**Episode 4**

#create table

hive> CREATE TABLE episodesIV\_counts

> (character STRING, dialouges STRING)

> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

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

#load data

hive> LOAD DATA LOCAL INPATH

> '/shrikant/episodeIV\_dialouges.txt'

> INTO TABLE episodesIV\_counts;

#Exports to HDFS directory

hive> INSERT OVERWRITE DIRECTORY '/user/root/output/exportIV'

> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

> SELECT character, count(character) FROM episodesIV\_counts group by character order by count(character) desc;

#check results in HDFS

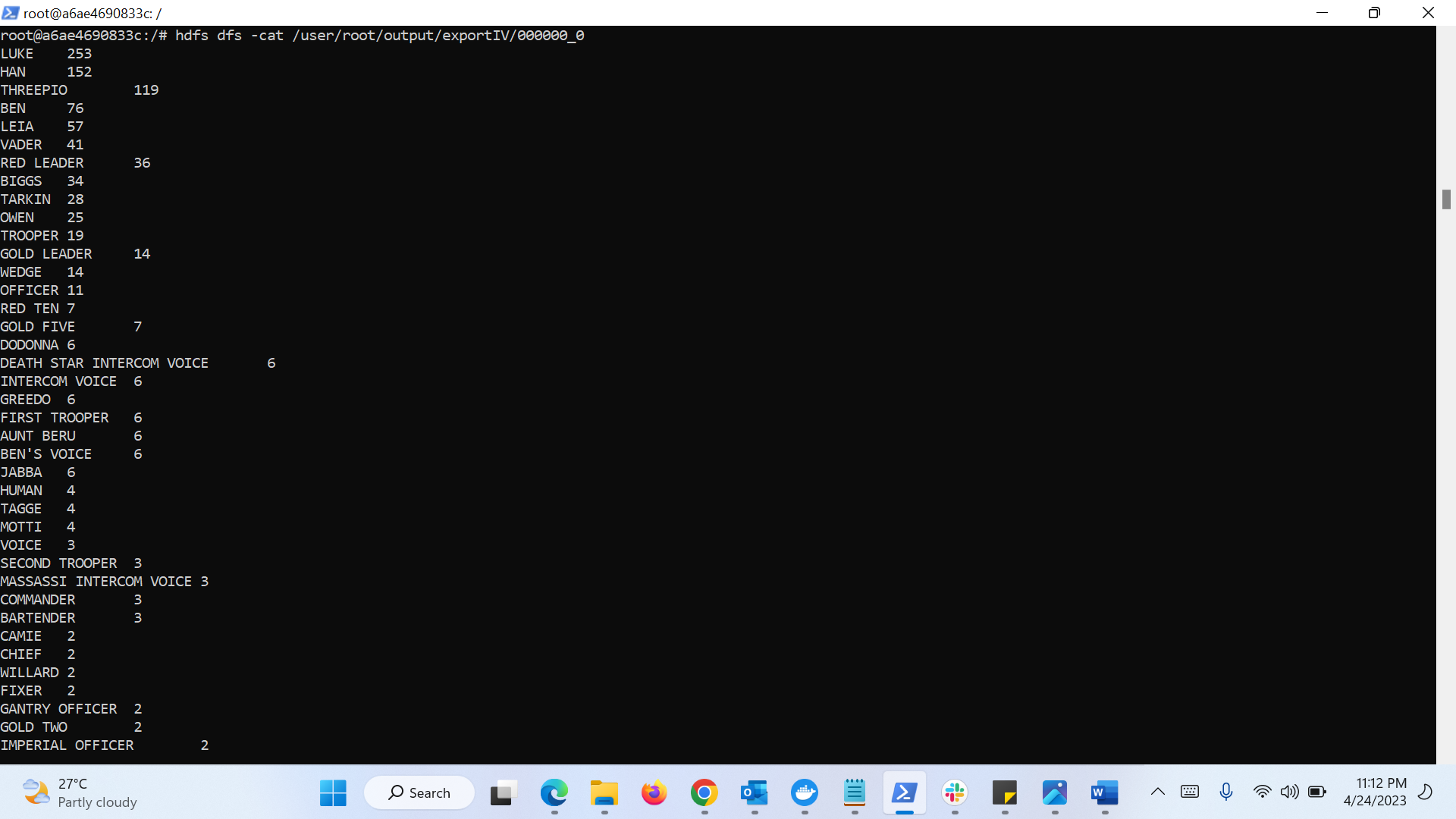
hdfs dfs -cat /user/root/output/exportIV/000000\_0

