**CSV File:** ep4\_activity2.csv

hive> show databases;

OK

default

Time taken: 0.298 seconds, Fetched: 1 row(s)

hive> create database episodeIV;

OK

Time taken: 0.062 seconds

hive> use episodeIV;

OK

Time taken: 0.149 seconds

hive> create table epV\_activity2

> (character STRING,dialog STRING)

> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> TBLPROPERTIES ("skip.header.line.count"="1");

OK

Time taken: 0.129 seconds

hive> DESCRIBE epV\_activity2;

OK

character string

dialog string

Time taken: 0.105 seconds, Fetched: 2 row(s)

hive> load data local inpath

> '/ep4\_activity2.csv'

> into table epV\_activity2;

Loading data to table episodeiv.epv\_activity2

OK

Time taken: 1.858 seconds

**Activity 1**

hive> select character , count(character) as no\_of\_lines from epV\_activity2 group by character order by no\_of\_lines;

Query ID = root\_20211130090316\_da984aa7-6836-4ff6-844a-5d0ded6dfe0c near 'no\_of\_lines'

Total jobs = 2haracter , count(character) as no\_of\_lines fromepV\_activity2 group by character order by no\_of\_lines;

Launching Job 1 out of 2

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1638242045094\_0012, Tracking URL = http://4c77d6c6ae20:8088/proxy/application\_1638242045094\_0012/

Kill Command = /usr/local/hadoop/bin/mapred job -kill job\_1638242045094\_0012

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2021-11-30 09:03:45,404 Stage-1 map = 0%, reduce = 0%

2021-11-30 09:03:58,401 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 10.84 sec

2021-11-30 09:04:11,666 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 18.56 sec

MapReduce Total cumulative CPU time: 18 seconds 560 msec

Ended Job = job\_1638242045094\_0012

Launching Job 2 out of 2

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1638242045094\_0013, Tracking URL = http://4c77d6c6ae20:8088/proxy/application\_1638242045094\_0013/

Kill Command = /usr/local/hadoop/bin/mapred job -kill job\_1638242045094\_0013

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2021-11-30 09:04:34,177 Stage-2 map = 0%, reduce = 0%

2021-11-30 09:04:42,830 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 3.16 sec

2021-11-30 09:04:52,527 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 7.83 sec

MapReduce Total cumulative CPU time: 7 seconds 830 msec

Ended Job = job\_1638242045094\_0013

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 18.56 sec HDFS Read: 62237 HDFS Write: 1530 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.83 sec HDFS Read: 9095 HDFS Write: 1333 SUCCESS

Total MapReduce CPU Time Spent: 26 seconds 390 msec

OK

FIRST CONTROLLER 1

ASSISTANT OFFICER 1

WOMAN CONTROLLER 1

CAPTAIN 1

STRANGE VOICE 1

SECOND THREEPIO 1

SECOND OFFICER 1

SECOND CONTROLLER 1

REBEL FIGHTER 1

REBEL CAPTAIN 1

PILOTS 1

PILOT 1

OFFICER 1

MAN'S VOICE 1

IMPERIAL SOLDIER 1

HOBBIE 1

HEAD CONTROLLER 1

LIEUTENANT 2

TRACKING OFFICER 2

IMPERIAL OFFICER 2

MEDICAL DROID 2

SENIOR CONTROLLER 2

COMMUNICATIONS OFFICER 2

INTERCOM VOICE 2

DERLIN 3

TRENCH OFFICER 3

ANNOUNCER 3

CONTROLLER 3

BEN'S VOICE 4

DACK 4

JANSON 4

BOBA FETT 4

OZZEL 5

EMPEROR 5

NEEDA 5

ZEV 6

DECK OFFICER 7

VEERS 7

WEDGE 8

BEN 11

RIEEKAN 13

CREATURE 21

PIETT 23

YODA 36

VADER 56

LANDO 61

THREEPIO 92

LEIA 114

LUKE 128

HAN 182

Time taken: 97.02 seconds, Fetched: 50 row(s)