



THE  
POWER  
TO KNOW.

# **SAS<sup>®</sup> Data Loader 2.3 for Hadoop**

vApp Deployment Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2015. *SAS® Data Loader 2.3 for Hadoop: vApp Deployment Guide*. Cary, NC: SAS Institute Inc.

### **SAS® Data Loader 2.3 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, July 2015

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](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/>.

---

## Contents

<b>Chapter 1 • Introduction</b>	<b>1</b>
About This Guide	1
Related Guides	1
What's New	2
About the Trial Edition	2
<b>Chapter 2 • Set Up and Run the vApp</b>	<b>3</b>
Before You Begin	3
Step 1: Download the Software for SAS Data Loader	5
Step 2: Copy Hadoop Configuration Files to Your Machine	7
Step 3: Configure the vApp in VMware Player Pro	8
Step 4: Configure SAS Data Loader	13
Step 5: Complete the Hadoop Configuration	17
Next Steps	19
<b>Chapter 3 • Migrate from Version 2.2 to 2.3</b>	<b>21</b>
What Gets Migrated?	21
Steps for Migrating	21
<b>Chapter 4 • Administration Tasks</b>	<b>25</b>
Tips for Running the vApp	25
Play the vApp and Start SAS Data Loader	26
Power Off the vApp	27
Check for vApp Updates	27
Manage Your License	28
Back Up and Restore	29
Configuring a New Version of Hadoop	31
Enable Logging and Download Log Files	32
<b>Chapter 5 • Troubleshooting</b>	<b>35</b>
Support Community	35
Troubleshooting Tips	35
<b>Recommended Reading</b>	<b>39</b>



## 1

# Introduction

<i>About This Guide</i> .....	1
<i>Related Guides</i> .....	1
<i>What's New</i> .....	2
<i>About the Trial Edition</i> .....	2

## About This Guide

Thank you for choosing SAS Data Loader for Hadoop. After you perform steps in this guide, SAS Data Loader for Hadoop will be running as a virtual machine, or vApp, on your computer.

Anyone who installs business client software can use this guide. To deploy the vApp successfully, you will need files from SAS, and a supported version of VMware Player Pro. The files from SAS comprise the SAS Data Loader for Hadoop vApp, and VMware Player Pro provides the environment where you will configure and run the vApp.

## Related Guides

The following guides are also available:

- *SAS Data Loader for Hadoop: User's Guide* is for business analysts and data stewards. This guide documents how to configure and use the directives.
- *SAS In-Database Products: Administrator's Guide* is for system and Hadoop administrators. Before you can complete the vApp deployment, an administrator must perform steps in this guide to install and configure the offering, SAS In-Database Technologies for Hadoop, on the Hadoop cluster.

If you need additional help with deploying or administering the SAS Data Loader for Hadoop, the support community is a great place to find answers.

[SAS Data Loader for Hadoop Community](#)

Join the community to ask questions and receive expert online support.

---

## What's New

For information about new features and enhancements, see *SAS Data Loader for Hadoop: User's Guide*

---

## About the Trial Edition

A trial edition of SAS Data Loader for Hadoop is available at the following web page:

[http://www.sas.com/en\\_us/software/data-management/data-loader-hadoop.html](http://www.sas.com/en_us/software/data-management/data-loader-hadoop.html)

If you install the trial edition, make sure you follow the instructions provided with the trial edition, and not this guide.

## 2

## Set Up and Run the vApp

<b>Before You Begin</b>	<b>3</b>
Review System Requirements	3
Make Sure You Have the SAS Software Order Email	3
VMware Player Pro Is Required	4
Get Information from Your Hadoop Administrator	4
Do You Want to Migrate from Version 2.2?	5
<b>Step 1: Download the Software for SAS Data Loader</b>	<b>5</b>
Download the Software with SAS Download Manager	5
Extract Files from the ZIP File	6
<b>Step 2: Copy Hadoop Configuration Files to Your Machine</b>	<b>7</b>
<b>Step 3: Configure the vApp in VMware Player Pro</b>	<b>8</b>
<b>Step 4: Configure SAS Data Loader</b>	<b>13</b>
<b>Step 5: Complete the Hadoop Configuration</b>	<b>17</b>
<b>Next Steps</b>	<b>19</b>

### Before You Begin

#### Review System Requirements

To view system requirements, go to the following location and search for *Data Loader*:

<https://support.sas.com/documentation/installcenter/94/index.html>

After the search completes, look for a link to system requirements on the results page.

#### Make Sure You Have the SAS Software Order Email

The Software Order Email for SAS Data Loader 2.3 vApp contains the license file and the link for downloading the SAS Download Manager. SAS Download Manager is used to download the required files from SAS.

## VMware Player Pro Is Required

To complete the deployment, a supported version of VMware Player Pro for Windows 64-bit is required. Visit the VMware website at one of the following locations.

- To purchase VMware Player Pro:  
<http://www.vmware.com/products/player>
- To download the trial version for personal or non-commercial use:  
<https://www.vmware.com/products/player/playerpro-evaluation.html>

## Get Information from Your Hadoop Administrator

To connect the vApp successfully to the Hadoop cluster, refer to this section to make sure that you have the right information before you start.

### Required for Installation

You must copy Hadoop related files to the machine where you run the vApp. Also, you must configure settings to connect to the Hadoop server.

Ask your Hadoop administrator for the following information.

What You Need	Details
<b>Files for Hadoop connectivity and configuration</b>	<p>The files are provided in two folders: <code>conf</code> and <code>lib</code>.</p> <p><b>Note:</b></p> <ul style="list-style-type: none"> <li>■ Ask your Hadoop administrator to make these folders available so that they can be copied to your machine. For the vApp to connect to your Hadoop cluster, your Hadoop administrator must deploy software in the Hadoop cluster. During that deployment, your Hadoop administrator can choose to collect these folders, which are required to connect to Hadoop.</li> <li>■ The folder names must be spelled as shown. The <code>conf</code> folder contains Hadoop client configuration files, including one or more JSON files with detailed information about the cluster. The <code>lib</code> folder contains Hadoop client JAR files.</li> </ul>
<b>Kerberos settings</b>	<p>These security settings are required if your Hadoop environment uses Kerberos as an authentication protocol.</p>
<b>Host name</b>	<p>The fully qualified network name of the HiveServer2 for your Hadoop cluster.</p>
<b>Port number</b>	<p>The port number of the HiveServer2 for your Hadoop cluster.</p>



What You Need	Details
<b>User ID and Password</b>	<p>The user name and password that SAS Data Loader for Hadoop uses to connect to the Hadoop server.</p> <p><b>Note:</b> These settings are required if your Hadoop environment does not use Kerberos as an authentication protocol.</p>
<b>Oozie URL</b>	<p>This setting is required if you want to use the <b>Copy Data to Hadoop</b> and <b>Copy Data from Hadoop</b> directives to move data between Hadoop and another data source.</p> <p><b>Note:</b> Additional configuration is required to use the directives. For more information, see <i>SAS Data Loader for Hadoop: User's Guide</i>.</p>

### Additional Security Settings

Your system administrator might perform the following security configuration on the machine where the vApp is deployed. The following information is not part of the vApp deployment, but it is good to know before you start.

- For your machine to access a Hadoop cluster secured by Kerberos, a host name (for your machine) must be added to the system `hosts` file on your machine.
- For single sign-on to work with Firefox and Chrome browsers, the browser must be configured to support Integrated Windows Authentication.

**Note:** Your Hadoop administrator can refer to *SAS In-Database Products: Administrator's Guide* for more information about these security settings.

## Do You Want to Migrate from Version 2.2?

Migrating enables you to use saved information, such as saved directives and profiles, with a newer version of SAS Data Loader. For more information, see [Chapter 3, "Migrate from Version 2.2 to 2.3," on page 21](#).

## Step 1: Download the Software for SAS Data Loader

### Download the Software with SAS Download Manager

- 1 Open your Software Order Email for SAS Data Loader 2.3 vApp, and click the URL 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, 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 Email. 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. Click **Next** to continue.
- 12 On the Specify Order Options page, accept the default selection, which downloads the complete order. Click **Next** to continue.
- 13 On the Specify SAS Software Depot Directory page, enter a new path for the new depot.

**TIP** SAS Data Loader for Hadoop 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.

**Note:** If necessary, you can click **Stop the Download Process**, and 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.

**TIP** Save the Software Order Email so that you can refer back to it as needed.

## Extract Files from the ZIP File

- 1 Locate the ZIP file in the directory where you downloaded your depot.  
`SAS-software-depot-directory\SAS_Data_Loader_for_Hadoop  
\2_3\VMWarePlayer`

- 2 Copy the ZIP file, and paste it into a directory, such as `C:\Program Files\SAS Data Loader\2.3`.
- 3 In Windows Explorer, navigate to the file location, and unzip it to a directory on your machine that VMware Player Pro can access:
  - If you use WinZip, right-click the ZIP file, and select **Open with WinZip**. In the WinZip application, click **Unzip**.
  - If you do not use WinZip, right-click the ZIP file, and select **Extract All**.

**TIP** Make sure the directories used by SAS Data Loader are secure. SAS Data Loader contains encrypted passwords and other sensitive information. Do not share the vApp install directories with other users, and protect it by making it accessible only to you.

---

## Step 2: Copy Hadoop Configuration Files to Your Machine

- 1 Use Windows Explorer to create a folder named `SASWorkspace` on your computer. This folder is the shared folder for the vApp. The folder name is case-sensitive, so enter the name exactly as shown.

