session or post-session SQL encounters errors: - Stop Session. The task stops when errors occur while executing pre-session or post-session SQL. - Continue. The task continues regardless of errors. By default, the task stops.
Error Log Type	Specifies the type of error log to create. You can specify flat file or no log. Default is none. You cannot log row errors from XML file sources. You can view the XML source errors in the session log. Do not use this property when you use the **SQL ELT Optimization** property.
Error Log File Directory	Specifies the directory where errors are logged. By default, the error log file directory is $PMBadFilesDir\.
Error Log File Name	Specifies error log file name. By default, the error log file name is PMError.log.
Log Row Data	Specifies whether or not to log transformation row data. When you enable error logging, the task logs transformation row data by default. If you disable this property, n/a or -1 appears in transformation row data fields.
Log Source Row Data	Specifies whether or not to log source row data. By default, the check box is clear and source row data is not logged.
Data Column Delimiter	Delimiter for string type source row data and transformation group row data. By default, the task uses a pipe (|) delimiter. **Tip:** Verify that you do not use the same delimiter for the row data as the error logging columns. If you use the same delimiter, you may find it difficult to read the error log file.

Advanced session properties in advanced mode

When you configure a mapping task or a dynamic mapping task that is based on a mapping in advanced mode, you can configure optional advanced session properties.

You can configure the following types of advanced session properties:

- General
- Custom

General properties

The following table describes the general properties:

Advanced session properties	Description
DateTime Format String	Date time format for the task. To specify seconds, enter `MM/DD/YYYY HH24:MI:SS`. To specify milliseconds, enter `MM/DD/YYYY HH24:MI:SS.MS`. To specify microseconds, enter `MM/DD/YYYY HH24:MI:SS.US`. To specify nanoseconds, enter `MM/DD/YYYY HH24:MI:SS.NS`. By default, the format specifies microseconds, as follows: MM/DD/YYYY HH24:MI:SS.US.
Override Mapping Task Timeout	Overrides the mapping task timeout set in the advanced configuration that is associated with the runtime environment.
Override Tracing	Overrides tracing levels set on an object level.

Custom properties

The following table describes the custom property:

Advanced custom properties	Description
advanced.custom.property	Configure custom properties to run the mapping. You can override the custom properties that the task uses after the job has started. The task also writes the override value of the property to the session log. Use `&:` to separate custom properties.

Parameter files

A parameter file is a list of user-defined parameters and their associated values.

Use a parameter file to define values that you want to update without having to edit the task. You update the values in the parameter file instead of updating values in a task. The parameter values are applied when the task runs.

You can use a parameter file to define parameter values in the following tasks:

Mapping tasks

Define parameter values for connections in the following transformations:

- Source
- Target
- Lookup
- SQL

Define parameter values for objects in the following transformations:

- Source

- Target
- Lookup

Also, define values for parameters in data filters, expressions, and lookup expressions.

Note: Not all connectors support parameter files. To see if a connector supports runtime override of connections and data objects, see the help for the appropriate connector.

Synchronization tasks

Define values for parameters in data filters, expressions, and lookup expressions.

PowerCenter tasks

Define values for parameters and variables in data filters, expressions, and lookup expressions.

You enter the parameter file name and location when you configure the task.

You can't use a parameter file in a mapping task that is based on a mapping in SQL ELT mode.

Serverless usage properties

Serverless usage properties define how a task requests resources from a serverless runtime environment

If the task runs in a serverless runtime environment, you can configure the following serverless usage properties:

Property	Description
Max Compute Units	Maximum number of serverless compute units corresponding to machine resources that the task can use. Overrides the corresponding property in the serverless runtime environment. By default, the maximum number of compute units is the value that is configured in the serverless runtime environment. If an administrator sets the maximum number of compute units in the serverless runtime environment to a value that is lower than the number configured in the task, the task requests the lower number.
Task Timeout	Amount of time in minutes to wait for the task to complete before it is terminated. The timeout ensures that serverless compute units are not unproductive when the task hangs. By default, the timeout is the value that is configured in the serverless runtime environment.

Schedules

You can run tasks manually or you can use schedules to run them at a specific time or interval such as hourly, daily, or weekly.

To use a schedule, you associate the task with a schedule when you configure the task. You can use an existing schedule or create a new schedule. If you want to create a schedule, you can create the schedule from the task's **Schedule** or **Runtime Options** page during task configuration.

When you create a schedule, you specify the date and time. You can configure a schedule to run associated assets throughout the day between 12:00 a.m. and 11:55 p.m. Informatica Intelligent Cloud Services might add a small schedule offset to the start time, end time, and all other time configurations. As a result, scheduled tasks and taskflows might start later than expected. For example, you configure a schedule to run

hourly until noon, and the schedule offset for your organization is 10 seconds. Informatica Intelligent Cloud Services extends the end time for the schedule to 12:00:10 p.m., and the last hourly task or taskflow starts at 12:00:10 p.m. To see the schedule offset for your organization, check the **Schedule Offset** organization property.

You can monitor scheduled tasks from the **All Jobs** page in Monitor. Scheduled tasks do not appear on the **My Jobs** page.

When you copy a task that includes a schedule, the schedule is not associated with the new task. To associate a schedule with the new task, edit the task.

If you remove a task from a schedule as the task runs, the job completes. Data Integration cancels any additional runs associated with the schedule.

Repeat frequency

The repeat frequency determines how often tasks run. You can set the repeat frequency to every N minutes, hourly, daily, weekly, biweekly, or monthly.

The following table describes the repeat frequency options:

Option	Description
Does not repeat	Tasks run as scheduled and do not repeat.
Every N minutes	Tasks run on an interval based on a specified number of minutes. You can configure the following options: - Repeat frequency. Select a frequency in minutes. Options are 5, 10, 15, 20, 30, 45. - Days. Days of the week when you want tasks to run. You can select one or more days of the week. - Time range. Hours of the day when you want tasks to start. Select All Day or configure a time range. You can configure a time range between 00:00-23:55. - Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.
Hourly	Tasks run on an hourly interval based on the start time of the schedule. You can configure the following options: - Repeat frequency. Select a frequency in hours. Options are 1, 2, 3, 4, 6, 8, 12. - Days. Days of the week when you want tasks to run. You can select one or more days of the week. - Time range. Hours of the day when you want tasks to start. Select All Day or configure a time range. You can configure a time range between 00:00-23:55. - Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.
Daily	Tasks run daily at the start time configured for the schedule. You can configure the following options: - Repeat frequency. The frequency at which you want tasks to run. Select Every Day or Every Weekday. - Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.
Weekly	Tasks run on a weekly interval based on the start time of the schedule. You can configure the following options: - Days. Days of the week when you want tasks to run. You can select one or more days of the week. - Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time. If you do not specify a day, the schedule runs regularly on the same day of the week as the start date.

Option	Description
Biweekly	Tasks run every two weeks based on the start time of the schedule. You can configure the following options: - Days. Days of the week when you want tasks to run. You can select one or more days of the week. You must select at least one day. - Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time. If you configure a biweekly schedule to start at 5 p.m. on a Tuesday and run tasks every two weeks on Mondays, the schedule begins running tasks on the following Monday.
Monthly	Tasks run on a monthly interval based on the start time of the schedule. You can configure the following options: - Day. Day of the month when you want tasks to run. You can configure one of the following options: - Select the exact date of the month, between 1-28. If you want the task to run on days later in the month, use the <n> <day of the week> option. - Select the <n> <day of the week>. Options for <n> include First, Second, Third, Fourth, and Last. Options for <day of the week> includes Day, and Sunday-Saturday. Tip: With the Day option, you can configure tasks to run on the First Day or the Last Day of the month. - Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.

Time zones and schedules

Informatica Intelligent Cloud Services stores time in Coordinated Universal Time (UTC). When you log in, Informatica Intelligent Cloud Services converts the time and displays it in the time zone associated with your user profile.

When you create a schedule, you select the time zone for the scheduler to use. You can select a time zone that is different from your time zone or your organization time zone.

Daylight Savings Time changes and schedules

Informatica Intelligent Cloud Services applies Daylight Savings Time changes to all tasks except biweekly tasks.

When Daylight Savings time goes into effect, tasks scheduled to run between 2:00 a.m. and 2:59 a.m., do not run the day that the time changes from 2:00 a.m. to 3:00 a.m. If a task is scheduled to run biweekly at 2 a.m., it will run at 3 a.m. the day of the time change and at 2 a.m. for the next run.

Daylight Savings Time does not trigger additional runs for tasks that are scheduled to run between 1:00 a.m. - 1:59 a.m. when Standard Time begins. For example, a task is scheduled to run every day at 1:30 a.m. When the time changes from 2 a.m. to 1 a.m., the task does not run again at 1:30 a.m.

Tip: To ensure that Informatica Intelligent Cloud Services does not skip any scheduled runs near the 2 a.m. time change, do not schedule jobs to run between 12:59 a.m. and 3:01 a.m.

Creating a schedule

You can create a schedule in Data Integration when you configure a task or linear taskflow. You can also create a schedule in Administrator if you have the appropriate permissions.

The following procedure describes how to create a schedule when you access the **Schedule** page from Data Integration during task or linear taskflow configuration.

1. Select **Run this task on a schedule**, and then click **New**.
2. Configure the following properties:

Property	Description
Schedule Name	Name of the schedule. Each schedule name must be unique within the organization. Schedule names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Schedule names are not case sensitive.
Description	Description of the schedule. Maximum length is 255 characters.
Starts	Date and time when the schedule takes effect. The date format is MM/DD/YYYY. Time appears in the 24-hour format. Click the calendar button to select the start date. The start date and time can affect the repeat frequency for tasks and taskflow jobs that repeat at regular intervals. For example, if the start date is November 10 and the repeat frequency is monthly, the schedule runs associated assets on the tenth day of each month. If the start time is 3:10 and the repeat frequency is hourly, the assets run every hour at 10 minutes past the hour. Default is the current date, current time, and time zone of the user that creates the schedule.
Time Zone	Select the time zone for the schedule to use. The time zone can differ from the organization time zone or user time zone.
Repeats	Repeat frequency for the schedule. Select one of the following options: - Does Not Repeat - Every N Minutes - Hourly - Daily - Weekly - Monthly Default is Does Not Repeat.

3. Click **Save** to save the schedule and return to the task configuration page.

Running a task on a schedule

Associate a task with a schedule on the **Schedule** or **Runtime Options** page when you configure the task. You can use an existing schedule or create a schedule.

1. Select **Run this task on a schedule**.
2. To specify whether to use an existing schedule or a new schedule, perform one of the following tasks:
   - To use an existing schedule, select the schedule that you want to use.

- To create a schedule to use for the task, click **New**, and then configure the schedule properties. For more information on creating a schedule, see the Administrator help.

3. Click **Save**.

Email notification

You can configure email notification for a task. When you configure custom email notification, Data Integration uses the custom email notification instead of the email notification options configured for the organization.

To configure email notification options, perform the following steps in the task wizard:

1. Specify whether to use the default email notification options that have been set for your organization or create custom email notification for the task. Configure email notification using the following options:

Field	Description
Use Default Email Notification Options for my Organization	Use the email notification options configured for the organization.
Use Custom Email Notification Options for this Task	Use the email notification options configured for the task. You can send email to different addresses based on whether the task failed, completed with errors, or completed successfully. Use commas to separate a list of email addresses. When you select this option, email notification options configured for the organization are not used.

2. Click **Save**.

Preprocessing and postprocessing commands

You can run preprocessing and postprocessing commands to perform additional jobs. The task runs preprocessing commands before it reads the source. It runs postprocessing commands after it writes to the target.

You can use the following types of commands:

- SQL commands. Use SQL commands to perform database tasks.
- Operating system commands. Use shell and DOS commands to perform operating system tasks.

If any command in the preprocessing or postprocessing scripts fail, the task fails.

Preprocessing and postprocessing SQL commands

You can run SQL commands before or after a task. For example, you can use SQL commands to drop indexes on the target before the task runs, and then recreate them when the task completes. Data Integration does not validate the SQL.

Use the following rules and guidelines when creating the SQL commands:

- Use any command that is valid for the database type. However, Data Integration does not allow nested comments, even if the database allows them.
- Use a semicolon (;) to separate multiple statements. Data Integration issues a commit after each statement.
- Data Integration ignores semicolons within comments. If you need to use a semicolon outside of comments, you can escape it with a backslash (\).

Preprocessing and postprocessing operating system commands

Data Integration can perform operating system commands before or after the task runs. For example, use a preprocessing shell command to archive a copy of the target flat file before the task runs on a UNIX machine.

You can use the following types of operating system commands:

- UNIX. Any valid UNIX command or shell script.
- Windows. Any valid DOS or batch file.

Enter multiple preprocessing or postprocessing commands as a single line without spaces.

If the Secure Agent is on a Windows machine, separate commands with an ampersand (&). If the Secure Agent is on a Linux machine, separate commands with a semicolon (;).

Monitoring a job

You can monitor tasks or taskflows that are currently running, have completed, or have stopped.

Monitor jobs on the following pages:

- Monitor the jobs that you initiated on the **My Jobs** page in Data Integration.
- Monitor running jobs in your organization on the **Running Jobs** page in Monitor.
- Monitor all jobs in your organization on the **All Jobs** page in Monitor.

For more information about monitoring jobs, see *Monitor*.

Data catalog discovery for sources

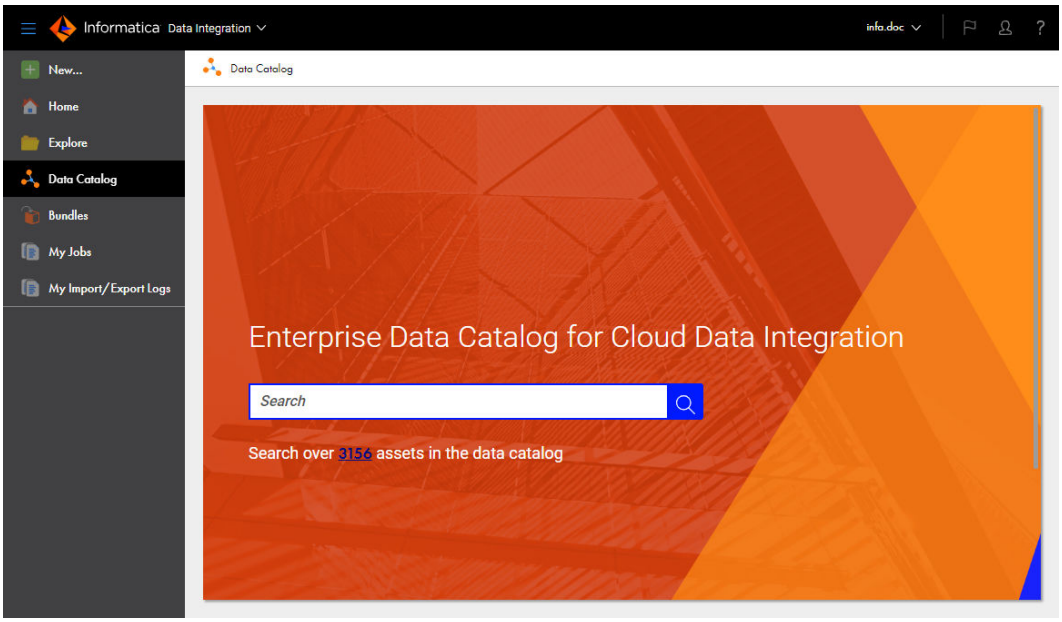
If your organization uses Enterprise Data Catalog and you have the appropriate license, you can perform a search against the catalog and discover catalog assets. You can use the assets that you discover as sources,

targets, and lookup objects in mappings and as sources in synchronization and file ingestion and replication tasks.

Note: Before you can use data catalog discovery, your organization administrator must configure the Enterprise Data Catalog integration properties on the **Organization** page in Administrator. For more information about configuring Enterprise Data Catalog integration properties, see the Administrator help.

Perform data catalog discovery on the **Data Catalog** page.

The following image shows the **Data Catalog** page:



The page displays a **Search** field and the total number of table, view, and flat file assets in the catalog.

In the **Search** field, enter a search phrase that might occur in the object name, description, or other metadata such as the data domain or associated business glossary term. When you select an object from the search results, Data Integration asks you where you want to use the object.

To use the object as a source in a synchronization or file ingestion and replication task, select **Create a new asset** and choose the task. Data Integration imports the connection if it does not exist in your organization. Data Integration then creates the task and adds the object to the task as the source object. You cannot add the object as a source in an existing task.

You can't use data catalog discovery to add objects to mappings in SQL ELT mode.

Catalog search

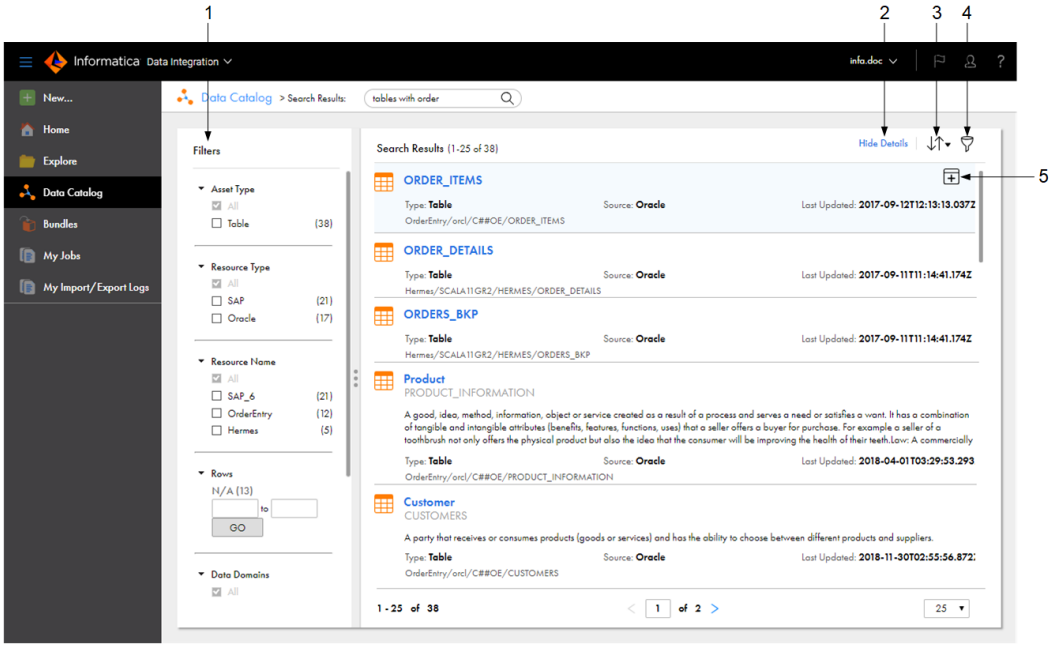
Use the search on the **Data Catalog** page to find an Enterprise Data Catalog object. Enter the object name, part of the name, or keywords associated with the object in the **Search** field, and then click the search icon. Data Integration returns all tables, views, and flat files in the catalog that match the search criteria.

You can use the `*` and `?` wildcard characters in the search phrase. For example, to find objects that start with the string "Cust", enter `Cust*` in the **Search** field.

You can also enter keyword searches. For example, if you enter `tables with order` in the **Search** field, Data Integration returns tables with "order" in the name or description, tables that have the associated business term "order," and tables that contain columns for which the "order" data domain is inferred or assigned.

For more information about Enterprise Data Catalog searches and search results, see the Enterprise Data Catalog documentation.

The following image shows an example of search results when you enter "tables with order" as the search phrase:



1. Filter search results.
2. Show or hide object details.
3. Sort search results.
4. Apply or remove all filters.
5. Use the selected object in a mapping, a synchronization task, or a file ingestion and replication task.

You can perform the following actions on the search results page:

Filter search results.

Use the filters to filter search results by asset type, resource type, resource name, number of rows, data domains, and date last updated.

Show details.

To display details about the object, click **Show Details**.

Sort results.

Use the **Sort** icon to sort results by relevance or name.

Open an object in Enterprise Data Catalog.

To open an object in Enterprise Data Catalog, click the object name. To view the object, you must log in to Enterprise Data Catalog with your Enterprise Data Catalog user name and password.

Use the object in a synchronization task, file ingestion and replication task, or mapping.

To use the object in a synchronization task, file ingestion and replication task or mapping, click **Use Object**. You can select an object if the object is a valid source, target, or lookup type for a mapping or a valid source type for the task. For example, you can select an Oracle table to use as the source in a new synchronization task, but you cannot select a Hive table.

When you select the object, Data Integration prompts you to select the task where you want to use the object and imports the connection if it does not exist.

Connection properties vary based on the object type. Data Integration imports most connection properties from the resource configuration in Enterprise Data Catalog, but you must enter other required properties, such as the connection name and password.

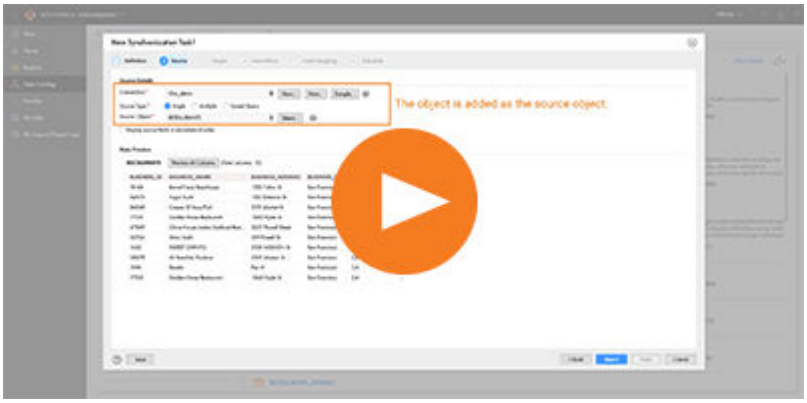
After you configure the connection or if the connection already exists, Data Integration adds the object to a new synchronization task, file ingestion and replication task, or to the inventory of a new or open mapping.

Discovering and selecting a catalog object

Discover and select a catalog object so that you can use the object as a source in a new synchronization or file ingestion and replication task.

Before you can use data catalog discovery, your organization administrator must configure the Enterprise Data Catalog integration properties on the **Organization** page in Administrator.

The following video shows you how to discover and select a catalog object as the source in a new synchronization task:



1. Open the **Data Catalog** page.
2. Enter the search phrase in the search field.

   For example, to find customer tables, you might enter "Customer," "Cust*," or "tables with customer."
3. On the search results page, click **Use Object** in the row that contains the object.

   You can select one object at a time.

   Data Integration prompts you to select where to use the object.
4. Select one of the following options:
   - To add the object to a new synchronization task, click **New Synchronization Task**.
   - To add the object to a new file ingestion and replication task, click **New File Ingestion and Replication Task**.
   - To add the object to a new mapping, click **New Mapping**.
   - To add the object to an open mapping, click **Add to an open asset**, and then select the mapping.

5. Click **OK**.

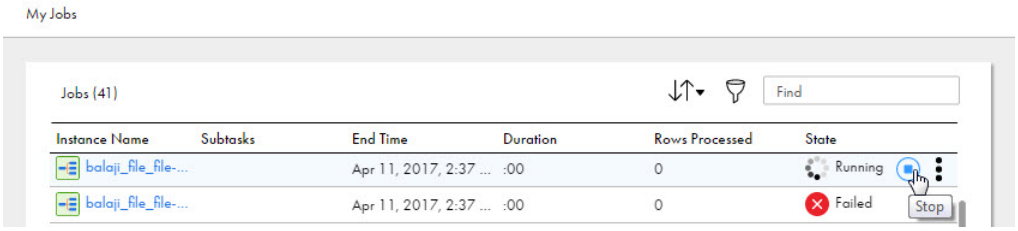
   If the connection does not exist in your organization, Data Integration prompts you to import the connection. Enter the missing connection properties such as the connection name and password.

   If you use the object in a synchronization or file ingestion and replication task, Data Integration creates the task with the object as the source. Configure other task properties such as the target, data filters, field mapping, and scheduling information.

Stopping a job

A job is an instance of a mapping, task, or taskflow. You can stop a running job on the **All Jobs**, **Running Jobs**, or **My Jobs** page.

1. Open Monitor and select **All Jobs** or **Running Jobs**, or open Data Integration and select **My Jobs**.
2. In the row that contains the job that you want to stop, click the **Stop** icon.



To view details about the stopped job, click the job name.

Guidelines for sources and targets in data integration tasks

Use the following rules and guidelines for sources and targets in data integration tasks:

- Field names must contain 65 characters or less.
- Field names must contain only alphanumeric or underscore characters. Spaces are not allowed.
- Field names cannot start with a number.
- Each field name must be unique within each source and target object.
- Data Integration truncates data if the scale or precision of a numeric target column is less than the scale or precision of the corresponding source column.

Rules and guidelines for flat file sources and targets

Use the following rules and guidelines for flat file sources and targets:

- All date columns in a flat file source must have the same date format. Rows that have dates in a different format than the one specified in the synchronization task definition are written to the error rows file.
- Each flat file target must contain all fields that will be populated by the synchronization task.

- The synchronization task truncates a flat file target before writing target data to the file.
  To avoid overwriting target data, you might use a post-session command to merge target data with a master target file in a different location.
- The flat file cannot contain empty column names. If a file contains an empty column name, the following error appears:

  ```
  Invalid header line: Empty column name found.
  ```
- Do not map binary fields when you use a flat file source or target in a synchronization task.
- Column names in a flat file must contain printable tab or ASCII characters (ASCII code 32-126). If the file contains a character that is not valid, the following error appears:

  ```
  Invalid header line: Non-printable character found. The file might be binary or might
  have invalid characters in the header line.
  ```
- You can use a tab, space, or any printable special character as a delimiter. The delimiter can have a maximum of 10 characters. The delimiter must be different from the escape character and text qualifier.
- For flat file sources and targets with multibyte data on Linux, the default locale must be UTF-8.

Rules and guidelines for database sources and targets

Use the following rules and guidelines for database sources and targets:

- You can use database tables as targets. You can use database tables, aliases, and views as sources.
- Relational targets must meet the minimum system requirements.
- The database user account for each database target connection must have DELETE, INSERT, SELECT, and UPDATE privileges.

CHAPTER 2

Mapping tasks

Use the mapping task to process data based on the data flow logic defined in a mapping.

When you create a mapping task, you select the mapping for the task to use. The mapping must already exist before you can create a mapping task for it. Alternatively, you can create a mapping task using a template.

If the mapping includes parameters, you can define the parameters when you configure the task or define the parameters when you run the task. You can use user-defined parameters for data filters, expressions, and lookup expressions in a mapping task. You define user-defined parameters in a parameter file associated with the task.

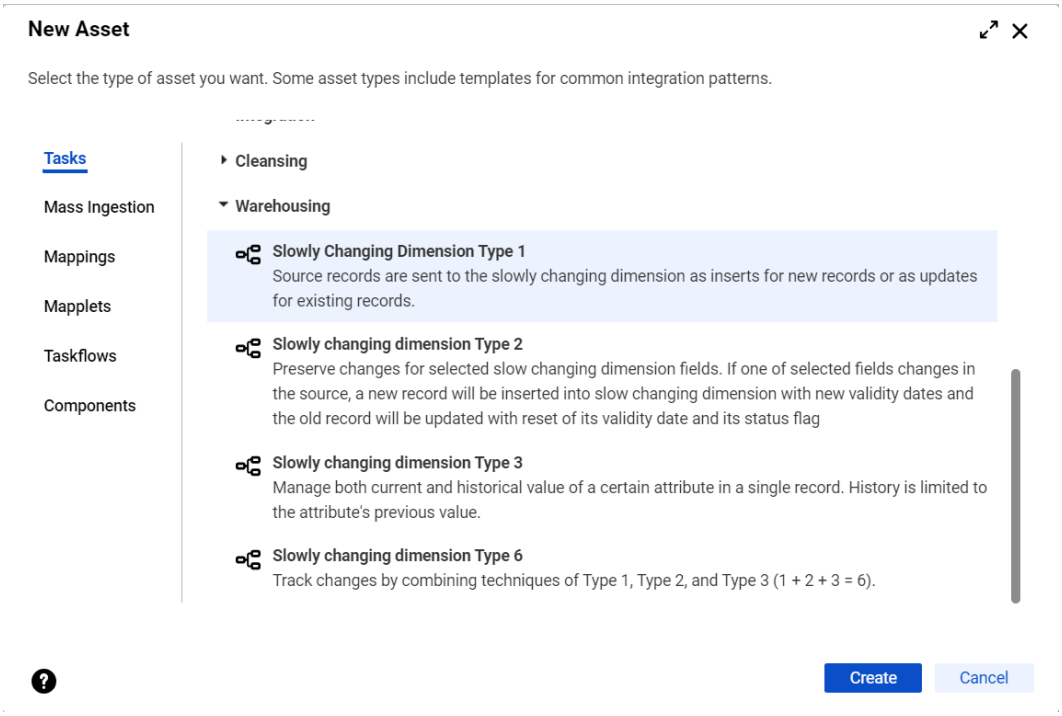
At run time, a mapping task processes task data based on the data flow logic from the mapping, the parameters defined in the task, and the user-defined parameters defined in a parameter file, when available.

Mapping task templates

Use a mapping task template to run a mapping task without creating a mapping beforehand.

Each mapping task template is based upon a mapping template. Use a mapping task template when the mapping on which the mapping task template is based suits your needs. When you select a mapping task template, Data Integration creates a copy of the template for you to use. When you define the mapping task in the task wizard, you save a copy of the mapping template on which the mapping task template is based.

Templates are divided into three categories: Integration, Cleansing, and Warehousing, as shown in the following image:



The templates range from simple templates that you can use to copy data from one source to another, to complex templates that you can use for data warehousing-related tasks.

Related objects

When a mapping includes a source that is a parameter and is configured for multiple objects, you can join related objects in the task.

You can join related objects based on existing relationships or custom relationships. Data Integration restricts the type of relationships that you can create based on the connection type.

Use the following relationships to join related objects:

Existing relationships

You can use relationships defined in the source system to join related objects. You can join objects with existing relationships for Salesforce, database, and some Data Integration Connectors connection types.

After you select a primary object, you select a related object from a list of related objects.

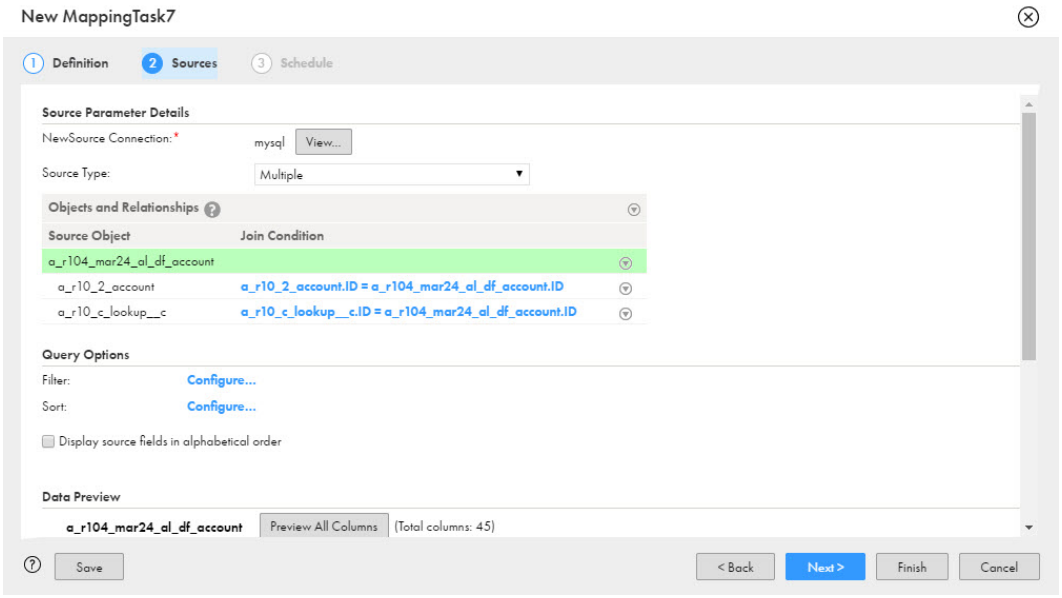
Custom relationships

You can use custom relationships to join multiple source objects. You can create custom relationships for the database connection type.

When you create a custom relationship for database objects, you create an inner, left outer, or right outer join on the source fields that you select.

To join source objects, you add the primary source object in the **Objects and Relationships** table. Then you add related objects, specify keys for the primary and related objects, and configure the join type and operator.

For more information about related source objects, see the Source Transformation section in *Transformations*.



Advanced relationships

You can create an advanced relationship for database sources when the source object in the mapping is a parameter and configured for multiple sources. You cannot create an advanced relationship between source objects that have been joined using a custom relationship.

When you create an advanced relationship, the wizard converts any relationships that you defined to an SQL statement that you can edit.

To create an advanced relationship, you add the primary source object in the Objects and Relationships table. Then you select fields and write the SQL statement that you want to use. Use an SQL statement that is valid for the source database. You can also add additional objects from the source.

Spark session properties

When you create a mapping task that is based on a mapping in advanced mode, you can configure optional Spark session properties.

The default values for Spark session properties are configured using best practices and the average computational requirements of in-house mapping tasks. If the default values do not fit the requirements of a specific mapping task, reconfigure the properties to override the default values.

To get an optimal set of Spark session properties for the mapping task, see "CLAIRE Tuning" on page 63.

The following table describes the Spark session properties:

Spark session property	Description
infaspark.sql.forcePersist	Indicates whether data persists in memory to avoid repeating read operations. For example, the Router transformation can avoid repeated read operations on output groups. Default is false.
spark.driver.extraJavaOptions	Additional JVM options for the Spark driver process. Default is `-Djava.security.egd=file:/dev/./urandom -XX:MaxMetaspaceSize=256M -XX:+UseG1GC -XX:MaxGCPauseMillis=500`.
spark.driver.maxResultSize	Maximum total size of serialized results of all partitions for each Spark action. Default is 4G.
spark.driver.memory	Amount of memory for the Spark driver process. Default is 4G.
spark.dynamicAllocation.maxExecutors	Maximum number of Spark executors if dynamic allocation is enabled. Default is 1000. The value is calculated automatically.
spark.executor.cores	Number of cores that run each Spark executor. Default is 2.
spark.executor.extraJavaOptions	Additional JVM options for Spark executors. Default is `-Djava.security.egd=file:/dev/./urandom -XX:MaxMetaspaceSize=256M -XX:+UseG1GC -XX:MaxGCPauseMillis=500`.
spark.executor.memory	Amount of memory for each Spark executor. Default is 6G.
spark.memory.fraction	Fraction of the heap that is allocated to the Spark engine. When set to 1, the Spark engine uses the full heap space except for 300 MB that is reserved memory. Default is 0.6.
spark.memory.storageFraction	Fraction of memory that the Spark engine uses for storage compared to processing data. Default is 0.5.
spark.rdd.compress	Indicates whether to compress serialized RDD partitions. Default is false.
spark.reducer.maxSizeInFlight	Maximum size of the data that each reduce task can receive from a map task while shuffling data. The size represents a network buffer to make sure that the reduce task has enough memory for the shuffled data. Default is 48M.
spark.shuffle.file.buffer	Size of the in-memory buffer that each map task uses to write the intermediate shuffle output. Default is 32K.

Spark session property	Description
spark.sql.autoBroadcastJoinThreshold	Threshold in bytes to use broadcast join. When the Spark engine uses broadcast join, the Spark driver sends data to Spark executors that are running on the advanced cluster, thereby avoiding shuffling and resulting in better performance. In some situations, like when a mapping task processes columnar formats or delimited files, broadcast join can cause memory issues at the Spark driver level. To resolve the issues, try reducing the broadcast join threshold to 10 MB, increasing the Spark driver memory, or disabling broadcast join. Default is 256000000. To disable broadcast join, set the value to -1.
spark.sql.broadcastTimeout	Timeout in seconds that is used during broadcast join. Default is 300.
spark.sql.shuffle.partitions	Number of partitions that Spark uses to shuffle data to process joins or aggregations. Default is 100.
spark.custom.property	Configure custom Spark session properties. Use `&:` to separate custom properties.

SQL ELT optimization

You can use SQL ELT optimization to push transformation logic to source databases or target databases for execution. Use SQL ELT optimization when using database resources can improve task performance.

When you run a task configured for SQL ELT optimization, the task converts the transformation logic to an SQL query. The task sends the query to the database, and the database executes the query.

The amount of transformation logic that you can push to the database depends on the database, transformation logic, and task configuration. The task processes all transformation logic that it cannot push to a database.

Use the **SQL ELT Optimization** advanced session properties to configure SQL ELT optimization for a task. You can configure the SQL ELT optimization type, how to run the task if SQL ELT optimization is not possible, whether to allow the task to create temporary views in the database, and other options.

SQL ELT Optimization advanced session properties aren't displayed for mapping tasks that are based on mappings in SQL ELT mode because mappings in SQL ELT mode are automatically configured to push transformation logic to the cloud data warehouse.

Note: SQL ELT optimization functionality varies depending on the support available for the connector. For more information, see the help for the appropriate connector.

SQL ELT optimization in advanced mode

You can run a mapping in advanced mode using full SQL ELT optimization to push all transformation logic to the source and target databases.

If full SQL ELT optimization isn't possible, Data Integration runs the mapping logic on the Data Integration Server and on an advanced cluster.

The following table describes the log files that you can use to troubleshoot the job:

Log file	Description
Session log	Indicates whether SQL ELT optimization is possible.
Mapping compilation log	Displays the SQL query that runs on the database.
Spark driver log	Displays the SQL query that runs on the database and the individual entries that track job progress on the database.

For more information about viewing log files, see *Monitor*.

