SAS® Data Loader 2.1 for Hadoop

User's Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2014. *SAS® Data Loader 2.1 for Hadoop: User's Guide*. Cary, NC: SAS Institute Inc.

SAS® Data Loader 2.1 for Hadoop: User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513–2414.

Printing 1, November 2014

SAS provides a complete selection of books and electronic products to help customers use SAS® software to its fullest potential. For more information about our products, visit **support.sas.com/bookstore** or call 1-800-727-3228.

Contents

1

Introducing SAS Data Loader

Hadoop for Everyone

The SAS Data Loader software enables you to transform, query, profile, and analyze big data in Hadoop, without moving that data. You can also apply the power of SAS Visual Analytics by copying Hadoop data to a separately licensed grid of SAS LASR Analytic Servers.

Now you can open the door to Hadoop regardless of your technical background. If you are a business analyst with little or no experience with Hadoop, you can use the wizard-based directives to merge, filter, and sort large distributed data sources. Programmers with experience in Hadoop and SAS also benefit from the simplicity of SAS Data Loader. Existing SAS DS2 programs and Hive SQL queries can be dropped into directives for repeat execution in Hadoop, with status monitoring in SAS Data Loader.

In addition to ease-of-use, SAS Data Loader also provides easy administration. The sftware installation and update processes are dramatically simplified by a new virtual machine architecture. The client software installs without interaction and without configuration files. When the web client informs you that an update is available, one click replaces the entire client, while retaining the existing client data.

Capabilities

The following wizard-based directives interact with your Hadoop cluster to deliver the following features:

Transform Data in Hadoop
: Select a source table, select one or more transformations, and select a target. Transformations include filter, manage columns, and summarize (aggregate) columns.

Query a Table in Hadoop
: Run aggregations on selected columns, filter source data, generate and edit a HiveQL query, or paste an existing HiveQL query.

Profile Data
: Select source columns from one or more tables to report uniqueness, incompleteness (null or blank), and patterns.

Saved Profile Reports
: List and open reports that were generated by the Profile Data directive.

Run Status
: List, control, and open all submitted directives. Stop and start directives, and open their log and generated code files.

Saved Directives
: List, open, and directives that were saved before they were run.

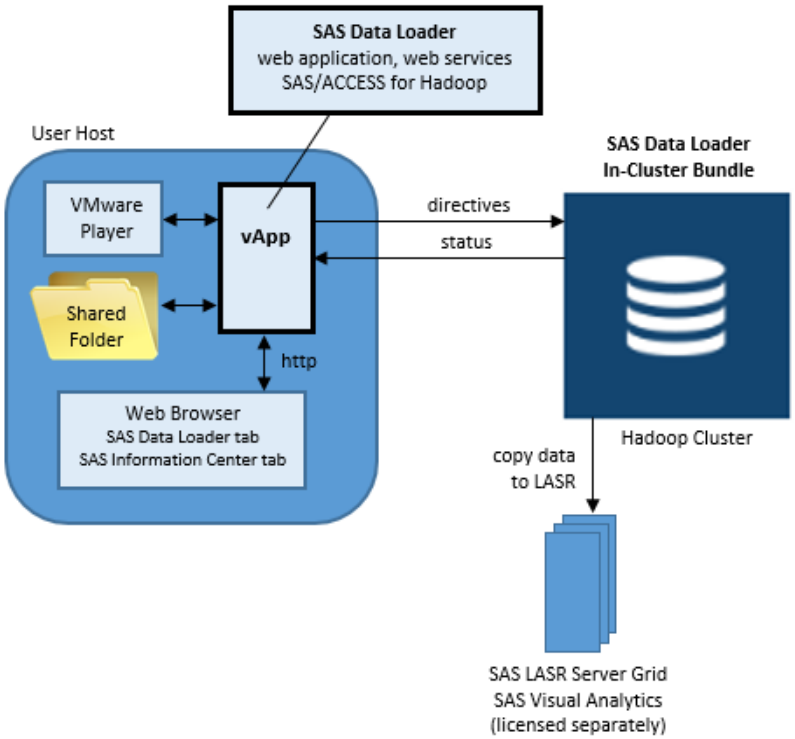
Load Data to LASR
: Loads specified Hadoop columns onto a grid of SAS LASR Analytic Servers for analysis using SAS Visual Analytics. SAS LASR Analytic Server and SAS Visual Analytics are separately licensed.

Run a SAS Program
: Paste or type a SAS DS2 program into the directive and then run that program in Hadoop. Supported SAS programs are those that utilize Data Step 2 languate elements, as defined in the *SAS DS2 Language Reference*.

Architecture

The following diagram illustrates the installed configuration of SAS Data Loader.

Figure 1.1 *SAS Data Loader Architecture*



On web clients, SAS Data Loader runs inside a virtual machine called a **vApp**. The vApp can be thought of as an enclosed entity with a single access point. The vApp runs in a separate and isolated Lynux operating environment. All client processing takes place inside the vApp. All of the files that are generated by SAS Data Loader are stored in the Shared Folder, which is outside of the vApp. This separation of the application from the data enables rapid and simple software updates.

Each instance of SAS Data Loader is accessed by a single user. The user works directly on the client computer on which SAS Data Loader was installed.

SAS Data Loader is displayed in a web browser, through a single HTTP address. the vApp accpets no other connections.

The vApp is started and stopped by the VMware Player Pro software, which is downloaded separately.

When the vApp is running, you enter into a web browser an HTTP address that is displayed by VMware Player Pro. The HTTP address opens the SAS Information Center. The SAS Information Center provides the button **Start SAS Data Loader**. Clicking the button opens SAS Data Loader in a new browser tab.

In addition to starting SAS Data Loader, the SAS Information Center also:

- specifies the version of Hadoop that is in use on the cluster
- announces the availability of software updates
- invokes the software update process
- starts and stops logging in the vApp

The SAS Data Loader web application works with the SAS In-Database Deployment Package for Hadoop to generate and execute directives in Hadoop. The client and the in-cluster bundle support the Cloudera or Hortonworks interfaces to the Hive database.

All of the client files that are generated and accessed by SAS Data Loader reside in a single Shared Folder. The contents of the Shared Folder, including saved directives and profile reports, remains unchanged by client software updates.

The directive Load Data to LASR is implemented when you define connections to an existing grid of SAS LASR Analytic Servers. The Load Data to LASR directive enables you to filter large tables out of Hadoop for further analysis using SAS Visual Analytics.

Note that the Load Data to LASR directive differs from Load to LASR capability that is provided in the SAS LASR Analytic Server software. Either method accomplishes the same task. SAS Data Loader uses a wizard-based approach.

Note: VMware Inc. provides the VMware Player Plus software for commercial applications and VMware Player, a free version, for non-commercial applications. See http://www.vmware.com/ to ensure that you download the version that is appropriate for your site. Both versions are fully supported by SAS Data Loader.

2

Get Started

Prerequisites

The installation of SAS Data Loader includes a client component and a Hadoop component. If you have not done so already, please install and configure both the client and Hadoop components before you proceed with this chapter. For complete instructions, refer to the *SAS Data Loader for Hadoop: Installation and Configuration Guide*.

Start the vApp and Open SAS Data Loader

Follow these steps to start the vApp and SAS Data Loader on your client computer.

1 Start VMware Player Plus by executing vmplayerplus.exe. The following path is typical for Windows hosts:

```
C:\Program Files (x86)\VMware\VMware Player Plus\x64\vmplayerplus.exe
```

2 In the VMware Player Plus window, click the current version of SAS Data Loader, and then click **Play Virtual Machine**.

3 In the window Data Loader – VMware Player Plus window, make note of the HTTP address that comprises the URL for SAS Data Loader.

4 Open a web browser and physically enter the SAS Data Loader URL into the address field. (You cannot cut and paste the URL.) Next, press the Enter key to open the SAS Data Loader Information Center in the web browser.

Note: To learn more about the features of the Information Center, see "About the SAS Information Center" on page 8.

CAUTION! The vApp is a single-user application. Concurrent use is not supported.

5 Under **Notifications** on the lower left, check to see whether a new version of SAS Data Loader is available. If the **Update** button is available, then see "Update SAS Data Loader" on page 58.

It is recommended that you always use the most recent version of the SAS Data Loader software.

6 Click **Start SAS Data Loader** to open the Directives page.

Close SAS Data Loader and Close the vApp

To close the vApp and close SAS Data Loader, simply close the VMware Player Plus window. Alternate methods include:

- Open the VMware Player Plus window and select **Suspend Guest** ▶ **Shut Down Guest**.
- Open the VMware Player Plus window and select **Player** ▶ **Power** ▶ **Shut Down Guest**

You can close the web browser tab for SAS Data Loader at any time without closing the web application or the vApp in which the web application runs. Closing the web browser tab for SAS Data Loader does not close the web application or the vApp. Any directives that are running will continue to run.

To reopen the web application in a new tab, select the SAS Information Center tab in the web browser and select **Start SAS Data Loader**. The status of any running directives will be updated.

You can check the status of running directives by opening the Run Status directive and clicking **Refresh**  . To learn more about directive status and stopping and restarting directives, see Chapter 4, “Run Status,” on page 27.

About the Directives Page

The Directives page enables you to work with Hadoop directives. For information about the tasks that are displayed on this page, see “Capabilities” on page 2.

In addition to the top-level tasks, you can also select the following menus and icons:

Help
Displays a link to the SAS Data Loader product documentation page on the SAS technical support website.

Configuration
Provides a short list of SAS Data Loader tasks that are used primarily to configure the directive Copy Data to LASR.

Sign Off
Closes SAS Data Loader, logs off of SAS, and displays the Logoff SAS web page.

About the SAS Information Center

The SAS Information Center is the launching point for SAS Data Loader. The web application also provides important notifications, configuration settings, and access to automated software updates for SAS Data Loader.

To open the SAS Information Center, you first need to to start the vApp on page 5, .

With the vApp running, you can open the SAS Information Center in the following ways:

- Click the browser tab for the SAS Information Center, if it is still available.
- Double-click the desktop icon for the SAS Information Center, if it is available in Windows.
- Open the VMware Player Plus window and enter the displayed URL in a browser tab.