**TIP** Remember the location of this folder. In a later step, you select this folder to be the shared folder, where VMware files and all files that are stored and referenced by the vApp are saved. Also, each vApp instance must have its own shared folder. You cannot create a shared folder on a shared drive and attempt to share it with multiple vApps.

- 2 In the `SASWorkspace` folder, create a new folder named `hadoop`. The folder name is case-sensitive, so enter the name exactly as shown.
- 3 Copy the two folders provided by your Hadoop administrator into the `SASWorkspace\hadoop` directory.

**TIP** The two folders, named `conf` and `lib`, contain the required files to configure the vApp to use the Hadoop cluster. The folder names are case-sensitive and must be named exactly as shown.

When you complete this step, you should have the following required folders on your machine:

```
SASWorkspace\hadoop\conf
```

```
SASWorkspace\hadoop\lib
```

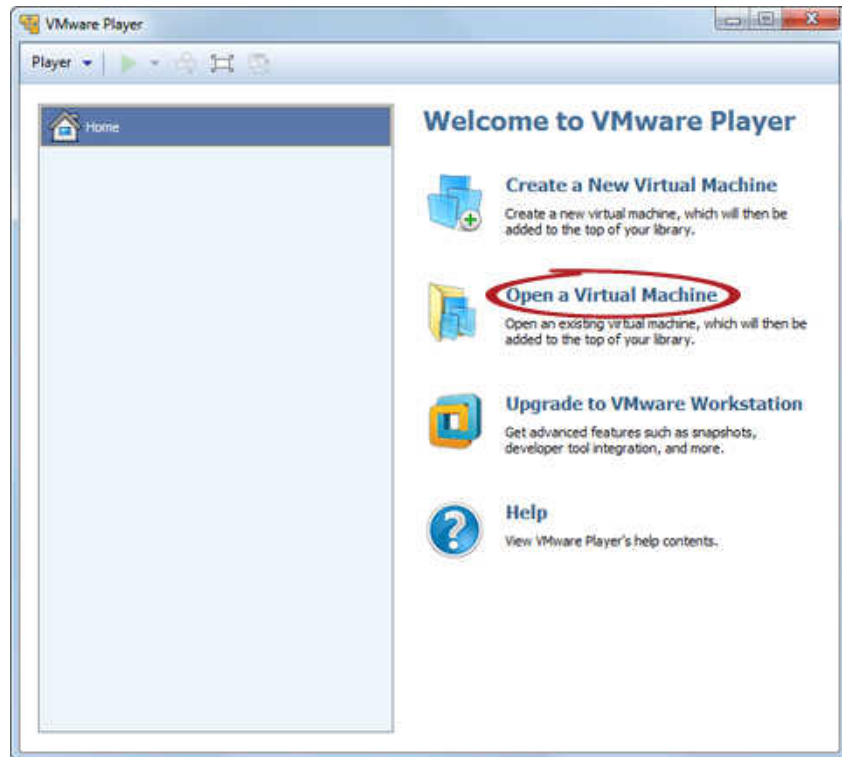
---

## Step 3: Configure the vApp in VMware Player Pro

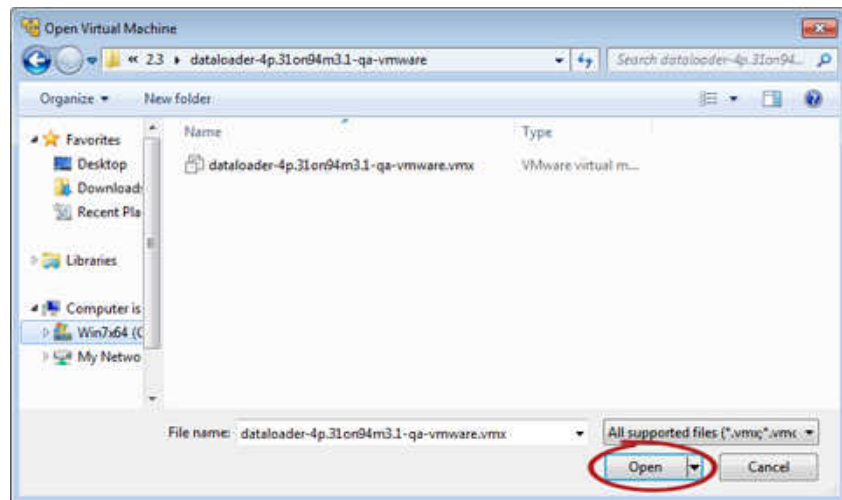
- 1 Open VMware Player Pro.

**TIP** If do not have VMware Player Pro, see “VMware Player Pro Is Required” on page 4.

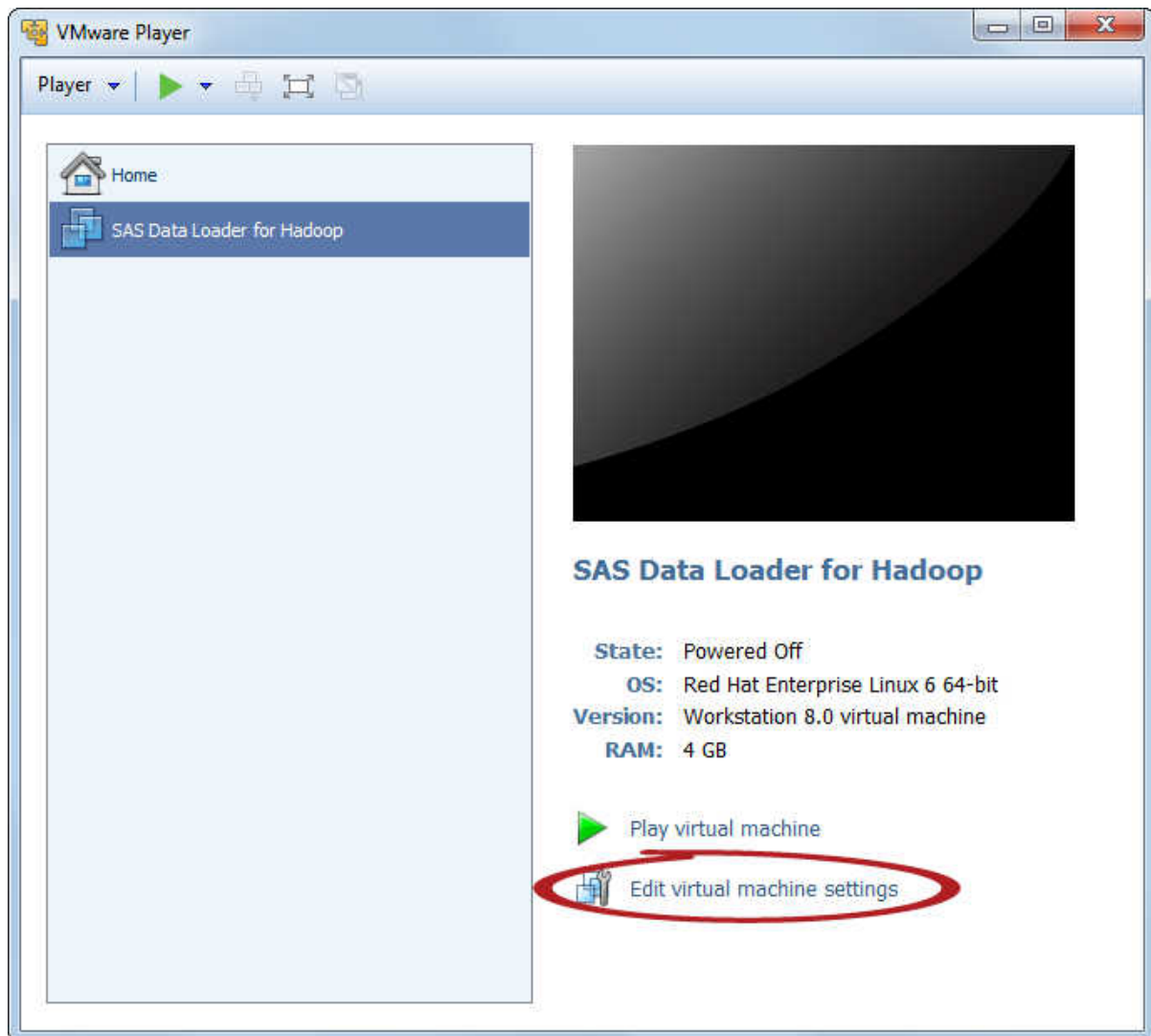
- 2 Click **Open a Virtual Machine**.



