

# Upgrading Big Data Management to Version 10.1.1 Update 2 for Hortonworks HDP

## Abstract

Big Data Management version 10.1.1 Update 2 supports Hortonworks HDP versions 2.3, 2.4, and 2.5. This article describes how to upgrade from version 10.1.1 to version 10.1.1 Update 2 for Hortonworks HDP 2.5.

## Supported Versions

- Informatica Big Data Management™ 10.1.1 Update 2

## Table of Contents

Overview. . . . .	2
Installation Packages. . . . .	4
Before You Upgrade. . . . .	5
Read the Release Notes. . . . .	5
Verify System Requirements. . . . .	5
Verify Upgrade Requirements. . . . .	5
Uninstall Big Data Management from the Hortonworks HDP Cluster. . . . .	6
Upgrade in the Hadoop Environment. . . . .	7
Download the Distribution Package on the Cluster. . . . .	7
Install Big Data Management on the Cluster. . . . .	7
Upgrade in the Domain Environment. . . . .	10
Upgrade on the Data Integration Service Machines. . . . .	10
Upgrade on the Developer Tool Machine. . . . .	10
After You Upgrade. . . . .	11
Update the hadoopEnv.properties File in the Domain Environment. . . . .	11
Update core-site.xml in the Hadoop Environment. . . . .	12
Update yarn-site.xml in the Domain Environment. . . . .	12
Import Security Certificates into the Domain Environment. . . . .	13
Roll Back 10.1.1 Update 2 in the Domain Environment. . . . .	14
Roll Back on the Data Integration Service Machines. . . . .	14
Roll Back on the Developer Tool Machine. . . . .	15

## Overview

Big Data Management version 10.1.1 Update 2 for Hortonworks HDP requires updates in the following environments:

### Domain environment

The domain environment includes the Informatica domain, which consists of nodes where the application services run and Informatica Developer (the Developer tool).

You cannot upgrade the Informatica services and clients directly to version 10.1.1 Update 2. You must upgrade to 10.1.1 and then apply 10.1.1 Update 2.

