A= LOAD 'hdfs:///user/Project\_module4/Project\_module4/module4.txt' USING PigStorage('\t') As (author:chararray,line:chararray);

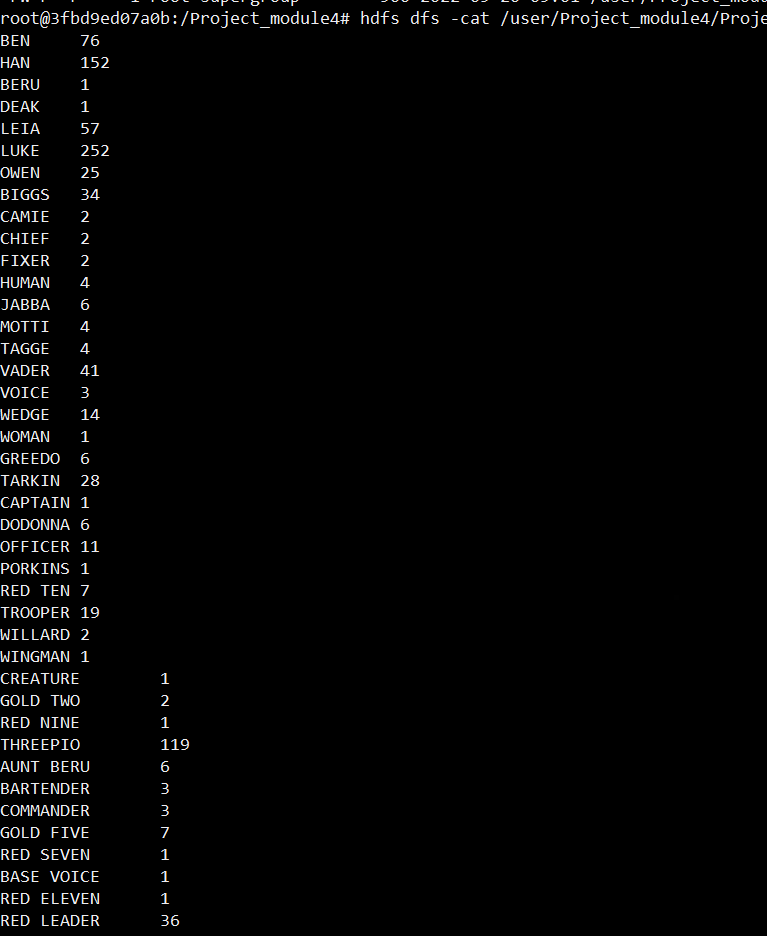
B= GROUP A by author;

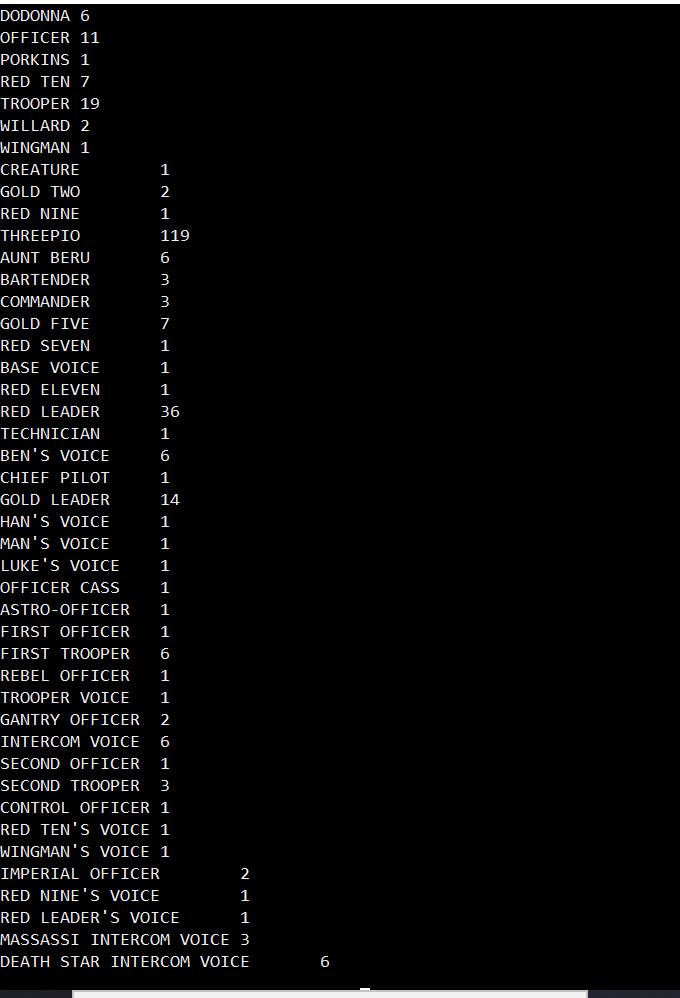
D= FOREACH B GENERATE (chararray)$0 as user,(chararray)COUNT($1) as number;