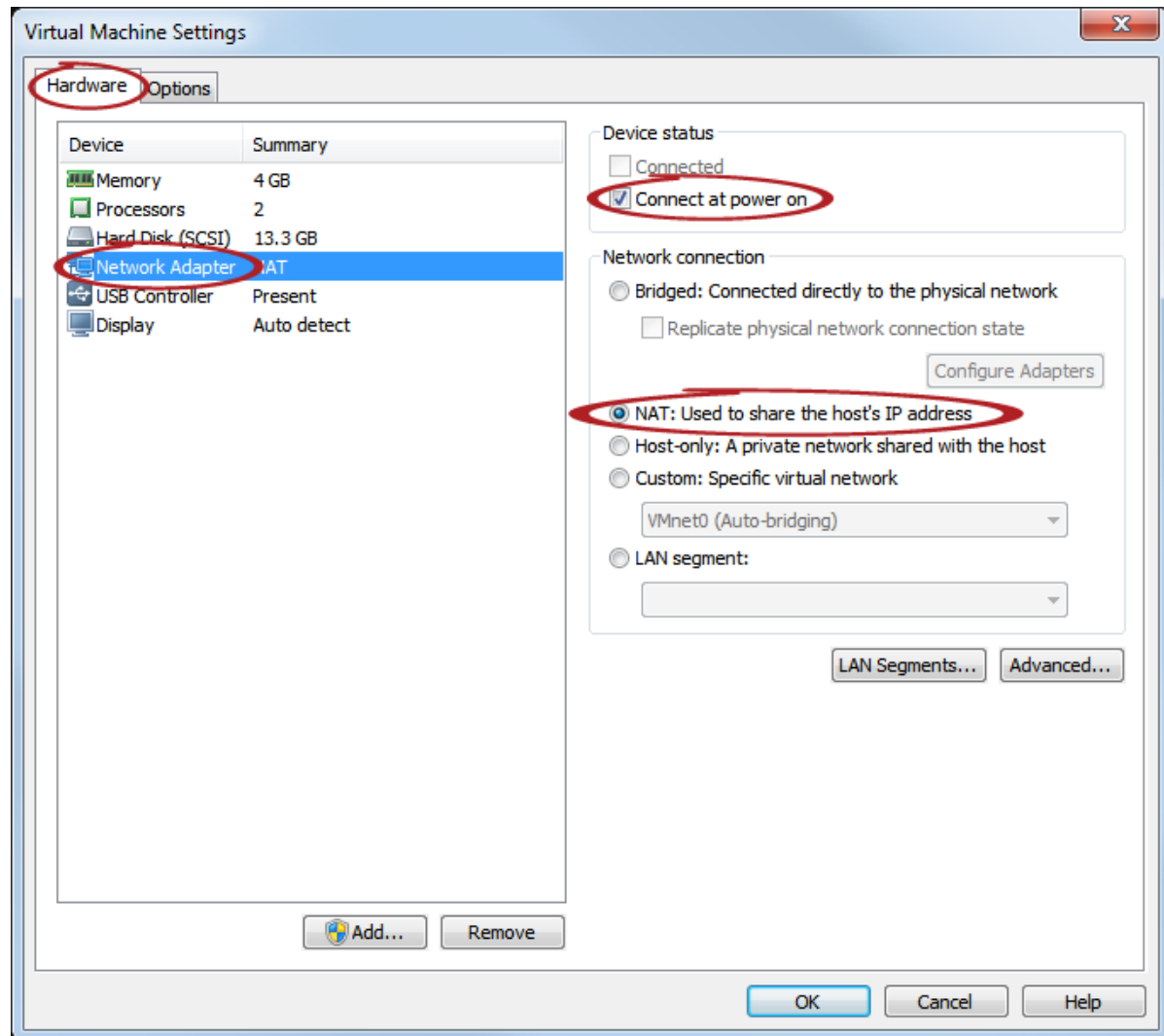
- 3 Navigate to the directory where you extracted files from the ZIP file. Select the VMX file for SAS Data Loader, and then click **Open**.



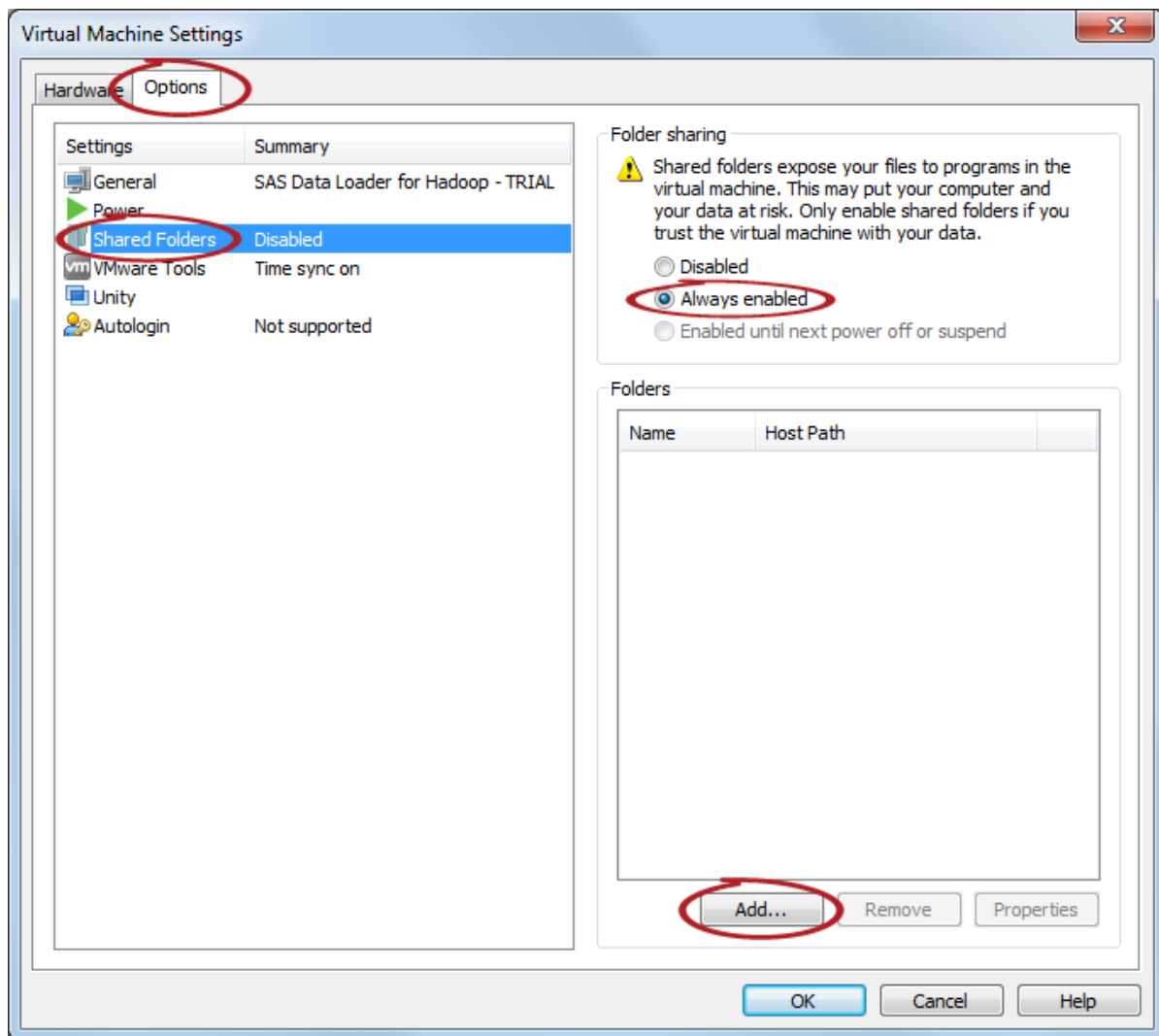
- 4 After SAS Data Loader for Hadoop displays in VMPlayer Pro, click **Edit virtual machine settings**.



- 5 In the Hardware tab, select **Network Adapter** in the left panel. In the right panel, select **Connect at power on** and **NAT: Used to share the host's IP address**.



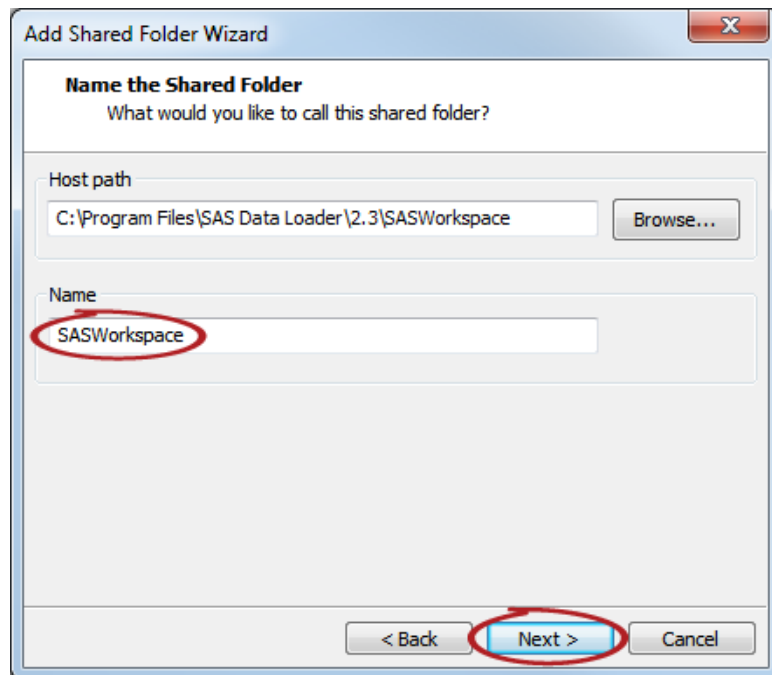
- 6 Click the **Options** tab, and then select **Shared Folders** in the left panel. In the right panel, select **Always enabled**, and then select the **Add** button.



