



THE
POWER
TO KNOW.

SAS[®] Data Loader 2.2 for Hadoop

vApp Deployment Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2015. *SAS® Data Loader 2.2 for Hadoop: vApp Deployment Guide*. Cary, NC: SAS Institute Inc.

SAS® Data Loader 2.2 for Hadoop: vApp Deployment Guide

Copyright © 2015, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication. The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

NOTICE: This documentation contains information that is proprietary and confidential to SAS Institute Inc. It is provided to you on the condition that you agree not to reveal its contents to any person or entity except employees of your organization or SAS employees. This obligation of confidentiality shall apply until such time as the company makes the documentation available to the general public, if ever.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202–1(a), DFAR 227.7202–3(a) and DFAR 227.7202–4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227–19 (DEC 2007). If FAR 52.227–19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

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

Printing 1, March 2015

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.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Other brand and product names are trademarks of their respective companies.

With respect to CENTOS third party technology included with the vApp ("CENTOS"), CENTOS is open source software that is used with the Software and is not owned by SAS. Use, copying, distribution and modification of CENTOS is governed by the CENTOS EULA and the GNU General Public License (GPL) version 2.0. The CENTOS EULA can be found at http://mirror.centos.org/centos/6/os/x86_64/EULA. A copy of the GPL license can be found at <http://www.opensource.org/licenses/gpl-2.0> or can be obtained by writing to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02110-1301 USA. The source code for CENTOS is available at <http://vault.centos.org/>.

With respect to open-vm-tools third party technology included in the vApp ("VMTOOLS"), VMTOOLS is open source software that is used with the Software and is not owned by SAS. Use, copying, distribution and modification of VMTOOLS is governed by the GNU General Public License (GPL) version 2.0. A copy of the GPL license can be found at <http://www.opensource.org/licenses/gpl-2.0> or can be obtained by writing to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02110-1301 USA. The source code for VMTOOLS is available at <http://sourceforge.net/projects/open-vm-tools/>.

With respect to VIRTUALBOX third party technology included in the vApp ("VIRTUALBOX"), VIRTUALBOX is open source software that is used with the Software and is not owned by SAS. Use, copying, distribution and modification of VIRTUALBOX is governed by the GNU General Public License (GPL) version 2.0. A copy of the GPL license can be found at <http://www.opensource.org/licenses/gpl-2.0> or can be obtained by writing to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02110-1301 USA. The source code for VIRTUALBOX is available at <http://www.virtualbox.org/>.

Contents

Chapter 1 • Introduction	1
Who Should Use This Guide	1
About vApp Deployment	1
How to Use This Guide	1
Chapter 2 • Deploying the SAS Data Loader for Hadoop vApp	3
Overview	3
Step 1: Review System Requirements	4
Step 2: Review Prerequisites	4
Step 3: Download and Expand the SAS Data Loader for Hadoop Client Software	5
Step 4: Create a Shared Folder	7
Step 5: Install and Configure VMware Player Pro	7
Step 6: Open and Configure SAS Data Loader: Information Center	11
Step 7: Copy Hadoop Configuration Files into the Shared Folder	15
Step 8: Start and Configure SAS Data Loader for Hadoop	16
Step 9: Set General Preferences	18
Chapter 3 • Post-Deployment Tasks	21
Overview: Post-Deployment Tasks	21
Updating the SAS Data Loader vApp	21
Updating Your Configuration	22
Usage Notes for VMware Player Pro	22
Closing and Reopening SAS Data Loader for Hadoop	23
About the Shared Folder	24
Recommended Reading	25

Introduction

<i>Who Should Use This Guide</i>	1
<i>About vApp Deployment</i>	1
<i>How to Use This Guide</i>	1

Who Should Use This Guide

This guide is for business analysts and data stewards who are using SAS Data Loader for Hadoop as a self-service way to prepare, integrate, and cleanse big data without writing code. This product enables users to run code and data quality functions on Hadoop, and then to load that data into memory for visualization and analysis. Thus, SAS Data Loader for Hadoop improves productivity, performance, and accuracy.

Note: If you are deploying the SAS Data Loader for Hadoop Trial Edition, do not refer to this guide. Instead, use the instructions that are provided at http://www.sas.com/en_us/software/data-management/data-loader-hadoop.html.

About vApp Deployment

SAS Data Loader for Hadoop runs inside a virtual machine called a vApp. The vApp is a complete and isolated operating environment that is accessed through a web browser. Each instance of SAS Data Loader for Hadoop is accessed by a single user. The vApp is started and stopped by a hypervisor application called VMware Player Pro.

The vApp architecture greatly simplifies the software installation and update processes. The installation runs without user input, and there are no configuration files or system options to configure after installation. The addition of site-specific client configuration settings is simple and quick. The software informs you when an update is available, and a single click installs the update.

How to Use This Guide

For business analysts, data stewards, and other users:

- This guide, the *SAS Data Loader for Hadoop: vApp Deployment Guide*, documents the installation, configuration, and settings of the SAS Data Loader for Hadoop vApp on the client machine. This document also includes the system requirements for the vApp.

Install the vApp after your system administrator has deployed the accompanying offering, SAS In-Database Technologies for Hadoop.

- The *SAS Data Loader for Hadoop: User's Guide* documents how to use the product's directives, and it shows examples of various tasks that you can perform. It also explains how to update your vApp and manage your license.

For system administrators and Hadoop administrators:

- The *SAS Data Loader for Hadoop: Administrator's Guide* documents the installation, configuration, and administration of the offering, SAS In-Database Technologies for Hadoop, on the Hadoop cluster. This document also includes the system requirements for this offering.

Note: SAS In-Database Technologies for Hadoop must be installed first and before the installation of the SAS Data Loader for Hadoop vApp in order for the vApp to communicate successfully with the Hadoop cluster.

Follow the instructions in [Chapter 2, "Deploying the SAS Data Loader for Hadoop vApp," on page 3](#) to deploy the SAS Data Loader for Hadoop vApp on your desktop and play it in VMware Player Pro.

When you complete all of the deployment steps, an instance of the SAS Data Loader for Hadoop web application will be running in a virtual machine on your desktop. The application will be configured to communicate with your Hadoop cluster.

Then refer to [Chapter 3, "Post-Deployment Tasks," on page 21](#) for general information about operating and managing the vApp.

To use the features of SAS Data Loader for Hadoop to interact with your Hadoop data, see *SAS Data Loader for Hadoop: User's Guide*.

2

Deploying the SAS Data Loader for Hadoop vApp

<i>Overview</i>	3
<i>Step 1: Review System Requirements</i>	4
<i>Step 2: Review Prerequisites</i>	4
<i>Step 3: Download and Expand the SAS Data Loader for Hadoop Client Software</i>	5
Introduction	5
Start with the Software Order Email	5
Download the Client Software for SAS Data Loader Using the SAS Download Manager	6
Expand the Contents of the Downloaded ZIP File	7
<i>Step 4: Create a Shared Folder</i>	7
<i>Step 5: Install and Configure VMware Player Pro</i>	7
<i>Step 6: Open and Configure SAS Data Loader: Information Center</i>	11
Open the Information Center	11
Complete the Basic Configuration	13
Complete Additional Steps for Kerberos	14
<i>Step 7: Copy Hadoop Configuration Files into the Shared Folder</i>	15
<i>Step 8: Start and Configure SAS Data Loader for Hadoop</i>	16
<i>Step 9: Set General Preferences</i>	18

Overview

This chapter takes you through the process of deploying the SAS Data Loader for Hadoop vApp on your desktop and playing it in VMware Player Pro.

Note: These instructions are for users who are deploying SAS Data Loader 2.2 for Hadoop. If you are deploying the SAS Data Loader for Hadoop Trial Edition, do not refer to this guide. Instead, use the instructions that are provided at http://www.sas.com/en_us/software/data-management/data-loader-hadoop.html.