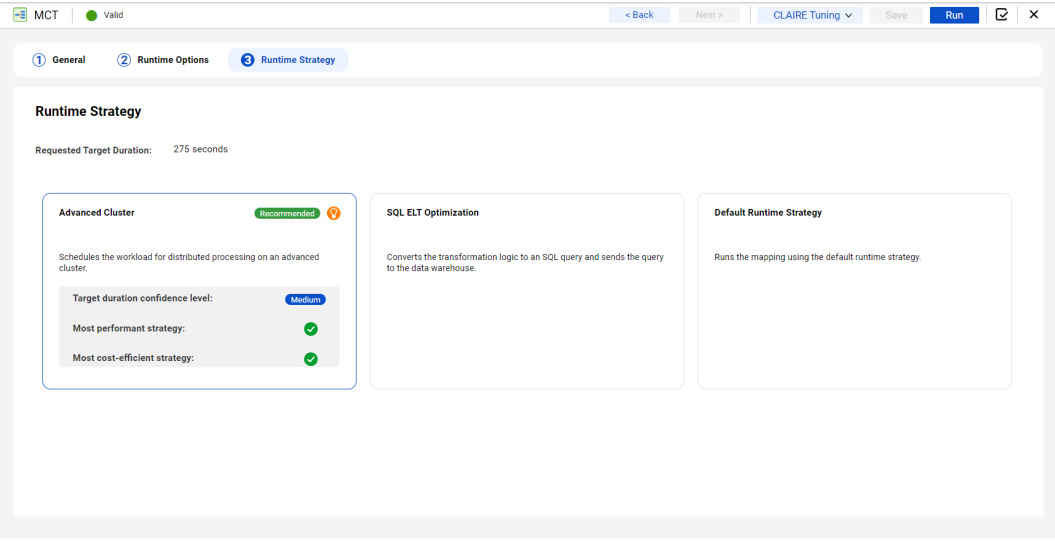
Runtime strategies

A runtime strategy determines whether the mapping task runs on an advanced cluster or using SQL ELT optimization. Runtime strategies are powered by CLAIRE, Informatica's AI engine. CLAIRE uses FinOps capabilities to identify the runtime strategy that saves the most on cloud infrastructure costs and optimizes job performance.

If CLAIRE recommendations are enabled in your organization, you can select a runtime strategy when you create a mapping task based on a mapping in advanced mode. CLAIRE analyzes the mapping and recommends one of the following runtime strategies to optimize cost, performance, or both:

- Advanced cluster. Schedules the workload for distributed processing on an advanced cluster.
- SQL ELT optimization. Converts the transformation logic to an SQL query and sends the query to the data warehouse.

The following image shows a recommendation that you might see:



The default runtime strategy runs the mapping task on an advanced cluster or using SQL ELT optimization based on whether SQL ELT optimization is configured on the **Runtime Options** page. If SQL ELT optimization is configured, the mapping task runs using SQL ELT optimization. Otherwise, the mapping task runs on an advanced cluster.

The target duration confidence level shows the likelihood that the job will run within the target duration that you specify. The confidence level can change based on the runtime strategy that you select. If you select the runtime strategy that is more performant, it's more likely that the mapping task will run within its target duration.

Note: Runtime strategy recommendations and the target duration confidence level might not appear for some mapping tasks. CLAIRE continues collecting metadata in the background and shows these details when it has enough confidence.

Runtime strategies aren't available for mapping tasks that run in a serverless runtime environment.

For more information about SQL ELT optimization, see "SQL ELT optimization" on page 44.

CDC runtime properties in mapping tasks

When you create or edit a mapping task that processes change data from a CDC source, you can set optional properties on the **CDC Runtime** page. These properties are used when the mapping task runs.

The following table describes the optional CDC runtime properties:

Property	Description
Maximum Rows Per Commit	Maximum number of change records in a UOW that the mapping task can process. When this maximum is reached, the UOW change records are committed on the target. Valid values are 0 through 999999999. Default is 0, which causes this property to be ignored.
Minimum Rows Per Commit	Minimum number of change records that must exist in a UOW before the PWX CDC Reader can request a real-time flush of the change records to the target. Valid values are 0 through 999999999. Default is 0, which causes this property to be ignored.
Maximum Latency in Seconds	Maximum number of seconds that the PWX CDC Reader waits for more change records before flushing change data from partially filled or empty blocks and returning control to Data Integration. Valid values are 2 through 360. Default is 2.
Real-time Flush Latency in Milliseconds	Number of milliseconds that must elapse before the PWX CDC Reader requests a real-time flush of uncommitted data in the change data stream to the targets defined in the mapping. Valid values are -1 through 999999999. Default is 2000. **Note:** To control commit processing, you can also use the **UOW Count**, **Maximum Rows Per Commit**, and **Minimum Rows Per Commit** properties.
Restart Point	Displays the restart point in read-only mode. The restart point indicates the position in the change data stream from which the PWX CDC Reader will start reading change records for a source table. Usually, you set the restart point before running the mapping task the first time to coincide with the time when the target table was materialized. You might also need to set the restart point to reprocess change records or to perform some testing. To set the restart point, click **Select**. If you do not specify a restart point, the mapping task starts or restarts from the end of log by default.

Property	Description
Restart Revision	Displays the restart-point revision number in read-only mode. When you navigate to the **CDC Runtime** page, open the **Select Restart Point** dialog box, and click **OK**, the revision number is incremented by 1, regardless of whether you changed the restart point. **Note:** During a single mapping-task edit session, the revision number is incremented by 1 only once, even if you return to the **Select Restart Point** dialog box multiple times. When a new revision is created, the CDC Connector ignores the restart information in the PM_REC_STATE recovery table and uses the specified restart point the next time you run the mapping task. **Note:** If you run the mapping task and then change to a different source object, you must either increment the restart revision number or create a new mapping task. Initial value is 0.
UOW Count	Number of units of work (UOWs) in the change data stream that the PWX CDC Reader processes before it requests a real-time flush of change records to the targets in the mapping. Valid values are -1 through 999999999. The values -1 and 0 cause this property to be ignored. Default is -1.
Update as Delete and Insert	Controls whether each SQL update that is captured from a source table is processed as an update or as a delete followed by an insert. You can select this property to handle changes to primary keys in source tables, in which case the insert record contains the key changes. Otherwise, leave this property cleared for normal processing of inserts, updates, and deletes. By default, this check box is not selected, which causes updates to be processed as update operations.
Continuous Extraction	Controls whether the mapping task runs continuously to capture change data from the source. Selecting this property indicates that the mapping task runs until it is stopped in the Monitor. By default, this check box is not selected and the mapping task stops after a single execution.

Note: You can also use the Informatica Intelligent Cloud Services Data Integration REST API to view and update CDC runtime options. The **mttask** API resource lets you retrieve or update mapping task options. For example, you might use **mttask** to change a selected option in multiple mapping tasks by using the API endpoint provided for the connector. For more information, see the *REST API Reference*.

Simultaneous task runs

You can run multiple instances of a mapping task at the same time.

You might want to enable simultaneous task runs to load target files concurrently.

You can use multiple task instances in a Parallel Paths step of a taskflow or in two different taskflows that run in parallel. You can also run multiple instances of a mapping task simultaneously using the REST API version 2 job resource. For more information about running tasks using the job resource, see *REST API Reference*.

To enable simultaneous task runs, select the **Allow the mapping task to be executed simultaneously** on the **Runtime Options** page when you configure the task.

Use caution when you configure mapping tasks to run simultaneously. Mapping features that change each time the task runs, such as in-out parameters and sequence generator values, might produce unexpected results when you run the task instances simultaneously.

Schema change handling

You can choose how Data Integration handles changes that you make to the schema of some data object types.

By default, if you make changes to the schema, Data Integration does not pick up the changes automatically. If you want Data Integration to refresh the data object schema every time the mapping task runs, you can enable dynamic schema handling.

A schema change includes one or more of the following changes to the data object:

- Fields are added.
- Fields are deleted.
- Fields are renamed.
- Field data type, precision, or scale is updated.

Data Integration automatically refreshes the schema for relational objects every time the task runs. If you want to dynamically refresh the schema for other object types, enable dynamic schema change handling in the **Advanced Options** area of the **Runtime Options** page when you configure the task.

The following table describes the schema change handling options:

Option	Description
Asynchronous	Default. Data Integration refreshes the schema when you edit the mapping or mapping task, and when Informatica Intelligent Cloud Services is upgraded.
Dynamic	Data Integration refreshes the schema every time the task runs. Applicable for source, target, and lookup objects of certain connector types. For some connector types, Data Integration can only refresh the schema if the data object is a flat file. If you select this option, the file object format must be delimited. Not applicable to hierarchical data. To see if a connector supports dynamic schema change handling, see the help for the appropriate connector.

If you update fields in the source object and you enable dynamic schema handling, be sure to update the Target transformation field mapping. Data Integration writes Null to the target fields that were previously mapped to the renamed or deleted source fields. If you use a target created at run time, update the target object name so that Data Integration creates a new target when the task runs. The task fails if Data Integration tries to alter a target created in a previous task run.

Dynamic schema handling options

When you enable dynamic schema change handling, you can select how Data Integration applies schema changes from upstream transformations to the target object. If the mapping contains more than one target, select the schema change handling for each target.

To select target schema options, the target field mapping must be automatic.

When you configure target schema options for objects that are created at runtime, Data Integration creates the target the first time you run the task. In subsequent task runs, Data Integration updates the target based on the schema change option that you select.

The schema change handling options available are based on the target connection. To see if a connector supports dynamic schema change handling, see the help for the appropriate connector.

The following table describes the schema handling options that you can select for each target type:

Schema handling option	Target type	Description
Keep Existing File Format	File	Data Integration fetches the most recent target schema at runtime and does not apply upstream schema changes to the target file.
Drop Current and Recreate	Database and file	For database targets, Data Integration drops the existing target table and creates a new target table with the schema from the upstream transformations on every run. For file targets, Data Integration updates the target schema to match the incoming schema on every task run.
Alter and Apply Changes	Database	Data Integration updates the target schema with additive changes to match the schema from the upstream transformations. It does not delete columns from the target.
Don't Apply DDL Changes	Database	Data Integration fetches the target schema at runtime and does not apply upstream schema changes to the target table.

Data Integration does not pass field constraints to the target. For example, the source contains fields S1 and S2 configured with the NOT NULL constraint. The target contains fields T1 and T2 also configured with the NOT NULL constraint. You select the **Alter and Apply Changes** schema handling option. When you run the task, fields S1 and S2 are written to the target with no constraints.

Dynamic schema change handling rules and guidelines

Enable dynamic schema change handling so that Data Integration refreshes the data object schema every time the mapping task runs.

Consider the following rules and guidelines when you enable dynamic schema change handling:

- Changes to the object schema take precedence over changes to the field metadata in the mapping. For example, you add a field to the source object and then edit the metadata of an existing field in the mapping. At run time, Data Integration adds the new field and does not edit the existing field.
- Data Integration resolves parameters before picking up the object schema.
- Data Integration treats renamed fields as deleted and added columns. If you rename a field, you might need to update transformations that reference the renamed field. For example, if you rename a field that is used in the lookup condition, the lookup cannot find the new field and the task fails.
- When you rename, add, or delete fields, you might need to update the field mapping. For example, if you delete all the previously mapped fields in a target object, you must remap at least one field or the task fails.
- Data Integration writes Null values to a target field in the following situations:
  - You rename a target field with automatic field mapping, and the field name does not match a source field.
  - You rename a source field with manual field mapping, and you do not remap the field to the target.
- If you delete a field from a source or lookup object and a downstream transformation references the field, the task fails.
- If you change a source or lookup field type, the task might fail if the new field type results in errors downstream. For example, if you change an integer field in an arithmetic expression to a string field, the expression is not valid and the task fails.

- If you change a target field type, Data Integration converts the data from the incoming field to the new target field type. If the conversion results in an error, Data Integration drops the row. For example if you change a string type to a date type where the string does not contain a date, Data Integration drops the row.

Schema mismatch handling

You can choose how Data Integration responds when it encounters a mismatch between a file and the selected schema.

When you run a mapping through a mapping task, you can configure an advanced option that specifies how Data Integration handles mismatched schemas:

- Skip mismatched files and continue. When Data Integration finds a schema mismatch, it stops searching for other errors in the same file and writes that error to the log. Data Integration doesn't process any other records from that file and continues with the next file.
- Stop on first mismatched file. Data Integration stops all processing when it finds a schema mismatch error, and it writes the error to the log. Data Integration does not roll backfiles processed before the error was found, and it doesn't process the file containing the error.

When you run a mapping in the Mapping Designer, Data Integration evaluates every file and skips the entire file when it encounters a schema mismatch.

Schema mismatch handling doesn't apply to mapping tasks that are based on mappings in SQL ELT mode.

Mapping task configuration

Complete the following steps to create a mapping task:

1. Create the mapping task.
2. Configure the source.
3. Configure the target.
4. Optionally, configure CDC runtime properties.
5. Define input parameters.
6. Optionally, define persistent values.
7. Optionally, configure runtime options.
8. Optionally, configure the runtime strategy.

As you configure the task, you can click **Save** to save your work at any time.

Creating a mapping task

Define general properties of the task on the **General** page.

1. To create a mapping task, click **New** > **Tasks**, and then complete one of the following steps:

- To create a mapping task based on a mapping, select **Mapping Task** and click **Create**.
- To create a mapping task using a template, expand the appropriate template category, select the template you want to use, and then click **Create**.

2. To edit a mapping task, on the **Explore** page, navigate to the mapping task. In the row that contains the task, click **Actions** and select **Edit**.
3. Configure the following fields:

Field	Description
Name	Name of the task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Task names are not case sensitive.
Location	Project or folder in which the task resides. If the **Explore** page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.
Description	Description of the task. Maximum length is 4000 characters.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task. **Note:** In a sub-organization, you can't run a mapping task on a Secure Agent group shared by the parent organization if the task is based on a mapping in SQL ELT mode or advanced mode.
Mapping	Mapping associated with the task. To select the mapping, click **Select** and navigate to the mapping that you want to use. To open the mapping, click **View**.

4. Configure the job priority and target duration if the mapping task is based on a mapping in advanced mode.

   The following table describes the properties:

Property	Description
Job Priority	Priority to schedule the job. High-priority jobs are scheduled sooner than medium-priority or low-priority jobs.
Target Duration (seconds)	Time in seconds to complete the job. You can view whether the target duration was met or exceeded when you monitor the job. The actual time to complete the job depends on the advanced cluster infrastructure and the runtime strategy.

5. Click **Next**.

Configuring sources

Configure source and lookup source parameters on the **Sources** page. The **Sources** page displays differently based on the basis for the task. If the mapping does not include source or lookup parameters, the **Sources** page does not appear.

You can add a single source object or multiple source objects based on the connection type and mapping configuration.

If the mapping specifies a connection parameter, and after you create the mapping task you edit the mapping to change the source, you might need to edit the task. Check the specific connection in the mapping task to determine if you need to reset it, and then validate the task again.

1. For each source or lookup source parameter, configure the following details as required:

Parameter detail	Description
Connection	Select a connection. To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**. **Tip:** Hover over the source connection tool tip to view more information about the connection.
Source type	Source type. The options available depend on the connection type.
Object	Select an object or enter a query. If a list of objects doesn't appear, click **Select**. The **Select Source Object** dialog box displays up to 200 objects. If the object you want to use doesn't appear, enter a search string to reduce the number of objects. **Tip:** Hover over the source object tool tip to view more information about the object.
Display Technical Field Names Instead of Labels	Displays technical names instead of business names. Not available for all connection types.
Add Currently Processed File Name	Adds the source file name to each row. Data Integration adds the CurrentlyProcesedFileName field to the source at run time. Available for parameterized source objects with flat file connections.

2. Configure formatting options if required.

   The properties available vary based on the connection type. For information about a particular connector's properties, see the help for the appropriate connector.

3. Optionally, expand the **Data Preview** area to preview the data.

   To display fields in alphabetical order, click **Display fields in alphabetical order**. By default, fields appear in the order returned by the source system.
   To download the preview results, click **Download**.

4. For parameterized source objects, configure data filters if necessary.
5. For parameterized source objects, configure sort options if necessary.
6. Configure field metadata if required.

You can configure field metadata for sources with certain connection types. To see if a connector supports field metadata configuration, see the help for the appropriate connector. You can't update field metadata in mapping tasks that are based on mappings in SQL ELT mode.

To edit field metadata, expand the **Field Data Types** area and select the row that contains the field that you want to edit. Configure the following attributes:

Data type attribute	Description
Retain existing fields at runtime	When enabled, the task uses the field metadata that is configured in the task. If field metadata changes after a task is saved, Data Integration uses the updated field metadata. Typically, this is the desired behavior. However, if the task uses a flat file connection and you want to retain the metadata used at design time, enable this option.
Type	The data type of the field.
Precision	Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.
Scale	Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a numeric data type is 65535. Not editable for all data types.

7. Optionally, configure advanced options for the selected connection.

   The options available depend on the connection.
   When you edit advanced options, you can restore default values for individual properties or for all advanced properties. Default values are defined in the mapping that the task is based on.

   For information about a particular connector's properties, see the help for the appropriate connector.

8. Optionally, configure partitions.

   The partitioning methods that you can select depend on the connection.

   If you defined partitions in the mapping, the partition details are displayed in the **Partitions** area. You can edit these details to override the partitions configured in the mapping at run time.

   You can't configure pass-through partitioning in a mapping task.

9. Click **Next**.

Configuring targets

The **Targets** page displays differently depending on the basis for the task.

The **Targets** page displays when the mapping includes parameters for target connections or target objects. The properties that you need to specify are based on the type of parameter.

1. For each target parameter, configure the following details:

Target parameter detail	Description
Connection	Select a connection. To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**. **Tip:** Hover over the target connection tool tip to view more information about the connection.
Object	Select a target object. If a list of objects does not appear, click **Select**. The **Select Target Object** dialog box displays up to 200 objects. If the object you want to use does not display, enter a search string to reduce the number of objects that display. **Tip:** Hover over the target object tool tip to view more information about the object.
Operation	Target operation, either insert, update, upsert, delete, or data driven. Not editable for all connection types.
Truncate target	Truncates the target object before inserting new rows. Applies to insert and data driven operations. Not available for all connection types.
Enable target bulk load	Uses the database bulk API to perform an insert operation. Use the bulk API to write large amounts of data to the database with a minimal number of API calls. Loading in bulk mode can improve performance, but it limits the ability to recover because no database logging occurs. Applies to insert operations. Not available for all connection types.
Display technical names instead of labels	Displays technical names instead of business names. Not available for all connection types.

2. Configure formatting options if required.

   The properties available vary based on the connection type. For information about a particular connector's properties, see the help for the appropriate connector.

3. Optionally, expand the **Data Preview** area to preview the data.

   To display fields in alphabetical order, click **Display mapplet fields in alphabetical order**. By default, fields appear in the order returned by the target system.
   To download the preview results, click **Download**.

4. Configure field metadata if required.

You can configure field metadata for targets with certain connection types. To see if a connector supports field metadata configuration, see the help for the appropriate connector. You can't update field metadata in mapping tasks that are based on mappings in SQL ELT mode.

To edit field metadata, expand the **Field Data Types** area and select the row that contains the field that you want to edit. Configure the following attributes:

Data type attribute	Description
Retain existing fields at runtime	When enabled, the task uses the field metadata that is configured in the task. If field metadata changes after a task is saved, Data Integration uses the updated field metadata. Typically, this is the desired behavior. However, if the task uses a flat file connection and you want to retain the metadata used at design time, enable this option.
Type	The data type of the field.
Precision	Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.
Scale	Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a numeric data type is 65535. Not editable for all data types.

5. Optionally, configure advanced options for the selected connection.

   The options available depend on the connection.
   When you edit advanced options, you can restore default values for individual properties or for all advanced properties. Default values are defined in the mapping that the task is based on.

   For information about a particular connector's properties, see the help for the appropriate connector.
6. Click **Next**.

Configuring CDC runtime properties

The **CDC Runtime** page displays if the mapping processes change data from a CDC source.

1. For each property, edit or reset the current value.
2. Click **Next**.

Configuring input parameters

The **Input Parameters** page displays differently depending on the basis for the task. If the mapping does not contain input parameters, the page doesn't display.

The **Input Parameters** page displays input parameters such as filter and lookup conditions, field mappings, and expressions. It also displays stored procedure connections. It doesn't display source, target, or lookup connections and objects.

1. Configure the parameters that display.

   Depending on the mapping data flow, you might need to configure some parameters before you configure other parameters. For more information, see *Mappings*.

   Warning: If the mapping contains a filter parameter, the task might display as valid even if you haven't configured the filter value. You must configure all parameters before the task is valid and you can run it.
2. If necessary, configure stored procedure details.

   This displays when a stored procedure requires a connection. Select a connection or optionally create a new one.
3. If necessary, configure lookup conditions.

   This displays when the lookup condition is parameterized.
4. If necessary, configure mapplet details.

   This displays when a mapplet requires a connection and the mapplet is configured to display on this page.

   For some connection types, you can select **Display technical field names instead of labels** to display technical names instead of business names.

   To display fields in alphabetical order, click **Display mapplet fields in alphabetical order**.
5. If necessary, configure field mappings.

   This displays when the field mapping is parameterized. Use the **Object** list to display fields from different objects.
   Map fields in one of the following ways:

   - To automatically map fields with the same name, click **Automap** > **Exact Field Name**. Or, to map fields with similar names, click **Automap** > **Smart Map**.
     You can undo all automapped field mappings by clicking **Smart Map** > **Undo Automap**. To unmap a single field, select the field to unmap and click **Actions** > **Unmap**.
   - To manually map fields, select and drag the source fields to the applicable target fields.
6. Click **Next**.

Configuring persistent values

The **Persistent Values** page displays if the mapping contains persisted values such as in-out parameters or sequences.

1. For each sequence value, edit or reset the current value.

   Note: If the Sequence transformation is configured to reset the value for each run, editing the current value in the mapping task has no effect.
2. Configure in-out parameters as necessary.

   In-out parameters don't apply to mapping tasks that are based on mappings in SQL ELT mode.
3. Click **Next**.

Configuring runtime options

On the **Runtime Options** page, configure optional runtime options for the task. Expand each area to see the options you can configure.

1. Specify whether to run the task on a schedule or without a schedule.

   Choose one of the following options:

   - If you want to run the task on a schedule, click **Run on a schedule**. Select the schedule you want to use or click **New** to create a schedule.
   - If you want to run the task without a schedule, click **Do not run on a schedule**.

2. Configure email notification options for the task.
3. Optionally, if the mapping contains a Source transformation that incrementally loads files, you can configure the time that the task uses to identify files to load. By default, the task loads files modified after the last job. To load files from a different time, choose one of the following override options:

   - **Full load of all files** to load all of the files in the source directory.

     Note: After you select this option, the mapping task might display the wrong date and time for the last load time, but the mapping task does perform a full load.

   - **Load files from selected time** to load the files modified after the date and time that you select.

   Note: If you want to run one job that reprocesses source files without affecting future jobs, run the mapping task with the advanced options configured to reprocess incrementally-loaded source files. For more information about reprocessing jobs, see “Reprocessing incrementally-loaded source files ” on page 66.

4. Optionally, in the **Advanced Options** area, configure the advanced options that are displayed for your connection:

Field	Description
Pre-Processing Commands	Commands to run before the task.
Post-Processing Commands	Commands to run after the task completes.
Maximum Number of Log Files	Number of session log files and import log files to retain. By default, Data Integration stores each type of log file for 10 runs before it overwrites the log files for new runs. **Note:** If a dollar sign ($) is present in a custom session log file name, for example, MyLog_ $CurrentTime, the file name is dynamic. If you customize the session log file name using a dynamic name, this property doesn't apply. To purge old log files, delete the files manually.
Schema Change Handling	Determines how Data Integration picks up changes to the object schema. Select one of the following options: - Asynchronous. Data Integration refreshes the schema when you update the mapping or mapping task, and after an upgrade. - Dynamic. Data Integration refreshes the schema every time the task runs. Default is Asynchronous.

Field	Description
Schema Mismatch Handling	Determines how Data Integration responds when records in a file don't conform with the schema. This setting is available when schema validation is enabled in the Source transformation. Select one of the following options: - Skip mismatched files and continue. Data Integration skips the entire file if at least one record in the file doesn't conform with the schema. Processing continues for all the remaining files. - Stop on first mismatched file. Data Integration stops processing when it encounters a file with at least one record that doesn't conform with the schema. Default is "Skip mismatched files and continue." For more information, see "Schema mismatch handling" on page 50. Schema mismatch handling doesn't apply to mapping tasks that are based on mappings in SQL ELT mode.
Dynamic Schema Handling	Determines how Data Integration applies schema changes from upstream transformations to the target object. Available when the schema change handling is dynamic and the field mapping is automatic. For each target, select how Data Integration updates the target schema. The options available are based on the target connection. For more information, see "Schema change handling" on page 48 or the help for the appropriate connector.

Not all options appear if the mapping task is based on a mapping in advanced mode.

5. Optionally, to create a parameter file based on the parameters and default values specified in the mapping on which the task is based, click **Download Parameter File Template** within the parameter file location.

   For more information about parameter file templates, see *Mappings*.

6. Optionally, if the mapping task contains parameters, you can use parameter values from a parameter file. Choose one of the following options:

- To use a parameter file on a local machine, select **Local**. Use this option if the task is based on a mapping in advanced mode. Enter the following information:

Field	Description

Parameter File Directory	**In a Secure Agent runtime environment:**
	Path for the directory that contains the parameter file, excluding the parameter file name. The Secure Agent must be able to access the directory.
	You can use an absolute file path or a path relative to one of the following $PM system variables:
	- $PMRootDir
	- $PMTargetFileDir
	- $PMSourceFileDir
	- $PMLookupFileDir
	- $PMCacheDir
	- $PMSessionLogDir
	- $PMExtProcDir
	- $PMTempDir
	By default, Data Integration uses the following parameter file directory:
	`<Secure Agent installation directory>/apps/Data_Integration_Server/data/userparameters`
	For mappings in advanced mode, Data Integration uses the following parameter file directory by default:
	`<Secure Agent installation directory>/apps/data/userparameters`
	In a serverless runtime environment:
	You can only use mounted directories configured in the data disk or their subdirectories for the following $PM system variables:
	- $PMLookupFileDir
	- $PMBadFileDir
	- $PMCacheDir
	- $PMStorageDir
	- $PMTargetFileDir
	- $PMSourceFileDir
	- $PMExtProcDir
	- $PMTempDir
	For more information about these system variables, see *Runtime Environments* in the Administrator help.

Parameter File Name	Name of the file that contains the definitions and values of user-defined parameters used in the task. You can provide the file name or the relative path and file name in this field. For the CDC and select mainframe and midrange connectors, you can specify a parameter file that includes connection overrides. In the parameter file, set the connection overrides for the parameters in the format of name=value pairs, using a semicolon (;) as the separator. For example: `$<ParameterName>=”User Name=jdoe;Password=mypassword”`

- To use a cloud-hosted file, select **Cloud Hosted**. Use this option if the task runs in a serverless runtime environment. Enter the following information about the file:

Field	Description
Connection	Connection where the parameter file is stored. You can use the following connection types: - Amazon S3 - Google Storage V2 - Azure Data Lake Store Gen2
Object	Name of the file that contains the definitions and values of user-defined parameters used in the task.

7. If the parameter file is stored on a local machine and the task runs a mapping in advanced mode, you can click **Download Parameter File** to preview the parameter file that the task will use.

   In a serverless runtime environment, the parameter file must be on a data disk.

8. Choose whether to run the task in standard or verbose execution mode, if this option is visible.

   If you select verbose mode, the mapping generates additional data in the logs that you can use for troubleshooting. Select verbose execution mode only for troubleshooting purposes. Verbose execution mode impacts performance because of the amount of data it generates.

   This option does not appear if the mapping task is based on a mapping in advanced mode.

9. Optionally, configure SQL ELT optimization.

   If you use the runtime options to configure SQL ELT optimization for a mapping task that is based on a mapping in advanced mode, you can't enable CLAIRE-powered runtime strategies on the **Runtime Strategy** page.

The following table describes the SQL ELT optimization properties:

Property	Description
SQL ELT Optimization Type	Type of SQL ELT optimization. Use one of the following options: - None. The task processes all transformation logic for the task. - To Source. The task pushes as much of the transformation logic to the source database as possible. Not available in advanced mode. - To Target. The task pushes as much of the transformation logic to the target database as possible. Not available in advanced mode. - Full. The task pushes all of the transformation logic to the source and target databases. The source and target schema must be the same. - $$PushdownConfig. The task uses the SQL ELT optimization type specified in the user-defined parameter file for the task. When you use $$PushdownConfig, ensure that the user-defined parameter is configured in the parameter file. When you use SQL ELT optimization, do not use the **Error Log Type** advanced session property. **Note:** SQL ELT optimization functionality varies depending on the support available for the connector. For more information, see the help for the appropriate connector. SQL ELT optimization doesn't apply to mapping tasks that are based on mappings in SQL ELT mode because mappings in SQL ELT mode are automatically configured to push transformation logic to the cloud data warehouse.
Optimization Context Type	Provides context about the mapping configuration for SQL ELT optimization. If you select an option other than None, Data Integration constructs a single query for SQL ELT optimization by combining multiple targets in the mapping based on the target configurations. If you select None, the query is not optimized. If Data Integration cannot apply the selected context, Data Integration uses the default SQL ELT optimization behavior. Select one of the following options: - None - SCD Type 2 merge - Multi-insert Not available in advanced mode. For more information, see the help for the appropriate connector.
SQL ELT Optimization Fallback Option	If full SQL ELT optimization is not available for the connection, choose how Data Integration handles SQL ELT optimization. Choose one of the following options: - Partial SQL ELT. Default. Data Integration pushes as much transformation logic as possible to the source and target database. The task processes any transformation logic that it can't push to a database. - Non SQL ELT. The task runs with no SQL ELT optimization. - Fail Task. Data Integration fails the task. Not available in advanced mode.

Property	Description
Create Temporary View	Allows the task to create temporary view objects in the database when it pushes the task to the database. Use when the task includes an SQL override in the Source Qualifier transformation or Lookup transformation. Disabled when the SQL ELT optimization type is None.
Create Temporary Sequence	Allows the task to create temporary sequence objects in the database. Disabled when the SQL ELT optimization type is None.

10. Optionally, if the task runs in a serverless runtime environment, configure serverless usage properties.
11. Optionally, for mappings in advanced mode, run "CLAIRE Tuning" on page 63 to get a set of recommended properties and values for Spark session properties.
12. If necessary, add other session properties now under **Advanced Session Properties**:
    a. Click **New Session Property**.
    b. Select a session property from the drop-down list.
    c. Configure the session property value.
13. Choose to enable cross-schema SQL ELT optimization.

    Cross-schema SQL ELT optimization doesn't apply to mapping tasks that are based on mappings in SQL ELT mode.
14. If you want to run multiple instances of the task at the same time, enable simultaneous runs of the mapping task.

    Some mapping features might produce unexpected results in simultaneous task runs.

Configuring the runtime strategy

On the **Runtime Strategy** page, you can get a recommendation for a runtime strategy that optimizes cost or performance.

To get a recommendation for a runtime strategy, the mapping task must be based on a mapping in advanced mode and CLAIRE recommendations must be enabled in your organization. If you select a runtime strategy other than the default, you can't configure SQL ELT optimization on the **Runtime Options** page.

Note: Runtime strategy recommendations and the target duration confidence level might not appear for some mapping tasks. CLAIRE continues collecting metadata in the background and shows these details when it has enough confidence.

1. Select one of the following runtime strategies:
   - Advanced cluster. Schedules the workload for distributed processing on an advanced cluster.
   - SQL ELT optimization. Converts the transformation logic to an SQL query and sends the query to the data warehouse.
   - Default runtime strategy. Runs the mapping task on an advanced cluster unless SQL ELT optimization is configured on the **Runtime Options** page.
2. Review the target duration confidence level.

   The target duration confidence level shows the likelihood that the job will run within the target duration that you configured on the **General** page.
3. Click **Save**.

CLAIRE Tuning

Use CLAIRE Tuning to tune a mapping task that runs on an advanced cluster.

CLAIRE, Informatica's AI engine, runs the mapping task several times and uses machine learning to assess the performance of each run. It uses the information to create a tuning recommendation for the set of Spark properties that optimizes task performance. CLAIRE Tuning considers parameters such as the complexity of the mapping, the size of the data, and the processing capacity on the advanced cluster.

You can run initial tuning or enable continuous tuning. When you run initial tuning, you can view the tuning recommendation to see a list of recommended Spark properties and their values. You can apply the recommendation to use the values in the mapping task. When you enable continuous tuning, CLAIRE silently monitors the mapping task and adjusts the Spark properties over time.

Continuous tuning is more effective if you run initial tuning first. During initial tuning, CLAIRE gets an optimized set of Spark properties that it can use as a baseline to make additional adjustments during continuous tuning.

If you run initial tuning on a mapping task that incrementally loads files, tuning runs on all of the source files. The recommended properties and values might not be optimal for future jobs that load and process only modified files.

Guidelines to get an accurate recommendation

Use the following guidelines to get an accurate recommendation during the tuning job:

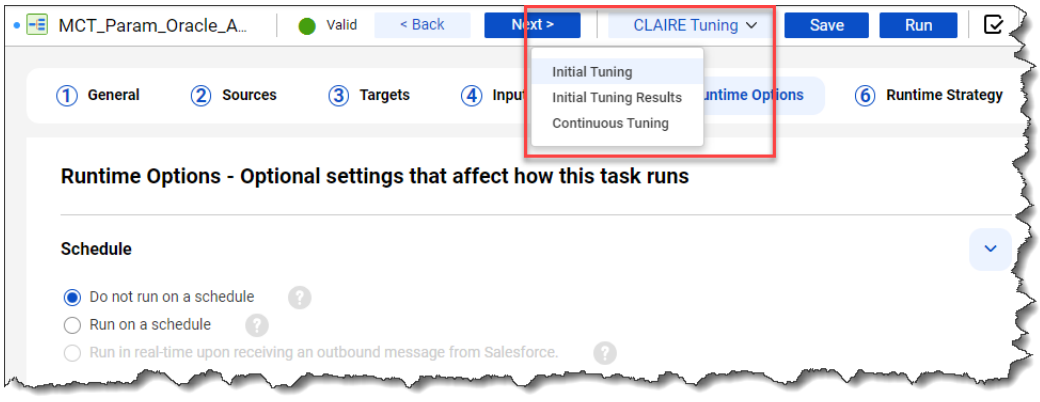
- Use sample data that closely matches the actual volume of the data that the mapping task will process.
- Make sure that the mapping logic handles duplicate data in the target. The tuning job will write data to the target multiple times.
- Set resource limits on your cloud environment by configuring the appropriate Spark properties before you tune the mapping task. Your cloud service provider charges you for the resources that each run uses.

  For example, if you know that you can allocate only 4 GB to the Spark driver, you can configure `spark.driver.memory=4G` in the mapping task. CLAIRE will honor the pre-defined Spark property to create a tuning recommendation for other Spark properties.

Configuring tuning

Configure CLAIRE Tuning in the mapping task details.

The following image shows where you can configure tuning in the mapping task details:



Initial tuning

Run initial tuning to get a tuning recommendation with a list of recommended Spark properties and their values.

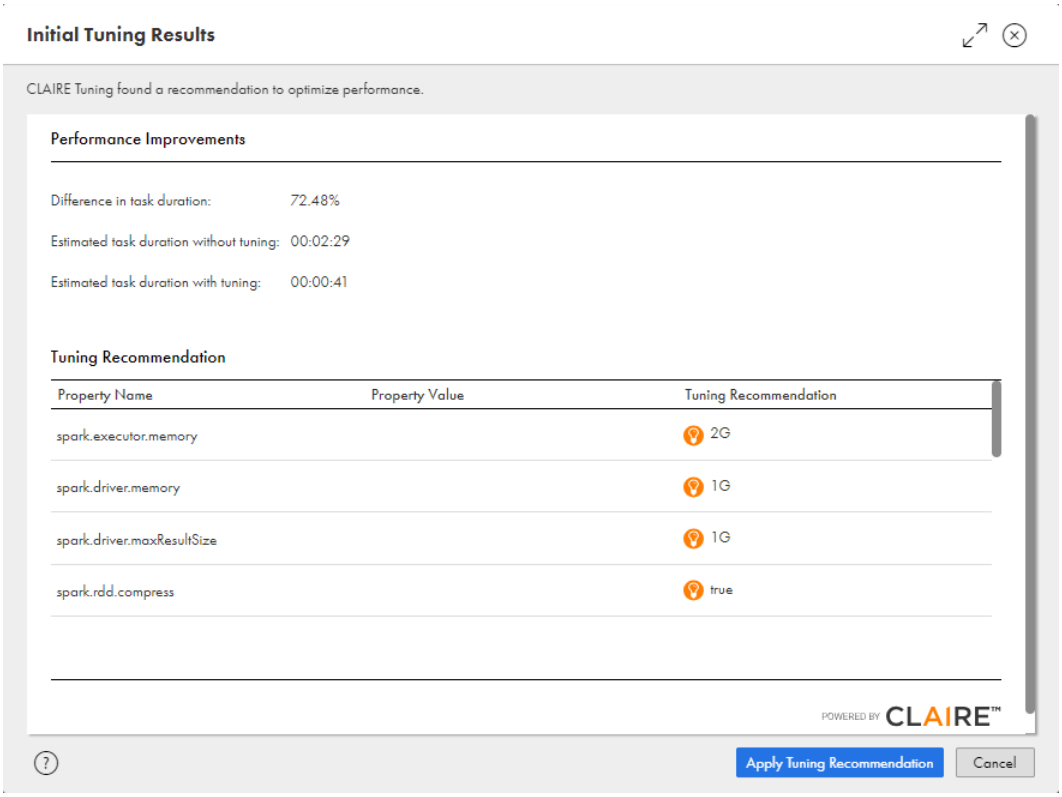
To configure initial tuning, set the number of times that CLAIRE runs the mapping task, with **10** being the minimum. Click **Tune** to begin tuning. When tuning begins, Data Integration creates a tuning job with multiple subtasks to represent each run of the mapping task. You must wait for all subtasks to complete before you can view the tuning results.

