Apache Spark Installation

=======================================================================

The Spark installation in standalone mode

**Step1:** Check Java Version

java -version

**Step2:** Download apache spark from from official website

[Apache Spark™ - Unified Analytics Engine for Big Data](https://spark.apache.org/)

**spark-2.4.5-bin-hadoop2.7.tgz**

**Step3:** Extract that tgz file using command line or we can directly extract

After extracting that file we can see folder

**Spark-2.4.5-bin-hadoop2.7**

**Step4:** The **bin** and **sbin** folders path in **Spark-2.4.5-bin-hadoop2.7**

We have to update that path in bashrc

Open bashrc using following command - **$nano** **~/.bashrc**

Add the following properties:

**export SPARK\_PATH = /home/hadoopuser/spark/spark-2.4.5-bin-hadoop2.7**

**export PATH = $PATH:$SPARK\_PATH/bin**

**export PATH = $PATH:$SPARK\_PATH/sbin**

**export PYSPARK\_PYTHON=python3**

**export PYSPARK\_DRIVER\_PYTHON=jupyter**

**export PYSPARK\_DRIVER\_PYTHON=notebook**

Now we have to save the bashrc file using following command

**source ~/.bashrc**

**Step5:** Now we have to generate the ssh keys using following command

**ssh-keygen -t rsa -p “”**

**Step6:** Now we have to copy the generated rsa keys to authorized keys using following command

**$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys1**

**Step7:** now we have to run spark by going to spark\_home and run following command

**$ sbin/start-all.sh**