7 Click **Next** to start the Add Shared Folder wizard.

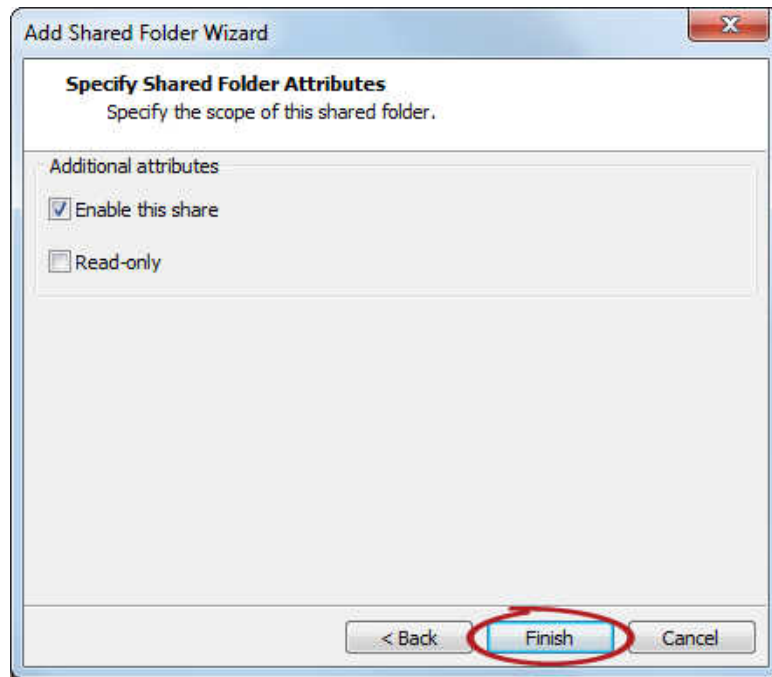


- 8 Click the **Browse** button, and locate your SASWorkspace folder. Make sure that the Name field contains **SASWorkspace**. The name is case-sensitive and must be spelled exactly as shown. Click **Next** to continue.



- 9 On the next screen, click **Finish** to accept the default selection.



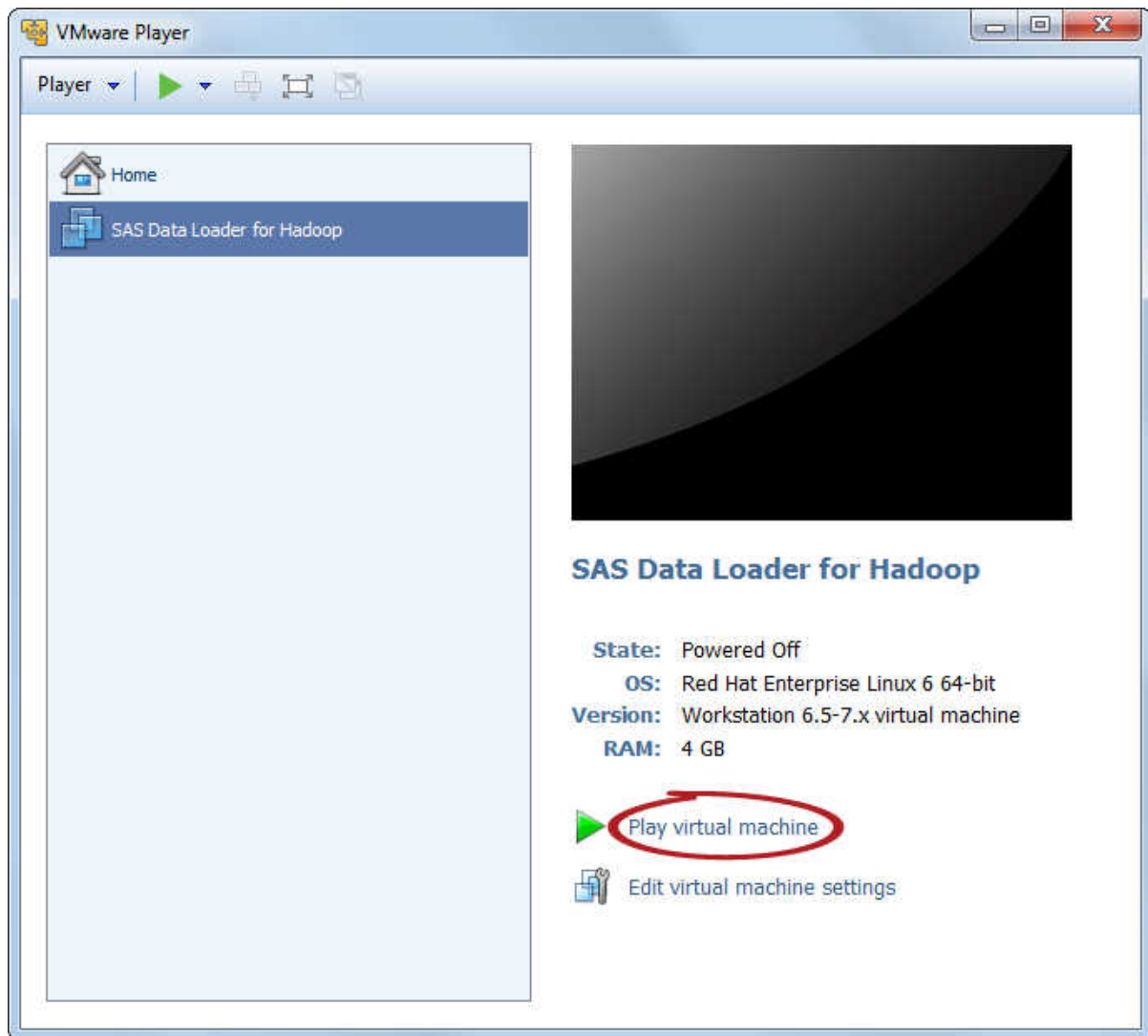


10 Click **OK** to close the Virtual Machine Settings window.

---

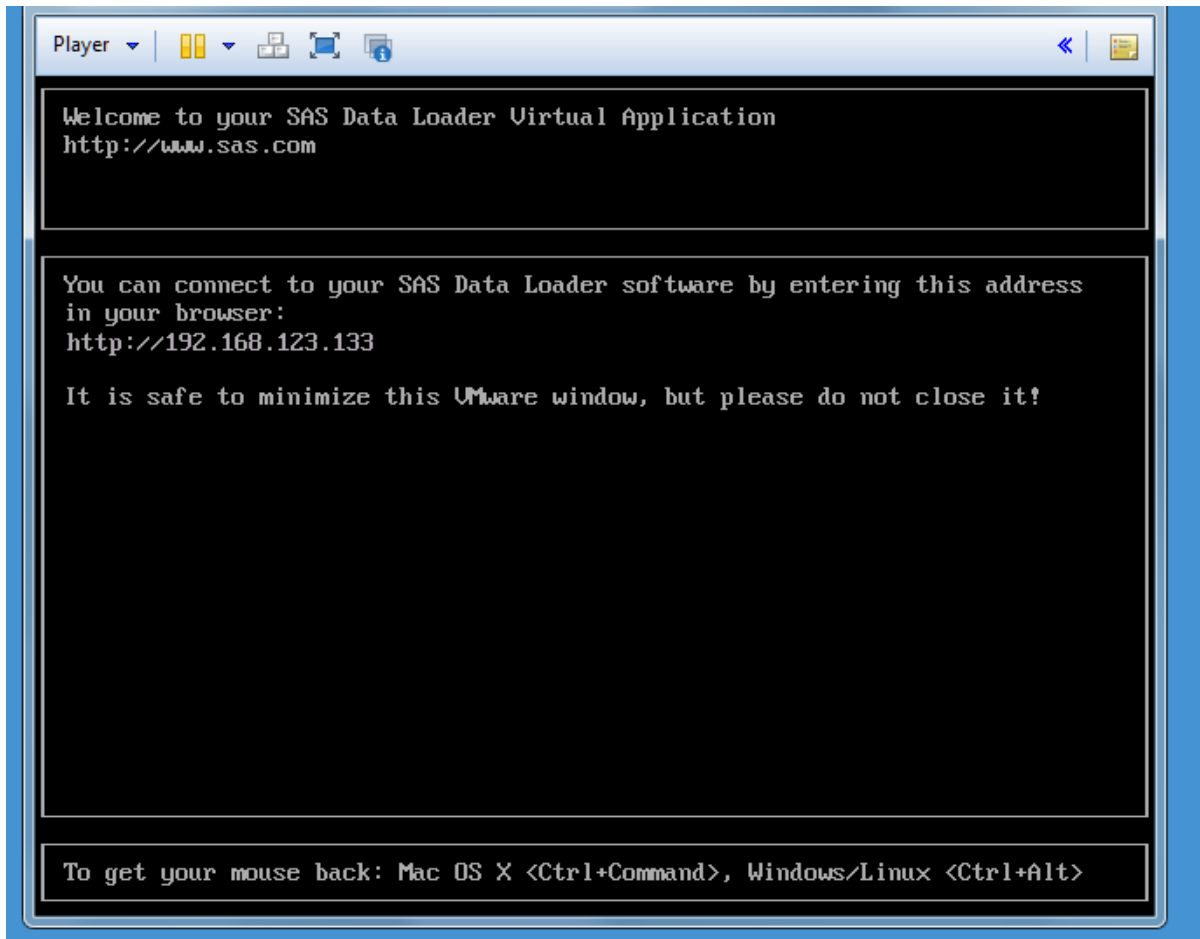
## Step 4: Configure SAS Data Loader

- 1 In VMware Player Pro, make sure **SAS Data Loader for Hadoop** is selected, and then click **Play virtual machine**.



**Note:** VMware Player Pro requires a minute or two to play the vApp.

- 2 When the vApp is ready, VMware Player displays the message:  
Welcome to your SAS Data Loader Virtual Application.