## Hadoop environment

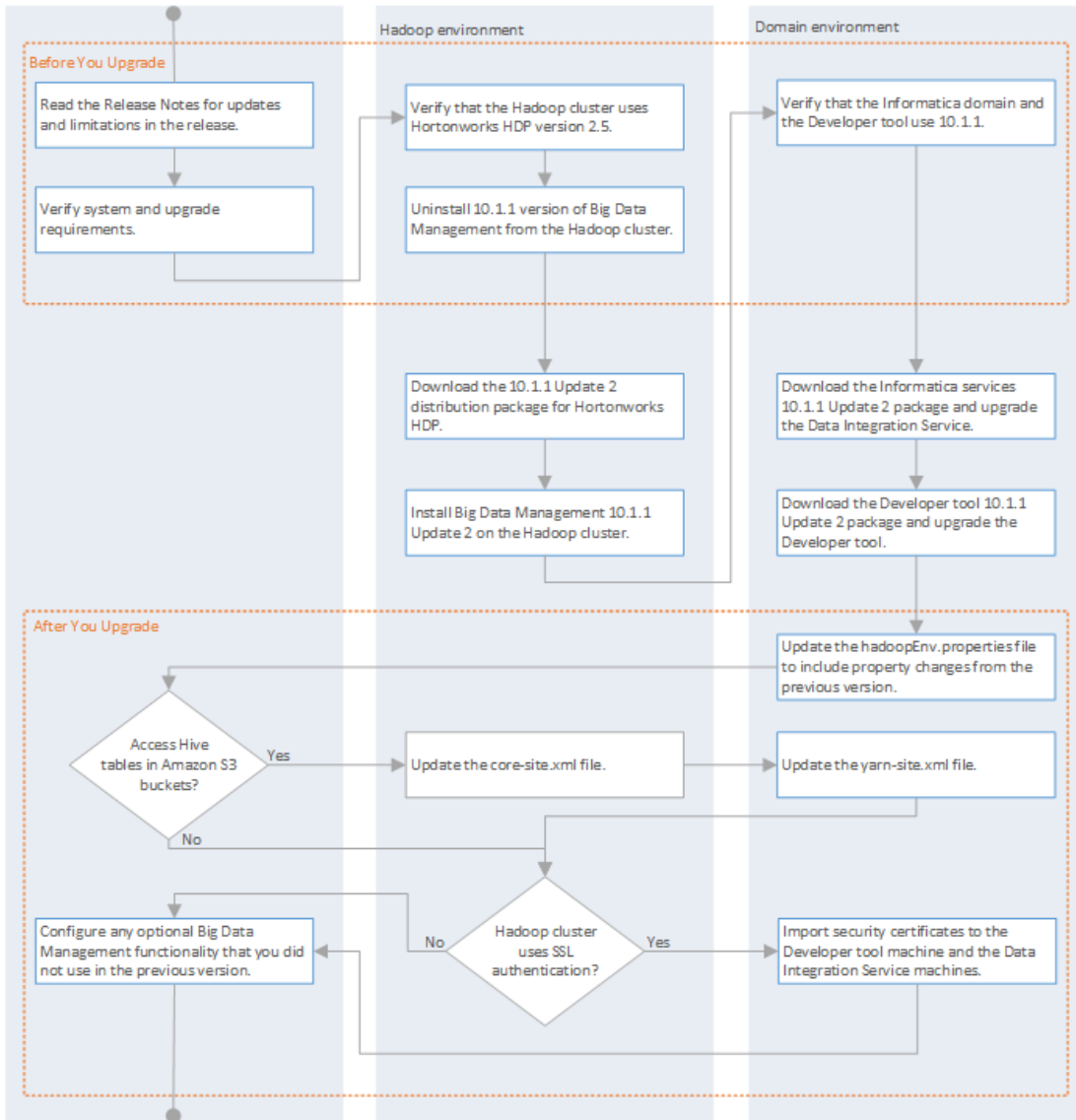
The Hadoop environment includes a Hadoop cluster that uses Hortonworks HDP version 2.5. Informatica distributes a Big Data Management package for Hortonworks HDP that you install on the Hadoop cluster.

You must uninstall previous version of Big Data Management from the Hadoop cluster and then install version 10.1.1 Update 2.

To upgrade Big Data Management for Hortonworks HDP, perform the following steps:

1. **Complete pre-upgrade tasks.**
  - In the Hadoop environment, verify the Hadoop distribution version and uninstall the previous version of Big Data Management from the Hadoop cluster.
  - In the domain environment, verify that the Informatica domain and the Developer tool use version 10.1.1.
2. **Complete the upgrade tasks.**
  - In the Hadoop environment, download the distribution package and install Big Data Management 10.1.1 Update 2 on the Hadoop cluster.
  - In the domain environment, download the packages to update the Data Integration Service machine and the Developer tool machine. Apply the update on the Data Integration Service and Developer tool machines. Use the `Input.properties` file to set the destination directory each time you run the upgrade.
3. **Complete post-upgrade tasks.**
  - If mappings read from or write to Hive tables in Amazon S3 buckets, configure the following files to access the Hive tables in S3 buckets:
    - In the Hadoop environment, update the `core-site.xml` file.
    - In the domain environment, update the `yarn-site.xml` file.
  - In the domain environment, update the `hadoopEnv.properties` file to include any configuration changes from the previous version. If the Hadoop cluster uses SSL authentication, import security certificates to the Developer tool machine and the Data Integration Service machines.

The following flowchart shows the upgrade task flow:



## Installation Packages

Informatica distributes the installation packages through download files.

