Informatica® Cloud Data Quality
April 2024

Cleanse assets

Informatica Cloud Data Quality Cleanse assets
April 2024
April 2024

© Copyright Informatica LLC 1998, 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, 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.

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: 2024-04-05

Table of Contents

Preface 5

Informatica Resources 5

- Informatica Documentation 5
- Informatica Intelligent Cloud Services web site 5
- Informatica Intelligent Cloud Services Communities 5
- Informatica Intelligent Cloud Services Marketplace 5
- Informatica Knowledge Base 6
- Informatica Intelligent Cloud Services Trust Center 6
- Informatica Global Customer Support 6

Chapter 1: Introduction to cleanse assets 7

Cleanse configuration 8

- Cleanse process flow 8
- Example: Customer data cleanup 9

Cleanse asset structure 10

- Configuration tab options 10
- Merge tab options 12

Working with steps 13

Configuring a cleanse asset 14

Cleanse instance configuration options 15

Adding a cleanse instance to the asset 16

Cleanse assets and mappings 16

Cleanse assets and dimensions 17

Chapter 2: Step configuration 18

Updating the character case of the input data 19

Removing instances of one or more values that you specify 19

Removing instances of values that appear in a dictionary 20

Removing character spaces from the data 21

Replacing instances of one or more values that you specify 21

Replacing instances of dictionary values with a value that you specify 22

Replacing instances of dictionary values with valid values 22

Dictionaries and cleanse assets 23

Rules and guidelines for search and replace operations 24

Chapter 3: Merge configuration 25

Merge configuration options 25

Merging records 26

Chapter 4: Test cleanse assets 27

Testing a cleanse instance 27

Rules and guidelines for test data 28

Testing the merge operation 28

Index 29

Preface

Refer to *Cleanse* for information on how to standardize the appearance of your data, replace incorrect values in your data, and remove unwanted values from your data. To standardize and update your data values, you configure one or more operations in a cleanse asset in Data Quality. Then, you add the asset to a Cleanse transformation in a mapping in Data Integration.

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.

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/

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

Introduction to cleanse assets

A Cleanse asset is a set of one or more data transformation steps that can standardize the form and content of your data. The asset comprises of one or more cleanse instances. Each instance in a cleanse asset contains the input fields on which you apply the cleanse operations. You can configure a cleanse asset to perform cleanse and merge operations on the input fields that a cleanse instance identifies in the asset.

You might decide to standardize your data to achieve the following goals:

Improve data consistency in a data set

Your data might use a range of data values to represent the same data point. For example, a field of country names might contain a mix of full names, ISO three-character codes, and other non-standard country identifiers. You standardize the values to ensure uniformity in the data field.

Fix errors in data

Your data might include correct and incorrect spellings of given data values. You standardize the data to find and fix the errors.

Comply with regulatory standards

You might need to verify that your data meets the latest regulatory standards that the industry or government defines. For example, you might decide to search a data set for potentially fraudulent accounts. To start the initiative, you standardize the data to replace inaccurate values with a single term, such as REVIEW. You can proceed to review the records that contain the new term.

Prepare for downstream data quality initiatives

You might standardize your data as the first step in a data quality project. For example, you standardize common business terms and postal address terms before you run a mapping with a Verifier transformation on your address data.

Merge data fields to effectively manage your data set

Your data set might contain the data information in the discrete fields, which you might decide to combine and add into a single field. For example, you might need a single full name field in addition to the first name and surname fields in your data set. You can merge the two separate fields into a single field.

You create and test a cleanse asset in Data Quality, and you add a cleanse asset to a Cleanse transformation in a mapping in Data Integration.

When the mapping runs, the Cleanse transformation applies the operations that you define in the cleanse asset to the fields that you select in the input data. The outputs from the transformation contain the standardized field values and also any merged fields that you configured in the asset.

Cleanse configuration

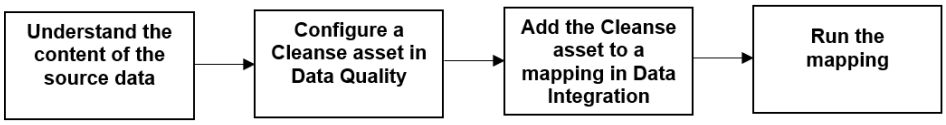
A cleanse asset is composed of instances and steps. An instance identifies the input fields on which you can perform cleanse and merge operations. A step is a discrete cleanse operation that you define for an input.

You define one or more cleanse steps for the input fields that an instance identifies. The input fields that you add to an instance and the steps that you define for the input fields depend on your business requirements and the content of your data. You can add one or more input fields to an instance.

Cleanse process flow

To cleanse your data, you configure and run assets in Data Quality and in Data Integration.

The following image shows the steps involved in the cleansing process:



The cleansing process includes the following steps:

1. Analyze the content of the source data, so that you can identify the fields that require cleanup. Work with a developer or data steward to understand the data sets. You will create inputs that represent the source data columns in the cleanse asset instances.

   During your analysis, perform the following steps:

   a. Verify your business requirements.

   b. Verify the content and structure of your data.

   c. Determine the sequence of the cleanse operations to apply to the data.

2. In Data Quality, configure a cleanse asset to translate your business requirements into one or more cleanse instances. The input fields within an instance are processed through the logic that you define in a particular instance.

   To configure the asset, perform the following steps:

   a. Add one or more instances for the input fields that need cleansing in the source data.

   b. Configure one or more steps for the input fields that each instance identifies.

   c. Optionally, configure the merge operation for the input fields.

3. In Data Integration, define a mapping that can run the cleansing operation:

   a. Add the Cleanse asset to a Cleanse transformation.

   b. Connect the cleanse asset input and output fields to the upstream and downstream objects in the mapping.

4. Run the mapping.

Note: The Cleanse transformation does not identify the instance on which each asset input originates. If you define multiple instances on the asset in Data Quality, make a record of the instances to which each input belongs. Use the record as a guide when you connect the asset inputs to the transformation input fields.

