



THE
POWER
TO KNOW.

SAS[®] Data Loader 2.1 for Hadoop

Installation and Configuration Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2014. *SAS® Data Loader 2.1: Installation and Configuration Guide*. Cary, NC: SAS Institute Inc.

SAS® Data Loader 2.1: Installation and Configuration Guide

Copyright © 2014, 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.

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, August 2014

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

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/>.

Contents

Chapter 1 • Introduction	1
Installing and Configuring SAS Data Loader for Hadoop	1
Requirements	1
Chapter 2 • Installing SAS Data Loader for Hadoop	3
Overview	3
Instructions for Microsoft Windows Users	4
Final Configuration	8
Configure SAS Data Loader to Access a Grid of SAS LASR Analytic Servers (Optional)	13
Chapter 3 • Configuring Hadoop	15
Introduction	15
In-Database Deployment Package for Hadoop	15
Hadoop Installation and Configuration	17
SASEP-SERVERS.SH Script	22
Hadoop Permissions	29
Appendix 1 • Hardware Virtualization	31
Recommended Reading	33
Index	35

Introduction

<i>Installing and Configuring SAS Data Loader for Hadoop</i>	1
<i>Requirements</i>	1

Installing and Configuring SAS Data Loader for Hadoop

SAS Data Loader for Hadoop has been designed to make installation and configuration simple. The web client software is installed as a vApp, which runs in a virtual machine that you download separately. Installation of the vApp is as simple as uncompressing a file and configuring the virtual machine. Any files that are required by the vApp are stored in a single shared folder on your client device. To upgrade to a new version, you simply replace the vApp. The shared folder of the previous vApp is available for the next version of the vApp with minimal migration.

Your Hadoop administrator must configure the SAS Embedded Process for Hadoop and provide you with a few files to copy onto your local device.

Here are the contents of this guide:

- [Chapter 2, “Installing SAS Data Loader for Hadoop,” on page 3](#). This chapter provides all the information that you need to install and configure SAS Data Loader for Hadoop.
- [Chapter 3, “Configuring Hadoop,” on page 15](#). This chapter is only for Hadoop administrators and contains information for configuring the SAS Embedded Process for Hadoop.

Requirements

The following are system requirements for installing and configuring SAS Data Loader for Hadoop:

- The SAS Data Loader for Hadoop compressed file downloaded to your software depot.
- A Microsoft Windows 7 64-bit operating system. This system must be capable of supporting a 64-bit virtual image. See [Hardware and Firmware Requirements](#) on the VMware website.

Note: When starting SAS Data Loader for Hadoop, if an error occurs stating that VT-x or AMD-v is not available, see [Appendix 1, “Hardware Virtualization,” on page 31](#).

- VMware Player Plus version 6.0+ for Windows. You can download VMware Player Plus 6.0 from: www.vmware.com.

Note: VMware Inc. provides VMware Player Plus for commercial applications and VMware Player, a free version, for non-commercial applications. See the website to ensure that you download the version that is appropriate for your site. SAS Data Loader for Hadoop fully supports both versions.

- Cloudera 5.0 or Hortonworks 2.0

Note: Both Hive 2 and YARN (MapReduce 2) are required. MapReduce 1 is not supported.

- One of the following web browsers:

- ☐ Microsoft Internet Explorer 9+
- ☐ Mozilla Firefox 14+
- ☐ Google Chrome 21+

- The SAS Data Loader for Hadoop virtual image is configured to use 8 GB of RAM and 2 processors.

- ☐ You can increase the RAM assigned to the SAS Data Loader for Hadoop virtual image, but do not allocate all memory to the virtual machine because it will have an impact on the operating system and other applications.
- ☐ You cannot increase the number of processors assigned to the SAS Data Loader for Hadoop virtual image.

- If you intend to upload data to SAS LASR Analytic Servers, you must first license, install, and configure a grid of SAS LASR Analytic Servers, version 6.3. See the *SAS Data Loader for Hadoop: User's Guide* for detailed information.

- ☐ The SAS LASR Analytic Servers must be registered on a SAS Metadata Server.
- ☐ SAS Visual Analytics 6.4 must be installed and configured on the SAS LASR Analytic Servers.
- ☐ When the grid of SAS LASR Analytic Servers is operational, you must generate and deploy Secure Shell (SSH) keys for SAS Data Loader. See [“Configure SAS Data Loader to Access a Grid of SAS LASR Analytic Servers \(Optional\)” on page 13](#) for more information.
- ☐ You must specify SAS LASR Analytic Server connection information in SAS Data Loader. See [Step 15 on page 11](#) for more information.
- ☐ The SAS LASR Analytic Servers must have memory and disk allocations that are large enough to accept Hadoop tables.

2

Installing SAS Data Loader for Hadoop

Overview	3
Instructions for Microsoft Windows Users	4
Unzipping the SAS Data Loader vApp	4
Configuring VMware Player Plus	4
Final Configuration	8
Configure SAS Data Loader to Access a Grid of SAS LASR Analytic Servers (Optional)	13

Overview

These instructions assume that you have downloaded the SAS Data Loader for Hadoop compressed file to your software depot as described in your welcome letter from SAS.

Regardless of the platform, the general instructions for installing and configuring SAS Data Loader for Hadoop are:

- 1 Unzip the SAS Data Loader for Hadoop compressed file.
- 2 Configure VMware Player Plus.
- 3 Verify with your Hadoop Administrator that your Hadoop system is properly configured. See [Chapter 3, “Configuring Hadoop,” on page 15](#) for more information.
- 4 Start SAS Data Loader for Hadoop in VMware Player Plus to finish the initial configuration.
- 5 If you intend to upload data to a SAS LASR Analytic Server grid, configure Secure Shell (SSH) keys for SAS Data Loader. See [“Configure SAS Data Loader to Access a Grid of SAS LASR Analytic Servers \(Optional\)” on page 13](#)

Instructions for Microsoft Windows Users

Unzipping the SAS Data Loader vApp

To unzip the SAS Data Loader for Hadoop vApp ZIP file:

- 1 Navigate to the SAS Data Loader for Hadoop vApp ZIP file in the following location of your SAS Software Depot: `\SAS Software Depot\SAS_Data_Loader_for_Hadoop\2_1\VMwarePlayer`.
- 2 Do one of the following:
 - a If WinZip is installed:
 - a Right-click the SAS Data Loader for Hadoop ZIP file and select **Open with WinZip**.
 - b In the WinZip application, click **Unzip** to unzip the compressed files to the current location of the zipped file.
 - b If WinZip is not installed, right-click the SAS Data Loader for Hadoop ZIP file and select **Extract All** to unzip the compressed files to the current location of the zipped file.

Wait for the files to expand before you continue.

Configuring VMware Player Plus

Overview

You must configure VMware Player to create a shared folder for data that is to be available both to the SAS Data Loader for Hadoop virtual image and to your host system.

Opening a Virtual Machine

To open a virtual machine:

- 1 Launch VMware Player Plus.
- 2 Click **Open a Virtual Machine**.
- 3 In the file browser window, navigate to the uncompressed SAS Data Loader for Hadoop virtual (.vmx) image.
- 4 Select the SAS Data Loader for Hadoop virtual image, and then click **Open**.

Sharing a Folder

You must use a virtual machine shared folder to enable SAS Data Loader for Hadoop to function properly. With a shared folder, you can easily share files among virtual machines and the host computer.