The following table lists the download packages required to upgrade Big Data Management:

Environment	Package
Hadoop environment	Download one of the following packages to install on the Hadoop environment: <ul style="list-style-type: none"><li>- <b>Informatica_1011U2_ambariservice_linux-x64t.tar.gz</b>. Contains Ambari Stack archives for Red Hat Enterprise Linux.</li><li>- <b>Informatica_1011U2_ambariservice_suse11-x64t.tar.gz</b>. Contains Ambari Stack archives for SUSE 11.</li></ul>
Domain environment	Download one of the following packages to update the Data Integration Service machine: <ul style="list-style-type: none"><li>- <b>Informatica_1011U2_server_linux_x64t.tar</b>. Contains updates for Red Hat Enterprise Linux.</li><li>- <b>Informatica_1011U2_server_suse11_x64t.tar</b>. Contains updates for SUSE 11.</li></ul>
Developer tool machine	Download the following package to update the Developer tool machine: <ul style="list-style-type: none"><li>- <b>Informatica_1011U2_Client_winem_x64t.tar</b>. Contains updates for Windows.</li></ul>

## Before You Upgrade

Before you upgrade, perform the following pre-upgrade tasks:

1. Read the Release Notes.
2. Verify system requirements.
3. Verify upgrade requirements.
4. Uninstall previous version of Big Data Management from the Hadoop cluster.

### *Read the Release Notes*

Read the Release Notes for updates to the installation and upgrade process. You can also find information about known and fixed limitations for the release.

### *Verify System Requirements*

Verify that your environment meets the minimum system requirements for the installation process, disk space requirements, port availability, and third-party software.

For more information, see the *Big Data Management 10.1.1 Update 2 Installation and Configuration Guide*.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica Network:

<https://network.informatica.com/community/informatica-network/product-availability-matrices/overview>

### *Verify Upgrade Requirements*

Verify the following requirements before you upgrade:

#### **Upgrade requirements for the domain environment**

- Verify that the Informatica domain and the Developer tool use version 10.1.1.
- Verify additional upgrade requirements as specified in the *Informatica Upgrade Guides*.

#### **Upgrade requirements for the Hadoop environment**

- Verify that the Hadoop cluster uses Hortonworks HDP version 2.5.
- Verify that the user who performs the installation can run sudo commands or has user root privileges.

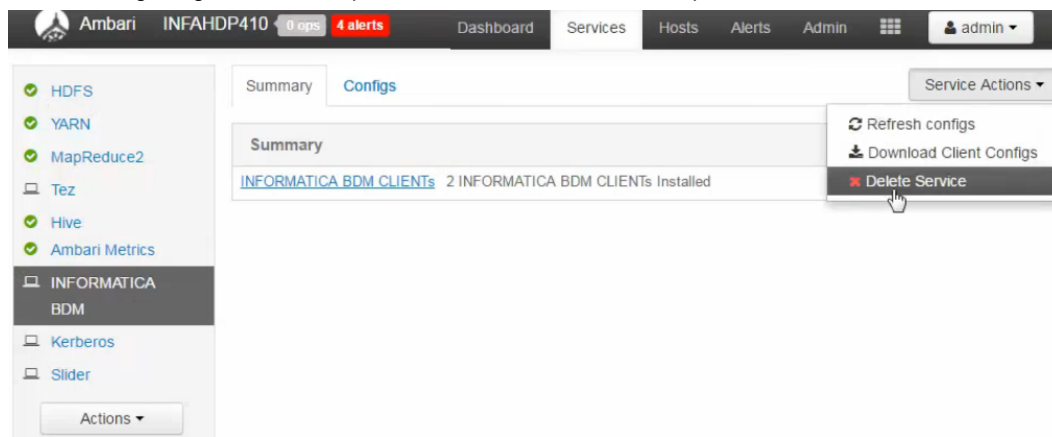
- Verify that the temporary folder on the local node has at least 2 GB of disk space.
- Configure a Secure Shell (SSH) connection between the machine that runs the Big Data Management installation and the nodes in the Hadoop cluster. Configure passwordless SSH for the root user.

## Uninstall Big Data Management from the Hortonworks HDP Cluster

Uninstall the previous version of Big Data Management from the Hadoop cluster. To uninstall the stack deployment of Big Data Management, you use the Ambari configuration manager to stop and deregister the Big Data Management service, and then perform manual removal of Informatica files from the cluster.

