# Installing Apache Zeppelin

**Date of Publish:** 2019-12-17

# Contents

# Install Apache Zeppelin Using Ambari

How to install Apache Zeppelin on an Ambari-managed cluster.

### Before you begin

Install Zeppelin on a node where Spark clients are already installed and running. This typically means that Zeppelin will be installed on a gateway or edge node.

Zeppelin requires the following software versions:

- HDP 3.0 or later.
- Apache Spark 2.0.
- Java 8 on the node where Zeppelin is installed.

The optional Livy server provides security features and user impersonation support for Zeppelin users. Livy is installed as part of Spark.

- After installing Spark, Livy, and Zeppelin, refer to "Configuring Zeppelin" in this guide for post-installation steps.

### Install Zeppelin Using Ambari

The Ambari installation wizard sets default values for Zeppelin configuration settings. Initially, you should accept the default settings. Later, when you are more familiar with Zeppelin, consider customizing the Zeppelin configuration settings.

To install Zeppelin using Ambari, add the Zeppelin service:

1. Click the ellipsis (…) symbol next to Services on the Ambari dashboard, then click Add Service.
2. On the Add Service Wizard under Choose Services, select Zeppelin Notebook, then click Next.
3. On the Assign Masters page, review the node assignment for Zeppelin Notebook, then click Next.
4. On the Customize Services page, review the default values, then click Next.
5. If Kerberos is enabled on the cluster, review the principal and keytab settings on the Configure Identities page, modify the settings if desired, then click Next.
6. Review the configuration on the Review page, then click Deploy to begin the installation.
7. The Install, Start, and Test page displays the installation status.
8. When the progress bar reaches 100% and a "Success" message appears, click Next.
9. On the Summary page, click Complete to finish installing Zeppelin.

To validate the Zeppelin installation, open the Zeppelin Web UI in a browser window. Use the port number configured for Zeppelin (9995 by default); for example:

```
http://<zeppelin-host>:9995
```

You can also open the Zeppelin Web UI by selecting Zeppelin Notebook > Zeppelin UI on the Ambari dashboard.

To check the Zeppelin version number, type the following command on the command line:

```
/usr/hdp/current/zeppelin-server/bin/zeppelin-daemon.sh --version
```

Zeppelin stores configuration settings in the /etc/zeppelin/conf directory. Note, however, that if your cluster is managed by Ambari you should not modify configuration settings directly. Instead, use the Ambari web UI.

Zeppelin stores log files in /var/log/zeppelin on the node where Zeppelin is installed.

# Enabling HDFS and Configuration Storage for Zeppelin Notebooks in HDP-2.6.3+

## Overview

When upgrading to HDP-2.6.3 and higher versions, additional configuration steps are required to enable HDFS storage for Apache Zeppelin notebooks.

HDP-2.6.3 introduced support for HDFS storage for Apache Zeppelin notebooks and configuration files. In previous versions, notebooks and configuration files were stored on the local disk of the Zeppelin server.

When upgrading to HDP-2.6.3 and higher versions, there are two options for configuring Zeppelin notebook and configuration file storage:

• Use HDFS storage (recommended) – Zeppelin notebooks and configuration files must be copied to the new HDFS storage location before upgrading. Additional upgrade and post-upgrade steps must also be performed, as described in the following section.

• Use local storage – Perform upgrade and post-upgrade steps to enable local storage.

Enabling HDFS storage makes future upgrades much easier, and also sets up the first step toward enabling Zeppelin High Availability. Therefore it is recommended that you enable HDFS for Zeppelin notebooks and configuration files when upgrading to HDP 2.6.3+ from earlier versions of HDP.

> **Note:**
>
> Currently HDFS and local storage are the only supported notebook storage mechanisms in HDP-2.6.3+. Currently VFSNotebookRepo is the only supported local storage option.

## Enable HDFS Storage when Upgrading to HDP-2.6.3+

Perform the following steps to enable HDFS storage when upgrading to HDP 2.6.3+ from earlier versions of HDP.

### Procedure

1. Before upgrading Zeppelin, perform the following steps as the Zeppelin service user.

   a. Create the /user/zeppelin/conf and /user/zeppelin/notebook directories in HDFS.