Each time that CLAIRE runs the mapping task, CLAIRE gathers task performance data to improve its recommendation for an optimal set of Spark properties.

Initial tuning results

When initial tuning is complete, you can view the tuning recommendation and the performance improvement. The improvement is measured in the amount of time that it takes for the mapping task to run using the recommended set of Spark properties.

The following image shows the tuning results for a particular mapping task:



You can apply the recommendation to use the Spark property values in the mapping task. You can also revert the Spark properties to their original values and apply the recommendation again.

Guidelines to apply a tuning recommendation

Use the following guidelines when you apply a tuning recommendation to make sure that job performance is optimal:

- Use the full set of Spark properties to achieve the performance improvement. Using a partial set of the recommended Spark properties might not be optimal.
- Do not edit the Spark properties in the mapping task in between the time that you begin tuning and the time that you apply the tuning recommendation. If you make significant changes to the Spark properties, tune the mapping task again.

Continuous tuning

Enable continuous tuning to silently monitor every run of the mapping task and adjust the Spark properties over time.

For example, you design a mapping task in your development environment and run initial tuning. When you migrate the mapping task to your production environment, you expect production loads to vary day-by-day. Continuous tuning analyzes the varying parameters to adjust the Spark properties.

During continuous tuning, CLAIRE analyzes all runs of the mapping task. The adjusted Spark properties override the Spark property values that are set in the mapping task. You can view the values of the adjusted Spark properties in the Spark driver and agent job logs.

Note: When you copy or import a mapping task with continuous tuning enabled, continuous tuning restarts from the Spark properties that are set in the mapping task.

Viewing and editing mapping task details

You can view details about a mapping task, such as the mapping used by the task.

The **Task Details** page includes the following information:

- The runtime environment used to run the mapping task.
- The date the task was created and the user who created the task.
- The last time the task was updated and the user who updated the task.
- The date of the last run.
- The name and image of the mapping on which the task is based and the date the mapping was last updated.
- Pre-processing and post-processing commands.

To view details for a mapping task, perform the following steps:

1. On the **Explore** page, navigate to the task.
2. In the row that contains the task, click **Actions** and select **View**.

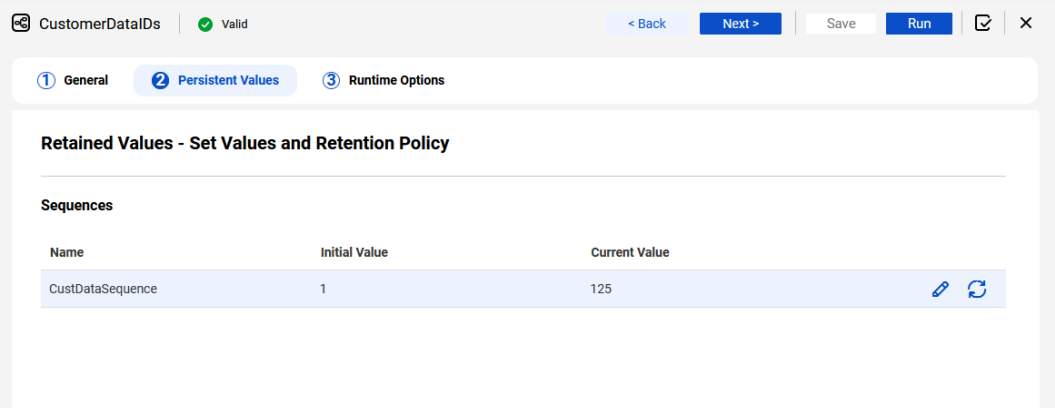
   On the **Task Details** page, you can click **Edit** to modify the mapping task.

Sequence values

When you run a mapping task that includes a Sequence transformation in the mapping, you can change the beginning value for the sequence.

To change the beginning value, you change the **Current Value** field in the **Sequences** or **Persistent Values** page in the mapping task wizard. The **Current Value** field shows the first value the task will generate in the sequence, based on the last value generated in the last task execution.

For example, the last time you ran the CustDataIDs task, the last value generated was 124. The next time the task runs, the first number in the sequence is 125 because the Sequence transformation is configured to increment by 1. If you want the sequence to begin with 200, you change the **Current Value** to 200.



Note: If the Sequence transformation is configured to reset the value for each run, editing the current value in the mapping task has no effect.

Running a mapping task

Run a mapping task after you configure all of the required settings.

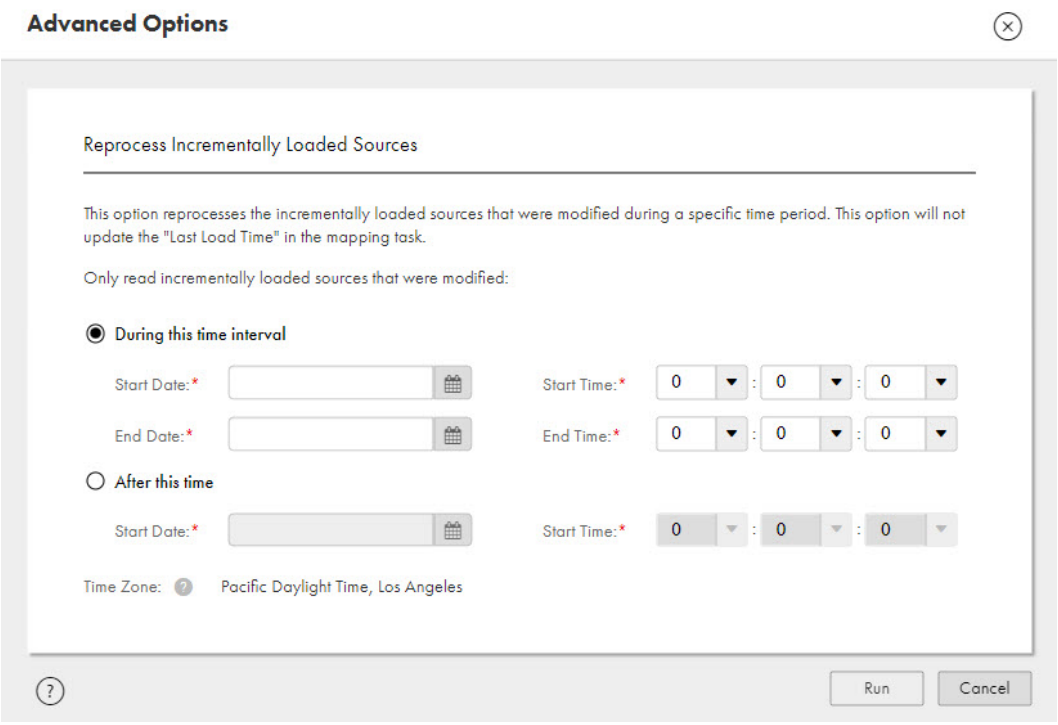
You can run a mapping task in the following ways:

- Manually. To run a mapping task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  You can also run a mapping task manually, with or without advanced options, from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**.
- On a schedule. To run a mapping task on a schedule, edit the task in the mapping task wizard to associate the task with a schedule.

Reprocessing incrementally-loaded source files

Reprocess source files to create a snapshot of the data from a specified time interval, to debug and discover the source of bad data in your target, or to restore deleted data. In the advanced options, you can configure the properties to reprocess incrementally-loaded source files.

A job that reprocesses incrementally-loaded files reads the files that were modified during the time that you specify. You can choose to reprocess files that were modified in a given time interval or all files modified after a given start time. Reprocessing applies only to Source transformations that you configure to incrementally-load files. The following image shows the reprocessing settings in **Advanced Options**:



You can reprocess source files when at least one source is configured to incrementally load files and the files have been loaded at least once. The advanced options are not available if the mapping does not have sources configured to incrementally load files, the mapping task has never run, or the mapping task is configured to perform a full load of all files.

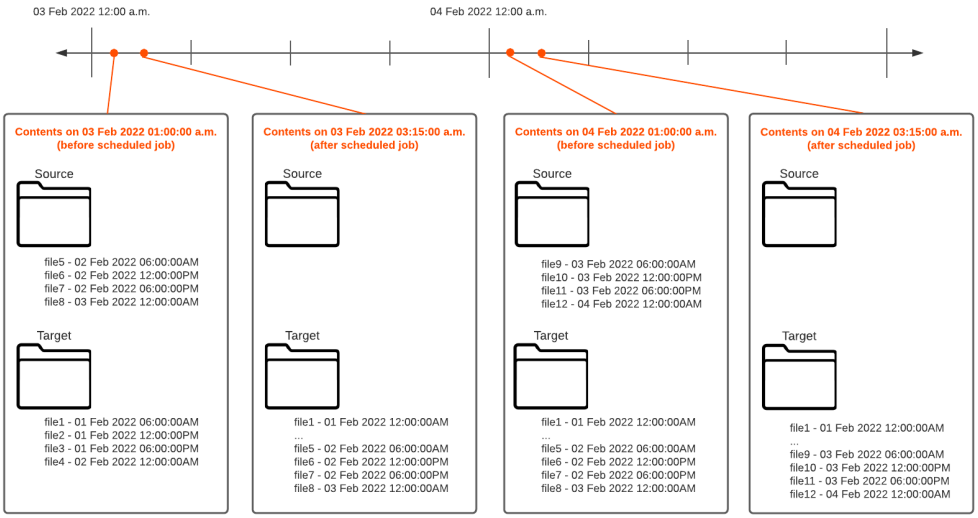
Reprocessing doesn't change the time that the mapping task uses in future jobs to identify modified files for incrementally-loaded sources. For example, you have a mapping task that incrementally loads files and is scheduled to run every Monday at 8 a.m. After you run a job with advanced options on a Thursday, the next scheduled job still checks for files modified after Monday at 8 a.m. If you want to change the time that the mapping task uses to identify modified files, override the incremental file load type in the mapping task configuration.

Note: Processing is unpredictable when the start or end time of a reprocessing job is within the repeated hour on the day that Daylight Saving Time ends.

Reprocessing incrementally-loaded source files example

You have a source directory that receives a new data file every six hours. In Data Integration, you have a mapping that incrementally loads these files to a target and a mapping task that is scheduled to run the mapping at 3:00 a.m. every day.

The following image shows the contents of the source and target directories over two days:



Mid-day on February 4, you realize that incorrect data was loaded the previous day, February 3, so your target contains incorrect data for the files created on February 2. After fixing the files in the source directory, you run a job with advanced options to reprocess the incorrect files.

In the advanced options, you choose to read incrementally-loaded source files that were modified between 02/02/2022 00:00:01 and 02/03/2022 00:00:00. You click **Run** to run the mapping task with the advanced options. The reprocessing job processes the files modified within the selected time interval and writes them to the target.

When the next scheduled job runs at 3 a.m. on February 5, it incrementally loads the files created on February 4, as usual.

CHAPTER 3

Dynamic mapping tasks

Use a dynamic mapping task to create and batch multiple jobs based on the same mapping.

A dynamic mapping task reduces the number of assets that you need to manage if you want to reuse a parameterized mapping. Instead of creating multiple mapping tasks, you can configure multiple jobs based on the same mapping in one task. Each job can have a different set of parameter values. You can also organize jobs in groups to batch jobs together and set the job run order.

When you create a dynamic mapping task, you select the mapping to use. The mapping that you select must contain at least one parameter. Do not use a mapping that contains field mapping parameters for hierarchical data. When you configure the task, you configure the value and scope of each parameter for each job.

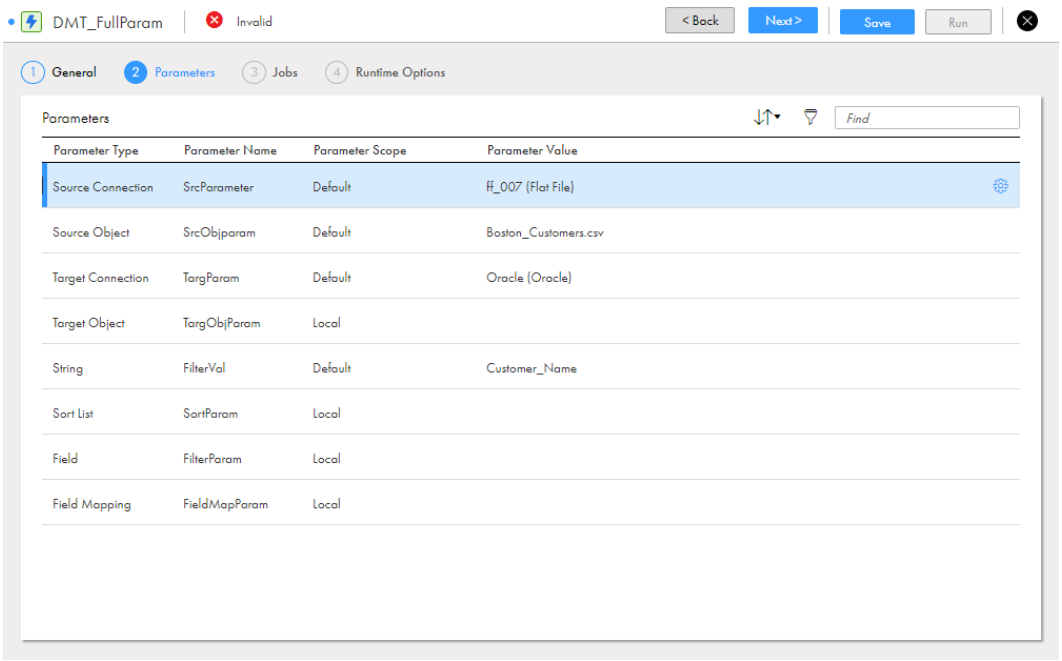
You cannot use CLAIRE tuning with a dynamic mapping task.

Parameters in dynamic mapping tasks

The **Parameters** page lists the parameters that are defined in the mapping that the dynamic mapping task is based on. You can configure the default value, and settings for each parameter on this page.

By default, Data Integration assigns local scope to each parameter and you configure parameters for each job. If you want to apply a default parameter value and settings to each job in the dynamic mapping task, configure default values and settings on the **Parameters** page.

The following image shows the **Parameters** page:



To configure a default parameter, click the row that contains the parameter and select **Default** scope. If the parameter has a default value assigned in the mapping, Data Integration lists the value in the **Parameter Value** column. You can override the default value or assign a different value to each job in the task. When you update the default parameter value on the **Parameters** page, Data Integration updates the value in jobs with the default parameter value. Data Integration does not update jobs with local scope or jobs where the default value was overridden.

You can sort the parameters on the page by parameter type or name. You can also filter the parameters on the page by parameter name, type, and scope.

Some parameter values depend on other parameters. For example, if a mapping contains a target connection and a target object parameter, you must configure the target connection parameter value before you configure the target object parameter value. Configure parameter values from the top down.

When you configure a source or lookup object with a database connection, you can select a single object or enter a custom query if the task is if the task is not based on a mapping in advanced mode. For more information about configuring data objects and custom queries, see *Transformations*.

Data Integration does not apply filter and sort parameter values unless the source is parameterized.

You cannot override parameters in a dynamic mapping task with values in a parameter file.

Parameter scope

The parameter scope determines how Data Integration applies the parameter values to the jobs in a dynamic mapping task.

The following table describes the parameter scope options:

Scope	Description
Default	Data Integration copies the value to the parameter in every job. You can override the default value for an individual job when you configure the job.
Local	You assign the value of the parameter in each job.

By default, Data Integration applies local scope to every parameter. When you select default scope, you must provide a value in the **Parameter Value** column.

The scope of in-out parameters and sequence parameters is always local.

Parameter settings

You can configure settings for parameters in a dynamic mapping task such as formatting options, advanced attributes, or field metadata.

The settings that you can configure vary based on the type of parameter and the parameter value. For example, the settings that you can configure are different for object and connection parameters.

If the parameter is a data object, you can preview data for the selected object.

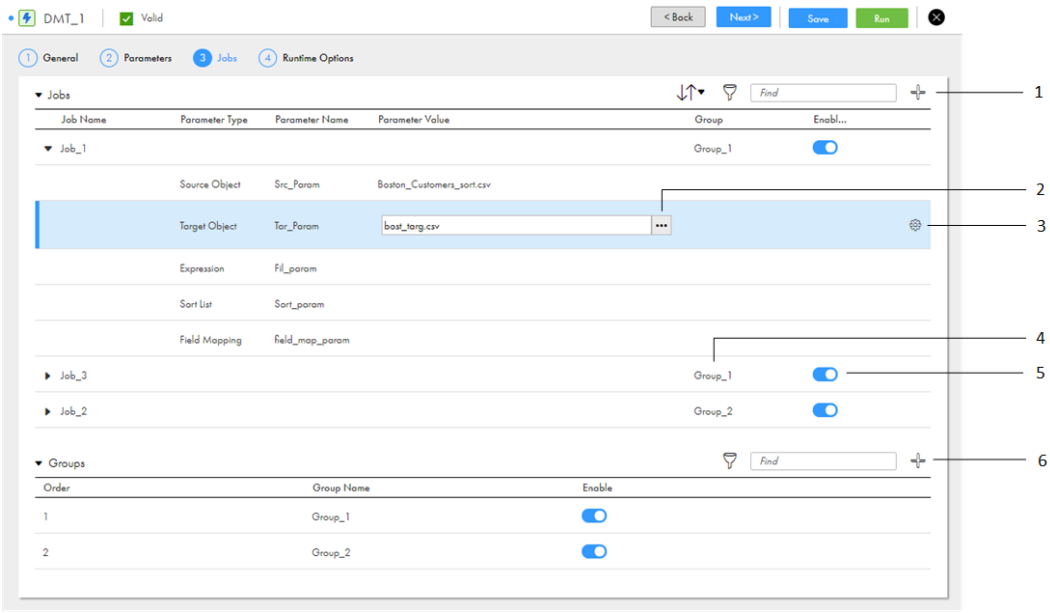
You can configure settings for parameters with default scope on the **Parameters** page. You configure settings for parameters with local scope within each job on the **Jobs** page.

Jobs and job groups

A dynamic mapping task can include multiple groups of jobs. A job is a single run of the mapping that the task is based on. Add jobs to a dynamic mapping task and organize them into groups on the **Jobs** page. You can configure a different set of parameter values for each job.

Assign each job to a group. A group is a set of jobs that run concurrently. Groups run sequentially, and all jobs in the previous group complete before the next group begins. You might set up groups when data in one group depends on the results of a previous group. There is no limit on the number of jobs and groups that you can add to a dynamic mapping task.

The following image shows the **Jobs** page:



1. Add a job
2. Select parameter value
3. Parameter settings
4. Add a job to a group
5. Enable/Disable job
6. Add a group

When you add a job, Data Integration displays each parameter in the mapping. If a parameter has default scope, Data Integration automatically applies the default value to the job. You can edit the parameter value and parameter settings for each job.

By default, Data Integration enables a job or group when you add it. You can disable individual jobs so that they won't run when you run the task. If you disable a group, no jobs in that group run when you run the task.

Job settings

You can configure settings for each job including stopping the job on an error or warning, advanced session properties, and pre and post processing commands. Configure job settings in the **Settings** window.

When you set the **Stop on** property for a job that is in a group, if the job encounters the configured state, the jobs in the group complete, and then the task stops.

To configure job settings, click **Settings** in the row that contains the job.

You can optionally configure the max compute units and task timeout serverless usage properties for a dynamic mapping task job in a serverless environment. Maximum compute units and task timeout settings are configured for serverless environments in Administrator but you can revise their values for a particular dynamic mapping task job.

You can use the max compute units property to increase or decrease the number of compute units corresponding to the machine resources that a job can use based on the Max Compute Units Per Task serverless environment property value. You might want to specify a smaller value for a simple job, or a larger value for a complex job. The Max Compute Units property value for a dynamic mapping task must be less than or equal to the Max Compute Units Per Task value in the serverless environment.

You can use the Task Timeout property to increase or decrease the amount of time in minutes to wait for a job to complete before it is terminated based on the Task Timeout serverless environment property value.

Schema change handling

Dynamic mapping tasks automatically use dynamic schema change handling. This means that Data Integration refreshes the schema of applicable source, target, and lookup objects every time you run the task.

A schema change includes one or more of the following changes:

- Fields are added.
- Fields are deleted.
- Fields are renamed.
- Field data type, precision, or scale is updated.

Data Integration updates the schema for data objects with connectors which support dynamic schema handling. A connector that supports dynamic schema handling in mapping tasks supports this functionality in dynamic mapping tasks. You can configure the applicable dynamic schema handling options similar to in a mapping task. To see if a connector supports dynamic schema change handling, see the help for the appropriate connector. Data Integration does not update the schema for hierarchical data objects.

If you update fields in the source object, be sure to update the Target transformation field mapping. Data Integration writes Null to the target fields that were previously mapped to the renamed or deleted source fields.

Schema change handling options

You can choose how Data Integration applies dynamic schema changes from upstream transformations to the target. If the mapping contains more than one target, select the schema change handling for each target.

Configure schema change handling options in the target object settings. Schema change handling options configured within a job override default schema options.

To select schema options, the target field mapping in the referenced mapping must be automatic.

The schema change handling options available are based on the target type. The following table describes the options that you can select for each target type:

Schema handling option	Target type	Description
Keep Existing File Format	File	Data Integration fetches the most recent target schema at runtime and does not apply upstream schema changes to the target file.
Drop Current and Recreate	Database and file	For database targets, Data Integration drops the existing target table and creates a new target table with the schema from the upstream transformations on every run. For file targets, Data Integration updates the target schema to match the incoming schema on every task run.
Alter and Apply Changes	Database	Data Integration updates the target schema with additive changes to match the schema from the upstream transformations. It does not delete columns from the target.
Don't Apply DDL Changes	Database	Data Integration fetches the target schema at runtime and does not apply upstream schema changes to the target table.

Data Integration does not pass field constraints to the target. For example, the source contains fields S1 and S2 configured with the NOT NULL constraint. The target contains fields T1 and T2 also configured with the

NOT NULL constraint. You select the Alter and Apply Changes schema handling option. When you run the task, fields S1 and S2 are written to the target with no constraints.

Configuring a dynamic mapping task

Use the dynamic mapping task wizard to create a dynamic mapping task.

To configure a dynamic mapping task, complete the following steps:

1. Create the dynamic mapping task.
2. Configure parameters.
3. Configure jobs.
4. Configure runtime options.

As you work through the task wizard, you can click **Save** to save your work at any time.

Defining a dynamic mapping task

1. To create a dynamic mapping task, click **New** > **Tasks**. Select **Dynamic Mapping Task** and click **Create**.

   To edit a dynamic mapping task, on the **Explore** page, navigate to the dynamic mapping task. In the row that contains the task, click **Actions** and select **Edit**.

2. In the **General Properties** area, configure the following properties:

Property	Description
Name	Name of the dynamic mapping task.
Location	Project folder in which the task resides. If the Explore page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.
Description	Description of the task.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task.
Mapping	Mapping associated with the task.

3. Click **Next**.

Configuring default parameters

The **Parameters** page lists the parameters in the mapping. Configure the default values and settings for each parameter on the **Parameters** page or configure local parameter values and settings when you configure the job.

Some parameter values depend on other parameters. For example, if a mapping contains a target connection and a target object parameter, you must configure the target connection parameter value before you configure the target object parameter value. Configure parameter values from the top down.

1. Click the row that contains the parameter. Select default scope.
2. For parameters with default scope, select or enter the value of the parameter.

   You can also override this value on the **Jobs** tab.
3. Click **Settings** to configure advanced attributes, formatting options, or field metadata.

   The available options change based on the parameter you're configuring.
4. Click **Next**.

Configuring jobs

Configure jobs and groups on the **Jobs** page.

1. In the **Jobs** area, click **Add**.
2. To rename the job, click the default name and enter another name.

   The default job name is Job_X where X is the sequential job number.
3. Expand the job and configure local parameters. You can also override default parameter values.
4. Select a group to assign the job to.

   Configure groups in the **Groups** area.
5. Optionally, click **Settings** and configure the job properties.

   You can configure the following properties:

Property	Description
Stop on	Data Integration stops the task if the job encounters an error or warning.
Pre-processing Commands	Commands to run before the task.
Post-processing Commands	Commands to run after the task.

6. Optionally, configure advanced session properties.
7. Optionally, configure the serverless usage properties for a dynamic mapping task job.

You can configure the following properties:

Property	Description
Max Compute Units	Increases or decreases the number of compute units corresponding to the machine resources that a job can use.
Task Timeout (in minutes)	Increases or decreases the amount of time in minutes to wait for a job to complete before it is terminated.

For more information about the Max Compute Units and Task Timeout (in minutes) properties, see "Job settings" on page 72.

Configuring groups

Create groups to batch jobs together.

1. In the **Groups** area, click **Add**.
2. To rename a group, click the default name and enter the new name.

   The default group name is Group_X, where X is the sequential group number.
3. To adjust the order in which groups run, click the up or down arrow in the row that contains the group that you want to move.
4. Click **Next**.

Configuring runtime options

Configure runtime options for the dynamic mapping task on the **Runtime Options** page.

1. Select to run the task manually or on a schedule.

   You must create the schedule in Administrator before you configure the task.
2. Configure email notification options. Select one of the following options:
   - Use the default email notification options for my organization.
   - Use custom email notification options. Enter the email addresses to use for job failure, warning, and success messages. Enter one or more valid email addresses. Separate email addresses with a comma (,) or semicolon (;).
3. Click **Save**.

CHAPTER 4

Synchronization tasks

Use the synchronization task to synchronize data between a source and a target. For example, you can read sales leads from your sales database and write them into Salesforce. You can also use expressions to transform the data according to your business logic or use data filters to filter data before writing it to targets.

You can use the following source and target types in synchronization tasks:

- Database
- Flat file
- Salesforce

Task operations

When you configure a synchronization task, you specify the task operation and the type of target. The available target types depend on the task operation that you select.

You can use one of the following task operations:

Insert

When you run a task with the Insert task operation, Data Integration inserts all source rows into the target. If Data Integration finds a source row that exists in the target, the row fails.

If you write data to a flat file target, Data Integration truncates the flat file before it inserts the source rows into the file.

Update

When you run a task with the Update task operation, Data Integration updates rows in the target that exist in the source. If Data Integration finds a row in the source that does not exist in the target, the row fails.

Upsert

When you run a task with the Upsert task operation, Data Integration updates all rows in the target that also exist in the source and inserts all new source rows in to the target.

If a source field contains a NULL value and the corresponding target field contains a value, Data Integration retains the existing value in the target field.

Delete

When you run a task with the Delete task operation, Data Integration deletes all rows from the target that exist in the source.

Synchronization task sources

You can add sources to a synchronization task when you configure the task or when you perform data catalog discovery. If you add a source when you configure the task and the source connection is a database connection, you can add a single object or multiple related objects.

You can add a source to a synchronization task in the following ways:

When you configure the task

Select the source connection and source object on the **Source** tab when you configure the synchronization task.

If the source connection is a database connection, you can use a single object as a source or multiple related objects as sources. Define relationships based on key columns or create a user-defined join condition.

Through data catalog discovery

If your organization administrator has configured Enterprise Data Catalog integration properties, you can perform data catalog discovery to find the source object in the catalog.

Search for the source object on the **Data Catalog** page, select the object in the search results, and then add it to a new synchronization task.

Rules and guidelines for multiple-object databases

Use the following rules and guidelines when you configure a multiple-object database:

- All objects must be available through the same source connection. All database tables in a multiple-object source must have valid relationships defined by key columns or user-defined join conditions.
- When you add multiple database tables as sources for a task, you can either create relationships or user-defined joins, but not both.
- The synchronization task wizard removes a user-defined join under the following conditions:
  - You remove one of two remaining database tables from the list of sources for the task.
  - You change the source connection from database to flat file or Salesforce.

Synchronization task targets

You can use a single object as a target for a synchronization task.

The target connections that you can use depend on the task operation you select for the task. For example, if you select the upsert task operation, you cannot use a flat file target connection because you cannot upsert records into a flat file target.

Flat file target creation

If a task has a flat file target, create the flat file before you save the task. Or, you can create the flat file target with the synchronization task wizard when all of the following are true:

- The source connection type is Salesforce, database, or ODBC.
- The source object is Single or Custom.

- The target connection type is Flat File.

The synchronization task wizard uses the source object name as the default name of the flat file target. It truncates the name of the flat file to the first 100 characters if the source name is too long. If the target name conflicts with the name of another target object, the following error appears:

```
Object named <object name> already exists in the target connection.
```

Database target truncation

You can configure a synchronization task to truncate a database target table before writing new data to the table when you configure the task to use an Insert task operation. By default, Data Integration inserts new rows without truncating the target table.

Salesforce targets and IDs for related objects

Data Integration identifies records of a Salesforce object based on one of the following types of IDs:

- Salesforce ID
  Salesforce generates an ID for each new record in a Salesforce object.
- External ID
  You can create a custom external ID field in the Salesforce object to identify records in the object. You might create an external ID to use the ID generated from a third-party application to identify records in the Salesforce object. You can use one or more external IDs to uniquely identify records in each Salesforce object.

If you create a task that writes to a Salesforce target, the source must provide either the Salesforce IDs or the external IDs for the records in the Salesforce target object and applicable related objects. To configure external IDs for related Salesforce objects, use **Related Objects** to select the external ID for each applicable related object. A related object is an object that is related to another object based on a relationship defined in Salesforce. The task uses the Salesforce ID or external ID to update changes to related objects.

If the source in a task contains external IDs for Salesforce objects, you must specify the external IDs for all related objects when you create the Salesforce target for the task. If you do not specify the external ID, Data Integration requires the Salesforce ID to identify records in each related object.

For more information about creating and using Salesforce external IDs, see the Data Integration Community article, "Using External IDs and Related Objects in Informatica Cloud".

Update columns

Update columns are columns that uniquely identify rows in the target table. Add update columns when the database target table does not contain a primary key and the synchronization task uses an update, upsert, or delete task operation.

When you run the synchronization task, the synchronization task uses the field mapping to match rows in the source to the database table. If the synchronization task matches a source row to multiple target rows, it performs the specified task operation on all matched target rows.

Column names in flat files

If the column name in a flat file source contains nonalphanumeric characters, starts with a number, or contains more than 75 characters, the synchronization task modifies the column name in the flat file target.

The synchronization task truncates column names to 75 characters. For a flat file source, the **Data Preview** area and the **Field Expression** dialog box show modified column names. For a flat file target, the synchronization task changes the column name in the flat file when it generates the file at run time.

Field mappings

Configure field mappings in a synchronization task to map source columns to target columns.

Configure field mapping on the **Field Mapping** page of the synchronization task wizard.

You must map at least one source column to a target column. Map columns with compatible data types or create field expressions to convert data types appropriately.

Depending on the task operation, the synchronization task requires certain fields to be included in the field mapping. By default, the synchronization task maps the required fields. If you configure the field mapping, ensure that the required fields remain mapped. If you do not map the required fields, the synchronization task fails.

The following table shows the required fields for each applicable task operation for a database target:

Required field	Task operations	Description
Primary Keys	Delete Update Upsert	Map primary key columns to enable the synchronization task to identify records to delete, update, or upsert in a database target.
Non-null fields	Insert Update Upsert	Map all fields that cannot be null in database.

When you configure field mappings, you can also perform the following tasks:

- Edit field data types.
- Add a mapplet to the field mapping.
- Create lookups.

Field data types

When you create a synchronization task, Data Integration assigns a data type to each field in the source and target. You can edit the field datatypes on the **Field Mapping** page of the synchronization task wizard. You can edit field data types for any source or target type except for Data Integration Connector sources and targets, and mapplets.

Mapplets in field mappings

You can add a mapplet to a field mapping. After you add a mapplet to a field mapping, you must map the source fields to the input fields of the mapplet and map the output fields of the mapplet to the target fields.

When a source field is mapped directly to a target field and you map an output field of a mapplet to the same target field, Data Integration concatenates the values of the source and output fields in the target field. Verify that the expression is correct for the target field.

Note: The names of the output fields of a mapplet do not match the source field names. Data Integration appends a number to the end of the source field name to determine the output field name. In addition, Data Integration may not display the output fields in the same order as the source fields.

Lookup conditions

A lookup returns values based on a lookup condition. You can create a lookup condition based on information in the source on the **Field Mapping** page of the synchronization task wizard. For example, for a SALES source database table, you might set the ITEM_ID column equal to ITEM_ID column in a ITEMS flat file, and have the lookup return the item name for each matching item ID.

When you create a lookup condition, you define the following components:

- Lookup connection and object. The connection and object to use to perform the lookup. When possible, use a native connection. For example, to perform a lookup on an Oracle table, use an Oracle connection instead of an ODBC connection.
- Source and lookup fields. The fields used to define the lookup condition. The synchronization task compares the value of the source field against the lookup field and then returns a value based on the match. You can define multiple conditions in a lookup. If you define more than one lookup condition, all lookup conditions must be true to find the match.
  For example, you define the following conditions for a lookup:

  ```
  SourceTable.Name = LookupTable.Name
  SourceTable.ID = LookupTable.ID
  ```

  The synchronization task performs the following lookup:

  ```
  Lookup (SourceTable.Name = LookupTable.Name, SourceTable.ID = LookupTableID)
  ```

Lookup return values

When you configure a lookup, you configure a lookup return value. The lookup return value depends on the return value properties that you define, such as multiplicity or a lookup expression.

A lookup return value is the value that Data Integration returns when it finds a match based on the lookup condition. If the lookup returns an error, Data Integration writes the row to the error rows file.

You can configure a lookup expression as part of the lookup return value. Configure a simple expression that uses the $OutputField variable to represent the lookup return value.

For example, the following expression adds 100 to each lookup return value:

```
$OutputField+100
```

As another example, you can use the concatenate operator (||) to append a string to a string lookup return value as follows:

```
'Mighty' || '$OutputField'
```

You can use parameters defined in a parameter file in a lookup expression.

The following table describes the lookup return value properties that you can configure:

Lookup Return Value Property	Description
Output Field	The field from the lookup table that you want to use.
Multiplicity	How Data Integration handles multiple return values. - Error If More Than 1 Output Value. Select if the synchronization task should display an error when the lookup condition returns multiple values. Data Integration rejects rows when multiple matches are found, writing them to the error rows file. This is the default. - Randomly Pick 1 Output Value. Select if the synchronization task should choose the first returned value when a lookup condition returns multiple values. Different systems might use different orders to return lookup values.
Expression	A simple expression that uses $OutputField to represent the selected output field. By default, Data Integration passes the lookup return value without alteration with the following expression: $OutputField.

Rules and guidelines for lookups

Use the following rules and guidelines when creating a lookup:

- If the lookup is on a flat file, the file must use a comma delimiter. You cannot use any other type of delimiter.
- When you configure a lookup, you can configure a simple lookup expression as part of the lookup return value. Use the $OutputField variable to represent the expression. If you use a lookup expression that does not include $OutputField, you negate the action of the lookup.
- Tasks with a flat file lookup that run by a Secure Agent on Windows 7 (64 bit) might not complete. To resolve the issue, configure a network login for the Secure Agent service.
- On the **Field Mapping** page, you can perform a lookup or create an expression for each source field. You cannot do both.
- Each task can contain one or more lookups. To avoid impacting performance, include less than six lookups in a task.
- When performing the lookup, the task performs an outerjoin and does not sort the input rows. The lookup performs a string comparison that is not case-sensitive to determine matching rows.
- The source field and lookup field in the lookup condition must have compatible data types. If the data types are not compatible, the following error appears:

  ```
  Source field [<source field name> (<source field data type>)] and lookup field
  [<lookup field name> (<lookup field data type>)] have incompatible data types.
  ```

  If you create multiple lookup conditions on a lookup field and the lookup source is a flat file, all source fields must have the same data type. The synchronization task uses the larger precision and scale of the source field data types as the precision and scale for the target field. If the source fields do not have the same data type, the following error appears:

  ```
  Lookup field <field name> in <file name> has conflict data types inferenced: <data
  type 1> and <data type 2>.
  ```
- You cannot include lookup fields of particular data types in a lookup condition. When the lookup field in a flat file has the Text or Ntext data type or the target field of a lookup has the Text or Ntext data type, the task fails.
- If you run a task with a lookup and the source field, lookup field, or output field of the lookup no longer exist in the lookup object, an error appears.

Configuring a synchronization task

Configure a synchronization task using the synchronization task wizard.

To configure a synchronization task, complete the following steps:

1. Complete the prerequisite tasks.
2. Create the synchronization task.
3. Configure the source.
4. Configure the target.
5. Optionally, configure data filters.
6. Configure field mappings.
7. Optionally, configure a schedule and advanced options.

As you work through the task wizard, you can click **Save** to save your work at any time. When you have completed the wizard, click **Finish** to save and close the task wizard.

Synchronization prerequisite tasks

Before you create a synchronization task, complete the following prerequisite tasks:

- Create database users.
  To write source data to a database target, the database administrator must create a database user account in the target database. Each database user account must have the DELETE, INSERT, SELECT, and UPDATE privileges.
- Verify that the sources and targets meet your requirements.

Defining a synchronization task

1. To create a synchronization task, click **New** > **Tasks**. Select **Synchronization Task** and click **Create**.

   To edit a synchronization task, on the **Explore** page, navigate to the synchronization task. In the row that contains the task, click **Actions** and select **Edit**.
2. In **Synchronization Task Details**, configure the following fields:

Field	Description
Task Name	Name of the synchronization task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Task names are not case sensitive.
Location	Project folder in which the task resides. If the **Explore** page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.

Field	Description
Description	Description of the synchronization task. Maximum length is 4000 characters.
Task Operation	Select one of the following task operation types: - Insert - Update - Upsert - Delete The list of available targets in a subsequent step depend on the operation you select.

3. Click **Next**.

Configuring the source

Select the source for the synchronization task. The steps to configure a source vary based on whether you use a single object or saved query as the source, or you use multiple database tables as the source.