Here are the general steps of the installation process, which this chapter covers in detail:

- 1 Review and comply with the system requirements.

- 2 Review and comply with the prerequisites.
- 3 Obtain the Software Order Email, and then download and expand the SAS Data Loader software.
- 4 Create a shared folder for files that are stored and referenced by SAS Data Loader for Hadoop.
- 5 Install the VMware Player Pro hypervisor, and configure it to play the vApp for SAS Data Loader.
- 6 Open the SAS Information Center web application, and configure it to run SAS Data Loader.
- 7 Copy Hadoop configuration files into the shared folder.
- 8 Start the SAS Data Loader web application and complete the configuration process.
- 9 Set the general preferences for SAS Data Loader.

Step 1: Review System Requirements

The client host for SAS Data Loader for Hadoop must meet the following hardware and software requirements:

- Operating system: 64-bit Windows 7 or later, or Windows Server 2008 R2 or later.
- System memory (RAM): Minimum 8 GB, but 16 gigabytes or more is recommended. The SAS Data Loader virtual machine requires at least 4 GB of available memory.
- Disk space: At least 30 GB of free hard-drive space.
- BIOS: Must be enabled for virtualization technology.
- Processors: at least 2 cores (4 logical processors), but 4 core (8 local processors) or more is recommended.
- Web browsers, without Kerberos security: Windows Internet Explorer 9 or later, Mozilla Firefox 14 or later, or Google Chrome 21 or later.
- Web browsers, with Kerberos security: Mozilla Firefox 14 or later, or Google Chrome 21 or later.
- Hypervisor : [VMware Player Pro](#) version 6 or 7.
- Hadoop: Cloudera CDH 5.2 or Hortonworks HDP 2.1.

Step 2: Review Prerequisites

Please meet the following prerequisites before you install and configure the vApp for SAS Data Loader:

- Contact your SAS administrator as needed to confirm that your Hadoop cluster is ready to accept connections from the vApp for SAS Data Loader. To enable these connections, the administrator must first install and configure SAS In-Database Technologies for Hadoop, as described in the *SAS Data Loader for Hadoop: Administrator's Guide*.
- During the configuration process, you need to identify the location of Hadoop JAR files that are stored on the client host. Contact your Hadoop administrator to ensure that the latest Hadoop JAR files have been copied to your computer.
- To use the directives Copy Data to Hadoop and Copy Data from Hadoop, you need to copy JDBC drivers from the Hadoop cluster to the shared folder on your client host, as directed by your Hadoop administrator. For details, see “Install JDBC Drivers and Add Database Connections” in Chapter 5 of *SAS Data Loader for Hadoop: User's Guide*. If the SAS Data Loader for Hadoop vApp is already started, be sure to restart it after you copy the files.
- If your site uses Kerberos security, your Hadoop administrator needs to configure Kerberos on the Hadoop cluster and on your client host, as described in the *SAS Data Loader for Hadoop: Administrator's Guide*. The Hadoop administrator also needs to provide Kerberos configuration values and the locations of the Kerberos files that were installed on your client host. If you do not have this information, contact your Hadoop administrator.
- Configure the supported browser on the client host to support Integrated Windows Authentication (IWA). This configuration process might have been performed by your Hadoop administrator. To confirm or add IWA support, see [Support for Integrated Windows Authentication](#) in the *SAS Intelligence Platform: Middle-Tier Administration Guide*.
- If your site intends to use the directive Load Data to LASR, then your SAS Administrator needs to install a grid of SAS LASR Analytic Servers, release 6.4 or later. For more information, including configuration prerequisites, see “Load Data to LASR” in Chapter 5 of *SAS Data Loader for Hadoop: User's Guide*.

Step 3: Download and Expand the SAS Data Loader for Hadoop Client Software

Introduction

If your computer meets the system requirements, and if the prerequisites are complete, then follow these steps to install the vApp for SAS Data Loader.

Start with the Software Order Email

The Software Order Email outlines the installation process and provides important contact and reference information. You will want to save this email for later reference.

Your Software Order E-Mail includes a license file. Save the license file to an appropriate directory for future reference during the configuration of SAS Data Loader for Hadoop. To keep the license file with the installed software, create

the install directory now, such as `C:\Program Files\SAS Data Loader for Hadoop\2.2`.

The installation steps that are introduced in the Software Order Email are described in this section in greater detail.

Download the Client Software for SAS Data Loader Using the SAS Download Manager

Download SAS Download Manager, and then use SAS Download Manager to download the client software for SAS Data Loader. Follow these steps:

- 1 Open the URL that is specified in the Software Order Email for downloading SAS Download Manager.
- 2 On the Downloads web page, click the version of SAS Download Manager that applies to Windows operating environments.
- 3 On the SAS Login web page, either enter your email address and password, or create a new profile.
- 4 In the SAS Download Manager table, locate the platform **Microsoft Windows for x64**. In that row, click the link in the **Request Download** column.
- 5 Click **Accept** to accept the license agreement for SAS Download Manager.
- 6 If you receive a pop-up message, click **Run** to begin the download.
- 7 In the Ready to Execute dialog box, click **Run**.
- 8 In the Choose Language dialog box, select a language for SAS Download Manager and click **OK**.
- 9 On the Order Information page, enter the order number and installation key that are provided in the Software Order E-Mail. You can copy and paste the installation key. When you are finished, click **Next**.
- 10 If you are prompted to do so, enter your user name and password and click **OK**.
- 11 On the Specify Order Details page, click the link to review your order. In the **Notes** field, add text that identifies this particular order, for future reference. iClick **Next** when you are ready to move ahead.
- 12 On the Specify Order Options page, accept the default selection, which downloads the complete order. Click **Next**.
- 13 On the Specify SAS Software Depot Directory page, enter a new path for a new depot, such as `C:\SAS Data Loader 2.2 Software Depot`. The SAS Data Loader for Hadoop client software must be installed in an empty depot directory.
- 14 On the Final Review page, review and print your order information, and click **Download**. SAS Download Manager proceeds to download your client software order. At any point you can click **Stop the Download Process**. You can restart the process later.
- 15 On the Download Complete page, click **Next**.

- 16 On the final page, review and print the download information, and then click **Finish** to close SAS Download Manager.

Expand the Contents of the Downloaded ZIP File

The SAS Download Manager installs a ZIP file in your SAS software depot. Follow these steps to expand the contents of that ZIP file:

- 1 Open the directory of your SAS Data Loader software depot. Locate the ZIP file in the following directory:

```
your-data-loader-software-depot\SAS Data Loader for Hadoop  
\2_2\VMwarePlayer
```

- 2 Copy the ZIP file and paste it into a program files directory, such as c:\Program Files\SAS Data Loader\2.2.
- 3 If you use WinZip, right-click the ZIP file in the program files directory and select **Open with WinZip**. In the WinZip application, click **Unzip** to expand the compressed files into the current directory.
- 4 If you do not use WinZip, right-click the ZIP file and select **Expand All** to expand the compressed files.
- 5 Wait for the files to expand before you continue.

Step 4: Create a Shared Folder

Create a folder on your local computer in a location that you will remember. You might want to create it within the same directory as the SAS Data Loader program files. This folder, which is referred to as the shared folder, will be used for all the files that are stored and referenced by SAS Data Loader for Hadoop. These files persist between vApp plays and between vApp updates.

You will need to refer to this folder in the next step when you configure VMware Player Pro.

