Readme Edureka VM !!

edureka!

Version 2.0

Please read this document carefully before using the EDUREKA VM.

Below are important points about this VM, please go through it without fail.

1) Hadoop and all other components are present in /usr/lib/

JDK	:	/usr/lib/jvm/jdk1.7.0_67
eclipse	:	/home/edureka/eclipse/java-mars/eclipse
hadoop	:	/usr/lib/hadoop-2.2.0
pig	:	/usr/lib/pig-0.12.0
hive	:	/usr/lib/hive-0.13.1-bin
hbase	:	/usr/lib/hbase-0.96.2-hadoop2
oozie	:	/usr/lib/oozie-4.0.0
sqoop	:	/usr/lib/sqoop-1.4.4
flume-ng	:	/usr/lib/flume-ng
spark	:	/usr/lib/spark-1.5.2
scala	:	/usr/lib/spark-1.5.2/build/scala-2.10.4

2) The paths of all the components are set.

JDK	:	.bashrc
hadoop	:	.bashrc
spark	:	.bashrc
scala	:	.bashrc
pig	:	/etc/profile.d/pig.sh
hive	:	/etc/profile.d/hive.sh
hbase	:	/etc/profile.d/hbase.sh
oozie	:	/etc/profile.d/oozie.sh
sqoop	:	/etc/profile.d/sqoop.sh

3) There is a LMS directory on Vm's Desktop which contains all the files for practical in your LMS module-wise.

4) Flume is already present. To execute flume, follow below commands:

i) cd /usr/lib/flume-ng/apache-flume-1.4.0-bin/bin/

ii) ./flume-ng agent -n TwitterAgent -c conf -f /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf

After sometime press ctrl + c to stop streaming the data. Go and check /user/flume/tweets on your HDFS, a flume file will be there which has streamed data.

5) To start Pig, type pig in the terminal, this will give you pig grunt shell.

6) To start Hive, type hive in the terminal, this will give you hive shell.

7) To start HBase, follow below commands:

i) cd /usr/lib/hbase-0.96.2-hadoop2/

ii) ./bin/start-hbase.sh

iii) hbase shell

In case hbase did not start or not working properly, follow below steps:

i) cd /usr/lib/hbase-0.96.2-hadoop2/

ii) sudo rm -r hbasestorage/*

iii) sudo rm -r logs/*

iv) ./bin/start-hbase.sh

v) hbase shell

8) To start Oozie, follow below steps:

i) cd /usr/lib/oozie-4.0.0/

ii) ./bin/oozie-start.sh

Now open mozilla browser in your VM, and check below URL, you should get Oozie Web Console.

localhost:11000

9) Sqoop is already present in VM. MySQL connector is present in /usr/lib/sqoop-1.4.4/lib/ directory.

10) When you are trying to access HDFS, you may get "NameNode is in SafeMode" Then go to terminal and give the command "hadoop dfsadmin -safemode leave " Now go and check your HDFS.

11) When you are closing the VM, use the option "Save the machine state", so that when you restart the VM you are at same place where you left and all your daemons are running.

12) In case you loose any daemon In case you loose any daemon, run below commands one by one:

Command: sudo service hadoop-master stop
Command: sudo service hadoop-master start
Command: hadoop dfsadmin -safemode leave

13) Below are the links for few documents that we prepared. These were for CDH3 VM, same steps are followed on CentOS also. (Just for reference)

i) Flume installation:
https://edureka.wistia.com/medias/hsby0ufqqo/download?media_file_id=48186073

ii) Sqoop installation:
https://edureka.wistia.com/medias/c1l2lcyilg/download?media_file_id=48186037

14) Run the command "spark-shell" in the terminal to start Apache Spark.

15) Eclipse IDE in this virtual machine has scala packages pre-installed.