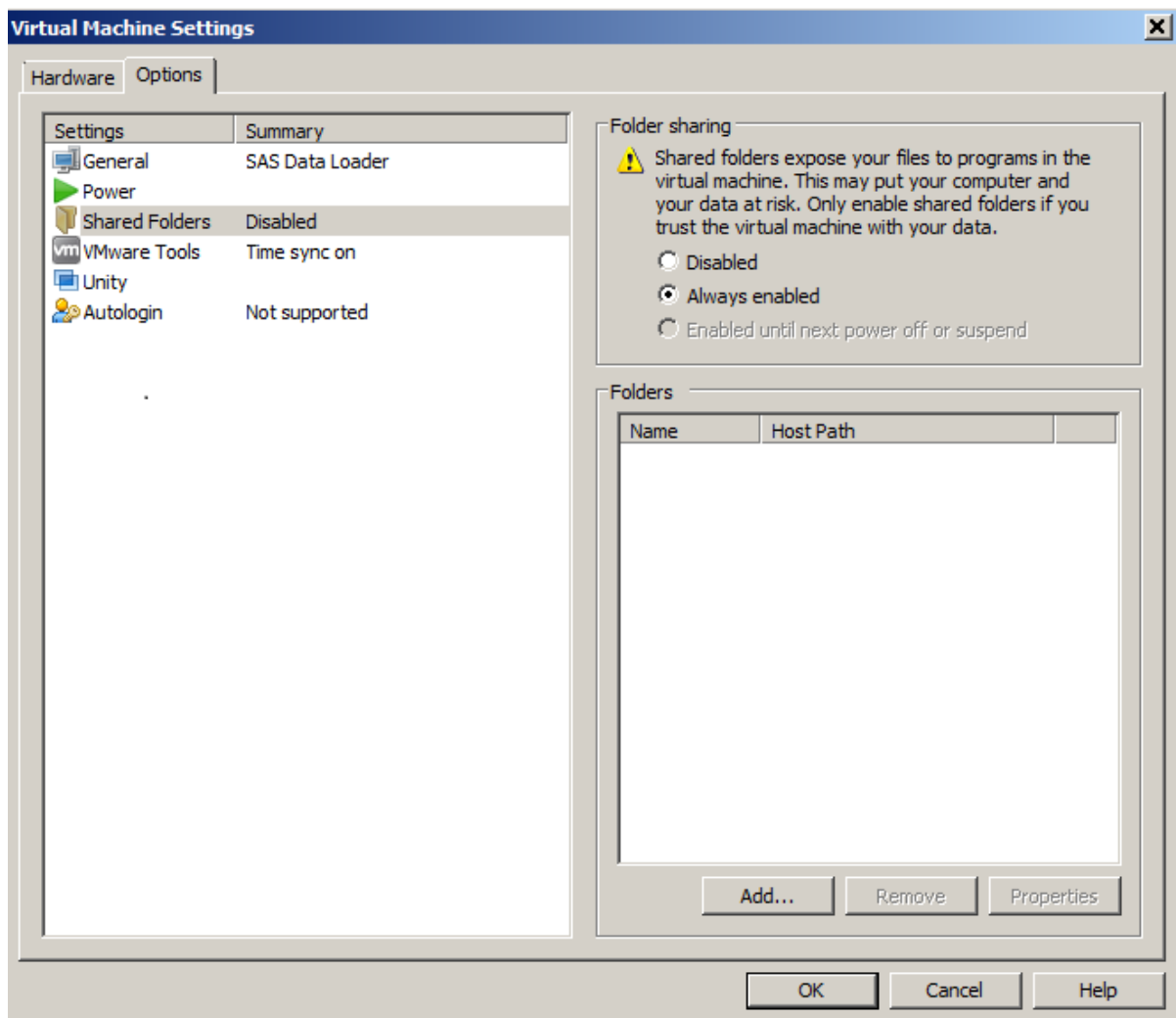
Note:

- You must have access permissions to add a network folder.
- Do not include a backslash (\) in the network folder name.
- The shared folder name is case-sensitive.

To share a folder from the virtual image to the host system:

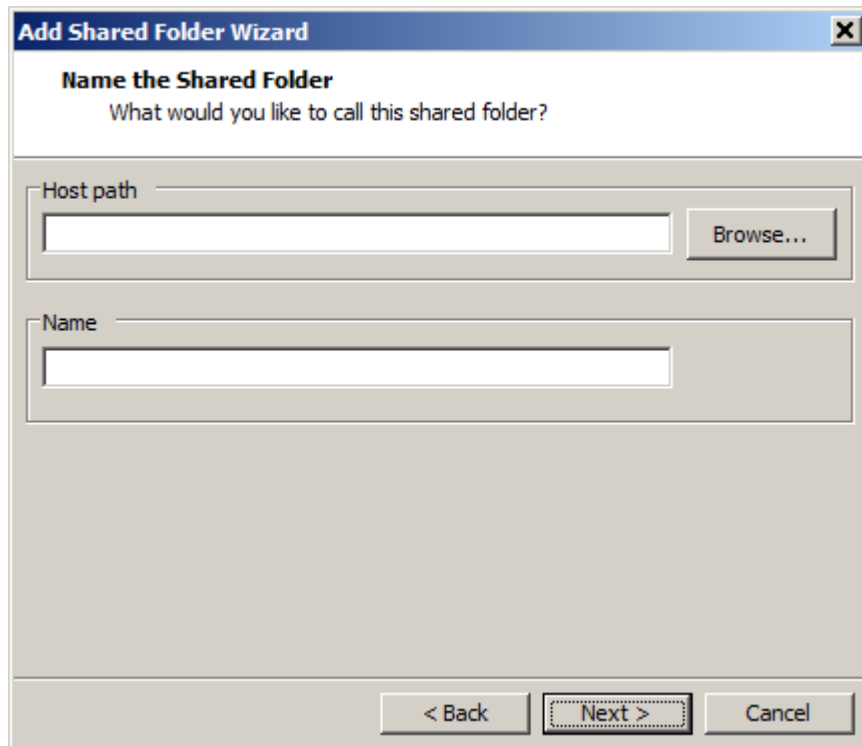
- 1 Click **Edit virtual machine settings**.
- 2 Select the **Options** tab.
- 3 Select **Shared Folders**, and then click **Always Enabled**.

Figure 2.1 Options



- 4 Click **Add** to open the Add Shared Folder Wizard window.
- 5 Click **Next** to open the Named the Shared Folder dialog box .

