

# Hortonworks Data Platform Windows

## Installing HDP for Windows

(May 21, 2013)

---

## Hortonworks Data Platform Win 1.1.0 (GA) : Installing HDP for Windows

Copyright © 2013 Hortonworks, Inc. Some rights reserved.

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.



Except where otherwise noted, this document is licensed under  
**Creative Commons Attribution ShareAlike 3.0 License.**  
<http://creativecommons.org/licenses/by-sa/3.0/legalcode>

## Table of Contents

1. Getting Ready to Install .....	1
1.1. Understand the Basics .....	1
1.2. Meet Minimum System Requirements .....	2
1.2.1. Hardware Recommendations .....	2
1.2.2. Operating Systems Requirements .....	2
1.2.3. Software Requirements .....	2
1.2.4. Database Requirements .....	4
1.3. Decide on Deployment Type .....	5
1.4. Collect Information .....	5
1.5. Prepare the Cluster Environment .....	6
1.5.1. Configure common time for all cluster nodes .....	6
1.5.2. Ensure that Windows networking uses IPv4 addresses .....	6
1.5.3. Optional - Create Hadoop user .....	7
1.5.4. Configure ports .....	7
1.5.5. Enable networking configurations for Workgroups .....	10
1.5.6. Enable networking configurations for Active Directory Domains .....	10
1.6. Define Cluster Configuration .....	20
2. Quick Start Guide for Single Node HDP Installation .....	24
3. Deploying HDP .....	28
3.1. Option I - Central Push Install Using Corporate Standard Procedures .....	28
3.2. Option II - Central Push Install Using Provided Script .....	29
3.3. Option III - Manual Install One Node At A Time .....	31
3.4. Optional - Install Client Host .....	33
4. Managing HDP on Windows .....	35
4.1. Starting the HDP Services .....	35
4.2. Stopping the HDP Services .....	36
5. Uninstalling HDP .....	38
5.1. Option I - Use Windows GUI .....	38
5.2. Option II - Use Command Line Utility .....	38

---

## List of Tables

1.1. HDFS Ports .....	7
1.2. MapReduce Ports .....	8
1.3. Hive Ports .....	9
1.4. Templeton Port .....	9
1.5. Configuration values for MSI installer .....	21

# 1. Getting Ready to Install

This section describes the information and materials you need to get ready to install the Hortonworks Data Platform (HDP) on Windows.

Use the following instructions before you start deploying Hadoop using HDP installer:

- [Understand the Basics](#)
- [Meet Minimum System Requirements](#)
- [Decide on Deployment Type](#)
- [Collect Information](#)
- [Prepare the Cluster Environment](#)
- [Define Cluster Configuration](#)

## 1.1. Understand the Basics

The Hortonworks Data Platform consists of three layers.

- **Core Hadoop:** The basic components of Apache Hadoop.
  - **Hadoop Distributed File System (HDFS):** A special purpose file system that is designed to work with the MapReduce engine. It provides high-throughput access to data in a highly distributed environment.
  - **MapReduce:** A framework for performing high volume distributed data processing using the MapReduce programming paradigm.
- **Essential Hadoop:** A set of Apache components designed to ease working with Core Hadoop.
  - **Apache Pig:** A platform for creating higher level data flow programs that can be compiled into sequences of MapReduce programs, using Pig Latin, the platform's native language.
  - **Apache Hive:** A tool for creating higher level SQL-like queries using HiveQL, the tool's native language, that can be compiled into sequences of MapReduce programs.
  - **Apache HCatalog:** A metadata abstraction layer that insulates users and scripts from how and where data is physically stored.
  - **Templeton:** A component that provides a set of REST-like APIs for HCatalog and related Hadoop components.
- **Supporting Components:** A set of components that allow you to monitor your Hadoop installation and to connect Hadoop with your larger compute environment.
  - **Apache Oozie:** A server based workflow engine which is optimized for running workflows that execute Hadoop jobs.

- **Apache Sqoop:** A component that provides a mechanism for moving data between HDFS and external structured datastores. Sqoop can be integrated with Oozie workflows.

For more information on the structure of the HDP, see [Understanding Hadoop Ecosystem](#).

## 1.2. Meet Minimum System Requirements

To run the Hortonworks Data Platform, your system must meet minimum requirements.

- [Hardware Recommendations](#)
- [Operating System Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)

### 1.2.1. Hardware Recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines.

You can see sample setups here: [Hardware Recommendations for Apache Hadoop](#).

### 1.2.2. Operating Systems Requirements

The following operating systems are supported:

- Windows Server 2008 R2 (64 bit)
- Windows Server 2012 (64 bit)

### 1.2.3. Software Requirements

Ensure that you install the following on every host machine in your cluster:

- **Microsoft Visual C++ 2010 Redistributable Package (64 bit)**
  1. Use the instructions provided [here](#) to download and install Microsoft Visual C++ 2010 Redistributable Package (64 bit).
- **Microsoft.NET framework 4.0**
  1. Use the instructions provided [here](#) to download and install Microsoft.NET framework 4.0.
- **JDK 6u31**

Use the instructions provided below to manually install JDK in your local environment:

1. Check the version. From a command shell or Powershell window, type:

```
java -version
```

2. (Optional): Uninstall the Java package if the JDK version is less than v1.6 update 31.
3. Download the Oracle 64-bit JDK for Windows from [here](#) and run the JDK Installer MSI.



### Important

Ensure that no whitespace characters are present in the installation directory's path.

For example, `C:\Program Files` is **not** allowed.

4. To validate if Java has been installed on your cluster host, go to the command shell and type `java -version`.
5. Create a system variable named `JAVA_HOME`. The value of this variable will be the full path to installation directory for JDK.
  - a. Open the **Control Panel** -> **System** pane and click on **Advanced system settings**.
  - b. Click on the **Advanced** tab.
  - c. Click the **Environment Variables** button.
  - d. Under **System variables**, click **New**.
  - e. Enter the **Variable Name** as `JAVA_HOME`.
  - f. Enter the **Variable Value**, as the installation path for the Java Development Kit.

For example, if your JDK is installed at `C:\Software\Java\jdk1.6.0_31`, then you must provide this path to the **Variable Value**.
  - g. Click **OK**.
  - h. Click **Apply Changes**.

- **Python 2.7**

Use the following instructions to manually install Python in your local environment:

1. Download Python from [here](#).
2. Update the `PATH` environment variable. Using Administrator privileges:
  - a. Open the **Control Panel** -> **System** pane and click on the **Advanced system settings** link.
  - b. Click on the **Advanced** tab.
  - c. Click the **Environment Variables** button.
  - d. Under **System Variables**, find `PATH` and click **Edit**.



### Note

If you do not have the item `PATH`, you may select to add a new variable and add `PATH` as the name and the location of the Python directory as the value.

- e. In the Edit windows, modify `PATH` by appending the installation path for your Python directory to the value of `PATH`.

For example, if Python executable is installed at `C:\Python27` then you must append this value to `PATH`.

- f. To validate your settings, from a command shell or Powershell window, type:

```
python
```

## 1.2.4. Database Requirements

- By default, Hive and Oozie use an embedded Derby database for its metastore.

To use an external database for Hive and Oozie metastores, Microsoft SQL Server database must be deployed and available in your environment.



### Important

Before using SQL server for Hive metastore, set up Microsoft SQL Server JDBC Driver using the instructions provided [here](#).

- Ensure that your database administrator creates the following databases and users.

Note down the database name and user account credentials. You need these details while configuring the HDP Installer.

- For Hive:

1. hive\_dbname



### Note

Create Hive database in SQL.

2. hive\_dbuser



### Note

Create Hive users on SQL and add them to the `sysadmin` role within SQL.

3. hive\_dbpasswd

- For Oozie:

1. oozie\_dbname



### Note

Create Oozie database in SQL.

2. oozie\_dbuser



### Note

Create Oozie users on SQL and add them to the `sysadmin` role within SQL.

3. oozie\_dbpasswd



### Important

Ensure that you set the security policy of Microsoft SQL server to use both SQL and Windows authentication. (By default, the security policy uses Windows authentication.)

## 1.2.4.1. (Optional) Install Microsoft SQL Server JDBC Driver

If you are using MS SQL Server for Hive and Oozie metastores, you must install the MS SQL Server JDBC driver.

1. Download the SQL JDBC JAR file [sqljdbc\\_3.0.1301.101\\_enu.exe](#).

2. Run the downloaded file.

(By default, the SQL JDBC driver file will be extracted at `C:\Users\Administrator\Downloads\Microsoft SQL Server JDBC Driver 3.0.`)

3. Copy and paste the `C:\Users\Administrator\Downloads\Microsoft SQL Server JDBC Driver 3.0\sqljdbc_3.0\enu\sqljdbc4.jar` file to `$HIVE_HOME/lib` (where `$HIVE_HOME` can be set to `D:\hadoop\hive-0.9.0`).

## 1.3. Decide on Deployment Type

While it is possible to deploy all of HDP on a single host, this is appropriate only for initial evaluation.

In general you should use at least three hosts: one master host and two slaves.

## 1.4. Collect Information

To deploy your HDP installation, you need to collect the following information:

- The network resolvable name for each host in your cluster. This can be the IP address or the hostname.

To determine the hostname for a particular cluster host, open the command shell on that cluster host and execute **hostname**.



### Important

The installer will fail if it cannot resolve the hostname of each cluster node.

- (Optional): To be able to use an external database for Hive or Oozie metastore, ensure that you have the hostname (for an existing instance), database name, and user account credentials for the SQL Server instance.



