

IBM InfoSphere QualityStage Standardization Rules  
Designer  
Version 11 Release 3

*Tutorial: Enhancing a product rule set  
in the Standardization Rules Designer*





IBM InfoSphere QualityStage Standardization Rules  
Designer  
Version 11 Release 3

*Tutorial: Enhancing a product rule set  
in the Standardization Rules Designer*



**Note**

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page 43.

---

# Contents

## **Tutorial: Enhancing a product rule set in the Standardization Rules Designer . . . 1**

Setting up the tutorial environment. . . . .	6
Copying the tutorial files to a directory . . . . .	6
Creating the tutorial project . . . . .	7
Importing the tutorial rule set . . . . .	7
Module 1: Creating a revision for a rule set and importing sample data . . . . .	9
Lesson 1.1: Viewing an SQA report . . . . .	9
Lesson 1.2: Viewing rule sets and creating a revision in the Standardization Rules Designer . . . . .	11
Lesson 1.3: Importing sample data. . . . .	12
Module 2: Classifying values . . . . .	14
Lesson 2.1: Assigning values to an existing class by adding classification definitions . . . . .	14
Lesson 2.2: Adding and populating a custom class. . . . .	18
Module 3: Adding a lookup table . . . . .	21
Lesson 3.1: Adding a lookup table. . . . .	21
Lesson 3.2: Importing and adding lookup table definitions. . . . .	22
Module 4: Adding and modifying rules . . . . .	25

Lesson 4.1: Modifying a rule to handle data correctly . . . . .	26
Lesson 4.2: Identifying unhandled records and customizing output columns. . . . .	28
Lesson 4.3: Adding a rule for the pattern with the most unhandled records . . . . .	30
Lesson 4.4: Adding a rule to split values into different output columns . . . . .	32
Lesson 4.5: Publishing a revision . . . . .	34

## **Appendix A. Product accessibility . . . 37**

## **Appendix B. Contacting IBM . . . . . 39**

## **Appendix C. Accessing the product documentation . . . . . 41**

## **Notices and trademarks . . . . . 43**

## **Index . . . . . 49**



---

## Tutorial: Enhancing a product rule set in the Standardization Rules Designer

In this tutorial, you will use the IBM® InfoSphere® QualityStage® Standardization Rules Designer to enhance a rule set that standardizes product data. When you standardize data, you implement data quality standards that normalize data values and prepare data for uses such as matching and reporting.

In this tutorial, you will use data from the fictional Sample Outdoor Company, which sells and distributes products to third-party retailer stores and consumers. For the last several years, the company has steadily grown into a worldwide operation, selling their line of products to retailers in nearly every part of the world.

The fictional Sample Outdoor Company recently acquired several new product lines. The company wants to integrate the data for these product lines into its current database, but the new data contains new types of information and is formatted inconsistently. The Sample Outdoor Company can use the IBM InfoSphere QualityStage Standardization Rules Designer to enhance the rule set that standardizes this type of data.

After the rule set is enhanced, the company will apply the rule set to a Standardize stage in a standardization job. When the standardization job is run, input data is standardized according to the logic that is specified in the enhanced rule set.

This tutorial guides you through some of the common tasks that you might complete when you enhance a rule set in the Standardization Rules Designer. The following steps illustrate the sequence of actions in the tutorial:

1. In module 1, you identify a rule set that requires enhancement and open a revision for that rule set in the Standardization Rules Designer. You also import sample data to use in the Standardization Rules Designer.
2. In module 2, you categorize the parts of the data. You add classification definitions that assign new values to existing classes and add a custom class for a new type of data. Figure 1 shows how each value in an example record for a fictional Sample Outdoor Company product can be assigned to a class.

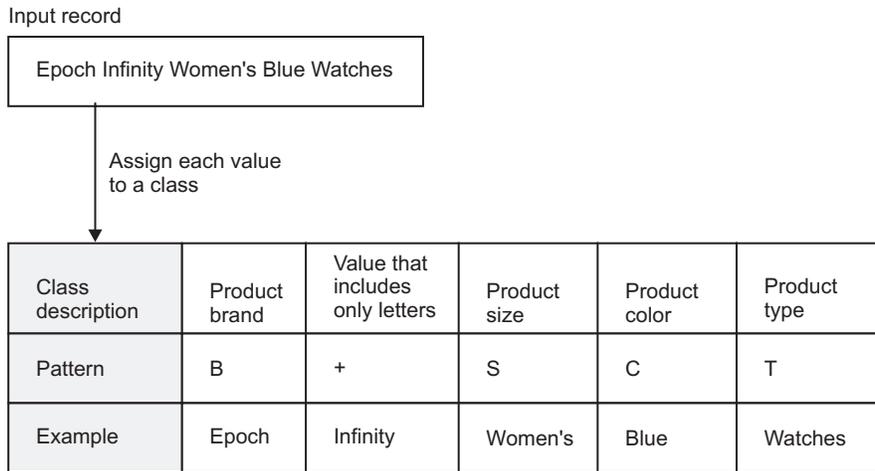


Figure 1. Each value in an example record is assigned to a class.

3. In module 3, you add a lookup table that converts alphabetic information about product colors to numeric color codes. Figure 2 shows part of the lookup table that is added in the Standardization Rules Designer.

Value (color)	Returned value (color code)
Blue	903
Silver	923
Grey	913

Figure 2. The lookup table converts an alphabetic color to a numeric color code.

4. In the first lesson of module 4, you modify a rule that was added in the Standardization Rules Designer. The rule is handling data incorrectly for some of the new product brands. Figure 3 shows the output with the current rule and the output after the rule is modified according to the data cleansing requirements of the fictional Sample Outdoor Company.

Input record

B	+	S	C	T
Epoch	Infinity	Women's	Blue	Watches

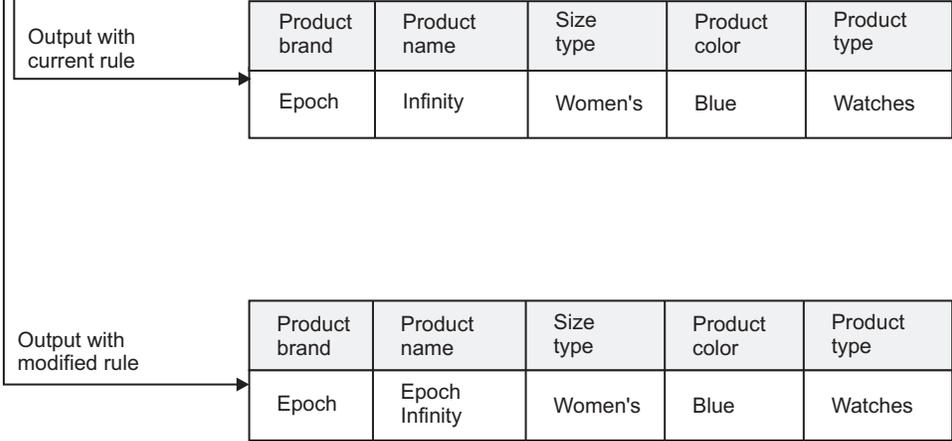
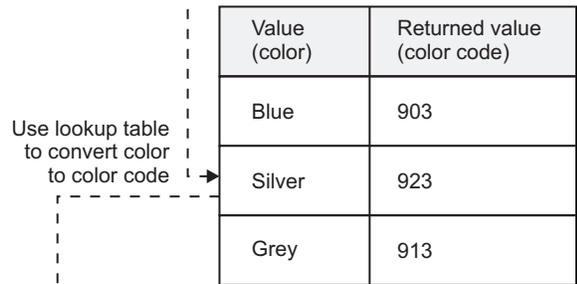


Figure 3. The current rule does not meet the data cleansing requirements of the fictional Sample Outdoor company. The modified rule duplicates the product brand name in the product name.

- 5. In the second and third lessons of module 4, you identify the most common pattern that is unhandled and add a rule to handle data that matches the pattern. Figure 4 shows how an example record for a fictional Sample Outdoor Company product is handled by the new rule.

Input record for unhandled pattern

B	+	S	+	C	T
Relax	Trendi	One	Size	Silver	Eyewear



Output

Product brand	Product name	Size type	Color code	Product type
Relax	Trendi	One Size	923	Eyewear

Figure 4. A new rule for the most common unhandled pattern uses the lookup table to convert color to color code. The rule also adds the product size values to the output column for size types.

6. In the fourth lesson of module 4, you create a rule that handles two distinct values that are concatenated in the input data. For example, if the input data contains the value 195cm, you can create a rule that splits the value into the values 195 and cm and places them in the appropriate output columns. Figure 5 shows how an example record for a fictional Sample Outdoor Company product is handled by this rule.

Input record for concatenated values

B	+	>	C	T	T
Hibernator	Pad	195cm	Grey	Sleeping	Bag

Output

Product brand	Product name	Size unit	Size type	Product color	Product type
Hibernator	Pad	195	cm	Grey	Sleeping Bags

Figure 5. The new rule splits a concatenated value into two distinct values and places them in the appropriate output columns. The rule also adds the product type values to the output column for product types.

### Learning objectives

By completing the modules, you will learn about the concepts and tasks for enhancing rule sets:

- Import sample data to see how records from the sample data are affected by changes to parts of the rule set
- Use classifications to categorize parts of your data
- Add lookup tables to compare or convert data to specified values
- Create rules that apply actions to a group of related records

### Time required

Before you begin the tutorial, you must set up your environment. The time that is required for setup depends on your current environment.