Example: Customer data cleanup

A customer data set might include multiple fields for customer contact data, including name, gender, and address fields. You can configure a single cleanse asset to perform cleanse operations on the different fields.

The customers data record in your organization might contain the following information:

Title	FirstName	MiddleName	Surname	Gender	Address	Country
Dr.	John	William	smith	Male	2101 massachusetts ave nw washington dc 20008-2811	usa
Mr.	Mr. Frances	folsom	cleveland	Man	18 broomfield ridge midleton co. cork p25 kn66	IRE
Miss	Miss. Abigail	powers	Fillmore	Female	shop 7 208 adelaide st brisbane city qld 4000	aus

You might want to perform the following operations on your record:

- Remove salutation data from the FirstName field.
- Change the character case to title case in the MiddleName and Surname fields.
- Replace `Man` with `Male` in the Gender field.
- Remove character spaces and change the character case to uppercase in the Address field and Country field.

Configure a cleanse asset with four instances, based on the types of cleanse operation that you want to perform on the data. Add one or more input fields to each instance and define the steps that you want to apply to the input fields in each case.

For example, to change the character case to title case in the MiddleName and Surname fields, create an instance in the asset to add two input fields. You can specify MiddleName and Surname as the input field names. Configure the step to perform the **Convert Case** operation on the fields that the instance identifies.

Likewise, you can configure an instance to change the character case of the address and country abbreviation data. Add a step to convert the case of the data to uppercase. Additionally, add a step to replace multiple character spaces with a single space. Provide suitable names for the inputs, such as Address and Country.

Note: The Country input does not contain any character spaces, and so the step logic to remove spaces does not alter the data in the associated input field.

Add the asset that you create to a Cleanse transformation in a mapping. When a mapping runs, the transformation applies the standardization operations that you define in the asset instances to the fields that you select.

Additionally, if you want to merge the cleansed data from the FirstName, MiddleName, and Surname fields into a single field, configure the asset to perform a merge operation on the input fields.

Create a merge record to add the fields that you want to merge. You can specify the merged field name as FullName.

Cleanse asset structure

A cleanse asset contains options on a **Definition** tab, a **Configuration** tab, and a **Merge** tab.

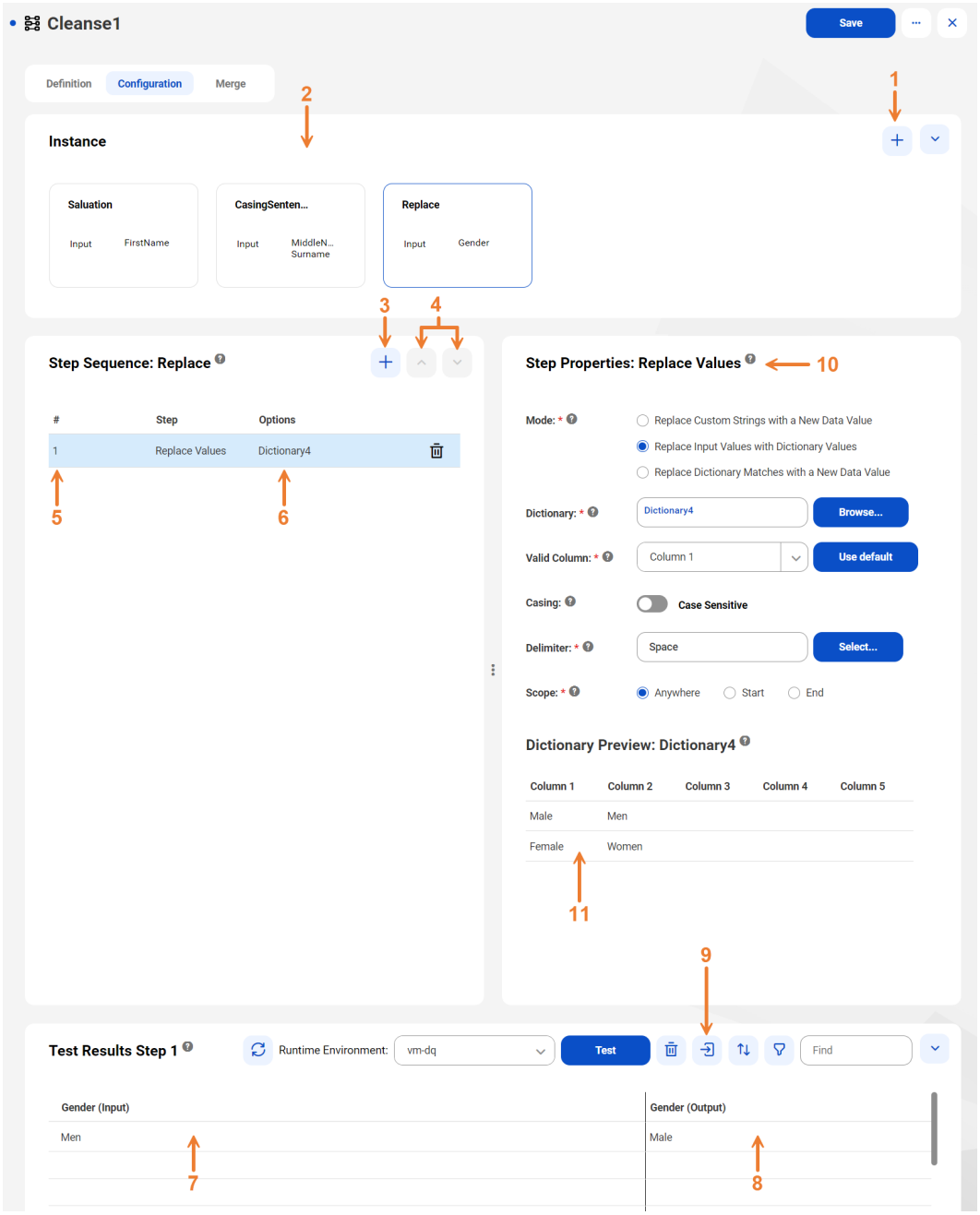
Use the Definition tab options to enter a name for the asset, optionally enter a description for the asset, and select the folder in which to store the asset. Use the Configuration tab options to configure the standardization operations that a mapping will perform. Optionally, use the Merge tab options to merge data from two or more fields into a single field.