To close the SAS Information Center, simply close the browser tab. You can reopen the SAS Information Center at any time.

To use the SAS Information Center, refer to theChapter 11, “Administration,” on page 56 chapter.

About the SAS Table Viewer

The SAS Table Viewer displays sample data and column information for a selected table. The viewer is available when you select source or target tables or when you view results or status. The viewer opens in a separate tab in the browser, so you can continue to reference the viewer while creating directives.

To open the viewer, click the icon **View a data sample**  .

In the viewer, you can click a column name to display the properties of that column. You can also click the checked box next to the column name to temporarily remove that

column from the sample data view. Click the empty box to restore the display of data for that column.

To change the number of sample rows that are displayed, change the value of the **Row Limit** field.

To refresh the sample data after a directive has operated on that table, click the **Refresh** icon  .

Column properties are defined as follows:

Index
: Column number.

Label
: A shortened version of the column name that can be added to the data values for that column. If a label is not assigned, then the column name is used as the label.

Length
: The size of the table cell (or variable value) in bytes.

Name
: Column name.

Type
: The type of the data in the column.

For information about data types and data conversions in SAS and Hadoop, see the chapter SAS/ACCESS Interface to Hadoop in the document *SAS/ACCESS Interface to Relational Databases: Reference*.

3

Transform Data in Hadoop

About the Transform Data in Hadoop Directive

The Transform Data in Hadoop directive enables the following transformations:

Filter data
: Reduce source table rows by applying logical operators to selected columns.

Manage columns
: Reduce source table columns by column name.

Summarize data
: Create a new row in the target table that contains summarized numeric data from a selected column.

A single directive can run multiple transformations. For example, a single directive can reduce the number of columns, filter the data in one or more columns, and then summarize data in one or more columns.

Directives can define a new target table, or use an existing table as the target. When a directive uses an existing target table, any existing data is dropped before the table receives new data.

Prerequisites

The directive Transform Data in Hadoop requires that you install and configure the SAS In-Database Deployment Package for Hadoop on your Hadoop cluster. The installation process is described in the

If you have not installed the SAS In-Database Deployment Package for Hadoop, or if service is interrupted during the execution of a transformation, the directive will fail and no code will be generated.

Create and Run a Transformation in Hadoop

The following example depicts the process of creating and running a directive that contains several transformations. The example opens a source table of customer information, selects columns for the target, and applies two filters.

1. Open SAS Data Loader by selecting the bookmark or favorite in your web browser. If you have not stored the web address in your browser, then see “Start the vApp and Open SAS Data Loader” on page 5.

2. In the SAS Data Loader page, click **Transform Data in Hadoop**.

3 In the Transform Data in Hadoop page, click the schema icon for **devSchema1** (for example). The tables in the schema are displayed.

4 Click the source table **customer1**, and then click **Next**.

> **TIP** To view sample contents of a source table, click the table, and then click **View a data sample** on the right side of the toolbar.

5 Click **Manage Columns**. An icon in the toolbar indicates the incomplete status of the transformation.

> **TIP** To change your selected schema, click **devSchema1** in the toolbar. To change your source table, click **customer1** in the toolbar.

6 To reduce the number of columns in the target, first click the double arrow to move all columns from Selected to Available. Then, to select a few columns for the target table, press **Ctrl** and click the columns **customerNumber**, **customerType**, and **custLastContactDateTime**. Click the single arrow to move the selected columns from **Available** to **Selected**.

Click **Add Another Transformation** to define a filter for your selected columns.

7 Click **Filter Data**.

8 Left-click or select from the drop-down menu **Select a Column**, and then click **customerType**. An icon indicates the type of the data in the selected column.

9 Right-click or select from the drop-down menu **Equal to**, and then click the logical operator **Contains**.

10 To enter a comparison value for the filter, click the text field to the right of **Equal to**, and then type `middleIncome`.

11 Click **Add Rule** to define a second filter.

12 Select the column **custLastContactString**, select the logical operator **After**, and then select the calendar date July 31, 2013. Click **Next** to configure a target table.

13 Click the schema **devSchema1**.

14 Click **New Table** to display the New Table window. Enter the name customerMidIncomeAfterJuly2013 for the new target table, and then click **OK** and **Next**.

15 Click **Start Transforming Data** to run your directive.

> **TIP** Click in the toolbar to change any aspect of your directive.

16 At the conclusion of the transformation, you can select **View Results**, **Error Details**, **Log**, or **Code**.

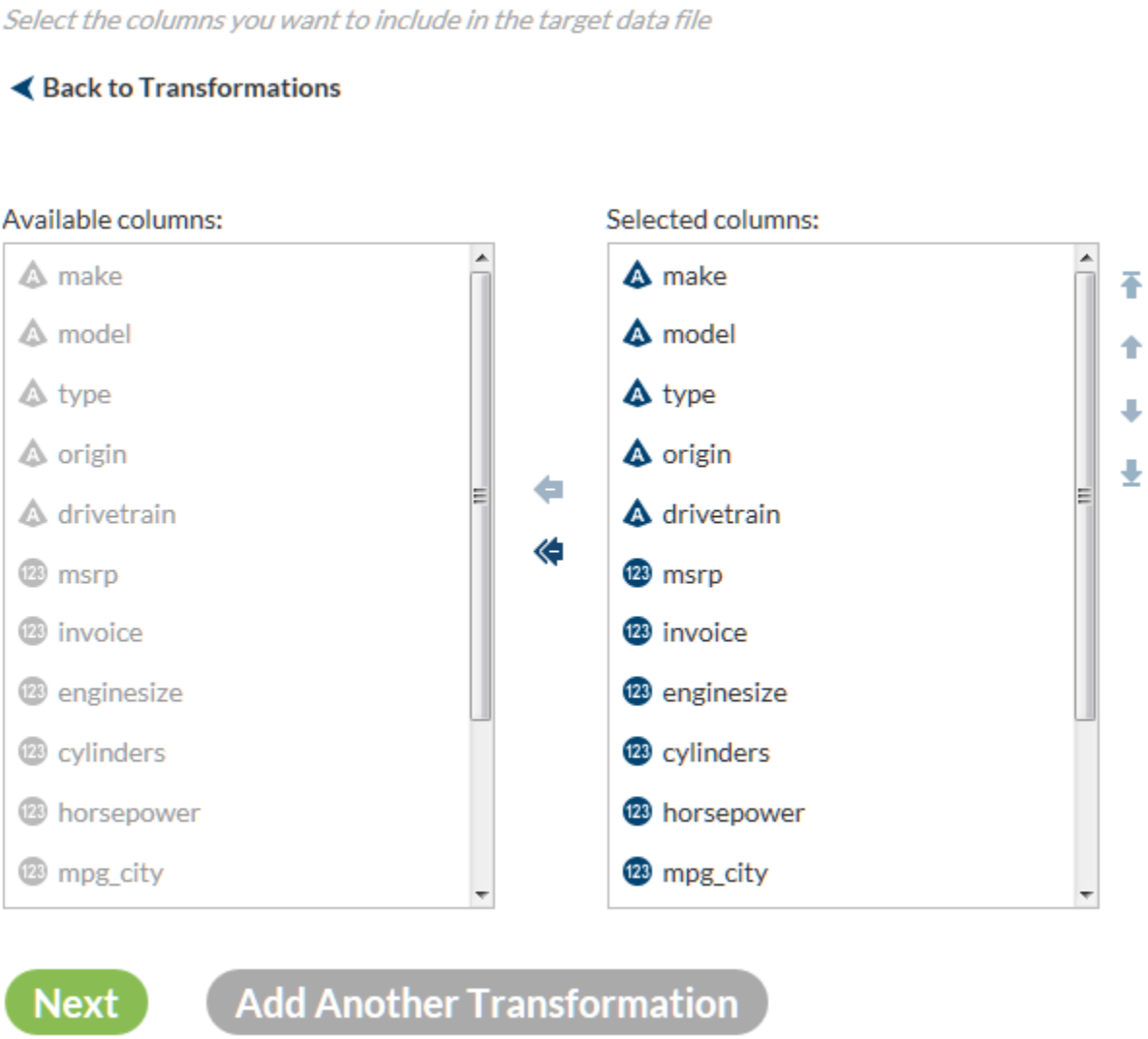
17 To save the directive, click **Save**.

About the Manage Columns Transformation

The Manage Columns transformation reduces the size of the target table by excluding columns that are irrelevant to your analysis. You simply choose the columns that you want in the target, and then order the columns from left-to-right as needed.

Note: To maximize the efficiency of your Filter transformation, use Manage Columns before you filter.

The Manage Columns page displays an Available columns box and a Selected columns box. Both boxes contain column names and data type icons.



The **Available columns** box provides a left-to-right ordered list of all of the columns in the source table. Blue columns can be selected for addition to the target table. Gray columns have already been selected for the target table.

The **Selected columns** box provides a left-to-right ordered list of all of the columns that will appear in the target table.

Click icons to select columns as follows:

Move all columns left (to Available) or right (to Selected)

Move selected columns left (to Selected) or right (to Available)

Use Shift+click to select a block of columns. Use Ctrl+Shift+Click to select separate columns.