The modules each take 20 - 60 minutes to complete.

### System requirements

The following components and applications must be installed on your system.

- IBM InfoSphere QualityStage Standardization Rules Designer
- IBM InfoSphere QualityStage with the following clients:
  - IBM InfoSphere DataStage® and QualityStage Designer
  - IBM InfoSphere DataStage and QualityStage Administrator

## Prerequisites

Before you begin this tutorial, you must understand data quality concepts such as standardization, classification, and rules for data cleansing. Knowledge about InfoSphere DataStage and QualityStage concepts such as jobs, stages, and reports might be helpful, but is not required.

**Notices:** The Sample Outdoor Company, GO Sales, any variation of the Great Outdoors name, and Planning Sample, depict fictitious business operations with sample data used to develop sample applications for IBM and IBM customers. These fictitious records include sample data for sales transactions, product distribution, finance, and human resources. Any resemblance to actual names, addresses, contact numbers, or transaction values, is coincidental. Other sample files may contain fictional data manually or machine generated, factual data compiled from academic or public sources, or data used with permission of the copyright holder, for use as sample data to develop sample applications. Product names referenced may be the trademarks of their respective owners. Unauthorized duplication is prohibited.

---

## Setting up the tutorial environment

Before you start this tutorial, you must prepare your system to run the tutorial.

### Before you begin

- Ensure that the IBM InfoSphere QualityStage server component and the IBM InfoSphere DataStage and QualityStage Administrator client component is installed and configured appropriately.
- Ensure that user credentials are mapped to the engine tier operating system user credentials and that your user account has the following sets of security roles:
  - The Suite Administrator and DataStage and QualityStage Administrator roles are required to create a project. If you do not have these roles, a user with these roles must create a project for the tutorial and assign you the DataStage and QualityStage Developer or DataStage and QualityStage Production Manager role for the project.
  - The Suite User and DataStage and QualityStage Administrator or DataStage and QualityStage User roles are required to log in to the Standardization Rules Designer.

If you cannot log in to the Standardization Rules Designer, consult your system administrator.

### About this task

The IBM InfoSphere QualityStage Standardization Rules Designer is a web-based application that interacts with the client, engine, and metadata repository tiers. Before you begin the tutorial, you must ensure that the sample data and tutorial rule set are available to the Standardization Rules Designer. You must also have a project that you can use for the tutorial.

## Copying the tutorial files to a directory

You copy the tutorial files to a directory that you use during the tutorial. To complete the tutorial, you must have access to the tutorial files on the computer where you use the Standardization Rules Designer.

## Procedure

1. Locate the `Standardization_Rules_Designer_product_tutorial.zip` file, which is on the installation media. In the directory that contains the installation media, the `Standardization_Rules_Designer_product_tutorial.zip` file is in the `parent_directory\TutorialData\QualityStage` directory. For example, the `Standardization_Rules_Designer_product_tutorial.zip` file might be in the `is-client\TutorialData\QualityStage` directory.
2. Extract the files from the `Standardization_Rules_Designer_product_tutorial.zip` file to a directory on the computer that you will use to connect to the Standardization Rules Designer.
3. If the directory that you extracted the files to is not on a computer where the client tier or engine tier is installed, copy the `Camping_Rules_for_Tutorial.isx` file to a computer with the client tier or engine tier.

## Creating the tutorial project

You create a new project for the tutorial to keep your tutorial exercises separate from the other work in InfoSphere QualityStage.

### Procedure

1. To start the IBM InfoSphere DataStage and QualityStage Administrator client, click **Start > All Programs > IBM InfoSphere Information Server > IBM InfoSphere DataStage and QualityStage Administrator**.
2. In the Attach to DataStage window, type your user name and password, and then click **Login**.
3. On the Projects page, click **Add**.
4. In the Add Project window, enter `StandardizationRulesDesignerTutorial` in the **Name** field, and then click **OK**.
5. If other users will complete the tutorial, assign the users an appropriate role for the project.
  - a. Select the `StandardizationRulesDesignerTutorial` project, and then click **Properties**.
  - b. Click the **Permissions** tab.
  - c. If the user that you want to add roles for is not in the list, add the user to the project.
  - d. From the list of project users, select the user that will complete the tutorial.
  - e. From the **User Roles** list, select **DataStage and QualityStage Developer** or **DataStage and QualityStage Production Manager**, and then click **OK**.
6. In the InfoSphere DataStage Administration window, click **Close**.

## Importing the tutorial rule set

You use the InfoSphere Information Server Manager to import the tutorial rule set into the project that you created for the tutorial.

### Before you begin

- “Copying the tutorial files to a directory” on page 6
- “Creating the tutorial project”

### Procedure

1. Click **Start > Programs > IBM InfoSphere Information Server > IBM InfoSphere Information Server Manager**.

2. Connect to the metadata repository that contains the tutorial project.
  - a. In the Information Server Manager application window, right-click in the Repository view, and then click **Add Domain**.
  - b. Specify the application server and an InfoSphere Information Server administrator user name and password.
  - c. Click **OK**.
3. In the Repository navigation tree, right-click the StandardizationRulesDesignerTutorial project, and then click **Import**.
4. Import the `Camping_Rules_for_Tutorial.isx` file.
  - a. Browse to the directory that contains the file. By default, the file is in the `Standardization_Rules_Designer_product_tutorial` directory where you copied the tutorial files.
  - b. Select the file, and then click **Open**.
  - c. Click **Import**.
  - d. Ensure that the StandardizationRulesDesignerTutorial project is selected, and then click **OK**.

---

## Module 1: Creating a revision for a rule set and importing sample data

In this module, you will identify a standardization rule set that requires enhancement and create a revision for that rule set in the Standardization Rules Designer. You will then import sample data to use in the Standardization Rules Designer.

### Learning objectives

After completing the lessons in this module, you will understand the relevant concepts and tasks for rule sets and sample data:

- Determine whether a rule set requires enhancement by using a Standardization Quality Assessment (SQA) report
- Create a revision for a rule set in the Standardization Rules Designer
- Import sample data

### Time required

This module takes approximately 20 minutes to complete.

### Prerequisites

Ensure that you completed the setup steps and have tutorial files and data loaded on your computer system.

## Lesson 1.1: Viewing an SQA report

In this lesson, you identify that a rule set does not meet standardization requirements for retail product data by viewing Standardization Quality Assessment (SQA) reports for that rule set. You can review your standardization results as part of quality control processes and ensure that requirements for data quality are met.

### Overview

The fictional Sample Outdoor Company applies the CAMPING.SET rule set to data about its retail product lines.

### Learn more about rule sets:

Rule sets check and normalize input data. You can use rule sets to help understand the content and structure of source data and to standardize data and prepare the data for matching. Rule sets contain the following elements, which are discussed in greater detail as you work with them in the tutorial:

- Classifications
- Lookup tables
- Rules
- Output columns

Rule sets are used in jobs that include an Investigate stage or Standardize stage.

In the past, the CAMPING.SET rule set standardized all retail product data correctly. However, the fictional Sample Outdoor Company wants to use the same

rule set for several new product lines that the company acquired. To determine whether the current rule set can standardize the new data effectively, the company has run SQA reports.

### Learn more about SQA reports:

Standardization Quality Assessment (SQA) reports assess the results of the standardization process. You can use the reports to verify that the results meet data quality objectives. SQA reports are generated by the SQA stage, which uses data that has been processed by a Standardize stage.

The high-level SQA report and low-level SQA report complement each other, although each type of report is generated separately:

#### Standardization Quality Assessment (SQA) report

Provides statistics about the standardization process, such as how much data is not standardized.

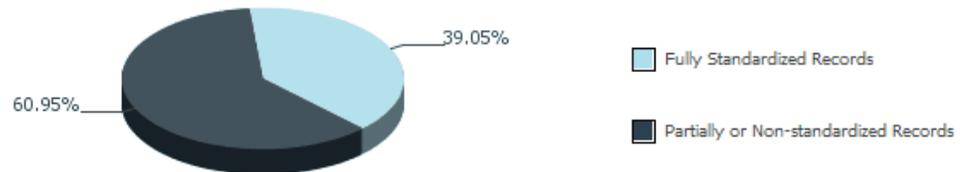
#### Standardization Quality Assessment (SQA) Record Examples report

Provides examples of processed records.

In this lesson, you will view SQA reports for a standardization job that applied the current CAMPING.SET rule set to data about new product lines. The reports are included in the tutorial files that you copied to a directory on your computer.

### Procedure

1. From the Standardization\_Rules\_Designer\_product\_tutorial directory on your computer, open the SQAreport SQA report.
2. Review the report. The standardization summary shows that only 39% of the records for new product lines were fully standardized by the standardization job that used the CAMPING.SET rule set.



3. From the Standardization\_Rules\_Designer\_product\_tutorial directory on your computer, open the SQA\_record\_examples SQA Record Examples report.
4. Review the report. The first set shows some of the unhandled records, which match different patterns. Patterns are comprised of characters, each of which represents the class of a value in the record. For example, the record TrailChef Canteen 2liters Brown Cooking Gear matches the pattern B+>CTT.