### Note

Create two database users, oozimetastore and hivemetastore, in SQL Server and assign the following database roles to each: db\_datareader, db\_datawriter, db\_ddladmin, and public.

## 1.5. Prepare the Cluster Environment

To deploy HDP across a cluster, you need to prepare your multi-node cluster deploy environment. Follow these steps to ensure each cluster node is prepared to be an HDP cluster node:

- [Ensure that all cluster nodes use a common time](#)
- [Ensure that Windows networking uses IPv4 addresses](#)
- [Optional - Create Hadoop user](#)
- [Configure ports](#)

To use the remote push install and remote services management scripts in a Workgroup cluster, use the following section to set up networking configurations:

- [Enable networking configurations for Workgroups](#)

To use the remote push install and remote services management scripts in an Active Directory Domain cluster, use the following section to set up networking configurations:

- [Enable networking configurations for Active Directory Domains](#)

### 1.5.1. Configure common time for all cluster nodes

The clocks of all the nodes in your cluster must be able to synchronize with each other. To configure this for Windows Server, use the instructions provided [here](#).

### 1.5.2. Ensure that Windows networking uses IPv4 addresses

Configure all the Windows Server nodes in your cluster to use IPv4 addresses only. You can either disable IPv6 or set the preference to IPv4.

Use the following Microsoft KB article to disable IPv6: [How to disable IP version 6 or its specific components in Windows](#).

### 1.5.3. Optional - Create Hadoop user

HDP installer takes the following actions to create hadoop user for your environment:

- If user `hadoop` does not exist, HDP installer automatically creates a local user with random password.
- If the user `hadoop` already exists, HDP installer will change the current password to a new random password. The random password is passed on the command line throughout the install process, then discarded. Administrator can change the password later, but it must be done both in the user configuration and in the service objects installed on each machine via Service Manager.

### 1.5.4. Configure ports

HDP uses multiple ports for communication with clients and between service components. To enable this communication, you will need to either open all ports or the specific ports that HDP uses.

To open specific ports only, you can set the access rules in Windows.

For example, the following command will open up port 80 in the active Windows Firewall:

```
netsh advfirewall firewall add rule name=AllowRPCCommunication dir=in action=allow protocol=TCP localport=135
```

For example, the following command will open up ports 49152-65535 in the active Windows Firewall:

```
netsh advfirewall firewall add rule name=AllowRPCCommunication dir=in action=allow protocol=TCP localport=49152-65535
```

The tables below specify which ports must be opened for which ecosystem components to communicate with each other.

Make sure that appropriate ports are opened before you install HDP.

**HDFS Ports:** The following table lists the default ports used by the various HDFS services.

**Table 1.1. HDFS Ports**

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
NameNode WebUI	Master Nodes (NameNode and any back-up NameNodes)	50070	http	Web UI to look at current status of HDFS, explore file system	Yes (Typically admins, Dev/ Support teams)	<code>dfs.http.address</code>
		50470	https	Secure http service		<code>dfs.https.address</code>

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
NameNode metadata service		8020/9000	IPC	File system metadata operations	Yes (All clients who directly need to interact with the HDFS)	Embedded in URI specified by <code>fs.default.name</code>
DataNode	All Slave Nodes	50075	http	DataNode WebUI to access the status, logs etc.	Yes (Typically admins, Dev/ Support teams)	<code>dfs.datanode.http.address</code>
		50475	https	Secure http service		<code>dfs.datanode.https.address</code>
		50010		Data transfer		<code>dfs.datanode.address</code>
		50020	IPC	Metadata operations	No	<code>dfs.datanode.ipc.address</code>
Secondary NameNode	Secondary NameNode and any backup Secondary NameNode	50090	http	Checkpoint for NameNode metadata	No	<code>dfs.secondary.http.address</code>

**MapReduce Ports:** The following table lists the default ports used by the various MapReduce services.

**Table 1.2. MapReduce Ports**

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
JobTracker WebUI	Master Nodes (JobTracker Node and any back-up JobTracker node )	50030	http	Web UI for JobTracker	Yes	<code>mapred.job.tracker.http.address</code>
JobTracker	Master Nodes (JobTracker Node)	8021	IPC	For job submissions	Yes (All clients who need to submit the MapReduce jobs including Hive, Hive server, Pig)	Embedded in URI specified by <code>mapred.job.tracker</code>
Task-Tracker Web	All Slave Nodes	50060	http	DataNode Web UI to access	Yes (Typically admins, Dev/	<code>mapred.task.tracker.http.address</code>

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
UI and Shuffle				status, logs, etc.	Support teams)	
History Server WebUI		51111	http	Web UI for Job History	Yes	mapreduce.history.server.http.address

**Hive Ports:** The following table lists the default ports used by the Hive services.



**Note**

These services are not used in a standard HDP installation.

**Table 1.3. Hive Ports**

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
HiveServer2	HiveServer2 machine (Usually a utility machine)	10001	thrift	Service for programmatically connecting to Hive	Yes	ENV Variable HIVE_PORT
Hive Server	Hive Server machine (Usually a utility machine)	10000	thrift	Service for programmatically connecting to Hive	Yes Clients who need to connect to Hive either programmatically or through UI SQL tools that use JDBC)	ENV Variable HIVE_PORT
Hive Metastore		9083	thrift	Service for programmatically connecting to Hive metadata	Yes Clients that run Hive, Pig and potentially M/R jobs that use HCatalog)	hive.metastore.uris

**Templeton Port:** The following table lists the default port used by the Templeton service.

**Table 1.4. Templeton Port**

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
Templeton Server	Any utility machine	50111	http	Web API on top of HCatalog and other Hadoop services	Yes	templeton.port

## 1.5.5. Enable networking configurations for Workgroups

The MSI installation scripts and many utility scripts within HDP require Powershell scripts to be enabled, on every host machine in your Hadoop cluster. Furthermore, the utility scripts (for starting and stopping the whole cluster with a single command) provided with HDP, requires remote scripting and trust to be enabled. Therefore, we strongly recommend that you complete the following three settings on every host in your cluster.

You can set these in Active Directory via Group Policies (for a Group including all hosts in your Hadoop cluster), or you can execute the given Powershell commands on every host in your cluster.



### Important

Ensure that the Administrator account on the Windows Server node has a password. The remote scripting below will not work if the Administrator account has an empty password.

#### Enable remote scripting using Powershell commands

1. On each host in the cluster, execute the following commands in a Powershell window with "Run as Administrator" elevation:

```
Set-ExecutionPolicy "AllSigned"
```

```
Enable-PSRemoting
```

```
Set-item wsman:localhost\client\trustedhosts -value "Host1,Host2"
```

The last argument is a list of comma-separated hostnames in your cluster (for example, "HadoopHost1, HadoopHost2, HadoopHost3").

2. On each host in the cluster, execute the following commands in a Powershell window with "Run as Administrator" elevation:

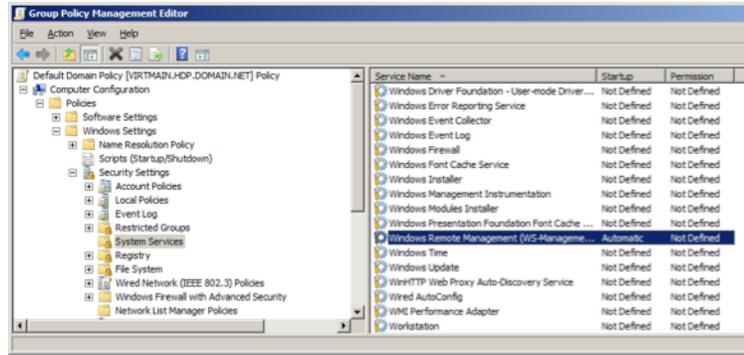
```
winrm quickconfig  
winrm set winrm/config/client '@{TrustedHosts="host1, host2, host3"}'
```

The last argument is a list of comma-separated hostnames in your cluster (for example, "HadoopHost1, HadoopHost2, HadoopHost3").

## 1.5.6. Enable networking configurations for Active Directory Domains

To enable remote scripting and to configure right domain policies for Windows Remote Management complete the following instructions on a domain controller machine (all actions are performed via **Group Policy Management\Default Domain Policy/Edit**):

1. Set the WinRM service to auto start.
  - Go to **Computer Configuration -> Policies -> Windows Settings -> Security Settings -> System Services -> Windows Remote Management (WS-Management)**.

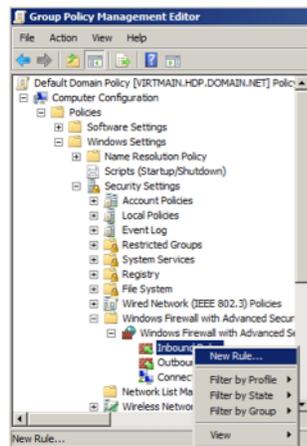


- Set Startup Mode to Automatic.