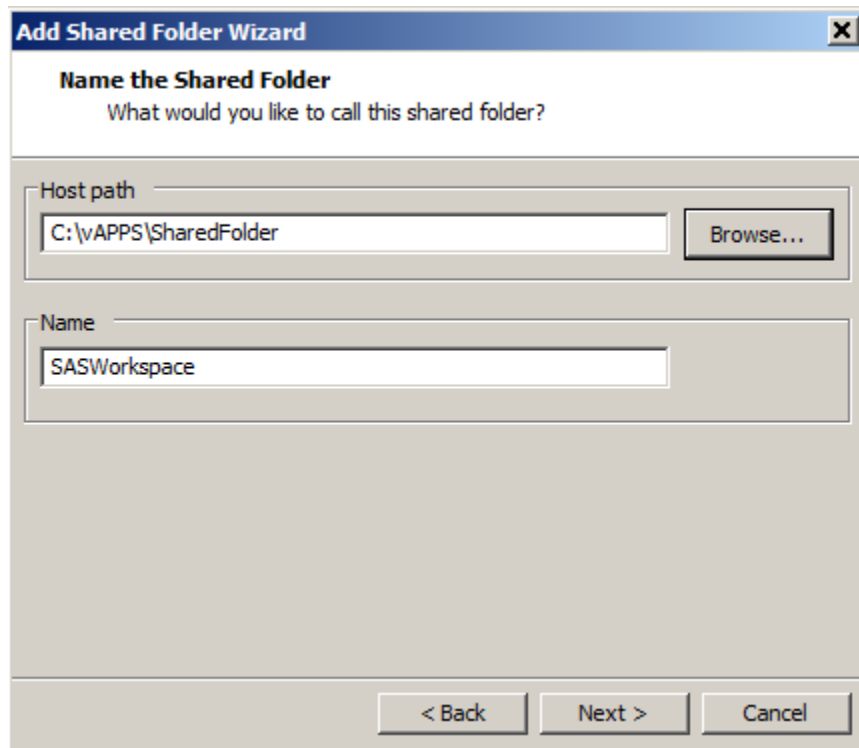
Figure 2.2 Add Shared Folder Wizard

The image shows a Windows-style dialog box titled "Add Shared Folder Wizard" with a close button (X) in the top right corner. The main heading is "Name the Shared Folder" with a subtitle "What would you like to call this shared folder?". Below this, there are two input sections. The first is labeled "Host path" and contains a text input field followed by a "Browse..." button. The second is labeled "Name" and contains a text input field. At the bottom of the dialog, there are three buttons: "< Back", "Next >" (which is highlighted with a dashed border), and "Cancel".

- 6 Click **Browse** to open the Browse For Folder dialog box.
- 7 In the Browse For Folder dialog box, choose a host path for the shared folder. The folder can be created anywhere. For example, creating it inside the folder where you have downloaded the SAS Data Loader for Hadoop vApp would group it with related files.
- 8 Click **Make New Folder**, and then enter the name SharedFolder. Click **OK** to return to the Named the Shared Folder dialog box.
- 9 Enter SASWorkspace (not any other name) for the shared folder name, and then click **Next**.

Note: The shared folder name is case-sensitive and must be entered exactly as described.

Figure 2.3 Shared Folder Name



10 Click **Finish**.

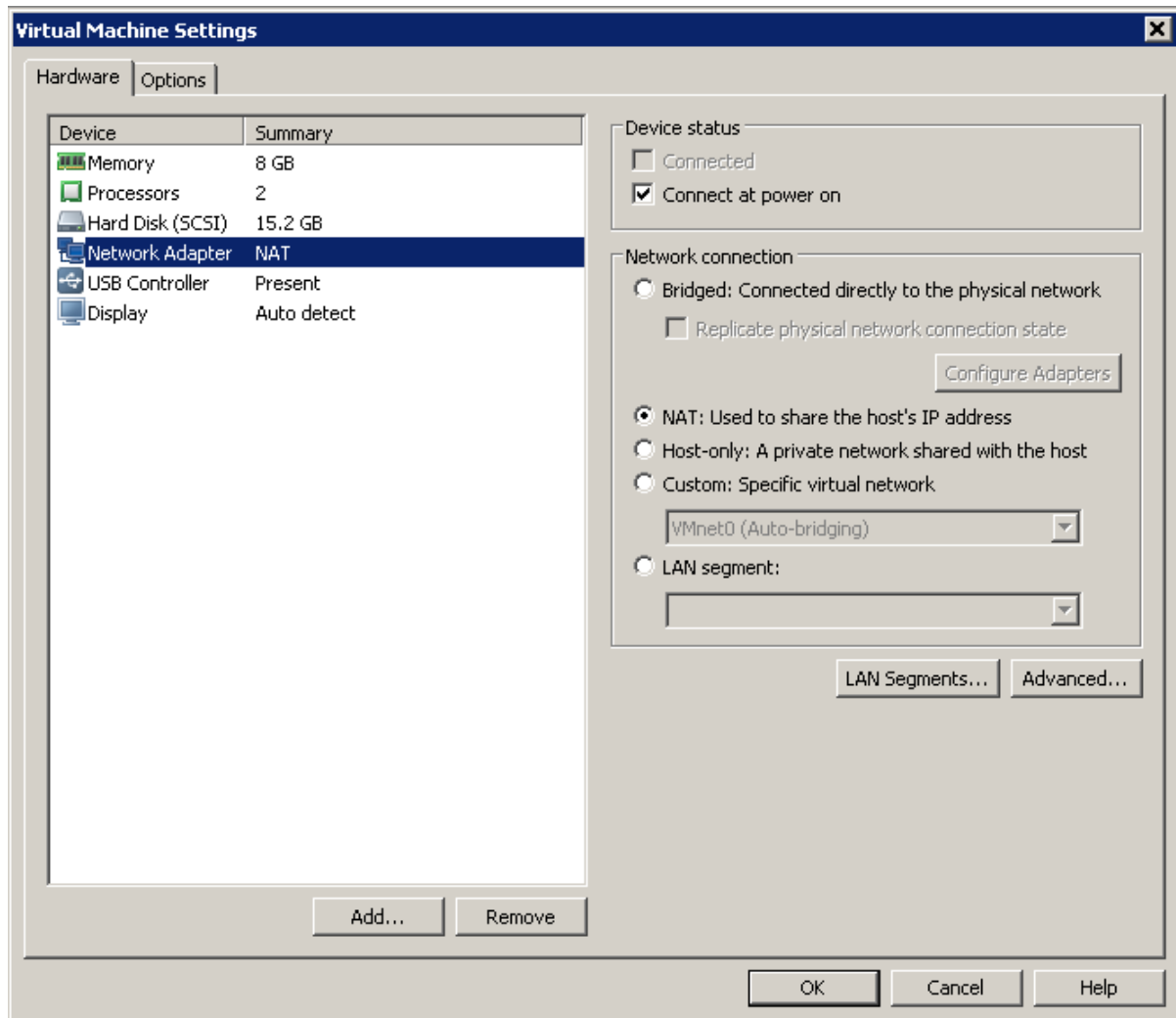
11 Click **OK** to close the Virtual Machine Settings dialog box.

Setting the Network Adapter

By default, the SAS Data Loader for Hadoop virtual image network adapter is set to NAT. You must use this value. Confirm that the network adapter is set to NAT by performing the following steps:

- 1 Click **Edit virtual machine settings**.
- 2 Select the **Hardware** tab.
- 3 Select **Network Adapter**.

Figure 2.4 Hardware



- 4 Select **NAT: Used to share the host's IP address..**
- 5 Click **OK**.

Final Configuration

Follow these steps to finalize your SAS Data Loader for Hadoop configuration:

- 1 Launch VMware Player Plus.
- 2 Select SAS Data Loader for Hadoop, and then click **Play virtual machine**.

Note: When starting SAS Data Loader for Hadoop, if an error occurs stating that VT-x or AMD-v is not available, see [Appendix 1, "Hardware Virtualization,"](#) on page 31.

- 3 The VMware Player displays a window listing the SAS Data Loader for Hadoop URL.

Note: If you click inside the VMware Player window, your cursor is disabled. Enter the appropriate keystrokes as described in the window to restore your cursor.

- 4 Open a web browser.
- 5 Type in the URL displayed in the VMware Player Plus window into the browser address bar, and then press the Enter key to display the SAS Data Loader for Hadoop Information Center in the browser.

Note: You cannot copy the URL from the VMware Player Plus window.

- 6 The SAS Data Loader for Hadoop Information Center displays the Settings dialog box.