**Note:** If an informational Removable Devices window appears, review the information about removable devices and click **OK**.

- 3 Locate the HTTP address to connect to the SAS Data Loader. For example, the HTTP address shown in the preceding image is `http://192.168.123.133`.
- 4 Open a web browser, and enter the HTTP address in the browser's address bar. Press **Enter** to continue.
- 5 The first time you open the SAS Data Loader: Information Center, the Settings window appears.

Click the **Browse** button to locate your license file. 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.

**TIP** The license filename has the format `SAS_vApp_order-number_license.txt`.

**Note:** The license for one instance of the vApp for SAS Data Loader for Hadoop applies to a single user on a single client machine.

- 6 Do one of the following steps:
  - If you received Kerberos information from your Hadoop administrator, go to [Step 7 on page 16](#).
  - If you did NOT receive Kerberos information from your Hadoop administrator, make sure you do NOT select **Run Data Loader in secure mode (use Kerberos authentication)**. Next, click **OK** to exit this window.
- 7 Perform the following steps to run SAS Data Loader in secure mode.
 

**CAUTION! Perform the following steps if you are certain that your Hadoop cluster 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.

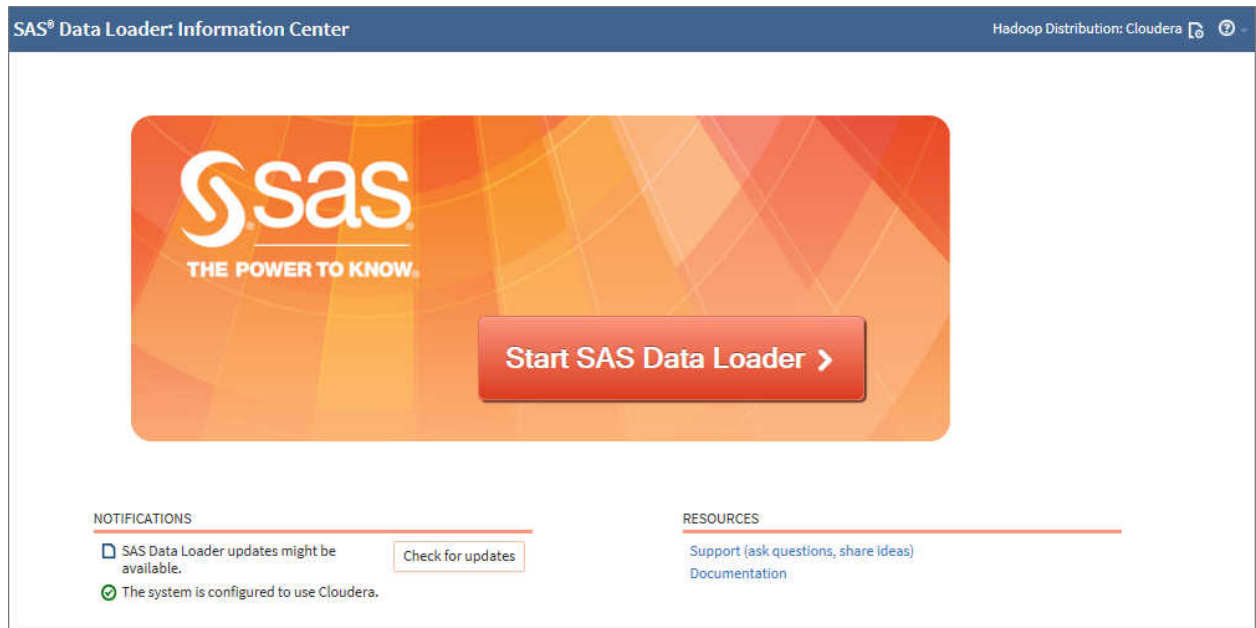
  - a Click **Run SAS Data Loader in secure mode**.
  - b Enter values for the following fields. Ask your Hadoop administrator for the values. All fields are required.
    - Host name
    - User ID
    - Kerberos realm
    - Kerberos configuration file
    - Host keytab file
    - SAS server keytab file
    - HTTP keytab

- Local JCE security policy jar
- US JCE security jar
- c Click **OK** to save the settings.

**Note:** On some systems, the update can take several minutes to complete.

## Step 5: Complete the Hadoop Configuration

- 1 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 [“Troubleshooting Tips” on page 35](#) for assistance.

- 2 SAS Data Loader opens in your web browser. The first time you open the application, the Configuration window appears.

**Note:** If you are using a backup, the Configuration window does not appear.

**TIP** To complete the configuration in the following steps, ask your Hadoop administrator for the values. Also, if you configured to run SAS Data Loader in secure mode as described in [Step 7 on page 16](#), the User ID and Password fields are not editable.

**Configuration**

**Hadoop Configuration**

LASR Analytic Servers

Databases

QKB

Profiles

General Preferences

Host: \*

Port: \*

User ID: \*

Password:

Oozie URL:

Schema for temporary file storage:

☒ Use the default schema

☐ Specify a different schema: ?

Hive storage location: ?

☒ Use the default Hive location

☐ Specify alternate storage location: /user/etlguest ...

SAS HDFS temporary storage location: ?

☒ Use the default temporary location

☐ Specify alternate storage location: /user/etlguest ...

OK Cancel

- 3 In the **Host** field of the Configuration window, enter the fully qualified name of the HiveServer2 on your Hadoop cluster.
- 4 In the **Port** field, enter the port number of the HiveServer2 on your Hadoop cluster.
- 5 In the **User ID** and **Password** fields, enter the name and password of the user account that is used to connect to the Hadoop cluster.

**Note:** The User ID and Password fields might not be editable, based the following:

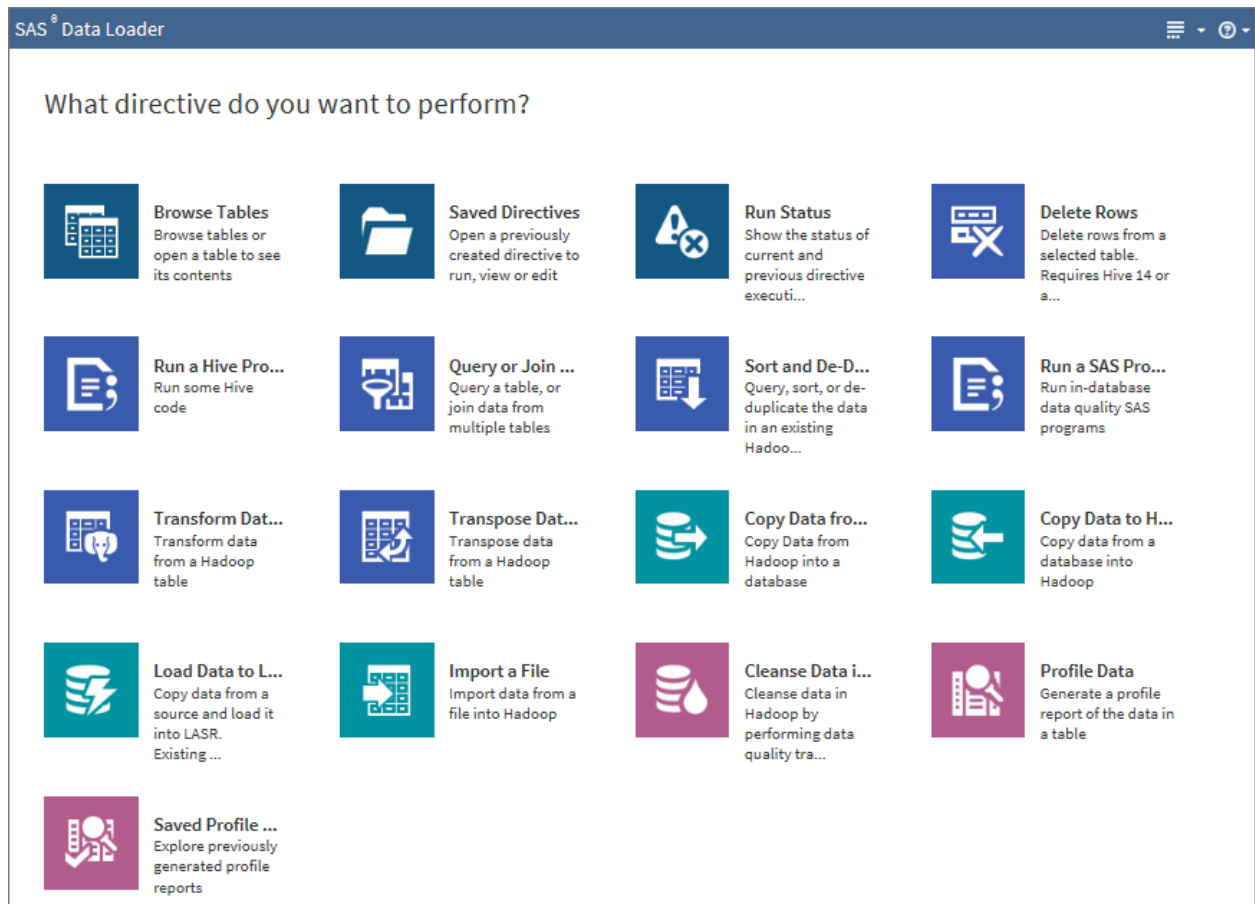
- If you configured to run SAS Data Loader in secure mode as described in [Step 7 on page 16](#), are not editable.
- If MapR is the Hadoop distribution, the **User ID** is not editable because the field is populated from a configuration file when you start the vApp. To change the User ID, you must enter a new value in the file `vApp-home\SASWorkspace\hadoop\conf\mapr-users.json`, and then restart the vApp. Finally, open the Hadoop Configuration panel and enter the new user ID.