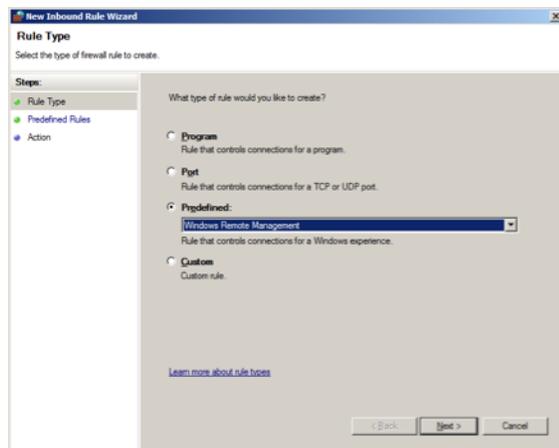


2. Add firewall exceptions to allow the service to communicate.

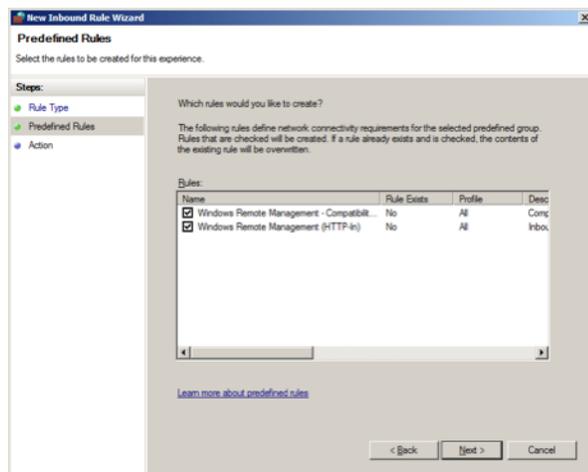
- Go to **Computer Configuration -> Policies -> Windows Settings -> Security Settings -> Windows Firewall with Advanced Security** .
- Right click on **Windows Firewall with Advanced Security** to create a new Inbound Rule.



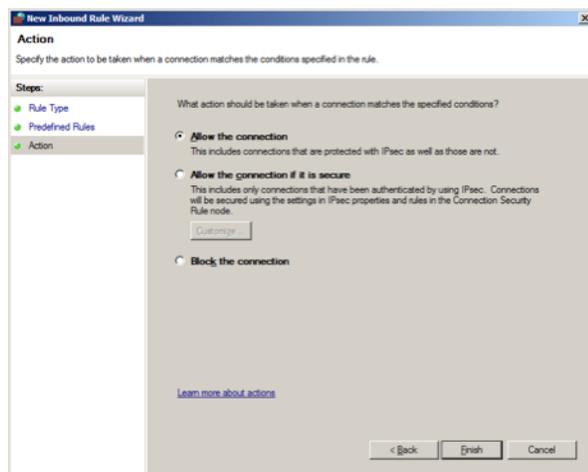
- Select the type of rule as **Predefined as Windows Remote Management** .



The Predefined rule will automatically create two rules as shown below:

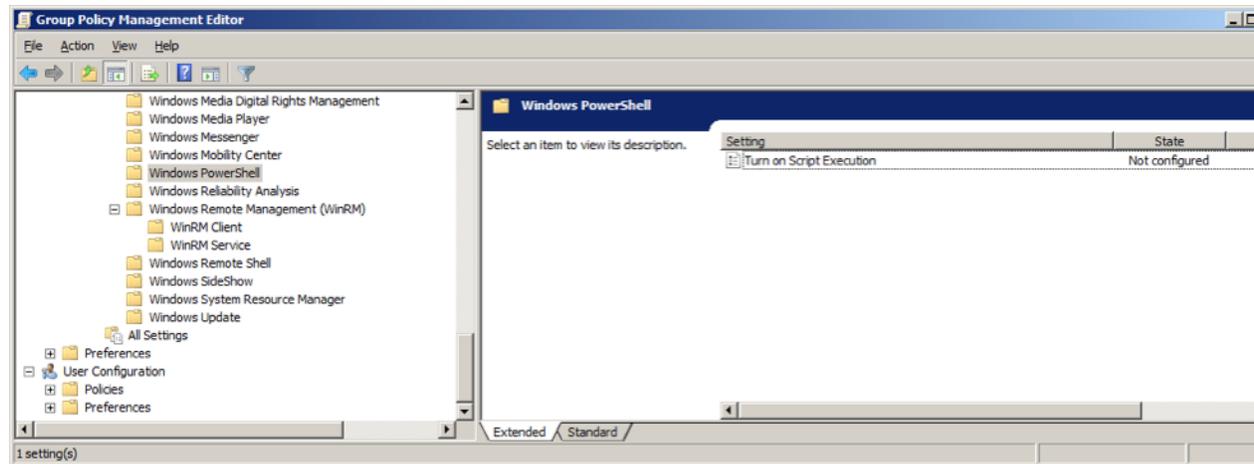


- Configure the Action as **Allow the connection** and click **Finish**.

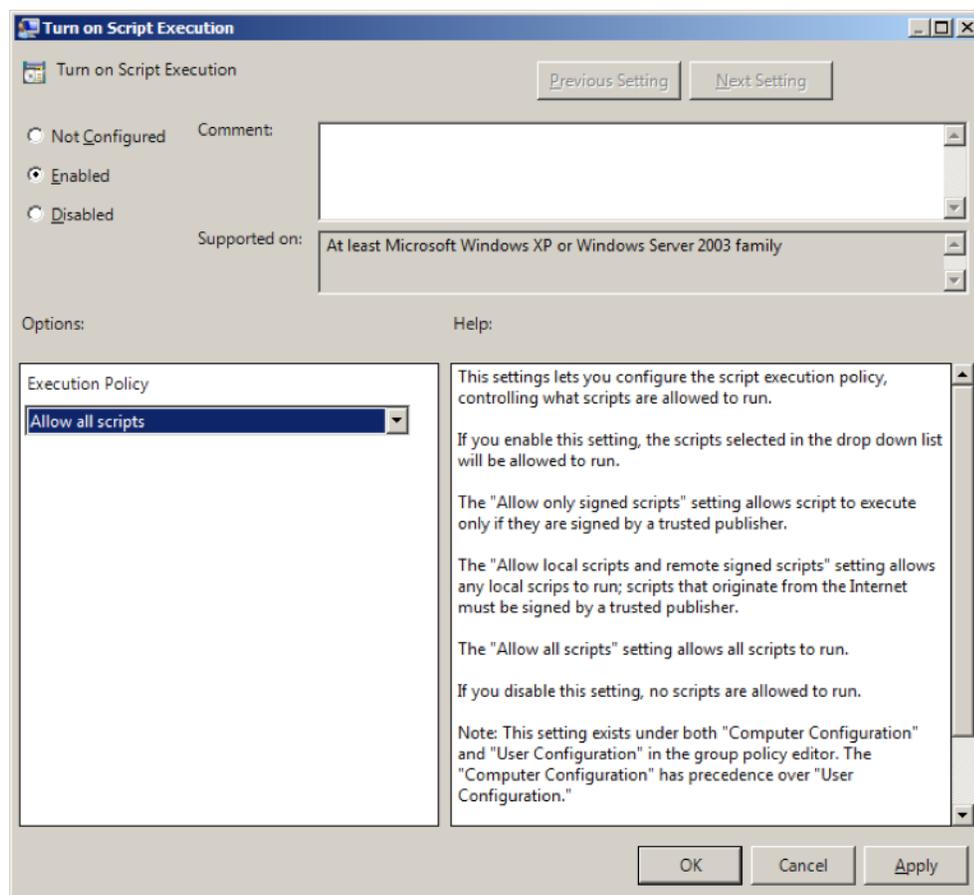


3. Set script execution policy.

- Go to **Computer Configuration -> Policies -> Administrative Templates -> Windows Components -> Windows PowerShell** .
- **Enable Script Execution** .

