Hortonworks Data Platform

Quick Start Guide: Single-Node Cluster
(Sep 30, 2015)

docs.hortonworks.com

Hortonworks Data Platform: Quick Start Guide: Single-Node Cluster

The Hortonworks Data Platform, powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner. The Hortonworks Data Platform consists of the essential set of Apache Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, Zookeeper and Ambari. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

Unlike other providers of platforms built using Apache Hadoop, Hortonworks contributes 100% of our code back to the Apache Software Foundation. The Hortonworks Data Platform is Apache-licensed and completely open source. We sell only expert technical support, training and partner-enablement services. All of our technology is, and will remain free and open source.

Please visit the Hortonworks Data Platform page for more information on Hortonworks technology. For more information on Hortonworks services, please visit either the Support or Training page. Feel free to Contact Us directly to discuss your specific needs.

Table of Contents

List of Tables

1. Quick Installation Guide Installing HDP for Windows on a Single-Node Cluster

1.1. Overview

Hortonworks Data Platform for Windows is typically deployed in a two or more node cluster environment. You should deploy HDP for Windows on a single-node cluster for evaluation purposes only. This Quick Start Guide provides you with quick and easy instructions to deploy HDP for Windows on a single-node cluster, walks you through the installation workflow, and provides some simple deployment instructions.

1.2. Quick Installation Workflow Overview

Use the following workflow to get started quickly:
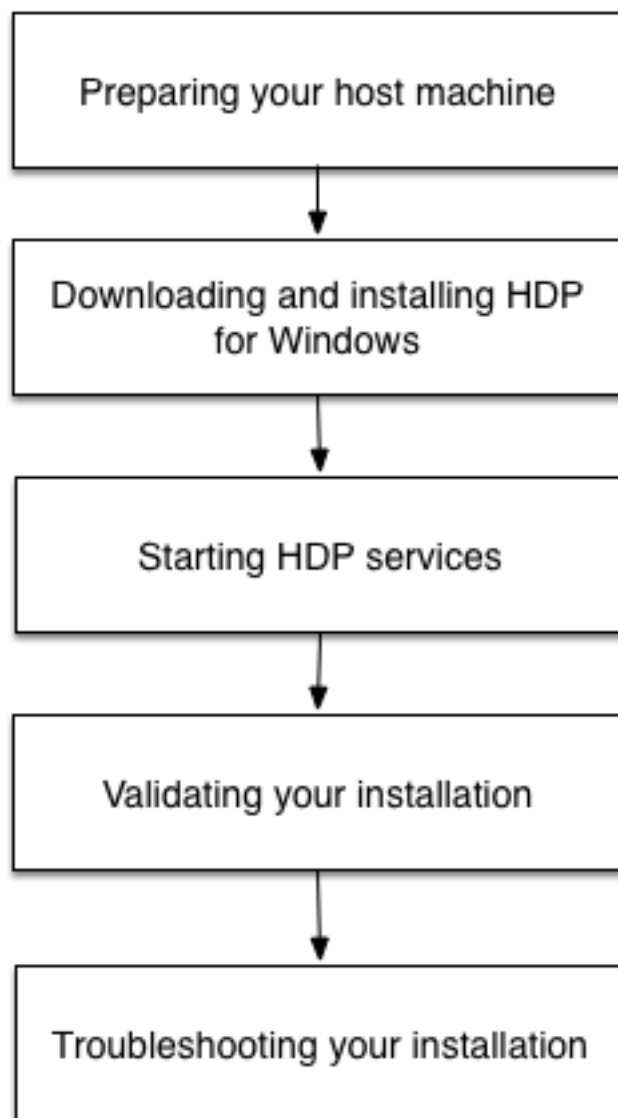
1.3. Preparing Your Host Machine

Before installing HDP for Windows, you should ensure that your host system meets the following minimum requirements.

For an evaluation installation, you should plan to do the following:

1. Ensure your system meets the minimum hardware and operating system requirements.

2. Download and install some additional required software.

3. Disable your firewalls.

1.3.1. System Requirements

You should ensure that your host system meets the following minimum requirements and software requirements.

Table 1.1. System Requirements

System Requirement	Details
Hardware requirements	5 GB free space on your system drive
Supported operating systems	• Windows Server 2008 R2 (64-bit) • Windows Server 2012 (64-bit) • Windows Server 2012 R1 (64-bit) • Windows Server 2012 R2 (64-bit)
Required software	• Python 2.7.x • Java JDK 1.7.x • Microsoft Visual C++ 2010 Redistributable Package (64-bit) • Microsoft .NET Framework 4.0

Download and install Visual C++ and the .NET Framework according to instructions on the Microsoft Websites. Follow the instructions below to install and configure Python and the Java JDK.

The .NET Framework 4.0 is already installed on Windows Server 2012 and later.

1.3.2. Installing and Configuring Python 2.7.x

1. Download Python and install it in a directory without whitespaces. For example: `c:\Python.`

2. As the administrator, update the PATH environment variable:

   a. From Control Panel > System, click the Advanced system setting link.

   b. Click Advanced, then click Environment Variables.

   c. In the System Variables field, select PATH and click Edit.

   d. After the last entry in the PATH value, enter a semi-con and add the Python installation directory path. For example, add: `;c:\Python27`

   e. Click OK to close the Environment Variable dialog.

3. From PowerShell or the command line, validate your environment variable settings by entering:

```
python -V Python 2.7.10
```

1.3.3. Installing and Configuring your Java JDK

1. Download the Java JDK and install it a directory without white spaces. For example, `c:\Java`.

   Ensure that the JDK folder is located inside the Java folder.