**Set 1 of 3 (60.95%, 167 of 274 total records) - 25 record examples shown**

Input Record	InputPattern	UnhandledPattern
<u>TrailChef</u> Water Bag 10 liters Clear Cooking Gear	B++^SCTT	B++^SCTT
<u>TrailChef</u> Canteen 2liters Brown Cooking Gear	B+>CTT	B+>CTT
<u>TrailChef</u> Kitchen Kit 15-piece Unspecified Cooking Gear	B++^SUTT	B++^SUTT
<u>TrailChef</u> Cup 330ml Silver Cooking Gear	B+>CTT	B+>CTT
<u>TrailChef</u> Cook Set 6-piece Silver Cooking Gear	B++^SCTT	B++^SCTT
<u>TrailChef</u> Deluxe Cook Set 9-piece Silver Cooking Gear	B+++^SCTT	B+++^SCTT
<u>TrailChef</u> Single Flame Single-burner Silver Cooking Gear	BS+S-SCTT	BS+S-SCTT
<u>TrailChef</u> Double Flame Double-burner Green Cooking Gear	BS+S-SCTT	BS+S-SCTT
<u>TrailChef</u> Kettle 1.5liters Ash Cooking Gear	B+@CTT	B+@CTT
<u>TrailChef</u> Utensils 4-piece Silver Cooking Gear	B+^SCTT	B+^SCTT
<u>Star Lite</u> Two-person Beige Tents	B+S-SCT	B+S-SCT
<u>Star Dome</u> Three-person Red Tents	B+S-SCT	B+S-SCT
<u>Star Gazer 2</u> Two-person Light Blue Tents	B+^S-SCCT	B+^S-SCCT
<u>Star Gazer 3</u> Three-person Red Tents	B+^S-SCT	B+^S-SCT

Because the patterns that are listed in the SQA Example Records report might not be the most common patterns, you might not want to add rules for all of those patterns. You can use the Standardization Rules Designer to identify the most common patterns and add rules for those patterns. When you add rules for common patterns, you affect the largest percentage of the unhandled data and improve the standardization process most efficiently.

You determine that the CAMPING.SET rule set does not meet the current standardization requirements for retail product data.

In the next lesson, you will view the CAMPING.SET rule set in the Standardization Rules Designer and create a revision for the rule set so that you can enhance it.

## Lesson 1.2: Viewing rule sets and creating a revision in the Standardization Rules Designer

In this lesson, you find the CAMPING.SET rule set in the Standardization Rules Designer and create a revision for that rule set. When you create a revision, you can enhance the rule set through multiple iterations, working at different times and in collaboration with other users.

### Overview

Because the CAMPING.SET rule set does not meet the data quality requirements of the fictional Sample Outdoor Company, the rule set must be enhanced. Before you can enhance this rule set, you must create a revision for the rule set in the Standardization Rules Designer.

### Learn more about revisions:

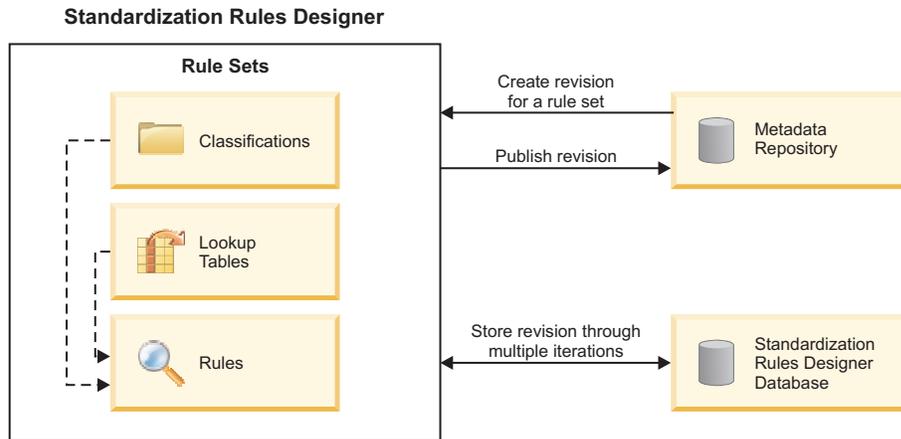
A revision is a collection of changes that were made to the rule set in the Standardization Rules Designer. One revision can be created for each rule set.

When you enhance a rule set in the Standardization Rules Designer, you work with a copy of the rule set that is stored in a database for the Standardization Rules Designer. You can keep a revision open for as long as you want, and the changes to the rule set can be viewed and modified by other users. When a revision is open, you can enhance the rule set in a collaborative way through multiple iterations that affect only the copy of the rule set. If you collaborate with other users, only one user can open the revision at a time.

When you want to update the actual rule set, you can publish the revision. When you publish the revision, the rule set in the metadata repository is updated with the changes that you made in the Standardization Rules Designer.

If you want to discard the current changes in the revision but keep the revision open, you can reset the revision. When you reset the revision, the version of the rule set in the Standardization Rules Designer database is replaced by the version in the metadata repository.

The following diagram shows how you can work with revisions throughout the lifecycle of a rule set in the Standardization Rules Designer.



### Procedure

1. To open the Standardization Rules Designer, in your browser, go to [https://host\\_name:secure\\_port\\_number/ibm/iis/qs/StandardizationRulesDesigner/](https://host_name:secure_port_number/ibm/iis/qs/StandardizationRulesDesigner/), where *host\_name* is the host name of the server where the Standardization Rules Designer is installed.
2. Log in to the Standardization Rules Designer. A list of servers that host IBM InfoSphere QualityStage projects is shown.
3. Expand the twistie for the server on which you created the StandardizationRulesDesignerTutorial project.
4. Expand **StandardizationRulesDesignerTutorial**.
5. Select the CAMPING.SET rule set, and then click **Edit**.
6. In the window for the message about editing the rule set, click **OK**. A revision is created for the CAMPING.SET rule set, and the Home page for the rule set is shown.

## Lesson 1.3: Importing sample data

In this lesson, you import sample data to use in the Standardization Rules Designer.

### Overview

The changes that you make to a rule set in the Standardization Rules Designer are informed by the sample data that you use. When you enhance a rule set in the Standardization Rules Designer, you can view how the changes affect the way that values and records in your sample data are handled. To work with your data, you must import it into the Standardization Rules Designer.

The fictional Sample Outdoor Company has prepared a sample data set to use in the Standardization Rules Designer. The sample data includes records for the new product lines that the fictional Sample Outdoor Company acquired.

### **Procedure**

1. In the Standardization Rules Designer, click the **Home** tab.
2. In the navigation pane, click **Import Sample Data**.
3. Next to the **Select a Source** field, click **Import**.
4. Import the product\_sample\_data.csv file:
  - a. Browse to the directory that contains the file. By default, the file is in the Standardization\_Rules\_Designer\_product\_tutorial directory where you copied the tutorial files.
  - b. Select the file, and then click **Open**.

The Standardization Rules Designer shows the raw records that were imported and example records that were processed by the separation characters and strip characters.

### **What's next**

In this module, you completed the following tasks:

- Determined that a rule set required enhancement by viewing unhandled records in a Standardization Quality Assessment (SQA) report
- Created a revision for a rule set in the Standardization Rules Designer
- Imported sample data

You can log out of the Standardization Rules Designer and complete the next module later. The revision remains open.

In the next module, you will strengthen the contextual information that the new retail product data provides by categorizing parts of the data.

---

## Module 2: Classifying values

In this module, you will strengthen the contextual information that the new data provides by categorizing parts of the data. You will add classification definitions that assign new values to existing classes and add a custom class to categorize the data.

The data for the new product lines that the fictional Sample Outdoor Company acquired includes the following types of values:

- New values to assign to an existing class
- A new type of value that the company wants to create a class for

Because some of the brand names for the new products are not assigned to the product brand class, rules that handle patterns with the product brand class do not handle the new data. In the first part of this module, you will identify how brand names for the new products are currently classified. To ensure that the brand names are handled by appropriate existing rules, you will add classification definitions that assign new product brands to the product brand class.

Additionally, some of the records for the product lines that the fictional Sample Outdoor Company recently acquired include information about the unit that is used to measure the size of the product, such as CM or CC. To categorize these values, you will add a custom class for unit of measure. After the class is added, you can write rules that properly handle the size data.

### Learning objectives

After completing the lessons in this module, you will understand the relevant concepts and tasks for classifying values:

- Identify the class that a value belongs to
- Add classification definitions
- Add a custom class

### Time required

This module takes approximately 30 minutes to complete.

### Prerequisites

Ensure that you completed the setup steps and that you have the tutorial files and sample data loaded on your computer.

## Lesson 2.1: Assigning values to an existing class by adding classification definitions

In this lesson, you identify how brand names for new product lines are currently classified. You then assign brand name values to an existing custom class by adding classification definitions.

### Overview

The fictional Sample Outdoor Company wants to ensure that records that contain the new product brand names are handled correctly. You can view the class that

new brand name values belong to, and then use that information to see which rules, if any, handle the values. You can then assign brand name values to a class by adding classification definitions.

**Learn more about classifications:**

In IBM InfoSphere QualityStage, records are represented as patterns. In the same way that a record is comprised of one or more values, patterns are comprised of one or more abstract characters, each of which represents a class. Classes strengthen the contextual information that patterns provide by identifying that the underlying value belongs to a particular category. For example, the fictional Sample Outdoor Company has records such as TRAKKER ZODIAK WOMEN'S GREEN WATCHES that are represented by the pattern B+SCT. The following table shows the contextual information that each class in the pattern B+SCT provides.