Figure 2.5 Settings

Note: See the *SAS Data Loader for Hadoop: Installation and Configuration Guide* for information about setting **Advanced** options.

Select the version of Hadoop that is used on your cluster.

- 7 Your software order e-mail provided you with a SAS installation data (SID) file to be downloaded to your local drive. Click **Browse** to locate and select this SID file, and then click **OK**. Your configuration is then updated, including the addition of the following folders to your shared folder:

- **Configuration**

- ☐ Contains sasdemo.pub, an ssh key file that must be moved to your SAS LASR Analytic Server if you want to upload data to the SAS LASR Analytic Server.

- **Configuration\DMServices**

- ☐ Contains an empty version of the configuration database. SAS Data Loader for Hadoop, when starting for the first time, creates default content for this database.
 - ☐ Contains Saved Directives and SAS Data Loader for Hadoop configuration information.

- **Configuration\HadoopConfig**

- ☐ Location into which Hadoop client configuration files are copied.

- **InClusterBundle**

- ☐ Contains the two self-extracting files (.sh) that must be run inside the Hadoop cluster.
 - ☐ Contains JAR files for the QKB Pack Tool and QKB Push Tool.

■ **Profiles**

- Location in which SAS Data Loader for Hadoop stores its profile reports.

■ **Logs**

- Location into which log files are written if you have enabled debugging.

- 8 The SAS Data Loader for Hadoop Information Center reloads (this might take several minutes) and displays a message instructing you to copy Hadoop configuration files to your shared folder. Click **Close**.

- 9 Contact your Hadoop Administrator who can provide you with the Hadoop cluster configuration files. You must place these files in your shared folder.

Your Hadoop administrator configures the Hadoop cluster that you use. Consult with your Hadoop administrator about how your particular Hadoop cluster is configured.

To connect to a Hadoop server, the following configuration files must be copied from the Hadoop cluster to `SharedFolder\Configuration\HadoopConfig`:

core-site.xml
hdfs-site.xml
hive-site.xml
mapred-site.xml
yarn-site.xml


Note: For a MapReduce 2 and YARN cluster, both the mapred-site.xml and yarn-site.xml files are needed.

- 10 Click **Start SAS Data Loader** to open SAS Data Loader in a new browser tab. The Configuration dialog box is displayed:




Figure 2.6 Configuration

The Configuration dialog box is divided into two main sections: Hadoop and LASR Analytic Servers.

Hadoop Section:

- Host:** A text input field.
- Port:** A text input field containing the value "10000".
- User ID:** A text input field.
- Schema for temporary file storage:**
 - ☒ Use the system's default schema
 - ☐ Specify another schema name: 

LASR Analytic Servers Section:

Name	Host	Port	
			  

At the bottom of the dialog are **OK** and **Cancel** buttons.


- 11** Enter the fully qualified host name of the Hadoop cluster to which you want to connect.
- 12** Enter the port of the Hadoop cluster to which you want to connect.
- 13** Enter the User ID for the Hadoop cluster to which you want to connect.
- 14** By default, the schema for temporary storage is the HIVE default schema on your cluster. You can select an alternative schema, but it must exist on the cluster.
- 15** To add a SAS LASR Analytic Server to which data can be uploaded, click  to open the LASR Server Configuration dialog box:

Figure 2.7 LASR Server Configuration

LASR Server Configuration

LASR Analytic Server Configuration

Name:

Description:

Host:

Port:

LASR authorization service location:

Metadata Configuration

Connection Profile

Host:

Port:

User ID:

Password:

Default Locations

Repository:

SAS folder for tables:

Library location:

LASR server tag:

OK Cancel

16 In the LASR Analytic Server Configuration section:

- a** Enter the server name and description in the **Name** and **Description** fields.
- b** In the **Host** field, enter the full network name of the host of the SAS LASR Analytic Server. A typical name is similar to saslasr03.us.ourco.com.
- c** In the **Port** field, enter the number of the port that the SAS LASR Analytic Server uses to listen to connections from SAS Data Loader. The default value is 10010.
- d** In the field **LASR authorization service location**, enter the HTTP address of the authorization service that is used by the SAS LASR Analytic Server to control access to services and data sources.

17 In the Metadata Configuration section:

- a** In the **Host** field, add the network name of the SAS Metadata Server that is accessed by the SAS LASR Analytic Server.
- b** In the **Port** field, add the number of the port that the SAS Metadata Server uses to listen for client connections. The default value of 8561 is normally left unchanged.
- c** In the **User ID** and **Password** fields, add the credentials that SAS Data Loader uses to connect to the SAS Metadata Server. These values are stored in encrypted form on disk.

Note: The Default Locations area specifies where tables are stored on the SAS LASR Analytic Server. You might need to obtain these values from your SAS administrator. The default location is also used to

determine where to register data table information in the SAS Metadata Server associated with the SAS LASR Analytic Server environment.

- d** In the **Repository** field, specify the name of the SAS Metadata Server repository on the SAS LASR Analytic Server that receives downloads from Hadoop. The default value is Foundation.
- e** In the **SAS folder for tables** fields, specify the path inside the repository that contains downloads from Hadoop. This is the location for registering SAS LASR Analytic Server tables in the SAS Metadata Server repository. The default value is /SharedData.
- f** In the **Library location** field, add the name of the SAS library that is referenced by SAS Data Loader for Hadoop.
- g** In the **SAS LASR Analytic Server tag** field, add the name of the tag that is associated with each table that is downloaded from Hadoop. The tag is required. It is used along with the table name as a unique identifier for tables that are downloaded from Hadoop.
- h** Click **OK** to return to the Configuration dialog box

18 Click **OK**. SAS Data Loader for Hadoop is displayed.

See the *SAS Data Loader for Hadoop: Installation and Configuration Guide* for detailed information about using SAS Data Loader for Hadoop. To close SAS Data Loader for Hadoop, simply close the browser tab in which the program is running.

Configure SAS Data Loader to Access a Grid of SAS LASR Analytic Servers (Optional)

The following procedure is required only if you intend to upload data to an existing SAS LASR Analytic Server grid. This procedure configures Secure Shell (SSH) keys for SAS Data Loader on your grid of SAS LASR Analytic Servers.

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

- 1** On the SAS LASR Analytic Server grid, create the user `sasdlldr1`, as described in the *SAS LASR Analytic Server: Administrator's Guide*.
- 2** Generate a public key and a private key for `sasdlldr1` and install those keys.
- 3** Copy the public key file from SAS Data Loader at `vApp-install-path\%vApp-instance%\Shared Folder\Configuration\sasdemo.pub`. Append the SAS Data Loader public key to the file `~sasdlldr1/.ssh/authorized_keys` on the head node of the grid.

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

3

Configuring Hadoop

Introduction	15
In-Database Deployment Package for Hadoop	15
Prerequisites	15
Overview of the In-Database Deployment Package for Hadoop	16
Hadoop Installation and Configuration	17
Hadoop Installation and Configuration Steps	17
Upgrading from or Reinstalling a Previous Version	17
Moving the SAS Embedded Process and SAS Hadoop	
MapReduce JAR File Install Scripts	19
Installing the SAS Embedded Process and SAS Hadoop	
MapReduce JAR Files	19
SASEP-SERVERS.SH Script	22
Overview of the SASEP-SERVERS.SH Script	22
SASEP-SERVERS.SH Syntax	23
Starting the SAS Embedded Process	27
Stopping the SAS Embedded Process	28
Determining the Status of the SAS Embedded Process	28
Hadoop Permissions	29

Introduction

Configuring the in-database deployment package for Hadoop is to be undertaken only by the Hadoop administrator and needs to be done only once for each Hadoop cluster. The Hadoop administrator will provide you with files to be copied onto your local device, as described in [Step 9 on page 10](#)

In-Database Deployment Package for Hadoop

Prerequisites

The following are required before you install and configure the in-database deployment package for Hadoop:

- You have working knowledge of the Hadoop vendor distribution that you are using.