Configuring a single object or saved query as the source

You can configure a single object or saved query as the source of a synchronization task.

1. On the **Source** page, select a connection.

   To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**.

2. To use a single source, select **Single**.

   To use a saved query, select **Saved Query**.

   You can use a saved query when you use a database connection.

3. If the connection includes less than 200 objects, select a source object or click **Select**.

   If the connection includes more than 200 objects, click **Select**.

   The **Select Source Object** dialog box displays up to 200 objects. If the object you want to use does not display, enter a search string to reduce the number of objects that display.

   Select an object and click **Select**.

4. To display technical names instead of business names, select **Display technical field names instead of labels**.

   This option is not available for all connection types.

5. To display source fields in alphabetical order, click **Display source fields in alphabetical order**.

   By default, fields appear in the order returned by the source system.

6. For a flat file or FTP/SFTP single source, click **Formatting Options**. Select a delimiter and text qualifier. Optionally, select an escape character.

   If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark.

   If you choose a delimiter for an FTP/SFTP flat file, Data Integration applies the delimiter to the local file, not the remote file, when previewing and reading data. If the remote and local files are not synchronized, you might see unexpected results.

7. If preview data does not display automatically, click **Show Data Preview** to preview data.

   The Data Preview area shows the first ten rows of the first five columns in the object. It also displays the total number of columns in the object.

   The Data Preview area does not display certain Unicode characters correctly. If the data contains binary data, the Data Preview area shows the following text:

   ```
   BINARY DATA
   ```

8. To preview all source columns in a file, click **Preview All Columns**.

   The file shows the first ten rows of the source.

9. Click **Next**.

Configuring multiple database tables as the source

You can configure multiple database tables as the source of a synchronization task.

1. On the **Source** page, select a database connection.

   To create a connection, click **New**. To edit a connection, click **View** and then click **Edit**.

2. Select **Multiple**.

   The Source Objects table displays.

3. Click **Add**.

4. In the **Select Source Objects** dialog box, select the objects you want to use.

   The dialog box displays up to 200 objects. If the objects that you want to use do not display, enter a search string to reduce the number of objects that display.

   When you select an object, it appears in the Selected Objects list. To remove an object from the Selected Objects list, press **Delete**.

5. Click **Select**.

   The selected sources display in the **Source Objects** table. To remove a source, in the **Actions** column, click **Remove**.

6. To display source fields in alphabetical order, select **Display source fields in alphabetical order**.

   By default, source fields appear in the order returned by the source system.

7. Create source relationships or create a user-defined join, and click **OK**.

   To create a relationship, perform the following steps:

   a. Select a database table and click **Create Relationship**.

   b. Select the source key for the table and then select the related source object and matching object key.

   c. Click **OK**.

   d. Match the primary key of the source table to the corresponding foreign key of the related database table.

   e. Create relationships as necessary to include all sources in the task.

   To create a user-defined join to join all database tables, perform the following steps:

   a. Select **User Defined Join** and define the join.

   b. Any existing relationships are added to the join condition. To ensure that you enter field names correctly, use the **Object** list and **Fields** list to add field names to the join statement.

   c. To save the user-defined join, click **OK**.

8. To preview source data, select the source in the Source Objects table. If preview data does not appear automatically, click **Show Data Preview**.

   The Data Preview area shows the first ten rows of the first five columns in the source. It also displays the total number of columns in the source.

   The Data Preview area does not display certain Unicode characters as expected. If the data contains binary data, the Data Preview area shows the following text:

   ```
   BINARY DATA
   ```

9. To preview all source columns in a file, select the source in the Source Objects table and click **Preview All Columns**.

   The file shows the first ten rows of the source.

10. Click **Next**.

Configuring the target

You can configure a single target for a synchronization task. The options that appear on the page depend on the task type and target type that you select for the task.

1. On the **Target** page, enter the following information:

Field	Description
Connection	Select a connection. The list of available connections depends on the task operation defined for the task. To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**.
Target Object	If the connection includes less than 200 objects, select a target object or click **Select**. If the connection includes more than 200 objects, click **Select**. The **Select Target Object** dialog box displays up to 200 objects. If the object you want to use does not appear, enter a search string to reduce the number of objects that display. Select a target object and click **OK**.
Display Technical Field Names Instead of Labels	Displays technical names instead of business names. Not available for all connection types.
Display Target Fields in Alphabetical Order	Displays target fields in alphabetical order instead of the order returned by the target system.
Formatting Options	For Flat File and FTP/SFTP connections only. Select a delimiter and text qualifier. Optionally, select an escape character. If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark. If you choose a delimiter for an FTP/SFTP flat file, Data Integration applies the delimiter to the local file, not the remote file, when previewing and reading data. If the remote and local files are not synchronized, you might see unexpected results.

Field	Description
Create Target	Flat File and relational database connections only. Creates a target file. You can create a target file when the source connection is Salesforce, database, or ODBC, and the source object is Single or Custom. Enter a file name and select the source fields that you want to use. By default, all source fields are used.
Truncate Target	Database targets with the Insert task operation only. Truncates a database target table before inserting new rows. - True. Truncates the target table before inserting all rows. - False. Inserts new rows without truncating the target table. Default is False.
Enable Target Bulk Load	Select this option to write data in bulk mode. The default value is false.

2. If preview data does not appear automatically, click **Show Data Preview** to preview data.

   The **Data Preview** area shows the first ten rows of the first five columns in the target. It also shows the total number of columns in the target.

   The **Data Preview** area does not display certain Unicode characters correctly. If the data contains binary data, the Data Preview area shows the following text:

   ```
   BINARY DATA
   ```

3. To preview all target columns in a file, click **Preview All Columns**.

   The file shows the first ten rows of the target.

4. Click **Next**.

Configuring the data filters

Use a data filter to reduce the number of source rows that the synchronization task reads for the task. By default, the synchronization task reads all source rows.

1. On the **Data Filters** page, choose whether to read all rows in sources or to read the first set of rows in sources.
   - To read all rows, select **Process all rows**.
   - To read the first set of rows, select **Process only the first** and enter a number.
2. To create a data filter, click **New**.
   - To create a simple data filter, select a source column and operator. Enter the value you want to use, and click **OK**.
   - To create an advanced data filter, click **Advanced**. Enter the field expression you want to use and click **OK**.

   You can use parameters defined in a parameter file in data filters. When you use a parameter in a data filter, start the data filter with the parameter. For example, use `$$Sales < 100000` instead of `100000 > $$Sales`.

   To delete a data filter, click **Delete**.
3. Click **Next**.

Configuring the field mapping

Configure field mappings to define the data that the synchronization task writes to the target.

1. On the **Field Mapping** page, configure field mappings.
2. If you included multiple source objects in the task, you can select each source object in the **Source** field to display the fields for the selected object. Or, you can view all source object fields.

   When displaying all source object fields, the **Sources** table displays field names grouped by source object. You can place the cursor over the Status icon for a source field to determine the following information:

   - Database table or Salesforce object to which the field belongs.
   - data type of a field.
3. Some source types allow you to configure field data types. To configure field data types for a source, click **Edit Types**.

   If the task includes more than one source, first select the source you want to edit.

   In the **Edit Field Datatypes** dialog box, configure the following data type attributes and click **OK**:

data type attribute	Description
Datatype	The data type of data in the column.
Precision	Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.
Scale	Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a numeric data type is 65535.

4. To add a mapplet, complete the following steps:
   a. Click **Add Mapplet**.
   b. In the **Add Mapplet** dialog box, select the mapplet.
   c. To display technical names instead of business names, select **Display technical field names instead of labels**.
   d. To display fields in alphabetical order, click **Display mapplet fields in alphabetical order**.

      By default, fields appear in the order specified by the mapplet.
   e. If necessary, select a connection for the mapplet.
   f. Click **OK**.
5. To configure field mappings, for **Mapping Selection**, select one of the following options:
   - **Source to Target**. Displays the source and target. Map source fields to the applicable target fields.
   - **Source to Mapplet**. Displays the source and the input fields of the mapplet. Map the source fields to the applicable input fields of the mapplet.
   - **Mapplet to Target**. Displays the output fields of the mapplet and the target fields. Map the output fields of the mapplet to the applicable target fields.

The **Clear Mapping**, **Automap**, and **Validate Mapping** buttons apply to the selected area of the field mapping.

6. To match fields with the same name, click **Automap** > **Exact Field Name**. Or, to match fields with similar names, click **Automap** > **Smart Map**.

   You can also select and drag the source fields to the applicable target fields.

   Data Integration caches field metadata. If the fields do not appear correctly, click **Refresh Fields** to update the cache and view the latest field attributes.

7. To configure field data types for a target, click **Edit Types**.

   This option is not available for all target types. If the task includes more than one target, first select the target you want to edit.

   In the **Edit Field Datatypes** dialog box, configure the following data type attributes and click **OK**:

data type attribute	Description
Datatype	The data type of data in the column.
Precision	Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.
Scale	Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a numeric data type is 65535.

8. To create an expression to transform data, click the **Add or Edit Expression** icon in the **Actions** column.

   In the **Field Expressions** dialog box, enter the expression you want to use and click **OK**.

   You can use parameters defined in a parameter file in expressions.

9. To create a lookup, click the **Add or Edit Lookup** icon.

   In the **Field Lookup** dialog box, configure the following properties and click **OK**:

Option	Description
Lookup Connection	Connection for the lookup object.
Lookup Object	Object on which you want to lookup a value.
Display Technical Field Names Instead of Labels	Displays technical names instead of business names. Not available for all connection types.
Display Fields in Alphabetical Order	Displays lookup fields in alphabetical order. By default, fields appear in the order returned by the lookup system.
Source Fields	Source column to use in the lookup condition.

Option	Description
Lookup Fields	The column in the lookup table to use in the lookup condition.
Output Field	The column in the lookup table that contains the output value.
Multiplicity	Determines how to handle cases when a lookup returns multiple values: - Error If More Than 1 Output Value. Select if the synchronization task should display an error when the lookup condition returns multiple values. Data Integration rejects rows when multiple matches are found, writing them to the error rows file. This is the default. - Randomly Pick 1 Output Value. Select if the synchronization task should choose the first returned value when a lookup condition returns multiple values. Different systems might use different orders to return lookup values.
Expression	A simple expression that uses the $OutputField variable to represent the lookup return value. Enter a simple expression, such as $OutputField*100. You can use parameters defined in a parameter file in lookup expressions. To return the lookup return value without an additional expression, use $OutputField.

10. To clear an expression or lookup and delete the field mapping, click the **Clear Expression/Lookup** icon next to the target field.
11. To clear all field mappings, click **Clear Mapping**.
12. To validate a mapping, click **Validate Mapping**.
13. Click **Next**.

Configuring a schedule and advanced options

On the **Schedule** page of the synchronization task wizard, you can specify to run a synchronization task manually or schedule it to run at a specific time or interval. You can create a schedule or use an existing schedule.

You can also configure email notifications and advanced options for the task on the **Schedule** page.

1. On the **Schedule** page, choose whether to run the task on a schedule or without a schedule.
2. To run a task on a schedule, click **Run this task on schedule** and select the schedule you want to use.

   To create a new schedule, click **New**. Enter schedule details and click **OK**.

   To remove the task from a schedule, click **Do not run this task on a schedule**.
3. If necessary, select a runtime environment to run the task.
4. Optionally, if the task runs in a serverless runtime environment, configure serverless usage properties.
5. Configure email notification options for the task.
6. Optionally, configure advanced options.

You can configure the following advanced options:

Advanced options	Description
Preprocessing Commands	Command to run before the task.
Postprocessing Commands	Command to run after the task completes.
Parameter File Name	Name of the file that contains the definitions and values of user-defined parameters used in the task.
Maximum Number of Log Files	Number of session log files, error log files, and import log files to retain. By default, Data Integration stores each type of log file for 10 runs before it overwrites the log files for new runs.
Update Columns	Database targets only. Temporary primary key columns to update target data. If the database target does not include a primary key column, and the task performs an update, upsert, or delete task operation, click Add to add a temporary key.
Upsert Field Name	The target field to use to perform upserts.

7. Choose whether to run the task in standard or verbose execution mode.

   If you select verbose execution mode, the mapping generates additional data in the logs that you can use for troubleshooting. It is recommended that you select verbose execution mode only for troubleshooting purposes. Verbose execution mode impacts performance because of the amount of data it generates.

8. Click **Finish**.

Viewing synchronization task details

You can view details about a synchronization task, including the source and target connections, the field mapping, and the associated schedule.

1. On the **Explore** page, navigate to the task.
2. In the row that contains the task, click **Actions** and select **View**.

   On the **Task Details** page, you can click **Edit** to modify the synchronization task or **Run** to run the task.

Running a synchronization task

You can run a synchronization task manually or on a schedule:

- To run a synchronization task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  You can also run a synchronization task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**.

- To run a synchronization task on a schedule, edit the task in the synchronization task wizard to associate the task with a schedule.

Rules and guidelines for running a synchronization task

Use the following rules and guidelines when you run a synchronization task:

- Verify that the source and target definitions are current. If the source or target no longer contains fields that are mapped in the field mapping, the synchronization task fails.
- You cannot run multiple instances of a synchronization task simultaneously. If you run a synchronization task that is already running, the synchronization task fails with the following error:

  ```
  Data synchronization task <Data Synchronization task name> failed to run. Another instance of the task is currently executing.
  ```

  If you configured the synchronization task to run on a schedule, increase the time interval between the scheduled tasks to prevent multiple instances of the synchronization task from running simultaneously. If you run the synchronization task manually, wait for the currently running instance of the synchronization task to complete before starting it again.

  You can view currently running synchronization tasks on the **All Jobs** or **Running Jobs** page in Monitor or on the **My Jobs** page in Data Integration.
- The synchronization task does not load any data into an IBM DB2 target if one or more records fails to load.
- When you use an active mapplet with a synchronization task that includes a saved query, the synchronization task ignores the configured target option for the task and inserts data to the target.

CHAPTER 5

Data transfer tasks

Use a data transfer task to transfer data from a source to a target. For example, you might use a data transfer task to transfer data from an on-premises database to a cloud data warehouse.

When you configure a data transfer task, you can augment the source data with data from a lookup source. Based on the source connection that you use, you can also sort and filter the data before you load it to the target.

To see if a data transfer task is applicable to the connectors you are using, see the help for the relevant connectors.

Task operations

When you configure a data transfer task, you specify the task operation. The operations available are based on the target that you select.

You can select the following operations:

Insert

Inserts all source rows into the target. If Data Integration finds a source row that exists in the target, the row fails.

If you write data to a flat file target, Data Integration truncates the flat file before it inserts the source rows into the file.

Update

Updates rows in the target that exist in the source. If Data Integration finds a row in the source that does not exist in the target, the row fails.

Upsert

Updates all rows in the target that also exist in the source and inserts all new source rows in to the target.

If a source field contains a NULL value and the corresponding target field contains a value, Data Integration retains the existing value in the target field.

Delete

Deletes all rows from the target that exist in the source.

Data transfer task sources

You can select a single source object to transfer data from or define a custom query.

The type of source and the formatting and advanced options that you can configure for the source depend on the source connection that you select. For example, you can create a custom query for database sources that support SQL queries. For a flat file source, you can configure formatting options such as the formatting type. For Salesforce sources, you can configure advanced options such as the SOQL filter condition, row limit, and bulk query.

For information about the options that you can configure for a source connection, see the help for the appropriate connector.

You can preview the data in the source. The preview returns the first 10 rows. To preview rows in alphabetical order, select **Display fields in alphabetical order**. Data Integration does not change the order of rows in the actual source. You can also download the preview results as a CSV file.

Source filters

Apply filter conditions to filter the source data that you transfer to the target.

You can apply the following types of filter conditions:

Simple

Select the source field, and configure the operator and value to use in the filter.

When you define more than one filter condition, the task evaluates them in the order that you specify. The task evaluates the filter conditions using the AND logical operator to join the conditions. It returns rows that match all the filter conditions.

Advanced

Create a filter expression using the expression editor. You enter one expression that contains all filters. You can use source fields and built-in functions in the expression.

For more information about advanced filters, see “Advanced data filters” on page 13.

Sort conditions

For certain source types, you can sort the source data to provided sorted data to the target.

When you sort data, you select one or more source fields to sort by. If you apply more than one sort condition, Data Integration sorts fields in the listed order.

To see if a connector supports sorting, see the help for the appropriate connector.

Custom queries

Create a custom query when you want to use a database source that you can't configure with a single object and lookup source. You might create a custom query to join multiple source tables.

To use a custom query as a source, select **Query** as the source type and then click **Select**. When you define the query, use SQL that is valid for the source database. You can use database-specific functions in the query.

When you create a custom query, enter an SQL SELECT statement to select the source columns you want to use. Data Integration uses the SQL statement to retrieve source column information.

Data Integration ensures that custom query column names are unique. If an SQL statement returns a duplicate column name, Data Integration adds a number to the duplicate column name as follows:

```
<column_name><number>
```

Tip: Test the SQL statement you want to use on the source database before you create a custom query. Data Integration doesn't display specific error messages for invalid SQL statements.

Second sources

When you configure a data transfer task, you can add a second source to use as a lookup source. Configure the lookup source on the **Second Source** page.

The task queries the lookup source based on the lookup condition that you specify and returns the result of the lookup to the target.

Select a second source when you want to augment the source data with a related value or values from the lookup source. For example, the source is an orders table that contains a customer ID field. You might retrieve the customer name and address from the lookup source so that you can include them in the target object. The task returns all fields from the lookup source.

To optimize performance, the task caches the lookup source. The cache remains static and does not change as the task runs. The task deletes the cache files after the task completes.

You can preview the data in the lookup source. The preview returns the first 10 rows. To preview rows in alphabetical order, select **Display fields in alphabetical order**. Data Integration does not change the order of rows in the actual source. You can also download the preview results to a CSV file.

You can also filter the data from both sources before writing it to the target.

Lookup condition

When you select a second source to use as a lookup source, you must configure one or more lookup conditions.

A lookup condition defines when the lookup returns values from the lookup source. When you configure a lookup condition, you compare the value of one or more fields from the original source with values in the lookup source.

A lookup condition includes an incoming field from the original source, a field from the lookup source, and an operator. To avoid possible naming conflicts, the data transfer task applies the prefix `SRC_` to the fields from the original source. If this results in a naming conflict for any field from the original source, the task applies the prefix `IN_SRC_` to the field from the original source.

For example, you might configure the following lookup condition when the original source contains the CustID field, the lookup source contains the CustomerID field, and you want to return values from the lookup source when the customer IDs match:

Lookup Field	Operator	Incoming Field
CustomerID	Equals	SRC_CustID

You can use the following operators in a lookup condition:

- Equals

- Not Equals
- Less Than
- Less Than or Equals
- Greater Than
- Greater Than or Equals

When you enter multiple conditions, the task evaluates the lookup conditions using the AND logical operator to join the conditions. It returns rows that match all the lookup conditions.

When you include multiple conditions, to optimize performance, enter the conditions in the following order:

1. Equals
2. Less Than, Less Than or Equals, Greater Than, Greater Than or Equals
3. Not Equals

The lookup condition matches null values. When an input field is NULL, the task evaluates the NULL equal to null values in the lookup.

If the lookup condition has multiple matches, the task returns any row.

Second source filters

You can apply filter conditions to filter the combined data.

You can configure the following types of filters:

Simple

To configure a filter condition, select a source field and configure the operator and value to use in the filter. You can select a field from either source. Fields from the original source are prefixed with the characters `SRC_` or `IN_SRC_`.

When you define more than one filter condition, the task evaluates them in the order that you specify. The task evaluates the filter conditions using the AND logical operator to join the conditions. It returns rows that match all the filter conditions.

Advanced

Create a filter expression using the expression editor. You enter one expression that contains all filters. You can use source fields and built-in functions in the expression.

For more information about advanced filters, see "Advanced data filters" on page 13.

Data transfer task targets

You can use a single object as a target for a data transfer task. Select a target object or create a new target object at run time.

The task operations that you can select depend on the target connection that you use. For more information about task operations for different target types, see the help for the appropriate connector.

You can preview the data in the target. The preview returns the first 10 rows. You can download the preview results to a CSV file. To preview rows in alphabetical order, select **Display fields in alphabetical order**. Data Integration does not change the order of rows in the actual target.

Database target truncation

You can configure a data transfer task to truncate a database target table before writing new data to the table when the task uses an Insert operation. By default, Data Integration inserts new rows without truncating the target table.

Update columns

Update columns are columns that uniquely identify rows in the target table. Add update columns when the database target table does not contain a primary key and the data transfer task uses an update, upsert, or delete operation.

When you run the data transfer task, the task uses the field mapping to match rows in the source to the database table. If the data transfer task matches a source row to multiple target rows, it performs the specified task operation on all matched target rows.

Field mapping

Configure field mapping in a data transfer task to map source fields to target fields. Configure field mapping on the **Field Mapping** page of the data transfer task wizard.

You must map at least one source field to a target field. If the task uses multiple sources, fields from the original source are prefixed with the characters `SRC_` or `IN_SRC_`.

You can configure the following field mapping options:

Options

Configure which fields to display. Click **Options** and select from the following display options:

- Show All
- Show Mapped
- Show Unmapped

Automap

Data Integration automatically links fields with the same name or similar name. Click **Automap** and select from the following mapping options:

- Exact Field Name. Data Integration matches fields of the same name.
- Smart Map. Data Integration matches fields with similar names. For example, if you have an incoming field `Cust_Name` and a target field `Customer_Name`, Data Integration automatically links the `Cust_Name` field with the `Customer_Name` field.
- Undo Automap. Data Integration clears fields mapping with Smart Map or Exact Field Name but does not clear manually mapped fields.

Actions

Additional field link options. Provides the following options:

- Map Selected. Links the selected incoming field with the selected target field.
- Unmap Selected. Clears the link for the selected field.
- Clear Mapping. Clears all field mappings.

After you map a field, if you want to configure a field expression, click the mapped field name. You can include fields and built-in functions in the expression but not user-defined functions.

When you create a target at run time, Data Integration maps the source fields to the target fields. You cannot unmap or edit the source fields mapped to the target but you can add fields to the target. You can also edit the mapped field expression and metadata and reorder the added fields. You cannot reorder the fields copied from the source.

Field data types

When you create a data transfer task, Data Integration assigns a data type to each field in the source and target. When you add a field to a target that you create at run time, you select the data type.

Configuring a data transfer task

To configure a data transfer task, perform the following steps:

1. Define the data transfer task.
2. Configure the source.
3. Optionally, configure a second source to use as a lookup source.
4. Configure the target.
5. Configure field mappings.
6. Configure runtime options.

As you work through the data transfer task wizard, you can click **Save** to save your work at any time. Use the **Validation** panel to validate the task. When you have completed the wizard, click **Exit** to close the task wizard.

Defining the data transfer task

1. To create a new data transfer task click **New** > **Tasks**. Select **Data Transfer Task** and click **Create**.

   To edit a data transfer task, on the **Explore** page, navigate to the data transfer task. In the row that contains the task, click **Actions** and select **Edit**.
2. Configure the following properties:

Property	Description
Name	Name of the data transfer task.
Location	Project and folder in which the task resides.

Property	Description
Description	Description of the data transfer task. Maximum length is 4000 characters.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task.

3. Click **Next**.

Configuring the source

You can configure a single object or custom query as the source of a Data Integration task.

1. On the **Source** page, select the source connection.

   To create a connection, click **New**. To edit a connection, click **View**, and the in **View Connection** dialog box, click **Edit**.
2. Select the source type. To use a single source, select **Single**. To use a custom query, select **Query**.

   You can use a custom query when you use a database connection.
3. Select the source object or enter the query.
4. For file sources, configure formatting options.
5. Configure data filters.
6. Configure sort conditions.

   Not all connectors support sorting. The **Sort** area appears if the source connector supports sorting.
7. If preview data does not appear automatically, expand the **Data Preview** area to preview the source data.
8. Click **Next**.

Configuring a second source

Optionally, configure a second source to use as a lookup source.

1. On the **Second Source** page, select **Yes** to add a second source to the task.

   If you do not want to configure a second source, select **No**.
2. If you add a second source, perform the following steps to configure the source:
   a. In the **Source Details** area, select **Augment Data with Lookup**.
   b. Select the source connection and source object.

      To create a connection, click **New**. To edit a connection, click **View**, and the in **View Connection** dialog box, click **Edit**.
   c. For file sources, configure formatting options.
   d. If preview data does not appear automatically, expand the **Data Preview** area to preview the source data.
   e. Configure one or more lookup conditions.
   f. Optionally, configure data filters for the combined sources.
3. Click **Next**.

Configuring the target

You can write data to a single target. You can select an existing target object or create a new object at run time. If you create a target at run time, the task operation is Insert.

1. On the **Target** page, configure the following properties as required:

Property	Description
Connection	Select a connection. To create a connection, click **New**. To edit a connection, click **View**, and the in **View Connection** dialog box, click **Edit**. To configure advanced properties for the connection, click **Advanced Options**. Not available for all connections types. For information about a particular connector's properties, see the help for the appropriate connector.
Object	Select a target object. Click **Select**. The **Select Target Object** dialog box displays up to 200 objects. If the object you want to use does not display, enter a search string to reduce the number of objects that display.
Target operation	Select one of the following task operations: - Insert - Update - Upsert - Delete - Data Driven The operations available depend on the target connection.
Formatting Options	For flat file and FTP/SFTP connections only. Select a delimiter and text qualifier. Optionally, select an escape character. If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark. If you choose a delimiter for an FTP/SFTP flat file, Data Integration applies the delimiter to the local file, not the remote file, when previewing and reading data. If the remote and local files are not synchronized, you might see unexpected results.
Truncate Target	Database targets with the Insert task operation only. Truncates a database target table before inserting new rows. - True. Truncates the target table before inserting all rows. - False. Inserts new rows without truncating the target table. Default is False.
Enable Target Bulk Load	Select this option to write data in bulk mode. The default value is false.

2. For database targets, configure update columns if necessary.
3. If preview data does not appear automatically, expand the Data Preview area to preview the target data.
4. Click **Next**.

Configuring the field mapping

Configure field mappings to define the data that the data transfer task writes to the target. Configure field mappings on the **Field Mapping** page.

1. To match fields with the same name, click **Automap** > **Exact Field Name**. Or, to match fields with similar names, click **Automap** > **Smart Map**.

   You can also select and drag the source fields to the applicable target fields.
2. To configure the field expression, click the mapped field. In the **Field Expression** window, enter the expression you want to use and click **OK**.
3. If you create a target at run time and want to add target fields, click **Add**. Configure the following field properties:

Property	Description
Name	Name of the field.
Type	The data type of the data in the column.
Precision	Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.
Scale	The number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a numeric data type is 65535.

4. Click **Next**.

Configuring runtime options

Configure runtime options on the **Runtime Options** page.

You can run a data transfer task manually or on a schedule.

1. To run a data transfer task on a schedule, select **Run on a schedule** and then select the schedule.

   Note: You must create the schedule in Administrator before you can select it in the task.
2. Configure email notification options for the task.
3. Click **Save**.

Running a data transfer task

You can run a data transfer task in one of the following ways:

- Manually. To run a data transfer task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  You can also run a data transfer task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**.
- On a schedule. To run a data transfer task on a schedule, edit the task in the data transfer task wizard to associate the task with a schedule.

CHAPTER 6

Replication tasks

Use the replication task to replicate data to a target. You might replicate data to back up the data or perform offline reporting.

You can replicate data in Salesforce objects or database tables to databases or flat files. You can configure a task to replicate all rows of a source object each time the task runs or to only replicate the rows that changed since the last time the task was run. You can use a replication task to reset target tables and create target tables.

Load types

The load type determines the type of operation to use when the replication task replicates data from the source to the target.

Use one of the following load types when you replicate data:

Incremental loads after initial full load

The first time the replication task runs, it performs a full load, replicating all rows of the source. For each subsequent run, the replication task performs an incremental load. In an incremental load, the replication task uses an upsert operation to replicate rows that changed since the last time the task ran. You can specify this load type when the task uses a Salesforce source and a database target.

Incremental loads after initial partial load

The replication task always performs an incremental load with this load type. The first time the replication task runs, the replication task processes rows created or modified after a specified point in time. For each subsequent run, the replication task replicates rows that changed since the last time the task ran. You can specify this load type when the task uses a Salesforce source and a database target.

Full load each run

The replication task replicates all rows of the source objects in the task during each run. You can specify this load type when the task uses a Salesforce or database source and a database or flat file target.

For information about incremental load, see the help for Salesforce Connector.

Full load

For a full load, the replication task replicates the data for all rows of the source objects in the task. Each time the task runs, the replication task truncates the target database tables or flat file and performs a full data refresh from the source.

Run a full load in the following situations:

- The replication task uses a database source.
- A Salesforce object in the replication task is configured to be nonreplicateable within Salesforce.
  If you run an incremental load on a replication task that contains nonreplicateable objects, the replication task runs a full load on the object. Contact the Salesforce administrator to get a list of replicateable Salesforce objects.
- The data type of a Salesforce field changed.
  If the replication task detects a data type change, you might need to reset the target table to create a table that matches the updated Salesforce object. Then run the replication task with full load to reload the data for all Salesforce objects included in the replication task. Alternatively, you can set the AutoAlterColumnType custom configuration property so that the target table column updates to match the Salesforce object. The AutoAlterColumnType property does not apply in certain situations, such as when the source and target data types are not compatible. For more information about the AutoAlterColumnType property, see the help for Salesforce Connector.

Replication task sources

You can use Salesforce and database sources in replication tasks. You can use database tables, aliases, and views as sources.

Replication task targets

You can replicate source data to database and flat file targets. The type of target affects how the replication task replicates data.

Replicate data to a database target

For a replication task configured with the full load type, the first time you run the task, the task creates the database tables. The replication task then writes the replicated data to the tables. During subsequent runs, the task truncates the database tables and then writes the source data to the tables.

For a replication task configured with the incremental load after initial full load type, the first time you run the task, the task creates the database tables. The replication task then writes the replicated data to the tables. During subsequent runs, the task performs an upsert operation.

For a replication task configured with the incremental load after initial partial load, the first time you run the task and for subsequent runs, the task performs an upsert to replicate source data from the specified period in time.

Replicate data to a flat file target

When you run a replication task for a flat file target, Data Integration performs a full load each time you run the task. When you run the task for the first time, the replication task creates the flat files. The replication

task then stores the files in the specified directory and writes the replicated data to the files. During subsequent runs, the replication task truncates the file and loads the data.

When the replication task includes flat file targets with multibyte data on Linux, the default locale must be UTF-8.

Reset a database target

To drop all of the target tables in the task, reset a relational target table in a replication task.

You might want to drop a target table when you want to re-create it based on the latest source object definition. When you reset target tables for a replication task, the task performs the following actions:

1. Drops all of the target tables included in the replication task from the database.
2. Sets the load type for the replication task to full load.

After you run the replication task to reset the target, you must run the replication task again to reload the data from all source objects included in the task. When you run the replication task after you reset the target, the replication task recreates each target table. The replication task then loads all of the data into the new table.

If the target table is damaged, the replication task might consistently fail to write to the target table. You might need to reset the relational target.

Resetting a target table

You can reset the relational target tables included in a replication task.

1. On the **Explore** page, navigate to the replication task.
2. In the row that contains the task, click **Actions** and select **Reset Target**.

Rules and guidelines for resetting a target table

Use the following rules and guidelines when you reset a target table:

- If you previously created indexes on a target table and reset the target table, the replication task drops the indexes and the target table. You must create the indexes again.
- If you try to reset a target table that does not exist in the target database, an error appears.
- The replication task drops the target table that was updated the last time the task ran. For example, if you change the prefix of the target table and do not run the replication task, the replication task resets the old target table.

Table and column names in a database target

The replication task replicates source objects and fields to target database tables and columns, respectively. In certain cases, the replication task does not give the target table and column names the same names as the source objects and fields.

The replication task might not give the same names in the following circumstances:

- You write replicated data to a database target and use a table name prefix.
  A table name prefix prevents you from overwriting database tables when you share a database account.
- You replicate case-sensitive data.
  When the replication task replicates data to a target database, it creates all table and column names in uppercase. If the target database is case sensitive, use uppercase table and column names when you query the database.
- You replicate objects with long object and field names.
  When a source object or field name contains more characters than the maximum allowed for the name in the target, the replication task truncates the table or column name in the target database.

Table name truncation

When you replicate a source object to a database, the replication task replicates the data to a database table with the same name as the source object.

If you replicate data to a database target and the length of the source object name exceeds the maximum number of characters allowed for the target table name, the replication task truncates the table name in the target database. It truncates the table name to the first X characters, where X is the maximum number of characters allowed for a table name in the target database.

Duplicate tables names from same replication task

If you replicate multiple source objects from the same replication task to the same database user account and truncation causes duplicate table names, the replication task replaces the last character of the duplicate table names with sequential numbers.

For example, the replication task contains the following Salesforce source objects:

```
TenLetters1234567890TenLettersXXX
TenLetters1234567890TenLettersYYY
TenLetters1234567890TenLettersZZZ
```

When you replicate the objects, the replication task creates the following truncated table names in the target database:

```
TenLetters1234567890TenLetters
TenLetters1234567890TenLetter1
TenLetters1234567890TenLetter2
```

Duplicate table names from different replication tasks

If you replicate multiple source objects with the same names from different replication tasks to the same database user account, the replication task creates one target table, and overwrites the table data each time you replicate one of the objects. If you run a full load, it overwrites the entire table. If you run an incremental load, it overwrites the changed rows.

To avoid overwriting tables, use a different target table name prefix for each replication task.

Column name truncation

If the length of the source field name exceeds the maximum number of characters allowed for a column name in a relational target, the replication task truncates the column name in the target database. It

truncates the column name to the first X characters, where X is the maximum number of characters allowed for a column name in the target database.

For example, the replication task creates a column name in an Oracle database based on the following 40-character field name of a Salesforce object:

```
TenLetters1234567890TenLettersXXXXXXXXXX
```

The replication task truncates the column name to the first 30 characters:

```
TenLetters1234567890TenLetters
```

If the truncation causes duplicate column names for the target table, the replication task replaces the last character of the duplicate column names with sequential numbers. The replication task also replaces the last character of duplicate table names from the same task.

Target prefixes

When you replicate data to a database table or flat file, the replication task names each database table or flat file based on the corresponding source object name.

By default, the replication task includes the target prefix SF_. For example, the default flat file name for the Account Salesforce object is `SF_ACCOUNT.CSV`. If you remove the default target prefix and do not specify another prefix, the replication task creates a flat file or database table with the same name as the corresponding source object.

You can use target prefixes to prevent overwriting data. For example, you and another user share a database user account. The other user ran a replication task on the Contact object from her Salesforce account. Her replication task created a database table named Contact in the shared database. You use no target prefix and run a replication task on the Contact object from your Salesforce account. The replication task overwrites the data in the existing Contact table with your data. If you use the SF_ prefix, the replication task creates a table named SF_CONTACT and does not overwrite the existing table named Contact.

Creating target tables

You can use Data Integration to create the database table for a target before you run the replication task. You might want to create the target table and then modify the table properties before the replication task loads the data into the table.

1. On the **Explore** page, navigate to the replication task.
2. In the row that contains the task, click **Actions** and select **Create Target**.

Replication task schedules

You can run a replication task manually or schedule it to run at a specific time or interval.

You might want to run a replication task manually for the following reasons:

- To verify that a replication task is configured properly.

- To replicate the data occasionally. You might not want to replicate data at regular intervals.

If you specify a schedule in the replication task wizard, you can select an existing schedule or create a schedule. You can include a repeat frequency to replicate the data at regular intervals.

If you remove a task from a schedule while the task is running, the task that is in progress completes and future tasks are cancelled.

Configuring a replication task

Configure a replication task to replicate data from a source to a target. When you configure a replication task, you specify the source connection, target connection, and the objects to replicate.

A replication task can replicate data from one or more Salesforce objects or database tables. When you configure the task, you can replicate all available objects through the selected connection, or you can select objects for replication by including or excluding a set of objects. You can also exclude rows and columns from the replication task.

Configure a replication task to run full or incremental loads. Perform a full load to replicate all rows for each object. Perform an incremental load to replicate rows that are new or changed since the last time you ran the task.

Associate a schedule with a replication task to specify when and how often the task runs. If you remove the replication task from a schedule as the task runs, the task completes. The replication task cancels any additional task runs associated with the schedule.

To configure a replication task, use the replication task wizard to perform the following steps:

1. Complete the prerequisite tasks.
2. Create the replication task.
3. Configure the source.
4. Configure the target.
5. Optionally, exclude fields.
6. Optionally, configure data filters.
7. Optionally, configure a schedule and advanced options.

As you work through the task wizard, you can click **Save** to save your work at any time. When you have completed the wizard, you can click **Finish** to save and close the task wizard.

Rules and guidelines for configuring replication tasks

Consider the following rules and guidelines when you configure a replication task:

- The names of source tables and fields can use a maximum of 55 characters.
- Multiple replication tasks cannot write to the same database table or flat file.
- You cannot configure a replication task when the source and target object are the same. If the source and target connections are the same, enter a target prefix to distinguish the source and target objects.

Replication prerequisite tasks

Complete the following prerequisite tasks before you create a replication task:

1. Verify that a database target exists.

   To replicate data to a database target, the database target must exist before you create the replication task. If a database does not exist, the database administrator must create a target database. The database must meet the minimum system requirements.

2. Create database users.

   To replicate data to a database target, the database administrator must create a database user account in the target database. Each database user account must have the CREATE, DELETE, DROP, INSERT, SELECT, and UPDATE privileges. You must have a database user account for each replication task that writes to that database. You can use the same database user account for multiple replication tasks. If you use the same database user account for multiple groups, ensure that the replication tasks do not overwrite data in the same target tables.

3. Create a directory for the flat files.

   To replicate data to a flat file, create a directory to store the flat files.