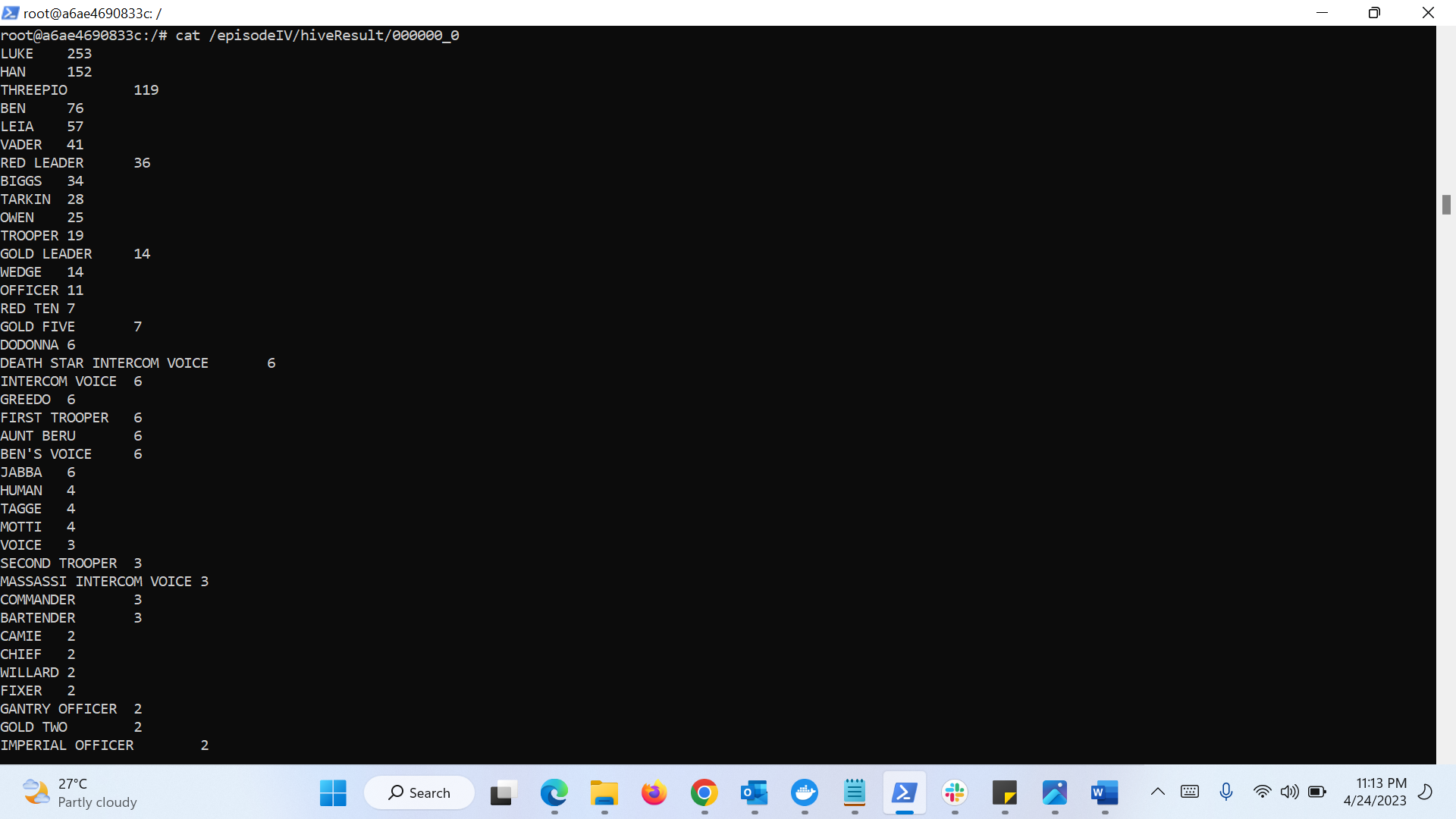
#check results in local

hdfs dfs -get /user/root/output/exportIV/000000\_0 /episodeIV/hiveResult

cat /episodeIV/hiveResult/000000\_0

Episode 4 results in HDFS



Episode 4 results in Local 

**Episode 5**

#create table

hive> CREATE TABLE episodesV\_counts

> (character STRING, dialouges STRING)

> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

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

#load data

hive> LOAD DATA LOCAL INPATH

> '/shrikant/episodeV\_dialouges.txt'

> INTO TABLE episodesV\_counts;

#Exports to HDFS directory

hive> INSERT OVERWRITE DIRECTORY '/user/root/output/exportV'

> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

> SELECT character, count(character) FROM episodesV\_counts group by character order by count(character) desc;

#check results in HDFS

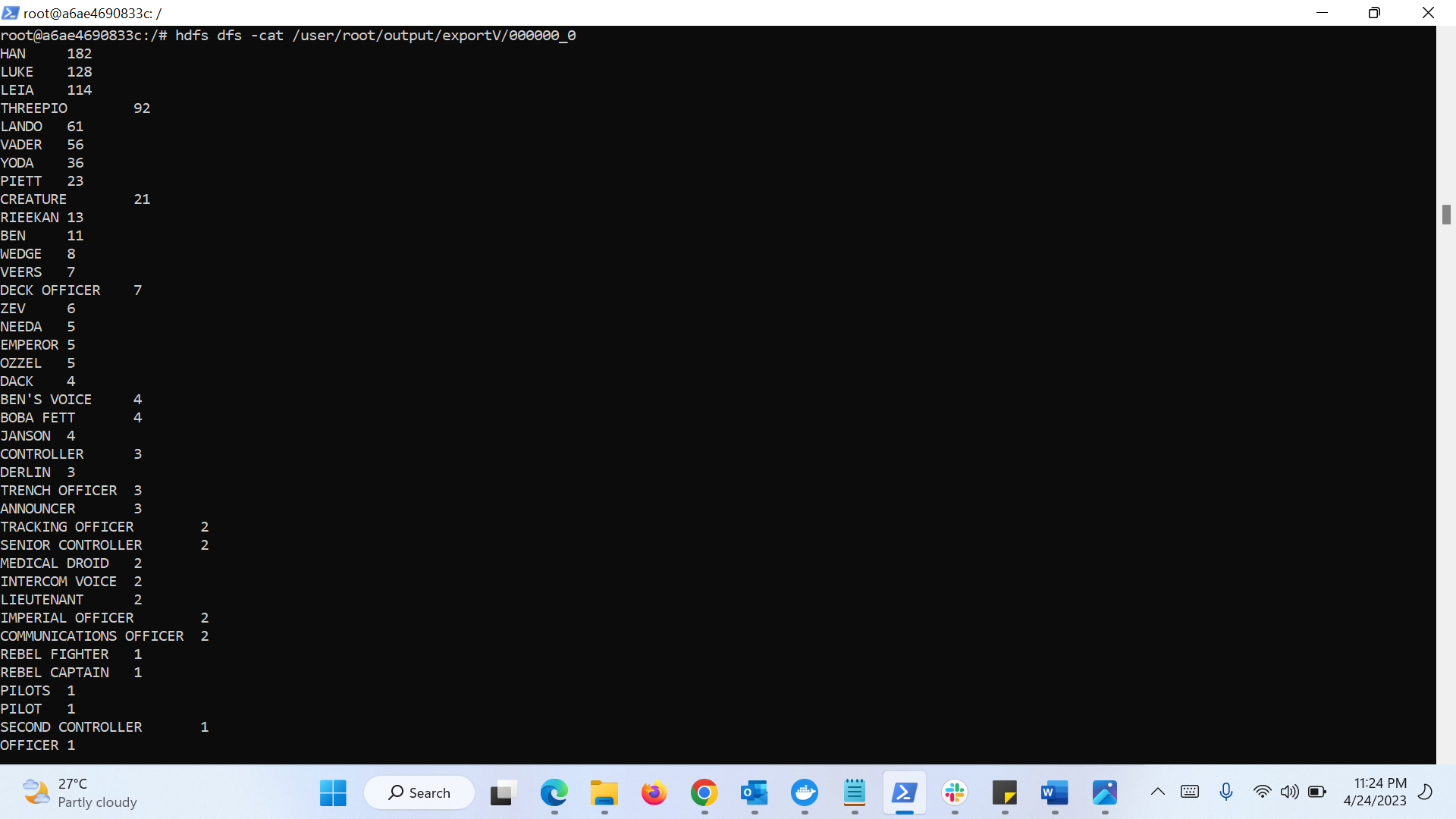
hdfs dfs -cat /user/root/output/exportV/000000\_0