1. In the Ambari configuration manager, select **INFORMATICA BDM** from the list of services.
2. Click the **Service Actions** dropdown menu and select **Delete Service**.

The following image shows the option in the **Service Actions** dropdown menu:



3. To confirm that you want to delete Informatica Big Data Management, perform the following steps:
  - a. In the **Delete Service** dialog box, click **Delete**.
  - b. In the Confirm Delete dialog box, type `delete` and then click **Delete**.
  - c. When the deletion process is complete, click **OK**.

Ambari stops the Big Data Management service and deletes it from the listing of available services. To fully delete Big Data Management from the cluster, continue with the next steps.

4. In a command window, delete the `INFORMATICA_BDM` folder from the following directory on the name node of the cluster: `/var/lib/ambari-server/resources/stacks/<Hadoop distribution>/<Hadoop version>/services/`
5. Delete the `INFORMATICA_BDM` folder from the following location on all cluster nodes where it was installed: `/var/lib/ambari-agent/cache/stacks/<Hadoop distribution>/<Hadoop version>/services`
6. Perform the following steps to remove RPM binary files:
  - a. Run the following command to determine the name of the RPM binary archive:
 

```
rpm -qa |grep Informatica
```
  - b. Run the following command to remove RPM binary files:
 

```
rpm -ev <output_from_above_command>
```

For example:

```
rpm -ev InformaticaHadoop-10.1.1-1.x86_64
```
7. Repeat the previous step to remove RPM binary files from each cluster node.
8. Delete the following directory, if it exists, from the name node and each client node: `/opt/Informatica/`.

9. Repeat the last step on each cluster node where Big Data Management was installed.
10. On the name node, restart the Ambari server.

## Upgrade in the Hadoop Environment

On the Hadoop cluster, verify that the destination directory for Informatica binary files is empty. Files that remain from previous installations can cause conflicts that lead to mapping run failures.

To upgrade Big Data Management in the Hadoop environment, complete the following steps:

1. Download the distribution package on the cluster.
2. Install Big Data Management on the cluster.

### *Download the Distribution Package on the Cluster*

The tar.gz file that you download includes the binaries required to run the installation.

1. Download the Big Data Management archive file `Informatica-<version>_ambariservice_<platform>-x64.tar.gz` to a location on the head node machine, and uncompress the archive.  
The archive file contains a folder named `INFORMATICABDM`. This folder contains Big Data Management binaries. It also contains configuration files that Ambari uses to deploy the software.
2. Copy the `INFORMATICABDM` folder to the following location on the name node: `/var/lib/ambari-server/resources/stacks/<Hadoop_distribution>/<version>/services/`
3. Restart the Ambari server.

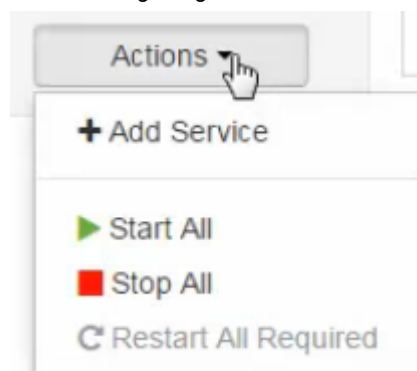
When Ambari restarts, it registers the presence of the Big Data Management package and makes it available to install as a service in the stack.

### *Install Big Data Management on the Cluster*

Install Big Data Management for Hortonworks HDP as a service in the Ambari stack.

1. In the Ambari management console, click the **Admin** tab and select **Stacks and Versions**.
2. Click the **Actions** button and select **Add Service**.

