STEPS FOR INSTALLATION (WINDOWS)

Complete Installations

MONGODB : https://youtu.be/dEm2AS5amyA?si=O5sKPWvQZyIT3cxt

HADOOP : https://www.youtube.com/watch?v=kUX6dCbdU3Q&t=470s

HIVE : https://youtu.be/CL6t2W8YC1Y?si=MAPOPDW3MgL-lzvY

VERSIONS

Java version : 1.8.0_202

Hadoop version : 3.2.4 (https://archive.apache.org/dist/hadoop/common/hadoop-3.2.4/hadoop-3.2.4.tar.gz)

Hive version : 3.1.2 (https://archive.apache.org/dist/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz)

INSTALLATIONS :

Hadoop Installation Guidelines

1. Uninstall all the previous versions of java and install - https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html (**Java SE Development Kit 8u202 – Windows x64)**
2. Create file ` (e.g., `**C:\Java**`), Extract and Copy the extracted folder to `C:\Java\jdk`
3. Followed step includes , extracting jre as `**C:\Java\jre**`
4. Set env variables as

   User variable : JAVA_HOME : **C:\Java\jdk**

   System variable path : **C:\Java\jdk\bin**
5. Confirm java installation from cmd : **java -version**
6. Download and extract **Hadoop - 3.2.4** as (**C:\Hadoop**)
7. Go to etc folder and make changes in necessary files:
   a) hadoop-env.cmd :

      `set JAVA_HOME=C:\Java\jdk`

   b) core-site.xml :

```
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

c) Create a folder in **C:\hadoop named data such that C:\hadoop\data** and two sub-folders named **datanode** and **namenode** i.e. **C:\hadoop\data\datanode** and **C:\hadoop\data\namenode**

d) Paste the below content in hdfs-site.xml

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>

<property>
<name>dfs.namenode.name.dir</name>
<value>C:\hadoop\data\namenode</value>
</property>

<property>
<name>dfs.datanode.name.dir</name>
<value>C:\hadoop\data\datanode</value>
</property>
</configuration>
```

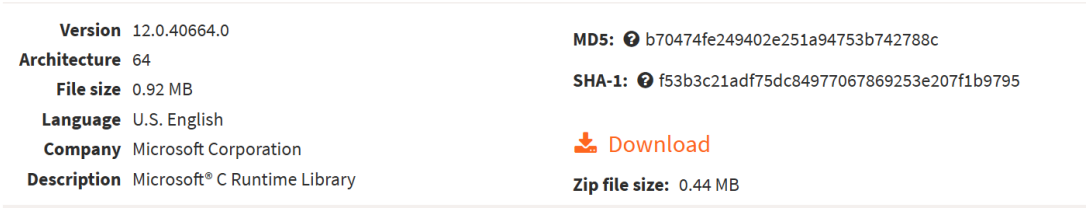
e) Mapred-site.xml

```
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>local</value>
</property>
</configuration>
```

f) Yarn-site.xml

```
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.shuffleHandler</value>
</property>
</configuration>
```

8. Delete the existing bin from Hadoop folder and replace it with new bin (https://drive.google.com/drive/folders/1iURNbow2IglhAhSy3sfY5xxVfAg33NBW)
9. Try running winutils.exe from this new bin folder , if it does not run and throws error,
10. Go to https://www.dll-files.com/msvcr120.dll.html



Download the above-mentioned file

Extract it and copy **mscvr120.dll** file and paste it in **C:\Windows\System32** such that **C:\Windows\System32\mscvr120.dll** exists

11. Setup env variables
    a) User variables:

       HADOOP_HOME : **C:\hadoop**

       HADOOP_USER_CLASSPATH_FIRST : true

    b) System variable path:

       C:\hadoop\bin

       C:\hadoop\sbin

12. Open command prompt in admin mode and

```
> hdfs namenode -format
> jps
> start-all.cmd
> jps
```

13. Check for localhost if working (localhost:9870) & (localhost:8088)
14. Done!!

Hive Installation Guidelines

1. Prerequisite: **Hadoop 3.2.4**
2. Download and Configure Apache Derby:
    a) Download **Derby 10.14.2.0** from

       https://db.apache.org/derby/derby_downloads.html

    b) Extract the file

    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** from https://archive.apache.org/dist/hive/hive-3.1.2/
    b) Extract the file
    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 (https://drive.google.com/file/d/1_h8dN46KK8zZf9yf2M2DBUfPR_IZl33T/view).
    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 verison **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 deamons**:
   a) Command to run deamons: `start-all.cmd`
   b) Commands to check deamons running: `jps`
10. Run Derby Server - Command: `StartNetworkServer -h 0.0.0.0`
11. To Initialize Hive Schema: Navigate to **C:\hive\bin** and run below command: `> hive --service schematool -dbType derby -initSchema`
12. Start Hive - Command: `hive`
13. You will be prompted with the Hive command line, ready to run SQL commands.