#check results in local

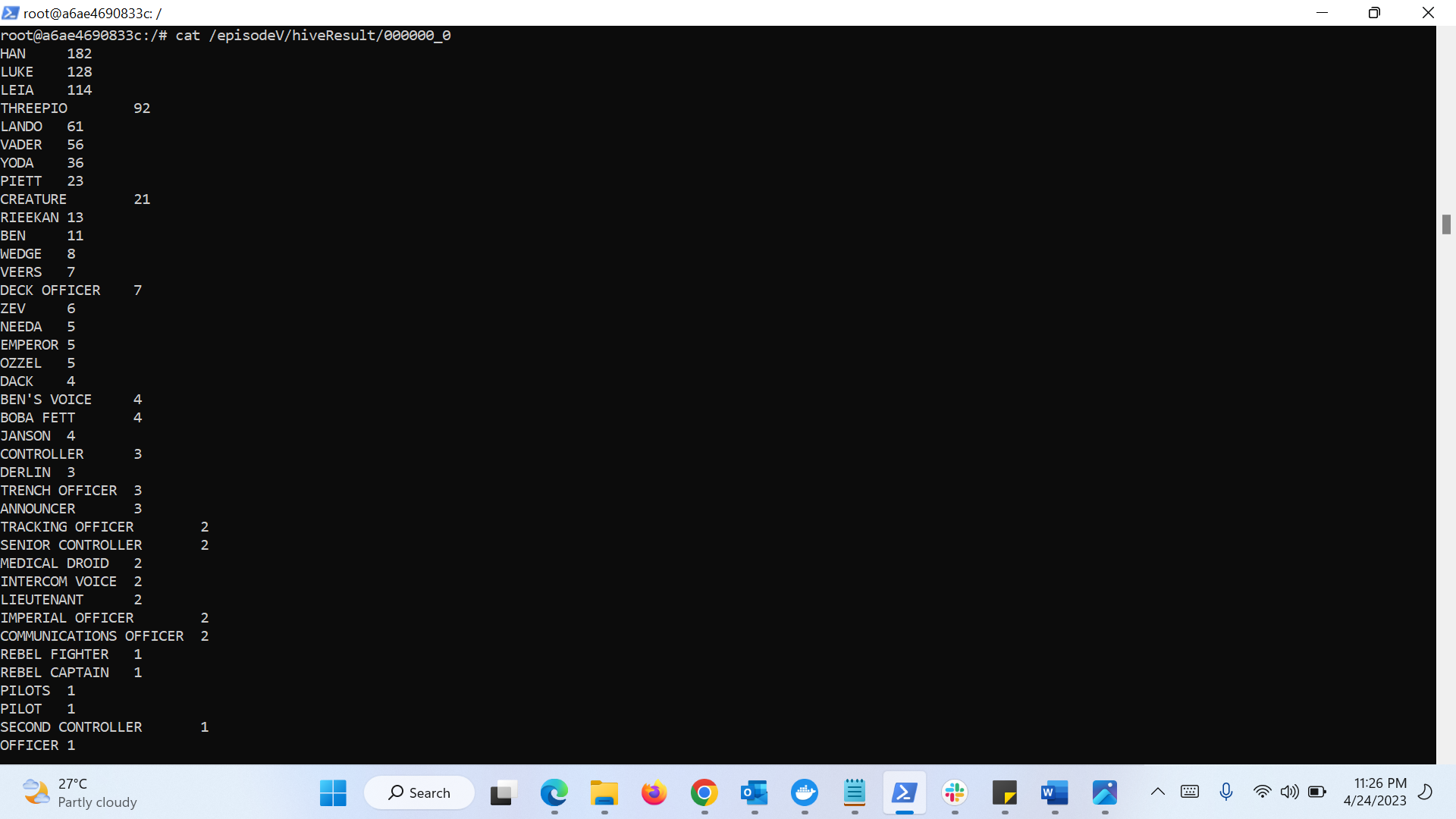
hdfs dfs -get /user/root/output/exportV/000000\_0 /episodeV/hiveResult

cat /episodeV/hiveResult/000000\_0

Episode 5 results in HDFS



Episode 5 results in Local



#create table

hive> CREATE TABLE episodesVI\_counts

> (character STRING, dialouges STRING)

> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

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

#load data

hive> LOAD DATA LOCAL INPATH

> '/shrikant/episodeVI\_dialouges.txt'

> INTO TABLE episodesVI\_counts;