4. Optionally, create a schedule.

   To run replication tasks at specified times or on regular intervals, create a schedule.

Defining a replication task

Define a replication task using the replication task wizard.

1. To create a replication task, click **New** > **Tasks**. Select **Replication Task** and then click **Create**.

   To edit a replication task, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Edit**.

2. In the Replication Task Details, configure the following properties:

Property	Description
Task Name	Name of the replication task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Task names are not case sensitive.
Location	Project folder in which the task resides. If the **Explore** page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.
Description	Description of the task. Maximum length is 255 characters.

3. Click **Next**.

Configuring the source

Configure the source on the **Source** page of the replication task wizard.

Note: Column names of a database source must not contain spaces or hyphens.

If you replicate a source with a name that includes a dollar sign ($), the replication task replaces the dollar sign with an underscore (_) in the target name.

1. In the **Source Details** area, select a connection.

   To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**.

2. To select the objects to replicate, select one of the following options:

   - **All Objects**. To replicate all objects in the database or Salesforce account, select **All Objects**.
   - **Include Objects**. To select the objects you want to include, click **Select**. In the **Include Source Objects** dialog box, select the objects to replicate and click **Select**.
   - **Exclude Objects**. To select the objects you want to exclude, click **Select**. In the **Exclude Source Objects** dialog box, select the objects to exclude and click **Select**. The task replicates all available objects except for the selected objects.

   The **Available Objects** area displays up to 200 objects. If the objects that you want to use do not display, enter a search string to reduce the number of objects that display.

   When you select an object, it displays in a list. To remove a selected object, select the object and press **Delete**.

3. If you want the replication task to stop processing when it encounters an error, click **Cancel processing the remaining objects**.

   If you want the replication task to continue to process a task after it encounters an error, click **Continue processing of the remaining objects**.

   By default, the replication task stops processing the task when it encounters an error.

4. To display technical names instead of business names for some source types, click **Display technical names instead of labels**.

5. Click **Next**.

Configuring the target

1. On the **Target** page, enter the following information:

Property	Description
Connection	Connection to the target object. To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**.
Target Prefix	Prefix that is added to Salesforce object names to create the flat file names or table names in a target database. By default, the prefix is SF_.

Property	Description
Load Type	Type of load. Select one of the following options: - **Incremental loads after initial full load**. Loads all data the first time the task runs. In subsequent runs, loads changed data only. - **Incremental loads after initial partial load**. Loads data created or modified after a specified period in time. If you select this option, enter the date and time, for example, August 29, 2015 at 2:00. The replication task uses the time zone that is set for the user. If the server on which the data resides is located in a different time zone, adjust the date and time accordingly. For example, the time zone for the user is Pacific Time and the time zone for the server is Eastern Time, which is three hours ahead of Pacific Time. The user wants the initial load to replicate data modified on the server after August 29, 2015 at 2:00 AM. Because the user's time zone is Pacific Time, the user specifies August 28, 2015 and 11:00 PM. - **Full Load each run**. Loads all data every time the task runs. This option is enabled for tasks with a Salesforce source and a relational target. For all other tasks, the replication task performs a full load.
Delete Options	Select one of the following options: - **Remove Deleted Columns and Rows**. Deletes columns and rows from the target if they no longer exist in the source. - **Retain Deleted Columns and Rows**. Retains columns and rows in the target that were removed from the source.
Commit Size	Number of rows to commit. Default for full load replication is 5,000 rows. Default for incremental load replication is 999,999,999.

2. Click **Next**.

Configuring the field exclusions

To limit the fields loaded in to a target, configure field exclusions for each source object. By default, the replication task loads all fields in to the target.

1. On the **Field Exclusion** page, click **Exclude Fields**.
2. In the **Field Exclusion** dialog box, select the source object that you want to use.
3. In the **Included Fields** list, select and move the fields that you want to exclude to the **Excluded Fields** list.
4. Click **OK**.

   The excluded fields display in the **Field Exclusion** page. To remove an excluded field, click **Delete** next to the field.
5. Click **Next**.

Configuring the data filters

By default, the replication task replicates all source rows to the target. To filter source rows that are replicated, configure data filters. If you replicate multiple source objects, create a different set of data filters for each object.

1. On the **Data Filters** page, enter the following details:

Property	Description
Row Limit	Select one of the following options: - Process all Rows. Replicates all rows of the source. - Process Only the First... Rows. Replicates the first X rows, where X is the number of rows. You might choose to process the first set of rows to test the task. You cannot specify a row limit on replication tasks with non-Salesforce sources. If you select a non-Salesforce source, the option is disabled.
Data Filters	Click **New** to create a data filter on a Salesforce or database source. You can create simple or advanced data filters.

2. Click the **Delete** icon next to the data filter to delete the filter.
3. Click **Next**.

Configuring a schedule and advanced options

Configure a schedule and advanced options for a replication task on the **Schedule** page of the task wizard.

1. On the **Schedule** page, choose whether to run the task on a schedule or without a schedule.
2. To run a task on a schedule, click **Run this task on schedule** and select the schedule you want to use.

   To create a new schedule, click **New**. Enter schedule details and click **OK**.
   To remove the task from a schedule, click **Do not run this task on a schedule**.
3. Optionally, if the task runs in a serverless runtime environment, configure serverless usage properties.
4. Configure email notification options for the task.
5. Optionally, enter the advanced options as required.

Advanced option	Description
High Precision Calculations	In calculated fields, allows for precisions of up to 28. Recommended for Salesforce calculation fields.
Use Float Semantic	When enabled, the task uses a target-specific floating point data type.
Preprocessing Commands	Commands to run before the task.
Postprocessing Commands	Commands to run after the task completes.
Maximum Number of Log Files	Number of session log files and import log files to retain. By default, Data Integration stores each type of log file for 10 runs before it overwrites the log files for new runs.

6. Choose whether to run the task in standard or verbose execution mode.

If you select Verbose mode, the mapping generates additional data in the logs that you can use for troubleshooting. It is recommended that you select verbose execution mode only for troubleshooting purposes. Verbose execution mode impacts performance because of the amount of data it generates.

7. Click **Finish**.

Viewing replication task details

You can view details about a replication task, including the load criteria, source and target connections, field exclusions, data filters, and the associated schedule.

1. On the **Explore** page, navigate to the task.
2. In the row that contains the task, click **Actions** and select **View**.

   On the **Task Details** page, you can click **Edit** to modify the replication task or **Run** to run the task.

Running a replication task

You can run a replication task in the following ways:

- Manually. To run a replication task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  You can also run a replication task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**.
- On a schedule. To run a replication task on a schedule, edit the task in the replication task wizard to associate the task with a schedule.

Rules and guidelines for running a replication task

Use the following guidelines when you run a replication task:

- You cannot run multiple instances of a replication task simultaneously. If you run a replication task that is already running, the replication task fails with the following error:

  ```
  Data replication task failed to run. Another instance of task <Data Replication task name> is currently replicating the same objects.
  ```

  If you configured the replication task to run on a schedule, increase the time interval between the scheduled jobs to prevent multiple instances of the replication task from running simultaneously. If you run the replication task manually, wait for the currently running instance of the replication task to complete before starting it again.
- If you run a replication task that excludes fields and that writes to a database target, the replication task drops any indexes defined on the excluded fields in the target table.
- If you replicate timestamp data, the replication task truncates the millisecond portion of the timestamp data.

CHAPTER 7

Masking tasks

Use masking tasks to mask the sensitive fields in source data with realistic test data for nonproduction environments. You can choose to create a subset of the sensitive source data that reconciles object relationships.

When you configure a masking task, choose the source and target and then select a masking rule for each field in the source you want to mask. You can also use inplace masking to mask the data in the same system from which the masking task reads the data.

A data masking rule is a type of masking that you can apply to a selected field. The type of masking rule that you apply depends on the type of the field that you need to mask. You can select built-in rules for masking fields such as Social Security numbers, credit card numbers, phone numbers, and dates. You can apply substitution values for fields such as names, cities, countries, or positions. You can mask fields with random values or with repeatable values.

For example, you might need to test a Human Resources application. You need realistic employee data to test with. You can mask the fields in an Employee table to create the test data.

You can apply masking parameters to some data masking rules. Masking parameters are options that you can apply to customize the rules.

If the source and target locations are different, you can create a subset of the source data. Define data subset criteria to selectively process source data. For example, you can use a data subset to create a small environment for testing and development. You can define the type of data that you want to include in the subset database. The data subset retains foreign key relationships from the source data.

Rules and guidelines for masking tasks

Consider the following rules and guidelines when you run masking tasks:

- You can run masking tasks on Salesforce sources.
- The Secure Agent must have access to Salesforce servers.
- Tasks that include data subset properties require a staging connection. You can create a staging connection on an H2 database.
- To improve batch processing, configure the *EnableSalesForceStagingResponse* flag in the Custom Configuration Details for the Secure Agent and set it to TRUE.

  Bulk operations that contain large amounts of data to read in a single query might encounter a connection reset at regular intervals. The task might fail because of the connection reset. Improved batch processing reduces the chances of a connection reset during a task run.

Masking task options

You can configure different options when you create a masking task.

You can choose a single source object or multiple source objects. You can configure the task operation that you want to perform in the target. You can choose whether you want to perform inplace masking. You can set filters to create a subset of data. Apply data masking rules to source objects and configure the rule parameters. When you schedule to run a masking task, you can configure email notifications and advanced options.

Source objects

You can add a single object or multiple related objects in a masking task.

You can add a single object that does not contain any related objects. You can add multiple objects that have an explicit relationship defined in Salesforce. For example, if you use the Opportunity object as a source, you can add the related Campaign as well. You can also add the RecordType object because it is related to the Campaign object. All Salesforce objects in a multiple-object source must have a predefined relationship in Salesforce.

If you select multiple source objects, you can choose an object and add the related parent, child, and self-reference objects manually. When a source object references to itself within a task, it is called self-reference relationship.

When you select multiple source objects, you can download the subset graph from the **Source** page to view the relationships between the related objects.

Schema graph

A schema graph contains the graphical representation of the relationships between the multiple source objects in a masking task.

You can see the visual images of the multiple source objects, the relationships, and the assignments.

You can view the graph from the **Source** page or from the **Data Filters** page.

To view the object relationships on the local system, download the schema graph. The graph downloads in the following format:

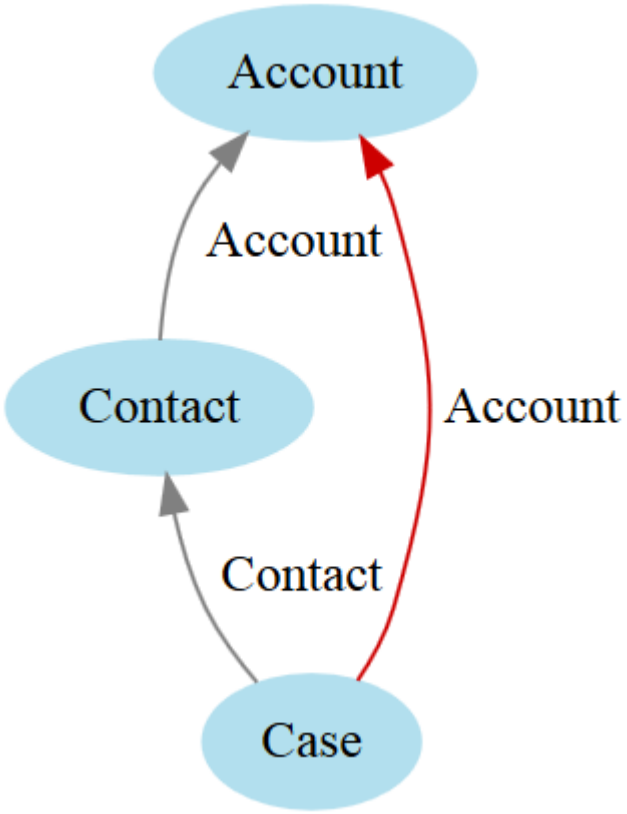
```
<filename>.dot
```

The DOT language defines a graph and you can directly view the graph in the browser. If you download the DOT format file, you must use a graph visualizer tool to render, view, and manipulate graphs in the DOT language. You can download and install any open source graph visualizer tool, such as Graphviz, Canviz, Grappa, and Beluging.

Schema Graph Example

Consider that you select the following source objects in a task: Account, Case, and Contact. The parent object is Account, and the child objects are Case and Contact.

The following image shows a sample schema graph:



The path from Account to Case is the default path that the task takes after applying the filter. Then the task loads the Contact and Account parent records that are related to the Case object.

If you want to load all the contacts related to the filtered accounts and then the cases, you need to select the path from Account to Contact objects and then to the Case object. You must ensure that the subset is selected for all the objects that are included in the task. Edges marked in red indicate that the child record is selected from a parent object.

Target task operations

You can select a task operation that you want to perform in the target.

When the target is same as the source, you can only perform an update operation.

When the target is different from the source, you can select the following task operations:

Insert

Ignores existing target data and inserts source data.

Update

Updates data in the target location based on source data.

Upsert

Updates existing target data. If data does not exist in the target, the masking task inserts data.

You can perform update, upsert, and insert operations on a partial sandbox. When you run an inplace masking task, you can perform an update operation. If the target is different from the source, you can perform insert and upsert operations.

Inplace masking

In inplace masking, you select the target same as the source.

When you configure a masking task, you can use inplace masking to mask the data in the same system from which the masking task reads the data. When you choose inplace masking, you can perform an Update operation in the target, but you cannot perform an Insert or an Upsert operation. You can apply data filters to a single source object when you select a Salesforce connection.

When you choose inplace masking in a masking task, the task does not create a custom field or an external ID for relationship reconciliation.

Update partial sandbox

You can update a partial sandbox if the record ID of the target sandbox matches the production or source record ID.

Salesforce upserts data based on the external ID values. A masking task creates fields for the external IDs in the target when you configure the task. If the external ID field is empty, Salesforce creates duplicate records. To avoid duplicate records in target for the upsert operation, you need the external IDs that the masking task creates. You can update the partial sandbox to load the external IDs to perform an upsert operation.

You can first update the partial sandbox to add the external IDs, and update the external ID values for the existing records in the target. Run the task with an upsert operation to update or insert records.

Refresh fields

You can refresh the fields whenever you change metadata in a masking task.

To refresh metadata, click **Refresh Fields** on the **Target** page.

You can refresh metadata when you make changes to the following fields in a masking task:

- Connection properties
- Source objects
- Metadata in Salesforce source
- Metadata in Salesforce target

Validation reports

You can view validation reports before you run a task.

When you click **Validation Report** on the **Target** page, the masking task validates whether the fields are visible in source and targets, the schema is proper, and then creates a validation report. The validation report is in plain ASCII text format.

The validation report lists mandatory relationships, fields, and field and relationship mismatch between source and target. The report lists the mandatory fields that are not available in the source that you need to fix. The report lists warning messages if the non-mandatory fields or relationships are not present in the source or target.

The validation report contains the following fields:

Object Name

The name of the source or target object.

Field Name

The name of the field in the source or the target object.

Data Type

The data type of the field.

Relationship

The relationship that is not present in the source or the target.

Staging database

You can perform data subset operations in a masking task. H2 is a database that you use to stage subsets of data.

The masking task uses the H2 database to stage record IDs, data subset for simple entities, entities with junction objects, multipath relationships, and lookup based relationships, and masking fields.

The H2 staging database installer is packaged with the Secure Agent. You can run the H2 package installer to install the staging connection. You can either start the staging database connection from the **Schedule** page of the masking task wizard, or manually run the H2 startup script.

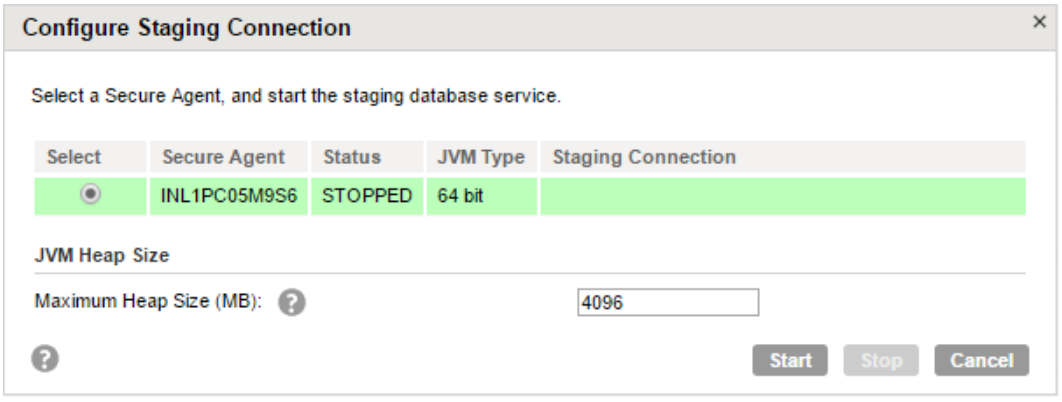
You can view the subset record count in the logs. The task reads the record IDs that are selected for subset from the staging database and uses standard API to read those records from the source. The task uses standard or bulk API to load data into the target.

Start the staging connection

You can start the staging database connection from the **Schedule** page before you run a masking task.

Select a Secure Agent and start the staging database service. The task uses the staging database to store the data during the subset operations.

The following image shows the staging connection details:



You can view the JVM type and configure the JVM heap size for a staging connection. The JVM type specifies the type of JVM that the staging database uses. The JVM heap size specifies the heap memory for the staging database. Configure the JVM heap size to process large amount of data in the staging database

in a short time. JVM type can be of 32-bit or 64-bit. When the JVM type is 32-bit, the default maximum heap size is 512 MB. When the JVM type is 64-bit, the default maximum heap size is 4096 MB. You can increase or decrease the heap size based on the amount of memory that the system can support.

You can select a Secure Agent and stop the staging database service.

H2 database configuration requirements

You must consider the maximum file size of the file system, cache size, and heap size when you configure the H2 database.

The H2 staging database has the following requirements:

- When you use a FAT or FAT32 file system, the data limit for H2 database is 4 GB.
- Configure the cache and heap size in the H2 startup script according to the source size and Secure Agent hardware configuration. An increase in the cache size improves staging, target load, and subset computation performance. Select the heap size based on the amount of physical memory that the system supports. The default cache size in the script is 2048 MB and the heap size is 4096 MB.
- The heap size must be at least twice the size of the cache memory.
- You can also configure the heap size for the staging connection on the **Schedule** page under **Data Subset** options.

The following example contains the section of code from the startup script file with the properties that you can configure:

```
@echo off
:: Script location
set H2_JAR_DIR=%~dp0
:: H2 Cache size in KBs
set H2_CACHE_SIZE=2097152
:: H2 Jar Name
set H2_JAR_PATH=%H2_JAR_DIR%h2-1.3.176.jar
:: H2 DB Name
set H2_DB=dmask
:: JVM path
set JVM_PATH=%H2_JAR_DIR%..\..\..\..\..\..\..\jre\bin
:: JVM Options. Initial and maximum heap size
set JVM_OPTS=-Xms128m -Xmx4g
```

You can change the H2_CACHE_SIZE value and the JVM_OPTS value to increase or decrease memory requirement for the H2 database. Higher memory allocation ensures better staging and subset computation performance.

Note: Do not make other changes to the script file. Changes to other properties can damage the file.

The H2 startup script is available in the following location:

```
<Agent installation directory>\apps\Data_Integration_Server\$$Version\ICS\main\tomcat\cmask
\h2_start.bat
```

Installing and configuring H2 database manually on Windows

After you install the Secure Agent, you must run the H2 staging database installer that is packaged with the Secure Agent. In Informatica Intelligent Cloud Services, you must create a connection to connect to the H2 database.

1. Browse to the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main
\tomcat\cmask
```

2. Run the `h2_start.bat` startup script.

   The database starts up and the Command Prompt displays the parameters that you need to configure the connection in Informatica Cloud. Keep the Command Prompt open.
3. Log in to Informatica Intelligent Cloud Services.
4. Perform one of the following steps:
   - In Administrator, select Connections.
   - In Data Integration, open a source or target object in a task.
5. Click **New Connection**.

   The **New Connection** page appears.
6. Enter the connection name for the H2 database.

   Connection names are not case sensitive. Connection names can contain alphanumeric characters, spaces, and the following special characters:

   ```
   _ . + -
   ```
7. Optionally, enter a description for the connection. Maximum length is 255 characters.
8. Select the **JDBC_IC (Informatica Cloud)** connection type.

   The JDBC_IC connection properties appears.
9. Enter the Secure Agent group that runs the masking task.
10. Enter the JDBC Connection URL that appears in the Command Prompt.
11. Enter the path to the JDBC jar directory that appears in the Command Prompt.
12. Enter the database schema.
13. Enter the user name and password that appears in the Command Prompt to connect to the H2 database.
14. To test the connection, click **Test Connection**.
15. To create the connection, click **Save**.

Installing H2 database manually on Linux

You can install H2 database on Linux and configure the connection in Informatica Intelligent Cloud Services.

Provide read, write, and execute access to the following directory:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main
\tomcat\cmask
```

1. Browse to the following location:

   ```
   <Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main
   \tomcat\cmask
   ```
2. To run the H2 startup script, enter the following command:

   ```
   nohup sh h2_start.sh &
   ```

   The H2 database starts up and lists all the parameter values that you need to configure the connection in Informatica Cloud.

   Note: To stop the database, identify the process ID and enter the following command:

   ```
   ps -ef | grep "h2", kill -9 processid
   ```
3. To view the parameter values, open `nohup.out` file in a text editor or run the following command:

   ```
   vi nohup.out
   ```
4. Log in to Informatica Intelligent Cloud Services.
5. Perform one of the following steps:

- In Administrator, select Connections.
- In Data Integration, open a source or target object in a task.

6. Click **New Connection**.

   The **New Connection** page appears.
7. Enter the connection name for the H2 database.

   Connection names are not case sensitive. Connection names can contain alphanumeric characters, spaces, and the following special characters:

   ```
   _ . + -
   ```
8. Optionally, enter a description for the connection. Maximum length is 255 characters.
9. Select the **JDBC_IC (Informatica Cloud)** connection type.

   The JDBC_IC connection properties appears.
10. Enter the Secure Agent group that runs the masking task.
11. Enter the JDBC Connection URL that appears in the `nohup.out` file.
12. Enter the path to the JDBC jar directory that appears in the `nohup.out` file.
13. Enter the database schema.
14. Enter the user name and password that appears in the `nohup.out` file to connect to the H2 database.
15. To test the connection, click **Test Connection**.
16. To create the connection, click **Save**.

Data subset

You can extract a subset of data from the source and move to the target in a masking task.

The masking task maintains primary and foreign key relationships in the subset data and reconciles the object relationships in the subset data.

Configure the following data subset options from the **Data Filters** page:

Data Filters

The data filter that you want to apply on the source. You can create a simple or an advanced data filter for an object. You can apply a filter to a single object in a task. You can apply multiple filters on the same object. You can add one advanced filter in a task. You can also use the filter values from a parameter file and mention the file name in the task.

Relationship Behavior

You can configure relationships when you select multiple source objects. When you perform a data subset operation, the masking task selects all the parent records of an object to maintain referential integrity. The task selects the child record if you configure to include the child objects. You can configure relationships of the child objects after you apply a data filter. You can select the child objects that you want to include in the data subset. You can view and download the schema graph that shows the graphical representation of the relationships between the source objects.

You can view the number of join operations that are required to compute a subset operation. View the sequence in which the task selects the records to create a data subset.

Subset Statistics

You can view the subset statistics such as total number of rows, the number of subset rows, and the subset size for each source object. The source contains large amount of data, but the target in which you want to create a subset might not contain enough space. To evaluate the target size, you can estimate the data subset. After you estimate the subset, you can view the estimated target size on all the masking task pages. If the estimated target size is large, you can update the task and estimate the subset again.

You can estimate the data subset size for multiple source objects.

Data subset options

Configure the data subset options on the **Schedule** page. You can view the data subset options if you select multiple source objects.

Configure the following data subset options:

Staging Connection

The connection that the task uses to run the data subset operation.

Source Lookup Batch Size

The number of records to retrieve from the Salesforce source in one SOQL query when the task writes to the target. Uses Salesforce standard API because the standard API limit is higher than the bulk API. Enter a number between 10 and 200 based on the SOQL character limit restriction for Salesforce.

Drop Staging Tables

Drops the staging tables even if there are error rows in the task.

When you configure data subset filters and run the masking task, the task runs through the staging, subset computation, target load, and staging drop phases. By default, if there are error rows, the task does not drop the staging tables. You can correct the errors and restart the task. The task resumes from the phase at which it failed. The staged data that is saved in the Secure Agent machine consumes some storage space, and you can choose to drop the staging tables even if there are errors rows present.

Automatic task recovery

In a masking task, you can estimate a subset. If you estimate the subset and then run the task, the task recovers from the previous stage and continues to the next stage.

When you configure data subset filters and run a masking task, the task runs through the staging, subset estimation, target load, and staging drop stages.

You can estimate the subset to evaluate the target subset size before you run the task. When you click **Estimate**, the task stages the records and estimates the subset. After you estimate the subset, if you click **Run**, the task resumes to load the target and then drops the staging tables.

If you save and run a task without estimating the subset, the task runs through all the stages and drops the staging tables at the end.

Every task has an associated staging schema. After the task runs through all the stages, the task drops staging tables if there are no errors. If you did not choose to drop the staging tables and if there are error rows present, the task does not drop the staging tables. If you run the same task after a few days, the task performs the data subset operation on the old data. To run the task with updated data, you must first reset the task. When you reset the task, the task status returns to the start stage. You can then estimate the subset and then run the task or directly run the task. In both cases, the task runs through staging and estimation stages and then loads the tables into the target.

Parameter files in data filters

In a masking task, you can use user-defined parameters in simple and advanced data filters.

When you use a parameter in a filter, start the filter with the parameter. Use two dollar signs to name the parameter in the following format: `$$<parameter>`

Save the parameter file local to the following directory:

```
<Secure Agent installation directory>/apps/Data_Integration_Server/data/userparameters
```

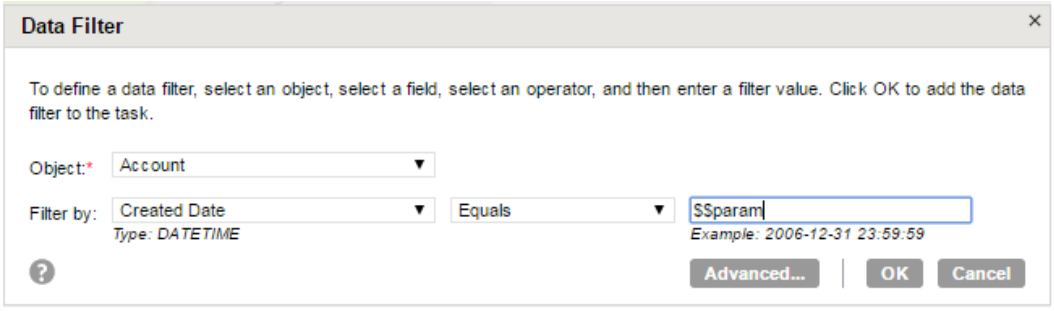
You can specify the parameter file name on the Schedule page of the task wizard. The parameter values are applied when the task runs.

Simple Filter Example

Consider that you apply a filter on the Account object. Configure the filter condition that the Created Date must be equal to $$param. Then create a parameter file with the following content:

```
$$param=('1991-10-03')
```

The following image shows a simple filter with the use of a parameter:

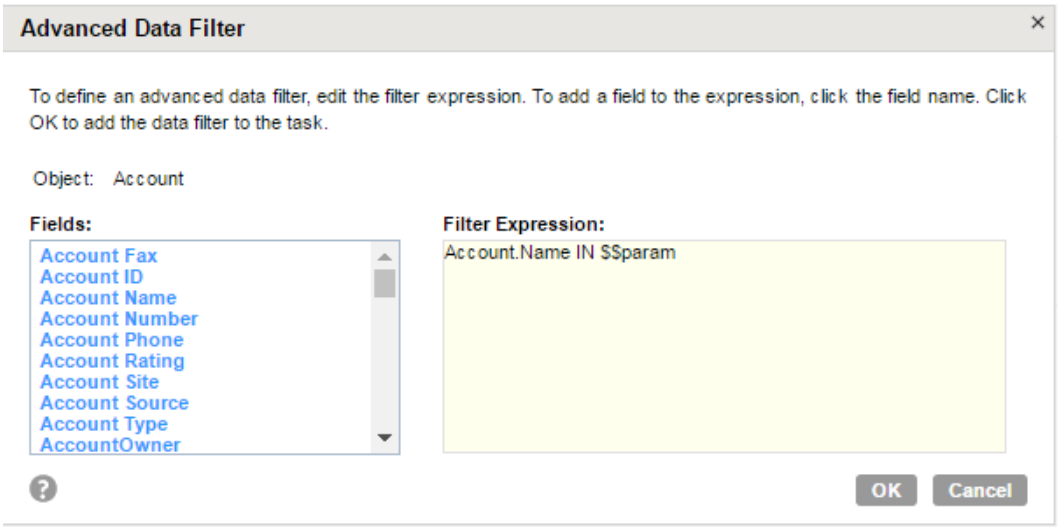


Advanced Filter Examples

Consider that you apply a filter on the Account object. In the Advanced Data Filter dialog box, you can specify the filter expression that the Account Name field must pick all the values from $$param. Then create a parameter file with the following content:

```
$$param=('Apple' , 'Microsoft')
```

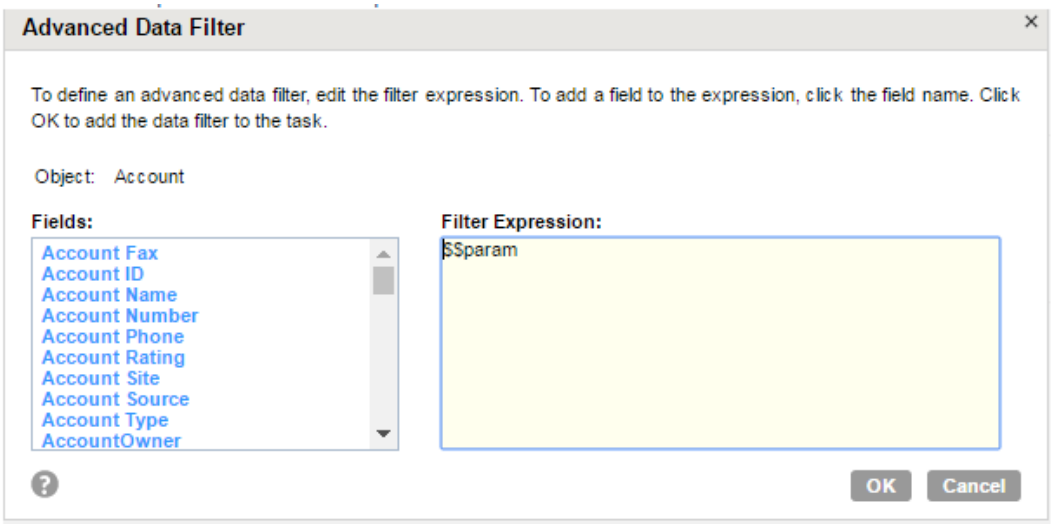
The following image shows an advanced filter with the use of parameter as a value:



You can also specify a filter expression as a parameter. Enter $$param as the filter expression. Then create a parameter file with the following content:

```
$$param=Name IN ('Apple' , 'Microsoft')
```

The following image shows an advanced filter with the use of a parameter as an expression:



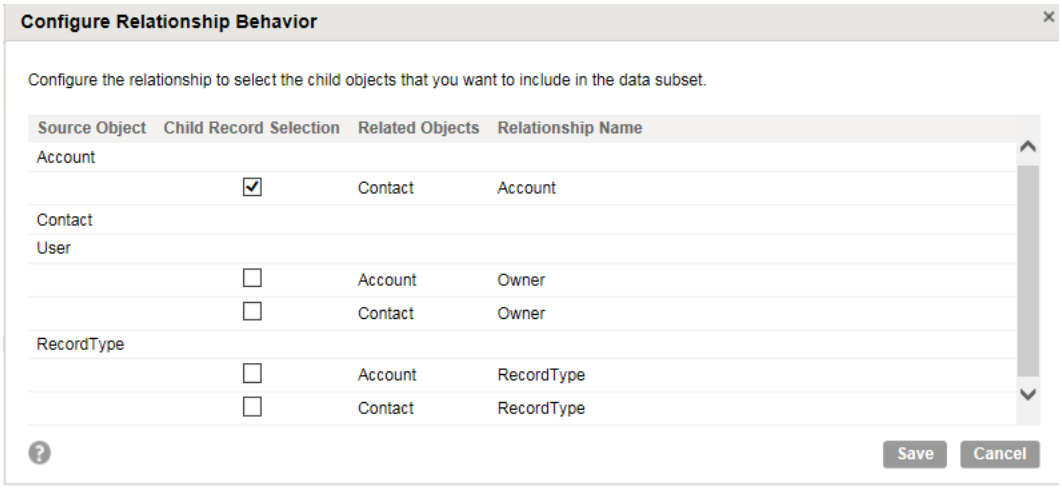
Configure relationship behavior

You can select the child objects that you want to include in the data subset.

When you define a filter criteria without configuring the child records selection, the task traverses through the object in the default path where the task includes required child records with minimal graph traversal. When you configure child records selection, the task follows the order in which you select the relationships. To configure child records selection, you can click **Configure** in the **Relationship Behavior** section on the **Data Filters** page.

You can view the number of join operations that are required to compute a subset operation. View the sequence in which the task selects the records to create a data subset.

The following image shows the child records that you can select for an object:



Data subset use cases for two objects

When you define a filter for an object in a task, the task selects a default path so that it can traverse through the entire graph at least once.

For every selected record, the task loads all the parent records to maintain referential integrity. You can configure relationship behavior to select child records from an object.

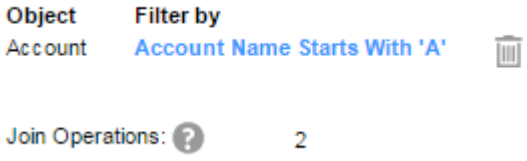
Consider that there are two objects Account and Contact. Account is the parent object, and Contact is the child object of Account. You can apply filter on the Account object or the Contact object. You can either use the default path that the task selects or configure the path. The number of paths traversed is minimal with the default path selection.

The number of subset rows is not the sum of all target rows that the task loads. The subset estimates show the number of unique records that are selected from all the relationships and loaded to target.

Case 1. Select the default path with filter on Account

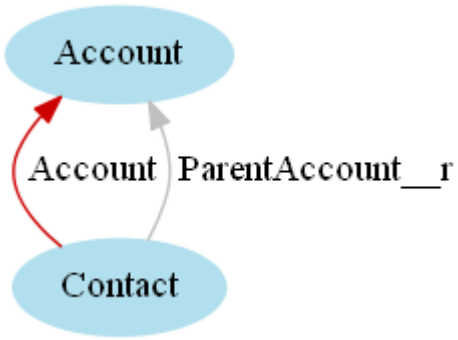
Consider that you apply a filter on the Account object with the condition that the account name starts with the letter A.

The following image shows the data subset filter criteria that you can configure:



The task traverses the default path from Account to Contact through the relationship Account. The number of join operations to compute the subset with the default path selection is two.

The following image shows the graphical representation of the relationship between Account and Contact objects:



The task first loads the records from the Account object on which the filter is applied. Then the task traverses through the default path from Account to Contact through the relationship Account and loads the records. To maintain referential integrity, the task traverses from Contact to Account through the relationship ParentAccount__r.

The following image shows the sequence in which the task selects the records:

Source	Target	Relationship	Rows
	Account		4
Account	Contact	Account	5
Contact	Account	ParentAccount__r	3

Based on the filter applied, the Account object has four rows. From Account to Contact, the Contact object has five rows through the relationship Account. From Contact to Account, the Account object has three rows through the relationship ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. If there are new records, the task adds the subset rows. In this use case, though the Account object shows total seven rows, the task loads four subset rows that are unique for the Account object from all the relationships. The task loads five subset rows for the Contact object.

The following image shows the subset statistics that you can estimate in a task:

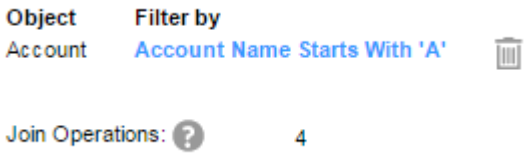
Subset Statistics

Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	5	10
Account	100	4	8
Total Size	203	9	18

Case 2. Select the configured path with filter on Account

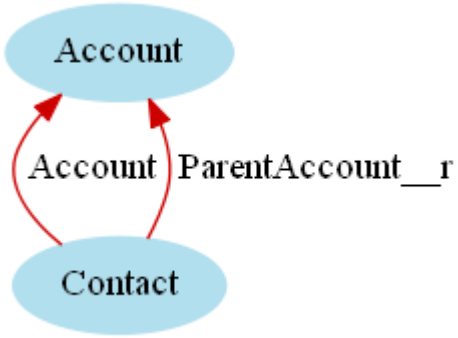
Consider that you apply a filter on the Account object with the condition that the account name starts with the letter A.

The following image shows the data subset filter criteria that you can configure:



Consider that you choose both the relationships Account and ParentAccount__r between Account and Contact objects. The number of join operations to compute the subset with the configured path is four.

The following image shows the graphical representation of the relationship between Account and Contact objects:



The task first loads the records from the Account object on which the filter is applied. Then the task traverses through the selected paths from Account to Contact through both the relationships Account and ParentAccount__r. To maintain referential integrity, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r.