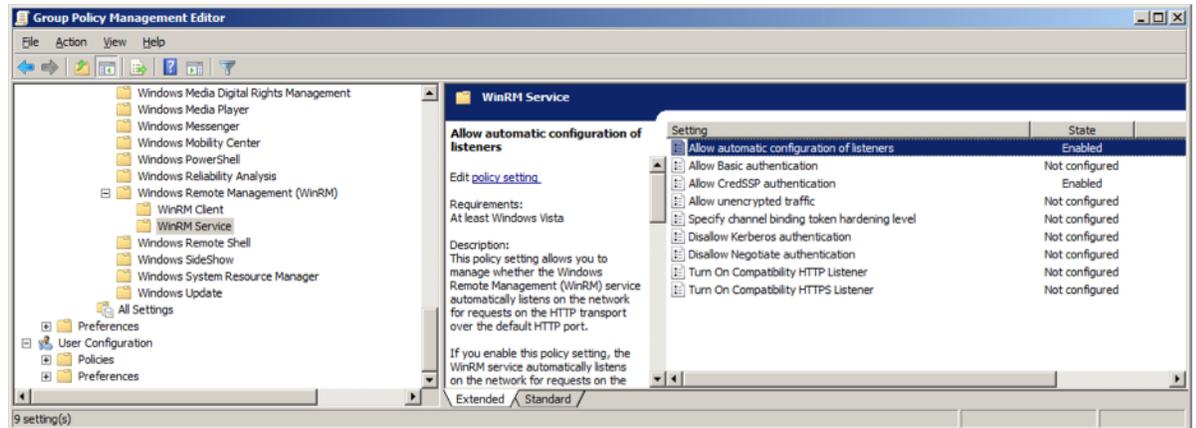


- **Set Execution Policy to Allow all scripts.**

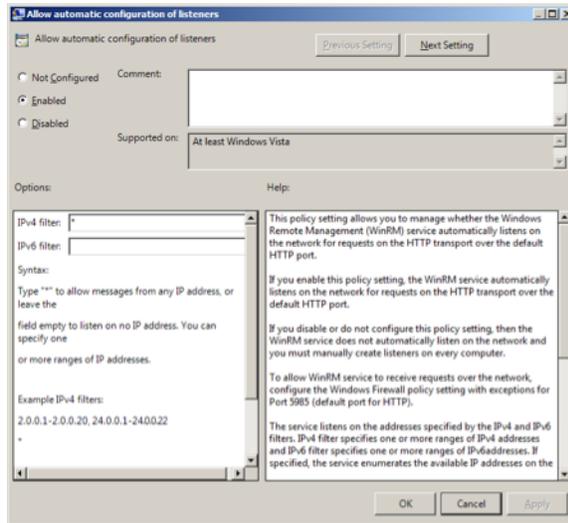


#### 4. Setup WinRM service.

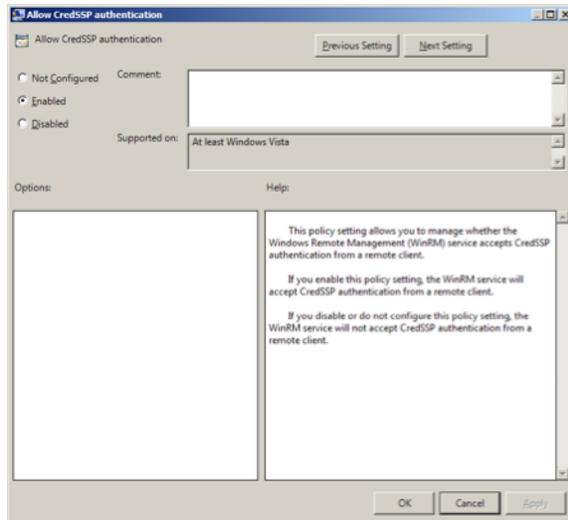
- Go to **Computer Configuration -> Policies -> Administrative Templates -> Windows Components -> Windows Remote Management (WinRM) -> WinRM Service**.



- Create a WinRM listener.
  - a. To allow automatic configuration of listeners, select **Enabled**.
  - b. Set **IPv4 filter** to \* (all addresses or specify range)

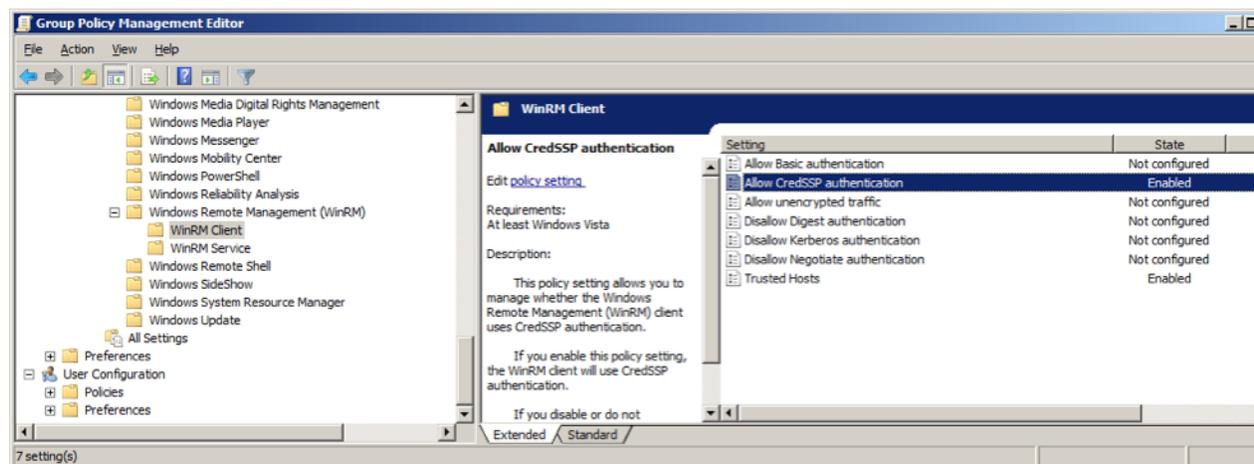


- c. Allow CredSSP authentication and click OK.

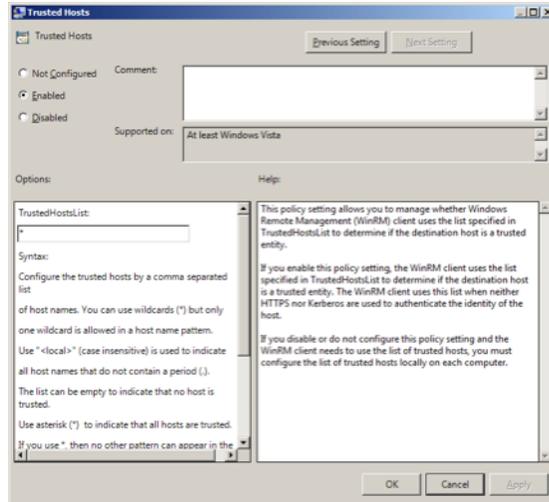


## 5. Setup WinRM client.

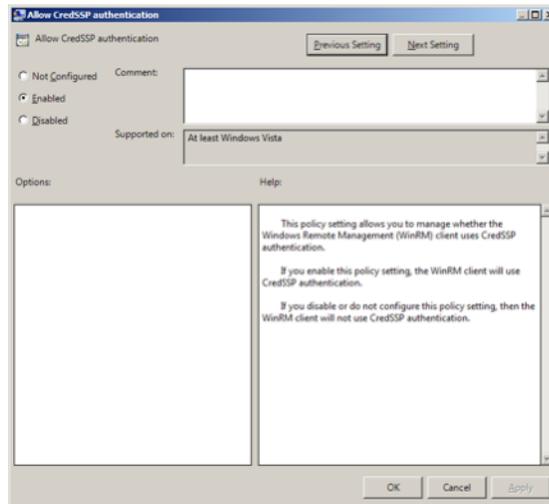
- Go to **Computer Configuration -> Policies -> Administrative Templates -> Windows Components -> Windows Windows Remote Management (WinRM) -> WinRM Client**.



- Configure the trusted host list (the IP addresses of the computers that can initiate connections to the WinRM service). To do this, set **TrustedHostsList** to \* (all addresses or specify range).

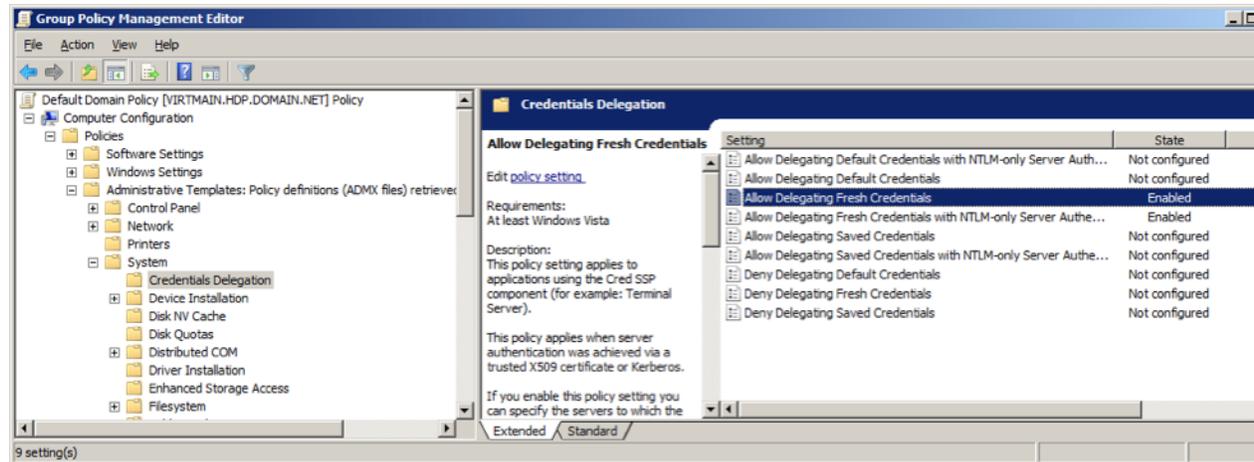


- Allow CredSSP authentication and click OK.

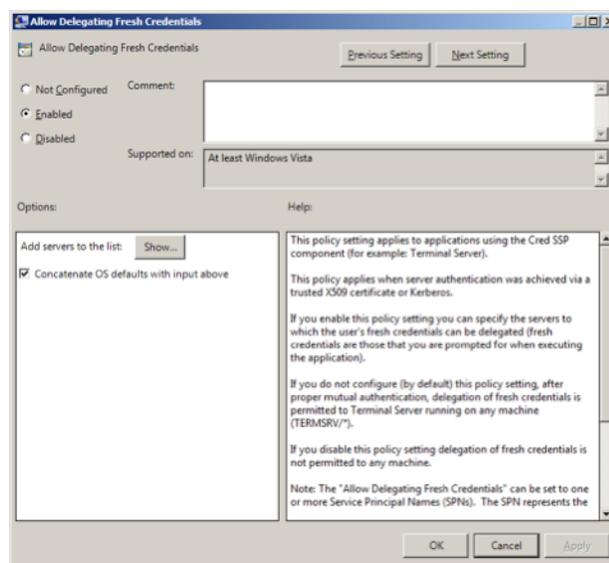


6. Enable credentials delegation.

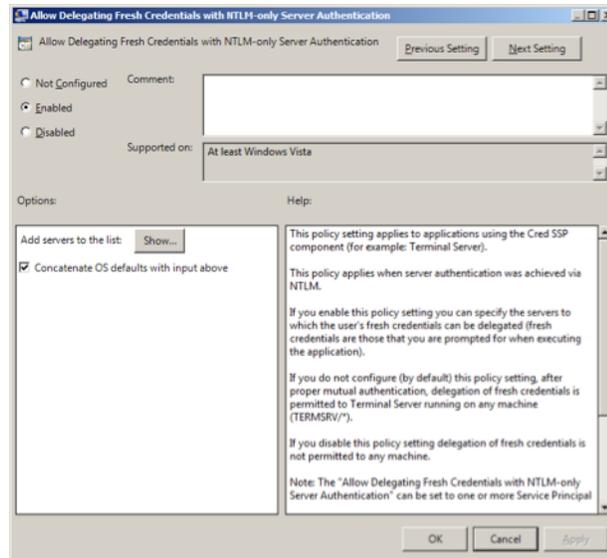
- Go to **Computer Configuration -> Policies -> Administrative Templates -> System -> Credentials Delegation**.



