

Assignment No. 4

Hive Installation Guidelines

1. Prerequisite: Hadoop 3.2.4 (Watch my video to install hadoop 3.2.4)
<https://youtu.be/kUX6dCbU3Q>
2. Download and Configure Apache Derby:
 - a) Download Derby 10.14.2.0 (It's Compatible to Hadoop 3.2.4) from https://db.apache.org/derby/derby_downloads.html
 - b) Extract the file using WinRar, 7Zip, or similar software.
 - c) Copy the extracted folder to `C:\` and rename it to `derby` (e.g., `C:\derby`).
3. Download and Configure Apache Hive Binaries:
 - a) Download Hive 3.1.2 (It's Compatible to Hadoop 3.2.4) from <https://archive.apache.org/dist/hive/hive-3.1.2/>
 - b) Extract the file using WinRar, 7Zip, or similar software.
 - c) Copy the extracted folder to `C:\` and rename it to `hive` (e.g., `C:\hive`).
 - d) Navigate to `C:\derby\lib` and copy all the `*.jar` files.
 - e) Paste the copied `*.jar` files into `C:\hive\lib`.
 - f) Download the exact `hive-site.xml` file from my Google Drive.
 - g) Copy the downloaded `hive-site.xml` file and paste it to `C:\hive\conf` folder.
4. Cross Check guava-x.y-jre file:
 - a) Make sure C:\hive\lib must have same version of guava-x.y-jar file as Hadoop version: C:\hadoop\share\hadoop\common\lib
For example: C:\hadoop\share\hadoop\common\lib\guava-27.0-jre.jar
Then C:\hive\lib\guava-27.0-jre.jar
 - b) If not just copy the Hadoop version guava-27.0-jre.jar and paste into C:\hive\lib and delete the other version guava-x.y-jre.jar file
5. Create a temporary folder named 'tmp' into C:\hive (e.g, C:\hive\tmp) and provide all rights by setting the properties of folder
6. Set Up Environmental Variables:
 - `DERBY_HOME` : `C:\derby`
 - `HIVE_HOME` : `C:\hive`
 - `HIVE_BIN` : `C:\hive\bin`
 - `HIVE_LIB` : `C:\hive\lib`
 - `HADOOP_USER_CLASSPATH_FIRST` : `true`
7. Download and Configure Wget:
 - a) Download wget zip files from <https://eternallybored.org/misc/wget/>
 - b) Extract the wget zip files.

- c) Copy the extracted folder to `C:\` and rename it to `wget` (e.g., `C:\wget`).
- d) Set the 'path' for wget in Environmental Variables: `C:\wget`.

8. Download Windows Version of Hive Bin:

- a) Navigate to a specific directory and create a folder of your choice (e.g., `C:\bin`).
- b) Open Windows Command Prompt and run:
`wget -r -np -nH --cut-dirs=3 -R index.html`

<https://svn.apache.org/repos/asf/hive/trunk/bin/>

- c) Navigate to `C:\bin` and its subdirectories until you reach the `bin` folder.
- d) Copy the inner `bin` folder.
- e) Navigate to `C:\hive`, delete the existing `bin` folder, and replace it with the downloaded `bin` folder.

9. Run the Hadoop daemons:

- a) Command to run daemons: `start-all.cmd`
- b) Commands to check daemons running: `jps`

10. Run Derby Server

- a) Command: `StartNetworkServer -h 0.0.0.0`

11. To Initialize Hive Schema:

- a) Navigate to `C:\hive\bin` and run below command: `hive --service schematool -dbType derby -initSchema`

12. Start Hive:

- a) Command: `hive`

13. Success!: You will be prompted with the Hive command line, ready to run SQL commands.

14. Congratulations!