The following image shows the selection sequence of the objects:

Source	Target	Relationship	Rows
	Account		5
Account	Contact	Account	5
Account	Contact	ParentAccount__r	5
Contact	Account	Account	4
Contact	Account	ParentAccount__r	3

Based on the filter applied, the Account object has five rows. From Account to Contact, the Contact object has ten rows with both the relationships Account and ParentAccount__r. From Contact to Account, the Account object has seven rows with the relationships Account and ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. If there are new records, the task adds the subset rows. In this use case, though the Account object shows total 12 rows, the task loads five subset rows that are unique for the Account object from both the relationships. Though the Contact object shows total 10 rows, the task loads seven subset rows that are unique for the Contact object from both the relationships.

The following image shows the subset statistics that you can estimate in a task:

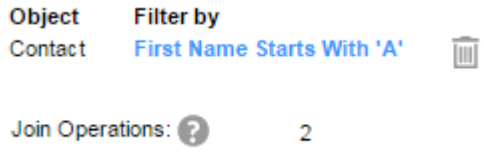
Subset Statistics

Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	7	14
Account	100	5	10
Total Size	203	12	24

Case 3. Select the default path with filter on Contact

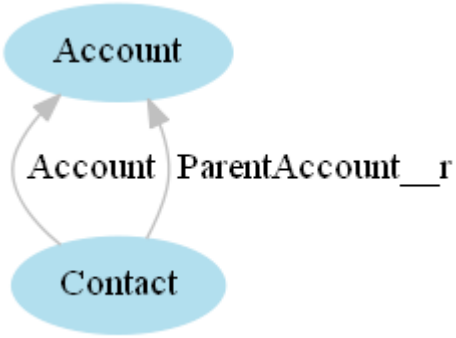
Consider that you apply a filter on the Contact object with the condition that the first names start with the letter A.

The following image shows the data subset filter criteria that you can configure:



Since the Account object is parent to the Contact object and the filter is applied on the Contact object, the task does not select a relationship for child record selection. To maintain referential integrity, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r. The number of join operations to compute the subset with the default path selection is two.

The following image shows the graphical representation of the relationship between Account and Contact objects:



The task first loads the records from the Contact object on which the filter is applied. Then the task traverses through the paths from Contact to Account through the relationships Account and ParentAccount__r, and loads the records.

The following image shows the sequence in which the task selects the records:

Source	Target	Relationship	Rows
	Contact		12
Contact	Account	Account	12
Contact	Account	ParentAccount__r	1

Based on the filter applied, the Contact object has 12 rows. From Contact to Account, the Account object has 13 rows through the relationships Account and ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. In this use case, though the Account object shows total 13 rows, the task loads 12 subset rows that are unique for the Account object from all the relationships. The task loads 12 subset rows for the Contact object.

The following image shows the subset statistics that you can estimate in a task:

Subset Statistics

Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	12	24
Account	100	12	24
Total Size	203	24	48

Case 4. Select the configured path with filter on Contact

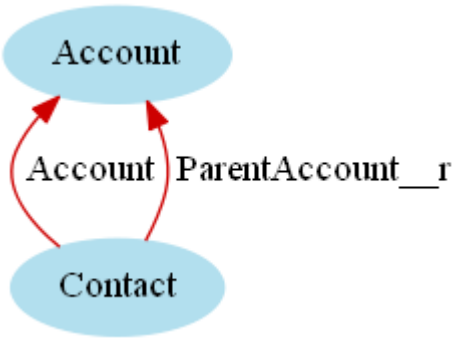
Consider that you apply a filter on the Contact object with the condition that the first names start with the letter A.

The following image shows the data subset filter criteria that you can configure:



Consider that you choose both the relationships Account and ParentAccount__r between Account and Contact objects. The number of join operations to compute the subset with the configured path selection is six.

The following image shows the graphical representation of the relationship between Account and Contact objects:



The task first loads the records from the Contact object on which the filter is applied. To maintain the referential integrity, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r. Then the task traverses from Account to Contact through the configured paths for child records selection with the relationships Account and ParentAccount__r and loads additional records. To maintain referential integrity for the additional records, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r.

The following image shows the sequence in which the task selects the records:

Source	Target	Relationship	Rows
	Contact		15
Contact	Account	Account	12
Contact	Account	ParentAccount__r	1
Account	Contact	Account	13
Account	Contact	ParentAccount__r	3
Contact	Account	Account	14
Contact	Account	ParentAccount__r	3

Based on the filter applied, the Contact object has 15 rows. From Contact to Account, the Account object has 13 rows through both the relationships Account and ParentAccount__r. From Account to Contact, the Contact object has 16 rows through both the relationships Account and ParentAccount__r. From Contact to Account, the Account object has 17 rows through the relationships Account and ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. In this use case, though the Account object shows total 30 rows, the task loads 14 subset rows that are unique for the Account object from both the relationships. Though the Contact object shows total 31 rows, the task loads 15 subset rows that are unique for the Contact object from both the relationships.

The following image shows the subset statistics that you can estimate in a task:

Subset Statistics

Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	15	30
Account	100	14	28
Total Size	203	29	58

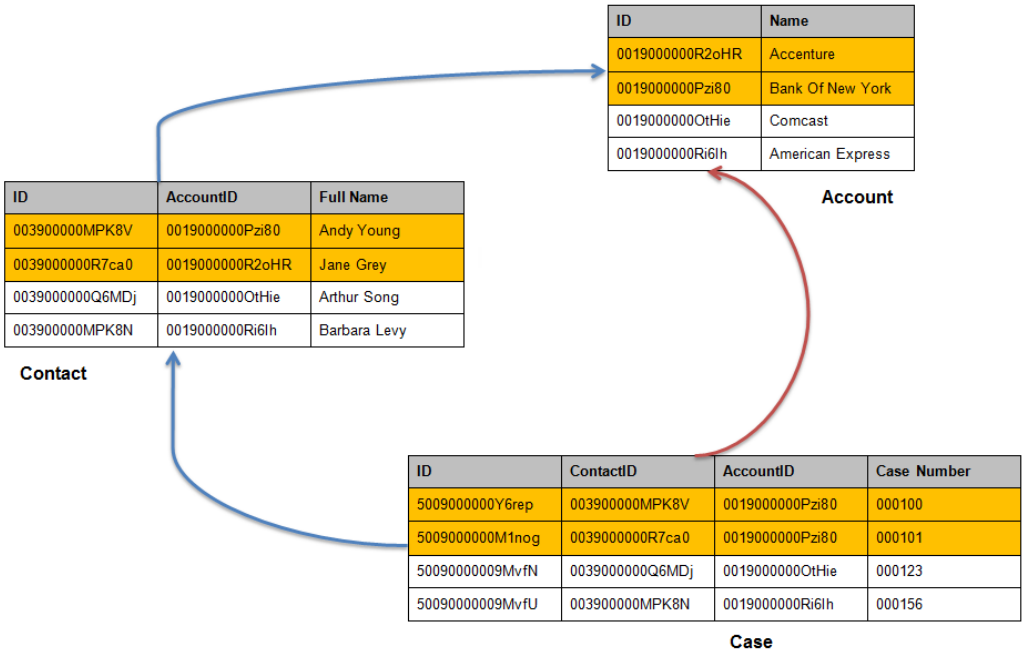
Data subset use cases for three objects

Consider that you selected the Account, Case, and Contact objects in a task. Account is the parent object, and Case and Contact are the child objects of Account. The Contact object is also parent to the Case object. Multipath relationship exists between Account and Case objects.

The task uses the default path to select the records. You can also configure the path to select the records.

Case 1. Default path

The following image shows the default path that the task chooses:



ID	Name
0019000000R2oHR	Accenture
0019000000Pzi80	Bank Of New York
0019000000OtHie	Comcast
0019000000Ri6Ih	American Express

Account

ID	AccountID	Full Name
003900000MPK8V	0019000000Pzi80	Andy Young
0039000000R7ca0	0019000000R2oHR	Jane Grey
0039000000Q6MDj	0019000000OtHie	Arthur Song
003900000MPK8N	0019000000Ri6Ih	Barbara Levy

Contact

ID	ContactID	AccountID	Case Number
5009000000Y6rep	003900000MPK8V	0019000000Pzi80	000100
5009000000M1nog	0039000000R7ca0	0019000000Pzi80	000101
50090000009MvfN	0039000000Q6MDj	0019000000OtHie	000123
50090000009MvfU	003900000MPK8N	0019000000Ri6Ih	000156

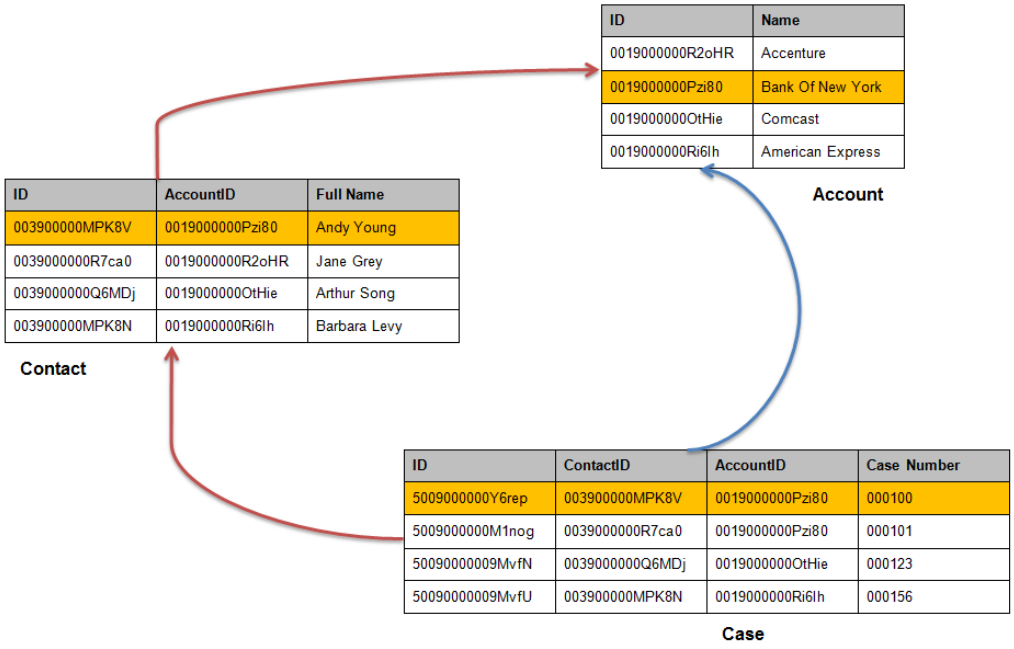
Case

Based on the default path selection, the task traverses to the Case object directly, and then traverses to the Contact object.

If you apply filter with Account Name equals to Bank of New York, the task marks the default minimal path Account -> Case to traverse through the complete graph. The task selects the Bank of New York account from the Account object and then marks the corresponding two records in the Case object based on the AccountID. The task traverses to the Contact object and selects the corresponding records based on ContactID. To maintain referential integrity, the task selects the additional account Accenture from the Account object.

Case 2. Configured path

The following images shows the path that you configure:



ID	Name
0019000000R2oHR	Accenture
0019000000Pzi80	Bank Of New York
0019000000OtHie	Comcast
0019000000Ri6Ih	American Express

Account

ID	AccountID	Full Name
003900000MPK8V	0019000000Pzi80	Andy Young
0039000000R7ca0	0019000000R2oHR	Jane Grey
0039000000Q6MDj	0019000000OtHie	Arthur Song
003900000MPK8N	0019000000Ri6Ih	Barbara Levy

Contact

ID	ContactID	AccountID	Case Number
5009000000Y6rep	003900000MPK8V	0019000000Pzi80	000100
5009000000M1nog	0039000000R7ca0	0019000000Pzi80	000101
50090000009MvfN	0039000000Q6MDj	0019000000OtHie	000123
50090000009MvfU	003900000MPK8N	0019000000Ri6Ih	000156

Case

If you configure to select the Contact -> Account and Case -> Account paths, the task first traverses to the Contact object and then traverses to the Case object.

If you apply filter with Account Name equals to Bank of New York, the task marks the corresponding ContactID in the Contact object based on the AccountID. For the selected ContactID in the Contact object, the task marks the corresponding Case ID in the Case object.

Data subset rows

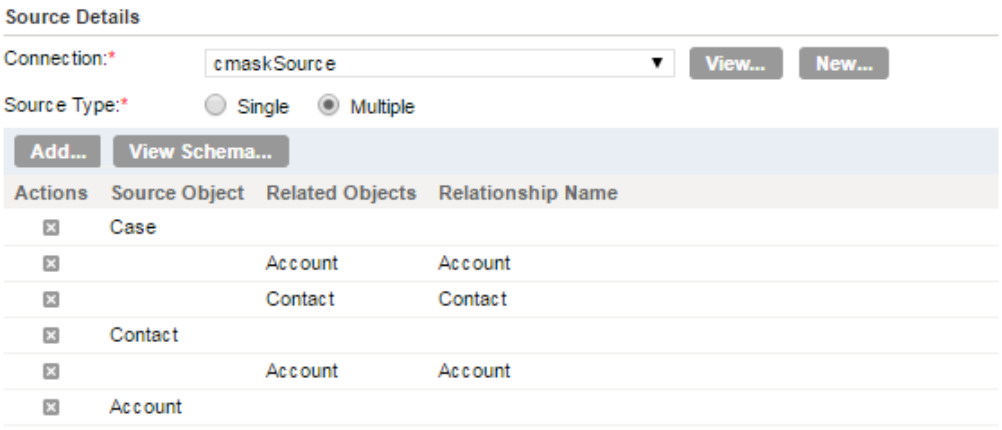
When you perform a data subset operation, the task can return the number of target subset rows that is more than the applied filter criteria.

When you run a data subset operation, the task selects the records in an object based on the filter criteria. The filtered records of the object can contain child records. The child records might have references to other records from the same parent object. In such cases, the task loads additional records along with filtered record to maintain referential integrity.

Data subset rows example

Consider that there are Account, Case, and Contact objects. Account is the parent object, and Case and Contact are the child objects of Account. The Contact object is also parent to the Case object.

The following image shows the source objects and related objects:



The following table shows the two sample cases and the respective contacts for the account ABC1:

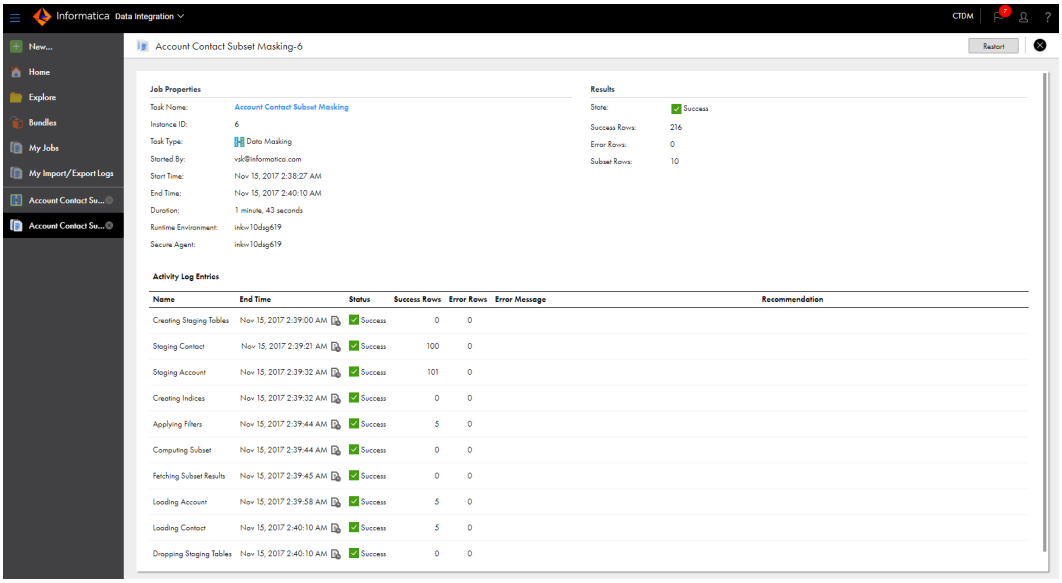
Case	Contact
00002541	Victor
00002542	Jack

Apply a subset filter on the Account object to load the target with ABC1 accounts. When you run the masking task, the task first applies filter on the account to load ABC1. The account ABC1 contains case 00002541 and case 00002542. If you enable the child record selection from the Account object to the Case object, the task loads case 00002542 and case 00002541.

The case 00002541 refers to the contact Victor, and the case 00002542 refers to the contact Jack. The parent account of Jack is XYZ2. To maintain the referential integrity for the contact Jack with the case 00002542, the task loads the additional XYZ2 account. Though you applied filter to load the account ABC1, the task loads both ABC1 and XYZ2 acounts.

The task loads six subset rows, two rows from each Account, Contact, and Case object.

The following image shows the masking task activity log entries for a data subset operation:



Refresh metadata

When you create a masking task, the task imports the source and target metadata. Over time, you might update the Salesforce objects and add or delete objects. You might also add or delete objects in the masking task.

Because of changes to Salesforce objects or objects in the task, the metadata imported when you created the task can get outdated.

If you run the same masking task at regular intervals, the metadata imported in the task might not be the latest. The masking task requires the latest metadata to define relationships between objects and to determine fields that you can mask.

A masking task might fail if it does not use the updated metadata in the Salesforce source and target.

You can refresh the metadata before you run a masking task to ensure that the source and target metadata in the task is up to date. The time required for a metadata refresh might differ based on the number of objects, latency, and the Salesforce API response time.

You can refresh the metadata in a masking task in one of the following ways:

Refresh the metadata without editing a task

When you refresh the metadata without editing the task, the refresh runs as a separate job. You cannot run an instance of a masking task and a metadata refresh of the task at the same time. If the refresh job fails at any point, the metadata does not update. So the source metadata and target metadata remain consistent.

Refresh the metadata without editing the task from the list of tasks on the **Explore** page or from the task view page. View the progress and status of the refresh job from the **My Jobs** page.

Refresh the metadata from within a masking task

You can refresh the source and target fields from within a masking task when you create or update a masking task. The **Target** page in the create task and edit task workflows includes an option to refresh

fields. You cannot view the progress of the refresh or perform other tasks during the refresh. You can continue to create or update and save the masking task after the refresh finishes. The refresh process might take some time, based on the number of objects and the size of the metadata.

Choose how you want to refresh the metadata based on the number of objects to refresh. As a best practice, if you need to update many objects, it is recommended that you refresh the metadata without editing the task. To update fewer objects or less metadata, you can edit the metadata from within the task.

Reset task

You can reset a masking task that has a different source and target and contains data filters if the task fails at any point and you want to restart the task from the first step.

A masking task with data filters performs different steps including staging data, subset computation, load to target, and drop staging tables. If a task fails at any of the steps, it continues from the point of failure when you restart the task.

For example, you configure a masking task with data filters and run the task. The task fails at the load to target step. If you restart the task at a later date, it skips the steps of data staging and subset computation. You want to rerun the task, but you want to ensure that the subset computation is accurate and includes any changes to the data. If the task skips the staging and subset computation steps, it uses the estimation and staged tables from the previous failed task run.

You might want to restart a failed task from the first step if you reinstall the staging database or the database file is corrupted and you use a different staging database. The staged files are not available on the new installed or different database. If you restart a task that failed at the load to target step, the task restarts from the same step but cannot access the previous staged tables. The task fails again. In such cases, reset the task before you restart the task. The status returns to start and when you restart the task, it stages and estimates the data again.

Choose to reset the task before you restart the task. The reset returns the task status to Start. When you restart the task, the task starts from the first step. It performs all steps of staging, subset computation, load to target, and drop staging tables, based on how you configure the task.

You can reset masking tasks that have different source and target connections and include data filters. Tasks that use the same source and target or do not include data filters do not require subset computation or staging tables.

Apply masking rules

You can apply a masking rule to a field from the **Masking** page.

You can select a rule from a list based on the data type of the source field. If some of the fields are different in source and target, the common fields are listed. The fields pick up the attributes, such as length, field type, and label, from the target connection.

When you select multiple source objects, the task lists fields from a single object at a time on the **Masking** page. Select the source objects individually if you want to apply masking rule to fields in different objects.

After you apply masking rules, you can configure masking rule properties. For each masking rule, you can configure preprocessing and postprocessing expressions.

You cannot apply masking rules to read-only objects.

Masking rule assignments

You can apply masking rules to the objects from the **Masking** page to mask the fields.

You can apply the masking rules to the objects based on the field data type. After you apply a masking rule to a field, you can configure the masking rule properties. You can either manually select the available data masking rules from the list for each field or assign the default masking rules to a set of fields at once. The masking task package contains default masking rules. To assign the default masking rules to the source objects, click **Default Assignment**.

You can clear default masking rule assignments and assign the rules manually. To delete a masking rule assignment, click **Clear Assignment**.

To use a relational dictionary in a custom substitution rule, the masking task must include the relational dictionary connection. To mask source data with unique substitution values, the task must have a storage connection. To add a relational dictionary or storage connection to the masking task, click **Configure Connections**.

Add mapplets

Add mapplets in a masking task to mask the target fields.

Use passive mapplets to perform a masking task. Assign a mapplet rule to a source object. Map the source fields to the input fields of the mapplet, and map the output fields of the mapplet to the target fields.

You can add multiple mapplets to an object. You can also add multiple instances of a mapplet to multiple objects.

You can add multiple instances of a mapplet to a single object. Informatica Cloud appends a unique number to identify each instance of the mapplet. You must configure each instance to the object before you run the task.

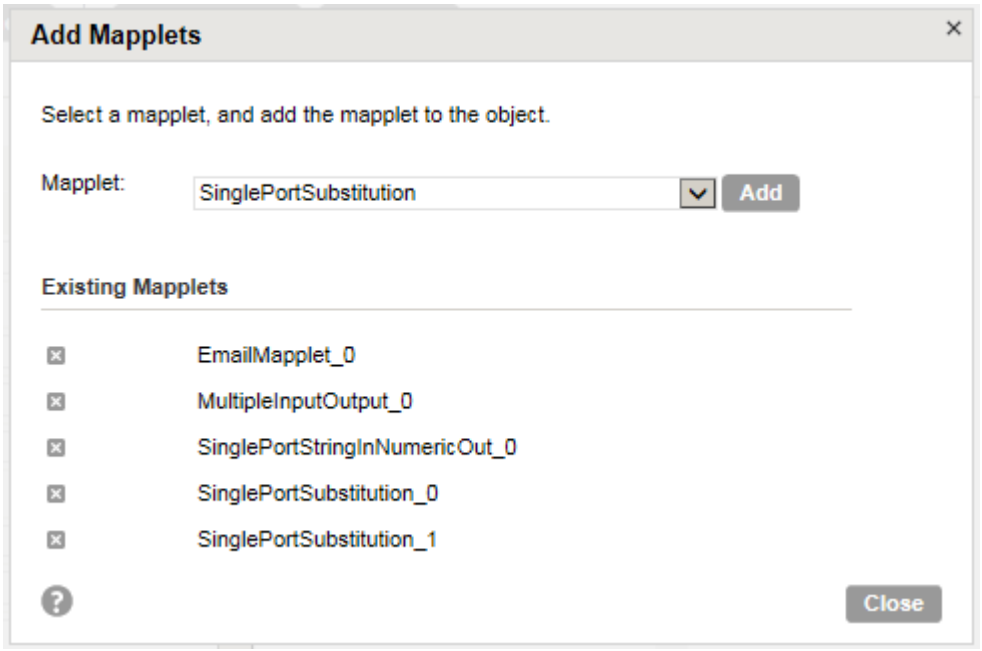
You can use a mapplet that requires an extra connection to a relational database or a flat file. Before you add the mapplet, you must add the connection.

If the dictionary information for the mapplet is in a flat file, the flat file must be present in the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\data
```

If the lookup connection for the mapplet is a flat file connection, the connection name must be the name of the flat file.

The following image shows the mapplets that you can add to a masking task:



You cannot use active mapplets.

Target fields

You can view the common and missing mandatory fields from the **Masking** page.

Common fields list the fields that are common in both source and target. You can assign masking rules to all the common fields.

The missing mandatory fields list the target mandatory fields that are missing in the source. To mask the missing mandatory fields, you can configure an expression or specify a value in the expression builder.

For example, you need to populate the target with data for testing purposes. You create a mandatory field called AlternatePhone_c in the target Account object that is not present in the source. When you run a masking task, you need to populate the missing mandatory target field with a value. The task fails because the mandatory field is missing in the source. You can enter a specific value or configure an expression for the missing mandatory field to populate the target.

Default masking rules package

You can assign default masking rules to the target fields.

The masking task package contains files with the default masking rules. After you install the Secure Agent, you can view the default_rules.xml, fields.properties, and salesforce_default_values.properties files in the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main
\dmask
```

The default_rules.xml file contains the configured rule properties for each masking rule. The fields.properties file contains the default masking rules for all the fields in the objects. When you apply default masking rules to the common fields, the task picks the default rules from the default_rules.xml and fields.properties files.

The salesforce_default_values.properties file contains the default values for the target mandatory fields that are missing in the source.

You can edit these files to change default values or create rules for default assignment.

Configure default rules parameters

You can edit the default rules files and configure the parameters for default assignments.

The following table describes the parameters that you can configure in the default_rules.xml file:

Parameter	Description
isSeeded	To configure repeatable output. Enter True or False. If you enter True, specify a seed value.
seedValue	A starting number to create repeatable output. Enter a number from 1 through 999.
keepCardIssuer	Masks a credit card number with the same credit card type. Enter True or False. If you enter False, specify the targetIssuer parameter.
targetIssuer	Masks the credit card numbers with the selected credit card type. You can enter the following credit card types: ANY, JCB, VISA, AMEX, DISCOVER, and MASTERCARD.
firstNameColumn	Name of the column to use as the first part of the email name. The email name contains the masked value of the column you choose.
firstNameLength	The maximum number of characters of the first name to include in the masked email addresses.
delimiter	Delimiter to separate the first name and the last name in masked email addresses. You can enter the following characters: . / - / _ If you do not want to separate the first name and last name in the email address, leave the delimiter blank.
lastNameColumn	Name of the column to use as the last part of the email name. The email name contains the masked value of the column you choose.
lastNameLength	The maximum number of the characters of the last name to include in the masked email address.
domainConstantValue	A domain string name to include in the masked email addresses.
useMaskFormat	Specifies if you want to use a mask format. Enter True or False.
maskFormat	Defines the type of character to substitute for each character in the source data. You can limit each character to an alphabetic numeric or alphanumeric character type. Use the following characters to define a mask format: A for alphabets, D for digits from 0 to 9, N for alphanumeric characters, X for any character, + for no masking, and R for the remaining characters in the string of any character type. R must appear at the end of the mask format.
useSrcFilter	Specifies if you want to skip masking some of the source characters. Enter True or False. If you enter True, you must specify the srcFilterOption and srcFilterStr parameters.
srcFilterOption	Defines a filter that determines which characters to mask in the source. Enter one of the following options: - Mask Only. Mask only the characters that you configure as source filter characters. - Mask All Except. Mask all characters except the characters you configure as source filter characters.

Parameter	Description
srcFilterStr	The source characters that you want to mask or the source characters that you want to skip masking. Each character is case-sensitive. Enter the source filter characters with no delimiters. For example, AaBbC.
usetargetFilter	Specifies if you want to limit the characters that can appear in the target. Enter True or False. If you enter True, you must specify the targetFilterOption and targetFilterStr parameters.
targetFilterOption	Defines a filter that determines which characters to use in target mask. Enter one of the following options: - Use Only. Limit the target to the characters that you configure as target filter characters. - Use All Except. Limits the target to all characters except the characters you configure as target filter characters.
targetFilterStr	The characters that you want to use in a mask or the characters that do not want to use in a mask, based on the values of target filter type. Each character is case-sensitive. Enter the target filter characters with no delimiters. For example, AaBbC.
useRange	Specifies whether you want to set a range for the masked data. Returns a value between the minimum and maximum values of the range depending on field precision. To define the range, configure the minimum and maximum ranges or configure a blurring range based on a variance from the original source value. You can configure ranges for string, date, and numeric data types.
minWidth	The minimum value of the range. You can specify the minimum width for date, string, and numeric data types.
maxWidth	The maximum value of the range. You can specify the minimum width for date, string, and numeric data types.
startDigit	Defines the first digit of the masked SIN.
startDigitValue	The value for the first digit of the masked SIN.
DicConn	The connection to the directory where the dictionary files are present. You must create a flat file connection with the directory that points to the dictionary files.
DicName	The dictionary that you want to select. The dictionary file must be present in the rdtmDir directory of the Secure Agent.
outputPort	The output port column from the dictionary.
useBlurring	Masks data with a variance of the source data if specify that you want to blur the target data.
blurringUnit	Unit of the date to apply the variance to. You can enter the following values: Year, Month, Day, Hour, Minute, or Second.
blurringOption	The unit of numeric blurring. Enter Fixed or Percent.
blurLow	The low boundary of the variance from the source. Enter the value for numeric and date data types.

Parameter	Description
blurHigh	The high boundary of the variance from the source. Enter the value for numeric and date data types.
expText	An expression that you can configure to mask the target data.
Preprocessing expression	An expression to define changes to make to the data before masking.
Preprocessing expression	An expression to define changes to make to the masked data before saving the data to the target.

Schedule options

You can run a masking task manually, or you can schedule the task to run at a specific time or at specified time intervals.

You must select the runtime environment that contains the Secure Agent to run the task. You can configure staging connections for a data subset operation. You can configure email notification options and advanced options before you run the task.

Email notification options

When you configure a masking task, you can set email notification options to receive the status of the task.

You can select the following email notification options:

Use the default email notification options for my organization

Use the email notification options for the organization.

Use custom email notification options for this task

Use the email notification options that you configure for the task. Use commas to separate the email addresses in a list. When you select this option, the masking task ignores email notification options for the organization.

You can send email to different addresses based on the status of the task:

- Failure Email Notification. Sends failure email notification when the task fails to complete.
- Warning Email Notification. Sends warning email notification when the task completes with errors.
- Success Email Notification. Sends success email notification when the task completes without errors.

Advanced options

You can configure the advanced options to optimize a masking task.

Configure the following advanced options:

Parameter File

Name of the file that contains the definitions and values of user-defined parameters used in the task.

Preprocessing Commands

Command to run before the task.

Postprocessing Commands

Command to run after the task completes.

Configuring a masking task

Configure a masking task to apply data masking to a source and to create a data subset.

To configure a masking task, perform the following steps:

1. Define the masking task.
2. Configure the source.
3. Configure the target.
4. Configure the data subset.
5. Define data masking rules.
6. Schedule the masking task.

When you configure a masking task, you can save your work after you enter all required properties. You can choose one of the following options:

- Save. Saves the masking task and keeps it open.
- Finish. Saves and closes the masking task.
- Cancel. Closes the masking task and discards changes made after the last save.

Prerequisites

Before you configure a masking task, perform the following tasks:

- Use API version 32.0 and above to perform the masking task.
- Set the EnableSalesForceStagingResponse flag to True in the runtime environment to improve batch processing.
- If the user that runs the task doesn't have sufficient privileges on the target, disable triggers, validations, lookup filters, and workflow rules on the target object before you run the masking task.
  If the user has the required privileges, you can disable rules and triggers from the advanced Salesforce options when you schedule the task. For information about advanced Salesforce options, see Masking Tasks with Salesforce in *Salesforce Connector*.
- Ensure that the source is synchronized with the target.
- Ensure that all the mandatory fields in the target are present in the source.
- Align the user profiles, and check the user permissions and the visibility of the objects in the target.

Step 1. Define the masking task

Create the masking task.

1. Open Data Integration and click **New** to open the **New Asset** window.

2. Select **Tasks** > **Masking Task**.
3. Click **Create**.

   The new task window opens on the **Definition** page.
4. Enter a masking task name and an optional description.
5. Click **Browse** and select a project or folder location to store the task.
6. Click **Next**.

   The **Source** page opens.

Step 2. Configure the source

To configure a source, select a connection on the **Source** page. You can edit a connection that you select. Alternatively, you can create another connection.

1. On the **Source** page, select a connection from the list of connections.

   The source must have a primary key.
2. Optional. Choose to edit the connection or create another connection.
   - To edit a connection, select the connection and click **View**. In the **View Connection** dialog box, click **Edit** and edit the connection details in the **Edit Connection** dialog box. Test the connection to verify that the connection is valid.
   - To create another connection, click **New**. In the **New Connection** dialog box, enter the connection information. Test the connection to verify that the connection is valid.
3. Choose a single source object or multiple source objects.
   - To select a single object from the list, click **Single**. Select a single source object from the list. You can preview the source object details in the **Data Preview** section.
   - To select many objects, click **Multiple**.

   A list of source objects appears.

   Note: Objects that you cannot update do not appear in the list of source objects. For example, you cannot update objects of type *isSfIdLookup()*, *isCreateable()*, *isupdateable()*, or *isreferenced()*. You cannot update objects with cyclic relationships. For example, you cannot update the User, Profile, Community, and Idea objects.
4. If you choose multiple source objects, perform the following steps:
   a. Click **Add**.

      The **Select Source Object** dialog box appears with a list of objects.
   b. Select a source object and click **Select**.
   c. Select the added object and click **Add**.

      The **Select Related Objects** dialog box appears.
   d. Select the related child, parent, or self-reference objects that you want to include in the source objects. The objects move to the list of selected objects.
   e. To add the related objects, click **Select**.

      You can view and download the schema graph in the DOT format to view the relationships between the related objects.
5. Click **Next**.

   The **Target** page opens.

Step 3. Configure the target

To configure a target, select a connection on the **Target** page. The target connection type must be the same as the source connection type.

1. On the **Target** page, you can choose to perform inplace masking or save the edited data to a different location.
   - To perform inplace masking, select **Same as source**. If you select the target same as the source, the **Connection** list is disabled. The **Same as source** check box is selected by default.
   - To save masked data to a different location, clear **Same as source**, and select a connection from the **Connection** list.

   The target object or objects are the same as the source.
2. Optional. You can choose to edit the connection or create another connection.
   - To edit a connection, select the connection and click **View**. In the **View Connection** dialog box, click **Edit** and edit the connection details in the **Edit Connection** dialog box. Test the connection to verify that the connection is valid.
   - To create another connection, click **New**. In the **New Connection** dialog box, enter the connection information. Test the connection to verify that the connection is valid.
3. From the **Task Operation** list, select the operation that you want to perform.

   On a partial sandbox, you can perform update, upsert, and insert operations. When you run an inplace masking task, you can perform an update operation. If the target is different from the source, you can perform insert and upsert operations.
4. Select a target field that can link to target records. Select an existing external ID, custom field, or unique field from the list, or create another target field. To create another target field, click **Create**. To add the target field, click **Create**.

   You can view the target field details and any errors or warnings. You can save the external IDs to perform another upsert operation.
5. If you change the source object or connection properties, click **Refresh Fields**.
6. If you want to validate the source and target fields, click **Validation Reports**.
7. Click **Next**.

   The **Data Filters** page opens.

Step 4. Configure the data subset

To configure a subset operation, use the **Data Filters** and **Relationship Behavior** options on the **Data Filters** page. Skip this step if you do not want to create a data subset.

1. On the **Data Filters** page, click **New** to create a data filter.
   - To create a simple data filter, select an object, a field to filter by, and an operator. Enter the value you want to use, and click **OK**.
   - To create an advanced data filter, click **Advanced**. Select an object. Select the field and create the filter expression. Click **OK**.
     If you create simple data filters for the same object, the simple data filters for the object merge with the advanced filter. For more information about configuring advanced data filters, see "Configuring advanced data filters" on page 14

   You can apply filters to a single object within a task.

Note: Fields that you cannot update do not appear in the filter. You cannot apply filters on the following field types: *TEXTAREA (RICH)*, *TEXTAREA (RICH)*, and *TEXT ENCRYPTED*.

2. To configure child relationships, click **Configure**.

   The **Configure Relationship Behavior** dialog box appears.

3. Enable the child records that you want to include in the subset.
4. Click **Save**.

   Join Operation displays the number of join operations that are required to create the subset. You can view and download the schema graph in the DOT format to view the relationships between the objects.

5. Optional. Click **View** to view the sequence in which a task with multiple objects selects objects to create the data subset.
6. Click **Next**.

   The **Masking** page opens.

Step 5. Define data masking rules

On the **Masking** page, choose the object and select masking rules to assign to each field in the target.

1. On the **Masking** page, select a source object to view the fields.

   The task lists the common fields and the missing mandatory fields.

2. To view information about a field in the source object, click **Status**.

   The field data type determines the masking rules that you can apply to it. Fields that you cannot mask do not display a masking rule list.

3. To use a dictionary in a masking rule or to configure unique substitution masking, click **Configure Connections** and add the relational or flat file dictionary or storage connection to the task.

   The **Is Unique** option does not appear if the task does not include a storage connection. You cannot choose a dictionary in a rule unless you add it to the masking task.

4. To assign a rule to a common field, select the rule in the **Masking Rule** list.

   If the rule you select requires additional parameters, a **Configure** button appears next to the rule.

5. To configure the masking rule properties, click **Configure**.

   Each masking rule can have different properties.

6. Configure the masking rule properties and click **Save**.

   When you select a mapplet rule, you must configure input and output fields of the mapplet.

7. To assign the default masking rules to the fields, click **Default Assignment**. To clear the masking rules assignment, click **Clear Assignment**.
8. To view and configure an expression for the mandatory fields that are missing in the source, click **Missing Mandatory Fields**.
9. In the **Actions** column, click **Configure Expression** and enter an expression in the expression builder. Click **OK**.
10. After you configure masking rules for all fields, click **Next**.

    The **Schedule** page opens.

Step 6. Schedule the masking task

Configure when to run the masking task from the **Schedule** page.

You can run the masking task manually or schedule it. You can schedule the masking task to run at a specific time or at specified time intervals.

Running the masking task immediately

You can run a masking task without scheduling it.

1. Click **Explore** to open the **Explore** page.
2. Choose to browse by projects or assets from the **Explore** list.

   Select **Assets** to view a list of all assets. Select **Projects** to view a list of all projects. You can then select a project to view the assets in the project.
3. You can run a task manually in one of the following ways:
   - Select the masking task that you want to run. Click the **Actions** icon and click **Run**.
   - Click to open the masking task that you want to run. In the task page, click **Run**.