You also need working knowledge of the Hadoop Distributed File System (HDFS), MapReduce 2, YARN, Hive, and HiveServer2 services. For more information, see the [Apache website](#) or the vendor's website.

- The HDFS, MapReduce, YARN, and Hive services must be running on the Hadoop cluster.
- You have root or sudo access. Your user name has Write permission to the root of HDFS.
- You know the location of the MapReduce home.
- You know the host name of the Hive server and the NameNode.
- You understand and can verify your Hadoop user authentication.
- You understand and can verify your security setup.
- You have permission to restart the Hadoop MapReduce service.
- In order to avoid SSH key mismatches during installation, add the following two options to the SSH `config` file, under the user's home `.ssh` folder. An example of a home `.ssh` folder is `/root/.ssh/`. `nodes` is a list of nodes separated by a space.

```
host nodes
    StrictHostKeyChecking no
    UserKnownHostsFile /dev/null
```

For more details about the SSH `config` file, see the SSH documentation.

- All machines in the cluster are set up to communicate with passwordless SSH. Verify that the nodes can access the node that you chose to be the master node by using SSH.

SSH keys can be generated with the following example.

```
[root@raincloud1 .ssh]# ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /root/.ssh/id_rsa.
Your public key has been saved in /root/.ssh/id_rsa.pub.
The key fingerprint is:
09:f3:d7:15:57:8a:dd:9c:df:e5:e8:1d:e7:ab:67:86 root@raincloud1
```

```
add id_rsa.pub public key from each node to the master node authorized
key file under /root/.ssh/authorized_keys
```

Overview of the In-Database Deployment Package for Hadoop

This section describes how to install and configure the in-database deployment package for Hadoop (SAS Embedded Process).

The in-database deployment package for Hadoop must be installed and configured before you can perform the following tasks:

- Run a scoring model in Hadoop Distributed File System (HDFS).

- Transform data in Hadoop and extract transformed data out of Hadoop for analysis.

The in-database deployment package for Hadoop includes the SAS Embedded Process and two SAS Hadoop MapReduce JAR files. The SAS Embedded Process is a SAS server process that runs within Hadoop to read and write data. The SAS Embedded Process contains macros, run-time libraries, and other software that is installed on your Hadoop system.

The SAS Embedded Process must be installed on all nodes capable of executing MapReduce 2 and YARN tasks. The SAS Hadoop MapReduce JAR files must be installed on all nodes of a Hadoop cluster.

Hadoop Installation and Configuration

Hadoop Installation and Configuration Steps

- 1 If you are upgrading from or reinstalling a previous release, follow the instructions in [“Upgrading from or Reinstalling a Previous Version” on page 17](#) before installing the in-database deployment package.

- 2 Move the SAS Embedded Process and SAS Hadoop MapReduce JAR file install scripts to the Hadoop master node.

For more information, see [“Moving the SAS Embedded Process and SAS Hadoop MapReduce JAR File Install Scripts” on page 19](#).

Note: Both the SAS Embedded Process install script and the SAS Hadoop MapReduce JAR file install script must be transferred to the same directory.

Note: The location where you transfer the install scripts becomes the SAS Embedded Process home.

- 3 Install the SAS Embedded Process and the SAS Hadoop MapReduce JAR files.

For more information, see [“Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files” on page 19](#).

Note: If you are installing the SAS High-Performance Analytics environment, you must perform additional steps after you install the SAS Embedded Process. For more information, see *SAS High-Performance Analytics Infrastructure: Installation and Configuration Guide*.

Upgrading from or Reinstalling a Previous Version

To upgrade or reinstall a previous version, follow these steps.

- 1 If you are upgrading from SAS 9.3, follow these steps. If you are upgrading from SAS 9.4, start with Step 2.

- a Stop the Hadoop SAS Embedded Process.

```
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.35/bin/sasep-stop.all.sh
```

SASEPHome is the master node where you installed the SAS Embedded Process.

- b** Delete the Hadoop SAS Embedded Process from all nodes.

```
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.35/bin/sasep-delete.all.sh
```

- c** Verify that the `sas.hadoop.ep.distribution-name`JAR files have been deleted.

The JAR files are located at *HadoopHome*/lib.

For Cloudera, the JAR files are typically located here:

```
/opt/cloudera/parcels/CDH/lib/hadoop/lib
```

For Hortonworks, the JAR files are typically located here:

```
/usr/lib/hadoop/lib
```

- d** Continue with Step 3.

2 If you are upgrading from SAS 9.4, follow these steps.

- a** Stop the Hadoop SAS Embedded Process.

```
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.*/bin/sasep-servers.sh  
-stop -hostfile host-list-filename | -host <">host-list<">
```

SASEPHome is the master node where you installed the SAS Embedded Process.

For more information, see [“SASEP-SERVERS.SH Script” on page 22](#).

- b** Remove the SAS Embedded Process from all nodes.

```
SASEPHome/SAS/SASTKInDatabaseForServerHadoop/9.*/bin/sasep-servers.sh  
-remove -hostfile host-list-filename | -host <">host-list<">  
-mrhome dir
```

Note: This step ensures that all old SAS Hadoop MapReduce JAR files are removed.

For more information, see [“SASEP-SERVERS.SH Script” on page 22](#).

- c** Verify that the `sas.hadoop.ep.apache*.jar` files have been deleted.

The JAR files are located at *HadoopHome*/lib.

For Cloudera, the JAR files are typically located here:

```
/opt/cloudera/parcels/CDH/lib/hadoop/lib
```

For Hortonworks, the JAR files are typically located here:

```
/usr/lib/hadoop/lib
```

3 Reinstall the SAS Embedded Process and the SAS Hadoop MapReduce JAR files by running the `sasep-servers.sh` script.

For more information, see [“Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files” on page 19](#).

Moving the SAS Embedded Process and SAS Hadoop MapReduce JAR File Install Scripts

Moving the SAS Embedded Process Install Script

The SAS Embedded Process install script is contained in a self-extracting archive file named `tkindbsrv-9.41_M2-n_lax.sh`. *n* is a number that indicates the latest version of the file. If this is the initial installation, *n* has a value of 1. Each time you reinstall or upgrade, *n* is incremented by 1. The self-extracting archive file is located in the `sharedFolder\InClusterBundle` directory.

Using a method of your choice, transfer the SAS Embedded Process install script to your Hadoop master node.

This example uses secure copy, and *SASEPHome* is the location where you want to install the SAS Embedded Process.

```
scp tkindbsrv-9.41_M2-n_lax.sh username@hadoop:/SASEPHome
```

Note: The location where you transfer the install script becomes the SAS Embedded Process home.

Note: Both the SAS Embedded Process install script and the SAS Hadoop MapReduce JAR file install script must be transferred to the same directory.

Moving the SAS Hadoop MapReduce JAR File Install Script

The SAS Hadoop MapReduce JAR file install script is contained in a self-extracting archive file named `hadoopmrjars-9.41_M2-n_lax.sh`. *n* is a number that indicates the latest version of the file. If this is the initial installation, *n* has a value of 1. Each time you reinstall or upgrade, *n* is incremented by 1. The self-extracting archive file is located in the `sharedFolder\InClusterBundle` directory.

