HDFS

HDFS

Hadoop
Distributed
File
System

Brief History of HDFS – GFS (google filesystem)

HDFS

HDFS Practice

Pre-requisite:

Hadoop running in pseudo-distributed or distributed mode.

HDFS

HDFS Practice

Commands begin with:

$ hdfs dfs

or

$ hadoop fs

HDFS Commands

PRACTICE 1 - format namenode
CAREFUL WITH EXISTING HDFS DATA!!!

$ hdfs namenode -format

<datetime> INFO namenode.NameNode: STARTUP_MSG:
(...)

HDFS Commands

PRACTICE 2 – start distributed filesystem

$ start-dfs.sh

HDFS Commands

PRACTICE 3 – list top level directory

$ hdfs dfs -ls /

HDFS Commands

PRACTICE 4 - Create directory structure

$ hdfs dfs -mkdir /input

sub-directory

$ hdfs dfs -mkdir /input/<initial><surname>

HDFS Commands

PRATICE 5 – create local file

$ date > ls.txt
$ ls / >> ls.txt
# check
$ cat ls.txt

PRACTICE 6 – move local file to HDFS

$ hdfs dfs -moveFromLocal ls.txt /input/<mydir>

HDFS Commands

PRATICE 7 – list recursively

$ hdfs dfs -ls -R /

PRATICE 8 – stop HDFS

$ stop-dfs.sh

HDFS Commands

PRATICE 9 – list (error – HDFS not running)

$ hdfs dfs -ls -R

PRATICE 10 – start HDFS

$ start-dfs.sh

HDFS Commands

PRACTICE 11 – assert data is still in HDFS

$ hdfs dfs -ls -R /

PRATICE 12 – list all commands

$ hdfs dfs

PRATICE 13 – access HDFS Web Interface

http://<ip address>:50070

HDFS Commands

PRACTICE 14 – stop HDFS

$ stop-dfs.sh

PRACTICE 15 – format namenode (!)

$ hdfs namenode -format

HDFS Commands

PRACTICE 16 – start HDFS

$ start-dfs.sh

PRACTICE 17 – assert data is gone

$ hdfs namenode -format

WebHDFS REST API

REST API