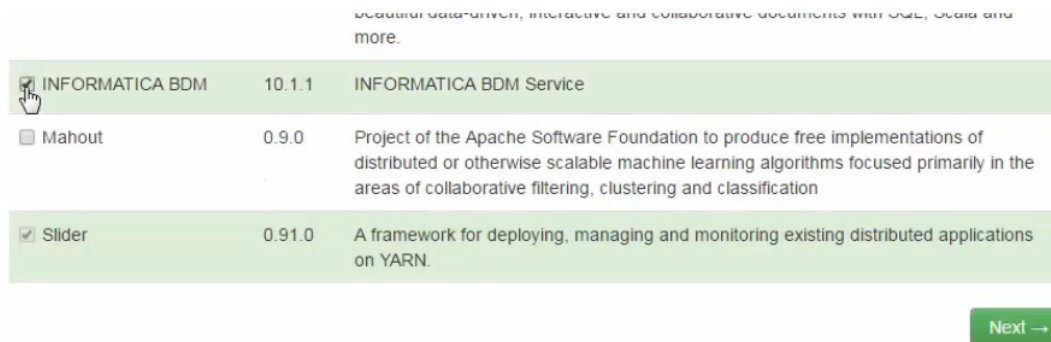
The following image shows the **Add Service** menu item:



The **Add Service Wizard** opens.

3. In the list of available services, select **INFORMATICABDM** and click **Next**.

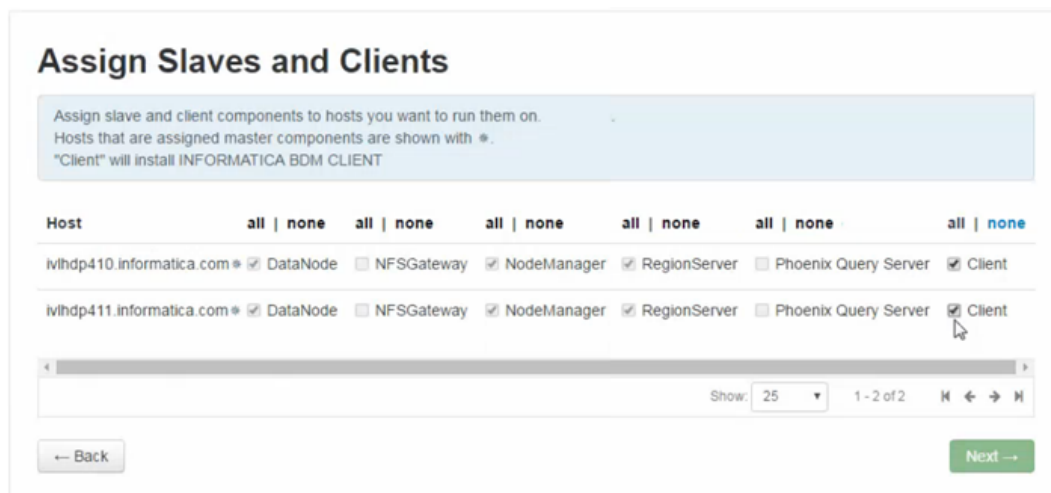
The following image shows the Big Data Management service in the list of available services:



The **Assign Slaves and Clients** screen opens.

- In the **Assign Slaves and Clients** screen, select **Client** for each node where you want to install Big Data Management.

The following image shows how to designate nodes as clients:



- In the **Customize Services** screen, click **Next**.
- A **Configurations** screen might appear with warning messages about configuration and environment details. You can ignore these warnings and click **Proceed Anyway**.

The **Configure Identities** screen displays default configuration properties.

- Click **Next**.

The **Review** screen displays a summary of the client installation to perform.

- Click **Deploy**.

- If the cluster uses Kerberos authentication, the **Admin session expiration error** dialog box appears. Enter Administrator principal authentication credentials and click **Save**.

The following image shows the dialog box populated with authentication credentials:



### Admin session expiration error

Missing KDC administrator credentials. Please enter admin principal and password.

Admin principal:

Admin password:

Save Admin Credentials ⓘ

10. Monitor client installation in the **Install, Start and Test** screen.

The following image shows the **Install, Start and Test** screen when installation is complete:

## Install, Start and Test

Please wait while the selected services are installed and started.