See Also

[“About the Shared Folder” on page 24](#)

Step 5: Install and Configure VMware Player Pro

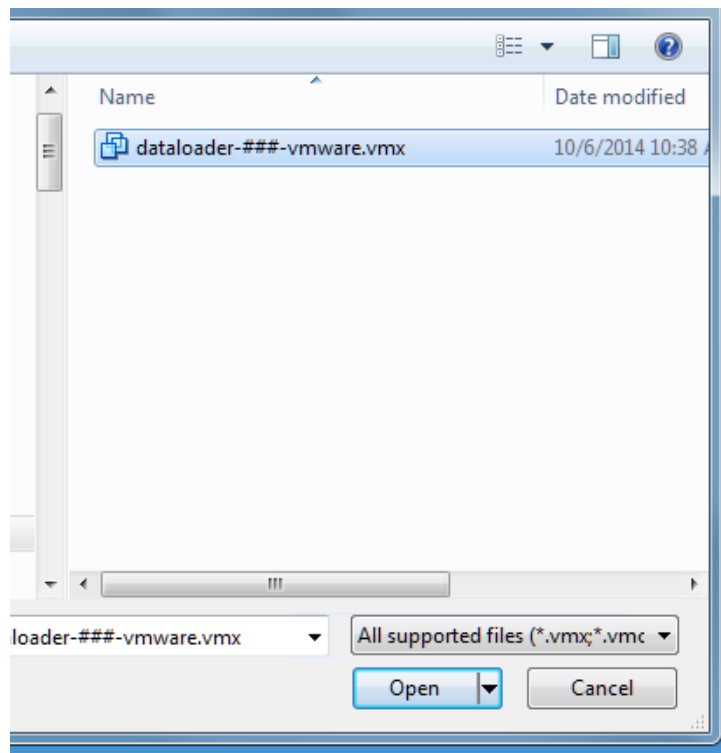
The VMware Player Pro hypervisor enables you to play and power-off the vApp for SAS Data Loader in your Windows operating environment. When you play the vApp, you run a guest operating system in a block of memory that is reserved for that purpose. The hypervisor provides a web address that enables you to open the SAS Data Loader: Information Center and SAS Data Loader for Hadoop web applications.

Follow these steps to install and configure VMware Player Pro:

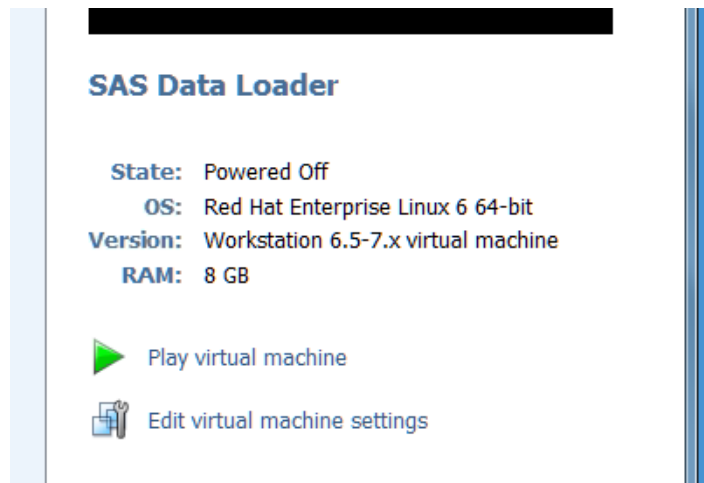
- 1 Purchase and download VMware Player Pro at <http://www.vmware.com/products/player>. A free version of the software, VMware Player, is also supported and is available for personal or non-commercial use.
- 2 Open VMware Player Pro.
- 3 Click **Open a Virtual Machine**.



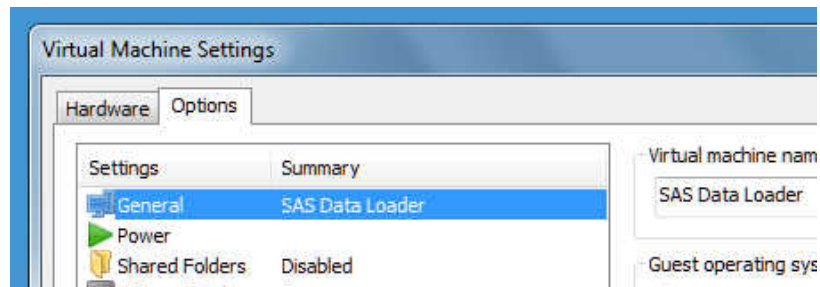
- 4 In the Open a Virtual Machine window, navigate to the directory where you expanded the ZIP file. Select the VMX file for SAS Data Loader, and then click **Open**.



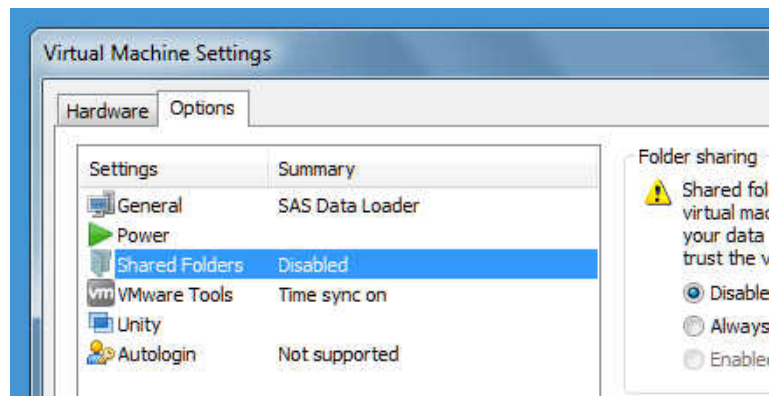
- 5 When the SAS Data Loader information is displayed in VMPlayer Pro, click **Edit virtual machine settings**.



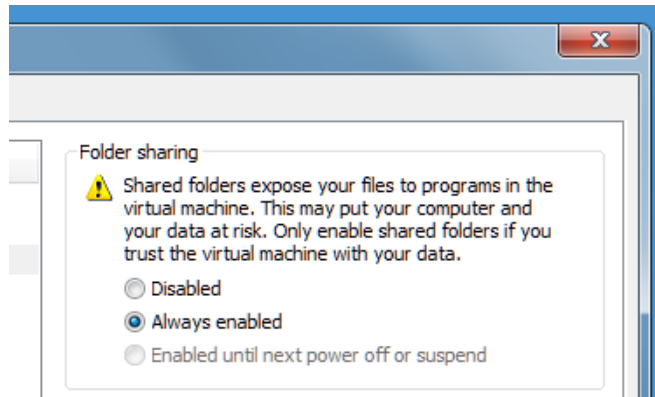
- 6 In the Virtual Machine Settings window, click the **Options** tab.



- 7 Under **Settings**, click **Shared Folders**.

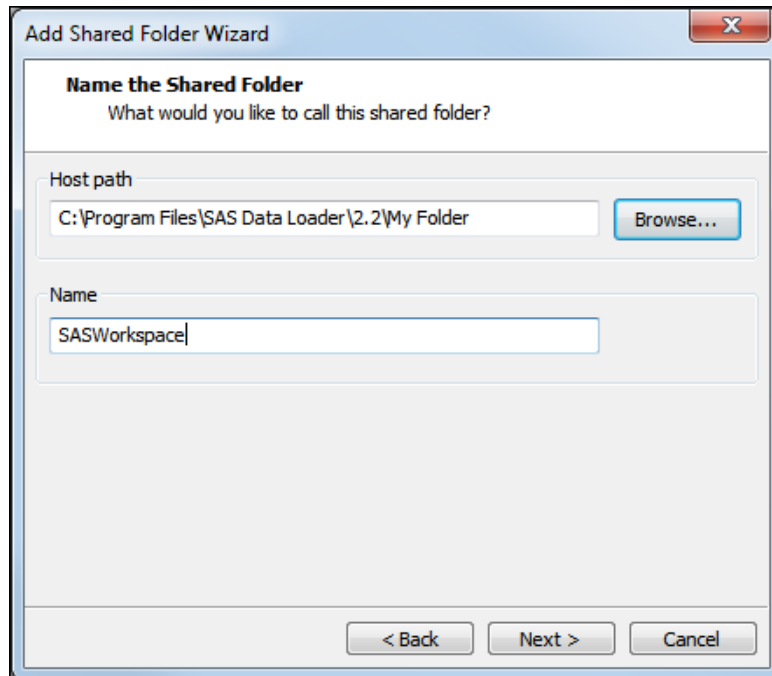


- 8 In the right panel under **Folder sharing**, select **Always enabled**, and then click **Add**.

