

# *Spark Installation Document*

**Distribution:** Apache

**Version:** 2.2.0

- 1) Copy the **Spark** directory from pendrive to home directory (/home/jpasolutions)
- 2) Open the terminal (Ctrl+Alt+t)
- 3) Change the directory to spark  
~\$ cd ~/spark
- 4) Extract the tar file under home directory (/home/jpasolutions)  
~\$ tar -xvzf spark-2.2.0-bin-hadoop2.7.tgz -C \$HOME
- 5) Set up the environment variables in ~/.bashrc file  
export SPARK\_HOME= /home/jpasolutions/spark-2.2.0-bin-hadoop2.7  
export PATH=\$PATH:\$SPARK\_HOME/bin
- 6) Environment variables to reflect on the current shell, source the file.  
~\$ source ~/.bashrc
- 7) Check the changes applied properly  
~\$ echo \$SPARK\_HOME  
~\$ which spark-shell
- 8) To access Hadoop files in Spark, HADOOP\_CONF\_DIR or YARN\_CONF\_DIR should be set.  
Check if the HADOOP\_CONF\_DIR is set on your system using below command  
~\$ echo \$HADOOP\_CONF\_DIR
- 9) Login into spark interactive shell  
~\$ spark-shell
- 10) To make sure the spark-shell started properly without errors, execute below commands  
~\$ print(sc)  
~\$ print(spark)