   You can view the progress and status of the job on the **My Jobs** page. You can also view and manage jobs from the **All Jobs** or **Running Jobs** page in Monitor.

Scheduling the masking task

Configure when to run the masking task from the **Schedule** page.

1. On the **Schedule** page, choose whether to run the masking task on a schedule or run manually without any schedule.
2. If you choose to run the task on a schedule, choose a schedule from the list or click **New** to create a new schedule.
3. To run the masking task, select the runtime environment that contains the Secure Agent to run the task.

   Note: You cannot use a cloud runtime environment to run a masking task.
4. Optionally, if the task runs in a serverless runtime environment, configure serverless usage properties.
5. If you selected multiple source objects, configure the data subset options. Select a staging connection that runs the data subset operation. Configure the staging connection and start the staging database service.
6. Optionally, select **Drop Staging Tables** to drop the staging tables even if there are error rows in the task.
7. Select an email notification option.
8. Optionally, configure the advanced options.
9. Optionally, configure the advanced Salesforce options if the API version is 32.0 and higher.
10. Click **Save** to save and keep the task open, or click **Finish** to save and close the task.

Masking task maintenance

You can view and maintain all masking tasks from the **Assets** list on the **Explore** page.

You can edit tasks, copy tasks, delete tasks, run tasks, estimate a data subset, and download XML mapping and validation reports. You can also edit permissions on a masking task.

Editing a masking task

You can edit a masking task if you want to make changes to the metadata.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Edit**. You can also open the task and click the **Edit** button on the task view page.

   You can edit the task information in the **Edit Task** window.
3. Click **Finish** to save the changes.

Running a masking task manually

You can start a masking task manually to run it immediately.

1. On the **Explore** page, navigate to the masking task.
2. You can run a task manually in one of the following ways:
   - In the row that contains the task, click **Actions** and select **Run**.
   - Open the masking task that you want to run. In the task page, click **Run**.

   You can view the progress and status of the job on the **My Jobs** page. You can also view and manage jobs from the **All Jobs** or **Running Jobs** page in Monitor.

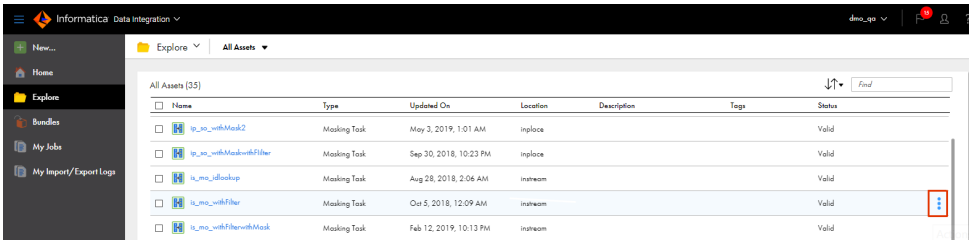
Refreshing the metadata in a masking task

You can refresh the metadata in a masking task if the metadata of the objects changes in Salesforce or you make changes to the list of objects in the masking task.

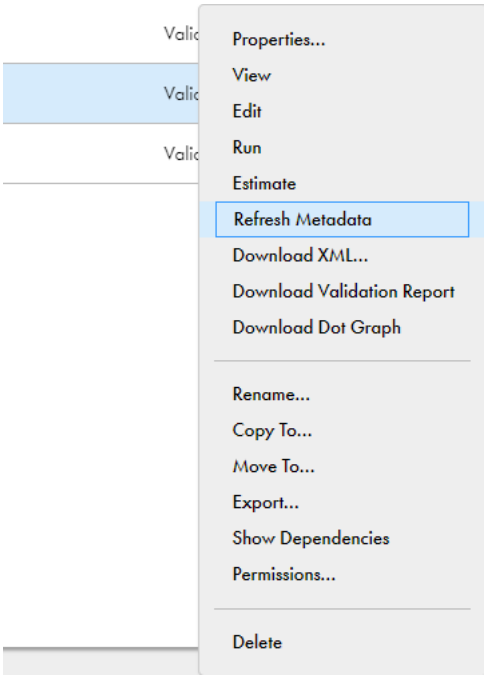
You can choose to refresh the metadata without editing a task. You can also refresh the metadata from the create or update workflow of a masking task. Choose how you want to refresh the metadata based on the number of objects to refresh.

1. On the **Explore** page, navigate to the masking task.
2. You can refresh the metadata in one of the following ways:
   - Refresh the metadata without editing the task. You can refresh the metadata without editing the task in one of the following ways:
     - In the row that contains the masking task, click **Actions** and select **Refresh Metadata**.

- Open the masking task. In the task view page, click **Actions** and select **Refresh Metadata**. The following image shows the **Actions** icon on the **Explore** page:



The following image shows the Refresh Metadata option available in the **Actions** menu:



You cannot refresh the metadata if an instance of the task is in progress. Wait for the task to finish and then refresh the metadata.

You can view the progress and status of the job on the **My Jobs** page. You can also view and manage jobs from the **All Jobs** or **Running Jobs** page in Monitor.

- Edit or update the task to refresh the metadata.

  Note: Use this option if you need to update a small number of objects.

  1. Open the masking task.
  2. Click **Edit** to open the task edit page.
  3. Click **Target** to open the **Target** page.
  4. Click **Refresh Fields**.
  5. Click **Save** and then click **Finish**.

Stopping a masking task

You can view the progress and status of a job on the **My Jobs** page. You can also view and manage jobs from the **All Jobs** or **Running Jobs** page in Monitor.

To stop a masking task that is currently running, click the **Stop** icon in the row that contains the job.

Resetting a masking task

Reset a failed masking task to ensure that the task starts from the first step when you restart the task.

Reset a failed task that includes a data filter and uses a different source and target. Tasks that do not include a data filter or use the same connection as the source and target do not require staging or subset computation.

1. On the **Explore** page, navigate to the masking task.
2. Open the masking task. In the task view page, click **Actions** and select **Reset Task**.

Configuring masking task permissions

You can configure permissions for a masking task. Set permissions for users or groups to read, create, update, delete, and run the masking task. You can also allow other users or groups to change permissions for the masking task.

By default, the user group defines the objects that a user can access.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Permissions**.

   The **Task Permissions** page opens with a list of users and groups that have permissions on the task.
3. To remove users or groups, select the users or groups and click **Remove** and then click **Save**.
4. To edit the level of permission assigned, select the user or group, change the permission level as required and click **Save**.
5. To add users or groups, click **Add**.
6. Choose the users or groups from the list of users or groups in the organization and click **Add**.

   The users or groups are added to the list.
7. Configure the required level of permission for each user or group and click **Save**.

Copying a masking task

You can copy a masking task to create another task with similar behavior.

When you copy a task, the masking task appends a number to the task name and creates a task name.

Note: When you copy a task with a schedule, the schedule is not copied.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Copy To**.
3. Choose the project or folder where you want to create a copy of the masking task and click **Select**.

   You cannot create a copy of a task in the same location.

Renaming a masking task

You can rename a masking task.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Rename**.
3. In the **Rename Asset** window, enter a name for the masking task and click **Save**.

   The name of the masking task is updated.

Deleting a masking task

You can delete a masking task at any time.

When you delete a task at run time, the masking task completes the task and then deletes it. You cannot retrieve a task after you delete it.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Delete**.
3. Click **Delete** to confirm that you want to delete the task.

Exporting a masking task

You can export a masking task. You can import the task into a different environment or use the information in an exported masking task to analyze errors.

If you export a masking task created before the April 2022 release, the exported task might not include the dictionary and storage connection information. When you import the task, the fields appear blank. To avoid the issue when you import the masking task into an environment with the April 2022 release or later, open and save the task before you export the task. The exported task then includes the dictionary and storage connection information and the fields appear in the imported task. The connections also appear on the **Uses** tab of the **Show Dependencies** page.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Export**.
3. Optional. Edit the job name if required.
4. Click **Export**.

   You can view and download the exported file from the **My Import/Export Logs** page. Click **Export** to view the export logs.

Downloading mapping XML

You can download a mapping XML file to view scope, control, and the workflows.

After you download the mapping XML, you can import the XML file into PowerCenter for debugging.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Download XML**.

Downloading validation reports

You can view validation reports after you run a masking task.

A validation report contains mandatory relationships, fields, and field and relationship mismatch between source and target.

1. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Download Validation Report**.

Dictionary files for data masking

The masking task uses a set of built-in dictionary files or the custom dictionaries that you create. When you configure a substitution masking operation, you select a dictionary that contains substitute values. The masking task performs a lookup on the dictionary selected and replaces the source data with data from the dictionary.

When you install or upgrade the Secure Agent in a runtime environment, the masking task downloads and saves the dictionary files to the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\data
```

You cannot edit or rename these files, but you can change the content within the specified file structure.

Note: The data in the dictionary files is test data.

Themasking task downloads the following dictionary files:

- informatica_mask_address.dic
- informatica_mask_cc_american_express.dic
- informatica_mask_cc_diners_club.dic
- informatica_mask_cc_american_express.dic
- informatica_mask_cc_diners_club.dic
- informatica_mask_cc_discover.dic
- informatica_mask_cc_jcb.dic
- informatica_mask_cc_master_card.dic
- informatica_mask_cc_visa.dic
- informatica_mask_countries.dic
- informatica_mask_email.dic
- informatica_mask_female_first_names.dic
- informatica_mask_first_names.dic
- informatica_mask_job_position.dic
- informatica_mask_last_names.dic
- informatica_mask_male_first_names.dic
- informatica_mask_states.dic
- informatica_mask_streets.dic
- informatica_mask_uk_ni.dic
- informatica_mask_us_telephone.dic
- informatica_mask_us_towns.dic
- informatica_mask_us_ssn.dic
- informatica_mask_us_zipcode.dic
- dict.csv
- defaultValue.xml

If you want to use a custom flat file dictionary, you must add a connection to the flat file dictionary. If there are multiple Secure Agents in a runtime environment, you must also copy the custom dictionary file to the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\data
```

You can use any of the flat file formats, such as `.txt, .dic,` and `.csv`. There is no limit on the maximum number of fields that can be present in a flat file dictionary. All the fields must have a column header, and the fields must be separated by a comma. The structure within the file must contain a sequential number and a value separated by a comma. A file can contain more than two columns. When you configure a substitution rule with custom dictionaries, you can select the dictionary column. To support non-English characters, you can use different code pages from a flat file connection when you configure a substitution rule with custom dictionaries.

The following text is a sample from a flat file dictionary that you can use to mask the credit card numbers:

```
SNO, CC_JCB
1, 3500-0003-7382-2377
2, 3500-0092-0490-3652
3, 3500-0077-6261-9918
4, 3500-0039-3695-5973
5, 3500-0089-8551-0603
6, 3500-0064-5387-7207
7, 3500-0030-0361-1582
8, 3500-0042-8477-2366
```

You can also use relational dictionaries to perform custom substitution masking. Create a connection to the relational database that contains the dictionary tables. To use the dictionary in a masking task, add the dictionary connection to the task. You can then select the required table and column to use in a masking rule.

Consistent masked output

You might want to use different tools to mask the source data.

You can use the following tools to generate the same masked output from the same source data:

- Informatica Intelligent Cloud Services
  You can use the following tools in Informatica Intelligent Cloud Services:
  - Masking tasks on Informatica Intelligent Cloud Services
  - Mapping tasks that contain mappings with the data masking transformation on Informatica Intelligent Cloud Services
  - Mappings that contain the data masking transformation on Informatica Intelligent Cloud Services
- Test Data Management (on-premise)
- PowerCenter mappings that contain the data masking transformation

Rules and guidelines

Consider the following rules and guidelines before you run mappings, tasks, or workflows to generate consistent masked output:

- Use substitution masking rules to generate consistent masked output.
- The masking rules must use the same dictionaries.
- The Repeatable option must be set to ON.
- Use the same seed value.

Substitution masking rules use values from dictionaries to create masked output. The default dictionaries on Informatica Intelligent Cloud Services and on-premise Test Data Management are the same. When you use the same substitution rule, the workflow uses the same dictionary to substitute source data. The same seed

value therefore ensures that the same substitute value is used for all rows provided the dictionaries are the same.

On Informatica Intelligent Cloud Services, the dictionary files are available at: `<Secure Agent installation directory>\apps\Data_Integration_Server\data`

In on-premise Test Data Management, the dictionary files are available at: `<Informatica installation directory>\server\infa_shared\LkpFiles`

The Repeatable option must be set to ON to ensure that the task or workflow repeats dictionary values for the same source value.

Example

Consider the following example:

The source data contains *First Name* and *Last Name* columns that you need to mask to ensure that you mask the full name in the target data.

You can use the following methods to generate the masked output:

- Run a masking task on Informatica Intelligent Cloud Services.
- Run a mapping task that contains a data masking transformation on Informatica Intelligent Cloud Services.
- Run a mapping that contains a data masking transformation on Informatica Intelligent Cloud Services
- Run a data masking plan in Test Data Management.
- Run a PowerCenter mapping that contains a data masking transformation.

Perform the following high-level tasks to generate the masked output:

1. Use the Substitution Name masking rule to mask the *First Name* column. Set the Repeatable option to ON. Enter a seed value.
2. Use the Substitution Last Name masking rule to mask the *Last Name* column. Set the Repeatable option to ON. Enter a seed value.
3. Use the default dictionaries available with the setup. Do not make changes to the dictionaries.

When you run the masking task, mapping, or mapping task on Informatica Intelligent Cloud Services, Test Data Management workflow, or PowerCenter mapping to generate output, you generate the same masked output for the same source data.

CHAPTER 8

Masking rules

A masking rule defines the logic that masks the data.

The type of masking rule that you can apply depends on the data type of the field that you need to mask. For example, if the field data type is numeric, you can define a masked value that is within a fixed or percent variance from the original value.

You can restrict the characters in a string to replace and the characters to apply in the mask. You can provide a range of numbers to mask numbers and dates.

Masking rules

You can select a masking rule for each source field on the **Masking** page that you want to mask. The rules that you can select from depend on the data type of the field that you want to mask.

The following table describes the masking rules that you can apply to source fields:

Masking rule	Description
Credit Card Masking	Masks a credit card number with a number that has a valid checksum. You can mask the string data type.
Email Masking	Masks an email address with a random email address. You can mask the string data type.
Email Advanced Masking	Masks an email address with a realistic email address from a first name, last name, and a domain name. You can mask the string data type.
IP Address Masking	Applies a built-in mask format to disguise an IP address. You can mask the string data type.
Key Masking	Creates repeatable results for the same source data. You can mask date, numeric, and string data types.
Nullification Masking	Changes input values in a field to null. You can mask date, numeric, and string data types.
Phone Masking	Masks a phone number with random numbers in the same format as the original number. You can mask the string data type.
Random Masking	Produces random, nonrepeatable results. You can mask date, numeric, and string data types.
SIN Masking	Masks a Social Insurance number. You can mask a string data type.

Masking rule	Description
SSN Masking	Masks a Social Security number with a valid number. You can mask the string data type.
Substitution Masking	Replaces a field with similar but unrelated data from a default dictionary file. You can mask the string data type.
Custom Substitution Masking	Replaces a field with similar but unrelated data from a custom flat file or relational dictionary. You can mask the string data type.
Dependent Masking	Replaces a field value with a value from a custom dictionary based on the values returned from the dictionary for another input column. You can mask the string data type.
URL Masking	Masks a URL by searching for the '://' string and masking the substring to the right of it. You can mask a string data type.
Custom Masking	Applies an expression to mask the target field. You can mask a string, numeric, and date data types.
Mapplet Masking	Applies masking rules from a PowerCenter mapplet. The mapplet contains the logic to mask the input fields and return data to the target.

Repeatable output

Repeatable masking output returns deterministic values. Use repeatable masking when you run a masking task more than once and you need to return the same masked values each time it runs.

Configure repeatable output if you have the same value in multiple source tables and you want to return the masked value in all of the target tables. The tables in the target database receive consistent masked values.

For example, customer John Smith has two account numbers, 1234 and 5678, and the account numbers are in multiple tables. The masking task always masks the name John Smith as Frank Martinez, the account number 1234 as 6549, and the account number 5678 as 3214 in all the tables.

Enter a seed value when you configure repeatable output. You can configure a dictionary file with data values that you can replace when you use substitution masking. When you configure repeatable output, the masking task returns the same value from the dictionary whenever a specific value appears in the source data.

Seed value

Apply a seed value to create repeatable output for masking output. The seed value is a starting point for generating masked values.

You can define a seed value from 1 through 999. The default seed value is 190. Apply the same seed value to a field to return the same masked data values in different source data. For example, if you have the same Cust_ID field in four tables and you want all of them to generate the output with the same masked values, apply the same seed value when you mask each field.

You can enter the seed value as a parameter. Seed value parameter names must begin with $$. You can include an underscore (_) in the name but you cannot include other special characters. Add the required parameter and value to the parameter file and specify the parameter file name at run time.

You cannot use a seed value for random and custom masking rules if the source data type is numeric.

Optimize dictionary usage

The **Optimize Dictionary Output** option increases the use of dictionary values for masking and reduces duplicate dictionary values in the target.

If you perform substitution masking or custom substitution masking, you can choose to optimize the dictionary usage. The workflow uses some values from the selected dictionary to mask source data. These dictionary values might be used for multiple entries so that all source data is masked in the target. The chances of using duplicate dictionary values reduces if you optimize dictionary usage. To optimize dictionary output, you must configure the masking rule for repeatable output.

Unique Substitution

Unique substitution masking ensures that each unique source value uses a unique dictionary value.

To mask a source value with a unique dictionary value, you can configure unique substitution masking. If a source value is masked with a specific dictionary value, then no other source value is masked with this dictionary value.

For example, the Name column in the source data contains multiple entries of `John`. If you configure repeatable masking, every entry of `John` takes the same dictionary value, such as `Xyza`. However, other source values might also be masked with the same dictionary value. A source entry `Jack` can also use the dictionary value `Xyza`. As a result, all entries of John and Jack use the same dictionary value. When you configure unique substitution masking, if all source values of `John` use the `Xyza` dictionary value, then no other source value uses the same dictionary value.

Unique substitution masking requires a storage connection for the storage tables. Storage tables contain the source to dictionary value mapping information required for unique substitution masking.

Note: If the source data contains more unique values than the dictionary, the masking fails because there are not enough unique dictionary values to mask all the source data.

Preprocessing and postprocessing expressions

When you configure masking rule properties, you can configure preprocessing and postprocessing expressions.

For each masking rule that you create, you can specify preprocessing and postprocessing expression parameters to apply changes before and after masking the data. A preprocessing expression defines the changes that you want to make to the data before masking. A postprocessing expression defines the changes that you want to make to the masked data before saving the data to the target. Use the expression builder to create an expression.

For example, the AccountName field can contain the name JOHN in uppercase and John in lowercase alphabetic characters. If you apply a masking technique and run the task, the task masks the same name with different values. If you want the same masked value in the target for both the names, you can apply preprocessing expression to convert all the uppercase alphabetic characters to lowercase. Use repeatable masking technique to mask with same values. To convert the masked values to uppercase alphabetic characters, apply postprocessing expression.

You cannot configure preprocessing and postprocessing expressions for custom masking.

Credit card masking

Credit card masking applies a built-in mask format to mask credit card numbers. You can create a masked number in the format for a specific credit card issuer.

The masking task generates a logically valid credit card number when it masks a valid credit card number. The length of the source credit card number must be from 13 through 19 digits. The input credit card number must have a valid checksum based on credit card industry rules.

The first six digits of a credit card number identify the credit card issuer. You can keep the original credit card issuer or you can select another credit card issuer to appear in the masking results.

The following examples of credit card numbers are in valid formats:

4567893453452

4567 8934 5345 2

4567-8934-5345-2

4567-8934-5345-2657

4567 8934 5345 2657

The following examples of credit card numbers are not in valid formats:

4563 x567 5674 7432

4563#7587 4666 9876

If the source does not contain a valid format, the task replaces with default values from the `defaultValue.xml` file.

Credit card parameters

You can mask a credit card with the same type of credit card number or you can replace the credit card number with a different type of credit card number.

The following table describes the parameters that you can configure for credit card masking:

Parameter	Description
Repeatable	Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190. You can enter the seed value as a parameter.
Keep Issuer	Returns the same credit card type for the masked credit card. For example, if the source credit card is a Visa card, generate a masked credit card number that is the Visa format.
Mask Issuer	Replaces the source credit card type with another credit card type. When you disable Keep Issuer, select which type of credit card to replace it with. You can choose credit cards such as AMEX, VISA, and MASTERCARD. Default is ANY.

Email masking

When you mask an email address, you can create a random email address or a realistic email address. Create a random email address by default.

When you configure email masking, you return random ASCII characters in the email address by default. For example, when you mask Georgesmith@yahoo.com, the results might be KtrIupQAPyk@vdSKh.BIC.

The source input data can contain alphanumeric characters, and the following special characters: _, ., and @

The following examples of email addresses are valid formats:

David_john.Waugh@gmail.com

DavidJohn1988@yahoo.com

The following examples of email addresses are not in valid formats:

david@john@gmail.com

david$%^&john@gmail.com

If the source does not contain a valid format, the task replaces with default values from the `defaultValue.xml` file.

You can configure email masking as repeatable between tasks, or you can configure email masking to be repeatable in one task. To return realistic email addresses, configure advanced email masking.

Advanced email masking

You can create a realistic email address from a first name, last name, and domain name.

When you configure advanced email masking, you can configure parameters to mask the user name and the domain name in the email address. For example, a source table might contain columns called First_Name and Last_Name. You can configure the email address to contain the first character of First_Name and seven characters of the last name. Define a domain name for the email address. The Masking task creates an address with the following syntax:

```
VSingh@mycompany.com
```

The following table describes the parameters you can configure for advanced email masking:

Parameter	Description
Repeatable	Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190. You can enter the seed value as a parameter.
First Name	Name of the column to use as the first part of the email name. The email name contains the masked value of the column you choose.
First Name Length	The number of characters in the first name to include in the email address.
Delimiter	Delimiter, such as a dot, hyphen, or underscore, to separate the first name and last name in the email address. If you do not want to separate the first name and last name in the email address, leave the delimiter blank.

Parameter	Description
Last Name	Name of the masked column to use in the email name. The email name contains the masked value of the column you choose.
Last Name Length	The number of characters in the last name to include in the email address.
Domain Name	A string value that represents an Internet Protocol (IP) resource such as gmail.com.

IP address masking

You can mask an IP address as another IP address.

The masking task masks an IP address as another IP address by splitting it into four numbers, separated by a period. The first number is the network. The masking task masks the network number within the network range.

The masking task masks a Class A IP address as a Class A IP Address and a 10.x.x.x address as a 10.x.x.x address. The masking task does not mask the class and private network address. For example, the masking task can mask 11.12.23.34 as 75.32.42.52. and 10.23.24.32 as 10.61.74.84.

You can configure repeatable output when you mask IP addresses. You must select **Repeatable** and enter a seed value.

Key masking

Key masking produces repeatable results for the same source data.

When you configure a field for key masking, the masking task creates a seed value for the field. The masking task uses the seed to create repeatable masking for the same source field values. Mask date, numeric, and string data types with key masking.

Key string masking

Configure key string masking to mask all or part of a string. To limit the masking output to certain characters, specify a mask format and result string replacement characters. If you need repeatable output, specify a seed value.

The following table describes the parameters that you can use with key masking:

Masking parameter	Description
Repeatable	Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190. You can enter the seed value as a parameter.

Masking parameter	Description
Mask Format	The type of character to substitute for each character in the source data. You can limit each character to an alphanumeric, numeric, or character type.
Filter Source	Determines whether to skip masking some of the source characters. Configure the Source Filter Type and the Source Filter Chars parameters when you enable this option. Default is disabled.
Source Filter Type	A filter that determines which characters to mask in the source. Use with the Source Filter Chars parameter. You must enable the Filter Source parameter to configure this parameter. Choose one of the following options: - Mask Only. Mask only the characters that you configure as source filter characters. - Mask All Except. Mask all characters except the characters you configure as source filter characters.
Source Filter Chars	The source characters that you want to mask or the source characters that you want to skip masking. Each character is case-sensitive. Enter the source filter characters with no delimiters. For example, `AaBbC`.
Filter Target	Determines whether to limit the characters that can appear in the target. Configure the Target Filter Type and the Target Filter Chars parameters when you enable this option. Default is disabled.
Target Filter Type	A filter that determines which characters to use in the target mask. Use with the Target Filter Chars parameter. You must enable the Filter Target parameter to configure this parameter. Choose one of the following options: - Use Only. Limit the target to the characters that you configure as target filter characters. - Use All Except. Limits the target to all characters except the characters you configure as target filter characters.
Target Filter Chars	The characters that you want to use in a mask or the characters that do not want to use in a mask, based on the values of target filter type. Each character is case-sensitive. Enter the target filter characters with no delimiters. For example, `AaBbC`.

Mask format

Configure a mask format to limit each character in the output field to an alphabetic, numeric, or alphanumeric character.

If you do not define a mask format, the masking task replaces each source character with any character. If the mask format is longer than the input string, the masking task ignores the extra characters in the mask format. If the mask format is shorter than the source string, the masking task does not mask the characters at the end of the source string.

When you configure mask a mask format, you must configure the source filter characters or target filter characters that you want to use the mask format with.

The following table describes mask format characters:

Character	Description
A	Alphabetical characters. For example, ASCII characters a to z and A to Z.
D	Digits. From 0 through 9.
N	Alphanumeric characters. For example, ASCII characters a to z, A to Z, and 0-9.

Character	Description
X	Any character. For example, alphanumeric or symbol.
+	No masking.
R	Remaining characters. R specifies that the remaining characters in the string can be any character type. R must appear as the last character of the mask.

Source filter characters

Configure source filter characters to choose the characters that you want to mask.

When you set the a character as a source filter character, the character is masked every time it occurs in the source data. The position of the characters in the source string does not matter, and you can configure any number of characters. If you do not configure source filter characters, the masking replaces all the source characters in the field.

The source filter are case sensitive. The masking task does not always return unique data if the number of source string characters is fewer than the number of result string characters.

Target filter characters

Configure target filter characters to limit the characters that can appear in a target column.

Masking replaces characters in the target with the target filter characters. For example, enter the following characters to configure each mask to contain the uppercase alphabetic characters A through F:

```
ABCDEF
```

To avoid generating the same output for different input values, configure a wide range of substitute characters, or mask only a few source characters. The position of each character in the string does not matter.

Key numeric masking

You can configure key masking for numeric values and generate deterministic output.

When you configure a field for key numeric masking, you can select a seed value for the field. When the masking task masks the source data, it applies a masking algorithm that requires the seed. You can change the seed value for a field to produce repeatable results if the same source value occurs in a different field.

Key date masking

Key date masking produces repeatable results for the same source date. Date masking always generates valid dates.

You can change the seed value for a field to return repeatable datetime values between the fields.

Key date masking can mask dates between the years 1753 and 9999. If the source year is in a leap year, the masking task returns a year that is also a leap year. If the source month contains 31 days, the masking task returns a month that has 31 days. If the source month is February, the masking task returns February.

When you perform key date masking on Salesforce data, the masking task can insert dates up to 4000. If the masked date value is more than 4000, the masking task fails and throws exception.

Nullification masking

Nullification masking changes field input values to null.

You can mask String, Date, and Numeric data types. The appropriate masking type appears based on the field type. For example, if the field type is Date, the Date Nullification masking type appears.

You can perform insert operations in nullification masking. You cannot perform upsert and update operations in nullification masking. Use custom masking with upsert and update operations to mask the field input values with null values.

Phone number masking

You can mask phone numbers with random numbers.

The masking task masks a phone number without changing the format of the original phone number. For example, the masking task can mask the phone number `(408) 382-0658` as `(607) 256-3106`.

The source data can contain numbers, spaces, hyphens, and parentheses. The following examples are valid formats of input phone numbers:

08040208950

080-4020-8950

(080)-4020-3797

(080)-(4020)-(3797)

(080)-4020-3797

(080) 4020 (3797)

Alphabetic and special characters are not masked. The following examples of input phone numbers are not in valid formats:

x80-4020-8950

x80-4020-x768

x80-4020/789

To support additional formats, you need to make changes to the Data Masking transformation. If the input data contains alphabetic and special characters, the task replaces with default values from the defaultValue.xml file.

You can configure repeatable output when you mask phone numbers. You must select **Repeatable** and enter a seed value.

Random masking

Random masking produces random, non-repeatable results for the same source data and masking rules.

Random masking does not require a seed value. The results of random masking are non-deterministic. Use random masking to mask string, numeric, and date data types.

Random string masking

Configure random masking to generate random output for string data types.

To configure limitations for each character in the output string, configure a mask format.

The following table describes the parameters that you can use with random string masking:

Masking parameters	Descriptions
Mask Format	The type of character to substitute for each character in the source data. You can limit each character to an alphanumeric, numeric, or character type.
Filter Source	Determines whether to skip masking some of the source characters. Configure the Source Filter Type and the Source Filter Chars parameters when you enable this option. Default is disabled.
Source Filter Type	Defines a filter that determines which characters to mask in the source. Use with the Source Filter Chars parameter. You must enable the Filter Source parameter to configure this parameter. Choose one of the following options: - Mask Only. Mask only the characters that you configure as source filter characters. - Mask All Except. Mask all characters except the characters you configure as source filter characters.
Source Filter Chars	The source characters that you want to mask or the source characters that you want to skip masking. Each character is case-sensitive. Enter the source filter characters with no delimiters. For example, `AaBbC`.
Filter Target	Determines whether to limit the characters that can appear in the target. Configure the Target Filter Type and the Target Filter Chars parameters when you enable this option. Default is disabled.
Target Filter Type	Defines a filter that determines which characters to use in the target mask. Use with the Target Filter Chars parameter. You must enable the Filter Target parameter to configure this parameter. Choose one of the following options: - Use Only. Limit the target to the characters that you configure as target filter characters. - Use All Except. Limits the target to all characters except the characters you configure as target filter characters.
Target Filter Chars	The characters that you want to use in a mask or the characters that do not want to use in a mask, based on the values of target filter type. Each character is case-sensitive. Enter the target filter characters with no delimiters. For example, `AaBbC`.

Random numeric masking

To mask numeric data, you configure a range of output values for a field.

The masking task returns a value between the minimum and maximum values of the range depending on field precision. To define the range, configure the minimum and maximum ranges or a blurring range based on a variance from the original source value.

The following table describes the parameters that you can configure for random masking of numeric data:

Masking parameter	Description
Range	A range that you want to set for the numeric data. Select the check box to enter a minimum and maximum range.
Minimum Range	The minimum value of the range.
Maximum Range	The maximum value of the range. The maximum value must be greater than the minimum value.
Blurring	A range of output values that are within a fixed variance or percent variance of the source data. Select the check box to enter blurring details.
Blurring Option	The unit of blurring. Select Fixed or Percent. Default is Fixed.
Low Bound	The low boundary of the variance from the source number.
High Bound	The high boundary of the variance from the source number.

Numeric blurring

To blur a numeric source value, select a fixed or percent variance, a high bound, and a low bound. The high and low bounds must be greater than or equal to zero.

The following table lists the masking results for blurring range values when the input source value is 66:

Blurring type	Low	High	Result
Fixed	0	10	Between 66 and 76
Fixed	10	0	Between 56 and 66
Fixed	10	10	Between 56 and 76
Percent	0	50	Between 66 and 99
Percent	50	0	Between 33 and 66
Percent	50	50	Between 33 and 99

Random date masking

Random date masking produces random, non-repeatable results for the same source date.

You can assign a minimum and a maximum date for the results. You can also configure blurring to define a variance limit for the date results.

The following table describes the parameters that you can configure for random date masking:

Masking parameter	Description
Range	A range that you want to specify for the date data. Select the check box to set minimum and maximum values for a datetime value.
Minimum Range	The minimum value to return for the selected datetime value. The default datetime format is MM/DD/YYYY HH24:MI:SS.
Maximum Range	The maximum value to return for the selected datetime value. The maximum datetime must be later than the minimum datetime.
Blurring	Mask a date as a variance of the source date.
Blurring Unit	Unit of the date to apply the variance to. Select the year, month, day, or hour. Default is year.
Low Bound	The low boundary of the variance from the source date.
High Bound	The high boundary of the variance from the source date.

Date blurring

To blur a datetime source value, select a unit of time to blur, a high bound, and a low bound. You can select year, month, day, or hour as the unit of time. By default, the blur unit is year.

For example, to restrict the masked date to a date within two years of the source date, select year as the unit. Enter two as the low and high bound. If a source date is 02 February, 2006, the masking task returns a date between 02 February, 2004 and 02 February, 2008.

SIN masking

SIN masking applies a built-in mask format to change Social Insurance numbers.

You can mask a Social Insurance number that is nine digits. The digits can be delimited by any set of characters. The following delimiters are valid:

```
space, no space, #, +, -, *, =, ~, !, @, $, %, ^, &, *, :, ;, ", ., /, and ,
```

If the number contains no delimiters, the masked number contains no delimiters. Otherwise the masked number has the following format:

```
xxx-xxx-xxx
```

The following examples of Social Insurance numbers are valid formats:

123456789

123 456 789

123-456-789

If the source does not contain a valid format, the task replaces with default values from the `defaultValue.xml` file.

You can define the first digit of the masked SIN. Enable **Start Digit** and enter the digit. The masking task creates masked SIN numbers that start with the number that you enter. You can configure repeatable

masking for Social Insurance numbers. To configure repeatable masking for SIN numbers, select **Repeatable** and enter a seed value.

The following table describes the parameters you can configure for SIN masking:

Masking parameter	Description
Repeatable	Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190. You can enter the seed value as a parameter.
Start Digit	When enabled, you can define the first digit of the masked SIN.
Start Digit Value	The value for the first digit of the masked SIN.

SSN masking

SSN masking applies a built-in mask format to change Social Security numbers.

The SSN masking accepts any SSN format that contains nine digits. You can delimit the digits with any characters. The following delimiters are valid:

```
space, no space, #, +, -, *, =, ~, !, @, $, %, ^, &, *, :, ;, ", ., /, and ,
```

For example, the SSN masking rule accepts the following format:

```
+=54-*9944$#789-,*()”
```

The following examples of Social Security numbers are valid formats:

123456789

123 45 6789

123-45-6789

If the source does not contain a valid format, the task replaces with default values from the `defaultValue.xml` file.

You can configure repeatable masking for Social Security numbers. You must select **Repeatable** and enter a seed value.

The masking task cannot return all unique Social Security numbers because it does not return valid Social Security numbers that the Social Security Administration has issued.

Substitution masking

Substitution masking replaces a column of data with similar but unrelated data from a dictionary.

The masking task provides dictionaries that contain sample data for substitution masking. When you configure substitution masking, select the dictionary that contains the type of substitute values that you

need. The masking task performs a lookup on the dictionary that you choose and replaces source data with data from the dictionary.

You can substitute data with repeatable or non-repeatable values. When you choose repeatable values, you must configure a seed value to substitute data with deterministic results.

Substitution masking parameters

To perform substitution masking, you can select one of the dictionaries that the masking task provides.

The following table describes the parameters that you can configure for substitution masking:

Parameter	Description
Repeatable	Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190. You can enter the seed value as a parameter.
Optimize Dictionary Usage	Increases the usage of masked values from the dictionary. Available if you choose the **Repeatable** option. The property is not applicable if you enable unique substitution.
Is Unique	Applicable for repeatable substitution. Replaces the target column with unique dictionary values for every unique source column value. If there are more unique values in the source than in the dictionary file, the data masking operation fails. Default is nonunique substitution.
Preprocessing Expression	Configure an expression in the rule to convert characters before the masking rule runs. For example, you might want to convert all characters to the same case before masking.
Postprocessing Expression	Configure an expression in the rule to convert characters in the masked output before the data is copied to the target.

Custom substitution masking

You can use custom dictionaries when you perform substitution masking. Create relational or flat file dictionaries to mask data with values from dictionaries other than the default dictionaries.

Create and add a flat file or relational dictionary connection to the masking task. For flat file dictionaries, you must create a flat file connection with the directory that points to the dictionary files. Add the connection to the relational or flat file dictionary from the **Configure** | **Connections** view.

When you configure a masking task, you can use the flat file or relational dictionary connection to perform custom substitution masking.

You can substitute data with repeatable or nonrepeatable values. When you choose repeatable values, the masking task produces deterministic results for the same source data and seed value. You must configure a seed value to substitute data with deterministic results. You can substitute more than one column of data with masked values from the same dictionary row.

You can configure the custom substitution masking rule to replace the target column with unique masked values for every unique source column value. To configure unique substitution masking, you must create a storage connection for the storage tables. Storage tables contain the source to dictionary value mapping information required for unique substitution masking.

When you configure the custom substitution masking rule, select the dictionary type, the connection, and then select the required dictionary file or table. You can then select the required column from the dictionary. To support non-English characters, you can use different code pages from a flat file connection.

The flat file connection code page and the Secure Agent system code page must be compatible for the masking task to work.

Custom substitution masking parameters

To perform custom substitution masking, select a custom dictionary that you create.

