1. **Java 8 Installation:**

**Follow steps 2->3->4->6 of this url**

**Url:** [**https://docs.datastax.com/en/jdk-install/doc/jdk-install/installOpenJdkDeb.html**](https://docs.datastax.com/en/jdk-install/doc/jdk-install/installOpenJdkDeb.html)

**Setting environment variable JAVA\_HOME:**

1. Open bashrc file -
   1. nano  .bashrc
2. Add the following line:
   1. export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64
   2. export JAVA\_HOME
   3. export PATH=$PATH:$JAVA\_HOME/bin
3. Run this command
   1. Source .bashrc
4. Open profile file:
   1. sudo gedit /etc/profile
5. Add the following lines:
   1. export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64
   2. export JAVA\_HOME
   3. export PATH=$PATH:$JAVA\_HOME/bin

1. **Install Hadoop:**

**Version: 3.3.3**

**In the given Url there is a step where you have to wget and give the link for hadoop**

**Give this link after wget:**

[**https://downloads.apache.org/hadoop/common/hadoop-3.3.3/hadoop-3.3.3.tar.gz**](https://downloads.apache.org/hadoop/common/hadoop-3.3.3/hadoop-3.3.3.tar.gz)

**Url:** [**https://phoenixnap.com/kb/install-hadoop-ubuntu**](https://phoenixnap.com/kb/install-hadoop-ubuntu)

**Note: In the next steps, Spark and Hive need to be installed using the same ‘hdoop’ user which was created during Hadoop Installation**

1. **Install spark:**

**Skip step 1 of this url,Follow from step 2**

**Url:** <https://computingforgeeks.com/how-to-install-apache-spark-on-ubuntu-debian/>

Note: During Step 3, if you face error when you use the command “start-master.sh”, try executing “**sudo chmod -R 777 /opt/spark**” and then repeat the step to start the master.

1. **Install Hive:**

**Url:** [**https://data-flair.training/blogs/apache-hive-installation/**](https://data-flair.training/blogs/apache-hive-installation/)

Note: In Step 4 of this document, do not put space after “=” sign while updating the .bashrc file

**Follow the above link till step 9**

**After that run the following commands**

**Cd $HIVE\_HOME/bin**

**HIVE**

**Hive shell will start**

**if it will show following error**

**(HiveException java.lang.RuntimeException: Unable to instantiate org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient)**

**Follow this link**

**Url:** [**https://sparkbyexamples.com/apache-hive/hive-hiveexception-runtimeexception-unable-to-instantiate-sessionhivemetastoreclient/**](https://sparkbyexamples.com/apache-hive/hive-hiveexception-runtimeexception-unable-to-instantiate-sessionhivemetastoreclient/)

1. **Install Python:**

**Url:** [**How to install anaconda on Ubuntu 20.04 LTS | Download and Install | 2021**](https://www.youtube.com/watch?v=PHkCmuzgHOo)

1. **Install Mysql:**

**Url:** [**https://phoenixnap.com/kb/install-mysql-ubuntu-20-04**](https://phoenixnap.com/kb/install-mysql-ubuntu-20-04)

1. **Install Airflow**

[**https://hevodata.com/learn/install-airflow/**](https://hevodata.com/learn/install-airflow/)

**Or**

* Download : <https://www.hackerxone.com/2021/10/15/steps-to-install-and-setup-apache-airflow-on-ubuntu-20-04-lts/>
* Open another terminal and run the following command:
  + airflow scheduler

**Web Server: localhost:8080**

**For unmet dependencies error in Ubuntu while installing any package**

* **sudo rm /var/lib/dpkg/lock**
* **sudo rm /var/lib/apt/lists/lock**
* **sudo rm /var/cache/apt/archives/lock**
* **sudo apt --fix-broken install**
* **sudo apt autoclean**
* **sudo apt update**
* **sudo apt upgrade**
* **sudo apt autoremove**
* **reboot**