**Hadoop Exercise-7 Q2**

**Question 2 ) Implement a simple map-reduce code for the wordcount problem using Python and then run the code using HDFS**

1. Start all Deamons

A screen shot of a computer

Description automatically generated

1. Directory Structure in Local Filesystem

/home/hadoop/hadoop/

├── apriori\_python/

├── input.txt

├── mapper.py

├── reducer.py

├── random\_input.txt

A screenshot of a computer

Description automatically generated

1. Directory Structure in HDFS (or) Browser Directory

/hadoop/ (user)

├── hadoop/ (parent folder)

├── apriori\_python/

├── input/

├── input.txt

├── output/

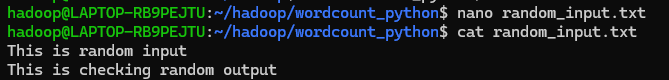
├── \_SUCCESS

├── part-00000

1. Set up Directory Structure



1. Create custom input to check the mapper and reducer programs



1. Create a Mapper in Python

NOTE: DO NOT REMOVE THE #! LINE. ‘ #!’ IS CALLED A SHEBAND AND INDICATES THE SCRIPT IS BEING INTERPRETED IN A PYTHON ENVIRONMENT INSTEAD OF A SHELL ENVIRONMENT.

A computer screen shot of a code

Description automatically generated

1. Create a Reducer in Python

NOTE: DO NOT REMOVE THE #! LINE. ‘ #!’ IS CALLED A SHEBAND AND INDICATES THE SCRIPT IS BEING INTERPRETED IN A PYTHON ENVIRONMENT INSTEAD OF A SHELL ENVIRONMENT.

A computer screen shot of a code

Description automatically generated

1. Check if they run locally using streaming

A computer screen with white text

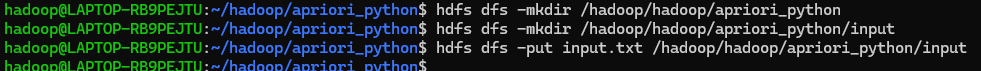
Description automatically generated

1. Create the actual Input file

A black screen with blue text

Description automatically generated

1. Create the directory structure in Hadoop for input



1. Change the permissions for the mapper and reducer file else it will give error while doing Hadoop streaming



1. Use Hadoop Streaming to run the job

-input : location of input in HDFS

-output : where you want the output to be stored in HDFS

-mapper : path of mapper in local filesystem (the dot before the first slash indicates cwd)

-reducer : path of reducer in local filesystem (the dot before the first slash indicates cwd)

A screen shot of a computer

Description automatically generated

1. View the results

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Stop all daemons

