**Installing Apache Hive 3.1.2 on Windows 10**

## **Pre-requisite**

## To install Apache Hive, we must have a Hadoop Cluster installed and running.

## Apache Derby

In addition, Apache Hive requires a relational database to create its Metastore (where all metadata will be stored). In this guide, we will use the Apache Derby database 4.

Since we have Java 8 installed, we must install Apache Derby 10.14.2.0 version which can be downloaded from the following link.

<https://downloads.apache.org/db/derby/db-derby-10.14.2.0/db-derby-10.14.2.0-bin.tar.gz>

Extract the content of the db-derby-10.14.2.0-bin.tar.gz archive into the desired installation directory.

## **Downloading Apache Hive binaries**

Download the apache-hive-3.1.2.-bin.tar.gz file from the below given link.

<https://downloads.apache.org/hive/hive-3.1.2/>

Extract the apache-hive.3.1.2-bin.tar.gz archive into a suitable directory.

## **Setting environment variables**

1. HIVE\_HOME: “E:\hadoop-env\apache-hive-3.1.2\”
2. DERBY\_HOME: “E:\hadoop-env\db-derby-10.14.2.0\”
3. HIVE\_LIB: “%HIVE\_HOME%\lib”
4. HIVE\_BIN: “%HIVE\_HOME%\bin”
5. HADOOP\_USER\_CLASSPATH\_FIRST: “true”

Now, we should edit the Path user variable to add the following paths:

* %HIVE\_BIN%
* %DERBY\_HOME%\bin

## **Configuring Hive**

## 4.1 Copy Derby libraries

Now, we should go to the Derby libraries directory (E:\your-hive-home\db-derby-10.14.2.0\lib) and copy all \*.jar files.

Then, we should paste them within the Hive libraries directory (E:\ \apache-hive-3.1.2\lib).

## 4.2 Configuring hive-site.xml

Now, we should go to the Apache Hive configuration directory (E:\ \apache-hive-3.1.2\conf) create a new file “hive-site.xml”. We should paste the following XML code within this file:

<?xml version="1.0"?>  
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
<configuration><property> <name>javax.jdo.option.ConnectionURL</name>   
<value>jdbc:derby://localhost:1527/metastore\_db;create=true</value>   
<description>JDBC connect string for a JDBC metastore</description>  
</property><property>   
<name>javax.jdo.option.ConnectionDriverName</name>   
<value>org.apache.derby.jdbc.ClientDriver</value>   
<description>Driver class name for a JDBC metastore</description>  
</property>  
<property>   
<name>hive.server2.enable.doAs</name>   
<description>Enable user impersonation for HiveServer2</description>  
<value>true</value>  
</property>  
<property>  
<name>hive.server2.authentication</name>   
<value>NONE</value>  
<description> Client authentication types. NONE: no authentication check LDAP: LDAP/AD based authentication KERBEROS: Kerberos/GSSAPI authentication CUSTOM: Custom authentication provider (Use with property hive.server2.custom.authentication.class) </description>  
</property>  
<property>  
<name>datanucleus.autoCreateTables</name>  
<value>True</value>  
</property>  
</configuration>

## **Starting Services**

## Hadoop Services

To start Apache Hive, open the command prompt utility as administrator. Then, start the Hadoop services using start-dfs and start-yarn commands (as illustrated in the Hadoop installation guide).

## Derby Network Server

Then, we should start the Derby network server on the localhost using the following command:

StartNetworkServer -h 0.0.0.0

## **Starting Apache Hive**

Now, let us try to open a command prompt tool and go to the Hive binaries directory (E:\hadoop-env\apache-hive-3.1.2\bin) and execute the following command:

hive

We will receive the following error:

'hive' is not recognized as an internal or external command, operable program or batch file.

This error is thrown since the Hive 3.x version is not built for Windows (only in some Hive 2.x versions). To get things working, we should download the necessary \*.cmd files from the following link:

h[ttps://svn.apache.org/repos/asf/hive/trunk/bin/](https://svn.apache.org/repos/asf/hive/trunk/bin/hive.cmd" \t "_blank). Note that, you should keep the folder hierarchy (bin\ext\util).

OR

You can download all \*.cmd files from the following GitHub repository

* <https://github.com/HadiFadl/Hive-cmd>

Now if we try to execute the “hive” command, we will receive the an error:

This issue can be solved by replacing the guava-19.0.jar stored in “E:\hadoop-env\apache-hive-3.1.2\lib” with Hadoop’s guava-27.0-jre.jar found in “E:\hadoop-env\hadoop-3.2.1\share\hadoop\hdfs\lib”.

*Note: This file is also uploaded to the GitHub repository mentioned above.*

Now, if we run hive command again, then Apache Hive will start successfully.

## **Initializing Hive**

After ensuring that the Apache Hive started successfully. We may not be able to run any HiveQL command. This is because the Metastore is not initialized yet. Besides HiveServer2 service must be running.

To initialize Metastore, we need to use schematool utility which is not compatible with windows. To solve this problem, we will use Cygwin utility which allows executing Linux command from windows.

**7.1. Creating symbolic links**

First, we need to create the following directories:

E:\cygdrive

C:\cygdrive

Now, open the command prompt as administrator and execute the following commands:

mklink /J E:\cygdrive\e\ E:\  
mklink /J C:\cygdrive\c\ C:\

These symbolic links are needed to work with Cygwin utility properly since Java may cause some problems.

**7.2. Initializing Hive Metastore**

Open Cygwin utility and execute the following commands to define the environment variables:

export HADOOP\_HOME='/cygdrive/e/hadoop-env/hadoop-3.2.1'  
export PATH=$PATH:$HADOOP\_HOME/bin  
export HIVE\_HOME='/cygdrive/e/hadoop-env/apache-hive-3.1.2'  
export PATH=$PATH:$HIVE\_HOME/bin  
export HADOOP\_CLASSPATH=$HADOOP\_CLASSPATH:$HIVE\_HOME/lib/\*.jar

We can add these lines to the “~/.bashrc” file then you don’t need to write them each time you open Cygwin.

Now, we should use the schematool utility to initialize the Metastore:

$HIVE\_HOME/bin/schematool -dbType derby -initSchema

**7.3. Starting HiveServer2 service**

Now, open a command prompt and run the following command:

hive --service hiveserver2 start

We should leave this command prompt open, and open a new one where we should start Apache Hive using the following command:

hive