100 % overall

Show: **All (2)** | [In Progress \(0\)](#) | [Warning \(0\)](#) | [Success \(2\)](#) | [Fail \(0\)](#)

Host	Status	Message
ivhdp410.informatica.com	100%	Success
ivhdp411.informatica.com	100%	Success

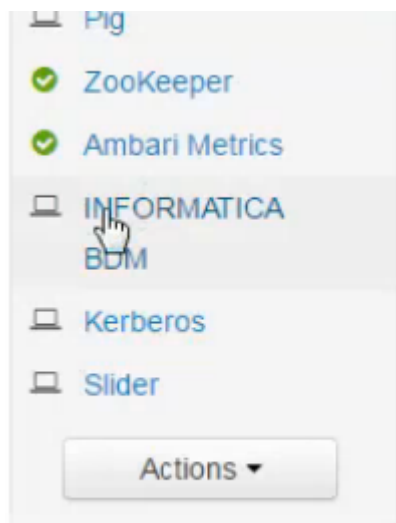
2 of 2 hosts showing - [Show All](#) Show: 25 1 - 2 of 2

Successfully installed and started the services.

11. Click **Complete**.

Big Data Management is installed on cluster nodes.

The following image shows the list of stack services after you install Big Data Management:



Big Data Management is installed as an Ambari stack service.

## Upgrade in the Domain Environment

In the Domain environment, complete the following steps to upgrade Big Data Management:

1. Download the 10.1.1 Update 2 server package and apply 10.1.1 Update 2 on each machine that hosts the Data Integration Service.
2. Download the 10.1.1 Update 2 client package and apply 10.1.1 Update 2 on the Developer tool machine.

### *Upgrade on the Data Integration Service Machines*

Upgrade each machine that hosts the Data Integration Service to apply 10.1.1 Update 2. Update the destination directory in the `Input.properties` file before you run the upgrade.

1. Create a temporary directory in a shared location on a local disk.
2. Download the `Informatica_<version>_server_<platform>-x64.tar.gz` file.
3. Extract the file to the machine from where you want to run the upgrade.
4. Update the `Input.properties` file with the following information:  
`DEST_DIR`. The destination directory on the machine where the Data Integration Service runs. For example,  
`DEST_DIR=C:\\Informatica\`
5. Type `installEBF.sh` from the command line to upgrade the Data Integration Service machine.

### *Upgrade on the Developer Tool Machine*

When you upgrade Big Data Management configuration on the Developer tool client to version 10.1.1 Update 2, the installer uses the `Input.properties` file to update the destination directory for the client upgrade.

1. Create a temporary directory in a shared location on a local disk.

2. Download the `Informatica_<version>_client_winem-x64.tar.gz` file.
3. Extract the file to the machine from where you want to where you want to run the upgrade.
4. Update the `Input.properties` file with the following information:  
`DEST_DIR`. The destination directory on the machine where you have installed the clients. For example,  
`DEST_DIR=C:\\Informatica\`
5. Type `installEBF.bat` from the command line to run the upgrade.

## After You Upgrade

After you install Big Data Management on the cluster and update the domain environment, perform additional tasks for Update 2 functionality.

After you upgrade, perform the following post-upgrade tasks:

1. In the domain environment, update the `hadoopEnv.properties` file to include any manual configuration that you performed in the previous version.
2. To access Hive tables in Amazon S3 buckets, update the `yarn-site.xml` file in the domain environment and the `core-site.xml` file in the Hadoop environment.
3. If the Hadoop cluster uses SSL authentication, import security certificates to the Developer tool machine and the Data Integration Service machines.