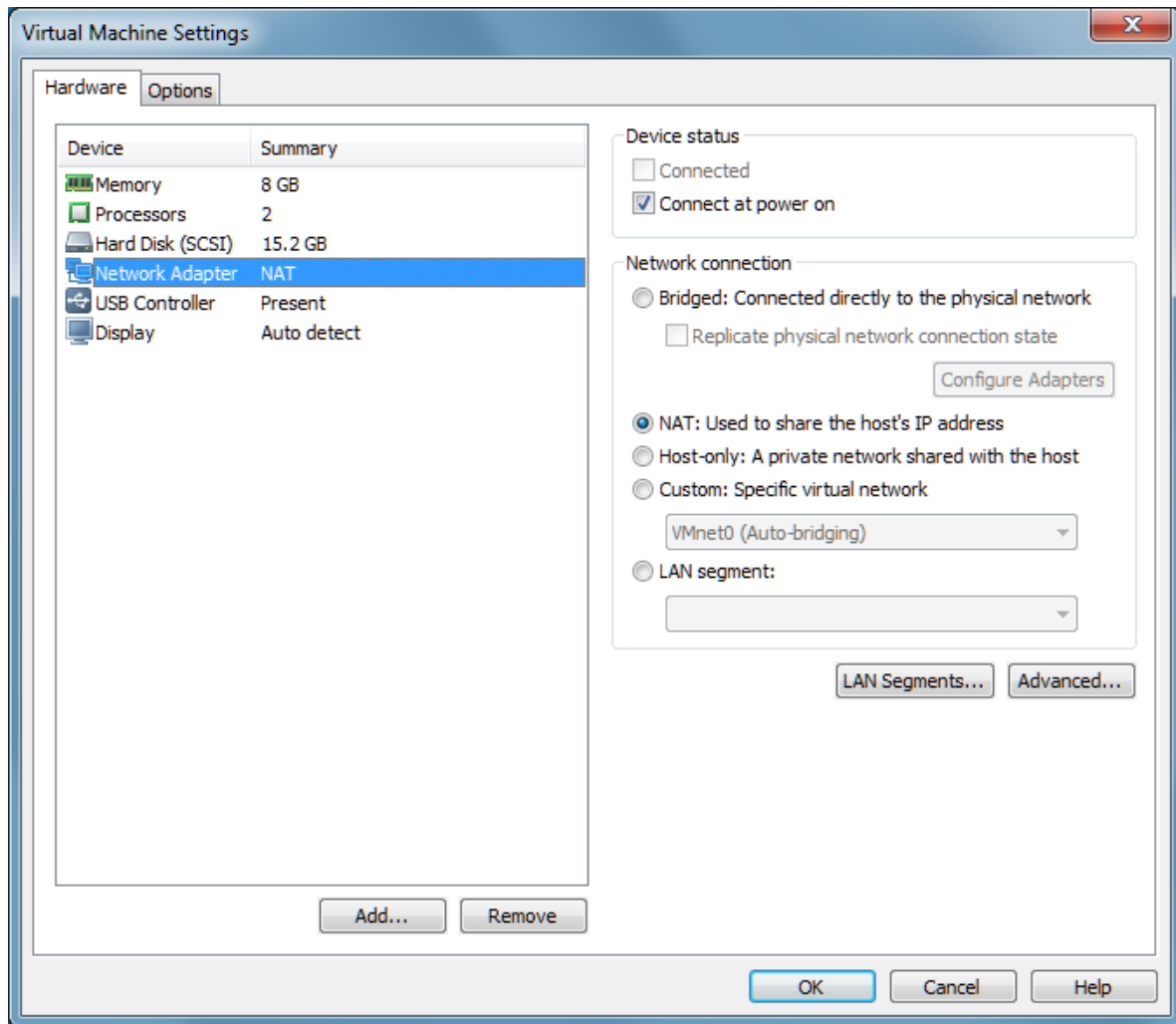


- 9 On the first page of the Add Shared Folders Wizard, click **Next**.
- 10 On the Name the Shared Folder page, click **Browse**, and navigate to the folder that you created in “[Step 4: Create a Shared Folder](#)”. Click **OK**.
- 11 On the Name the Shared Folder page, change the contents of the **Name** field to **SASWorkspace**. Spell this name exactly as shown. Then click **Next**.

Note: The host path can point to a folder of any name that you choose. However, you must specify **SASWorkspace** in the **Name** field.



- 12 On the Specify Shared Folder Attributes page, click **Finish** to accept the default selection and close the wizard.
- 13 In the Virtual Machine Settings window, click the **Hardware** tab. Under **Device**, click **Network Adapter**. Verify or select **Connect at power on** and **NAT: used to share the host's IP address**.



14 Click **OK** to close the Virtual Machine Settings window.

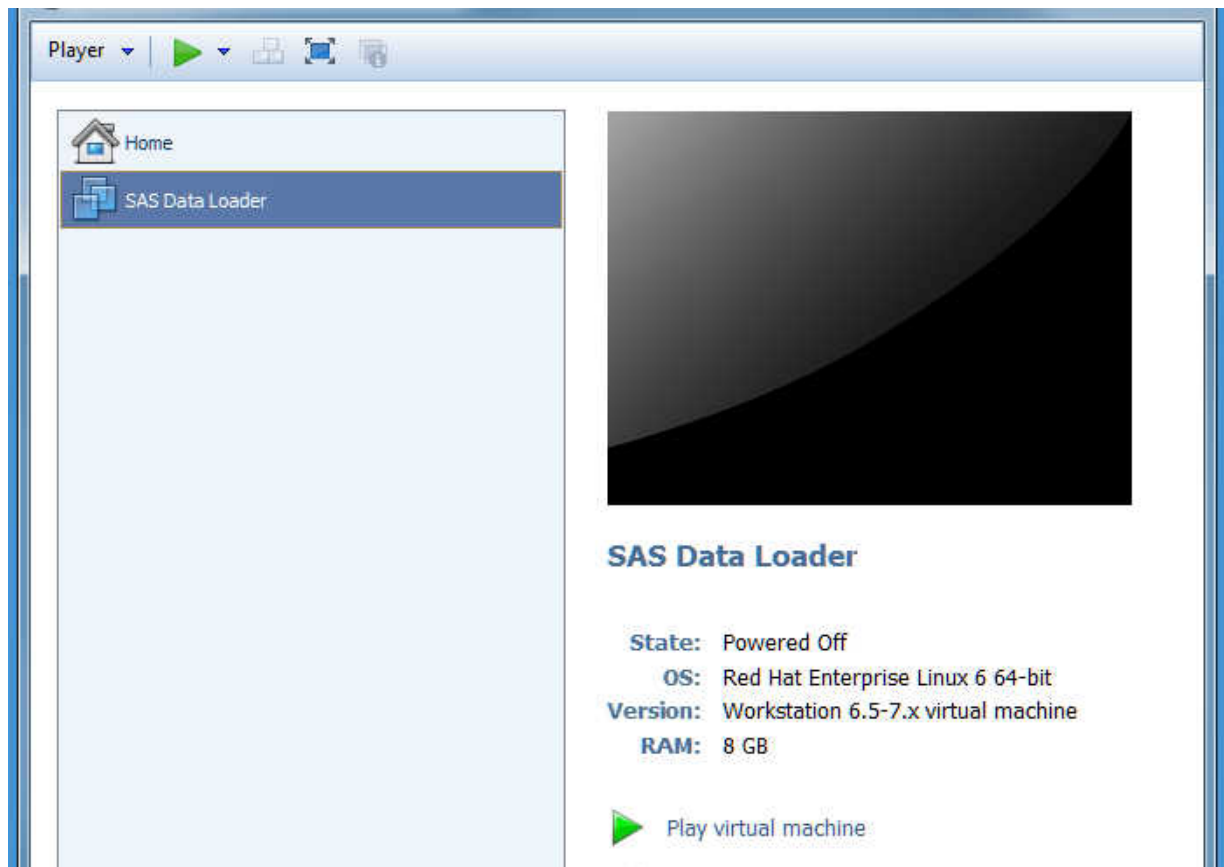
Step 6: Open and Configure SAS Data Loader: Information Center