The following table describes the parameters that you can configure for substitution masking:

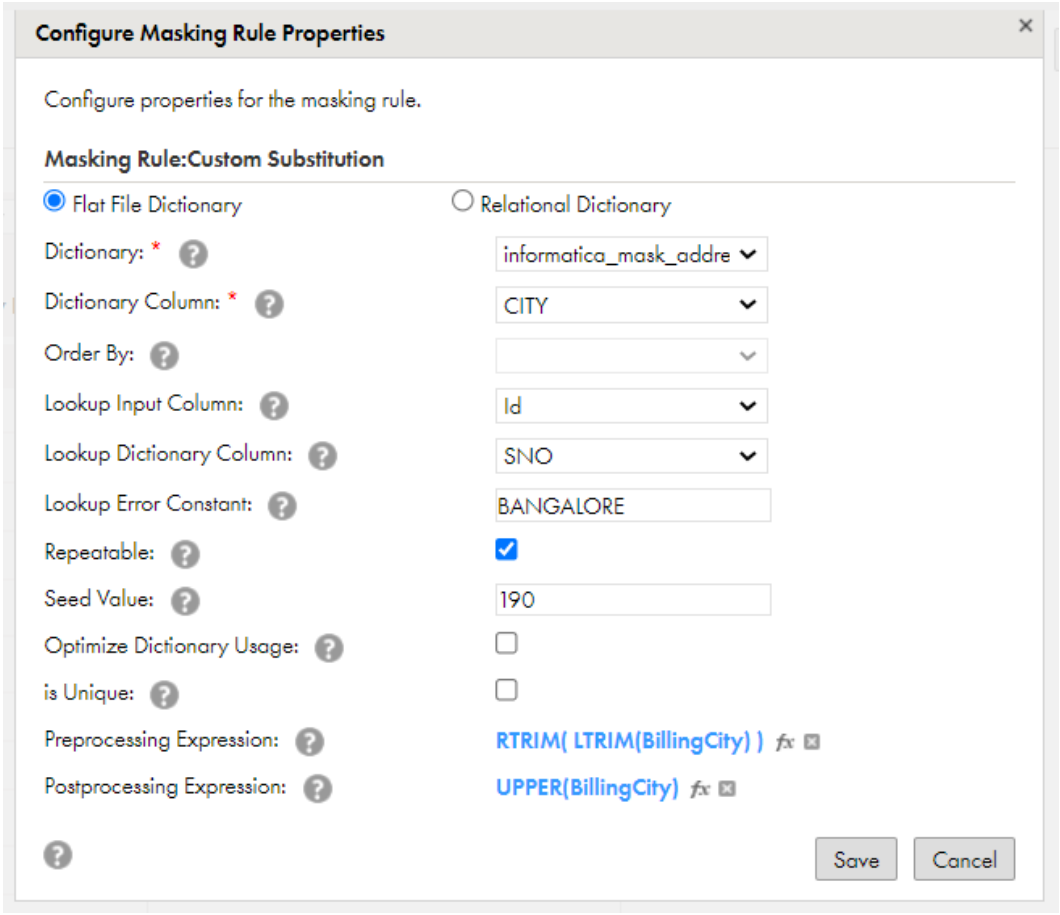
Parameter	Description
Flat File Dictionary Relational Dictionary	Choose the type of custom dictionary to use.
Dictionary	**Select the required dictionary from the list.** You must have added the dictionary connection to the masking task. The list includes relational or flat file dictionaries depending on what you choose. For flat file dictionaries, the dictionary file must be present for all the Secure Agents in a runtime environment in the following location: `<Secure Agent installation directory>\apps\Data_Integration_Server\data`
Dictionary Column	The output column from the custom dictionary. For a flat file dictionary, you can select a dictionary column if the flat file contains column headers.
Order By	Applicable for relational dictionaries. The dictionary column on which you want to sort entries. Specify a sort column to generate deterministic results even if the order of entries in the dictionary changes. For example, if you move a relational dictionary and the order of entries changes, sort on the serial number column to consistently mask the data. **Note:** The column that you choose must contain unique values. Do not use columns that can contain duplicate values to sort the data.
Lookup Input Port	Optional. The source input column on which you perform a lookup operation with the dictionary.
Lookup Dictionary Port	Required if you enter a lookup Input Column value. The dictionary column to compare with the input port. The source is replaced with values from the dictionary rows where the Lookup Input and Lookup Dictionary values match.
Lookup Error Constant	Optional. A constant value that you can configure when there are no matching values for the lookup condition from the dictionary. Default is an empty string.
Repeatable	Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190. You can enter the seed value as a parameter.
Optimize Dictionary Usage	Increases the usage of masked values from the dictionary. Available if you choose the **Repeatable** option. The property is not applicable if you enable unique substitution.

Parameter	Description
Is Unique	Applicable for repeatable substitution. Replaces the target column with unique dictionary values for every unique source column value. If there are more unique values in the source than in the dictionary file, the data masking operation fails. Default is nonunique substitution.
Preprocessing Expression	Configure an expression in the rule to convert characters before the masking rule runs. For example, you might want to convert all characters to the same case before masking.
Postprocessing Expression	Configure an expression in the rule to convert characters in the masked output before the data is copied to the target.

Custom substitution lookup example

Consider that you apply substitution masking on the S_City column and you select a dictionary file with city names, identification numbers, and serial numbers. Select CITY as the dictionary column. The lookup input port is Id and the lookup dictionary port is SNO. If there are no matching values between the Id and SNO columns, the task uses the error constant BANGALORE as the lookup value.

The following image shows the substitution parameters for masking with custom dictionaries:



Custom substitution dictionary lookup use cases

The task performs dictionary lookup in custom substitution masking in the following cases:

- Case 1. If there are valid target lookup records in a dictionary for all the corresponding source records, the task picks all the values from the dictionary and replaces in the target.
- Case 2. If there are some records in the source for which there are multiple lookup values in a dictionary, the task picks one of the lookup values from the dictionary and substitutes with the source value.
- Case 3. if some of the source values are same as the lookup values in a dictionary, the target contains the same data as the source.
- Case 4. If the source records do not have a lookup value in a dictionary and if you specify a valid error constant, the task uses the error constant for all the failed lookup conditions.
- Case 5. If the source records do not have a lookup value in a dictionary and if you do not specify a valid error constant, the task fails and generates an exception.

Dependent masking

Dependent masking replaces a column of data with values from a custom dictionary that you use to mask data in another column. To use dependent masking, at least one other source column must be masked with a custom substitution rule.

For example, mask a Name column in the source data with a custom substitution rule. Configure the rule to mask the values with values from the Name column in a Personal_Information dictionary.

You can configure dependent masking on another column to mask the source with values from a corresponding column in the same dictionary. For example, apply dependent masking on the Age column. Choose the Name column as the dependent column. You can then select a corresponding column from the Personal_Information dictionary as the dependent output column. If you select the Age column from the dictionary, the masking rule uses the age value that corresponds to the name value.

Dependent masking parameters

To apply dependent masking on a source column, at least one column must be masked with a custom substitution rule.

The following table describes the parameters that you can configure for dependent masking:

Property	Description
Dependent Column	The input column configured for custom substitution masking that you want to relate to the source column. Choose a column from the list. Columns that you configure with substitution masking appear in the list.
Dependent Output Column	The dictionary column to use to mask the source data column. Lists the columns in the dictionary used to mask the dependent column. Choose the required column from the list of dictionary columns.

URL masking

You can configure URL masking to mask a source URL address.

The masking task parses a URL by searching for the '://' string and parsing the substring to the right of it. The source URL must contain the '://' string. The source URL can contain numbers and alphabetic characters.

The task cannot mask a URL without protocols, such as `http://`, `https://`, and `ftp://`. If the source does not contain a protocol, the task replaces with default values from the `defaultValue.xml` file.

The masking task does not mask the protocol of the URL. For example, if the URL is `http://www.yahoo.com`, the masking task can return `http://MgL.aHjCa.VsD/`. The masking task can generate a URL that is not valid.

You can configure repeatable output when you mask a URL address. You must select **Repeatable** and enter a seed value.

Note: The masking task always returns ASCII characters for a URL.

Custom masking

Custom masking applies an expression to mask the target data. Use custom masking to mask string, numeric, and date data types.

When you apply custom masking to a field, click **Configure** and enter the expression. You can select the source fields, operators, and functions to build an expression. When you select a function, you can view the function description and the syntax.

You can concatenate data from multiple source fields to create a masked value for the target field. For example, you need to create a login name. The source has FirstName and LastName fields. Use substitution masking to mask the first and last names. In the Login field, configure an expression to concatenate the first letter of the first name with the last name:

```
CONCAT(SUBSTR(FirstName,1,1),LastName)
```

To mask field input values with null values, use custom masking. In the expression builder, enter single quotes separated by a space in the following format: `' '`

For more information about configuring expressions, see "Field expressions" on page 17

Mapplet masking

You can assign a mapplet rule to the source fields to mask the output target fields.

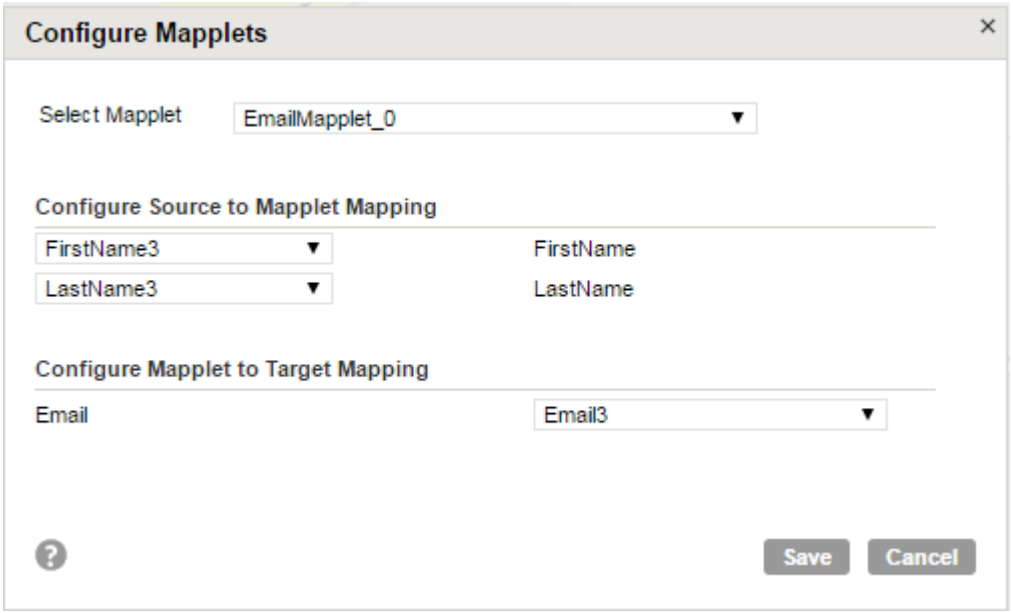
A mapplet can contain multiple input and multiple output ports. A task fails if you do not configure any of the mapplet input or output ports that you add to a source object.

After you add a mapplet rule and assign the rule to a field, you must configure the mapplet parameters. Map the source fields to the input fields of the mapplet, and map the output fields of the mapplet to the target fields.

For example, an email mapplet contains the logic to concatenate the first name and last name of the source object to generate an email ID. Apply the email mapplet masking rule to the source fields. Map the FirstName3 input field of the source to the FirstName field of the mapplet. Map the LastName3 input field of

the source to the LastName field of the mapplet. Map the Email output field of the mapplet to the Email3 field of the target.

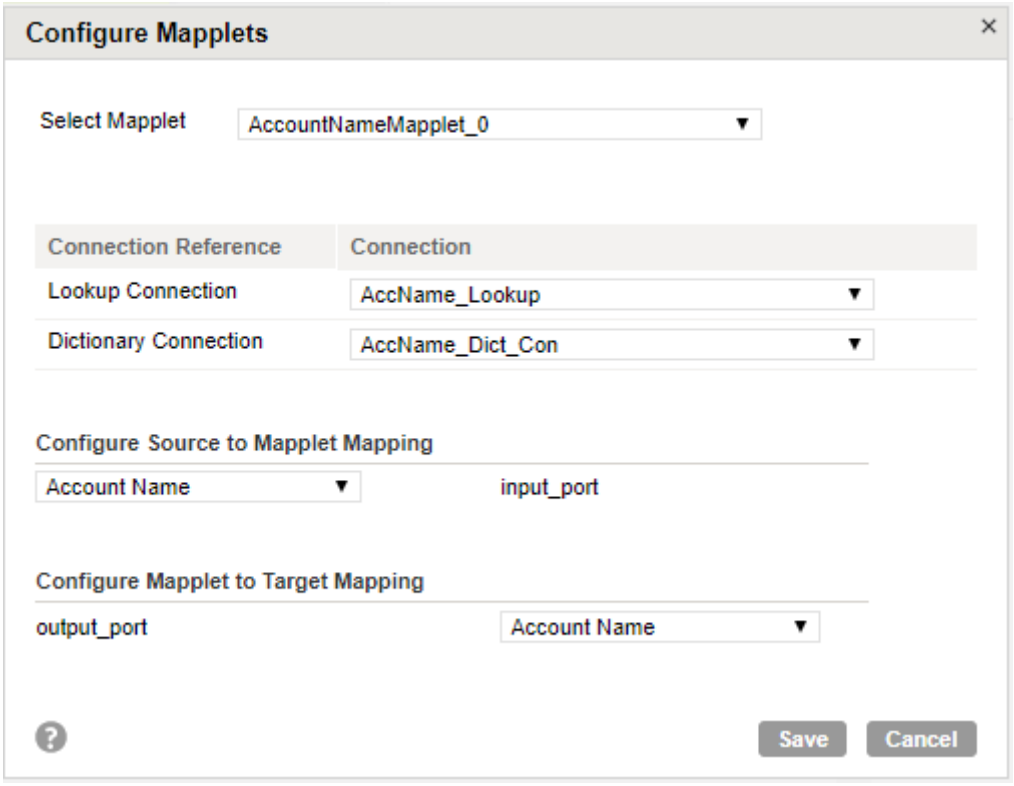
The following image shows the mapplet parameters that you can configure:



You can use a passive mapplet that requires an extra connection to a relational database or a flat file. For example, mapplets that contain an SQL transformation, lookup transformation, or a data masking transformation that uses a dictionary connection. Before you add the mapplet, you must create the connection. When you configure a mapplet that requires an extra connection, you must configure the dictionary, SQL, or lookup connections. You select the connection reference based on the type of connection that the mapplet contains.

For example, you want to mask an account name with an AccountNameMapplet mapplet and the mapplet has connections to a dictionary and a relational database. After you add the mapplet and the connections, configure and assign the mapplet to the target. After you select the AccountNameMapplet mapplet, select the AccName_Lookup connection to perform the lookup operation. Select the AccName_Dict_Con connection to read the values from the dictionary connection. Map the Account Name input source field to the input mapplet field. Map the mapplet output port to the Account Name target field.

The following image shows the mapplet that contains the dictionary and lookup connections:



If the dictionary information for the mapplet is in a flat file, the flat file must be present in the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\data
```

If the lookup connection for the mapplet is a flat file connection, the connection name must be the name of the flat file.

CHAPTER 9

PowerCenter tasks

Use the PowerCenter task to run a PowerCenter session in Data Integration.

To run a PowerCenter session in Data Integration, you create a workflow for the session in the PowerCenter Workflow Manager. You create a PowerCenter task in Data Integration. When you configure the task, you upload the PowerCenter XML file that contains the workflow.

If you want to make any changes to a session that is used in a PowerCenter task, you need to make the changes in PowerCenter. You can export the revised PowerCenter XML file and then edit the PowerCenter task to upload the updated XML file.

PowerCenter workflows

To use a PowerCenter workflow for a PowerCenter task, the workflow objects must be objects that the Data Integration PowerCenter task supports.

Consider the following rules when you use a PowerCenter workflow for a PowerCenter task:

- The PowerCenter XML file must only contain one workflow.
- The workflow must contain one Session task with one mapping.
- The workflow cannot include task types other than Session tasks.
- Do not edit the XML file after you've exported the workflow from PowerCenter. Instead, change the workflow in PowerCenter and then export it again.
- The session can contain up to 64 partitions for sources and targets.
- The session can use pre-session and post-session commands.
- The session must have the Enable High Precision session property enabled.
- The mapping must contain a source definition and target definition.
- The mapping cannot contain an IIF expression with values of different data types, such as the following IIF expressions:

```
[IIF(ANNUALREVENUE >0,NAME)]
[IIF(emplid_offset = 'Y',LINE_NO + 1,LINE_NO]
```

- The mapping cannot include reusable objects such as reusable transformations or shortcuts because Data Integration doesn't use a repository like PowerCenter does, so reusable objects cannot be stored.

Supported transformations and mapping objects

The mapping objects in a workflow must be supported by Data Integration.

A mapping can include the following source and target types:

- Flat file
- FTP/SFTP
- Database
- Salesforce
- SAP
- Web service
- Most add-on connectors

To find out if the add-on connector you use supports PowerCenter tasks, see the help for the appropriate Data Integration connector.

A mapping can include the following transformations:

- Aggregator transformation
- Data Masking transformation
- Expression transformation
- Filter transformation
- HTTP transformation
- Java transformation
- Joiner transformation
- Lookup transformation
- Normalizer transformation
- Router transformation
- Salesforce Lookup transformation
- Salesforce Picklist transformation
- Salesforce Merge transformation
- SAP IDOC Interpreter transformation
- SAP IDOC Prepare transformation
- Sequence Generator transformation
- Sorter transformation
- Stored Procedure transformation
- Transaction Control transformation
- Union transformation
- Update Strategy transformation
- Web Services Consumer transformation
- XML Parser transformation with file or database sources
- XML Generator transformation with file or database sources

If the workflow contains transformations or mapping objects other than the objects listed above, the workflow upload to Data Integration might fail.

Exception handling in stored procedures

When a mapping that you want to use in a PowerCenter task contains a Stored Procedure transformation, the stored procedure must include exception handling. Exception handling can be as complex as necessary. Or, you can use the following simple example:

```
Exception
when NO_DATA_FOUND
then NULL;
END;
```

For example, you have the following stored procedure in a PowerCenter workflow:

```
CREATE OR REPLACE PROCEDURE SP_GETSAL_WITH_EXCEPTION (EMP_ID NUMBER, EMP_NAME OUT
VARCHAR, SAL OUT NUMBER)
  AS
  BEGIN
  SELECT EMPNAME INTO EMP_NAME FROM EMPLOYEE WHERE EMPID=EMP_ID;
SELECT SALARY INTO SAL FROM EMPLOYEE WHERE EMPID=EMP_ID;
```

Before you export the workflow, add exception handling as follows:

```
CREATE OR REPLACE PROCEDURE SP_GETSAL_WITH_EXCEPTION (EMP_ID NUMBER, EMP_NAME OUT
VARCHAR, SAL OUT NUMBER)
  AS
  BEGIN
  SELECT EMPNAME INTO EMP_NAME FROM EMPLOYEE WHERE EMPID=EMP_ID;
SELECT SALARY INTO SAL FROM EMPLOYEE WHERE EMPID=EMP_ID;
Exception
when NO_DATA_FOUND
then NULL;
END;
```

Pre-session and post-session commands

You can use pre-session and post-session SQL or shell commands in a workflow that you want to use in a PowerCenter task.

You might use a pre-session or post-session command to start FTP/SFTP scripts or stored procedures, rename or archive files, or run post-processing commands. Configure pre-session and post-session commands in the PowerCenter session.

When you configure a pre-session or post-session command, you can enter a single command or you can call a batch file that contains a set of commands. If you use a batch file, be sure to use complete paths or directories. When you configure the pre-session or post-session command in PowerCenter, enter the complete path or directory along with the file name, such as `c:/IC PowerCenter Task Commands/pre-session1.bat`.

Rules and guidelines for PowerCenter sources and targets

Observe the following rules and guidelines for sources and targets in a PowerCenter task:

- Do not include Nvarchar2 columns in Oracle targets. Due to an ODBC driver limitation, the PowerCenter task truncates the last half of Nvarchar2 data before writing it to Oracle targets.
- Do not write Decimal data of 2147483648 or larger to Microsoft SQL Server or ODBC Integer(10) columns. Doing so can cause unexpected results.

FTP/SFTP connections for PowerCenter tasks

If you create a PowerCenter task with an FTP/SFTP target connection and the IS_STAGED option is enabled for the underlying PowerCenter session, Data Integration writes the flat file to the remote machine and the following local directory:

```
<Secure Agent installation directory>/apps/Data_Integration_Server/data
```

For PowerCenter tasks, Data Integration ignores the Local Directory property specified in the FTP/SFTP connection. Instead, it uses properties specified in the PowerCenter session. To change the local directory or default local filename, change the Output File Directory and Output Filename session properties in PowerCenter. Then export the workflow from PowerCenter to an XML file and re-import the XML file into Data Integration.

Web Service connections for PowerCenter tasks

When a PowerCenter XML file contains Web Service connection information, you can configure a Web Service connection in the PowerCenter task. If you configure a different connection type, the PowerCenter task uses Web Service connection information that is saved in the workflow.

Parameters in PowerCenter tasks

Use mapping parameters and mapping variables in place of values that you want to update without having to edit the workflow. You can include parameters and variables in any transformation that Data Integration supports.

The workflow XML includes the default mapping parameter and mapping variable values. You can edit the values when you configure the task or override them with values in a parameter file.

You cannot add or delete parameters or variables from the PowerCenter workflow in Data Integration.

If the workflow uses a parameter file, the parameter file name is uploaded with the workflow. Copy the parameter file from the PowerCenter directory to a location that is accessible by the Secure Agent or to a cloud-hosted directory. Also, you must update the parameter file headers to use the Data Integration project and PowerCenter task names. For more information about parameter files, see *Mappings*.

PowerCenter task configuration

To create and configure a PowerCenter task, export the workflow from the PowerCenter Repository Manager to an XML file and then upload the XML file in to Data Integration. After you upload the file, map the uploaded connections to Data Integration connections.

You can update an existing PowerCenter task to use a different PowerCenter XML file. When you upload a new PowerCenter XML file to an existing PowerCenter task, the PowerCenter task deletes the old XML file and updates the PowerCenter task definition based on new XML file content.

Configuring a PowerCenter task

Perform the following tasks to create a PowerCenter task in Data Integration.

1. To create a PowerCenter task, click **New** > **Tasks**. Select **PowerCenter Task** and then click **Create**.

   To edit a PowerCenter task, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Edit**.

2. In the Task Details area, configure the following fields:

Field	Description
Task Name	Name of the PowerCenter task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Task names are not case sensitive.
Location	Project folder in which the task resides. If the **Explore** page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.
Description	Description of the PowerCenter task. Maximum length is 255 characters.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task.
Workflow XML File	PowerCenter workflow XML file associated with the task. Only the first 30 characters of the XML file name appear. To upload a file, click Upload XML File. After you upload the workflow XML file, the connections and transformations appear in the Workflow XML File Details area. To download the workflow XML file from Data Integration, click Download XML File. You might download a file to import the workflow to the PowerCenter Workflow Manager for review.

3. Optionally, if the task runs in a serverless runtime environment, configure serverless usage properties.
4. In the Schedule Details area, choose whether to run the task on a schedule or without a schedule. Choose one of the following options:
   - To run a task on a schedule, select **Run this task on schedule** and select the schedule you want to use.
   - To create a new schedule, select **New**. Enter schedule details and click **OK**.
   - To run the task manually, select **Do not run this task on a schedule**.
5. Optionally, if the workflow contains parameters or variables, you can use values from a parameter file. Choose one of the following options:

- To use a parameter file on a local machine, select **Local**. enter the following information:

Field	Description
Parameter File Directory	Absolute path for the directory that contains the parameter file, excluding the parameter file name. The directory must be accessible by the Secure Agent. If you do not enter a location, Data Integration uses the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/userparameters`
Parameter File Name	Name of the file that contains the definitions and values of user-defined parameters and variables used in the task. You can provide the file name or the relative path and file name in this field.

- To use a cloud-hosted file, select **Cloud Hosted**. Enter the following information about the file:

Field	Description
Connection	Connection where the parameter file is stored. You can use the following connection types: - Amazon S3 - Google Storage V2 - Azure Data Lake Store Gen2
Object	Name of the file that contains the definitions and values of user-defined parameters and variables used in the task.

6. Optionally, if you want to create a parameter file based on the parameters and default values specified in the mapping on which the task is based, click **Download Parameter File Template**. For more information about parameter file templates, see *Mappings*.
7. Configure email notification options for the task.
8. In the Connections area, select a **Connection** for each connection reference. A connection reference is a source, target, or lookup connection defined in the workflow XML file.

   Alternatively, to create a connection, click **New**. To edit a connection, click **View** and then click **Edit**.

   The Transformations area displays all transformations defined in the workflow XML file.
9. If the mapping contains parameters, you can edit the values in the Mapping Parameters area.
10. If the mapping contains variables, you can edit the values of the variables in the Mapping Variables area.
11. Click **Save**.
12. To run the PowerCenter task, click **Run**.

    You can also run the task from the **Explore** page.

Running a PowerCenter task

Perform the following tasks before you run a PowerCenter task:

- Ensure that the source and target definitions are current. If the source or target no longer contains fields that are mapped in the field mapping, the PowerCenter task fails.

- If the PowerCenter workflow uses the $PMSourceFileDir\ or $PMTargetFileDir variables to specify the source or target file directory location, you must copy the source or target files to the following directory:
  `<Secure Agent installation directory>/apps/Data_Integration_Server/data`

  If you do not move the source or target files, the task fails.
- If the PowerCenter workflow uses a parameter file, update the parameter file headers with the Data Integration project and task names. Ensure that you have saved the parameter file in a location that is accessible by the Secure Agent or in your cloud-hosted directory. For more information about parameter files, see *Mappings*.

Note: You cannot run multiple instances of a PowerCenter task simultaneously. If you run a PowerCenter task that is already running, the PowerCenter task fails.

You can run a PowerCenter task manually or on a schedule:

- To run a PowerCenter task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  You can also run a PowerCenter task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**.
- To run a PowerCenter task on a schedule, edit the task in the PowerCenter task wizard to associate the task with a schedule.

INDEX

A

add sources
  multiple objects 114
advanced data filters
  configuring 13, 14
  deleting 13, 14
advanced email masking
  description 156
advanced options
  configuring for synchronization tasks 90
advanced relationships
  in mapping tasks 42
AND
  reserved word 20
automatic task recovery
  estimate subset 121

C

CHR function
  inserting single quotation mark 18
CLAIRE tuning
  guidelines 63, 65
CLAIRE Tuning
  continuous tuning 65
  initial tuning 64
Cloud Application Integration community
  URL 10
Cloud Developer community
  URL 10
column names
  duplicates 106
columns
  names, truncating 106
  target database, replication 104
commands
  in synchronization and replication tasks 33
comments
  adding to field expressions 19
configure
  staging connection 117
constants
  definition 18
credit card issuer
  description 155
credit card masking
  overview 155
custom masking
  expression 169
custom substitution masking
  custom dictionaries 165
  parameters 166

D

data catalog discovery
  discovering and selecting objects 37
  searching for objects 35
  synchronization task sources 35
data filters
  configuring 12
  configuring advanced filters 13, 14
  configuring for replication tasks 111
  configuring simple filters 13
  deleting 13
  operators 14
  rules for configuring 15
  synchronization task 87
  variables 15
Data Integration community
  URL 10
data integration tasks
  rules for flat file sources and targets 38
data masking
  common fields 136
  email notification 139
  inplace masking 116
  missing mandatory fields 136
  overview 113
  refresh fields 116
Data Masking
  parameter files 122
data subset
  automatic task recovery 121
  configure relationships 120
  data filters 120
  options 121
  relationship behavior 120
  schema graph 120
  subset rows 131
data transfer task
  configuration 98
  configuring a lookup source 99
  configuring a second source 99
  configuring the source 99
  configuring the target 100
  creation 98
  description 93
  field mapping 101
  lookup condition 95
  lookup source 95
  operations 93
  running 101
  runtime options 101
  scheduling 101
  second source 95
  second source filters 96
  sort conditions 94
  source filters 94

data transfer task *(continued)*
targets 96
data transfer task sources 94
data types
configuring 80
database sources and targets
rules in synchronization tasks 39
database user accounts
requirements for synchronization 83
user privileges for replication tasks 108
databases
configuring multiple source tables for synchronization tasks 85
Daylight Savings Time
schedules 31
DD_DELETE constant
reserved word 20
DD_INSERT constant
reserved word 20
DD_REJECT constant
reserved word 20
DD_UPDATE constant
reserved word 20
default assignment
masking rules 136
delete criteria
description 109
delete task operation 77
deleting
advanced data filters 13, 14
data filters 13
dictionaries
substitution masking 164
dictionary files
data masking 149
DOT format
download 114
download
mapping XML 148
validation reports 148
dynamic mapping task
groups 71
job groups 76
job settings 72
jobs 71
parameters 71
dynamic mapping task jobs
configuring 75
dynamic mapping tasks
configuring 74
creating 74
input and in-out parameters 75
parameters 69
runtime options 76
Dynamic mapping tasks
configuring groups 76
editing 74
overview 69
dynamic schema handling 48, 50

E

elastic mappings
CLAIRE tuning 64
CLAIRE Tuning 63
email masking
description 156
email notification
configuring for tasks 33
Enterprise Data Catalog
discovering and selecting objects 37
finding synchronization task sources 35
searching for objects 35
exception handling
in stored procedures for PowerCenter tasks 174
expressions
configuring for a lookup 81
rules 19
:EXT reference qualifier
reserved word 20
external IDs
for related Salesforce objects 79

F

FALSE constant
reserved word 20
field data type 98
field exclusions
configuring for replication tasks 110
field expressions
comments, adding 19
components 18
creating 17
literals 18
reserved words 20
rules 19
rules for validating 17
syntax 18
field mapping
configuration 101
field mappings
configuring for synchronization tasks 88
required fields 80
understanding for synchronization tasks 80
field metadata 16
fields
required in field mappings 80
flat files
column names in synchronization tasks 80
creating target files in synchronization tasks 78
directory for replication tasks 108
directory for synchronization tasks 83
editing metadata in mapping tasks 16
rules for sources and targets in data integration tasks 38
FTP/SFTP connections
for PowerCenter tasks 175
full load
replication task load type 102
functions
definition 18

H

H2 database
configure 118
data subset 117
install 118
Linux 119
requirements 118
staging 117
Windows 118

I

IDs
  types for objects 79
in-out parameters
  in dynamic mapping tasks 75
  in mapping tasks 56
:INFA reference qualifier
  reserved word 20
Informatica Global Customer Support
  contact information 11
Informatica Intelligent Cloud Services
  web site 10
inplace masking
  update 116
input parameters
  in dynamic mapping tasks 75
  in mapping tasks 56
insert task operation 77
IP address masking
  description 157

J

jobs
  monitoring 34
  scheduling 29

K

key masking
  date 159
  description 157
  numeric 159
  string 157

L

$LastRunDate
  data filter variable 15
$LastRunTime
  data filter variable 15
linear taskflows
  creating schedules 32
literals
  definition 18
  single quotation mark requirement 18
  string and numeric 18
:LKP reference qualifier
  reserved word 20
load types
  for replication 102
  full load 102
lookups
  configuring return values and expressions 81
  configuring the condition 81
  rules for creating 82

M

maintenance outages 11
mapping in advanced mode
  runtime strategy 45
mapping tasks
  advanced options 66
  advanced session properties for mappings 21
  advanced session properties for Visio templates 21
  advanced session properties in advanced mode 27
  CDC runtime properties 46
  CLAIRE tuning 64
  CLAIRE Tuning 63–65
  configuring 50
  configuring schedules, email, and advanced options 57
  configuring sources 52
  creating 50
  editing 50
  in-out parameters 56
  input parameters 56
  joining related sources 41
  overview 40
  reprocessing incrementally-loaded files 66
  running 66
  Sequence configuration 56, 66
  serverless usage properties 29
  Spark session properties 42
  SQL ELT optimization 44
  templates 40
  using parameter files 28
  viewing details 65
mapplets
  in synchronization tasks 81
mask all except
  key masking 157
mask only
  key masking 157
masking
  advanced email 156
  advanced options 139
  configuring rules 143
  configuring the source 141
  configuring the target 142
  credit cards 155
  defining the task 140
  email addresses 156
  IP addresses 157
  overview 113
  phone numbers 160
  random 160
  rules descriptions 152
  scheduling a task 144
  SSN 164
  URL 169
  validation reports 116
masking parameters
  random date 162
  random numeric 161
  random string 161
masking rule
  mapplet 169
  postprocessing expression 154
  preprocessing expression 154
masking rules
  description 152
  source field 134
masking task
  data subset 120
  metadata refresh 133
  refresh metadata 145
  reset 134
  resetting 147

Masking task
copying 147
deleting 148
exporting 148
rename 147
running immediately 145
steps to create 140
masking tasks
serverless usage properties 29
Masking tasks
using parameter files 28
:MCR reference qualifier
reserved word 20
monitoring jobs 34
multiplicity
configuring for a lookup 81

N

NOT
reserved word 20
NULL constant
reserved word 20

O

operating system commands
rules for replication tasks 34
rules for synchronization tasks 34
operations
for synchronization tasks 77
operators
definition 18
OR
reserved word 20
$OutputField
lookup expression
description 81
lookup return value variable 81

P

parameter file
configuring for mapping tasks 57
parameter files
overview 28
parameter scope
dynamic mapping tasks 71
parameter settings
dynamic mapping tasks 71
parameters
user defined 28
parameters in PowerCenter tasks 175
partial sandbox
update 116
phone numbers
masking 160
postprocessing commands
SQL commands 34
tasks 33
powercenter tasks
serverless usage properties 29
PowerCenter tasks
configuring 175, 176
description 172
PowerCenter tasks *(continued)*
editing 176
parameters 175
rules for PowerCenter workflows 172
rules for PowerCenter XML files 172
rules for sources and targets 174
running 177
supported transformations and mapping objects 173
Web Service connections 175
PowerCenter XML files
rules for PowerCenter tasks 172
supported transformations and mapping objects 173
pre- and post-session commands
PowerCenter tasks
using pre-session and post-session commands 174
preprocessing and postprocessing commands
configuring for mapping tasks 57
configuring for synchronization tasks 90
rules for operating system commands 34
rules for SQL commands 34
preprocessing commands
SQL commands 34
tasks 33
prerequisites
for synchronization tasks 83
replication tasks 108
PROC_RESULT variable
reserved word 20

Q

quotation marks
inserting single using CHR function 18

R

random masking
date 162
numeric 161
string 161
related objects
configuring external IDs 79
repeat frequency
schedules 30
replication
source types 103
target types 103
replication tasks
configuring 107
configuring data filters 111
configuring schedules and email 111
configuring sources 108
configuring targets 109
creating 108
creating target tables 106
database targets 103
editing 108
field exclusions 110
flat file targets 104
load types 102
preprocessing and postprocessing commands 33
prerequisites 108
replication process 102
rules for creating 107
rules for running 112
running 112

replication tasks *(continued)*
serverless usage properties 29
viewing details 112
reserved words
list 20
resetting
target tables 104
rules
configuring 143
rules for sources and targets 38
runtime options
dynamic mapping tasks 76

S

Salesforce IDs
for objects 79
schedule
configure 139
configuring for a PowerCenter task 176
schedules
configuring 32
creating 32
Daylight Savings Time 31
repeat frequency 30
time zones 31
scheduling
mapping tasks 57
Masking task 144
replication tasks 111
synchronization tasks 90
taskflows 29
tasks 29
Schema graph
DOT format 114
multiple source objects 114
:SD reference qualifier
reserved word 20
:SEQ reference qualifier
reserved word 20
simple data filters
configuring 13
Social Security masking
description 164
sort conditions
data transfer task 94
source filter characters
description 159
source filter type
description 157
source filters
data transfer tasks 94, 96
source objects
multiple 114
single 114
sources
configuring for mapping tasks 52
configuring for replication tasks 108
configuring for synchronization task 84
configuring multiple database tables for synchronization tasks 85
configuring saved query sources 84
configuring single-object sources 84
rules for multiple-object database in synchronization tasks 78
:SP reference qualifier
reserved word 20
SPOUTPUT
reserved word 20
SQL commands
rules for replication tasks 34
rules for synchronization tasks 34
SQL ELT optimization
mapping tasks 44
SSN masking
description 164
status
Informatica Intelligent Cloud Services 11
stopping a running job 38
stored procedures
exception handling for PowerCenter tasks 174
string literals
single quotation mark requirement 18
substitution masking
description 164
parameters 165
synchronization tasks
adding mapplets 81
configuring 83
configuring schedules, email, and advanced options 90
creating 83
creating flat file targets 78
data filters 87
description 77
editing 83
preprocessing and postprocessing commands 33
prerequisites 83
rules for database sources and targets 39
rules for multiple-object database 78
rules for running 92
running 91
serverless usage properties 29
sources overview 78
sources, configuring 84
targets overview 78
targets, configuring 86
task operations 77
truncating database targets 79
understanding field mappings 80
update columns 79
using parameter files 28
viewing details 91
syntax
for field expressions 18
system status 11

T

table names
duplicates 105
tables
resetting for replication 104
target database, replication 104
user-defined indexes 104
target
refresh fields 116
target databases
column names, truncating 106
for data replication 103
requirements for a synchronization task 83
requirements for replication tasks 108
resetting tables for replication 104
table and column names, replication 104
target filter characters
description 159

target flat files
creating in synchronization tasks 78
target prefix
description 109
guidelines 106
target tables
creating for replication tasks 106
rules for resetting 104
target task operations
insert 115
update 115
upsert 115
targets
configuring for mapping tasks 54
configuring for replication tasks 109
configuring for synchronization tasks 86
truncating in synchronization tasks 79
task operations
data transfer tasks 93
taskflows
monitoring 34
scheduling 29
tasks
configuring email notification 33
configuring field expressions 17
creating schedules 32
monitoring 34
scheduling 29
stopping a running job 38
:TD reference qualifier
reserved word 20
time zones
description 31
transformations
in PowerCenter tasks 173
TRUE constant
reserved word 20
trust site
description 11

U

update columns
adding 79
data transfer task 97
update task operation 77
upgrade notifications 11
upsert task operation 77
URL masking
description 169
user parameters 28
using parameter files
advanced filter 122
simple filter 122

V

validation
rules 17
validation report
target 116
variables
for data filters 15

W

Web Service connections
and PowerCenter tasks 175
web site 10
workflow XML file
downloading from a PowerCenter task 176
WORKFLOWSTARTTIME variable
reserved word 20