You can also configure any additional Big Data Management functionality that you did not configure in 10.1.1. For example, you might want to configure the following functionality:

### Run-time engines

If you did not configure all run-time engines in 10.1.1, you can get instructions from the Informatica How-To Library article *Configuring Run-time Engines for Big Data Management*:  
<https://kb.informatica.com/h2/HowTo%20Library/1/1005-Configuring-Run-Time-Engines-For-BDM-H2L.pdf>

### Sqoop connectivity

If you did not configure Sqoop connectivity in 10.1.1, you can get instructions from the Informatica How-To Library article *Configuring Sqoop Connectivity for Big Data Management*:  
<https://kb.informatica.com/h2/HowTo%20Library/1/1004-Configuring-Sqoop-Connectivity-For-BDM-H2L.pdf>

## Update the `hadoopEnv.properties` File in the Domain Environment

When you run the upgrade on a machine that hosts the Data Integration Service, the upgrade process creates a new `hadoopEnv.properties` file and backs up the existing configuration file.

After the upgrade, edit the 10.1.1 Update 2 version of the `hadoopEnv.properties` file to include any manual configuration that you performed in the previous version for the corresponding distribution.

You can find the backup `hadoopEnv.properties` file in the following location:

```
<Informatica installation directory>/EBFs/EBF-1011U2/services/shared/hadoop/<Hadoop distribution name>_<version>/infaConf
```

You can find the 10.1.1 Update 2 `hadoopEnv.properties` file in the following location:

```
<Informatica installation directory>/services/shared/hadoop/<Hadoop distribution name>_<version>/infaConf
```

## Update core-site.xml in the Hadoop Environment

Update the core-site.xml file in the Hadoop environment to enable access to Hive tables in Amazon S3 buckets.

When you use Apache Ambari to update properties, updates are propagated to each node in the cluster. After you make updates, you can restart the Hadoop cluster and services through Apache Ambari.

### Configure Access to Hive Tables in Amazon S3 Buckets

Configure the AWS access key to run mappings with sources and targets on Hive tables in Amazon S3 buckets.

**Note:** To use a Hive table as a target on Amazon S3, grant write permission to the bucket through bucket policies, or add these properties to the configuration file. You must add these properties to the core-site.xml file on the Hadoop environment and to the yarn-site.xml file on the domain environment.

Configure the following properties:

#### **fs.s3n.awsAccessKeyId**

The ID for the Blaze and Spark engines to connect to the Amazon S3n file system. For example,

```
<property>
  <name>fs.s3n.awsAccessKeyId</name>
  <value>[Your Access Key]</value>
</property>
```

#### **fs.s3n.awsSecretAccessKey**

The password for the Blaze and Spark engines to connect to the Amazon S3n file system.

```
<property>
  <name>fs.s3n.awsSecretAccessKey</name>
  <value>[Your Access Id]</value>
</property>
```

When you use a server side encryption protocol for Hive buckets, configure properties to enable access to encrypted Hive buckets.

Configure the following properties:

#### **fs.s3.enableServerSideEncryption**

Set to TRUE to enable server side encryption. For example,

```
<property>
  <name>fs.s3.enableServerSideEncryption</name>
  <value>TRUE</value>
</property>
```

#### **fs.s3n.server-side-encryption-algorithm**

The encryption algorithm that you use to encrypt Hive buckets. For example,

```
<property>
  <name>fs.s3n.server-side-encryption-algorithm</name>
  <value>AES256</value>
</property>
```

## Update yarn-site.xml in the Domain Environment

Update the yarn-site.xml file on the domain to configure functionality such as access to Hive tables in Amazon S3 buckets.

### Configure Access to Hive Tables in Amazon S3 Buckets

Configure the AWS access key to run mappings with sources and targets on Hive tables in Amazon S3 buckets.

**Note:** To use a Hive table as a target on Amazon S3, grant write permission to the bucket through bucket policies, or add these properties to the configuration file. You must add these properties to the core-site.xml file on the Hadoop environment and to the yarn-site.xml file on the domain environment.

Configure the following properties:

**fs.s3n.awsAccessKeyId**

The ID for the Blaze and Spark engines to connect to the Amazon S3n file system. For example,

```
<property>
  <name>fs.s3n.awsAccessKeyId</name>
  <value>[Your Access Key]</value>
</property>
```

**fs.s3n.awsSecretAccessKey**

The password for the Blaze and Spark engines to connect to the Amazon S3n file system.

```
<property>
  <name>fs.s3n.awsSecretAccessKey</name>
  <value>[Your Access Id]</value>
</property>
```

When you use a server side encryption protocol for Hive buckets, configure properties to enable access to encrypted Hive buckets.