Open the Information Center

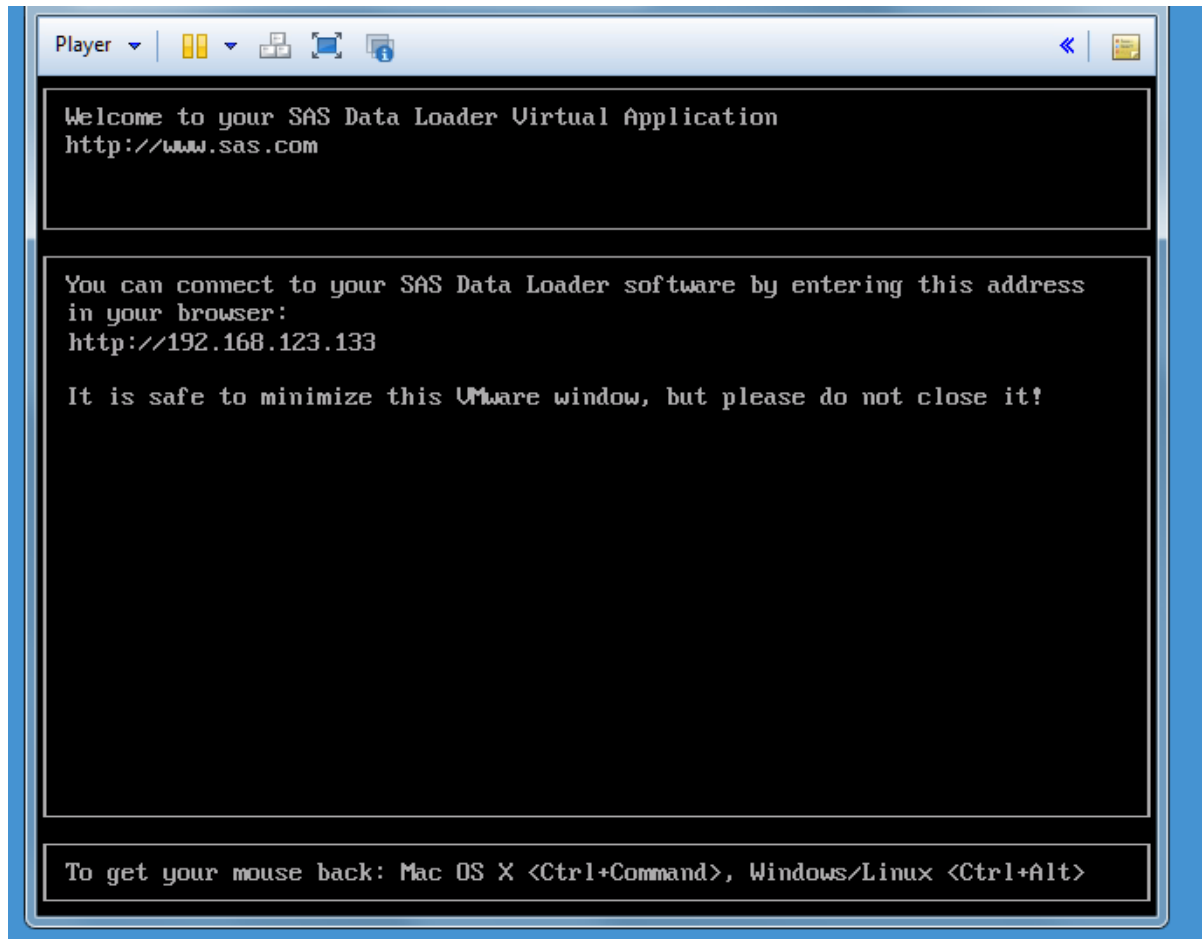
SAS Data Loader: Information Center is a web application that enables you to configure SAS Data Loader for Hadoop, download vApp updates, access product documentation, and start the SAS Data Loader for Hadoop web application.

Follow these steps to open SAS Data Loader: Information Center:

- 1 Open VMware Player Pro if it is not already open, and click **SAS Data Loader**.
- 2 When SAS Data Loader appears, click **Play virtual machine**.



- 3 VMware Player Pro requires a minute or two to play the vApp. When the vApp is ready, VMware Player displays the message *Welcome to your SAS Data Loader Virtual Application*. (If an informational Removable Devices window appears, review the information about removable devices and click **OK**.)

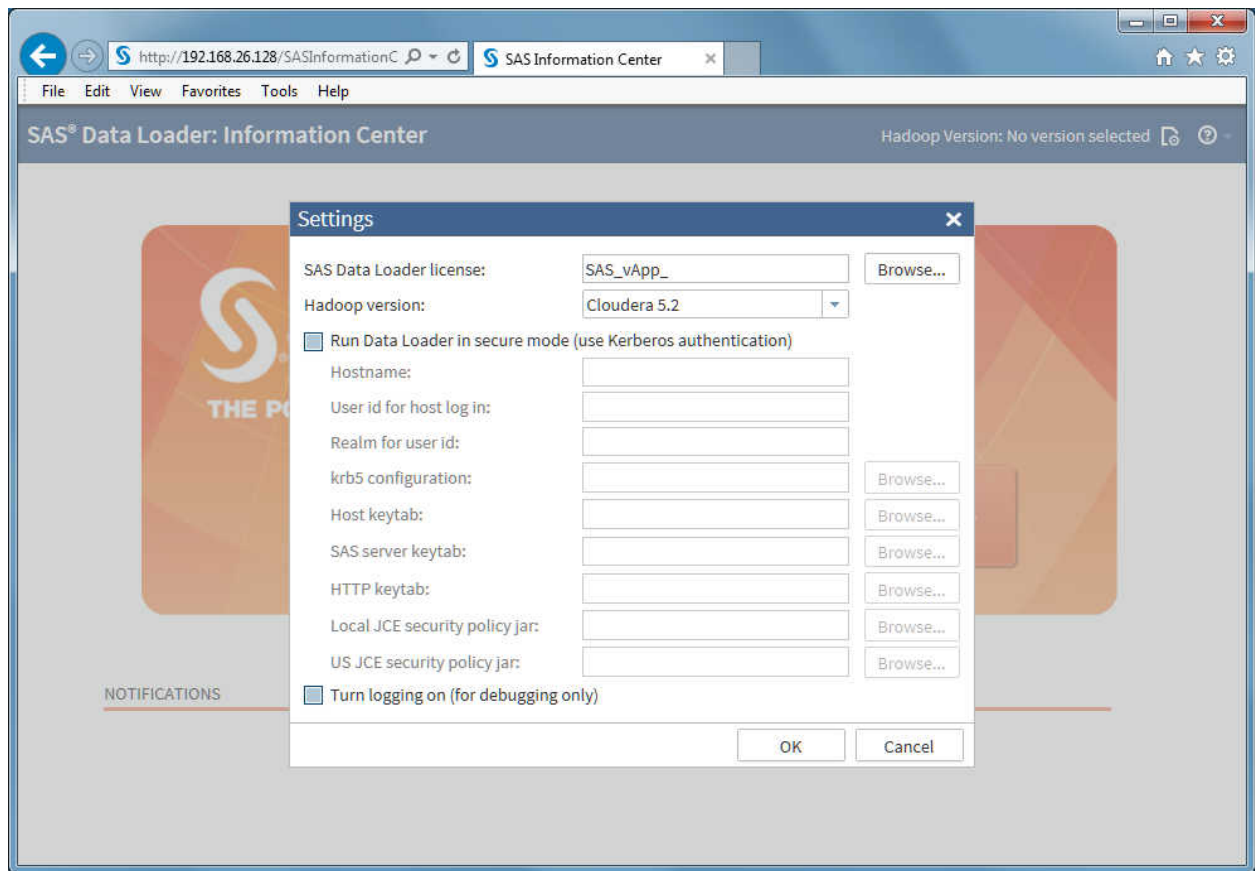


- 4 In the window SAS Data Loader – VMware Player Pro, locate the HTTP address to connect to the SAS Data Loader
- 5 Open a supported browser, and enter the HTTP address in the browser's address bar. Press **Enter** to open SAS Data Loader: Information Center.