#Exports to HDFS directory

hive> INSERT OVERWRITE DIRECTORY '/user/root/output/exportVI'

> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

> SELECT character, count(character) FROM episodesVI\_counts group by character order by count(character) desc;

#check results in HDFS

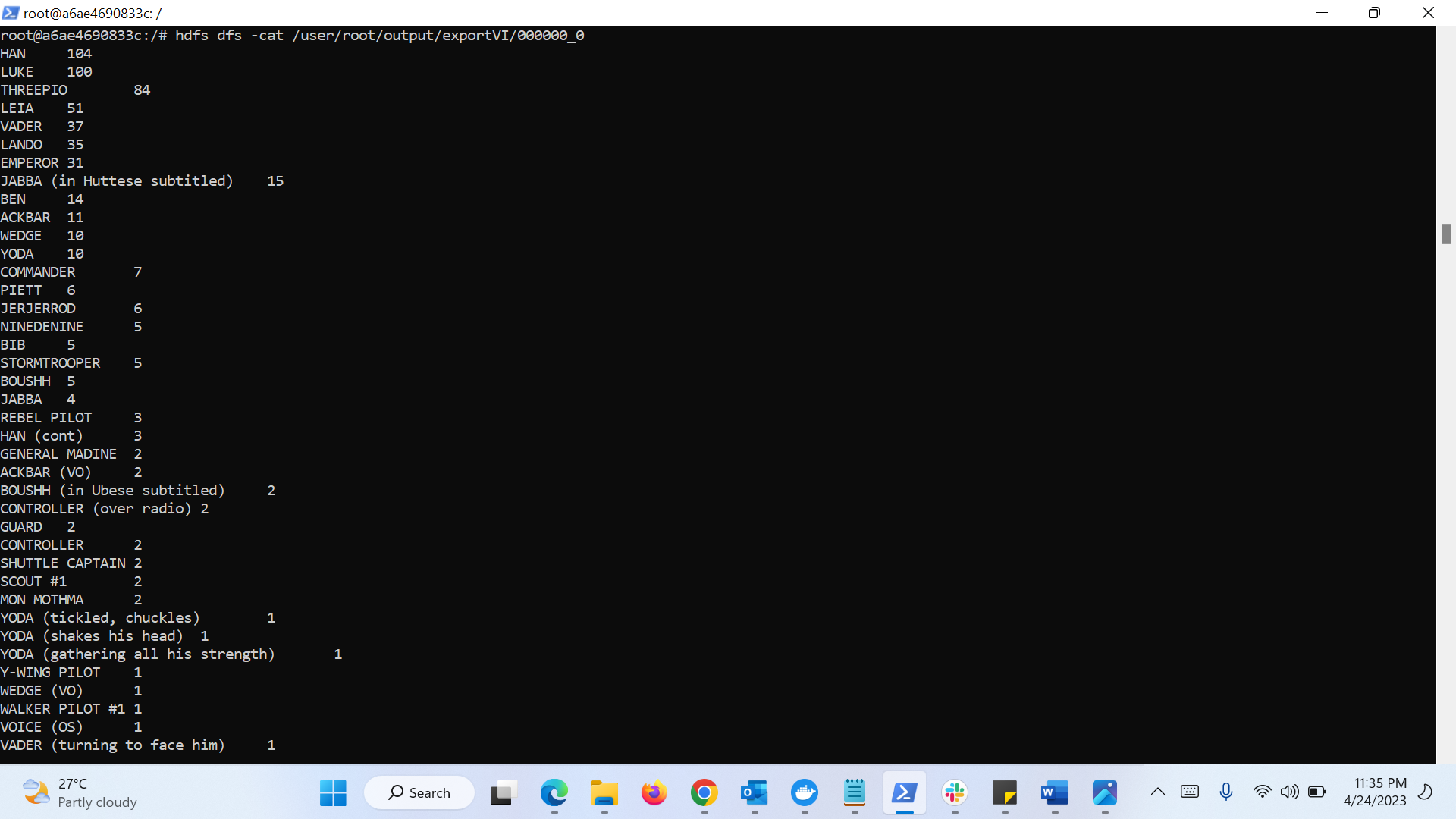
hdfs dfs -cat /user/root/output/exportVI/000000\_0

#check results in local

hdfs dfs -get /user/root/output/exportVI/000000\_0 /episodeVI/hiveResult

cat /episodeVI/hiveResult/000000\_0

Episode 6 results in HDFS



Episode 6 results in Local

