7c2e274a4619

-- Load input file from HDFS

inputFile = LOAD 'hdfs:///user/shilpa/episodeIV\_dialouges.txt' USING PigStorage('\t') AS (name:chararray, line:chararray);

-- Group data by name

grpd = GROUP inputFile BY name;

-- Count the occurence of each word (Reduce)

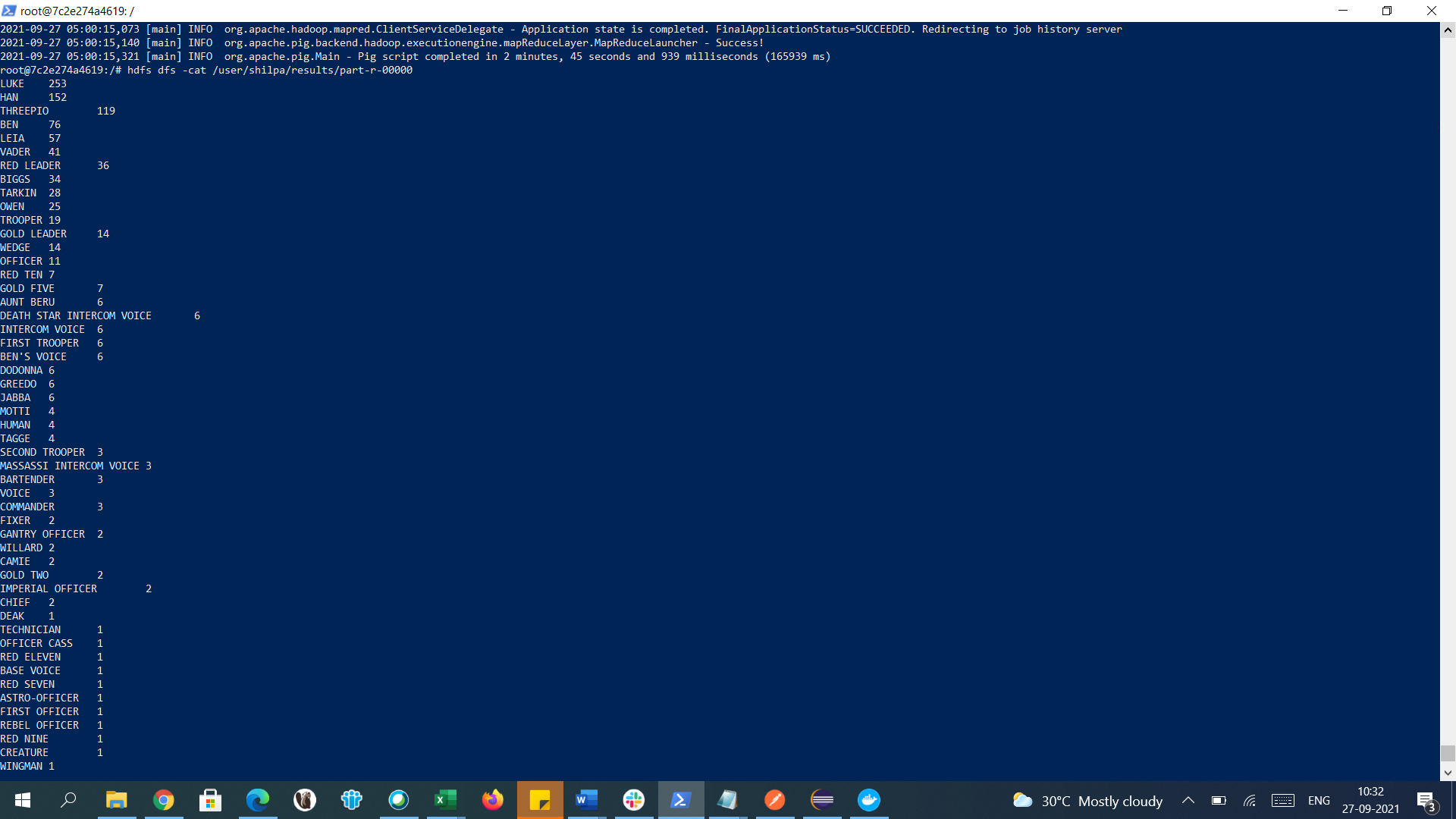
cntd = FOREACH grpd GENERATE group, COUNT(inputFile.name) AS cnt;

--Sort data

sorted = ORDER cntd BY cnt DESC;

-- Store the result in HDFS

STORE sorted INTO 'hdfs:///user/shilpa/results';



-- Load input file from HDFS

inputFile = LOAD 'hdfs:///user/shilpa/episodeV\_dialouges.txt' USING PigStorage('\t') AS (name:chararray, line:chararray);

-- Group data by name

grpd = GROUP inputFile BY name;

-- Count the occurence of each word (Reduce)

cntd = FOREACH grpd GENERATE group, COUNT(inputFile.name) AS cnt;

--Sort data

sorted = ORDER cntd BY cnt DESC;

-- Store the result in HDFS

STORE sorted INTO 'hdfs:///user/shilpa/results1';



inputFile = LOAD 'hdfs:///user/shilpa/episodeVI\_dialouges.txt' USING PigStorage('\t') AS (name:chararray, line:chararray);

-- Group data by name

grpd = GROUP inputFile BY name;

-- Count the occurence of each word (Reduce)

cntd = FOREACH grpd GENERATE group, COUNT(inputFile.name) AS cnt;

--Sort data

sorted = ORDER cntd BY cnt DESC;

-- Store the result in HDFS

STORE sorted INTO 'hdfs:///user/shilpa/results2';