*Table 1. Each one-character class label indicates that the underlying value belongs to a particular category.*

<b>B</b>	<b>+</b>	<b>S</b>	<b>C</b>	<b>T</b>
Product Brand	Value that includes only letters	Product Size	Product Color	Product Type

Information about values is provided by the following types of classes:

- Default classes provide basic information about the type of the value, such as whether the value is comprised of alphabetic characters, numeric characters, or some combination of both. Default classes are represented by a range of characters that is listed in the Standardization Rules Designer.
- Custom classes provide stronger contextual information about the type of the value, such as whether an alphabetic value is the name of a product or the name of a brand. The one-character label for custom classes can be any letter in the Latin alphabet or 0, which indicates a null class.

**Tasks in this lesson**

1. "Identifying the class that a value belongs to"
2. "Assigning values to a class" on page 17

**Identifying the class that a value belongs to**

You can identify the default or custom class to which a brand name value is currently assigned. If the brand name values are not assigned to the B Product Brand custom class, records that contain those values might not be handled by the appropriate rules.

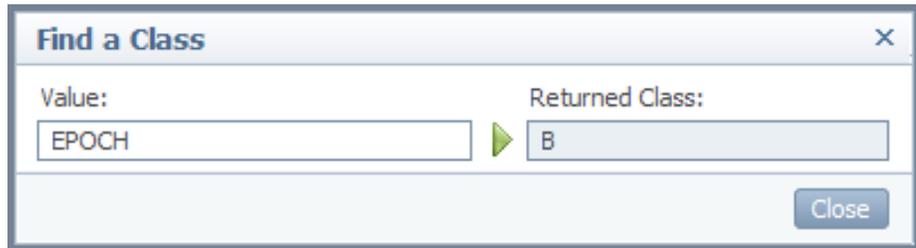
The product lines that the fictional Sample Outdoor Company acquired include the following product brand names:

- EPOCH
- ANTONI
- POLAR

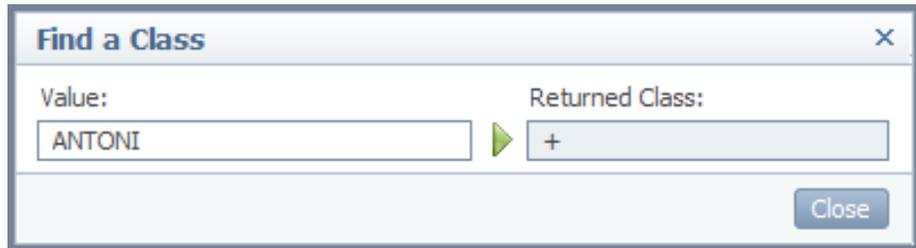
To ensure that records that contain these brand names are handled properly, you must first identify what class the values belong to. You can also view which patterns contain the B Product Brand class and view rules that handle records with those patterns.

**Procedure**

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. Click **Find a Class**.
3. Identify the class that the brand name values belong to:
  - a. In the **Value** field, enter the value EPOCH, and then click . The **Returned Class** field shows that this value is assigned to the B Product Brand custom class.



- b. Repeat step 3a for the values ANTONI and POLAR. The **Returned Class** field shows that these values are assigned to the + default class.



- c. Click **Close**.
4. To access the patterns in the data and the rules for those patterns, open a rule group:
  - a. Click the **Rules** tab.
  - b. Select the Input\_Overrides rule group, and then click **Open**.
5. Expand **Pattern Rule**. A list of patterns in the data is shown.

▼ Pattern Rule		<a href="#">Add Pattern...</a>
Rule	Frequency	Unhandled
▶  B+SCT	26.64%	0%
▶  ++SCT	9.12%	9.12%
▶  B+S+CT	8.39%	8.39%
▶  B++SCT	4.74%	4.74%
▶  B+^SCT	4.74%	4.74%

6. Expand **B+SCT**, and then expand **Copy product data to output columns**. A list of example records that are handled by this rule is shown. The list of records includes records that contain the value EPOCH. Because this value is assigned to the B Product Brand class, the rule for this pattern handles records that contain the value.

7. Expand **++SCT**, and then expand **Unhandled Records**. A list of example records that match this pattern is shown. The list of records includes records that contain the values ANTONI and POLAR. Because these values are not assigned to the B Product Brand class, records that contain the values are not handled by the appropriate rule.
8. To exit the Input\_Overrides rule group, click **Select Rule Group** at the top of the Rules page.

To ensure that records that contain the values ANTONI and POLAR are handled by the appropriate rule, you assign these values to the B Product Brand class.

### Assigning values to a class

You can assign the new brand values to the product brand class by adding classification definitions manually.

#### Learn more about classification definitions:

Classification definitions include the following elements:

**Value** The input value that is defined.

#### Standard Value

A standardized spelling or representation of the value that an action or condition in a rule can use instead of the input value.

**Class** The class that the value is assigned to. The class is represented by a one-character class label.

#### Similarity Threshold

The degree of variation that is allowed in the spelling or representation of the value. The similarity threshold must be in the range 700 - 900. A similarity threshold of 900 indicates that only the exact value in the definition is affected by the definition.

If the similarity threshold is lower than 900, the definition might affect values that are similar to the value in the definition. For example, if you create a classification definition for the value GRAMS and set the similarity threshold to 800, the definition also affects the value GRAM.

#### Procedure

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. Click **Define Values**.
3. Add a definition for a product brand value:
  - a. Enter the value ANTONI.
  - b. From the **Class** list, select **B**.
  - c. Expand **Status of Definitions** and verify that **Active definition** is selected. An inactive definition does not affect the value.
  - d. Click **OK**.
4. To access the patterns in the data and the rules for those patterns, open a rule group:
  - a. Click the **Rules** tab.
  - b. Select the Input\_Overrides rule group, and then click **Open**.
5. Expand **Pattern Rule**. A list of patterns in the data is shown.

6. Expand **B+SCT**, and then expand **Copy product data to output columns**. A list of example records that are handled by this rule is shown. The list of records now includes records that contain the value ANTONI.
7. Expand **++SCT**, and then expand **Unhandled Records**. A list of example records that have this pattern is shown. The list of records still includes records that contain the value POLAR.

## Lesson 2.2: Adding and populating a custom class

In this lesson, you add a custom class to designate a new type of value. You then assign an initial set of values to this class by adding multiple classifications at the same time.

### Overview

The fictional Sample Outdoor Company assigns all values that relate to size to the S Product Size custom class. The class includes both clothing sizes, such as WOMENS, JUNIOR, and LARGE, and units of measure such as CM and ML. To distinguish clothing sizes from units of measure, the company wants to add a separate custom class for units of measure.

### Learn more about custom classifications:

A custom classification provides stronger contextual information about a value than a default classification. For example, if a definition does not assign the value TRAILCHEF to a custom class, the value is assigned to the + default class. This default class indicates that the value is a single alphabetic word. If a definition assigns the value to the Product Brand custom class, which is represented by the character B, the value provides more contextual information. When this information is provided, you can write rules that address a specific subset of the data and therefore handle that data more effectively.

### Tasks in this lesson

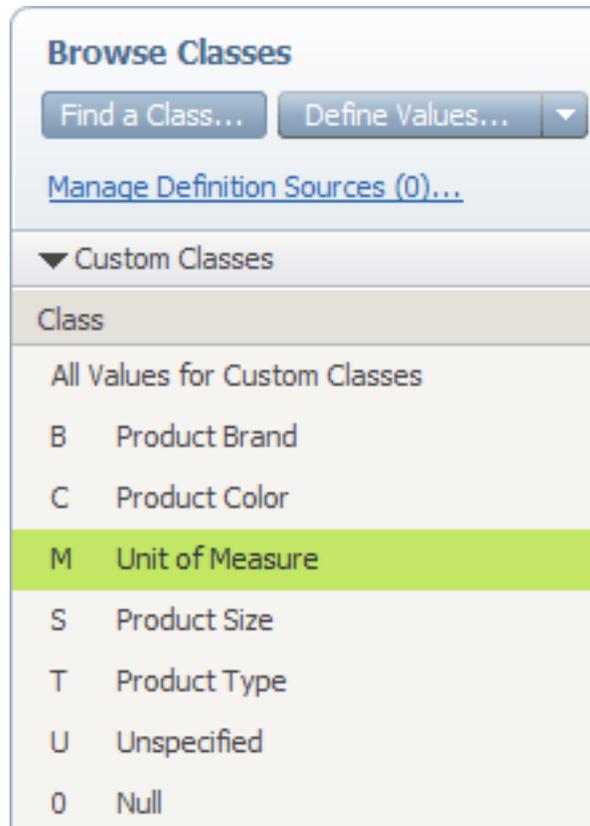
1. "Adding a custom class"
2. "Adding multiple classification definitions" on page 19

### Adding a custom class

To create a class for units of measure, you can add the M custom class from the list of available classes.

### Procedure

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. In the **Custom Classes** section, click **Add Custom Class**.
3. From the **Available Classes** list, select **M**.
4. In the **Description** field, enter Unit of Measure.
5. Click **Add**. An entry for the new class is shown in the list.



6. Click **Close**.

### Adding multiple classification definitions

You can assign more than one value to the new M class by adding multiple classification definitions at the same time.