Using a method of your choice, transfer the SAS Hadoop MapReduce JAR file install script to your Hadoop master node.

This example uses Secure Copy, and *SASEPHome* is the location where you want to install the SAS Hadoop MapReduce JAR files.

```
scp hadoopmrjars-9.41_M2-n_lax.sh username@hadoop:/SASEPHome
```

Note: Both the SAS Embedded Process install script and the SAS Hadoop MapReduce JAR file install script must be transferred to the same directory.

Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files

To install the SAS Embedded Process, follow these steps.

Note: Permissions are needed to install the SAS Embedded Process and SAS Hadoop MapReduce JAR files. For more information, see [“Hadoop Permissions” on page 29](#).

- 1 Log on to the server using SSH as root with sudo access.

```
ssh username@serverhostname
sudo su - root
```

- 2 Move to your Hadoop master node where you want the SAS Embedded Process installed.

```
cd /SASEPHome
```

SASEPHome is the same location to which you copied the self-extracting archive file. For more information, see [“Moving the SAS Embedded Process Install Script” on page 19](#).

- 3 Use the following script to unpack the `tkindbsrv-9.41_M2-n_lax.sh` file.

```
./tkindbsrv-9.41_M2-n_lax.sh
```

n is a number that indicates the latest version of the file. If this is the initial installation, *n* has a value of 1. Each time you reinstall or upgrade, *n* is incremented by 1.

Note: If you unpack in the wrong directory, you can move it after the unpack.

After this script is run and the files are unpacked, the script creates the following directory structure where **SASEPHome** is the master node from Step 1.

```
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/misc
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/sasexe
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/utilities
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/build
```

The content of the

SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin directory should look similar to this.

```
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/sas.ep4hadoop.template
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/sasep-servers.sh
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/sasep-common.sh
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/sasep-server-start.sh
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/sasep-server-status.sh
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/sasep-server-stop.sh
SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/InstallTKIndbsrv.sh
```

- 4 Use this command to unpack the SAS Hadoop MapReduce JAR files.

```
./hadoopmrjars-9.41_M2-1_lax.sh
```

After the script is run, the script creates the following directory and unpacks these files to that directory.

```
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/ep-config.xml
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/
sas.hadoop.ep.apache023.jar
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/
sas.hadoop.ep.apache023.nls.jar
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/
sas.hadoop.ep.apache121.jar
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/
sas.hadoop.ep.apache121.nls.jar
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/
sas.hadoop.ep.apache205.jar
SASEPHome/SAS/SASACCESStoHadoopMapReduceJARFiles/9.41_M2/lib/
sas.hadoop.ep.apache205.nls.jar
```


- 5 Use the `sasep-servers.sh -add` script to deploy the SAS Embedded Process installation across all nodes. The SAS Embedded Process is installed as a Linux service.

TIP There are many options available when installing the SAS Embedded Process. We recommend that you review the script syntax before running it. For more information, see [“SASEP-SERVERS.SH Script” on page 22](#).

Run the `sasep-servers.sh` script.

```
cd $ASEPHOME/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin
./sasep-servers.sh -add
```

TIP There are many options available when installing the SAS Embedded Process. We recommend that you review the script syntax before running it. For more information, see [“SASEP-SERVERS.SH Script” on page 22](#).

During the install process, the script asks whether you want to start the SAS Embedded Process. If you choose `y` or `Y`, the SAS Embedded Process is started on all nodes after the install is complete. If you choose `n` or `N`, you can start the SAS Embedded Process later by running `./sasep-servers.sh -start`.

Note: When you run the `sasep-servers.sh -add` script, a user and group named `sasep` is created. You can specify a different user and group name with the `-epuser` and `-epgroup` arguments when you run the `sasep-servers.sh -add` script.

Note: The `sasep-servers.sh` script can be run from any location. You can also add its location to the `PATH` environment variable.

Note: Although you can install the SAS Embedded Process in multiple locations, the best practice is to install only one instance.

Note: The SAS Embedded Process runs on all the nodes that are capable of running a MapReduce task. In some instances, the node that you chose to be the master node can also serve as a MapReduce task node.

Note: If you install the SAS Embedded Process on a large cluster, the SSHD daemon might reach the maximum number of concurrent connections. The `ssh_exchange_identification: Connection closed by remote host` SSHD error might occur. Follow these steps to work around the problem:

- 1 Edit the `/etc/ssh/sshd_config` file and change the `MaxStartups` option to the number that accommodates your cluster.
- 2 Save the file and reload the SSHD daemon by running the `/etc/init.d/sshd reload` command.
- 6 If this is the first install of the SAS Embedded Process, a restart of the Hadoop YARN or MapReduce service is required.

This enables the cluster to reload the SAS Hadoop JAR files (`sas.hadoop.ep.*.jar`).

Note: It is preferable to restart the service by using Cloudera Manager or Hortonworks Ambari.

- 7 Verify that the SAS Embedded Process is installed and running. Change directories and then run the `sasep-servers.sh` script with the `-status` option.

```
cd $ASEPHOME/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin
./sasep-servers.sh -status
```

This command returns the status of the SAS Embedded Process running on each node of the Hadoop cluster. Verify that the SAS Embedded Process home directory is correct on all the nodes.

Note: The `sasep-servers.sh -status` script cannot run successfully if the SAS Embedded Process is not installed.

- 8 Verify that the `sas.hadoop.ep.apache*.jar` files are now in place on all nodes.

The JAR files are located at `HadoopHome/lib`.

For Cloudera, the JAR files are typically located here:

```
/opt/cloudera/parcels/CDH/lib/hadoop/lib
```

For Hortonworks, the JAR files are typically located here:

```
/usr/lib/hadoop/lib
```

- 9 Verify that an `init.d` service with a `sas.ep4hadoop` file was created in the following directory.

```
/etc/init.d/sas.ep4hadoop
```

View the `sas.ep4hadoop` file and verify that the SAS Embedded Process home directory is correct.

The `init.d` service is configured to start at level 3 and level 5.

Note: The SAS Embedded Process needs to run on all nodes that you chose during installation.

- 10 Verify that configuration files were written to the HDFS file system.

```
hadoop fs -ls /sas/ep/config
```

Note: If you are running on a cluster with Kerberos, you need a Kerberos ticket. If not, you can use the WebHDFS browser.

Note: The `/sas/ep/config` directory is created automatically when you run the install script.

SASEP-SERVERS.SH Script

Overview of the SASEP-SERVERS.SH Script

The `sasep-servers.sh` script enables you to perform the following actions.

- Install or uninstall the SAS Embedded Process and SAS Hadoop MapReduce JAR files on a single node or a group of nodes.
- Start or stop the SAS Embedded Process on a single node or on a group of nodes.
- Determine the status of the SAS Embedded Process on a single node or on a group of nodes.

- Write the installation output to a log file.
- Pass options to the SAS Embedded Process.
- Create a HADOOP_JARZIP file in the local folder. This ZIP file contains all required client JAR files.

Note: The sasep-servers.sh script can be run from any folder on any node in the cluster. You can also add its location to the PATH environment variable.

Note: You must have sudo access to run the sasep-servers.sh script.