- 6 In the **Oozie URL** field, enter the URL. The URL is similar to the following example: `http://host_name:11000/oozie`.

**Note:** The additional settings on the Configuration window, such as **Schema for temporary file storage**, are not required for deployment. For more

information about these settings, see *SAS Data Loader for Hadoop: User's Guide*.

- 7 Click **OK** to close the **Configuration** window and to access the main page of SAS Data Loader.



## Next Steps

Now that you have deployed SAS Data Loader for Hadoop, you can start exploring software features by using the sample data, or by using your own data.

- For a quick introduction on how to work with data, check out the video tutorials for [Using SAS Data Loader for Hadoop](#).
- Refer to the *SAS Data Loader for Hadoop: User's Guide* for information about using the directives.
- To use the directives **Copy Data to Hadoop** and **Copy Data from Hadoop**, you need to copy JDBC drivers from the Hadoop cluster to your shared folder. For more information, see *SAS Data Loader for Hadoop: User's Guide*.
- To use the directive **Load Data to LASR**, your SAS Administrator needs to install a grid of SAS LASR Analytic Servers. For more information, see *SAS Data Loader for Hadoop: User's Guide*.

- To learn about administrative tasks, see [Chapter 4, “Administration Tasks,”](#) on [page 25](#) for more information.



## 3

## Migrate from Version 2.2 to 2.3

<i>What Gets Migrated?</i> .....	21
<i>Steps for Migrating</i> .....	21

### What Gets Migrated?

Follow the steps in this section to keep and use existing content from SAS Data Loader 2.2 for Hadoop with your deployment of SAS Data Loader 2.3 for Hadoop.

You can migrate the following information so that it is available when you use SAS Data Loader 2.3 for Hadoop.

- Saved directives and profile reports
- Configuration settings
- History information about job runs
- User preferences
- User ID and other Hadoop configuration settings (for environments that do not use Kerberos authentication)

**Note:** Passwords are not migrated, so you must re-enter any passwords. Also, information that is entered in the SAS Data Loader: Information Center Settings panel is not migrated. You must re-enter the license information and, if the Hadoop cluster is secured by Kerberos, the Kerberos information.

### Steps for Migrating

**TIP** Before you start the vApp (release 2.3) for the first time, copy the appropriate directories from your previous shared folder to the new SASWorkspace shared folder for release 2.3. This is an important step for migrating successfully. If you do start the vApp before you copy the directories, you have to start the migration process over. For more information, see [“Troubleshooting Tips” on page 35](#).

Perform the following steps to migrate from a previous release.

- 1 Power off your vApp, if it is running. For more information, see [“Power Off the vApp” on page 27](#).
- 2 Go to [“Step 1: Download the Software for SAS Data Loader” on page 5](#), and perform the steps to download and unzip the required software from SAS onto your machine.
- 3 Go to [“Step 2: Copy Hadoop Configuration Files to Your Machine” on page 7](#), and perform the steps to set up the directory structure and save files that are required for release 2.3 and later.
- 4 Move contents from a previous release so that they are migrated to release 2.3.
  - a Using Windows Explorer, navigate to the shared folder that you set up for the previous release.

**TIP** To find out the directory path of the shared folder for a vApp, go to VMware Player Pro interface, and select **Edit virtual machine settings** for the vApp. In the Virtual Machine Settings window, select the **Options** tab, and then select **Shared Folders**. Look for the host path of the shared folder in the right panel.

- b Copy the folders that are in the shared folder for the previous release, and paste them to the SASWorkspace folder for release 2.3.

The following folders are required:

- **Configuration**
- **SASData**

**Note:** The `Configuration\HadoopConfig` subdirectory that was part of previous releases is not used in release 2.3.

Each of the following folders is optional. If you want to use the existing JDBC drivers and profiles, copy the following folders.

- **JDBCDrivers**
- **Profiles**

**Note:** For more information about JDBC drivers and profiles, see *SAS Data Loader for Hadoop: User's Guide*.

- 5 Go to [“Step 3: Configure the vApp in VMware Player Pro” on page 8](#), and perform the steps to save the required settings in VMware Player Pro.
- 6 Go to [“Step 4: Configure SAS Data Loader” on page 13](#), and perform the steps to apply the license file for release 2.3, and apply Kerberos settings, if your Hadoop cluster is configured to use Kerberos.
- 7 Go to [“Step 5: Complete the Hadoop Configuration” on page 17](#), and perform the steps to start SAS Data Loader 2.3 for Hadoop and complete the Hadoop configuration.
- 8 Perform the following steps to migrate all saved profiles.

**Note:** After you perform the following steps, you are not presented with this wizard again.

- a Click **Migrate Profiles**.

## Welcome to SAS® Data Loader for Hadoop

Welcome to SAS® Data Loader for Hadoop, where you can manage big data on your own terms with self-service data integration and data quality.

### Migrate SAS® Data Loader Profiles

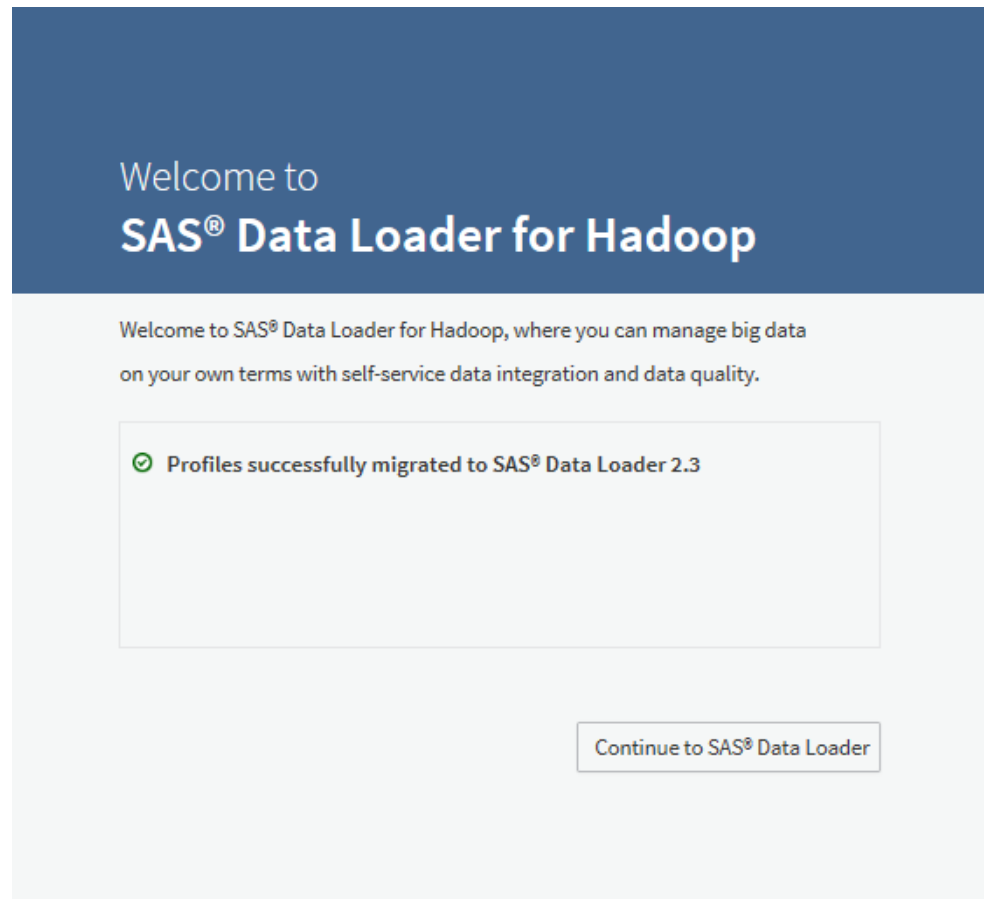
Profile content has been detected and may be migrated.

By migrating this content, all saved profiles will appear in SAS® Data Loader 2.3.

Migrate Profiles

Continue without Migrating

- b** Click **Continue to SAS Data Loader**.



After you complete the profiles migration, the SAS Data Loader for Hadoop interface is displayed.

- 9 Verify that saved information was migrated. Click **Saved Directives** or **Saved Profile Reports**.

**Note:**

- If you use the directives **Copy Data to Hadoop** and **Copy Data from Hadoop**, you must enter your password in the Database Configuration panel. For more information, see “Chapter 9, Maintaining SAS Data Loader” in *SAS Data Loader for Hadoop: User's Guide*.
- If you use the directive **Load Data to LASR**, you must enter your password in the LASR Server Configuration panel. For more information, see “Chapter 9, Maintaining SAS Data Loader” in *SAS Data Loader for Hadoop: User's Guide*.