2. Add the JAVA_HOME environment variable:

   a. Click Control Panel > System and click the Advanced system settings link.

   b. Click Advanced, then click Environment Variables.

   c. Add JAVA_HOME as a new system environment variable and specify the Java Development Kit installation path as the value, and then click OK. For example, `c:\Java\jdk1.7.0_51`.

3. From the command line, validate the JAVA_HOME environment variable. Enter: `Echo %JAVA_HOME%`

   The path you specified as the JAVA_HOME value should be returned. For example: `c:\Java\jdk1.7.0_51`

4. As the administrator, update your PATH environment variable:

   a. In the System Variables field, select PATH and click Edit.

   b. After the last entry in the PATH value, enter a semi-color and add the installation path to your JDK. For example: `...;c\Java\jdk1.7.0_51\bin`.

   c. Click OK.

5. From the command line, validate your PATH environment variable edit by entering: `java version`

   The return value should b the JDK version details. For example: `Java version "1.7.0."`

1.3.4. Disabling your Windows firewall

Disable your Windows firewall so that all the ports are available for use:

1. Follow the instructions in this Microsoft Technet article to disable the firewall.

1.4. Downloading and Installing HDP for Windows

1. Download the HDP for Windows installation zip file.

2. From the command line, launch the HDP for Windows installer by entering:

```
runas /user:administrator "cmd /C msiexec /lv c:\hdplog.txt /i
PATH_to_MSI_file MSIUSEREALADMINDETECTION=1"
```

where PATH_to_MSI_file is the location of the downloaded MSI file. For example:

```
runas /user:administrator "cmd /C msiexec /lv c:\hdplog.txt /i c:\
MSI_INSTALL\hdp-2.3.0.0.winpkg.msi MSIUSEREALADMINDETECTION=1"
```

3. The HDP Setup install window displays. Provide the following information:

   • "Hadoop" user password

   • Hive and Oozie database names, user names, and passwords

   • Specify that you want to use a Derby database

   • Specify that you want delete existing HDP data

4. In addition to the Main components, for a single-node cluster deployment, you can click the Additional components tab, and install the following components:

   • Datafu

- Falcon

- Flume

- HBase

- Knox

- Phoenix

- Ranger

- Slider

- Spark

- Storm

5. Click Install to complete your installation. This can take up to 20 minutes, depending on which components you are installing.

1.5. Starting HDP services

After installation, you need to start each HDP service for the first time. You can do this from the Hadoop command line that was added to your host system desk top during installation.

1.5.1. Starting HDP services from the command line

Note

To perform any administrative tasks, you need to be logged in as the Hadoop super user, using the password you created during installation.

1. From the Hadoop command line, navigate to the HDP install directory.

2. Enter: `%HADOOP_NODE%\start_local_hdp_services`

1.5.2. Manually started individual HDP services

If all your services do not start successfully the first time, you can either rerun the above command, or you can manually start the HDP services using the control panel.

1. Launch Control Panel.

2. Click Administrative Tools > Services.

3. Select the service you want to started, and click Start.

1.6. Validating your Installation

Once you have installed HDP services, you should verify that they are working as expected. You can do this by running provided smoke tests.

1. Create a smoke test user directory in HDFS, if one does not already exist. From the Hadoop command line created during installation:

```
%HADOOP_HOME% \bin\hdfs dfs –mkdir –p /user/smoketest
        %HADOOP_HOME% \bin\hdfs dfs –chown –R smoketest
```

2. Establish that you want to run your smoketests as a hadoop user:

```
runas /user:hadoop
```

3. Run your smoketests:

```
“cmd /K %HADOOP_NODE%\Run-SmokeTests.cmd”
```

1.7. Troubleshooting your Installation

If you do experience errors in your installation, or in with the services themselves, it is useful to know where to find configuration and error log files.

Table 1.2. Component configuration and log file locations

Component	Configuration file location	Log file location
Hadoop (HDFS, YARN, MapReduce)	C:\hdp\hadoop-2.7.1.2.3.0.0-2543\etc\hadoop	C:\hadoop\logs\hadoop
ZooKeeper	C:\hdp\zookeeper-3.4.6.2.3.0.0-2543\conf	C:\hadoop\logs\zookeeper
Hive	C:\hdp\hive-1.2.1.2.3.0.0-2543\conf	C:\hadoop\logs\hive
HBase	C:\hdp\hbase-1.1.1.2.3.0.0-2543\conf	C:\hadoop\logs\hbase
WebHCat	C:\hdp\hive-1.2.1.2.3.0.0-2543\hcatalog\etc\webhcat	C:\hadoop\logs\webhcat
Oozie	C:\hdp\oozie-4.2.0.2.3.0.0-2543\oozie-win-distro\conf	C:\hadoop\logs\oozie
Storm	C:\hdp\storm-0.10.0.2.3.0.0-2543\conf	C:\hadoop\logs\storm
Knox	C:\hdp\knox-0.6.0.2.3.0.0-2543\conf	C:\hadoop\logs\knox
Flume	C:\hdp\flume-1.5.2.2.3.0.0-2543\conf	C:\hdp\flume-1.5.2.2.3.0.0-2543\bin\logs
Pig	C:\hdp\pig-0.15.0.2.3.0.0-2543\conf	No log files because no service is running
Sqoop	C:\hdp\sqoop-1.4.6.2.3.0.0-2543\conf	No log files because no service is running
Tez	C:\hdp\tez-0.7.0.2.3.0.0-2543\conf	No log files because no service is running