You configure the standardization operations as a sequence of steps that the mapping follows at run time.

Configuration tab options

Use the Configuration tab options to configure the type of cleansing operations that a mapping will perform on the input data. Each instance that you add identifies one or more input fields to which a mapping can apply the cleanse operations. You add instances to configure different cleansing operations for different input fields within a single asset.

The following image shows the cleanse asset options in the **Configuration** tab:



The Configuration tab includes the following options:

1. Add Instance option.
   Adds a cleanse instance to the asset. You can add one or more cleanse instances.

   Define one or more input fields in each instance that you add.

2. Instance panel.
   Displays the cleanse instances that you add to the asset. You can edit or delete an instance.

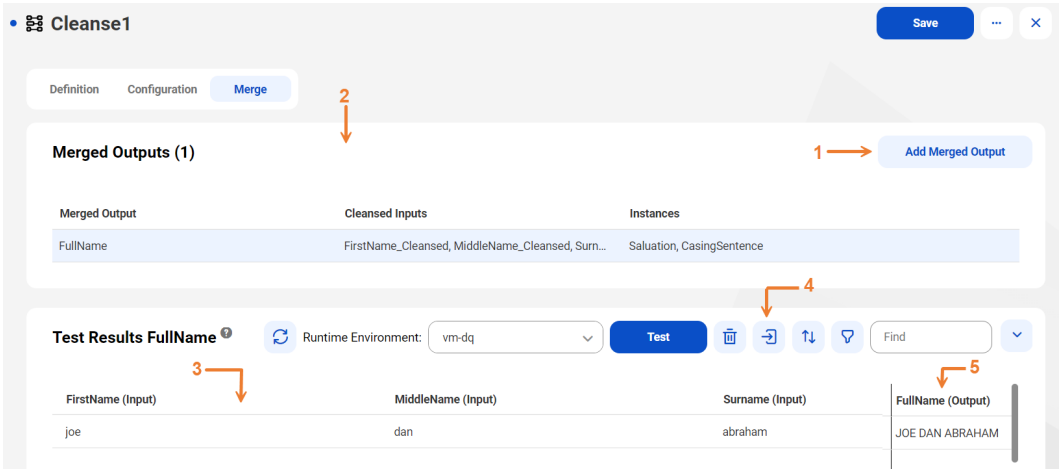
   Expand or collapse the instance panel to show or hide the instances that you add. You might collapse the instance panel when you configure a step in an instance, so that the step configuration options are easier to view.

3. Add Step option.
   Adds a step to the asset. A step describes a standardization operation that a mapping can apply to an input data field.
4. Up and Down options.
   Moves a step that you select up or down within the step sequence. The mapping performs standardization operations in the order that you specify.
5. Step sequence.
   Defines the order in which the mapping applies each step to the input field.
6. Operation options.
   Lists the main element in the operation that you define within the step. For example, if you use a dictionary in a step, the list displays the dictionary name.
7. Test input fields.
   Contains the test data that you enter for the instance that you select.
8. Test output fields.
   Contains the result of the test for the instance that you select. The asset displays an output field for each input field that you add to an instance.
9. Import file option.
   Imports data to the test panel.
10. Step properties.
    Displays the configurable properties that you use to define the step.
11. Dictionary data.
    Shows the contents of any dictionary that you select in the step.

Merge tab options

Use the Merge tab options to merge multiple input fields into a single output field. You can merge any combination of the fields to suit your data requirements.

The following image shows the options in the Merge tab:



The Merge tab includes the following options:

1. Add Merged Output option.
   Opens the merge configuration options. Use the options to identify the fields to merge.
2. Merged outputs panel.

Lists the merged fields that the asset will create at run time. Also lists the input fields in each case and the instances to which the inputs belong. You can edit or delete a merged field. When you hover over the merged field, the Edit and Delete options appear.

3. Test input fields.
   Contains input data that you enter to verify the structure of the merged field that you select.
4. Import file option.
   Imports data to the test panel.
5. Test output fields.
   Contains the result of the test.

Working with steps

When you configure a cleanse asset, you define one or more steps for the input fields that an instance identifies. A step describes a standardization operation that a mapping can apply to an input field on a Cleanse transformation. The mapping applies the steps in a sequence that you define in the cleanse asset.

You can configure multiple steps for the input fields in a cleanse asset. The mapping applies the steps to one or more input fields that you select in the Cleanse transformation.

When you add a step to a cleanse asset, you choose from the following options:

Convert Case

Updates the character case of the data in an input field. For example, you can convert a field of mixed-case characters to uppercase characters. You can select or clear an option to ignore uppercase values in the input data.

Remove Values

Searches an input field for one or more values that you specify, and removes the values.

You can enter one or more values that the mapping will remove. Or, you can use a dictionary to identify the values that the mapping will remove.

Remove Spaces

Removes redundant character spaces from an input field.

You can remove leading and trailing spaces, and you can replace multiple spaces with a single space.

Replace Values

Searches an input field for one or more values that you specify, and replaces the values. You can find and replace values in the following ways:

- Enter one or more values that the mapping can search for. Enter a single replacement value for every value that the mapping finds.
- Use a dictionary to identify multiple strings that the mapping can search for. Use the same dictionary to specify a replacement value for each value that the mapping finds.
- Use a dictionary to identify multiple strings that the mapping can search for. Enter a single replacement value for every value that the mapping finds.

Configuring a cleanse asset

To create an asset in Data Quality, click **New**. When you click **New**, Data Quality prompts you to select an asset type. When you select Cleanse, Data Quality opens a new cleanse asset.