Move one or more target columns left or right 1 column position

Move one or more target columns to the first column positions (leftmost)

Move one or more target columns to the last column positions (rightmost)

About the Filter Data Transformation

Reduce, Reduce, Reduce

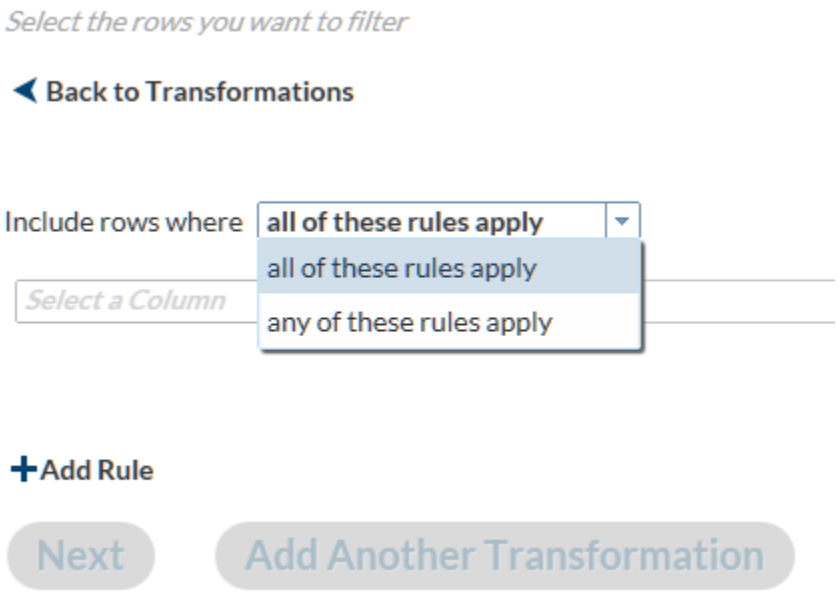
When you filter a table, you reduce the number of rows in the resulting target table. This much smaller target table contains only those roles that are relevant to your analysis. Excluding extraneous data improves the accuracy and efficiency of your analysis.

To filter a table, you apply rules to columns. Each rule applies a logical comparison to each value in that column. If the rule evaluates to TRUE, the associated source row is written into the target. You can apply a series of filters to produce the optimal target table for analysis.

For maximum efficiency, exclude irrelevant columns using the Manage Columns transformation, before you run your Filter transformation.

Specifying Multiple Rules

To specify multiple filters, first select a value for **Include rows where**.



Choose **all of these rules apply** if you want source rows to be written to the target only when they evaluate to TRUE for all of your rules. Otherwise, choose **any of these rules apply**, whereby one TRUE rule writes the row to the target.

Note: When you apply multiple rules to a single column, each rule runs against the original source column, not the resulting column from the previous transformation.

About the Logical Operators

The Filter Data Transformation plugs each row value in a selected column into a logical statement. If the statement is true, then that row is written into the target table. The logical statement takes the following form:

source-value operator comparator

Where:

source-value
is a row value from the selected column in the source table

operator
is a logical operator such as Equal To or Contains

comparator
is a statement that provides the basis for the logical comparison. The data type of the comparator must match the data type of the source value. The statement can be an expression that resolves to a value of the appropriate data type.

The data type of the source column can be character, numeric, or datetime. Each of these data types has unique logical operators. Some logical operators are common to all data types.

The following table describes the logical operators for each data type.

Table 3.1 *Logical Operators in the Filter Transformation*

Operator	Source ColumnData Types	Description and Example

Equal To	The Equal To operator is available for use with all source data types, which include the following:Character Numeric Datetime	The source value is accepted and its row is written to the target table only when the source value exactly matches the comparator.
		Character values can be case-sensitive. Blank spaces are included in the comparison.
		Datetime values in the comparator use the SAS format DATETIME(w.p).
		`Gender Equal To Male`
		`PrefCustomer Equal To 1`
		`SaleDate Equal To 5/1/2014`

Operator	Source Column Data Types			Description and Example

Not Equal To	A	123		Accepts the source row when the column value is anything other than the comparator.
				`Region Not Equal To EuropeNumChildren Not Equal To 0`
				`SaleDate Not Equal To 11/25/2013`

Null	A	123		Accepts the source row when the column value is NULL or if no source value is present. `CreditScore NullAnnualIncome Null`
Not Null	A	123		Accepts the source row when the column value is present and when the value is not NULL. `PostalCode Not NullPhoneNumber Not Null`
In	A	123		Accepts the source row when the column value is included in its entirety within the comparator. The comparator consists of a list of constant values. The list consists of a vertical list of individual entries, without commas. Blank spaces are interpreted literally. Case sensitivity can be enabled. `CarManuf In BMW VW BenzWaistSize In 32 34 36 38`

Operator	Source ColumnData Types	Description and Example
Not In	 	Accepts the source row when the column value is not included anywhere within the comparator's list of constant values. `City Not In New York Chicago Los AngelesWaistSize Not In 32 34 36 38`
Like	 	Accepts the source row when the column value matches the result of an expression in the comparator. The source value and the comparator are compared on a character-by-character basis. For character columns, case sensitivity can be enabled. Use the pattern-matching character `%` to indicate any string of characters. Use the underscore character _ to indicate any single character in that position. Note that trailing blank characters are written to the target table when using `%` at the end of the comparator. Use the word `escape` to include literal instances of `%` and _ in the comparator. `SalesRegion Like NorthAmer%AnnualSales Like 199_CustSatisfaction Like 100 escape %`

Operator	Source ColumnData Types		Description and Example
Not Like			Accepts the source row when the column value does not match the result of an expression in the comparator. The source value and the comparator are compared on a character-by-character basis. For character columns, case sensitivity can be enabled. Pattern-matching characters `%` and `_` and `escape` are valid as described for the Like operator. `Sports Not Like %ballFootballFieldLength Not Like 100%`
Contains			Accepts the source row when the column value is found within the character string of the comparator. Case sensitivity can be enabled. `Address Contains ILLicenseNumber Contains 7227`
Not Contains			Accepts the source row when the column value is not found within the character string of the comparator. Case sensitivity can be enabled. `Month Not Contains OctNovDecSalesMonthly Not Contains 0`
Between			Accepts the source row when the column value or date is between the two values or dates in the comparator, but is not equal to either. `GradeAverage Between 87.5 93DailySales Between December 20, 2014 December 27, 2014`

Operator	Source ColumnData Types	Description and Example
Greater Than	123	Accepts the source row when the column value is greater than the value of the comparator. `AnnualSales GreaterThan 100000`
Greater Than Or Equal To	123	Accepts the source row when the column value is equal to the comparator or greater than the comparator. `CarsInFamily Greater Than or Equal To 3`
Less Than	123	Accepts the source row when the column value is less than the value of the comparator. `GamerAge Less Than 30`
Less Than Or Equal To	123	Accepts the source row when the column value is equal to the value of the comparator, or less than the value of the comparator. `SalesYear Less Than Or Equal To 2010`
After		Accepts the source row when the column date is later than the date in the comparator. `HomePurchaseDate After January 1, 2013`
Before		Accepts the source row when the column date is earlier than the date in the comparator. `BirthDate Before March 17, 1980`

Operator	Source ColumnData Types	Description and Example
On Or After		Accepts the source row when the column date is later than, or the same date of, the date in the comparator. `DailySales On Or After January 1, 2014`
On Or Before		Accepts the source row when the column date is earlier than, or the same date of, the date in the comparator. `DailySales On Or Before December 31, 2013`

About the Summarize Rows Transformation

Use the Summarize Rows Transformation to groups rows and generate summaries for numeric columns. In the target table, new columns store summary values.

Take, for example, a source table of automobile data. You could group target rows by make and type. For the source columns RetailPrice and InvoicePrice, you could generate Max, Min, and Mean values for each distinct group.

The summary types are also known as aggregations. The available summaries are defined as follows:

Count
: the number of rows in the group that contain valid values.

Count Distinct
: the number of unique values in the column for each group

Corrected Sum of Squares
: measures variability or dispersion around the mean. To learn more about this (and other) statistical summaries, see the *Introduction to Statistical Modeling with SAS/STAT Software*.

Covariance
: measures the strength of the correlation of the values in the group. A positive value indicates that values move in the same direction within the group. A negative value indicates that values move in opposite or random directions.

Max
: the maximum value in the column for each group

Mean
: the calculated center value between the maximum and minimum values in the group

Min
: the minimum value in the group

Number of Missing Values
: the number of rows in the group that contain a blank or NULL value

Range
: the difference between the lowest and highest values in the group.

Standard Deviation
: measures the degree of variance, or the degree in which the values in the group deviate from the mean. A small value indicates little deviation. The standard deviation is the square root of the Variance.

Standard Error
: measures the applicability or accuracy of the mean as it applies to the values in the group. A small value indicates that the mean is a more accurate reflection of the values in the group.

Sum
: adds the values in the group

Variance

The average of the squared differences from the mean, which measure diversity in the group

4

Run Status

About the Run Status Directive

The Run Status directive provides a record of current and previous directive runs. Each run is recorded by name, status, start time, end time, and run time. An action menu for each directive enables you stop or start the directive, display the run log, and display the SAS code that was generated for the run.

For Profile Data directives that complete successfully, you can display the profile report (unless the report has been deleted in the Saved Profiles directive.)

For completed transformations or queries, you can view a sample of the target table.

By default, the Run Status page displays all of the directives that ran in the past 30 days. The most recent runs appear at the top of the list.

When you click the **Refresh** icon in the web page, you receive updates for all running directives, including any that were started or completed after you opened Run Status.

If any running directive generates an error message when Run Status is displayed in the browser, then an error message is displayed in the Error Details dialog box.

Examine and Start a Previous Directive

Follow this example to learn the capabilities of the Run Status directive.

1. Open see “Start the vApp and Open SAS Data Loader” on page 5.
2. Click **Run Status**. The Run Status page displays a list of directives that are running or have run. The **Show** pull-down enables you to modify the display to see the job runs that you are interested in.
3. If no directives are listed, then you can create and run a test job. See “Create and Run a Transformation in Hadoop” on page 12.
4. Reports are identified by the given name or by the generic name of the directive (such as Transform Data in Hadoop.) Given names area applied when you save a directive.
5. The Status column value can be In Progress, Stopped, Failed, or Successful.
6. The Action menu appears at the right side of each directive. You can select **Stop** for running directives and **Start** for directives that are not running.

   Note: If you select Stop, your directive can retain its In Progress status. In this situation, the directive is stopping, but it has yet to reach a suitable stopping point. Click the Refresh icon periodically as needed until the status changes to Stopped, or reopen Run Status later to confirm the Stopped status.
7. For a transformation or query with a Successful status, you can select View Results from the Action menu. In response, SAS Data Loader opens a new browser tab for

the SAS Table Viewer. The viewer displays a sample of data in the target table and provides information about the columns in the table. For further information about the SAS Table Viewer, see “About the SAS Table Viewer” on page 8.