Complete the Basic Configuration

Follow these steps to complete the basic configuration of the SAS Data Loader: Information Center:

- 1 The first time you open the Information Center, the Settings window appears. In the **SAS Data Loader license** field, specify the location of your SAS Data Loader license file. The license filename has the format `SAS_vApp_order-number_license.txt`. The file is located in the `sid_files` subdirectory of the software depot where you downloaded the SAS Data Loader software, and it is also attached to your Software Order Email.
- 2 Right-click in the **Hadoop version** field and select a version from the list.



- 3 If your site *does* use Kerberos security, go to [“Complete Additional Steps for Kerberos” on page 14](#).

Note: If you are unsure about the use of Kerberos security at your site, contact your network administrator or Hadoop administrator.

If your site *does not* use Kerberos security, follow these steps:

- a Make sure that you have not selected **Run Data Loader in secure mode**.
- b Click **OK** to save your settings. SAS Data Loader: Information Center proceeds to update configuration files for a minute or two.
- c SAS Data Loader: Information Center displays a message telling you to copy configuration files from the Hadoop cluster to the shared folder. Click **Close**, and continue to [“Step 7: Copy Hadoop Configuration Files into the Shared Folder” on page 15](#) for instructions.

Complete Additional Steps for Kerberos

As one of the prerequisites for installing SAS Data Loader for Hadoop, your Hadoop administrator was asked to provide you with the information that you need to configure Kerberos security. The administrator was also asked to deliver certain files to you or install those files on your client host.

In the Setup window for SAS Data Loader: Information Center, follow these steps to configure Kerberos security on your client host:

- 1 Click **Run SAS Data Loader in secure mode**.

CAUTION! Do not click “Run Data Loader in secure mode” and click “OK” unless you are certain that you will configure a Hadoop cluster that uses Kerberos authentication. After you click **Run Data Loader in secure mode** and click **OK**, you cannot reconfigure your vApp to connect to an unsecured Hadoop cluster. If you need to configure an unsecured Hadoop cluster at that point, you are required for reasons of security to download a new vApp.

- 2 In the **Hostname** field, enter the name of your client host as it was defined by your Hadoop administrator. This host name is different from your normal network host name.
- 3 In the **User ID for host login** field, enter your normal login ID.
- 4 In the **Realm for user ID** field, enter the Kerberos realm for your host name, as provided by your administrator. The Kerberos realm is similar to a Windows domain.
- 5 In the **krb5 configuration** field, enter the location of your Kerberos configuration file.
- 6 In the **Host keytab** field, enter the location of the host keytab file. The three keytab files authenticate the host, SAS, and the vApp HTTP server to one another and to the Active Directory authentication provider.
- 7 In the **SAS server keytab** field, enter the location of the SAS server keytab file.
- 8 In the **HTTP keytab** field, enter the location of the keytab file of the HTTP server in the vApp for SAS Data Loader.
- 9 In the **Local JCE security policy jar** field, enter the location of this Java archive file. This Java Cryptology Extension defines the implementation of encryption at your site.
- 10 In the **US JCE security policy jar** field, enter the location of the United States JCE JAR file.
- 11 Click **OK**. SAS Data Loader: Information Center proceeds to update configuration files for a minute or two.
- 12 SAS Data Loader: Information Center displays a message telling you to copy configuration files from the Hadoop cluster to the shared folder. Click **Close**, and continue to the next topic for instructions.

Step 7: Copy Hadoop Configuration Files into the Shared Folder

With the assistance of your Hadoop administrator, locate the following files on your Hadoop cluster:

- core-site.xml
- hdfs-site.xml
- hive-site.xml

- mapred-site.xml
- yarn-site.xml

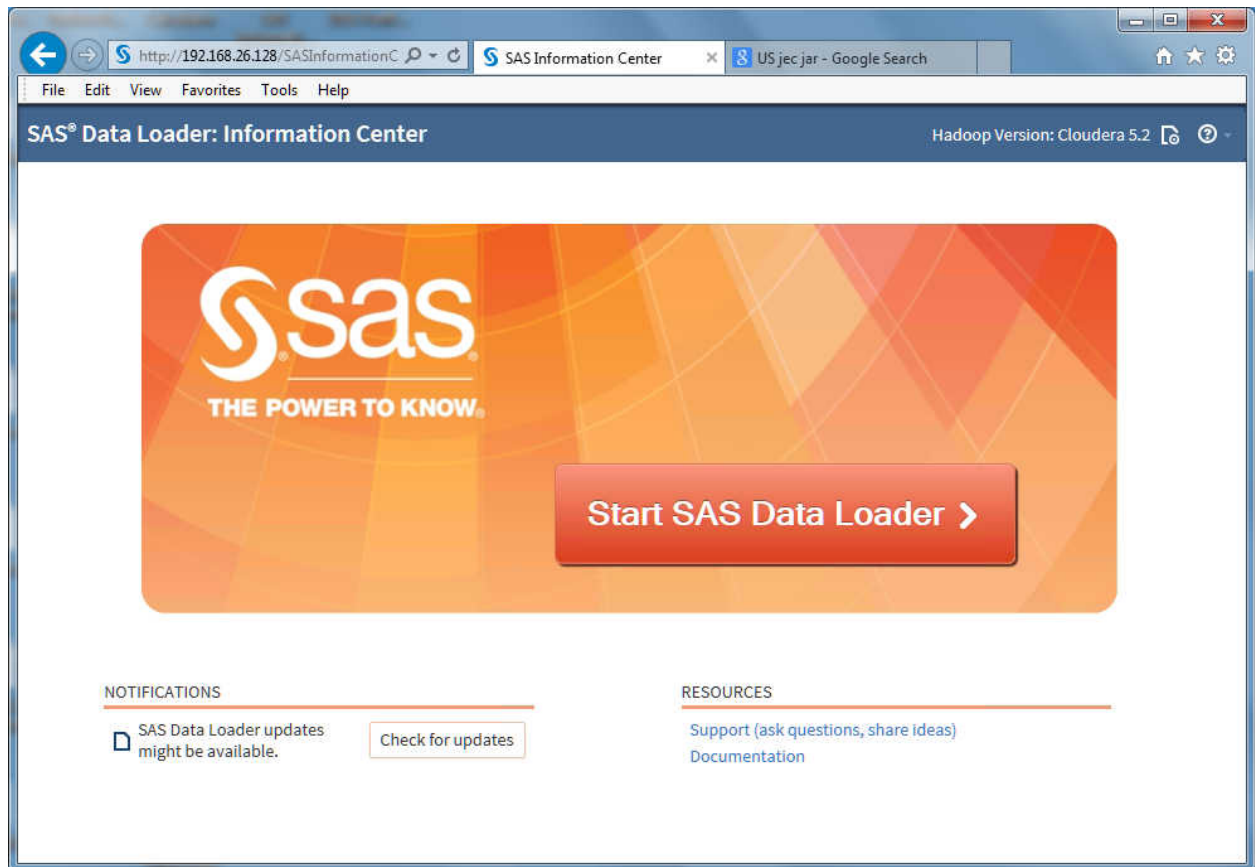
Copy these files and paste them in the following location under the shared folder that you created in “[Step 4: Create a Shared Folder](#)”:

`shared-folder-path\Configuration\HadoopConfig`

Step 8: Start and Configure SAS Data Loader for Hadoop

Follow these steps to start and configure SAS Data Loader for Hadoop:

- 1 Open the [SAS Data Loader for Hadoop: Information Center](#) if it is not already open.
- 2 In the SAS Data Loader: Information Center, click **Start SAS Data Loader**.