The asset displays a **Definition** tab, a **Configuration** tab, and a **Merge** tab. Use the Definition tab to define the name and the location of the asset. Use the Configuration tab to configure the steps that the mapping will apply to the input data. Use the Merge tab to create new fields from on two or more fields across the instances.

Note: You can also open the Definition, Configuration, and Merge tabs when you open the asset from the Explore page.

1. To create a cleanse asset, click **New** > **Cleanse**.
2. On the **Definition** tab, enter a name for the cleanse asset.
3. Optionally, enter a description.
4. Select the location in which to save the cleanse asset.
5. Optionally, select a data quality dimension to represent the type of data quality issue that you want the asset to examine.

   For more information about dimensions, see "Cleanse assets and dimensions" on page 17.
6. Save the cleanse asset.

   The **Asset References** panel appears. The **Asset References** panel shows additional assets that are referenced by Data Quality assets to perform transformation tasks. When you create a new asset, there are no referenced assets.
7. Optionally, add a tag to the cleanse asset. You can search for assets with a common tag on the Explore page.
8. Select the Configuration tab.

   Data Quality displays the configuration workspace for the cleanse asset. Data Quality also displays a validation message to indicate that the step configuration is incomplete.

   When you create a cleanse asset, the Instance panel in the **Configuration** view displays a single default cleanse instance. By default, the instance contains a single input field. The input field name is the name of the asset. Optionally, you can add input fields and modify the default input field name and instance name.
9. Optionally, add one or more cleanse instances. For information about how to add a cleanse instance to the asset, see "Adding a cleanse instance to the asset" on page 16.
10. Add a step to each cleanse instance in the asset.

    Select one of the following step types:

    - Convert Case
    - Remove Values
    - Remove Spaces
    - Replace Values
11. Configure the properties on the step to create an operation that the mapping can apply to the data.
12. Optionally, add one or more additional steps to each instance.

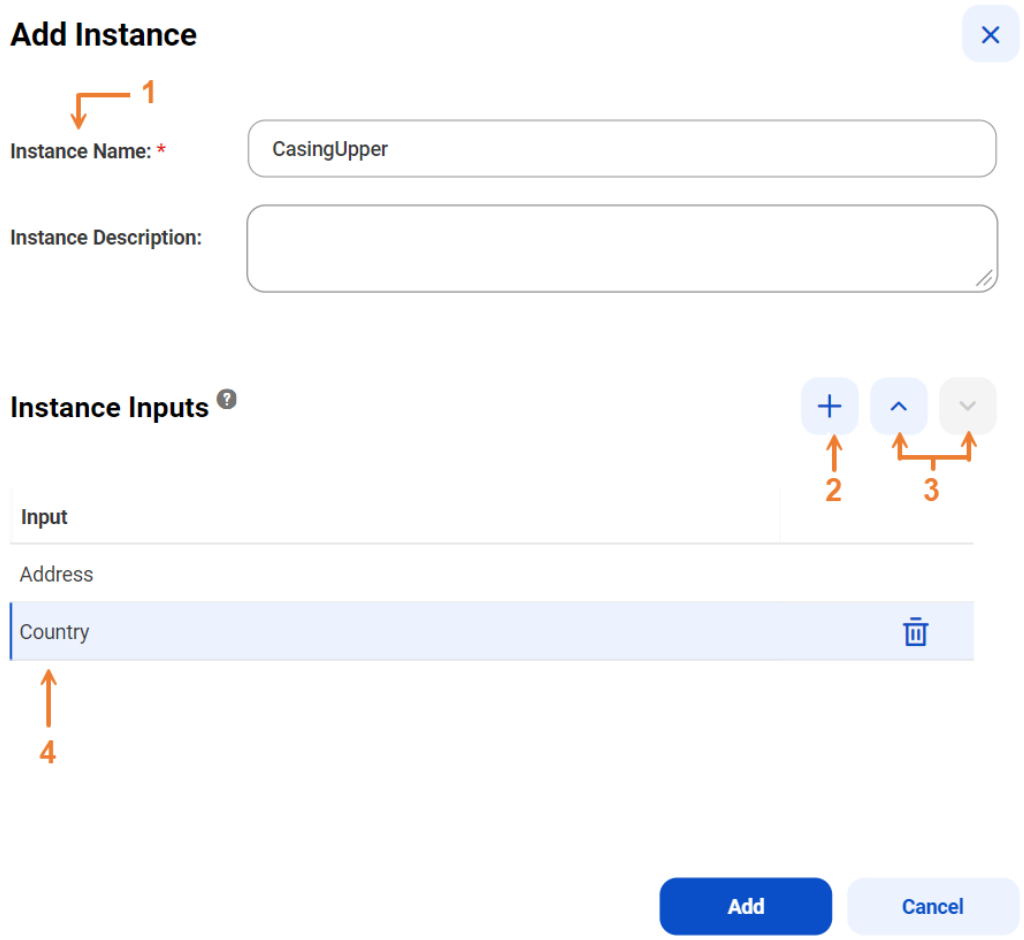
    Configure the properties on each step that you add.
13. Optionally, configure a merge operation for two or more input fields. For information about configuring the fields to merge, see "Merging records" on page 26.
14. Save the asset.

After you configure the cleanse asset, test the asset with sample data.

Cleanse instance configuration options

When you configure a cleanse instance, you add one or more input fields on which to perform cleanse operations.

The following image shows the options on the **Add Instance** dialog box:



The dialog box contains the following options:

1. Instance name.
   Specifies the name of the instance that you add.

   Enter an instance name that represents the type of input data fields that you add to the instance or the type of operation to perform on the input data.
2. Add Input option.
   Adds an input field to the instance. You can add one or more input fields.
3. Up and Down options.
   Moves an input field that you select up or down within the instance input.

4. Input fields.
   Displays the input fields that you add to the instance. You enter the input data field names.

   Choose names for each input that represent the data on the input fields that the Cleanse transformation will read.