Configure the following properties:

**fs.s3.enableServerSideEncryption**

Set to TRUE to enable server side encryption. For example,

```
<property>
  <name>fs.s3.enableServerSideEncryption</name>
  <value>TRUE</value>
</property>
```

**fs.s3n.server-side-encryption-algorithm**

The encryption algorithm that you use to encrypt Hive buckets. For example,

```
<property>
  <name>fs.s3n.server-side-encryption-algorithm</name>
  <value>AES256</value>
</property>
```

## *Import Security Certificates into the Domain Environment*

When you use custom, special, or self-signed security certificates to secure the Hadoop cluster, the Developer tool and the Data Integration Service that connect to the cluster require these certificates to be present on the machine.

To connect to the Hadoop cluster to develop a mapping, the Developer tool requires security certificate aliases on the machine that hosts the Developer tool. To run a mapping, the machine that hosts the Data Integration Service requires these same certificate alias files.

Perform the following steps from the Developer tool host machine, and then repeat them from the Data Integration Service host machine:

1. Run the following command to export the certificates from the cluster:

```
keytool -export -alias <alias name> -keystore <custom.truststore file location> -file
<exported certificate file location> -storepass <password>
```

For example,

```
<java home>/jre/bin/keytool -export -alias <alias name> -keystore ~/custom.truststore -
file ~/exported.cer
```

The command produces a certificate file.

2. Choose whether the domain is SSL-enabled or not SSL-enabled.

- If the domain is SSL-enabled, then import the certificate alias file to the following locations:
  - The client truststore at the location `<Informatica installation directory>\clients\DeveloperClient\shared\security\infa_truststore.jks`.
  - The domain truststore at the location `<Informatica installation directory>/services/shared/security/infa_truststore.jks`
- If the domain is not SSL-enabled, then run the following command to copy the exported certificate file to the client machine and import the certificate file to java cacerts:

```
keytool -import -trustcacerts -alias <alias name> -file <exported certificate file location> -keystore <java cacerts location> -storepass <password>
```

For example,

```
<java home>/jre/keytool -import -alias <alias name> -file ~/exported.cer -keystore <Informatica installation directory>/java/jre/lib/security/cacerts
```

3. Set the following JVMOption property for the Data Integration Service:

```
-Djavax.net.ssl.trustStore=<truststore location>
```

## Roll Back 10.1.1 Update 2 in the Domain Environment

You can roll back the 10.1.1 Update 2 domain upgrade. The installer rolls back the Big Data Management updates from the Data Integration Service machines and the Developer tool machine to version 10.1.1.

You cannot roll back Big Data Management 10.1.1 Update 2 from the Hadoop cluster.

### *Roll Back on the Data Integration Service Machines*

You can roll back Big Data Management updates on the Data Integration Service machines. The installer uses the `Input.properties` file to update the destination directory and rolls back the Big Data Management updates to the previously installed version. Run the installer on each machine that hosts the Data Integration Service to roll back.

1. Edit the `Input.properties` file with the following information:

Option	Description
DEST_DIR	The path to the Informatica installation directory.
ROLLBACK	Enter ROLLBACK=1 to roll back the installation. Default is ROLLBACK=0.

2. Type `installEBF.sh` from the command line to roll back the upgrade.

## Roll Back on the Developer Tool Machine

If you want to roll back the 10.1.1 Update 2 client upgrade, the installer uses the `Input.properties` file to update the destination directory and rolls back the Big Data Management updates to the previously installed version.

1. Edit the `Input.properties` file with the following information:

Option	Description
DEST_DIR	The path to the Informatica installation directory.
ROLLBACK	Enter ROLLBACK=1 to roll back the installation. Default is ROLLBACK=0.

2. Type `installEBF.bat` from the command line to roll back the upgrade.

## Author

Informatica Documentation Team