If you add a new custom class in the Standardization Rules Designer, no values are assigned to that class initially. For example, to populate the M custom class, you add classification definitions for values that represent a unit of measure. After values are assigned to the class, you can add or modify rules to handle unit of measure values.

In the retail product data set, the following values represent a unit of measure:

- CC
- CM
- M
- ML
- MM

#### Procedure

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. Click **Define Values**.
3. In the lower right of the Define Values window, click **Define Multiple**.
4. Select the **Apply class to all definitions** check box.
5. From the **Select Class** list, select **M**. In the Class column, **M** is selected automatically for each value that you define.

6. Enter CC in the first field in the Value column, then click outside the field. A second row is enabled.
7. Enter the following values by using the method from step 6.
  - CM
  - M
  - ML
  - MMEnter each value in a separate row.
8. Expand **Status of Definitions** and verify that **Active definition** is selected. An inactive definition does not affect the value.
9. Click **OK**.

To view the values that are assigned to the M custom class, expand **Custom Classes** and select **M**.

### **What's next**

In this module, you completed the following tasks:

- Identified the class that a value belongs to
- Added classification definitions
- Added a custom class

You can log out of the Standardization Rules Designer and complete the next module later. The revision remains open.

In the next module, you will add a lookup table that converts alphabetic information about product colors to numeric color codes.

---

## Module 3: Adding a lookup table

In this module, you will add a lookup table that converts alphabetic information about product colors to numeric color codes.

When the fictional Sample Outdoor Company sells products to third-party retailers, the company provides product information that includes the color of the product. As the company has expanded, the number and range of product colors has grown.

To avoid confusion about the growing number of colors, the company assigned a numeric color code to each color in its database. The company also created a set of lookup table definitions to convert the alphabetic product color information to the numeric code for that color.

In this module, you will add a new lookup table and import lookup table definitions into the Standardization Rules Designer. You will then add lookup table definitions for new colors that are not included in the current lookup table. In module 4, you will use this lookup table in a rule.

### Learning objectives

After completing the lessons in this module, you will understand the relevant concepts and tasks for lookup tables:

- Add a lookup table
- Import lookup table definitions
- Add lookup table definitions

### Time required

This module takes approximately 20 minutes to complete.

### Prerequisites

Ensure that you completed the setup steps and that you have the tutorial files and sample data loaded on your computer.

## Lesson 3.1: Adding a lookup table

In this lesson, you add a lookup table to the CAMPING.SET rule set that converts alphabetic information about product colors to numeric color codes.

### Overview

The fictional Sample Outdoor Company wants to use a lookup table to convert alphabetic information about product colors to numeric color codes. You can add a lookup table to the rule set in the Standardization Rules Designer. Later, you can add a rule that references the lookup table as part of an action.

### Learn more about using lookup tables:

Lookup tables contain definitions that rules can reference as part of actions or conditions. Actions or conditions might use a lookup table in the following ways:

- An action or condition can compare a particular value to a value in the lookup table. For example, a condition can stipulate that a rule handles a record only when a value in a particular position in the record is in the lookup table.
- An action can convert a particular value to a value in the lookup table. For example, an action might use a lookup table that contains geographic information to convert numeric place codes to place names.

#### Procedure

1. In the Standardization Rules Designer, click the **Lookup Tables** tab.
2. Click **Add Lookup Table**.
3. Enter information about the product color lookup table:
  - a. In the **Name** field, enter `Color_Code`.
  - b. In the **Description** field, enter Converts product color to color code.
  - c. Click **OK**.

## Lesson 3.2: Importing and adding lookup table definitions

In this lesson, you will import and add lookup table definitions to the `Color_Code` lookup table.

#### Overview

The fictional Sample Outdoor Company created a CSV file that contains lookup table definitions for the `Color_Code` lookup table. You can import these definitions into the lookup table that you added in the Standardization Rules Designer.

Recently, the company added a new color code to its product line and changed the color code for an existing product color. You can update the lookup table by adding lookup table definitions in the Standardization Rules Designer.

#### Learn more about lookup table definitions:

Lookup table definitions include the following elements:

**Value** The input value that is defined.

#### Returned Value

A value that an action or condition in a rule can use instead of the input value.

#### Similarity Threshold

The degree of variation that is allowed in the spelling or representation of the value. The similarity threshold must be in the range 700 - 900. A similarity threshold of 900 indicates that only the exact value in the definition is affected by the definition.

If the similarity threshold is lower than 900, the definition might affect values that are similar to the value in the definition. For example, if you create a lookup table definition for the value `GRAMS` and set the similarity threshold to 800, the lookup table definition also affects the value `GRAM`.

#### Procedure

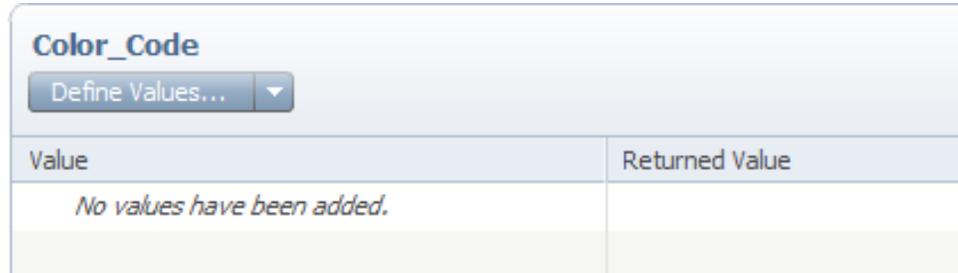
1. "Importing lookup table definitions" on page 23
2. "Adding lookup table definitions" on page 23

## Importing lookup table definitions

You can import lookup table definitions from one or more files into a lookup table in the Standardization Rules Designer. Lookup table definitions that you do not import must be added manually in the Standardization Rules Designer.

### Procedure

1. In the Standardization Rules Designer, click the **Lookup Tables** tab.
2. From the list of lookup tables, select the `Color_Code` table. A list of values in the lookup table is shown in the right pane. Because no values are defined for this lookup table, the list is empty.

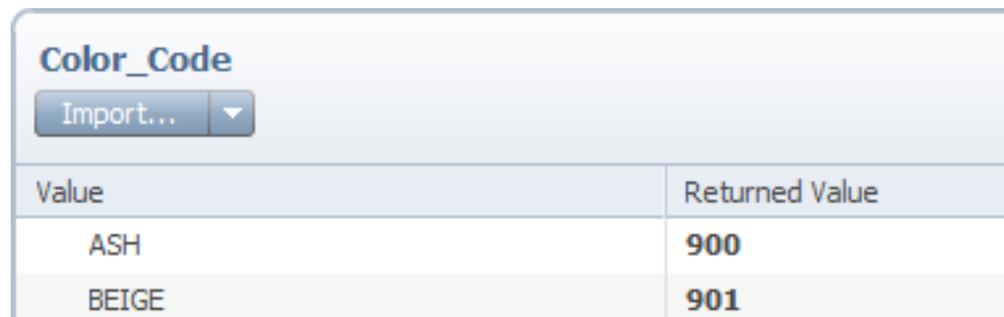


The screenshot shows a window titled "Color\_Code" with a "Define Values..." button. Below the button is a table with two columns: "Value" and "Returned Value". The table is currently empty, with the text "No values have been added." displayed in the "Value" column.

Value	Returned Value
<i>No values have been added.</i>	

3. From the **Define Values** list, select **Import**.
4. Import the `color_code_lookup.csv` file:
  - a. Browse to the directory that contains the file. By default, the file is in the `Standardization_Rules_Designer_product_tutorial.zip` directory where you copied the tutorial files.
  - b. Select the file, and then click **Open**.
  - c. In the Import Definitions window, click **OK**.

The lookup table definitions are imported and the right pane shows the values in the lookup table that are defined.



The screenshot shows the same "Color\_Code" window, but now with an "Import..." button. The table below has two rows of data: "ASH" with a returned value of "900", and "BEIGE" with a returned value of "901".

Value	Returned Value
ASH	900
BEIGE	901

## Adding lookup table definitions

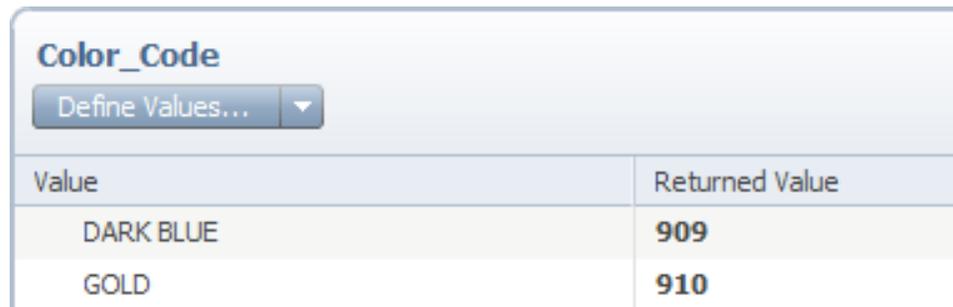
You can add lookup table definitions manually in the Standardization Rules Designer.

### Procedure

1. For the `Color_Code` table, select **Define Values** from the **Import** list.
2. Add a lookup table definition for the new value by completing the fields:
  - a. In the **Value** field, enter the product color GOLD.
  - b. In the **Returned Value** field, enter the color code 910.
  - c. Expand **Define Value As** and ensure that **Active definition** is selected. An inactive definition does not affect the value.