SASEP-SERVERS.SH Syntax

sasep-servers.sh

```
-add | -remove | -start | -stop | -status | -restart
<-mrhome path-to-mr-home>
<-hdfsuser user-id>
<-epuser>epuser-id
<-epgroup>epgroup-id
<-hostfile host-list-filename | -host <">host-list<">>
<-epscript path-to-ep-install-script>
<-mrscript path-to-mr-jar-file-script>
<-options "option-list">
<-log filename>
<-version apache-version-number>
<-getjars>
```

Arguments

-add

installs the SAS Embedded Process.

Note The -add argument also starts the SAS Embedded Process (same function as -start argument). You are prompted and can choose whether to start the SAS Embedded Process.

Tip You can specify the hosts on which you want to install the SAS Embedded Process by using the -hostfile or -host option. The -hostfile or -host options are mutually exclusive.

See [-hostfile and -host option on page 25](#)

-remove

removes the SAS Embedded Process.

Tip You can specify the hosts for which you want to remove the SAS Embedded Process by using the -hostfile or -host option. The -hostfile or -host options are mutually exclusive.

See [-hostfile and -host option on page 25](#)

-start

starts the SAS Embedded Process.

Tip You can specify the hosts on which you want to start the SAS Embedded Process by using the `-hostfile` or `-host` option. The `-hostfile` or `-host` options are mutually exclusive.

See [-hostfile and -host option on page 25](#)

-stop

stops the SAS Embedded Process.

Tip You can specify the hosts on which you want to stop the SAS Embedded Process by using the `-hostfile` or `-host` option. The `-hostfile` or `-host` options are mutually exclusive.

See [-hostfile and -host option on page 25](#)

-status

provides the status of the SAS Embedded Process on all hosts or the hosts that you specify with either the `-hostfile` or `-host` option.

Tips The status also shows the version and path information for the SAS Embedded Process.

You can specify the hosts for which you want the status of the SAS Embedded Process by using the `-hostfile` or `-host` option. The `-hostfile` or `-host` options are mutually exclusive.

See [-hostfile and -host option on page 25](#)

-restart

restarts the SAS Embedded Process.

Tip You can specify the hosts on which you want to restart the SAS Embedded Process by using the `-hostfile` or `-host` option. The `-hostfile` or `-host` options are mutually exclusive.

See [-hostfile and -host option on page 25](#)

-mrhome *path-to-mr-home*

specifies the path to the MapReduce home.

-hdfsuser *user-id*

specifies the user ID that has Write access to HDFS root directory.

Default hdfs

Note The user ID is used to copy the SAS Embedded Process configuration files to HDFS.

-epuser *epuser-name*

specifies the name for the SAS Embedded Process user.

Default sasep

-epgroup *epgroup-name*

specifies the name for the SAS Embedded Process group.

Default sasep

-hostfile *host-list-filename*

specifies the full path of a file that contains the list of hosts where the SAS Embedded Process is installed, removed, started, stopped, or status is provided.

Default If you do not specify -hostfile, the sasep-servers.sh script will discover the cluster topology and uses the retrieved list of data nodes.

Tip You can also assign a host list filename to a UNIX variable, **sas_ephosts_file**.

```
export sasep_hosts=/etc/hadoop/conf/slaves
```

See [“-hdfsuser *user-id*” on page 24](#)

Example `-hostfile /etc/hadoop/conf/slaves`

-host <">*host-list*<">

specifies the target host or host list where the SAS Embedded Process is installed, removed, started, stopped, or status is provided.

Default If you do not specify -host, the sasep-servers.sh script will discover the cluster topology and uses the retrieved list of data nodes.

Requirement If you specify more than one host, the hosts must be enclosed in double quotation marks and separated by spaces.

Tip You can also assign a list of hosts to a UNIX variable, **sas_ephosts**.

```
export sasep_hosts="server1 server2 server3"
```

See [“-hdfsuser *user-id*” on page 24](#)

Example `-host "server1 server2 server3"`
`-host bluesvr`

-epscript *path-to-ep-install-script*

copies and unpacks the SAS Embedded Process install script file to the host.

Restriction Use this option only with the -add option.

Requirement You must specify either the full or relative path of the SAS Embedded Process install script, `tkindbsrv-9.41_M2-n_lax.sh` file.

Example `-epscript /home/hadoop/image/current/tkindbsrv-9.41_M2-2_lax.sh`

-mrscript *path-to-mr-jar-file-script*

copies and unpacks the SAS Hadoop MapReduce JAR files install script on the hosts.

Restriction Use this option only with the -add option.

Requirement You must specify either the full or relative path of the SAS Hadoop MapReduce JAR file install script, `hadoopmrjars-9.41_M2-n_lax.sh` file.

Example `-mrscrip /home/hadoop/image/current/tkindbsrv-9.41_M2-2_lax.sh`

-options "*option-list*"

specifies options that are passed directly to the SAS Embedded Process. The following options can be used.

-trace *trace-level*

specifies what type of trace information is created.

- 0 no trace log
- 1 fatal error
- 2 error with information or data value
- 3 warning
- 4 note
- 5 information as an SQL statement
- 6 critical and command trace
- 7 detail trace, lock
- 8 enter and exit of procedures
- 9 tedious trace for data types and values
- 10 trace all information

Default 02

Note The trace log messages are stored in the MapReduce job log.

-port *port-number*

specifies the TCP port number where the SAS Embedded Process accepts connections.

Default 9261

Requirement The options in the list must be separated by spaces, and the list must be enclosed in double quotation marks.

-log *filename*

writes the installation output to the specified filename.

-version *Apache-version-number*

specifies the Hadoop version of the JAR file that you want to install on the cluster. The *apache-version-number* can be one of the following values.

0.23

installs the SAS Hadoop MapReduce JAR files that are built from Apache Hadoop 0.23 (sas.hadoop.ep.apache023.jar and sas.hadoop.ep.apache023.nls.jar).

1.2

installs the SAS Hadoop MapReduce JAR files that are built from Apache Hadoop 1.2.1 (sas.hadoop.ep.apache121.jar and sas.hadoop.ep.apache121.nls.jar).

2.0

installs the SAS Hadoop MapReduce JAR files that are built from Apache Hadoop 0.2.3 (sas.hadoop.ep.apache023.jar and sas.hadoop.ep.apache023.nls.jar).

2.1

installs the SAS Hadoop MapReduce JAR files that are built from Apache Hadoop 2.0.5 (sas.hadoop.ep.apache205.jar and sas.hadoop.ep.apache205.nls.jar).

Default

If you do not specify the -version option, the sasep.servers.sh script will detect the version of Hadoop that is in use and install the JAR files associated with that version. For more information, see [“Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files” on page 19](#).

Interaction

The -version option overrides the version that is automatically detected by the sasep.servers.sh script.

-getjars

creates a HADOOP_JARZIP file in the local folder. This ZIP file contains all required client JAR files.

Starting the SAS Embedded Process

There are three ways to manually start the SAS Embedded Process.

Note: Root authority is required to run the sasep-servers.sh script.

- Run the sasep-servers.sh script with the -start option on the master node.

This starts the SAS Embedded Process on all nodes. For more information about running the sasep-servers.sh script, see [“SASEP-SERVERS.SH Syntax” on page 23](#).

- Run sasep-server-start.sh on a node.

This starts the SAS Embedded Process on the local node only. The sasep-server-start.sh script is located in the *SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/* directory. For more information, see [“Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files” on page 19](#).

- Run the UNIX service command on a node.

This starts the SAS Embedded Process on the local node only. The service command calls the init script that is located in the */etc/init.d* directory. A symbolic link to the init script is created in the */etc/rc3.d* and */etc/rc5.d* directories, where 3 and 5 are the run level at which you want the script to be executed.