Note: When starting SAS Data Loader for Hadoop, if an error occurs stating that VT-x or AMD-v is not available, see “Troubleshoot the vApp Start Process” in Chapter 7 of *SAS Data Loader for Hadoop: User's Guide*.

- 3 The SAS Data Loader web application opens in a new tab in your web browser. The first time you open the application, the Configuration window appears:

- 4 In the **Host** field of the Configuration window, enter the fully qualified name of the host that supports your Hadoop cluster.

Note: Contact your Hadoop administrator as needed to determine Hadoop configuration values.

- 5 In the **Port** field, enter the number of the Hadoop port on the host that supports your cluster.
- 6 In the **User ID** field, enter the name of the user account that will be used to connect to the Hadoop cluster.
- 7 In the **Oozie URL** field, enter the URL to the Oozie Web Console, which is an interface to the Oozie server. The URL is similar to the following example: `http://host_name:port_number/oozie/`. Oozie is a workflow scheduler system that is used to manage Hadoop jobs.
- 8 In the **Schema for temporary file storage** field, either accept the Hive default schema or click **Specify a different schema** and enter the name of an existing Hadoop schema.
- 9 If you intend to use the directive Load Data to LASR (to copy data to an existing grid of SAS LASR Analytic Servers), then click **LASR Analytic Servers**. For additional steps, see “Load Data to LASR” in Chapter 5 of *SAS Data Loader for Hadoop: User’s Guide*.
- 10 At this point you can configure connections to the databases that you will use to copy data to and from Hadoop. To configure database connections now, see “Install JDBC Drivers and Add Database Connections” in Chapter 5 of *SAS Data Loader for Hadoop: User’s Guide*.
- 11 Click **QKB** to view the default locale, which is English. To change the default locale, right-click and select from the list.


12 To configure the processing of profile jobs, click **Profiles** and see “Configure Profile Jobs” in Chapter 4 of *SAS Data Loader for Hadoop: User's Guide*. Profile jobs report on the structure and quality of the data in one or more Hadoop tables.

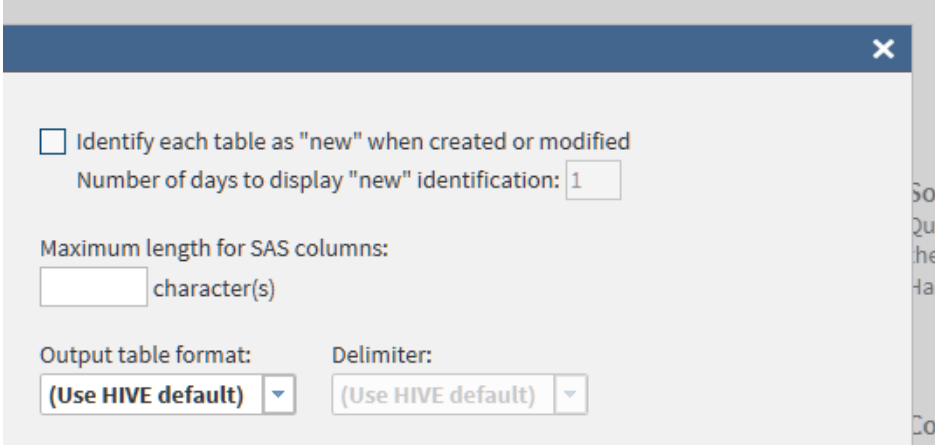
13 Click **OK** to close the **Configuration** window.

To configure general preferences, see “[Step 9: Set General Preferences](#)” on [page 18](#).

Step 9: Set General Preferences

Follow these steps to set general preferences in SAS Data Loader for Hadoop:

- 1** If the **Configuration** window is not already open, click the More icon  in the SAS Data Loader window and click **Configuration**.
- 2** In the **Configuration** window, click **General Preferences**.



- 3** Select **Identify each table as “new”...** to display a new icon with all new source and target tables. Also specify the **Number of days** to display the new icon. The default value is 1.
- 4** Specify a **Maximum length for SAS columns**. This maximum prevents errors and manages table size when character data types are read into SAS or written from SAS using SAS/ACCESS. You can specify any integer value between 1 and 32767. Use caution when setting this value, since data truncation can occur if the specified length is too small to accommodate your data.
- 5** Click **Output table format** to display the following list of available formats:
 - The **Hive default** format is the format that is specified in the Hadoop cluster.
 - **Text**
 - **Parquet** is a structured format that supports the efficient processing of columns with Impala.

- **Orc**, the Optimized Row Columnar format, improves processing efficiency in Hive.
- **Sequence**, or SequenceFile, is a key-value format that is used with MapReduce.

Click the format that you prefer. All of the target tables that you generate will use the selected format.

- 6 If you prefer the text format, then you can also choose a **Delimiter** to separate values in tables. You can choose the Hive default, comma, tab, space, or other. If you choose **Other**, then you enter the delimiter of your choice. The delimiter can be any single character, or a 3-digit octal number, beginning with a backslash. Valid values range from \000 to \177 (1 to 127).
- 7 Click **OK** to save your selections and close the Configuration window.

3

Post-Deployment Tasks

<i>Overview: Post-Deployment Tasks</i>	21
<i>Updating the SAS Data Loader vApp</i>	21
<i>Updating Your Configuration</i>	22
<i>Usage Notes for VMware Player Pro</i>	22
<i>Closing and Reopening SAS Data Loader for Hadoop</i>	23
Closing and Reopening SAS Data Loader: Information Center	23
Closing and Reopening the SAS Data Loader for Hadoop Browser Tab	23
Powering Off the SAS Data Loader for Hadoop vApp	23
Reopening SAS Data Loader	23
<i>About the Shared Folder</i>	24

Overview: Post-Deployment Tasks

This section provides the following information about ongoing operation and management of the vApp:

- [“Updating the SAS Data Loader vApp”](#)
- [“Updating Your Configuration”](#)
- [“Usage Notes for VMware Player Pro”](#)
- [“Closing and Reopening SAS Data Loader for Hadoop”](#)
- [“About the Shared Folder”](#)

See Also

SAS Data Loader for Hadoop: User's Guide for detailed information about using the SAS Data Loader for Hadoop functionality.



Updating the SAS Data Loader vApp

All of the client software for the SAS Data Loader for Hadoop runs inside the vApp. The vApp is a virtual machine that runs a separate operating system. All of the files that are accessed by the vApp are stored in a Shared Folder that resides in this host operating environment. This architecture enables you to

install vApp updates with one-button simplicity. Each update completely replaces the entire vApp. After the vApp update, there are no configuration or migration procedures.

vApp updates require less than 15 minutes, given reasonable broadband capacity. When you update the vApp, you might also see an Information Center link to notes that describe the release's changes.

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

- 1 Open the browser tab for the SAS Data Loader: Information Center if it is not already open. See [“Closing and Reopening SAS Data Loader: Information Center” on page 23](#).
- 2 Locate the **Notifications** section in the bottom left corner of SAS Data Loader: Information Center.
- 3 To check to see whether a vApp update is available, click **Check for Updates**.
- 4 If a vApp update is available, open the **Run Status** directive to ensure that you have named and saved your jobs. If jobs are still running, click **Refresh**  to see their current status.
- 5 For any running directives, either wait for them to complete, or select the **Stop** option from the action menu .
- 6 Close the **SAS Data Loader** tab in the web browser.
- 7 Return to SAS Data Loader: Information Center and click **Update**. The software update process stops the vApp, replaces the vApp, and then starts the new vApp in the VMware hypervisor.
- 8 When the SAS Data Loader: Information Center indicates that the vApp update is complete, click **Start SAS Data Loader**.

