**Creation of txt files**

root@4c77d6c6ae20:/# vim episodeIV\_dialouges.txt

root@4c77d6c6ae20:/# vim episodeV\_dialouges.txt

root@4c77d6c6ae20:/# vim episodeVI\_dialouges.txt

**Navigating into Project\_Activity1 folder for 3 text files**

root@4c77d6c6ae20:/# hdfs dfs -mkdir /user/root/Project\_Activity1

root@4c77d6c6ae20:/# hdfs dfs -ls

Found 2 items

drwxr-xr-x - root supergroup 0 2021-11-30 03:40 Project\_Activity1

drwxr-xr-x - root supergroup 0 2021-08-23 10:44 input

root@4c77d6c6ae20:/# hdfs dfs -put ./episodeIV\_dialouges.txt /user/root/Project\_Activity1

root@4c77d6c6ae20:/# hdfs dfs -put ./episodeV\_dialouges.txt /user/root/Project\_Activity1

root@4c77d6c6ae20:/# hdfs dfs -put ./episodeVI\_dialouges.txt /user/root/Project\_Activity1

root@4c77d6c6ae20:/# hdfs dfs -ls /user/root/Project\_Activity1

Found 3 items

-rw-r--r-- 1 root supergroup 67670 2021-11-30 03:39 /user/root/Project\_Activity1/episodeIV\_dialouges.txt

-rw-r--r-- 1 root supergroup 43657 2021-11-30 03:40 /user/root/Project\_Activity1/episodeVI\_dialouges.txt

-rw-r--r-- 1 root supergroup 49890 2021-11-30 03:39 /user/root/Project\_Activity1/episodeV\_dialouges.txt

root@4c77d6c6ae20:/#

**activity1.pig**

inputFile = LOAD 'hdfs:///user/root/Project\_Activity1/episodeIV\_dialouges.txt' AS (name:chararray, line:chararray);

ranked = RANK inputDialogues;

words = FOREACH inputFile GENERATE FLATTEN(TOKENIZE(line)) AS word;

grouped = GROUP words BY word;

wordcount = FOREACH grouped GENERATE group, COUNT(words);

STORE wordcount INTO 'hdfs:///user/root/results';

**pig command result:**

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

3.3.1 0.17.0 root 2021-11-30 04:19:33 2021-11-30 04:22:26 GROUP\_BY

Success!

Job Stats (time in seconds):

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

job\_1638242045094\_0004 1 1 n/a n/a n/a n/a n/a n/a n/a n/a grouped,inputFile,wordcount,words GROUP\_BY,COMBINER hdfs:///user/root/results,

Input(s):

Successfully read 0 records from: "hdfs:///user/root/Project\_Activity1/episodeIV\_dialouges.txt"

Output(s):

Successfully stored 0 records in: "hdfs:///user/root/results"

root@4c77d6c6ae20:/# hdfs dfs -cat /user/root/results/part-r-00000

- 1

4: 1

BEN 76

HAN 152

RED 49

TAR 1

TEN 7

TWO 2

AUNT 6

BASE 1

BERU 7

CASS 1

DEAK 1

FIVE 7

GOLD 23

LEIA 57

LUKE 253

NINE 1

OVER 1

OWEN 25

STAR 8

WARS 2

BEN'S 6

BIGGS 34

CAMIE 2

CHIEF 3

DEATH 7

FIRST 7

FIXER 2

HAN'S 1

HUMAN 4

JABBA 6

MAN'S 1

MOTTI 4

PILOT 1

REBEL 1

SEVEN 1

TAGGE 4

TEN'S 1

VADER 41

VOICE 34

WEDGE 14

WOMAN 1

ELEVEN 1

GANTRY 2

GREEDO 6

LEADER 50

LUKE'S 1

NINE'S 1

SECOND 4

TARKIN 28

CAPTAIN 1

CONTROL 1

DODONNA 6

EPISODE 1

OFFICER 20

PORKINS 1

TROOPER 29

WILLARD 2

WINGMAN 1

CREATURE 1

IMPERIAL 2

INTERCOM 16

LEADER'S 1

MASSASSI 3

THREEPIO 119

BARTENDER 3

COMMANDER 3

WINGMAN'S 1

TECHNICIAN 1

ASTRO-OFFICER 1

0

root@4c77d6c6ae20:/#



Solution Displayed