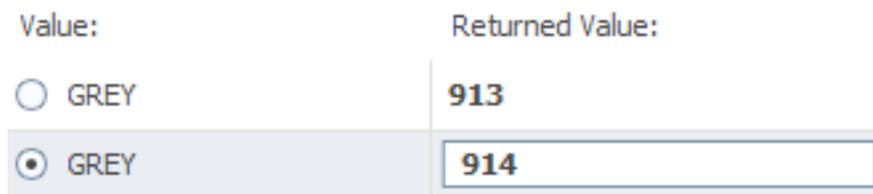
d. Click **OK**.

The right pane shows the value that you added a definition for.



Value	Returned Value
DARK BLUE	909
GOLD	910

3. Create a new lookup table definition for a value by repeating step 2 on page 23, but entering GREY for the value and 914 for the returned value.
4. Verify that the new lookup table definition is the active definition:
  - a. From the list of lookup tables, select the Color\_Code table. A list of values in the lookup table is shown in the right pane.
  - b. From the list of values, select GREY. A list of definitions for the value is shown on the Define Value page. The definition that you added is selected, and is therefore the active definition.



Value:	Returned Value:
<input type="radio"/> GREY	913
<input checked="" type="radio"/> GREY	914

### What's next

In this module, you completed the following tasks:

- Added a lookup table
- Imported lookup table definitions
- Added lookup table definitions

In the next module, you will create and modify rules that handle the retail product records for the fictional Sample Outdoor Company.

---

## Module 4: Adding and modifying rules

In this module, you will add and modify rules to ensure that records for new retail product data are standardized correctly.

The fictional Sample Outdoor Company wants to add or modify rules so that the records from their new product lines are handled properly. To do so, it will use the new custom class and lookup table that were created earlier in the tutorial.

First, the company looks at the rule that handles the most common pattern in the data. The records that match this pattern are handled relatively well, but some products require that the product name include the product brand name. The rule must be modified to ensure that the product names for some products include the product brand name.

Next, the company looks at the second most common pattern in the data, B+S+CT, which is not handled at all. The company must add a rule to handle this data. This rule will convert the product color to product color code by using the lookup table that was added to the Standardization Rules Designer.

When the company looks at the remaining unhandled patterns, the company finds a pattern in which unit of measure and amount are concatenated into one value. For example, the values 195 and CM are concatenated into 195CM. The company adds a rule for this pattern that splits the data into the appropriate output columns.

After rules are added and modified, the company decides update the actual rule set by publishing the revision.

### Learning objectives

After completing the lessons in this module you will understand the concepts and tasks for adding and modifying rules and publishing revisions:

- Modify a rule
- Identify unhandled patterns and view example records for those patterns
- Customize the output columns that are available in the Standardization Rules Designer
- Add a basic rule by mapping values in a record to the appropriate output columns
- Add actions that reference a lookup table
- Add a rule that splits a concatenated value into two distinct values
- Publish a revision

### Time required

This module takes approximately 60 minutes to complete.

### Prerequisites

- Ensure that you completed the setup steps and have the tutorial files and sample data loaded on your computer system.
- Complete Module 2, which includes creating the unit of measure custom class.
- Complete Module 3, which includes adding a lookup table that converts product colors to numeric color codes.

## Lesson 4.1: Modifying a rule to handle data correctly

In this lesson, you ensure that all records that match the most common pattern are handled correctly by modifying the rule that handles those records.

### Overview

The standardization goals of the fictional Sample Outdoor Company require that the name of some products include the product brand name. For the most common pattern in the data, the company added a rule that maps values in the input records to the correct output columns. However, the ProductName output column does not include the product brand name for any brands. The rule must be modified for some brands to add the brand name to the ProductName output column.

### Learn more about rules and rule groups:

Rules are processes that standardize groups of related records. Rules can apply to records that match the same pattern or to exact strings of text. When you add or modify a rule, you map values in the input records to output columns, specify actions that manipulate the data, and identify conditions to ensure that rules apply only to the correct records.

A rule group is a collection of rules that are applied to records at the same point in the standardization process. To ensure that rules are applied in a particular order, you can organize the rules into rule groups in the Standardization Rules Designer. Then, you can invoke the rule groups from the pattern-action specification (previously called the pattern-action file).

### Procedure

1. Click the **Rules** tab, select the Input\_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.

▼ Pattern Rule		<a href="#">Add Pattern...</a>	
Rule	Frequency	Unhandled	
▶  B+SCT	32.48%	0%	
▶  B+S+CT	8.39%	8.39%	
▶  B++SCT	4.74%	4.74%	
▶  B++UCT	4.01%	4.01%	
▶  ++SCT	3.28%	3.28%	

3. Expand **B+SCT**, and then expand **Copy product data to output columns**, which is the only rule for the pattern. The rule maps values in the input records to output columns.

Note that the frequency of the rule is equal to the frequency of the pattern.

▼ B+SCT	32.48%
▼ Copy product data to output columns	32.48%

This information indicates that the rule applies to all of the records that match this pattern.

4. If the example record on the Define Rule page is not ANTONI BELLA JUNIOR BLUE EYEWEAR, select ANTONI BELLA JUNIOR BLUE EYEWEAR from the list of example records. If ANTONI BELLA JUNIOR BLUE EYEWEAR is not in the list of example records, increase the number of records in the list.

- a. In the upper-right corner of the Browse Rules pane, click  .
- b. From the **Records to Display** list, select **100**.
- c. Click **OK**. The list of example records includes ANTONI BELLA JUNIOR BLUE EYEWEAR. You can select the record from the list.

5. From the example record, drag the value ANTONI, which is a product brand name, to the ProductName output column. To ensure that the product brand name is included before the product name in the output column, you might need to drag the BELLA product name after the product brand name.

The product brand name is in the ProductName output column and is separated from the rest of the product name by a space.

Example Record

ANTONI	BELLA	JUNIOR	BLUE	EYEWEAR
--------	-------	--------	------	---------

ANTONI	ANTONI BELLA		JUNIOR	BLUE
ProductBrand	ProductName	SizeUnit	SizeType	ProductColor
EYEWEAR				
ProductType	TypeCode	ColorCode	InputPattern	UnhandledPattern

6. In the ProductName output column, right-click the ANTONI value and click **Edit Action**. In the Edit Action window, you can manipulate the data that is sent to the output column.
7. In the **Look up the Object** section, click **Yes**.
8. Specify a list of the product brand names to add to the ProductName output column:
  - a. From the **Source** list, select **List**.
  - b. In the **List** table, enter the product brand names that must be added to the product name. The following brands include the product brand name in the product name:
    - EPOCH
    - FIREFLY
    - HAILSTORM
    - ANTONI

The table lists the required brand names.

List:

*Value	Returned Value	Similarity Threshold
EPOCH	EPOCH	900 (Exact) ▼
FIREFLY	FIREFLY	900 (Exact) ▼
HAILSTORM	HAILSTORM	900 (Exact) ▼
ANTONI	ANTONI	900 (Exact) ▼
Enter Text...		900 (Exact) ▼

- c. From the **If Found** list, select **Convert to Returned Value**. If an input record contains a product brand name that is in the list, the product brand name is added to the ProductName output column.
  - d. From the **If Not Found** list, select **Do Nothing (Stop Action)**. If you select this option, any product brand names that are not in the list will not be added to the ProductName output column.
  - e. Click **OK**.
9. Click **Apply**.

## Lesson 4.2: Identifying unhandled records and customizing output columns

In this lesson, you identify records that are not handled by rules in the rule set so that you can add rules for those records. You then hide output columns that you will not need when you add a rule for the unhandled records.

### Tasks in this lesson

1. "Identifying unhandled records"
2. "Customizing output columns" on page 29

### Identifying unhandled records

You can identify records that match a common pattern and are not handled by the rule set. When you add rules to the rule set, you can focus on these records.

#### Learn more about unhandled records:

Unhandled records are records that no rule in the rule set applies to. Unhandled records can be records from the sample data or records that are generated by the system.

When you use the Standardization Rules Designer, you manage your standardization processes most effectively when you add rules that affect a large number of records in your data. After you identify unhandled records, you can add rules that apply to those records.

#### Procedure

1. Click the **Rules** tab, select the Input\_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.

▼ Pattern Rule		<a href="#">Add Pattern...</a>
Rule	Frequency	Unhandled
▶  B+SCT	32.48%	0%
▶  B+S+CT	8.39%	8.39%
▶  B++SCT	4.74%	4.74%
▶  B++UCT	4.01%	4.01%
▶  ++SCT	3.28%	3.28%

- Expand **B+SCT**. Note that the frequency of the rule is equal to the frequency of the pattern.

▼  B+SCT	32.48%
▼  Copy product data to output columns	32.48%

This information indicates that the rule applies to all of the records that match this pattern.

- Expand **B+S+CT**, which is the pattern with the second-highest frequency. No rules exist for this pattern. All of the records are unhandled.

Because no rules have been created for this pattern, and the pattern has a high frequency, you can affect many records in your data by adding a rule for this pattern. In the next lesson, you will add a rule for this pattern.