Because the SAS Embedded Process init script is registered as a service, the SAS Embedded Process is started automatically when the node is rebooted.

Stopping the SAS Embedded Process

The SAS Embedded Process continues to run until it is manually stopped. The ability to control the SAS Embedded Process on individual nodes could be useful when performing maintenance on an individual node.

There are three ways to stop the SAS Embedded Process.

Note: Root authority is required to run the `sasep-servers.sh` script.

- Run the `sasep-servers.sh` script with the `-stop` option from the master node.

This stops the SAS Embedded Process on all nodes. For more information about running the `sasep-servers.sh` script, see [“SASEP-SERVERS.SH Syntax” on page 23](#).

- Run `sasep-server-stop.sh` on a node.

This stops the SAS Embedded Process on the local node only. The `sasep-server-stop.sh` script is located in the `$SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/` directory. For more information, see [“Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files” on page 19](#).

- Run the UNIX `service` command on a node.

This stops the SAS Embedded Process on the local node only.

Determining the Status of the SAS Embedded Process

You can display the status of the SAS Embedded Process on one node or all nodes. There are three ways to display the status of the SAS Embedded Process.

Note: Root authority is required to run the `sasep-servers.sh` script.

- Run the `sasep-servers.sh` script with the `-status` option from the master node.

This displays the status of the SAS Embedded Process on all nodes. For more information about running the `sasep-servers.sh` script, see [“SASEP-SERVERS.SH Syntax” on page 23](#).

- Run `sasep-server-status.sh` from a node.

This displays the status of the SAS Embedded Process on the local node only. The `sasep-server-status.sh` script is located in the `$SASEPHome/SAS/SASTKInDatabaseServerForHadoop/9.41_M2/bin/` directory. For more information, see [“Installing the SAS Embedded Process and SAS Hadoop MapReduce JAR Files” on page 19](#).

- Run the UNIX `service` command on a node.

This displays the status of the SAS Embedded Process on the local node only.

Hadoop Permissions

The person who installs the SAS Embedded Process must have sudo access.

Appendix 1

Hardware Virtualization

An error stating that VT-x or AMD-v is not available indicates that changes need to be made to the BIOS (or firmware) of your system before you can use SAS Data Loader for Hadoop. In general, this error message indicates one of two things: that your system does not support virtualization or that the option to use virtualization needs to be enabled. To remedy this situation, you must perform three tasks:

- verify that your computer supports virtualization
- change the virtualization option
- restart your machine into the BIOS menus

Follow these steps:

- 1 Verify that your computer supports virtualization. Typically, newer computers support virtualization, however, there are exceptions to this. Determine whether your x64-based machine has an Intel or AMD processor installed. Follow the steps below to locate this information.
 - a On a Windows machine, press the Windows key and the R key on your keyboard at the same time. The **Run** dialog box appears.
 - b In the **Open** field of the dialog box, type msinfo32 and click **OK**.
 - c In the System Information window, ensure that System Summary is selected in the left panel.
 - d In the right panel, find System Type and ensure that you have an x64-based machine. Next, find Processor. The manufacturer of the processor is shown here.
- 2 Once you have located the manufacturer name, download, and use the tool that corresponds to your processor. These tools provide a brief description of your computer's capabilities and whether the virtualization technology is supported on your machine.
 - Download the Intel tool at: https://downloadcenter.intel.com/Detail_Desc.aspx?DwnldID=7838
 - Download the AMD tool at: http://download.amd.com/techdownloads/AMD-VwithRVI_Hyper-V_CompatibilityUtility.zip
- 3 If you have determined that your computer supports virtualization, visit [virtualization hardware extensions page](#) to learn about enabling Intel VT and AMD-V virtualization hardware extensions in BIOS. The page provides the general process for entering the BIOS and changing the virtualization setting.

Note: You must restart your computer during this process.

The BIOS varies greatly by the make and model of your computer. To obtain information about how to navigate through your specific BIOS, contact the support site for the manufacturer of your computer.

Recommended Reading

- *SAS Data Loader for Hadoop: User's Guide*
- *SAS In-Database Products: Administrator's Guide*
- *SAS Hadoop Configuration Guide for Base SAS and SAS/ACCESS*

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

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

Index

C

configuration
Hadoop 17

H

Hadoop
in-database deployment package 15
installation and configuration 17
permissions 29
SAS/ACCESS Interface 15
starting the SAS Embedded Process 27
status of the SAS Embedded Process 28
stopping the SAS Embedded Process 28
unpacking self-extracting archive files 19

I

in-database deployment package for Hadoop
overview 16
prerequisites 15
installation
Hadoop 17
SAS Embedded Process (Hadoop) 16, 19
SAS Hadoop MapReduce JAR files 19

P

permissions
for Hadoop 29
publishing
Hadoop permissions 29

R

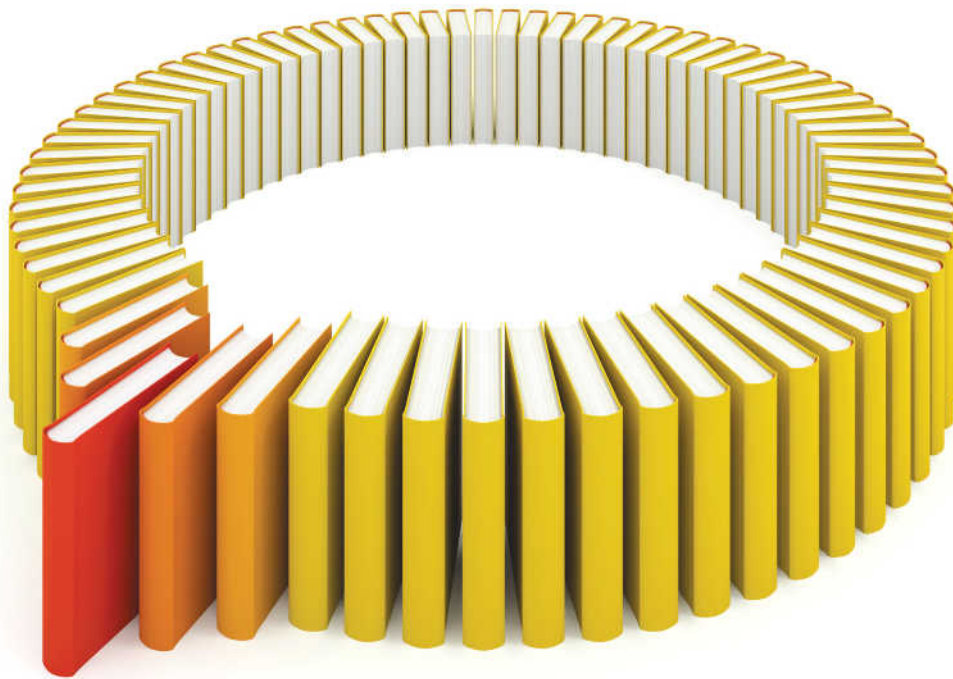
reinstalling a previous version
Hadoop 17

S

SAS Embedded Process
controlling (Hadoop) 22
Hadoop 15
SAS Foundation 15
SAS Hadoop MapReduce JAR files 19
SAS/ACCESS Interface to Hadoop 15
sasep-servers.sh script
overview 22
syntax 23
self-extracting archive files
unpacking for Hadoop 19

U

unpacking self-extracting archive files for Hadoop 19
upgrading from a previous version
Hadoop 17



Gain Greater Insight into Your SAS[®] 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