- Select **Enabled** to allow delegation fresh credentials.
- Under **Options** click on **Show**. Set **WSMAN** to \* (all addresses or specify range). Click on **Next Setting**.

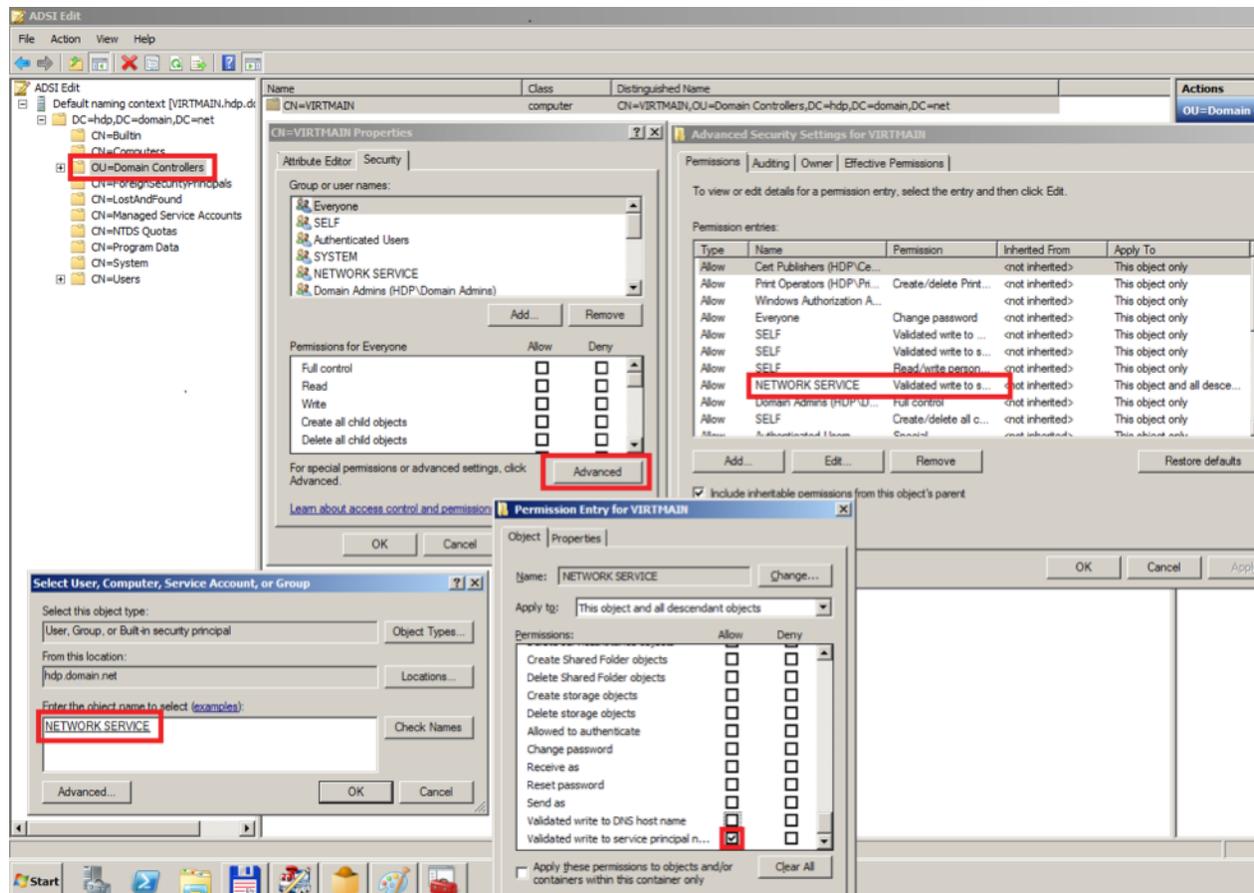


- Select **Enabled** to allow delegation fresh credentials with NTLM-only server authentication.
- Under **Options** click on **Show**. Set **WSMAN** to \* (all addresses or specify range). Click on **Finish**.



#### 7. Enable creating WSMAN SPN.

- Go to **Start-> Run**. In the dialog box, type `ADSIEdit.msc` and click **Enter**.
- Expand **OU=Domain Controllers** menu item and select **CN=domain controller hostname**. Go to **Properties -> Security -> Advanced -> Add**.
- Enter **NETWORK SERVICE**, click **Check Names**, then **Ok**. In the Permission Entry select **Validated write to service principal name**. Click **Allow** and **OK** to save your changes.



## 8. Restart WinRM service and update policies.

- On the domain controller machine, execute the following commands in PowerShell:

```
Restart-Service WinRM
```

- On other hosts in domain, execute the following commands:

```
gpupdate /force
```

- Ensure that SPN-s WSMAN is created for your environment. Execute the following command on your domain controller machine:

```
setspn -l $Domain_Controller_Hostname
```

You should see output similar to the following:

```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) 2009 Microsoft Corporation. All rights reserved.

PS C:\Users\Administrator> Restart-Service WinRM
PS C:\Users\Administrator> setspn -l UIRTMMAIN
Registering SPN: UIRTMMAIN/hdp.domain.net for CN=UIRTMMAIN,OU=Domain Controllers,DC=hdp,DC=domain,DC=net:
WSMAN/UIRTMMAIN
WSMAN/UIRTMMAIN.hdp.domain.net
E31B6C55EB04/UIRTMMAIN.hdp.domain.net
ldap/UIRTMMAIN.hdp.domain.net/ForestDnsZones.hdp.domain.net
ldap/UIRTMMAIN.hdp.domain.net/DomainDnsZones.hdp.domain.net
DNS/UIRTMMAIN.hdp.domain.net
GC/UIRTMMAIN.hdp.domain.net/hdp.domain.net
RestrictedKrbHost/UIRTMMAIN.hdp.domain.net
RestrictedKrbHost/UIRTMMAIN
HOST/UIRTMMAIN/HDP
HOST/UIRTMMAIN.hdp.domain.net/HDP
HOST/UIRTMMAIN
HOST/UIRTMMAIN.hdp.domain.net
HOST/UIRTMMAIN.hdp.domain.net/hdp.domain.net
E314235-4B06-11D1-AB04-00C04FC2DCD2/e6665522-1123-472f-b422-bd2d496c734e/hdp.domain.net
ldap/UIRTMMAIN/HDP
ldap/e6665522-1123-472f-b422-bd2d496c734e._msdcs.hdp.domain.net
ldap/UIRTMMAIN.hdp.domain.net/HDP
ldap/UIRTMMAIN
ldap/UIRTMMAIN.hdp.domain.net
ldap/UIRTMMAIN.hdp.domain.net/hdp.domain.net
PS C:\Users\Administrator>
```

9. Check the WSMAN SPN on other host in domain. Execute the following command on any one of your host machines:

```
setspn -l $Domain_Controller_Hostname
```

You should see output similar to the following:

```
Administrator: Windows PowerShell
ldap/UIRTMMAIN.hdp.domain.net/hdp.domain.net
PS C:\Users\Administrator> setspn -l UIRTMMAIN
Registering SPN: UIRTMMAIN/hdp.domain.net for CN=UIRTMMAIN,OU=Domain Controllers,DC=hdp,DC=domain,DC=net:
WSMAN/UIRTMMAIN
WSMAN/UIRTMMAIN.hdp.domain.net
RestrictedKrbHost/UIRTMMAIN
RestrictedKrbHost/UIRTMMAIN.hdp.domain.net
HOST/UIRTMMAIN/HDP
HOST/UIRTMMAIN.hdp.domain.net/HDP
HOST/UIRTMMAIN
HOST/UIRTMMAIN.hdp.domain.net
PS C:\Users\Administrator>
```

## 1.6. Define Cluster Configuration

Use the following instructions to configure HDP installer for your cluster:

1. Create a `clusterproperties.txt` file.
2. Add the properties to the `clusterproperties.txt` file as described in the table given below:



### Important

- Ensure that all the properties in the `clusterproperties.txt` file are separated by a new line character.
- Ensure that the directory paths do not contain any whitespace character.

For example, `C:\Program Files\Hadoop` is an invalid directory path for HDP.

- Use Fully Qualified Domain Names (FQDN) for specifying the network host name for each cluster host. The FQDN is a DNS name that uniquely identifies the computer on the network. By default, it is a concatenation of the host name, the primary DNS suffix, and a period.

- When specifying the host lists in the `clusterproperties.txt` file, if the hosts are multi-homed or have multiple NIC cards, make sure that each name or IP address by which you specify the hosts are the preferred name or IP address by which the hosts can communicate among themselves. In other words, these should be the addresses used internal to the cluster, not those used for addressing cluster nodes from outside the cluster.