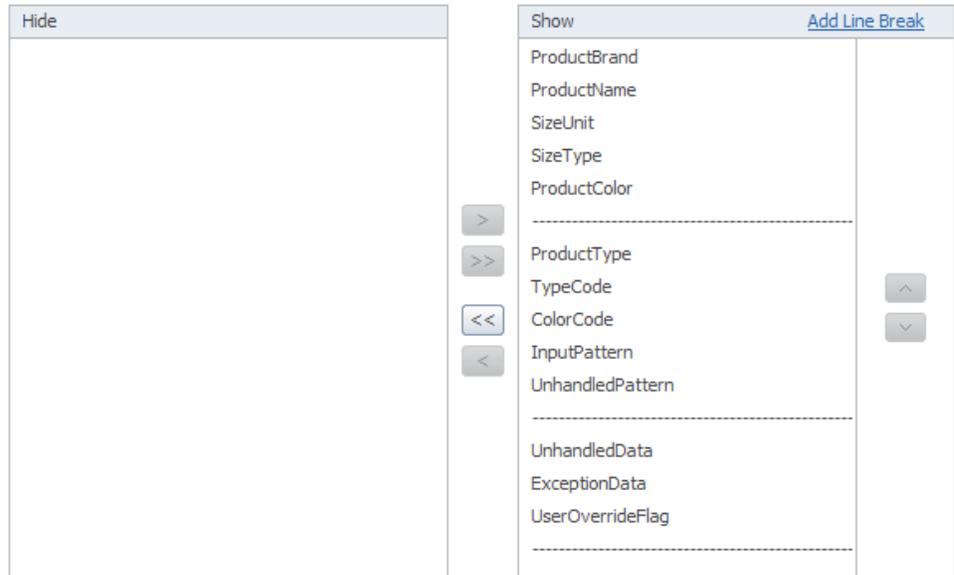
## Customizing output columns

You can hide output columns that are not required for a rule.

The rule that the fictional Samples Outdoor Company wants to add for the B+S+CT pattern does not require the TypeCode or InputPattern output columns. To simplify the pane where the rule is added, the company wants to remove the output columns that are not needed from the Define Rule page.

### Procedure

- From the **B+S+CT** pattern, click **Unhandled Records**. An example record and list of output columns is shown.
- On the Unhandled Records page, click **Customize Output Columns**. A list of the columns that are hidden and the columns that are displayed is shown.



3. Hide the output columns that are not required:
  - a. Select the TypeCode output column and click  to move the output column to the list of columns that are hidden.
  - b. Repeat step 3a for the InputPattern output column.
  - c. Click **OK**.

The Define Rule page does not show the output columns that you just hid.

### Lesson 4.3: Adding a rule for the pattern with the most unhandled records

In this lesson, you add a rule to handle the second most common pattern in the data.

#### Overview

The fictional Sample Outdoor Company wants to add a rule for a pattern that matches more than 8% of the records in the sample data. The rule will include an action that converts the product color to product color code by using the lookup table that was added to the Standardization Rules Designer in module 3.

#### Learn more about actions:

An action is a part of a rule that specifies how the rule processes a record. An action can specify what information is moved from the input data to the output columns and how that data is manipulated. In the Standardization Rules Designer, you can use an action to manipulate data in the following ways:

- Specify the part of the input data that is acted on. The part can be an exact value, standard value, one or more characters in the value, or a literal.
- Compare or convert the selected part of the input data to values in a lookup table or list.
- Add the output data to an output column and specify a leading separator between that data and other data in the column.

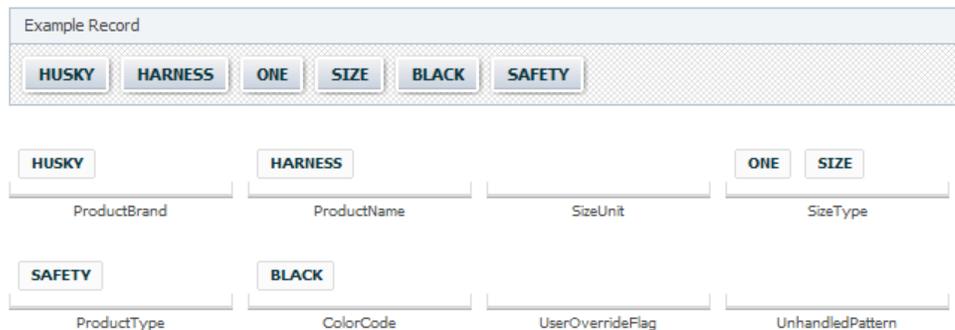
## Procedure

1. Click the **Rules** tab, select the Input\_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.
3. Expand **B+S+CT**, and then click **Unhandled Records**. An example record and list of output columns is shown.
4. If the example record on the Define Rule page is not HUSKY HARNESS ONE SIZE BLACK SAFETY, select HUSKY HARNESS ONE SIZE BLACK SAFETY from the list of example records. If HUSKY HARNESS ONE SIZE BLACK SAFETY is not in the list of example records, increase the number of records in the list.
  - a. In the upper-right corner of the Browse Rules pane, click .
  - b. From the **Records to Display** list, select 100.
  - c. Click **OK**. The list of example records includes HUSKY HARNESS ONE SIZE BLACK SAFETY. You can select the record from the list.
5. To map the values from the example record to the appropriate output columns, drag the values to the column. The following table shows the output column for each value.

Table 2. Output columns for each value in the example record

Value	Output column
HUSKY	ProductBrand
HARNESS	ProductName
ONE	SizeType
SIZE	SizeType
SAFETY	ProductType
BLACK	ColorCode

The values from the example records are mapped to the output columns.



6. In the ColorCode output column, right-click BLACK, then click **Edit Action**. In the Edit Action window, you can manipulate the data that is sent to the output column.
7. In the **Look up the Object** section, click **Yes**.
8. Use the Color\_Code lookup table to convert product colors to numeric color codes:
  - a. From the **Source** list, select **Lookup Table**.
  - b. From the **Lookup Table** list, select Color\_Code.
  - c. From the **If Found** list, select **Convert to Returned Value**.

- d. From the **If Not Found** list, select **Do Nothing (Stop Action)**. If you select this option, any product colors that are not in the lookup table will not be added to the ColorCode output column.
  - e. Click **OK**.
9. Apply the rule to the pattern:
    - a. Click **Apply as New Rule**.
    - b. In the **Rule Name** field, enter Convert color to color code.
    - c. Click **OK**.

The rule is added and is shown under the **B+S+CT** pattern.

▼  B+S+CT	8.39%
▶  Convert color to color code	8.39%

## Lesson 4.4: Adding a rule to split values into different output columns

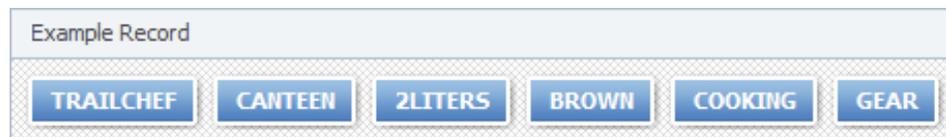
In this lesson, you add a rule that splits a value and maps each part of the split value to a different output column.

### Overview

Records in input data might contain information that is concatenated. For example, the input data might contain the value 195CM. This value concatenates 195, which represents an amount, and CM, which represents a unit of measure. To distinguish the amount from the unit of measure, you can add a rule that maps each part of the input value to a different output column.

### Procedure

1. Click the **Rules** tab, select the Input\_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.
3. Expand **B+>CTT**, and then click **Unhandled Records**. An example record is shown. One of the values in the record, which is represented in the pattern as >, includes leading digits followed by letters.



4. If the example record on the Define Rule page is not TRAILCHEF CANTEEN 2LITERS BROWN COOKING GEAR, select TRAILCHEF CANTEEN 2LITERS BROWN COOKING GEAR from the list of example records. If TRAILCHEF CANTEEN 2LITERS BROWN COOKING GEAR is not in the list of example records, increase the number of records in the list.
  - a. In the upper-right corner of the Browse Rules pane, click  .
  - b. From the **Records to Display** list, select **100**.
  - c. Click **OK**. The list of example records includes TRAILCHEF CANTEEN 2LITERS BROWN COOKING GEAR. You can select the record from the list.

- Map the values from the example record to the appropriate output columns by dragging the values to the column. The following table shows the output column for each value.

Table 3. Output columns for each value in the record

Value	Output column
TRAILCHEF	ProductBrand
CANTEEN	ProductName
2LITERS	SizeUnit
BROWN	ProductColor
COOKING	ProductType
GEAR	ProductType

The values from the example records are mapped to the output columns. However, the value in the SizeUnit output column contains a value that belongs in the SizeType output column.

- In the SizeUnit output column, right-click 2LITERS, and then click **Edit Action**. In the Action windows, you can manipulate the data that is sent to the output column.
- Edit the action to add only the leading digits to the SizeUnit output column:
  - From the **Object** list, select **Prefix**.
  - Click **All leading digits**. The **Example Result** field shows the value 2.
- Click **Specify action for remaining characters**, and then click **OK**. The **Actions** window opens. Options are selected that specify an action for the characters that you did not map to the SizeUnit output column.

- Map the remaining characters to the SizeType output column:
  - From the **Output Column** list, select **SizeType**.
  - Click **OK**.

The values are shown in the correct output columns.