Updating Your Configuration

See Chapter 7, “Client Administration,” in *SAS Data Loader for Hadoop: User's Guide* for information about changing or updating your initial configuration.

Usage Notes for VMware Player Pro

You can close and reopen the SAS Data Loader web application without shutting down the vApp. The vApp continues to play until you shut it down in VMware Player Pro.

Do not close the SAS Data Loader – VMware Player Pro window while the vApp is playing.

Note that if the vApp is playing, and if you click in the window SAS Data Loader – VMware Player Pro, the cursor disappears. This behavior is expected; it

ensures that you have to physically enter the web address in a web browser to open SAS Data Loader: Information Center. To restore your cursor, click Ctrl+Alt.

CAUTION! Do not pause or suspend the vApp for SAS Data Loader. VMware Player Pro provides a capability to suspend vApps. Suspending the vApp for SAS Data Loader is not supported. Do not select **Player ► Suspend Guest**. Suspending the vApp can interrupt communications between the SAS Data Loader web client and the Hadoop cluster. Use the Power Off/Shutdown Guest option instead.

Closing and Reopening SAS Data Loader for Hadoop

Closing and Reopening SAS Data Loader: Information Center

While the vApp is still playing, you can close the tab for SAS Data Loader: Information Center at any time without closing the browser tab for SAS Data Loader for Hadoop.

To reopen SAS Data Loader: Information Center, enter the HTTP address (from the display in VMware Player Pro) in the browser's address bar.

Closing and Reopening the SAS Data Loader for Hadoop Browser Tab

While the vApp is still playing, you can close the browser tab for SAS Data Loader at any time. Any jobs that are running on the Hadoop cluster continue to run, and their run status continues to be collected.

To reopen SAS Data Loader after you close its browser tab, open SAS Data Loader: Information Center and click **Start SAS Data Loader**.

Powering Off the SAS Data Loader for Hadoop vApp

To completely close SAS Data Loader, you need to power off (shut down) the vApp in VMware Player Pro, as follows:

- 1 In the browser, close the tab for SAS Data Loader if it is open.
- 2 In the SAS Data Loader – VMware Player window, click **Player ► Power ► Shut Down Guest**. (The term guest refers to the guest operating system that runs the vApp.)
- 3 In the VMware Player dialog box, click **Yes** to confirm that you want to power off the vApp.

Reopening SAS Data Loader

To reopen SAS Data Loader after it has been powered off, follow these steps:

- 1 Open VMware Player Pro.
- 2 Select **SAS Data Loader**, and then select **Play virtual machine**.
- 3 In the window SAS Data Loader – VMware Player Pro, locate the HTTP address to connect to the SAS Data Loader. Enter the HTTP address in the browser's address bar. The SAS Data Loader: Information Center opens in a new tab in your browser.
- 4 In the SAS Data Loader: Information Center, click **Start SAS Data Loader**. The SAS Data Loader web application opens in a new tab in your browser.

About the Shared Folder

The shared folder contains all of the files that are stored and referenced by the SAS Data Loader for Hadoop. The folder is located outside of the vApp. For example, it can be located in the same directory as your SAS Data Loader program files. The contents of the shared folder persist across vApp updates, which eliminates the need for migration or configuration during those updates.

You create and configure the shared folder during the deployment process. (See [“Step 4: Create a Shared Folder” on page 7](#) and [“Step 5: Install and Configure VMware Player Pro” on page 7](#).) In the computer's file system, you can give the folder any name that you choose. However, **SASWorkspace** must be specified in the **Name** field of the vApp's shared folder settings in VMware Player Pro.

The subfolders that are created inside the shared folder are as follows:

Configuration

contains `sasdemo.pub`, which is an SSH key file. The file is moved to a grid of SAS LASR Analytic Servers when Hadoop data is to be loaded into those servers for analysis using the Load Data to LASR directive.

Configuration\DMServices and HadoopConfig

contains a database of Hadoop configuration data and all saved directives. The database is populated automatically after your initial configuration of the SAS Data Loader: Information Center.

InClusterBundle

contains two self-extracting script files (*.sh) that administrators use to deploy the SAS In-Database Technologies for Hadoop across the Hadoop cluster. A JAR file (*.jar) contains utilities that deploy the SAS Quality Knowledge Base across the Hadoop cluster.

Logs

contains the log files that are generated when you enable vApp logging in the **Settings** menu in SAS Data Loader: Information Center.

JDBCDrivers

contains the drivers that enable the loading of data into and out of Hadoop.

Profiles

contains all of the profiles that are created with the Profile Data directive.

SASData

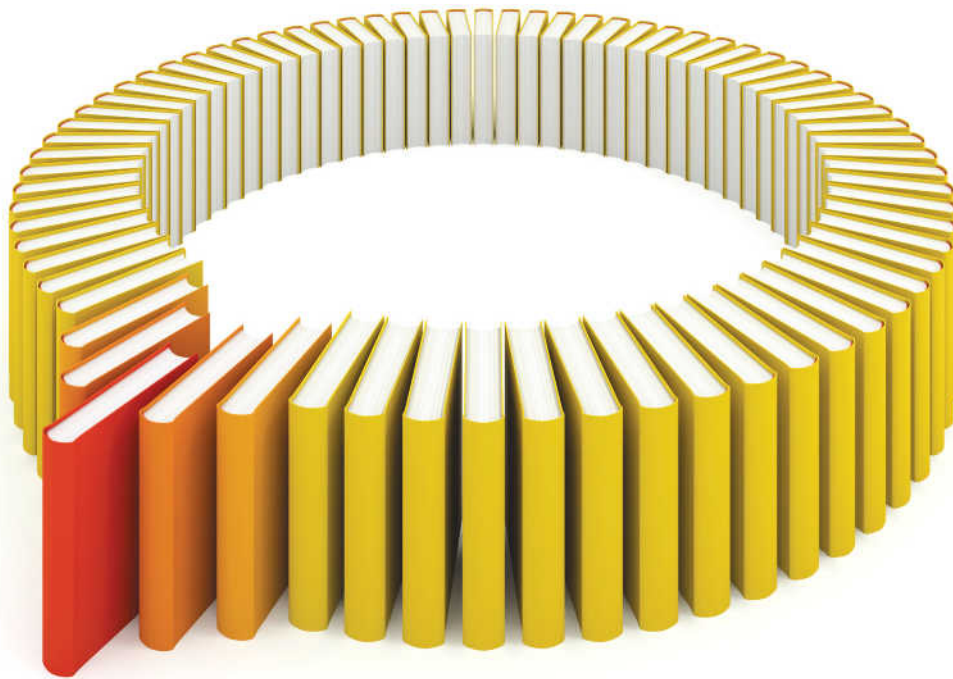
contains the SAS data that enables the SAS instance on the client host to communicate with the SAS instances on the Hadoop cluster.

Recommended Reading

- *SAS Data Loader for Hadoop: User's Guide*
- *SAS Data Loader for Hadoop: Administrator's Guide*
- *SAS 9.4 DS2 Language Reference*
- *SAS/ACCESS for Relational Databases: Reference*
- *SAS 9.4 In-Database Products: Administrator's Guide*
- *SAS Quality Knowledge Base for Contact Information 23: Installation and Configuration* (see the online Help for usage information)
- *Introduction to SAS® and Hadoop Course Notes*
- *The Little SAS® Book: A Primer*

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



Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

 support.sas.com/bookstore
for additional books and resources.


THE POWER TO KNOW.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. © 2013 SAS Institute Inc. All rights reserved. S107969US.0613