   Note: The Maximum string length property applies to the asset inputs and outputs in test operations and at run time. You can set a maximum string length of 10,000 characters. The default value on the property is 255 characters.

Adding a cleanse instance to the asset

You can add one or more instances to a cleanse asset. By default, a cleanse asset contains a single cleanse instance. The Instance panel on the **Configuration** tab displays the cleanse instances.

1. Open a cleanse asset.
2. On the Configuration tab, click the **Add Instance** option present in the Instance panel.

   The **Add Instance** dialog box appears.
3. Enter a name for the instance.
4. Select the option to add the input fields to the instance.
5. Click **Add**.

   The asset adds the cleanse instance to the Instance panel.
6. Open the Instance panel to view the instance that you created.
7. Optionally, you can edit a cleanse instance to add additional input fields or to modify the input field name or instance name. You can also delete the instance that you added.

Cleanse assets and mappings

The cleanse assets that you create are available to the users who create mappings in Data Integration. You or other users add a Cleanse transformation to a mapping and add a cleanse asset to the transformation. The Cleanse transformation applies the logic in the cleanse asset to one or more input fields that you specify.

A Cleanse transformation works in a similar way to a Mapplet transformation. A Data Integration user connects the transformation inputs and outputs to other objects in the mapping in the same manner as a mapplet connects to other mapping objects. When the Data Integration user runs the mapping, the Cleanse transformation applies the cleanse asset logic to the input fields and generates an output field that you can connect to a downstream transformation.

After the mapping runs, you can evaluate the data output to determine if the data meets the standardization objectives.

Rules and guidelines for cleanse assets and mappings

Consider the following rules and guidelines when you work with mappings and cleanse assets:

- A cleanse asset is a reusable object. A Data Integration user can add a cleanse asset to transformations in multiple mappings. A Cleanse transformation is a nonreusable object. You select a Cleanse transformation when you configure a mapping.

- A cleanse asset is a read-only object in Data Integration. The cleanse asset does not change unless you update it in Data Quality.
- To view the list of cleanse assets that you created in Data Quality, select the **Explore** option.
- If you update a cleanse asset that you previously added to a transformation, you must synchronize the transformation with the cleanse asset.

  To synchronize the transformation, open the transformation in the mapping and select the Cleanse properties. Data Integration displays a message that prompts you to synchronize the transformation with the cleanse asset.

Cleanse assets and dimensions

The data quality issues that you may find in your data can fall into a range of common categories. Data Quality assets can identify the categories as dimensions. When you configure an asset in Data Quality, you can use the **Dimension** property to indicate the type of data quality issue that you want the asset to examine.

Find the Dimension property on the **Definition** tab of the asset.

For more information about data quality dimensions, see the *Introduction* in Data Quality documentation.

CHAPTER 2

Step configuration

Steps define the standardization operations that a mapping applies to input fields and the order in which the operations run. You add one or more steps to each cleanse instance in the asset. Each sequence of steps applies to the fields in the instance that you select.

You can configure steps to perform the following operations:

Update the character case of the data

For example, you might want to convert a set of postal address records to uppercase. The postal carrier in your country might declare a preference for address data in uppercase characters.

Select **Convert Case** as the step type, and select the uppercase option.

Remove instances of a single value from the data

For example, you might want to remove the country identifier from an address data field, so that the country name does not appear on envelopes that you mail to domestic addresses.

Select **Remove Values** as the step type, and select **Remove Custom Strings** as the operation. Next, enter the value to remove.

Remove instances of multiple values from the data

For example, you might want to remove state or province names from a field that also contains the state or province abbreviations.

Select **Remove Values** as the step type, and select **Remove Dictionary Values** as the operation. Next, select a dictionary that contains the state or province names.

Remove character spaces from the data

You can clean up a field that contains extraneous and redundant character spaces.

Select **Remove Spaces** as the step type, and select the options for leading spaces, trailing spaces, and multiple spaces.

Replace instances of a single value in the data

For example, you might want to update a vendor name in your address records.

Select **Replace Values** as the step type, and select **Replace Custom Strings with a New Data Value** as the operation. Next, enter the old vendor name and the new name.

Replace instances of multiple values with corresponding new values

For example, you might want to use abbreviated street descriptors in your address data, such as ST, RD, and AVE. The postal carrier in your country might declare a preference for abbreviated descriptors.

Select **Replace Values** as the step type, and select **Replace Input Values with Dictionary Values** as the operation. Next, select a dictionary that contains the full versions of the street descriptors and the abbreviated versions. Select the column of abbreviations as the valid column.

Replace multiple values in the data with a single value

For example, your address data might contain multiple values that are no longer valid. The values might identify customer coupons that have expired.

Select **Replace Values** as the step type, and select **Replace Dictionary Values with a New Data Value** as the operation. Next, select a dictionary that contains the coupon codes. Additionally, enter a value that the operation can return, such as EXPIRED.

Updating the character case of the input data

You can configure a step to change the character case in a data field.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.
2. Select the **Convert Case** option.
3. In the step properties pane, select one of the following casing styles:
   - Upper Case
   - Lower Case
   - Toggle Case
   - Sentence Case
   - Title Case
4. If you select the Lower Case, Toggle Case, Sentence Case, or Title Case option, specify one or more delimiters for the input data. Optionally, select one or more dictionaries to apply to the input data.

   If you select a dictionary, select a valid column for the dictionary. In addition to converting the character case, the asset replaces any value in a non-valid dictionary column with the corresponding value in the valid column.
5. Optionally, select or clear the option to leave uppercase words unchanged.
6. Save the asset.

Removing instances of one or more values that you specify

You can configure a step to remove instances of one or more values that you specify from a data field.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.
2. Select the **Remove Values** option.
3. In the step properties pane, select **Remove Custom Strings**.
4. Verify the delimiter that the mapping will use to recognize discrete values in the input field. You can specify one or more delimiters.

   The input delimiter determines how the mapping reads the input data when you select the Start or End option as the scope.
5. Set the scope for the search operation.