**Table 1.5. Configuration values for MSI installer**

Configuration Property Name	Description	Example value	Mandatory/Optional/Conditional
HDP_LOG_DIR	HDP's operational logs will be written to this directory on each cluster host. Ensure that you have sufficient disk space for storing these log files.	d:\hadoop\logs	Mandatory
HDP_DATA_DIR	HDP data will be stored in this directory on each cluster node. You can add multiple comma-separated data locations for multiple data directories.	d:\hdp\data	Mandatory
NAMENODE_HOST	The FQDN for the cluster node that will run the NameNode master service.	NAMENODE_MASTER.acme.com	Mandatory
SECONDARY_NAMENODE_HOST	The FQDN for the cluster node that will run the Secondary NameNode master service.	SECONDARY_NN_MASTER.acme.com	Mandatory
JOBTRACKER_HOST	The FQDN for the cluster node that will run the JobTracker master service.	JOBTRACKER_MASTER.acme.com	Mandatory
HIVE_SERVER_HOST	The FQDN for the cluster node that will run the Hive Server master service.	HIVE_SERVER_MASTER.acme.com	Mandatory
OOZIE_SERVER_HOST	The FQDN for the cluster node that will run the Oozie Server master service.	OOZIE_SERVER_MASTER.acme.com	Mandatory
TEMPLETON_HOST	The FQDN for the cluster node that will run the Templeton master service.	TEMPLETON_MASTER.acme.com	Mandatory
SLAVE_HOSTS	A comma separated list of FQDN for those cluster nodes that will run the DataNode and TaskTracker services.	slave1.acme.com, slave2.acme.com, slave3.acme.com	Mandatory
DB_FLAVOR	Database type for Hive and Oozie metastores (allowed databases are SQL Server and Derby). To use default embedded Derby instance, set the value of this property to <code>derby</code> . To use an existing SQL Server instance as the metastore DB, set the value as <code>mssql</code> .	mssql or derby	Mandatory
DB_HOSTNAME	FQDN for the node where the metastore database	sqlserver1.acme.com	Mandatory

Configuration Property Name	Description	Example value	Mandatory/Optional/Conditional
	service is installed. If using SQL Server, set the value to your SQL Server hostname. If using Derby for Hive metastore, set the value to HIVE_SERVER_HOST.		
DB_PORT	This is an optional property required only if you are using SQL Server for Hive and Oozie metastores. By default, database port is set to 1433.	1433	
HIVE_DB_NAME	Database for Hive metastore. If using SQL Server, ensure that you create the database on the SQL Server instance.	hivedb	Mandatory
HIVE_DB_USERNAME	User account credentials for Hive metastore database instance. Ensure that this user account has appropriate permissions.	hive_user	Mandatory
HIVE_DB_PASSWORD		hive_pass	Mandatory
OOZIE_DB_NAME	Database for Oozie metastore. If using SQL Server, ensure that you create the database on the SQL Server instance.	ooziedb	Mandatory
OOZIE_DB_USERNAME	User account credentials for Oozie metastore database instance. Ensure that this user account has appropriate permissions.	oozie_user	Mandatory
OOZIE_DB_PASSWORD		oozie_pass	Mandatory

The following snapshot illustrates a sample `clusterproperties.txt` file:

```
#Log directory
HDP_LOG_DIR=d:\hadoop\logs

#Data directory
HDP_DATA_DIR=d:\hdp\data

#Hosts
NAMENODE_HOST=NAMENODE_MASTER.acme.com
SECONDARY_NAMENODE_HOST=SECONDARY_NAMENODE_MASTER.acme.com
JOBTRACKER_HOST=JOBTRACKER_MASTER.acme.com
HIVE_SERVER_HOST=HIVE_SERVER_MASTER.acme.com
OOZIE_SERVER_HOST=OOZIE_SERVER_MASTER.acme.com
TEMPLETON_HOST=TEMPLETON_MASTER.acme.com
SLAVE_HOSTS=slave1.acme.com, slave2.acme.com, slave3.acme.com

#Database host
DB_FLAVOR=derby
DB_HOSTNAME=DB_myHostName

#Hive properties
HIVE_DB_NAME=hive
HIVE_DB_USERNAME=hive
HIVE_DB_PASSWORD=hive

#Oozie properties
```

```
OOZIE_DB_NAME=oozie  
OOZIE_DB_USERNAME=oozie  
OOZIE_DB_PASSWORD=oozie
```

## 2. Quick Start Guide for Single Node HDP Installation

Use the following instructions to deploy HDP on a single node Windows Server machine:

### 1. Install the necessary prerequisites

#### a. Install **Microsoft Visual C++ 2010 Redistributable Package (64 bit)**

Use the instructions provided [here](#) to download and install Microsoft Visual C++ 2010 Redistributable Package (64 bit).

#### b. Install **Microsoft.NET framework 4.0**

Use the instructions provided [here](#) to download and install Microsoft.NET framework 4.0.

#### c. Install **Java JDK 6u31**

Use the following instructions to install the JDK:

- i. Download the [Oracle JDK](#) and install to a directory path that has no whitespace characters in its path. For example, "C:\Program Files\Java\" is not a valid path. "C:\Software\Java\" is a valid path.
- ii. Create a system variable named `JAVA_HOME`. The value of this variable will be the full path to installation directory for JDK.

A. Open the **Control Panel** -> **System** pane and click on **Advanced system settings**.

B. Click on the **Advanced** tab.

C. Click the **Environment Variables** button.

D. Under **System variables**, click **New**.

E. Enter the **Variable Name** as `JAVA_HOME`.

F. Enter the **Variable Value**, as the installation path for the Java Development Kit.

For example, if your JDK is installed at `C:\Software\Java\jdk1.6.0_31`, then you must provide this path to the **Variable Value**.

G. Click **OK**.

H. Click **Apply Changes**.

#### d. Install **Python 2.7**

- i. Download Python from [here](#).

- ii. Update the `PATH` environment variable. Using Administrator privileges:

- A. Open the **Control Panel** -> **System** pane and click on the **Advanced system settings** link.
- B. Click on the **Advanced** tab.
- C. Click the **Environment Variables** button.
- D. Under **System Variables**, find `PATH` and click **Edit**.
- E. In the Edit windows, modify `PATH` by appending the installation path for your Python directory to the value of `PATH`.

For example, if Python executable is installed at `C:\Python27\` then you must append this value to `PATH`.

- F. To validate your settings, from a command shell or Powershell window, type:

```
python
```

## 2. Prepare the single node machine

- a. Collect Information - Get the hostname of the server you are installing on.

To determine the hostname for a particular cluster host, open the command shell on that cluster host and execute `hostname`. The output of the command is the exact hostname you will use to identify the cluster machine.

```
> hostname  
WIN-RT345SERVER
```

- b. Configure Firewall

HDP uses multiple ports for communication with clients and between service components.

If your corporate policies require maintaining per server firewall, you must enable the ports listed [here](#). Use the following command to open these ports:

```
netsh advfirewall firewall add rule name=AllowRPCCommunication dir=in  
action=allow protocol=TCP localport=$PORT_NUMBER
```

- For example, the following command will open up port 80 in the active Windows Firewall:

```
netsh advfirewall firewall add rule name=AllowRPCCommunication dir=in  
action=allow protocol=TCP localport=135
```

- For example, the following command will open ports all ports from 49152 to 65535. in the active Windows Firewall:

```
netsh advfirewall firewall add rule name=AllowRPCCommunication dir=in  
action=allow protocol=TCP localport=49152-65535
```

If your networks security policies allow you open all the ports, use the following instructions to turn off Windows Firewall: [http://technet.microsoft.com/en-us/library/cc766337\(v=ws.10\).aspx](http://technet.microsoft.com/en-us/library/cc766337(v=ws.10).aspx)

### 3. Specify the configuration for HDP on the single node machine

Create a `clusterproperties.txt` file.

The `clusterproperties.txt` is a text file and contains parameter definitions (like the hostnames of the nodes in your cluster, the roles for each of them, etc.).

Start creating the `clusterproperties.txt` file by copying the example text provided at the bottom of this section and modify it according to the hostname of your machine

The following snapshot illustrates a sample `clusterproperties.txt` file:

```
#Log directory
HDP_LOG_DIR=c:\hadoop\logs

#Data directory
HDP_DATA_DIR=c:\hdp\data

#Hosts (Roles for the host machines in your cluster)
NAMENODE_HOST=${Hostname for your single node cluster}
SECONDARY_NAMENODE_HOST=${Hostname for your single node cluster}
JOBTRACKER_HOST=${Hostname for your single node cluster}
HIVE_SERVER_HOST=${Hostname for your single node cluster}
OOZIE_SERVER_HOST=${Hostname for your single node cluster}
TEMPLETON_HOST=${Hostname for your single node cluster}
SLAVE_HOSTS=${Hostname for your single node cluster}

#Database host
DB_FLAVOR=derby
DB_HOSTNAME=${Hostname for your single node cluster}

#Hive properties
HIVE_DB_NAME=hive
HIVE_DB_USERNAME=hive
HIVE_DB_PASSWORD=hive

#Oozie properties
OOZIE_DB_NAME=oozie
OOZIE_DB_USERNAME=oozie
OOZIE_DB_PASSWORD=oozie
```

More details on the `clusterproperties.txt` configuration options are available [here](#).

### 4. Install and start HDP

- a. Download the HDP for Windows MSI from [here](#).