Clearing Run Status

To clear all of the reports from the Run Status page, click the Clear All icon at the top of the page. Clearing reports permanently removes the reports from the vApp database.

About Unsaved Directives

If you run a directive without saving it, the directive is displayed in Run Status like any other directive. When processing stops on the unsaved directive, you can select Open from the action menu. You can then edit and save the unsaved directive.

About Incomplete Directives

An incomplete directive is one that you have stopped using the Action menu , or one whose status is Failed. Incomplete directives do not update their target tables. Nor do they include log information or code from the Hadoop database (Hive.)

5

Save and Manage Directives

Reuse Saved Directives

Use Saved Directives to open and manage your existing directives. After you open a saved directive, you can modify, save, save as, and execute the directive. You can also duplicate saved directives without opening and clicking the Save As icon.

Example: Open and Manage Saved Directives

Follow these steps to learn how to open, modify, execute, and resave a directive.

1 OpenSAS Data Loader on page 5.

2 Click **Saved Directives**.

3 In the Saved Directives page, click a directive to edit and execute the directive.

4 To manage saved directives, click the Manage Directives icon  .

5 In the Manage Save Directives window, select a directive and then click an icon. You can open, duplicate, delete, refresh, or rename the selected directive. Click the Refresh icon to update the name and last-modified date of the selected directive. You can refresh all of the saved directives by clicking the Refresh icon when you have no directives selected.

6

Query a Table in Hadoop

About the Query a Table in Hadoop Directive

Use this directive to group, summarize, and aggregate selected columns, filter source rows, generate a HiveQL query, edit the query as needed, and execute the directive in Hadoop. Aggregation data is stored in new columns in the target.

If you have an existing HiveSQL query, you can paste it directly into the directive, and add filters or aggregations as needed.

Prerequisites

The directive Query a Table in Hadoop requires that you install the SAS In-Database Deployment Package for Hadoop on your Hadoop cluster. The installation process is described in the SAS Data Loader for Hadoop: Installation and Configuration Guide.

Query a Table

Follow these steps to query a table in Hadoop.

1. Open SAS Data Loader by selecting the bookmark or favorite in your web browser. If you have not stored the web address in your browser, then see “Start the vApp and Open SAS Data Loader” on page 5.
2. In the Directives page, click **Query a Table in Hadoop**.
3. In the Data Sources page, click the schema that contains the table that you want to query.
4. In the Source Data page, click the table that you want to query, and then click **Next**.
5. In the Summarize Rows page, click the **Group rows by** field, and then click the column that you want to use as the primary grouping in your target table. For example, if you are querying a table of product sales data, then you could group by product type.

   Note: If you intend to paste a HiveSQL query into this directive, then you can click **Next** to bypass the pages for summaries and filters to reach the Code page.
6. To generate multiple aggregations, you can add additional groups. The additional groups will appear in the target table as nested subgroups. Each group that you define will receive its own aggregations.

   To add a group, click **Add Column**, and then repeat the previous step to select a different column than the first group. In a table of product sales data, you could choose a second group by selecting the column product_code.
7. In **Summarize columns**, select the first numeric column that you want to aggregate.
8. In **Aggregations**, select one of the following:

Count
: Specifies the number of rows that contain values in each group.

Count Distinct
: Specifies the number of rows that contain distinct (or unique) values in each group.

Max
: Specifies the largest value in each group.

Min
: Specifies the smallest value in each group.

Sum
: Specifies the total of the values in each group.

9 In **New column name**, either accept the default name or specify a new name for the target column that will contain the values of the aggregation.

10 To specify additional aggregations, click **Add Column** and either accept or change the default values.

11 When the aggregations are complete, click **Next**.

12 In the Filter Rows page, either click **Next** to apply all source rows to the target, or select **Specify rows**. Filtering source rows improves performance and focuses the results of the directive.

13 To filter source rows, first specify how you want to apply your filter rules. In the field **Include rows where**, select all of these rules apply to specify that all rules must be true to apply the row to the target. Select **any of these rules apply** to apply a source row to the target when one or more rules are true.

14 To create the first rule, click **Select a column** and choose the column to which you will apply your rule.

15 To specify a logical operator for your rule, click the middle field and select from the list. The logical operators that are available for your rules depend on the data type of

your column. Columns can be numeric, character, or datetime. To learn about the available logical operators, see Table 3.1 on page 18.

16 To add another rule, click **Add Rule**. When your filter rules are complete, click **Next**.

17 In the Sort page, specify how rows will be ordered top-to-bottom in the target table. Choose a column and either ascending or descending values in that column. Click **Add Column** to specify additional sort passes for rows that share the same value of the preceding sort column. Click **Next** when your row sort order is complete.

18 In the Target Table page, click the schema or data source that is to receive the target table.

19 In the target table list, you can click a target table, or create a new table or to save the query as a view.

The names of new tables must meet the naming conventions of SAS and Hadoop.

When you save a query as a view, the results of the directive are displayed in the SAS Table Viewer on page 8 without saving the results to a target table on disk.

When your target selection is complete, click **Next**.

20 In the Code page, SAS Data Loader builds a HiveQL query. To edit the query, click **Edit HiveQL Code**. As you edit, you can click **Reset Code** to restore the original HiveQL query that was generated by SAS Data Loader. Click **Next** when your query is complete.

Note: Edit your HiveQL code with care. The code in the editor is the exact code that will be executed by the directive, regardless of previous selections.

21 Review your directive and click the pages of the directive as needed to make corrections.

22 Click **Start querying data** to execute your directive. To monitor the progress of your directive, see the Run Status directive.

7

Profile Data

Overview

The Profile Data directive enables you to generate a profile report of the data in a table. You can choose from all the columns in a table to create a custom profile of only the data that you need to view.

Create a Profile

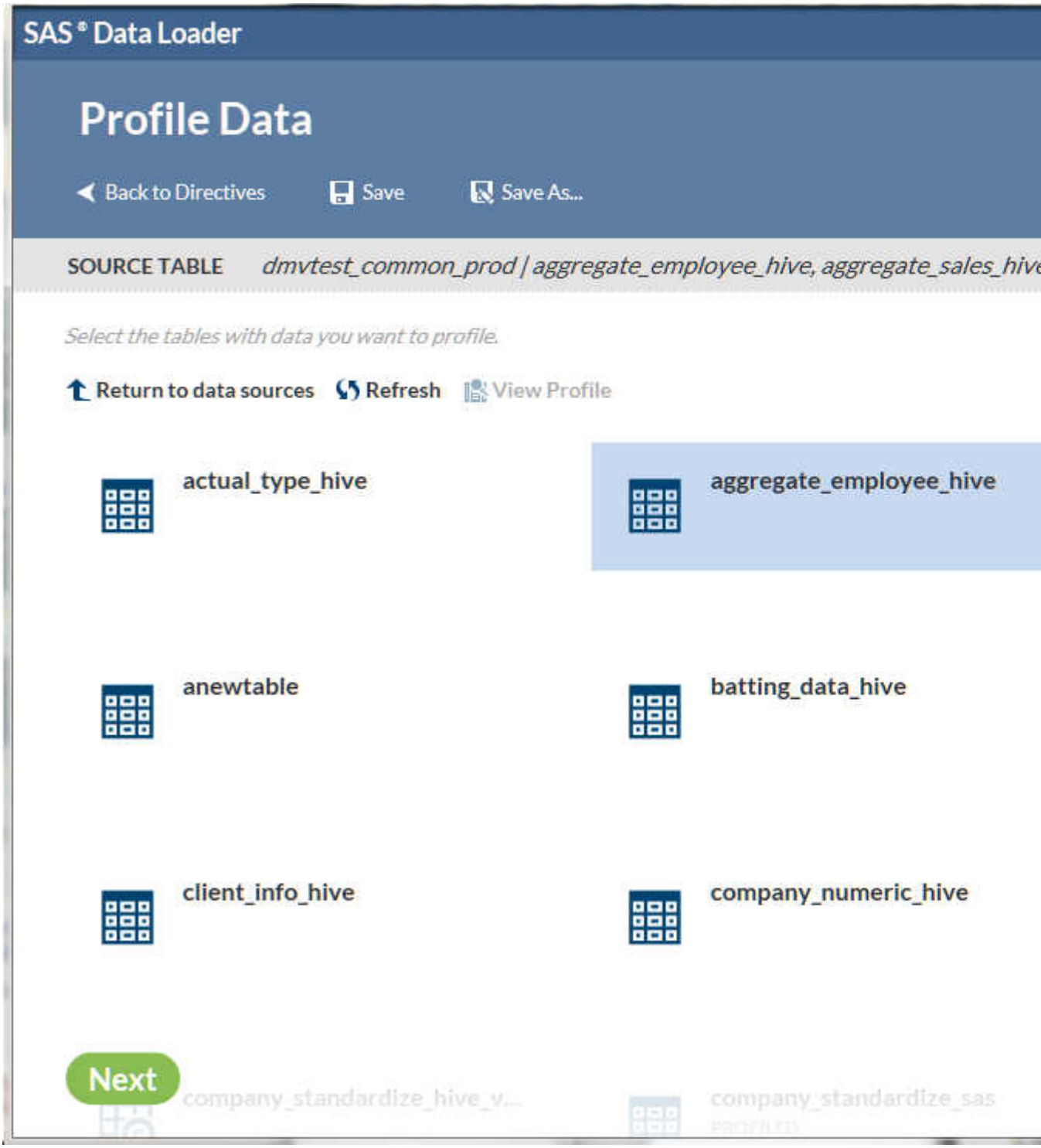
The following example demonstrates the process of creating a profile:

1 In the top-level Directives page, click the **Profile Data** directive.

2

3 In the Profile Data directive, click a data source to display its tables, click the table or tables for the profile report, and then click **Next**.