You can set the following options:

- Anywhere. Removes any instance of the value that occurs in the input field. Can remove multiple instances of the same value in a field.
- Start. Removes any instance of the value that is not preceded by another value in the input field.
- End. Removes any instance of the value that is not followed by another value in the input field.

6. Enter a string to search for and remove in the **Custom Strings** field.

   Note: The search operation is case-sensitive.

7. Add the string to the list of strings to search for.
8. Optionally, add one or more additional strings to the list.
9. Save the asset.

Removing instances of values that appear in a dictionary

You can configure a step to remove instances of values that appear in a dictionary. You might use a dictionary when the list of values to remove is too long to enter as custom values. You might also use a dictionary when you will reuse the list of values in multiple cleanse steps.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.
2. Select the **Remove Values** option.
3. In the step properties pane, select **Remove Dictionary Values**.
4. Select the dictionary to use in the step.

   Click **Browse** to select the dictionary. You can preview the dictionary contents in the Step Properties pane.

5. Select the valid column for the dictionary. The valid column contains the versions of the dictionary values that you designate as the preferred or correct values for the current operation.

   Note: The search operation ignores the values in the valid column. However, if a value in the valid column also appears in another dictionary column, the search operation selects the value for removal.

6. Select or clear the option to perform a case-sensitive comparison between the dictionary data and the input data.
7. Verify the delimiter that the mapping will use to recognize discrete values in the input field. You can specify one or more delimiters.

   The input delimiter determines how the mapping reads the input data when you select the Start or End option as the scope.

8. Set the scope for the search operation.

   You can set the following options:

   - Anywhere. Removes any instance of the value that occurs in the input field. Can remove multiple instances of the same value in a field.
   - Start. Removes any instance of the value that is not preceded by another value in the input field.
   - End. Removes any instance of the value that is not followed by another value in the input field.

9. Save the asset.

Removing character spaces from the data

You can configure a step to remove extraneous and redundant character spaces from a data field. The step does not remove instances of single spaces between values in the field.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.
2. Select the **Remove Spaces** option.
3. Select one or both of the following options:
   - Leading and trailing spaces. Removes one or more character spaces at the start or the end of the string.
   - Multiple spaces. Replaces instances of two or more spaces with a single space.
4. Save the asset.

Replacing instances of one or more values that you specify

You can configure a step to replace instances of one or more values that you specify with a single new value.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.
2. Select the **Replace Values** option.
3. In the step properties pane, select **Replace Custom Strings with a New Data Value**.
4. Verify the delimiter that the mapping will use to recognize discrete values in the input field. You can specify one or more delimiters.

   The input delimiter determines how the mapping reads the input data when you select the Start or End option as the scope.
5. Set the scope for the search operation.

   You can set the following options:
   - Anywhere. Replaces any instance of the value that occurs in the input field. Can replace multiple instances of the same value in a field.
   - Start. Replaces any instance of the value that is not preceded by another value in the input field.
   - End. Replaces any instance of the value that is not followed by another value in the input field.
6. Enter a value to search for and replace in the **Custom Strings** field.

   Note: The search operation is case-sensitive.
7. Add the string to the list of strings to search for.
8. Optionally, add one or more additional strings to the list.
9. Enter the replacement value in the **New Data Value** field.
10. Save the asset.

Replacing instances of dictionary values with a value that you specify

You can configure a step to replace instances of values that appear in a dictionary with a value that you enter in the step properties. You might use a dictionary when the list of values to replace is too long to enter as custom values. You might also use a dictionary when you will reuse the list of values in multiple cleanse steps.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.
2. Select the **Replace Values** option.
3. In the step properties pane, select **Replace Dictionary Matches with a New Data Value**.
4. Select the dictionary to use in the step.

   Click **Browse** to select the dictionary. You can preview the dictionary contents in the Step Properties pane.
5. Select the valid column for the dictionary. The valid column contains the versions of the dictionary values that you designate as the preferred or correct values for the current operation.

   Note: The search operation ignores the values in the valid column. However, if a value in the valid column also appears in another dictionary column, the search operation selects the value for replacement with the new value.
6. Select or clear the option to perform a case-sensitive comparison between the dictionary data and the input data.
7. Verify the delimiter that the mapping will use to recognize discrete values in the input field. You can specify one or more delimiters.

   The input delimiter determines how the mapping reads the input data when you select the Start or End option as the scope.
8. Set the scope for the search operation.

   You can set the following options:

   - Anywhere. Replaces any instance of the value that occurs in the input field. Can replace multiple instances of the same value in a field.
   - Start. Replaces any instance of the value that is not preceded by another value in the input field.
   - End. Replaces any instance of the value that is not followed by another value in the input field.
9. Enter the replacement value in the **New Data Value** field.
10. Save the asset.

Replacing instances of dictionary values with valid values

You can configure a step to search for instances of values that appear in a dictionary and to replace each value with a corresponding valid value from the dictionary. You can use a dictionary when the list of values to replace is too long to enter as custom values. You might also use a dictionary when you will reuse the list of values in multiple cleanse steps.

1. On the **Configuration** tab of the cleanse asset, click **Add Step**.

2. Select the **Replace Values** option.
3. In the step properties pane, select **Replace Input Values with Dictionary Values**.
4. Select the dictionary to use in the step.

   Click **Browse** to select the dictionary.
5. Select the valid column for the dictionary.

   You can preview the dictionary contents in the Step Properties pane.

   Note: The valid column represents the preferred version of the values in the dictionary. The step searches for values that appear in other columns in the dictionary and replaces them with the corresponding values from the valid column.
6. Select or clear the option to perform a case-sensitive comparison between the dictionary data and the input data.
7. Verify the delimiter that the mapping will use to recognize discrete values in the input field. You can specify one or more delimiters.

   The input delimiter determines how the mapping reads the input data when you select the Start or End option as the scope.
8. Set the scope for the search operation.