Open a command prompt with Administrator privileges and execute the MSI installer command:

```
msiexec /i "<$MSI_PATH>" /lv "<$PATH_to_Installer_Log_File>"  
HDP_LAYOUT="<$PATH_to_clusterproperties.txt_File>" HDP_DIR="<  
$PATH_to_HDP_Install_Dir>" DESTROY_DATA="<Yes_OR_No>"
```

The following example illustrates the command, with parameters, to launch the installer:

```
msiexec /i "hdp-win-1.1.msi" /lv "hdp.log" HDP_LAYOUT="C:\config\  
clusterproperties.txt" HDP_DIR="C:\hdp\hadoop" DESTROY_DATA="no"
```

As shown in the example above, the following command parameters should be modified to match your files and directories:

- HDP\_LAYOUT: Absolute path to cluster properties file
- HDP\_DIR: Install directory for HDP
- DESTROY\_DATA: Whether to preserve or delete existing HDP data

b. Start all HDP services on the single machine.

In a command prompt, navigate to the HDP install directory. This is the 'HDP\_DIR' setting from the msiexec command.

Then execute 'start\_local\_hdp\_services.cmd'.

```
cd <$PATH_TO_HDP_DIR>  
start_local_hdp_services
```

c. Validate the install by running the full suite of smoke tests.

```
Run-SmokeTests
```

## 3. Deploying HDP

Use any one of the following options to deploy Hadoop cluster in your environment:

- [Option I: Central push install using corporate standard procedures](#)
- [Option II: Central push install using provided script](#)
- [Option III: Manual Install one node at a time](#)

### 3.1. Option I - Central Push Install Using Corporate Standard Procedures

Many Windows Data Centers have standard corporate procedures for performing centralized push-install of software packages to hundreds or thousands of computers at the same time. In general, these same procedures will work for doing a centralized push-install of HDP to a Hadoop cluster.

If your Data Center already has such procedures in place, then follow this simple checklist:

1. Identify the hosts that will constitute the Hadoop cluster nodes, and configure them for centralized push-install, according to your standard procedures.
2. Complete all the prerequisites provided in the section [Prepare the Environment](#).



#### Note

In many cases, your standard procedures will cover all the suggestions in this section. In some cases, you may need to make additional configuration changes for Hadoop to run correctly. In particular, you will want to enable Remote Scripting from your administrative Master Node to the other nodes in the cluster, since many Powershell scripts are used for Hadoop cluster management and administration. To enable Remote Scripting, see the instructions provided [here](#).

3. Extract the HDP-Win zip folder from [here](#).

Identify the MSI and sample `clusterproperties.txt` file.



#### Note

Downloaded MSI includes full source, binary-only MSI, and documentation for all components. Hortonworks recommends using the binary-only MSI for faster downloads.

4. Create your own custom `clusterproperties.txt` file, following the instructions in the section [Configure the HDP Installer](#).



### Important

When specifying the host lists in the `clusterproperties.txt` file, if the hosts are multi-homed or have multiple NIC cards, make sure that each name or IP address by which you specify the hosts are the preferred name or IP address by which the hosts can communicate among themselves. In other words, these should be the addresses used internal to the cluster, not those used for addressing cluster nodes from outside the cluster.

5. Using your standard procedures, push both the MSI and the custom `clusterproperties.txt` file to each node in the cluster. Alternatively, you can place the MSI and the `clusterproperties.txt` file in a network share location accessible via CIFS file path from each node in the cluster.



### Note

Ensure that you place the two files together in the same target directory on each node.

6. Continuing to use your standard procedures, remotely execute on each node the `msiexec` command documented in section [Install from the MSI \[32\]](#). This will cause the MSI to install itself, using the parameters in the `clusterproperties.txt` file.
7. Examine the return results and/or logs from your standard procedures to ensure that all nodes were successfully installed.
8. Smoke test your installation using the instructions provided in the [Validate the Install \[33\]](#) section.

## 3.2. Option II - Central Push Install Using Provided Script

If your Data Center does not have established procedures for doing centralized push-install, then you may either follow the below suggested procedure, or you may avoid the issue of centralized installation by manually installing HDP on each individual node in the cluster, documented in section [Installing HDP manually, one node at a time](#).

You can choose to use the helper install script `push_install_hdp.ps1`. The `push_install_hdp.ps1` script only installs one machine at a time, and does all the machines in the cluster in sequence. This is sufficient for a small test cluster, and it does not require any shared storage. It pushes the files to each node using that node's "Administrative Shares", so make sure that the Admin Shares feature is turned on for all hosts in the cluster, and the Administrator user id you will use to run the install script has privileges to write to the Admin Shares on the other hosts.

1. Identify the hosts that will constitute the Hadoop cluster nodes and follow the instructions in section [Prepare the Environment](#) to configure them all for centralized push-install.
2. Extract the HDP-Win zip folder from [here](#) and identify the MSI and example `clusterproperties.txt` file.



## Note

Downloaded MSI includes full source, binary-only MSI, and documentation for all components. Hortonworks recommends using the binary-only MSI for faster downloads.

3. Create your own custom `clusterproperties.txt` file using the instructions in section [Configure the HDP Installer](#).
4. Place the MSI and custom `clusterproperties.txt` file in a convenient local subdirectory on the host from which you are running the push install.
  - **Advanced:** You may also place in this subdirectory other files that you want pushed to all the cluster hosts.
5. On the master install node, run a command prompt in Administrator mode (to use Administrator privileges), and execute the install scripts with the following parameters:
  - `source_path`: Local or network share directory path where all the installable files reside, including the MSI, the `clusterproperties.txt` file, and any other files you want pushed to the cluster nodes.
  - `target_path`: A single absolute directory path, the same for all target machines, for the install directory (`$HADOOP_NODE`). The path should be specified as it would be locally on a target host. For example, `D:\hadoop\`.
  - `clusterpropfile`: Simple file name of the custom `clusterproperties.txt` file.

The `clusterproperties.txt` resides in the `$source_path` directory. So, ensure that you provide the filename (without any path) as the value for the `clusterpropfile` parameter.

- `files_list`: A single string containing a comma-delimited list of all simple file names to be pushed to all nodes.

Typically these will include the MSI file, the custom `clusterproperties.txt` file, etc.

Ensure that all files are in the `$source_path` directory. These files will be copied to `$target_path` directory.

You may not input paths for these files, only simple names. Ensure that you include the custom `clusterproperties.txt` file.

- `msiexec` command line token sequence: Complete `msiexec` command must be provided just as though you were entering it on a PowerShell command line of one of the target servers, after the files have been pushed.

Start with the word `msiexec` and provide the entire command line, each argument in its own string object.

The `msiexec` command line must be constructed by referring to section [Install from the MSI \[32\]](#).

- Ensure that `<$MSI_PATH>` and `<$PATH_to_clusterproperties.txt_File>` arguments are specified as simple filenames (assume that the `msiexec` command will be run from the context of the `target_path`).
- The `<$PATH_to_Installer_Log_File>` argument may be an absolute path or will be interpreted relative to the `target_path`.
- The `HDP_DIR` argument is different from the `target_path`:
  - `target_path` is the location on each host where the MSI and `clusterproperties.txt` files will be copied to while preparing for the installation.
  - `HDP_DIR` is the location on each host where the HDP components will actually be installed by `msiexec`.

The syntax for `msiexec` command line token sequence parameter is given below:

```
msiexec /i "<$MSI_PATH>" /lv "<$PATH_to_Installer_Log_File>"  
HDP_LAYOUT="<$PATH_to_clusterproperties.txt_File>" HDP_DIR="<  
$PATH_to_HDP_Install_Dir>" DESTROY_DATA="<Yes_OR_No>"
```



### Note

The `push_install_hdp.ps1` script, will deploy the MSI and other files to each host in the cluster by writing (pushing) them from the install master host to the administrative share corresponding to the `target_path` argument on each host. (e.g., a target path of `D:\hadoop\` will cause the admin share `\\hostname\D:\hadoop\` to be written to on each host.)

6. The installer script will return error messages or successful completion results to the Install Master host. These messages will be printed out at the end of the script execution. Examine these return results to ensure that all nodes were successfully installed.
7. Smoke test your installation using the instructions provided in the [Validate the Install \[33\]](#) section.

## 3.3. Option III - Manual Install One Node At A Time

Use the following instructions to deploy Hadoop using HDP:

1. Complete all the prerequisites provided [here](#).
2. Download the HDP for Windows MSI from [here](#).



### Note

Downloaded MSI includes full source, binary-only MSI, and documentation for all components. Hortonworks recommends using the binary-only MSI for faster downloads.

3. Use the instructions provided [here](#) to complete the configuration for HDP installer.
4. Install from MSI.
  - a. Launch the MSI installer with the `clusterproperties.txt` file created previously.



### Important

This MSI must be executed on each and every cluster node and must use the same `clusterproperties.txt` file.

- b. On each node, run a command prompt in Administrator mode (to use Administrator privileges), and execute the following command:

```
msiexec /i "<$MSI_PATH>" /lv "<$PATH_to_Installer_Log_File>"  
HDP_LAYOUT="<$PATH_to_clusterproperties.txt_File>" HDP_DIR="<  
$PATH_to_HDP_Install_Dir>" DESTROY_DATA="<Yes_OR_No>"
```

Ensure that you provide appropriate values for the following mandatory command line option:

- **HDP\_LAYOUT:** Mandatory parameter. Provide location of the `clusterproperties.txt` file (For example, `d:\config\clusterproperties.txt`).