Example Record  Pass through

TRAILCHEF CANTEEN 2LITERS BROWN COOKING GEAR

TRAILCHEF CANTEEN 2 LITERS BROWN

ProductBrand ProductName SizeUnit SizeType ProductColor

COOKING GEAR

ProductType TypeCode ColorCode InputPattern UnhandledPattern

10. Apply the rule to the pattern:
  - a. Click **Apply as New Rule**.
  - b. In the **Name** field, enter Split size values.
  - c. Click **OK**.

The rule is added and is shown under the **B+>CTT** pattern.

▼  B+>CTT	2.18%
▶  Split size values	2.18%

## Lesson 4.5: Publishing a revision

In this lesson, you publish a revision to the metadata repository so that the enhanced rule set can be used in a standardization job.

### Overview

After you make changes to a rule set in the Standardization Rules Designer, you can apply those changes to the rule set in the metadata repository by publishing the revision. After you publish the revision, the rule set must be provisioned in the Designer client before the rule set can be applied in a standardization job.

The fictional Sample Outdoor Company has updated the classifications, lookup tables, and rules in the version of the CAMPING.SET rule set that is stored on the Standardization Rules Designer database. To ensure that the version of the rule set that is used in standardization jobs reflects these changes, you publish the revision to the metadata repository.

### Procedure

1. In the Standardization Rules Designer, click the **Home** tab.
2. In the navigation pane, click **View and Publish Revision**.
3. Optional: In the Properties pane, enter information about the rule set.
  - a. In the **Notes** field, enter a description of the changes that you made to the rule set since the revision was published.
  - b. Click **Apply**.
4. Click **Publish**. In the window for the message about publishing changes, click **Yes**.

The rule set in the metadata repository is updated with the changes that you made in the Standardization Rules Designer.

### What's next

In this module, you completed the following tasks:

- Modified a rule

- Identified unhandled patterns and viewed example records for those patterns
- Managed the output columns that are available in the Standardization Rules Designer
- Added a basic rule by mapping values in a record to the appropriate output columns
- Added actions that reference a lookup table
- Added a rule that splits a concatenated value into two distinct values
- Published a revision

You have completed the tutorial for enhancing a product rule set in the IBM InfoSphere QualityStage Standardization Rules Designer.

You can now complete one or more of the following tasks:

- Complete the tutorial for enhancing an address rule set in the Standardization Rules Designer.
- Import your own sample data and use the Standardization Rules Designer to enhance a rule set that applies to the data
- Learn more about IBM InfoSphere QualityStage and the Standardization Rules Designer by reading the *IBM InfoSphere QualityStage User's Guide*.



---

## Appendix A. Product accessibility

You can get information about the accessibility status of IBM products.

The IBM InfoSphere Information Server product modules and user interfaces are not fully accessible.

For information about the accessibility status of IBM products, see the IBM product accessibility information at [http://www.ibm.com/able/product\\_accessibility/index.html](http://www.ibm.com/able/product_accessibility/index.html).

### Accessible documentation

Accessible documentation for InfoSphere Information Server products is provided in an information center. The information center presents the documentation in XHTML 1.0 format, which is viewable in most web browsers. Because the information center uses XHTML, you can set display preferences in your browser. This also allows you to use screen readers and other assistive technologies to access the documentation.

The documentation that is in the information center is also provided in PDF files, which are not fully accessible.

### IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.



---

## Appendix B. Contacting IBM

You can contact IBM for customer support, software services, product information, and general information. You also can provide feedback to IBM about products and documentation.

The following table lists resources for customer support, software services, training, and product and solutions information.

*Table 4. IBM resources*

<b>Resource</b>	<b>Description and location</b>
IBM Support Portal	You can customize support information by choosing the products and the topics that interest you at <a href="http://www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server">www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server</a>
Software services	You can find information about software, IT, and business consulting services, on the solutions site at <a href="http://www.ibm.com/businesssolutions/">www.ibm.com/businesssolutions/</a>
My IBM	You can manage links to IBM Web sites and information that meet your specific technical support needs by creating an account on the My IBM site at <a href="http://www.ibm.com/account/">www.ibm.com/account/</a>
Training and certification	You can learn about technical training and education services designed for individuals, companies, and public organizations to acquire, maintain, and optimize their IT skills at <a href="http://www.ibm.com/training">http://www.ibm.com/training</a>
IBM representatives	You can contact an IBM representative to learn about solutions at <a href="http://www.ibm.com/connect/ibm/us/en/">www.ibm.com/connect/ibm/us/en/</a>



---

## Appendix C. Accessing the product documentation

Documentation is provided in a variety of formats: in the online IBM Knowledge Center, in an optional locally installed information center, and as PDF books. You can access the online or locally installed help directly from the product client interfaces.

IBM Knowledge Center is the best place to find the most up-to-date information for InfoSphere Information Server. IBM Knowledge Center contains help for most of the product interfaces, as well as complete documentation for all the product modules in the suite. You can open IBM Knowledge Center from the installed product or from a web browser.

### Accessing IBM Knowledge Center

There are various ways to access the online documentation:

- Click the **Help** link in the upper right of the client interface.
- Press the F1 key. The F1 key typically opens the topic that describes the current context of the client interface.

**Note:** The F1 key does not work in web clients.

- Type the address in a web browser, for example, when you are not logged in to the product.

Enter the following address to access all versions of InfoSphere Information Server documentation:

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ/
```

If you want to access a particular topic, specify the version number with the product identifier, the documentation plug-in name, and the topic path in the URL. For example, the URL for the 11.3 version of this topic is as follows. (The ⇒ symbol indicates a line continuation):

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/⇒  
com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html
```

**Tip:**

The knowledge center has a short URL as well:

```
http://ibm.biz/knowctr
```

To specify a short URL to a specific product page, version, or topic, use a hash character (#) between the short URL and the product identifier. For example, the short URL to all the InfoSphere Information Server documentation is the following URL:

```
http://ibm.biz/knowctr#SSZJPZ/
```

And, the short URL to the topic above to create a slightly shorter URL is the following URL (The ⇒ symbol indicates a line continuation):

```
http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/⇒  
common/accessingiidoc.html
```

## Changing help links to refer to locally installed documentation

IBM Knowledge Center contains the most up-to-date version of the documentation. However, you can install a local version of the documentation as an information center and configure your help links to point to it. A local information center is useful if your enterprise does not provide access to the internet.

Use the installation instructions that come with the information center installation package to install it on the computer of your choice. After you install and start the information center, you can use the **iisAdmin** command on the services tier computer to change the documentation location that the product F1 and help links refer to. (The `⇒` symbol indicates a line continuation):

### Windows

```
IS_install_path\ASBServer\bin\iisAdmin.bat -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

### AIX® Linux

```
IS_install_path/ASBServer/bin/iisAdmin.sh -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

Where `<host>` is the name of the computer where the information center is installed and `<port>` is the port number for the information center. The default port number is 8888. For example, on a computer named `server1.example.com` that uses the default port, the URL value would be `http://server1.example.com:8888/help/topic/`.

## Obtaining PDF and hardcopy documentation

- The PDF file books are available online and can be accessed from this support document: <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1>.
- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss>.

---

## Notices and trademarks

This information was developed for products and services offered in the U.S.A. This material may be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

### Notices

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:**

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. \_enter the year or years\_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

## Privacy policy considerations

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session or persistent cookies. If a product or component is not listed, that product or component does not use cookies.

*Table 5. Use of cookies by InfoSphere Information Server products and components*

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
Any (part of InfoSphere Information Server installation)	InfoSphere Information Server web console	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	User name	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
Any (part of InfoSphere Information Server installation)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	No personally identifiable information	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Enhanced user usability</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled

Table 5. Use of cookies by InfoSphere Information Server products and components (continued)

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
InfoSphere DataStage	Big Data File stage	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	<ul style="list-style-type: none"> <li>• User name</li> <li>• Digital signature</li> <li>• Session ID</li> </ul>	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled
InfoSphere DataStage	XML stage	Session	Internal identifiers	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere DataStage	IBM InfoSphere DataStage and QualityStage Operations Console	Session	No personally identifiable information	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere Data Click	InfoSphere Information Server web console	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	User name	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere Data Quality Console		Session	No personally identifiable information	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled
InfoSphere QualityStage Standardization Rules Designer	InfoSphere Information Server web console	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	User name	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	<ul style="list-style-type: none"> <li>• User name</li> <li>• Internal identifiers</li> <li>• State of the tree</li> </ul>	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled
InfoSphere Information Analyzer	Data Rules stage in the InfoSphere DataStage and QualityStage Designer client	Session	Session ID	Session management	Cannot be disabled

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at <http://www.ibm.com/privacy> and IBM's Online Privacy Statement at <http://www.ibm.com/privacy/details> the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at <http://www.ibm.com/software/info/product-privacy>.

## Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com)<sup>®</sup> are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

The following terms are trademarks or registered trademarks of other companies:

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java<sup>™</sup> and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACS<sup>Link</sup>, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS<sup>Link</sup> licensee of the United States Postal Service.

Other company, product or service names may be trademarks or service marks of others.



---

# Index

## C

customer support  
    contacting 39

## L

legal notices 43

## P

product accessibility  
    accessibility 37  
product documentation  
    accessing 41

## R

rule set extensions  
    product data tutorial 1

## S

software services  
    contacting 39  
Standardization Rules Designer  
    product data tutorial 1  
support  
    customer 39

## T

trademarks  
    list of 43







Printed in USA

SC19-4333-00