   You can set the following options:

   - Anywhere. Replaces any instance of the value that occurs in the input field. Can replace multiple instances of the same value in a field.
   - Start. Replaces any instance of the value that is not preceded by another value in the input field.
   - End. Replaces any instance of the value that is not followed by another value in the input field.
9. Save the asset.

Dictionaries and cleanse assets

A dictionary is a reference data set that a cleanse step can use to evaluate data. Use dictionaries to verify that the data values on a data source or another object in a mapping are accurate and correctly formatted.

When you run a mapping with a Cleanse transformation that specifies a dictionary, the transformation compares the input field data to the data in the dictionary. If the transformation finds a match between an input value and a dictionary value, the transformation performs an action that you define in the corresponding step in the cleanse asset.

At least one dictionary column must contain the set of standard or preferred values for your current data project. The other columns can contain alternative versions of the values. The column that contains the standard or preferred values is called the valid column.

In search and replace operations, a Cleanse transformation searches every column in the dictionary except the valid column that the asset specifies. If you want the search operation to include the valid column data, add a copy of the valid column to the dictionary. By default, the valid column is the first or left-most column in a dictionary. You can create a dictionary that contains two identical columns of data.

A dictionary might contain public terms, such as telephone area codes or address abbreviations. Or, a dictionary might contain values that are specific to an organization, such as employee codes or product codes. You can populate a dictionary with any combination of values that suits your project. The data in each column does not need to be formally correct.

Rules and guidelines for search and replace operations

When you configure a step to remove or replace values in an input data field, you can enter the values as properties on the step. You can also use a dictionary to specify the values.

The combination of dictionary values and custom values that you select and the order in which you select them can affect the search and replace operations at run time.

Consider the following rules and guidelines when you configure the properties on a step:

Searching in single-value and multi-value strings

- A search operation looks for complete values in an input field. The search operation ignores any value that occurs within a larger string. For example, a search operation will ignore the value DATA in the string METADATA.
- In a step to remove or replace values, the search operation applies the scope criterion that you select to the values in the input field. The search option does not reapply the scope as the operation progresses.

  For example, if you enter the values HARD and DISK as separate search terms in a step and you set the scope to Start, the search operation removes HARD from the string HARD DISK DRIVE. The value DISK remains at the start of the output string.

  If you set the scope to Anywhere in the same step, the search operation will remove both HARD and DISK from the input field.
- You can remove or replace more than one value in a string. For example, you can configure a step to remove the values VIRTUAL PRIVATE from VIRTUAL PRIVATE NETWORK.

Selecting the valid column in a dictionary

- When you configure a step to remove dictionary terms, the search operation ignores the terms in the valid column that you select.

  You can use a dictionary that contains a single column of data to remove values or to replace data values with a custom value. To do so, select any other column number on the Valid Column field as the valid column.
- You can use a dictionary to find values in the input data and to replace the values with values from the valid column in the dictionary. When you use a dictionary to find and replace values, select a dictionary that contains at least two columns.
- You can review the dictionary data in the Step properties pane and select the valid column that you need.

Differences in search criteria for dictionaries and custom strings

- When you configure a step to remove or replace a list of custom strings, the search operation searches for the strings in the order in which you enter them in the Step properties pane.

  For example, let's say that you first enter SITE and then enter WEB SITE as custom strings, and you set the scope to Anywhere. The search operation removes or replaces the value SITE first and cannot subsequently remove the string WEB SITE. The value WEB remains in the output data.

  If you enter WEB SITE before SITE in the same step, the search operation will remove both terms from the input field.
- When you configure a step to remove or replace a list of dictionary values with a new value, the search operation prioritizes longer strings in the dictionary.

  For example, let's say that you select a dictionary that contains the strings SITE and WEB SITE, and you set the scope to Anywhere. The search operation removes or replaces the value WEB SITE and also removes the value SITE.

CHAPTER 3

Merge configuration

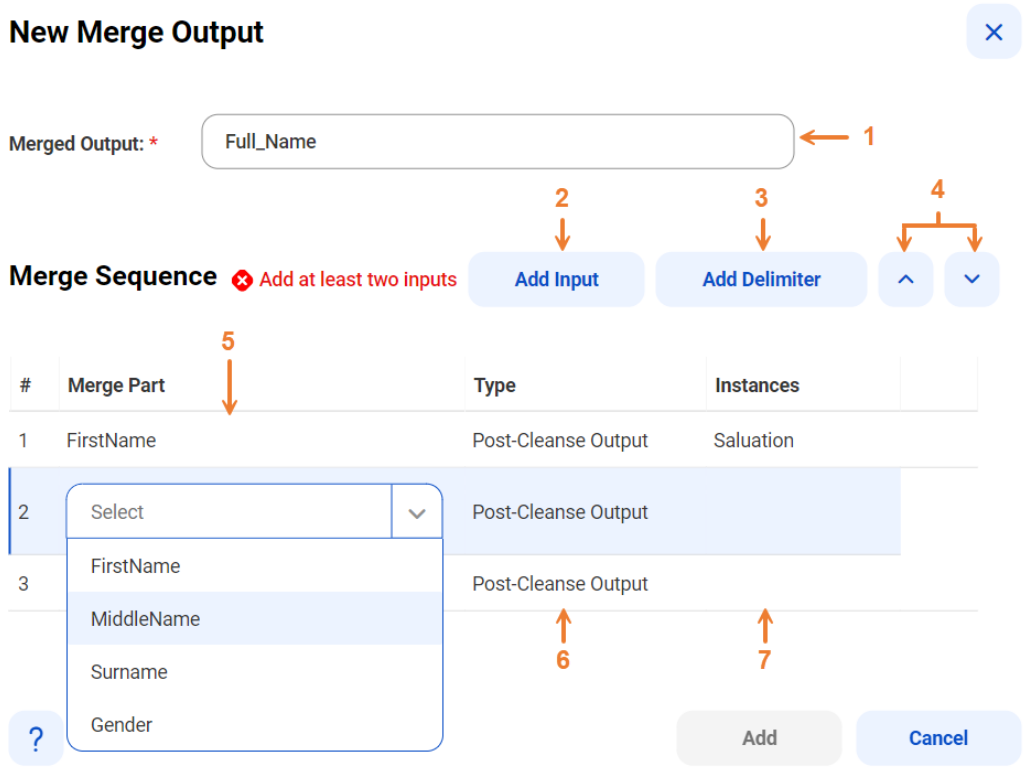
A cleanse asset contains a Merge tab where you can define merge operations for input fields. Use the Merge tab to create a field based on one or more fields across the instances or within an instance.

