**Exercise n**

*Get started with Hadoop*

**Prior Knowledge**

Unix Command Line Shell

Simple Python

**Learning Objectives**

Understand how to start and stop HDFS, and transfer files into and out of the Hadoop filesystem.

Understand how to create a Python mapper/reducer and execute this on a single-node setup of Hadoop.

Stop and start YARN.

**Software Requirements**

(see separate document for installation of these)

* Apache Hadoop 2.7.0
* Python 2.7.x
* Nano text editor or other text editor

**Part A: Hadoop File System (HDFS)**

1. Make sure you are running the Ubuntu VM, and start a fresh terminal window.
2. Type  
   sudo su hduser  
   You will be prompted  
   [sudo] password for oxclo:  
   Type *oxclo* and enter.
3. Type cd and then enter, to switch to the home directory of user *hduser.*
4. Now let’s format the HDFS filesystem:  
   hadoop namenode -format
5. You should see a lot of output ending something similar to this:

15/10/22 09:15:20 INFO namenode.FSNamesystem: Retry cache on namenode is enabled

15/10/22 09:15:20 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis

15/10/22 09:15:20 INFO util.GSet: Computing capacity for map NameNodeRetryCache

15/10/22 09:15:20 INFO util.GSet: VM type = 64-bit

15/10/22 09:15:20 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB

15/10/22 09:15:20 INFO util.GSet: capacity = 2^15 = 32768 entries

15/10/22 09:15:20 INFO namenode.FSImage: Allocated new BlockPoolId: BP-420615264-127.0.1.1-1445501720083

15/10/22 09:15:20 INFO common.Storage: Storage directory /usr/local/hadoop\_store/hdfs/namenode has been successfully formatted.

15/10/22 09:15:20 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0

15/10/22 09:15:20 INFO util.ExitUtil: Exiting with status 0

15/10/22 09:15:20 INFO namenode.NameNode: SHUTDOWN\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SHUTDOWN\_MSG: Shutting down NameNode at oxclo/127.0.1.1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

1. Now let’s start the Hadoop filesystem. Type:   
     
   start-dfs.sh  
     
   You should see output like:

15/10/22 09:18:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Starting namenodes on [localhost]

localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-oxclo.out

localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-oxclo.out

Starting secondary namenodes [0.0.0.0]

0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-oxclo.out

15/10/22 09:19:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

1. Now let’s make a directory:  
     
   hadoop fs -mkdir -p /usr/hduser/input
2. And check it worked:

hadoop fs -ls -R /

You should see:

15/10/22 09:24:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

drwxr-xr-x - hduser supergroup 0 2015-10-22 09:23 /usr

drwxr-xr-x - hduser supergroup 0 2015-10-22 09:23 /usr/hduser

drwxr-xr-x - hduser supergroup 0 2015-10-22 09:23 /usr/hduser/input