Figure 7.1 *Selected Data Source and Tables*

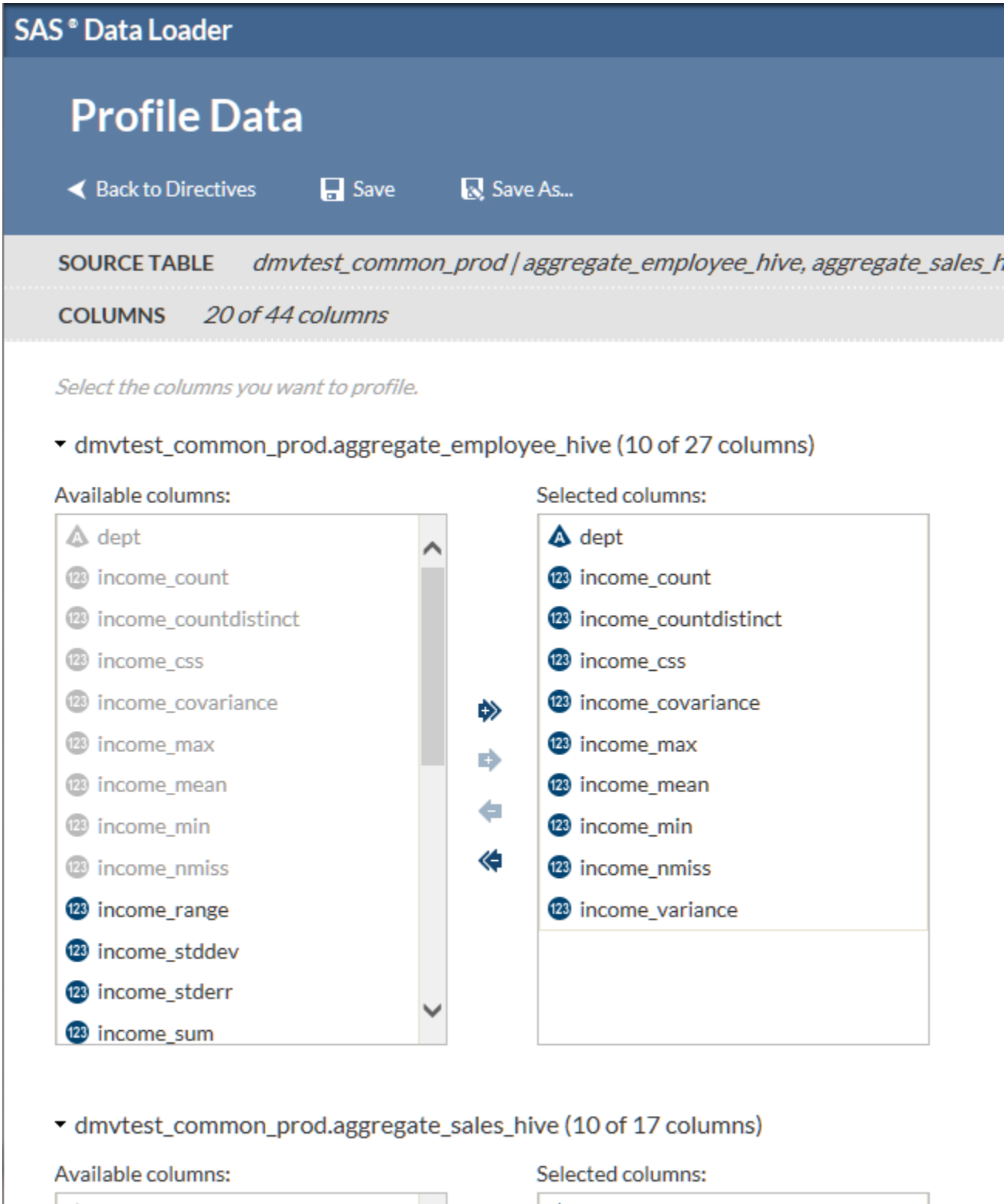


TIP To view sample data in a table, click the table and click the Table Viewer. For more information about the SAS Table Viewer, see “About the SAS Table Viewer” on page 8.

4 The Columns page displays the total number of columns that will be processed in the profile report. If you selected more than one table for your report, the tables are listed by name. Click one of the tables to display the columns that will be included in the profile report.

5 In the **Selected Columns** list box, click columns and click the single left arrow icon to remove columns from the profile report. The removed columns appear in the **Available Columns** list box.

Figure 7.2 Select Columns from Source Tables for the Profile Report



When the column selection is complete, click **Next**.

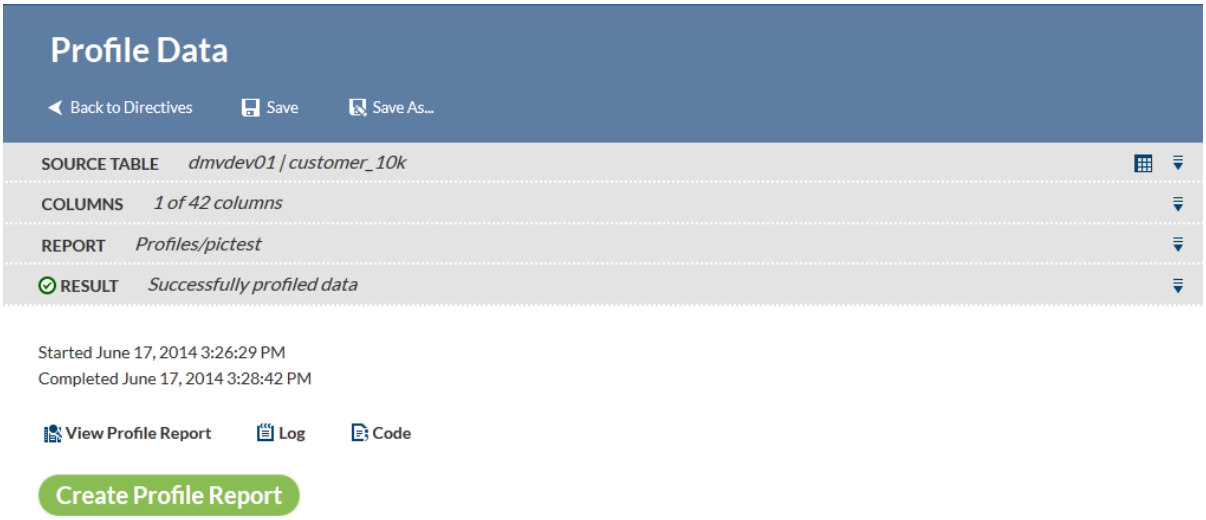
6 In the Report page, click **Report name** and enter a name for the profile report. Click the folder icon at the **Report location** field to icon to change the storage location of the profile report.

Figure 7.3 *Profile Location*

7 After specifying a name and location, click **Next**, and then click **Create Profile Report**.

8 After successfully creating the profile report, a screen similar to the following is displayed:

Figure 7.4 Profile Data



The following actions are available:

View Profile Report
enables you to view the Profile Report. See Chapter 8, “Saved Profile Reports,” on page 43 for more information about the profile report.

Log
displays the SAS log that is generated during the creation of the profile.

Code
displays the SAS code that generates the profile.

8

Saved Profile Reports

Overview

The Saved Profile Reports directive enables you to view the results of previously executed data profile directives and to create notes about the results. The profiles are created with the Profile Data directive. The profile reports and notes are stored as xml documents on the file system. Saved Profile Reports displays these xml files in a readable format.

View a Saved Profile

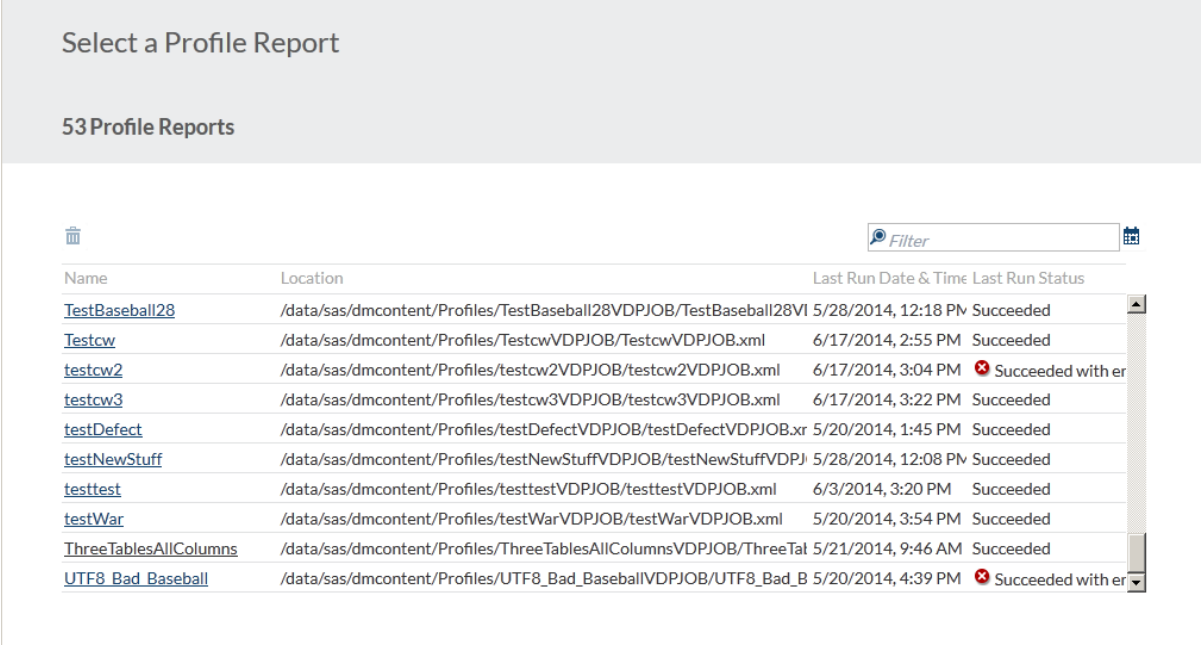
Using the Interface

Opening Saved Profile Reports

The following example demonstrates the process of opening a saved profile report:

1 On the Directives page, click the **Saved Profile Reports** directive to display a new browser tab for SAS Data Loader – Profile Reports:

Figure 8.1 *Profile Report List*



2 You can filter the list of reports using the following methods:

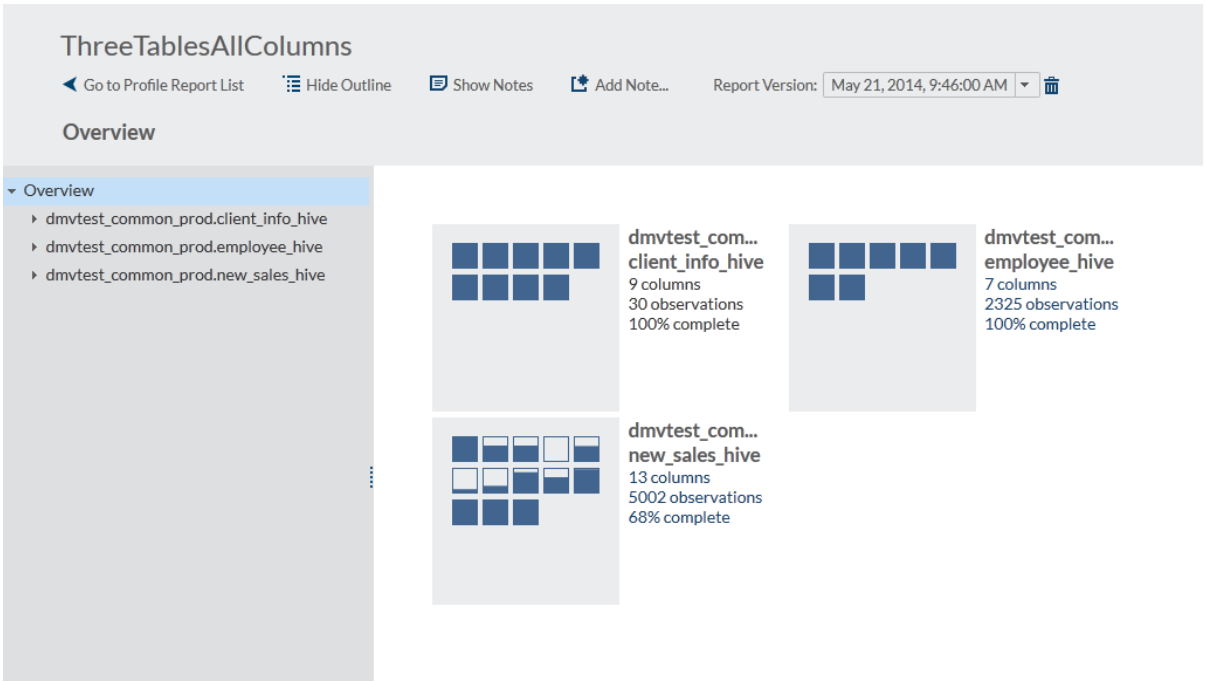
- Click **Filter list by date** and select a date. This filter displays profile reports that were generated on or after the selected date.
- Typing a text string into the search field.

3 Click **Close Filter** to remove the filter and restore the full list.

4 To delete profile reports, select one or more reports and click **Delete** .

If you open a profile report that contains multiple tables, click **Show Outline** , which displays

Figure 8.2 *Table Overview*

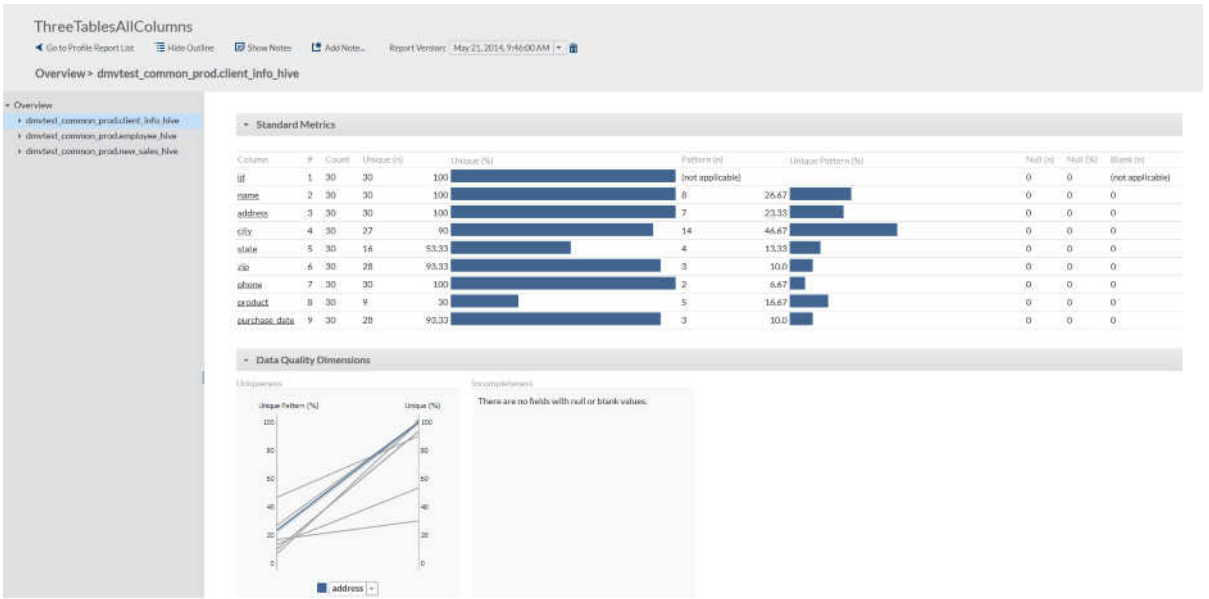


Note: If a profile contains a single table, opening the profile takes you directly to the details view described in Step 6.

5 To open a profile report, click its name. The report opens in a new browser tab.

6 Select a table beneath **Overview** or click directly on the table icon to display detailed table information in the right pane:

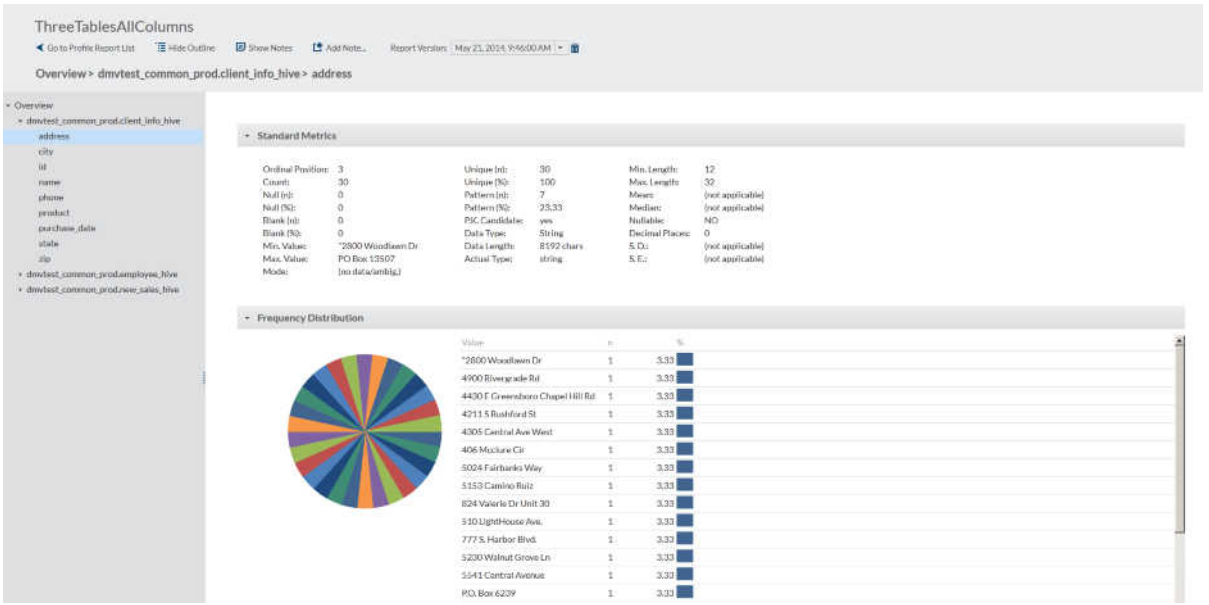
Figure 8.3 *Table Display*



7 Click ▸ to expand the table and display columns.

8 Select a column to display detailed column information in the right pane:

Figure 8.4 *Column Display*



The following actions are available:

Go to Profile Report List
returns you to the Profile Report List.

Show Outline
displays or hides the outline in the left pane.

Show Notes
displays or hides notes in the right pane. You can filter the notes by typing a text string into the filter field.

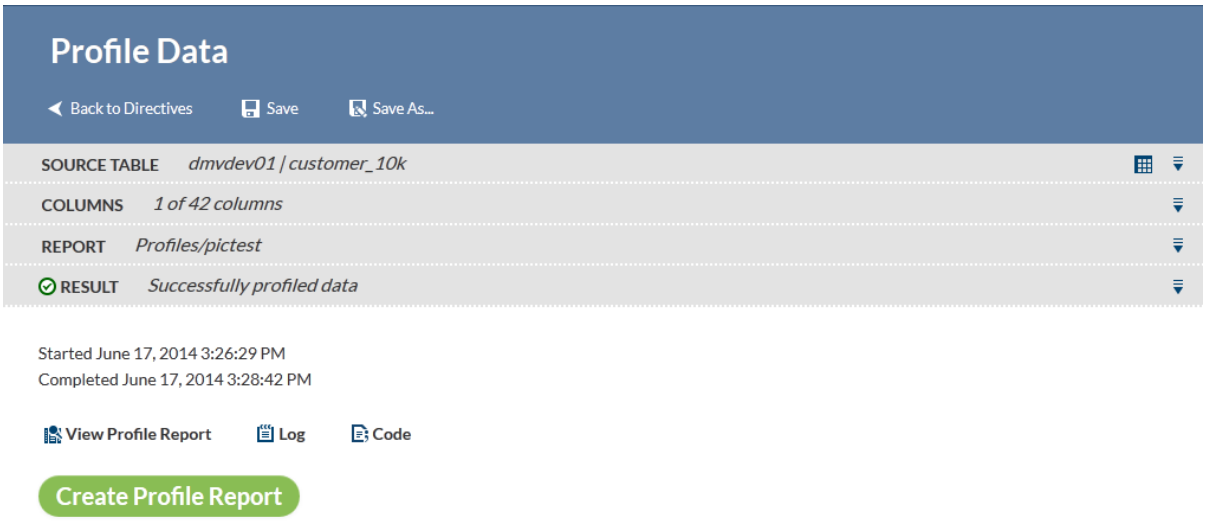
Add Note
opens a dialog box in which you can add a note.