### Important

The path to the `clusterproperties.txt` file must be absolute. Relative paths will not work.

Optionally, you can also use the following command line options:

- **HDP\_DIR:** Install directory for HDP (For example, `d:\hdp`). Default value is `<$Default_Drive>/hdp`.
- **DESTROY\_DATA:** Specifies whether to preserve or delete existing data in target data directories (allowed values are `undefined`(default), `yes`, and `no`).

The `DESTROY_DATA` parameter takes care of the following conditions:

- During installation, when `HDP_DATA_DIR` has data from previous installation if `DESTROY_DATA` is set to `undefined`, installation will fail.
- During installation, if `DESTROY_DATA` is set to `no`, the installer will reuse the existing data and would not format the NameNode.



### Note

Installer does not check for the data correctness.

- During installation, if `DESTROY_DATA` is set to `yes`, installation will remove previous data and format the NameNode.

- During installation, if no data exists in `$HDP_DATA_DIR` then the `HDP_DATA_DIR` is created irrespective of the value of `DESTROY_DATA` and NameNode is formatted.

The following example illustrates the command to launch the MSI installer:

```
msiexec /i "hdp-win-1.1.msi" /lv "hdp.log" HDP_LAYOUT="D:\config\  
clusterproperties.txt" HDP_DIR="D:\hdp\hadoop" DESTROY_DATA="no"
```

- (Optional):** Configure compression for HDFS. Download the `zlib1.dll` from [here](#). Copy the downloaded file to either `$HADOOP_HOME\lib\native` or to `C:\Windows\System32`. To use `GzipCodec`, ensure that you copy the downloaded file to `C:\Windows\System32`.

#### 5. Validate the install.

- Use the instructions provided [here](#) to start the HDP Services.
- On a cluster node, open a command shell and execute the smoke test command script as shown below:

```
cd %HADOOP_NODE_INSTALL_ROOT%  
Run-SmokeTests
```

The smoke tests validate the installed functionality by executing a set of tests for each HDP component.



#### Note

It is recommended to re-install HDP, if you see installation failures for any HDP component.

## 3.4. Optional - Install Client Host

A client host has all the HDP JAR files on it for communicating with Hive, HDFS, etc. Note that you will not find any HDP service running on the client host machine.

Use the following instructions to install a client host:

- Copy existing `clusterproperties.txt` file from any host machine in your cluster.
- Run the HDP installer from the client host. Execute the following command on your client host machine:

```
msiexec /i "<$MSI_PATH>" /lv "<$PATH_to_Installer_Log_File>" HDP_LAYOUT="<  
$PATH_to_clusterproperties.txt_File>" HDP_DIR="<$PATH_to_HDP_Install_Dir>"  
DESTROY_DATA="<Yes_OR_No>"
```

Ensure that you provide appropriate values for the following mandatory command line option:

- **HDP\_LAYOUT:** Mandatory parameter. Provide location of the copied `clusterproperties.txt` file on your client host machine (For example, `d:\config\clusterproperties.txt`).



### Important

The path to the `clusterproperties.txt` file must be absolute. Relative paths will not work.

Optionally, you can also use the following command line options:

- **HDP\_DIR:** Install directory for HDP (For example, `d:\hdp`). Default value is `<$Default_Drive>/hdp`.
- **DESTROY\_DATA:** Specifies whether to preserve or delete existing data in target data directories (allowed values are `undefined`(default), `yes`, and `no`).

## 4. Managing HDP on Windows

This section will describe how to manage HDP on Windows services.

### 4.1. Starting the HDP Services

The HDP on Windows installer sets up Windows services for each HDP component across the nodes in a cluster. Use the instructions given below to start HDP services from any host machine in your cluster.

Complete the following instructions as the administrative user:

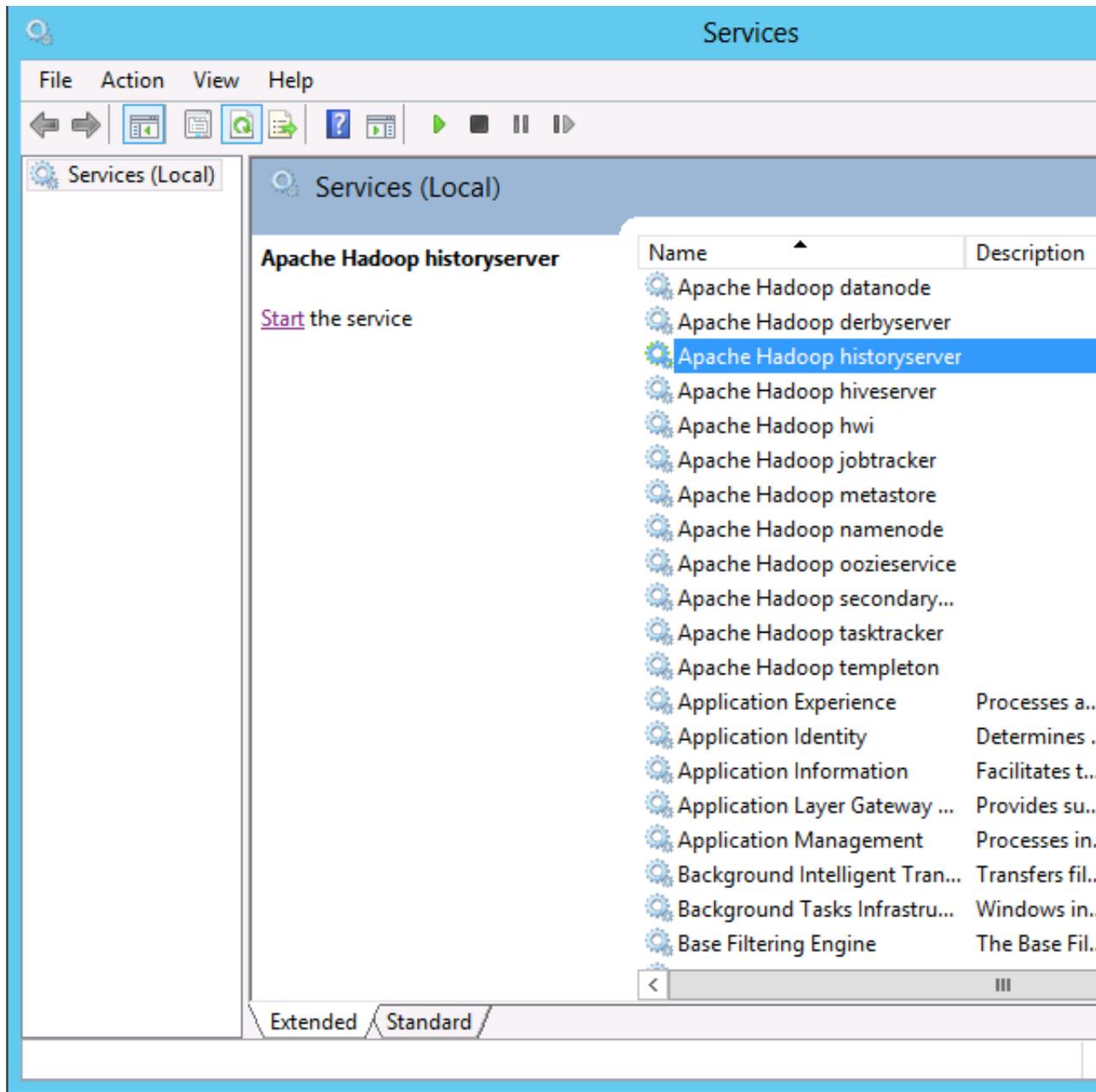
1. Start the HDP cluster.

Navigate to the install directory (as specified by the `HADOOP_NODE_INSTALL_ROOT` environment variable) and execute the following command from any host machine in your cluster.

```
cd %HADOOP_NODE_INSTALL_ROOT%  
start_remote_hdp_services.cmd
```

2. Open the **Services** administration pane.

Against all the services that are installed successfully, you should see the following message as highlighted in the screenshot below:



## 4.2. Stopping the HDP Services

The HDP on Windows installer sets up Windows services for each HDP component across the nodes in a cluster. Use the instructions given below to stop HDP services from any host machine in your cluster.

Complete the following instructions as the administrative user:

1. Stop the HDP cluster.

Navigate to the install directory and execute the following command from any host machine in your cluster.

```
cd %HADOOP_NODE_INSTALL_ROOT%  
stop_remote_hdp_services.cmd
```

## 5. Uninstalling HDP

Choose one of the following options to uninstall HDP:

- [Option I - Use Windows GUI](#)
- [Option II - Use command line utility](#)

### 5.1. Option I - Use Windows GUI

1. Open the **Programs and Features** Control Panel Pane.
2. Select the program listed: **Hortonworks Data Platform 1.1.0 for Windows**.
3. With that program selected, click on the **Uninstall** option.

### 5.2. Option II - Use Command Line Utility

1. On each cluster host, execute the following command from the command shell:

```
msiexec /x "$MSI_PATH" /lv "$PATH_to_Installer_Log_File"
```

where

- `$MSI_PATH` is the full path to MSI.
  - `$PATH_to_Installer_Log_File` is full path to Installer log file.
2. Optionally, you can also specify if you want delete the data in target data directories.

To do this, use the `DESTROY_DATA` command line option as shown below:

```
msiexec /x "$MSI_PATH" /lv "$PATH_to_Installer_Log_File" DESTROY_DATA="yes"
```



#### Note

During uninstall if `DESTROY_DATA` is not specified or set to `no`, data directories are preserved as well as the `hadoop` user that owns them.