   ```
   hdfs dfs -ls /user/zeppelin
   drwxr-xr-x   - zeppelin hdfs          0 2018-01-20 04:17 /user/zeppelin/
   conf
   drwxr-xr-x   - zeppelin hdfs          0 2018-01-20 03:40 /user/zeppelin/
   notebook
   ```

   b. Copy all notebooks from the local Zeppelin server (for example, /usr/hdp/2.5.3.0-37/zeppelin/notebook/) to the /user/zeppelin/notebook directory in HDFS.

   ```
   hdfs dfs -ls /user/zeppelin/notebook
   drwxr-xr-x   - zeppelin hdfs          0 2018-01-19 01:40 /user/zeppelin/
   notebook/2A94M5J1Z
   drwxr-xr-x   - zeppelin hdfs          0 2018-01-19 01:40 /user/zeppelin/
   notebook/2BWJFTXKJ
   ```

c. Copy the interpreter.json and notebook-authorization.json files from the local Zeppelin service configuration directory (/etc/zeppelin/conf) to the/user/zeppelin/conf directory in HDFS.

```
hdfs dfs -ls /user/zeppelin/conf
-rw-r--r--   3 zeppelin hdfs      284091 2018-01-22 23:28 /user/zeppelin/
conf/interpreter.json
-rw-r--r--   3 zeppelin hdfs      123849 2018-01-22 23:29 /user/zeppelin/
conf/notebook-authorization.json
```

2. Upgrade Ambari.

3. Upgrade HDP and Zeppelin. During the upgrade, verify that the following configuration settings are present in Ambari for Zeppelin.

```
zeppelin.notebook.storage =
 org.apache.zeppelin.notebook.repo.FileSystemNotebookRepo
zeppelin.config.fs.dir = conf
```

If necessary, add or update these configuration settings as shown above.

4. After the upgrade is complete:

a. Log on to the Zeppelin server and verify that the following properties exist in the /etc/zeppelin/conf/zeppelin-site.xml file. The actual value for the keytab file and principal name may be different for your cluster.

```
<property>
  <name>zeppelin.server.kerberos.keytab</name>
  <value>/etc/security/keytabs/zeppelin.server.kerberos.keytab</value>
</property>
<property>
  <name>zeppelin.server.kerberos.principal</name>
  <value>zeppelin@EXAMPLE.COM</value>
</property>
```

b. Check the Zeppelin Interpreter page to see if any interpreter (e.g. the Livy interpreter) is duplicated. This may happen in some cases. If duplicate interpreter entries are found, perform the following steps:

1. Backup and delete the interpreter.json file from HDFS (/user/zeppelin/conf/interpreter.json) and from the local Zeppelin server.
2. Restart the Zeppelin service.
3. Verify that the duplicate entries no longer exist.
4. If any custom interpreter settings were present before the upgrade, add them again via the Zeppelin interpreter UI page.

c. Verify that your existing notebooks are available on Zeppelin.

**Note:** When an existing notebook is opened for the first time after the upgrade, it may ask you to save the interpreters associated with the notebook.

## Use Local Storage when Upgrading to HDP-2.6.3+

Perform the following steps to use local notebook storage when upgrading to HDP 2.6.3+ from earlier versions of HDP.

### Procedure

1. Upgrade Ambari.

2. Upgrade HDP and Zeppelin. During the upgrade, verify that the following configuration settings are present in Ambari for Zeppelin.

```
zeppelin.notebook.storage =
  org.apache.zeppelin.notebook.repo.VFSNotebookRepo
zeppelin.config.fs.dir = file:///etc/zeppelin/conf
```

If necessary, add or update these configuration settings as shown above.

3. After the upgrade is complete:

   a. Copy your notebooks and the notebook-authorization.json file from the previous Zeppelin installation directory to the new installation directory on the Zeppelin server machine.

   b. Verify that your existing notebooks are available on Zeppelin.

   **Note:** When an existing notebook is opened for the first time after the upgrade, it may ask you to save the interpreters associated with the notebook.