Report Version
enables you to select the version of the report by date.

Opening a Profile from Profile Data

At the end of the procedure to create a data profile using the Profile Data directive, a window similar to the following is displayed:

Figure 8.5 *Profile Data*



Clicking **View Profile Report** takes you directly to Profile Data on a new tab. See Chapter 7, “Profile Data,” on page 37 for more information about the Profile Data directive.

9

Run a SAS Program in Hadoop

A Directive for SAS DS2

The purpose of the directive Run a SAS Program is to enable you to run SAS DS2 programs on an Hadoop cluster. DS2 is a SAS proprietary programming language that is appropriate for advanced data manipulation. DS2 language elements support additional data types, ANSI SQL types, programming structure elements, and user-defined methods and packages. Several DS2 language elements accept embedded FedSQL syntax. Runtime-generated queries can exchange data interactively between DS2 and Hadoop. This allows SQL preprocessing of tables, which effectively combines the power of the two languages.

It is important to note that the directive Run a SAS Program supports SAS DS2 programs only. The programs that you run in this directive are required to use the syntax that is defined in the *SAS DS2 Language Reference*.

To run SAS DS2 programs, install and configure the SAS In-Database Deployment Package for Hadoop on your Hadoop cluster, as described in the *SAS Data Loader for Hadoop: Installation and Configuration Guide*.

For a complete set of releated documents, see “Recommended Reading” on page 69.

Paste and Run a SAS DS2 Program

Follow these steps to create and execute a directive that runs a SAS DS2 data cleansing program on a Hadoop source table.

1 OpenSAS Data Loader.

2 In the Directives page, click **Run a SAS Program**  .

3 Develop a SAS DS2 program, using the syntax that is described in the *SAS DS2 Language Reference*. You can enter program text directly into the text box.

4 To paste your SAS DS2 program into the text box,m , or enter the program text directly.

5 In the Code page, right-click and select **Paste**.

Note: In the pop-up menu, the following options are invalid: Navigate out of code (backward), Navigate out of code (forward), and Syntax Help.

6 Verify that your entire program is now present in the text editor, edit the program as needed, and then click **Next**.

7 In the Result page, click **Start SAS Program**. The directive runs and generates selectable Log icon  . The final status of the directive is portrayed by an icon in the Result banner.

You can monitor the execution of your SAS program in theChapter 4, “Run Status,” on page 27directive.

8 Click **Save** or **Save As** to store your directive in the local Shared Folder.

10

Load Data to LASR

Copy Tables to SAS for Analysis

Use the Load Data to LASR directive to copy Hadoop tables to a grid of SAS LASR Analytic Servers. On the SAS LASR Analytic Servers, you can analyze tables using software such as SAS Visual Analytics.

Note: The Load Data to LASR directive in SAS Data Loader is distinct and separate from the Load to LASR capability that is provided by the SAS LASR Analytic Server.

Prerequisites

Before you can use the Load Data to LASR directive, you must first license, install, and configure a grid of SAS LASR Analytic Servers, version 2.4 or later.

SAS Visual Analytics 6.4 or later is required to be installed and configured on the SAS LASR Analytic Servers.

When the grid of SAS LASR Analytic Servers is operational, you generate and deploy Secure Shell (SSH) keys for SAS Data Loader. For additional information, see “Configure SSH Keys on SAS LASR Analytic Servers” on page 61.

Note: The public key from SAS Data Loader needs to be copied and installed on the head node of the SAS LASR Analytic Server grid each time you update SAS Data Loader.

Another requirement for using the Load Data to LASR directive is to specify SAS LASR Analytic Server connection information in SAS Data Loader. See “Add or Change Connections to SAS LASR Analytic Servers” on page 62.

The SAS LASR Analytic Servers need to be configured to start automatically.

The SAS LASR Analytic Servers must have memory and disk allocations that are large enough to accept Hadoop tables. The Load Data to LASR directive does not check the SAS LASR Analytic Servers for available memory or disk space.

The Load Data to LASR directive moves entire tables. To improve performance, you can filter the rows and manage the columns before you load the table to LASR. To reduce table size, use the directive Transform Data in Hadoop or Query Data in Hadoop.

Load a Target Table into LASR

Follow these steps to create and run the Load Data to LASR directive:

1 OpenSAS Data Loader on page 5.

2 In the Directives page, click **Load Data to LASR**.

3 In the Source Table page, click the schema that contains the source table that you want to load. Clickng the schema displays the tables in that schema. Click the table that you want to load onto your grid of SAS LASR Analytic Servers, and then click **Next**.

4 In the Target Table page, click the SAS LASR Analytic Server that you want to receive the target table. Clicking displays target table configuration fields and controls.

5 As needed, change the name in the **Target table name** field. The field defines the name of the table on the grid of SAS LASR Analytic Servers.

6 Click boxes as needed to replace any existing table of the same name or to compress the target table on the grid of SAS LASR Analytic Servers.

7 Click the **Locations** link to view or change the default storage options for the target table on the grid of SAS LASR Analytic Servers.

8 In the Locations window, you can change the SAS folder, the library name, and the required tag that accompanies the table name.

9 In the Target Table page, click **Next**.

10 In the Result page, click **Start loading data**. SAS proceeds to generate code for the directive and display the **Code** icon  . Click the icon to open or save the text of the SAS code that comprises the directive.

11 During the execution of the directive, the Result page will display the **Log** icon  . Click the icon to open or save the SAS log file that is generated during the execution of the directive.

12 At the conclusion of the directive, the Result banner receives a status icon that indicates the success or failure of the directive. To view the target table on the SAS LASR Analytic Server, click the **View Results** icon  .

11

Administration

Introduction

The administrative tasks in this chapter apply after the initial installation and configuration of SAS Data Loader. To install and configure SAS Data Loader, refer to the *SAS Data Loader for Hadoop: Installation and Configuration Guide*.

Protect the vApp Directory

SAS Data Loader contains encrypted passwords and other sensitive information. Do not share the vApp install directory with other users. Also, be sure to protect that directory by making it accessible only to you.

Troubleshoot Directives

Unsupported Hive Data Types and Values

The Hive database in Hadoop identifies table columns by name and data type. To access a column of data,SAS Data Loader first converts the Hadoop column name and data type into its SAS equivalent. When the transformation is complete, SAS Data Loader writes the data into the target table using the original Hadoop column name and data type.

If your target data is incorrectly formatted, then you might have encountered a data type conversion error.

The Hive database in Hadoop supports a Boolean data type. SAS does not support the Boolean data type in Hive at this time. Boolean columns in source tables will not be available for selection in SAS Data Loader.

The Bigint data type in Hive supports integer values larger than those that are currently supported in SAS. Bigint values that exceed +/-9,223,372,036,854,775,807 generate a stack overflow error in SAS.

Hive Limit of 127 Expressions per Table

Due to a limitation in the Hive database, tables can contain a maximum of 127 expressions. When the 128th expression is read, the directive fails and the SAS log receives a message similar to the following:

```
ERROR: java.sql.SQLException: Error while processing statement: FAILED:
Execution Error, return
       code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask
ERROR: Unable to execute Hadoop query.
ERROR: Execute error.
SQL_IP_TRACE: None of the SQL was directly passed to the DBMS.
```

The Hive limitation applies anytime a table is read as part of a directive. For SAS Data Loader, the error can occur in aggregations, profiles, when viewing results, and when viewing sample data.

Discover New Columns Added after Directive Execution

When you add columns to a source table, any directives that need to use the new columns need to discover them. To make the new columns visible in a directive, open the Source Table page, click the source table again, and click **Next**. The new columns will then be available for use in the body of the directive, in a transformation or query, for example.

About Session Time-out

SAS Data Loader records periods of inactivity in the user interface. After 4 hours of continuous inactivity, the current web page receives a session time-out warning message in a dialog box. If you do not provide input within three minutes after you receive the warning, the current web page is replaced by the Session Timeout page.

You can restart your session by clicking the text **Return to the SAS Data Loader application**.

When a session terminates, any directives that you did not save or run are lost.

To open an unsaved directive that you ran before your session terminated, follow these steps:

1. Open the Run Status directive.
2. Locate the entry for your unsaved directive.
3. If the unsaved directive is still running, click the Refresh button.
4. If the directive continues to run, either click **Stop** in the action menu, or wait for the completion of the run.
5. In the action menu, select **Open** to open the directive.
6. In the open directive, select **Save** from the title bar.

Update SAS Data Loader

Follow these steps to check for the availability of vApp software updates, and to download and install updates.

1. Open or start VMware Player Plus. In VMware Player Plus, note the HTTP address of SAS Data Loader.
2. Type the HTTP address from VMware Player Plus into a web browser and press the Enter key.
3. In the SAS Information Center, locate the Notifications section on the bottom left. If a software update is available, do not click **Update** immediately. Instead, return to SAS

Data Loader, open Run Status, and make sure that you have named and saved your directives. If directives are still running, click **Refresh** to see their current status.

4 For any running directives, either wait for them to complete, or select the **Stop** option from the action menu .

5 Close the SAS Data Loader tab in the web browser.

6 Return to the Information Center and click **Update**. The software update process stops the vApp.

About Hadoop Client JAR Files and Client Configuration Files

SAS Data Loader uses client Java Archive (JAR) files and XML client configuration files to connect the client to Hadoop. Both the JAR and XML files are produced and distributed by Hadoop vendors such as Cloudera and Hortonworks.