## 4

## Administration Tasks

<i>Tips for Running the vApp</i> .....	25
<i>Play the vApp and Start SAS Data Loader</i> .....	26
<i>Power Off the vApp</i> .....	27
<i>Check for vApp Updates</i> .....	27
<i>Manage Your License</i> .....	28
Renew Your License .....	28
Request a Temporary License .....	28
<i>Back Up and Restore</i> .....	29
Tips for Configuring Backup and Restore .....	29
What Gets Backed Up? .....	29
Configure the Backup Location .....	29
Restore: Configure a New vApp to Use a Backup .....	30
<i>Configuring a New Version of Hadoop</i> .....	31
Before You Begin .....	31
General Steps for Configuring a New Version of Hadoop .....	31
<i>Enable Logging and Download Log Files</i> .....	32
Turn On Logging .....	32
Download Log Files .....	33

---

### Tips for Running the vApp

- You can close and open SAS Data Loader in a web browser without shutting down the vApp. The vApp continues to play until you shut it down in VMware Player Pro. If you close the browser tab for SAS Data Loader while the vApp is playing, any jobs on the Hadoop cluster continue to run. Also, their run status continues to be collected.
- After SAS Data Loader displays in your browser, you do not have to keep the browser tab for SAS Data Loader: Information Center open. While the vApp is still playing, you can close the tab for the Information Center at any time and still access SAS Data Loader.
- Do not close VMware Player Pro while the vApp is playing.
- Shut down SAS Data Loader for Hadoop as described in [“Power Off the vApp” on page 27](#).

**Note:** If you shut down vApp processes with Windows Task Manager, the vApp might not restart properly. If this happens, see “[Troubleshooting Tips](#)” on page 35 for more information.

- If the vApp is playing, and 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 the SAS Data Loader: Information Center. To restore your cursor, press Ctrl+Alt.
- You must restart the vApp after you configure a connection to a new database.
- You must restart the vApp after you change the version of Hadoop. A change can include a version update of an existing Hadoop vendor, or if you configure the vApp to use a different Hadoop vendor.

**CAUTION! If you suspend a guest, services might be interrupted** VMware Player Pro provides a capability to suspend a guest, which is the guest operating system that runs the vApp. In VMware Player Pro, do not select **Suspend Guest** or **Player ► Power ► Suspend Guest**. Suspending the vApp can interrupt communications between the SAS Data Loader web client and the Hadoop cluster. To resolve a suspended vApp, select **Player ► Power ► Restart Guest**.

---

## Play the vApp and Start SAS Data Loader

If you powered down the vApp, refer to the following instructions to restart the vApp and start SAS Data Loader for Hadoop.

- 1 Open VMware Player Pro.
- 2 Click **SAS Data Loader for Hadoop** and then, when it appears, click **Play virtual machine**.

VMware Player Pro requires a minute or two to play the vApp. When the vApp is ready, the VMware Player displays the message *Welcome to your SAS Data Loader Virtual Application*.

**Note:** If an informational Removable Devices window appears, review the information about removable devices and click **OK**.

- 3 In the window SAS Data Loader – VMware Player Pro, locate the HTTP address to connect to the SAS Data Loader.
- 4 Open a web browser, and enter the HTTP address in the browser’s address bar. Press the Enter key to continue. The SAS Data Loader: Information Center opens in a new tab in your browser.
- 5 In the SAS Data Loader: Information Center, click **Start SAS Data Loader**. SAS Data Loader opens in a new tab in your browser.

---

## Power Off the vApp

Perform the following steps to power off the vApp in VMware Player Pro:

- 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**.



**Note:** 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.

---

## Check for vApp Updates

Follow these steps to check for the availability of vApp software updates, and to download and install updates. If you have reasonable broadband capacity, vApp updates can take less than 15 minutes. When you update the vApp, you might also see an Information Center link to notes that describe the release's changes.

- 1 Open the browser for the SAS Data Loader: Information Center if it is not already open.
- 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 Player Pro.
- 8 When the SAS Data Loader: Information Center indicates that the vApp update is complete, click **Start SAS Data Loader**.

---


## Manage Your License

During installation, you selected a SAS installation data file (SID). The SID file contains your license, and it is delivered as an attachment to the Software Order Email. The license remains valid for a year after the receipt of the Software Order Email. Refer to this section for information about how to renew your license.

### Renew Your License

If you receive messages that your license is about to expire, perform the following steps to renew your license.

**Note:** Expiration messages are not displayed in the SAS Information Center if you use the Firefox web browser.

- 1 Contact your SAS Installation Representative to renew your license. When you renew the license, you receive a Renewal Order Email that contains a new SID file. Save the SID file from your Renewal Order Email to a directory on the computer that hosts the vApp.
- 2 If necessary, open SAS Data Loader: Information Center.
- 3 In the SAS Information Center, click the Settings icon  in the top right corner.
- 4 In the Settings window, click **Browse**, and then navigate to the directory that contains the new license file. Next, select the license file, and click **Open**.
- 5 In the Settings window, click **OK**.
- 6 In the Applying Settings Changes window, click **Yes** to confirm your selection.
- 7 To begin using the new license, simply open SAS Data Loader for Hadoop in a web browser. The new license remains valid for the time period that is specified in your Renewal Order Email.

### Request a Temporary License

If you need an emergency license, perform the following steps to download a temporary SID file that extends 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 can 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 [“Renew Your License” on page 28](#).



---

## Back Up and Restore

### Tips for Configuring Backup and Restore

Directives and other information from an existing vApp deployment can be saved, and then applied to a new vApp deployment.

Here are some tips to remember before you get started:

- To configure the vApp to back up directives, you add a new shared folder, named **backup**, to be the location where directives are stored. After this shared folder is added, data is saved to it each time a user clicks **Back Up Directives** in the SAS Data Loader interface.
- To restore directives, you configure a new vApp with a new shared folder that points to the location of the **backup** shared folder. An important step is to re-name the shared folder **restore** before you save it.
- The folder names **backup** and **restore** are case-sensitive, so enter the folder names exactly as shown.
- Only one backup at a time is supported. If a user initiates a backup, any previous backup information in the **backup** shared folder is overwritten.

### What Gets Backed Up?

The following information is saved each time a user initiates a backup.

- Saved directives and profile reports
- Configuration settings
- History information about job runs
- User ID and other Hadoop configuration settings (for environments that do not use Kerberos authentication)

**Note:** Passwords are not backed up, so you must re-enter any passwords. Also, information that is entered in the SAS Data Loader: Information Center Settings window is not backed up. You must re-enter the license information and if the Hadoop cluster is secured by Kerberos, the Kerberos information.

### Configure the Backup Location

Perform the following steps to specify the backup location for the vApp:


- 1 Create a directory on your computer to store the backup information.
- 2 In VMware Player Pro, select the virtual machine for SAS Data Loader for Hadoop, and then click **Edit virtual machine settings**.
- 3 Click the **Options** tab, and then select **Shared Folders** in the left panel.
- 4 In the **Folders** section, click **Add**.

- 5 Follow the steps in the Add Shared Folder Wizard to add the folder that you created in [Step 1 on page 29](#) as a shared folder.

Specify information for the following fields:

- **Host Path:** Enter the directory location for the folder or use the **Browse** button to select the directory.
- **Name:** Enter `backup` as the name of the shared folder. The folder name is case-sensitive, so enter the folder name exactly as shown.

- 6 Save your changes.

Each time you perform a backup, information is saved in the backup location that you specified. You initiate a backup in the SAS Data Loader user interface by clicking the **More** menu , and then selecting **Back Up**

**Directives.**

## Restore: Configure a New vApp to Use a Backup

Before you proceed with the restore process, ensure that you have completed these prerequisites:

- You have the required software and Hadoop configuration files for the vApp, as described in [Chapter 2, “Set Up and Run the vApp,” on page 3](#).
- A backup location has been defined, and a backup has been performed.

To restore your SAS Data Loader for Hadoop environment, complete the following steps:

- 1 Create a directory for the new vApp and add the software files, as described in [“Extract Files from the ZIP File” on page 6](#).
- 2 Create the shared folder for the vApp, as described in [“Step 2: Copy Hadoop Configuration Files to Your Machine” on page 7](#).
- 3 Configure the virtual machine, and add the shared directory for the vApp, as described in [“Step 3: Configure the vApp in VMware Player Pro” on page 8](#).
- 4 Before you close the Virtual Machine Settings window, add the restore location to the shared directories list.

**a** In the **Folders** section, click **Add**.

- b** Follow the steps in the Add Shared Folder Wizard to add the folder that you created in [“Configure the Backup Location” on page 29](#) as a shared folder.

Specify information for the following fields:

- **Host Path:** Enter the directory location for the backup folder or use the **Browse** button to select the directory.
- **Name:** Enter `restore` as the name of the shared folder.

**c** Save your changes.

- 5 Continue with the standard start-up process, as described in [“Step 4: Configure SAS Data Loader” on page 13](#).

**Note:** Settings for the Information Center are not included in the backup and restore process. You must re-enter any required configuration settings for the Information Center, such as the license file location and Kerberos settings for your Hadoop environment.

- 6 When you access SAS Data Loader for Hadoop, information from your previous environment, including saved directives and configuration settings, is available from the new vApp. Perform the following steps, as needed:
  - If you use the directives **Copy Data to Hadoop** and **Copy Data from Hadoop**, you must enter your password in the Database Configuration panel. For more information, see “Chapter 9, Maintaining SAS Data Loader” in *SAS Data Loader for Hadoop: User’s Guide*.
  - If you use the directive **Load Data to LASR**, you must enter your password in the LASR Server Configuration panel. For more information, see “Chapter 9, Maintaining SAS Data Loader” in *SAS Data Loader for Hadoop: User’s Guide*.

---

## Configuring a New Version of Hadoop

### Before You Begin

If you upgrade the version of your existing Hadoop vendor, or configure a new Hadoop vendor, you can configure SAS Data Loader for Hadoop to access it.