You configure the Merge tab options to merge the post-cleansed input fields. When you run a test on the Merge tab, the merged output column in the test panel displays the result of the merged record that you select.

Merge configuration options

To merge one or more input data fields into a single field, create a cleanse asset and configure the options in **Merge** tab.

The following image shows the options on the **New Merge Output** dialog box:



The dialog box contains the following options:

1. Merged output field name.
   Specifies the merged field name. You enter a name for the merged output field.
2. Add Input option.
   Adds an input field whose data you want to merge. You can add two or more input data fields.
3. Add Delimiter option.
   Specifies the delimiter to use between the input fields that you merge. The default delimiter is a character space.

   You can select a delimiter from a list of built-in delimiters or add a custom delimiter. Select the delimiters from the **Select Delimiter** dialog box.
4. Up and Down options.
   Moves an input field that you select up or down within the merge sequence.
5. Merged part names.
   Displays the lists of the fields that you want to merge.
6. Type.
   Identifies the type of the input field considered for merging. The operation merges the cleansed version of each input field.
7. Instances.
   Indicates the instance from which you added the input field.

Merging records

You can configure a cleanse asset to merge two or more fields into a single new field.

1. Select the **Merge** tab on the asset.
2. On the **Merge** tab, click **Add Merged Output**.

   The **New Merge Output** dialog box appears.
3. Enter a name for the new field.
4. To select the input fields to merge, click **Add Input**.
5. Optionally, to change the delimiter to apply between the merged fields, click **Add Delimiter**. The default delimiter is a character space.
6. Click **Add**.

   The **Merge** tab displays the new field.

CHAPTER 4

Test cleanse assets

Test the asset in Data Quality to verify that the asset logic generates the results that you expect.

Testing a cleanse instance

Test a cleanse asset instance to verify that the data flows through the instance in the ways that you expect. You test a single cleanse instance at a time.

1. Open the cleanse asset.
2. Select the **Configuration** tab, and select an instance.
3. Select a step in the step sequence that you configured in the instance.

   The test will run from the first step in the sequence through the step that you select. To test all steps in the instance, select the final step.

   Note: If any step that you select includes a dictionary, the Secure Agent downloads all of the dictionaries that the asset specifies when you run the test.
4. Select a Secure Agent to run the test.

   To refresh the list of active Secure Agents, click the **Refresh** icon.
5. Enter data values in the test panel, or import the data to test. To import the data, click the **Import** option in the test panel.

   Note: Configure the asset as completely as you can before you enter or import the test data. The asset may discard your test data if you make further changes to the configuration.

   For more information see "Rules and guidelines for test data" on page 28.
6. Click **Test**.

   The Output column displays the test results. The results represent the application of the step logic in the sequence that the instance specifies.
7. Verify that the steps read the input values and write the output values that you expect.

   You can sort and filter the test data, and you use the Find feature to search for values in the test data.
8. Optionally, repeat the steps to test other instances in the asset.

Rules and guidelines for test data

You can import data to the test panel in the cleanse asset and save the test data in the asset configuration.

Consider the following rules and guidelines when you import or enter test data:

- The import option supports CSV and Microsoft Excel files.
- You can import up to 200 consecutive rows of data from a delimited file. You can specify the row at which the import starts.

  Note: Before you import, check the file for column headings. If the first row in the import file contains column headings, start the import at line 2 or lower.
- The Maximum string length property applies to the data in the test panel in addition to the strings that the asset reads and writes at run time. When you test the asset, import or enter a string on each row that does not exceed the value on the property.

  You can set a maximum string length of 10,000 characters. The default maximum length is 255 characters.
- If you import a CSV file that contains multiple columns or uses a text qualifier, verify that the file uses a delimiter or a text qualifier that the Secure Agent recognizes. By default, the *Comma* option is the delimiter for the column data. By default, the *No quotes* option is the text qualifier for the data. You can update the delimiter and text qualifier characters when you select the data to import. The Delimiter and Text Qualifier options are not required when you import a Microsoft Excel file.
- The Secure Agent saves the data that you import to the asset when you save the asset. If you change an option in the asset configuration, you may lose any unsaved test data.

Testing the merge operation

Test the merge field records that you configured in the **Merge** tab.

1. Select the **Merge** tab on the asset.
2. Select the merged field that you want to test.
3. Select a Secure Agent to run the test.

   To refresh the list of active Secure Agents, click the **Refresh** icon.
4. Enter data values in the test panel, or import the data to test. To import the data, click the **Import** option in the test panel.
5. Click **Test**.

   The Output column displays the test results. The results represent the application of the step logic that you define in the **Configuration** tab for the input fields.
6. Verify the result of the test.

   You can sort and filter the test data, and you use the Find feature to search for values in the test data.

INDEX

C

cleanse assets
  and dimensions 17
  and mappings 16
  creating a cleanse asset 14
  steps 13, 18
  testing a cleanse asset 27
Cloud Application Integration community
  URL 5
Cloud Developer community
  URL 5

D

Data Integration community
  URL 5

I

Informatica Global Customer Support
  contact information 6
Informatica Intelligent Cloud Services
  web site 5

M

maintenance outages 6

S

status
  Informatica Intelligent Cloud Services 6
steps
  configuration 18
  in cleanse assets 13
system status 6

T

trust site
  description 6

U

upgrade notifications 6

W

web site 5