The client JAR files are obtained from vendors and pre-installed with SAS Data Loader. At install time, in the SAS Information Center, you select the version of Hadoop that is implemented on your Hadoop cluster. Your selection enables the appropriate JAR files. Within the available supported versions of Hadoop, you can change versions at any time. To change versions, see the next topic “Change the Version of Hadoop”.. New client JAR files for currently unsupported versions of Hadoop will be delivered in updates to SAS Data Loader.

The XML client configuration files are separate from the client JAR files. The XML files are copied from the Hadoop cluster onto the SAS Data Loader client during the installation of SAS Data Loader. The XML files are stored here: ***vApp-install-path**vApp-instance*\SharedFolder\Configuration\HadoopConfig**.

Change the Version of Hadoop

If you move to a different version of Cloudera or Hortonworks, and if that version is supported by SAS Data Loader, then follow these steps to select a different version.

1 Open the SAS Information Center on page 8.

2 Click the Settings menu .

3 In the Settings window, change the value of the **Hadoop version** field.

Change the Hadoop Server Connection

Follow these steps if you reconfigure your Hadoop server or move to a different Hadoop server:

1 OpenSAS Data Loader on page 5.

2 Click the More menu in the top right corner and select **Configuration**.

3 In the Configuration window, change as needed the following fields:

- **Host** specifies the network name of the server
- **Port** specifies the port number that the server listens to for connections from SAS Data Loader.
- **UserID** specifies the user name that SAS Data Loader uses to connect to the Hadoop server.
- **Schema for temporary file storage** specifies an existing schema on the Hadoop server that stores temporary files and tables. To obtain the name of a

non-default schema, open the data sources page in a directive such as Transform Data in Hadoop or Query a Table in Hadoop.

When your changes are complete, click **OK** to apply your changes and close the Configuration window.

Configure SSH Keys on SAS LASR Analytic Servers

Follow these steps when you want to use the Load Data to LASR directive. These steps configure the Secure Shell (SSH) keys for SAS Data Loader on your grid of SAS LASR Analytic Servers. SSH enables secure communication between the server processes in a grid without specifying a password.

After you install SSH keys, add server connection information to SAS Data Loader, as described in the next topic.

Note: Repeat the last step of this procedure if you replace your current version of SAS Data Loader with a new version. Do not repeat the last step after software updates, using the **Update** button in the SAS Information Center.

1. On the SAS LASR Analytic Server grid, create the user `sasdldr1`. To learn how to add users to the grid of SAS LASR Analytic Servers, refer to the *SAS LASR Analytic Server: Reference Guide*.

2. Generate a public key and a private key for `sasdldr1` and install those keys, again as described in the *SAS LASR Analytic Server: Reference Guide*

3. Copy the public key file from SAS Data Loader at ***vApp-install-path**vApp-instance*\Shared Folder\Configuration\sasdemo.pub**. Append the SAS Data Loader public key to the file `~sasdldr1/.ssh/authorized_keys` on the head node of the grid.

   CAUTION! Repeat this last step each time you replace your current version of SAS Data Loader.

Add or Change Connections to SAS LASR Analytic Servers

Follow these steps to define connections to a grid of SAS LASR Analytic Servers. Define these connections only if you want to use the Load to LASR directive.

To enable communication between SAS Data Loader and the grid of SAS LASR Analytic Server, you also need to add SSH keys, as described in the preceding topic.

Note: The grid of SAS LASR Analytic Servers must be fully configured, including the connections to a SAS Metadata Server, before you can configure connections in SAS Data Loader.

1 OpenSAS Data Loader on page 5.

2 Click the More menu  in the top right corner and select **Configuration**.

3 If you are adding a new In the Configuration window, click the plus icon  . If you are changing an existing connection, click that connection in the list.

4 In the LASR Server Configuration window, enter or change your choice of server name and description in the **Name** and **Description** fields.

5 In the **Host** field, add or change the full network name of the host of the SAS LASR Analytic Server. A typical name is similar to saslaser03.us.ourco.com.

6 In the **Port** field, add or change the number of the port that the SAS LASR Analytic Server uses to listen to connections from SAS Data Loader. The default port number is 10010.

7 In the field **LASR authorization service location**, add or change the HTTP address of the authorization service.

8 Under Connection Profile, in the lower of the two **Host** fields, add or change the network name of the SAS Metadata Server that is accessed by the SAS LASR Analytic Server.

9 In the lower of the two **Port** fields, add or change the number of the port that the SAS Metadata Server uses to listen for client connections. The default value 8561 is frequently left unchanged.

10 In the **User ID** and **Password** fields, add or change the credentials that SAS Data Loader will use to connect to the SAS Metadata Server. These values are stored in encrypted form on disk.

11 The Default Locations section specifies where tables will be stored on the SAS LASR Analytic Server by the directive Load Data to LASR. You might need to obtain these values from your SAS administrator. In the **Repository** field, specify the name of the repository on the SAS Metadata Server that is associated with the SAS LASR Analytic Server. The default value Foundation might suffice.

12 In the field **SAS folder for tables**, specify the path inside the repository that will contain the downloads from Hadoop. The default value /SharedData might suffice.

13 In the **Library location** field, add or change the name of the SAS library that will be referenced by the Load Data to LASR directive.

14 In the **LASR server tag** field, add or change the name of the tag that will be associated with each table that is downloaded from Hadoop. The tag is required. It used in along with the table name to uniquely identify tables that are downloaded from Hadoop.

15 Review your entries and click **OK** to return to the Configuration window.

At this point you can define a connection to another SAS LASR Analytic Server.

Enable Logging inside the vApp

For debugging purposes, you can enable logging inside the vApp. To maintain performance, logging is not recommended under normal circumstances.

SAS recommends that you enable logging only when you are directed to do so by your SAS Technical Support representative.

Inside the vApp, log files are generated by a SAS Object Spawner and a series of SAS Workspace Servers. The SAS Object Spawner creates a new instance of the SAS Workspace Server for each HTTP session. When logging is enabled, the SAS Object Spawner generates the log files ObjectSpawner_console_vsasmaster.log and ObjectSpawner_YYYY-MM-DD_localhost_PID.log. The SAS Workspace generates the log file SASApp_WorkspaceServer_YYYY-MM-DD_localhost_PID.log.

The log files are stored in the following location:***vApp-path**vApp-instance*\Shared Folder\Logs**.

Follow these steps to enable or disable logging inside the vApp:

1. Check the Run Status on page 27directive to ensure that no directives are running. This is important because the SAS Object Spawner and other services restart when you enable logging. The same services also restart when you disable logging.
2. Open the SAS Information Center on page 8.
3. Click the Settings icon  .
4. In the Settings window, click **Advanced**.
5. To activate logging, click **Turn logging on (for debugging only)**. To deactivate logging, click **Turn logging off**. Click **OK**.
6. In the window Applying Settings Changes, click **Yes** to confirm your selection. Click **No** to make no change.

Manage Your License

During installation, your installer identified the local storage location of a SAS installation data file (SID). The SID file contains your license. The SID file is delivered as an attachment to the Software Order E-mail.

If your initial license has yet to be applied to SAS Data Loader, follow the renewal steps below.

To check to see if a license has been identified, open the SAS Information Center and click the Settings icon in the top right corner. In the Settings window, if the check box Apply New License has been selected, then your license has been identified to SAS Data Loader.

The license remains valid for the renewal period that is specified in the Software Order E-mail or Renewal Order E-mail.

Ten days before the expiration of your license, the Configuration menu in SAS Data Loader  displays a message that state the number of days available prior to the expiration of the license. On the expiration date, the message states that your license has expired.

On the day of the expiration of your license, the SAS Information Center web page will begin to display the following message each time you open SAS Data Loader:

```
SETINIT ExpirationYour SETINIT for this product is nearing expiration.
```

The word SETINIT is a name that refers to the SAS license.

Note: The expiration message is not displayed in the SAS Information Center when you use the Firefox web browser.

When you begin to receive expiration messages, you should contact your SAS Installation Representative to renew your license. Upon renewal, you will receive a Renewal Order E-mail that contains a new SID file. Follow these steps to identify the new SID file in SAS Data Loader:

1 Save the SID file from your Renewal Order E-mail to a directory on the computer that hosts the vApp.

2 If necessary, start the vApp and open the SAS Information Center.

3 If SAS Data Loader is currently open in a web browser, first close the web browser tab to stop SAS Data Loader. (Doing so will not have a negative impact on any running directives.) Second, open the SAS Information Center by clicking the icon on your desktop. Another way to open the SAS Information Center is to enter into a web browser the web address that is displayed in the VMware Player Plus window.

4 In the SAS Information Center, click the Settings icon  in the top right corner.

5 In the Settings window, click **Apply New License**, and then click **Browse**.

6 In your file browser, navigate to the directory that contains the license file, click the license file, and click **Open**.

7 In the Settings window, click **Yes**.

8 To begin using the new license, simply simply open SAS Data Loader on page 5in a web browser.

You have now renewed your license for SAS Data Loader. The new license will remain valid for the time period that is specified in your Renewal Order E-mail.

Emergency SID Files

Follow these steps to download a temporary SID file that will extend the use of your licensed SAS software products for six days:

1 In a web browser, open the SAS Install Center, at http://support.sas.com/documentation/installcenter/index.html.

2 Under **Site and Account Data** on the right side of the page, select **Request a Temporary License Extension**. You may also select **Resend the SAS Installation Data**.

3 After you receive your temporary SID file, identify that file to SAS Data Loader as described in “Manage Your License”.

Recommended Reading

- *SAS Data Loader for Hadoop: Installation and Configuration Guide*
- *SAS/ACCESS for Relational Databases: Reference*
- *SAS 9.4 In-Database Products: Administrator's Guide*

For a complete list of SAS books, go to support.sas.com/bookstore. If you have questions about which titles you need, please contact a SAS Book Sales Representative:

SAS Books
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-3228
Fax: 1-919-677-8166
E-mail: sasbook@sas.com
Web address: support.sas.com/bookstore

Index

M

P

Q

R

S

T

U

V