- Before you configure the vApp, your Hadoop administrator must configure the offering, SAS In-Database Technologies for Hadoop, on the Hadoop cluster. For more information, your Hadoop administrator can refer to *SAS In-Database Products: Administrator’s Guide*.
- Your Hadoop administrator must provide two folders that you will copy to your machine. The folders are `conf` and `lib`. Ask your Hadoop administrator to make these folders available.

### General Steps for Configuring a New Version of Hadoop

**Note:** If you have configured the vApp with a MapR distribution of Hadoop and want to configure a different Hadoop vendor, such as Cloudera CDH, do not follow the steps below. Instead, shut down the vApp as described in [“Power Off the vApp” on page 27](#), and then reinstall the vApp as described in [Chapter 2, “Set Up and Run the vApp,” on page 3](#).

Configuring a new version of Hadoop includes the following steps:

- 1 Shut down the vApp as described in [“Power Off the vApp” on page 27](#).
- 2 Copy the two folders provided by your Hadoop administrator into the `SASWorkspace\hadoop` directory. The two folders, named `conf` and `lib`, contain the required files to configure the vApp to use the Hadoop cluster. When you copy these files, choose to overwrite the existing `conf` and `lib` folders.

- 3 Power on the vApp as described in “Play the vApp and Start SAS Data Loader” on page 26, and perform the following steps for the new version of Hadoop:
  - a After you start the Information Center, enter new Kerberos settings, if necessary.
  - b After you start SAS Data Loader, click the **More** menu , and then select **Configuration**. In the Configuration window, expand **Hadoop Configuration** and update values as needed.


---

## Enable Logging and Download Log Files

For debugging purposes, you can enable logging. SAS recommends that you enable logging only when directed to do so by your SAS Technical Support representative. To maintain performance, logging is not recommended under normal circumstances.

### Turn On Logging

Perform the following steps to enable or disable logging inside the vApp.

- 1 Check the **Run Status** 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 Data Loader: Information Center.
- 3 Click the Settings icon .
- 4 To activate logging, select the **Turn logging on (for debugging only)** check box. To deactivate logging, make sure the check box is not selected. Click **OK**.
- 5 In the Applying Settings Changes window, click **Yes** to confirm your selection.

#### Note:

- Log files are stored in the following location: `vApp-path\vApp-instance\SASWorkspace\Logs`.
- 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 Server generates the log file `SASApp_WorkspaceServer_YYYY-MM-DD_localhost_PID.log`.

## Download Log Files

Perform the following steps to download and view log files generated by the vApp.

- 1 Open the browser for the SAS Data Loader: Information Center if it is not already open.
- 2 Click the **Help** icon in the top right, and then click **Download log file**.
- 3 Look for the prompt that enables you to open or save the file that contains the log files. Choose to **Open**.
- 4 Unzip the file to access the logs.



## 5

## Troubleshooting

<i>Support Community</i> .....	35
<i>Troubleshooting Tips</i> .....	35

---

### Support Community

If you need additional help with deploying or administering the SAS Data Loader for Hadoop, the support community is a great place to find answers.

[SAS Data Loader for Hadoop Community](#)

Join the community to ask questions and receive expert online support.

---

### Troubleshooting Tips

If something is not working, refer to the following table for tips and solutions.

Issue	Tips
When I run the vApp and open the Information Center, an error states that the system could not find the required Hadoop files.	<p>Be sure to create a shared folder named <code>SASWorkspace</code>, and copy the <code>conf</code> and <code>lib</code> to <code>SASWorkspace\hadoop</code>. Next, exit the Information Center and restart the vApp to continue the process.</p> <p><b>Note:</b> The folder name <code>SASWorkspace</code> is case-sensitive, so enter the folder name exactly as shown.</p>

---

Issue	Tips
vApp fails to start in VMware Player Pro, and an error message states that Intel VT-x or AMD-v is not available.	<p>This message indicates that running a virtual machine (virtualization) is not supported or configured in your firmware. To resolve this issue, run a utility to configure your machine.</p> <p>First, determine whether your machine has an Intel or AMD processor:</p> <ol style="list-style-type: none"> <li>1 Press the Windows key and the R key on your keyboard at the same time. The Run dialog box appears.</li> <li>2 In the <b>Open</b> field of the dialog box, enter <b>msinfo32</b> and click <b>OK</b>.</li> <li>3 In the System Information window, ensure that <b>System Summary</b> is selected in the left panel.</li> <li>4 In the right panel, find <b>System Type</b> and ensure that you have a 64-bit computer. Next, find <b>Processor</b> to determine the processor type, Intel, or AMD.</li> <li>5 Download one of the following:</li> </ol> <p>Next, download, and run the utility for your machine processor:</p> <ul style="list-style-type: none"> <li>■ Download the <a href="#">Intel tool</a>.</li> <li>■ Download the <a href="#">AMD tool</a>.</li> </ul> <p>Finally, visit the <a href="#">virtualization hardware extensions</a> page to enable Intel and AMD virtualization hardware extensions. To obtain information about how to navigate through your specific BIOS, contact the support site for the manufacturer of your computer.</p> <p><b>Note:</b> For additional information about virtualization support, refer to the <a href="#">VMware Knowledge Base</a>.</p>
Migration from a previous version is not working.	<p>Before you start the vApp (release 2.3) for the first time, copy the appropriate directories from your previous shared folder to the new SASWorkspace shared folder for release 2.3.</p> <p>If you encounter this problem, try the following steps:</p> <ol style="list-style-type: none"> <li>1 Shut down the vApp (release 2.3) as described in <a href="#">“Power Off the vApp” on page 27</a>.</li> <li>2 Remove the folder where you unzipped the files for release 2.3.</li> <li>3 Continue with the migration, as described in <a href="#">“Steps for Migrating” on page 21</a>.</li> </ol>
vApp does not restart properly after you end vApp processes with Windows Task Manager.	<p>If vApp processes were shut down with Windows Task Manager, the vApp might not restart and the following message might be displayed: <i>Service Temporarily Unavailable</i>.</p> <p>If you encounter this problem, try the following steps:</p> <ol style="list-style-type: none"> <li>1 Shut down the vApp as described in <a href="#">“Power Off the vApp” on page 27</a>.</li> <li>2 Power on the vApp as described in <a href="#">“Play the vApp and Start SAS Data Loader” on page 26</a>.</li> </ol>



Issue	Tips
Single sign-on is not working in my browser.	<p>SAS Data Loader for Hadoop supports the Firefox and Chrome browsers for single sign-on. The browser must be configured to support Integrated Windows Authentication (IWA).</p> <p>Contact your system or Hadoop administrator. For more information, refer to <i>SAS In-Database Products: Administrator's Guide</i>.</p>



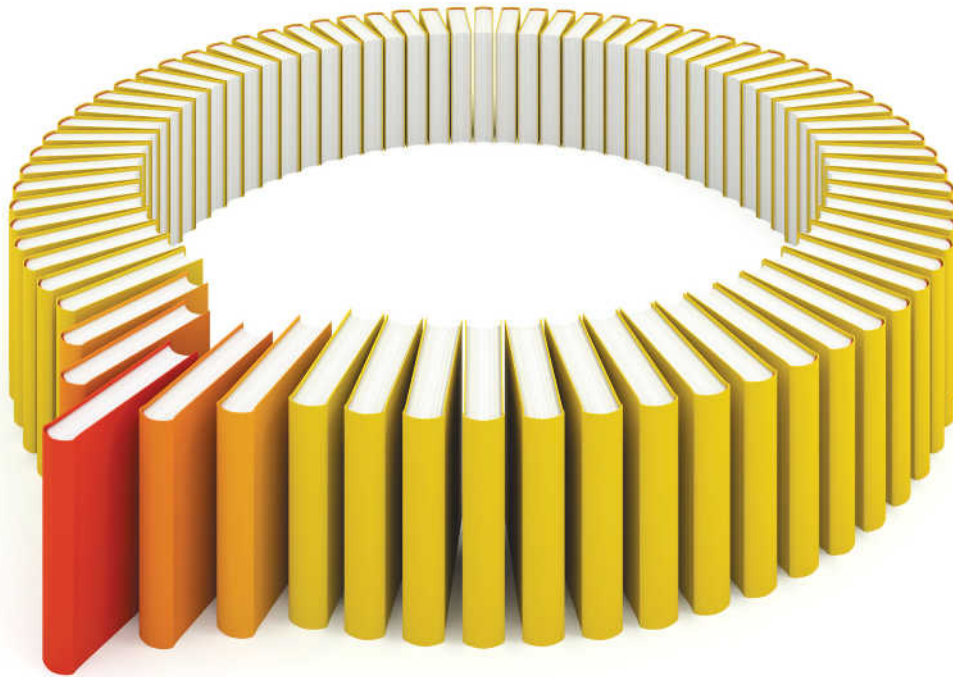
## Recommended Reading

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

For a complete list of SAS publications, go to [sas.com/store/books](http://sas.com/store/books). If you have questions about which titles you need, please contact a SAS Representative:

SAS Books  
SAS Campus Drive  
Cary, NC 27513-2414  
Phone: 1-800-727-0025  
Fax: 1-919-677-4444  
Email: [sasbook@sas.com](mailto:sasbook@sas.com)  
Web address: [sas.com/store/books](http://sas.com/store/books)





# Gain Greater Insight into Your SAS<sup>®</sup> Software with SAS Books.

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



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

