W**eek 3 A): PIG EXECUTION : NORMAL TEXT FILE (HOW TO EXECUTE & RUN THE PROGRAM LOCALLY & TEST IT ON HADOOP)**

1. Create input.txt file :

**Content :**

hi how are you

i am good

hope you doing good too

how about you.

i am in manipal

studying Btech in Data science.

1. Transfer to HDFS: hdfs dfs -put /home/hdoop/input.txt bda1/
2. Create Pigscript file: sudo gedit pigscript.pig

**Content :** record = load '/bda1/input.txt/';

store record into '/bda1/out';

pig -x mapreduce pigscript.pig

1. hdfs dfs -ls bda1/input
2. hdoop@hadoop-master:~$ hdfs dfs -ls /bda1/out

**Output:** Found 2 items

-rw-r--r-- 2 hdoop supergroup 0 2024-01-12 15:05 /bda1/out/\_SUCCESS

-rw-r--r-- 2 hdoop supergroup 112 2024-01-12 15:05 /bda1/out/part-m-00000

1. hdoop@hadoop-master:~$ pig -x mapreduce pigscript.pig
2. HadoopVersion PigVersion UserId StartedAt FinishedAt Features

3.3.6 0.17.0 hdoop 2024-01-12 15:05:13 2024-01-12 15:06:55 UNKNOWN

**Output:**

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs

job\_1705045524875\_0010 1 0 n/a n/a n/a n/a 0 0 0 0 record MAP\_ONLY /bda1/out,

Input(s):

Successfully read 0 records from: "/bda1/input.txt"

Output(s):

Successfully stored 0 records in: "/bda1/out"

Counters:

Total records written : 0

Total bytes written : 0

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1705045524875\_0010

1. hdoop@hadoop-master:~$ hdfs dfs -cat /bda1/out/part-m-00000

**Output:**

hi how are you

i am good

hope you doing good too

how about you.

i am in manipal

studying Btech in Data science.

**Week 3 B): WORD COUNT PROGRAM:**

1. sudo gedit wordcount.pig

**Content:**

--LOAD THE DATA

records = LOAD '/pig1/input.txt';

-- SPLIT EACH LINE OF TEXT AND ELIMINATE NESTING

terms = FOREACH records GENERATE FLATTEN(TOKENIZE((chararray) $0)) AS word;

--GROUP SIMILAR TERMS

grouped\_terms = GROUP terms BY word;

--COUNT THE NUMBER OF TUPLES IN EACH GROUP

word\_counts = FOREACH grouped\_terms GENERATE COUNT(terms), group;

--STORE THE RESULT

STORE word\_counts INTO '/pig1/output';

1. pig : (it will take default as MAPREDUCE)
2. grunt>pig wordcount.pig

**Output:**

024-01-12 15:43:25,571 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)

2024-01-12 15:43:25,673 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.

2024-01-12 15:43:25,673 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2024-01-12 15:43:25,706 [main] INFO org.apache.pig.Main - Pig script completed in 2 minutes, 48 seconds and 931 milliseconds (168931 ms)

1. grunt>run wordcount.pig

**Output:**

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

3.3.6 0.17.0 hdoop 2024-01-12 16:25:21 2024-01-12 16:27:08 GROUP\_BY

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs

job\_1705045524875\_0014 1 1 n/a n/a n/a n/a n/a n/a n/an/a grouped\_terms,records,terms,word\_counts GROUP\_BY,COMBINER /pig1/output,

Input(s):

Successfully read 0 records from: "/pig1/input.txt"

Output(s):

Successfully stored 0 records in: "/pig1/output"

Counters:

Total records written : 0

Total bytes written : 0

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1705045524875\_0014

1. grunt> pwd

hdfs://192.168.159.101:9000/user/hdoop

1. grunt> cd /pig1/output
2. grunt> ls

hdfs://192.168.159.101:9000/pig1/output/\_SUCCESS<r 2> 0

hdfs://192.168.159.101:9000/pig1/output/part-r-00000<r 2> 127

1. grunt> cat part-r-00000

2 i

2 am

1 hi

2 in

1 are

2 how

1 too

2 you

1 Data

2 good

1 hope

1 you.

1 Btech

1 about

1 doing

1 manipal

1 science.

1 studying