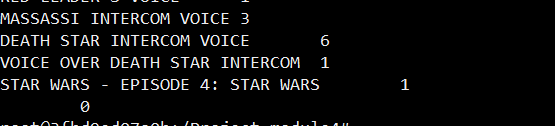
E = ORDER D BY number ASC;

Dump E

STORE D INTO 'hdfs:///user/Project\_module4/Project\_module4/result\_module67.txt';







Activity2Pig:-

A= LOAD 'hdfs:///user/Project\_module4/Project\_module4/module6.txt' USING PigStorage('\t') As (author:chararray,line:chararray);

B= GROUP A by author;

D= FOREACH B GENERATE (chararray)$0 as user,(chararray)COUNT($1) as number;

E = ORDER D BY number ASC;

Dump E

STORE D INTO 'hdfs:///user/Project\_module4/Project\_module4/result\_module69.txt';

Hive:-

Create TABLE file2

(ID STRING,conv STRING)

ROW FORMAT DELIMITED FIELDS TERMINATED by '\t';

LOAD DATA LOCAL INPATH

'/Project\_module4/module4.txt'

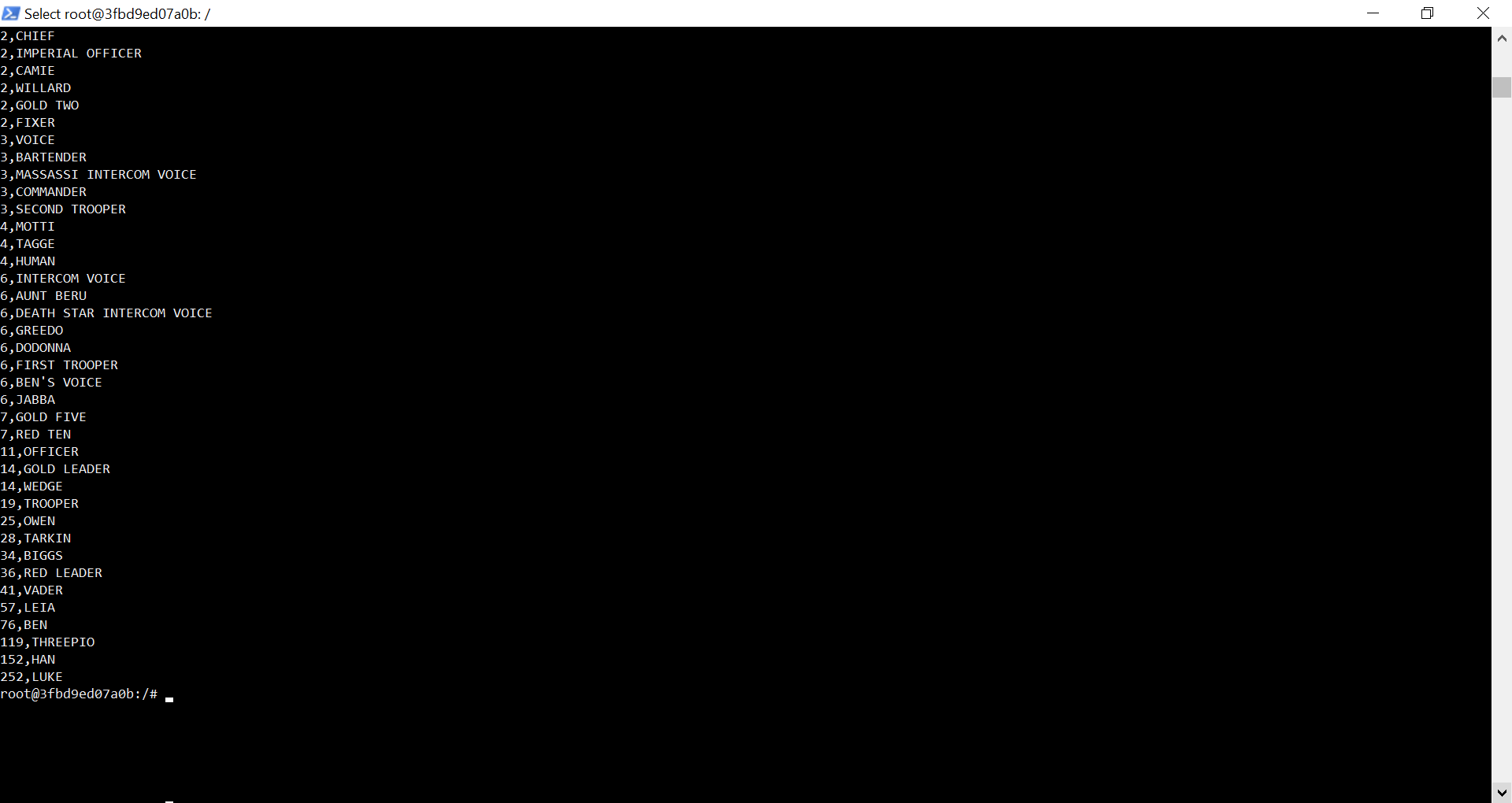
INTO TABLE file2;

hive> INSERT OVERWRITE DIRECTORY '/user/Project\_module4/result\_hive1.txt'

> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> select count(\*) AS count1, ID FROM file2 GROUP BY ID

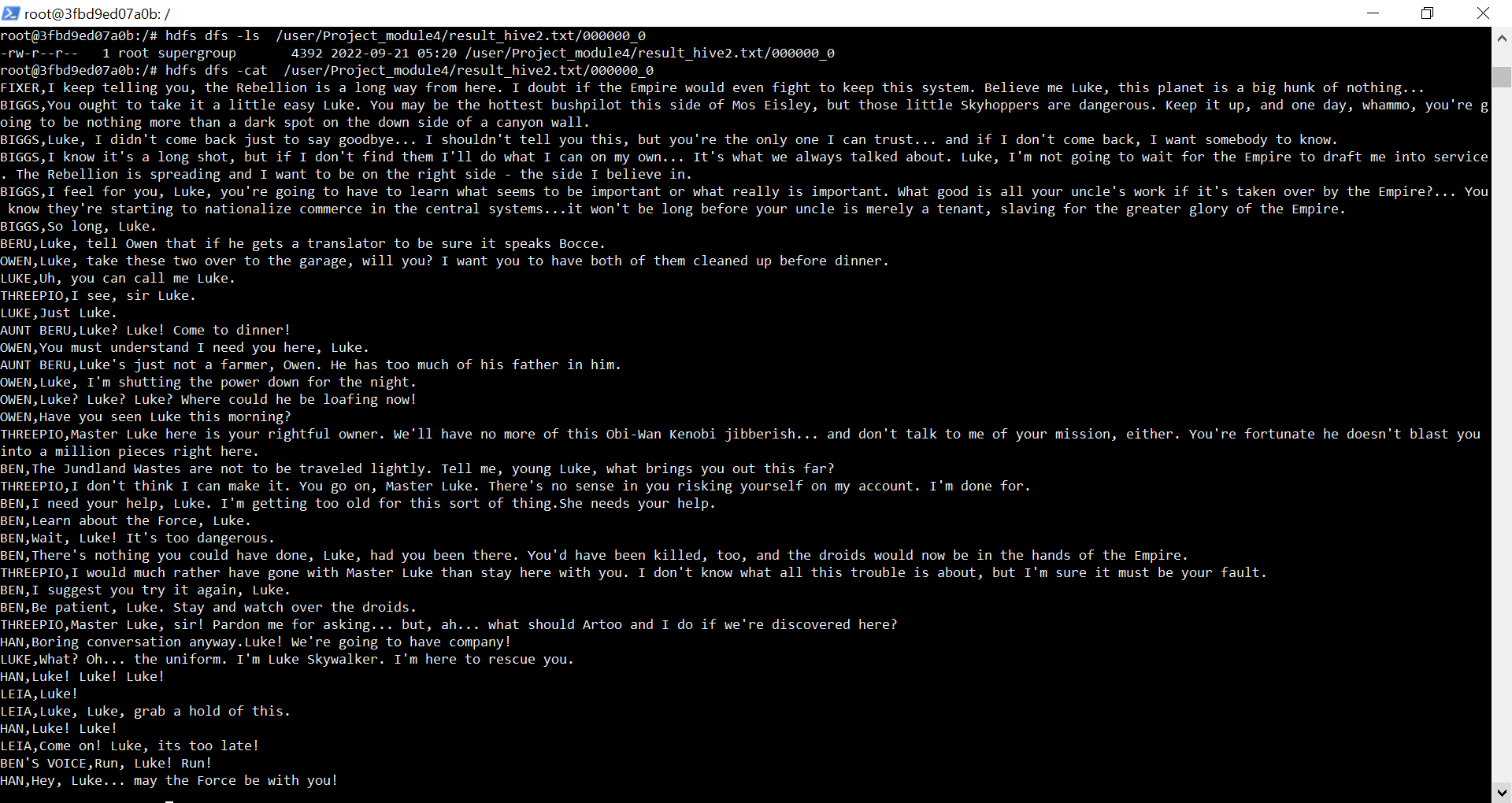


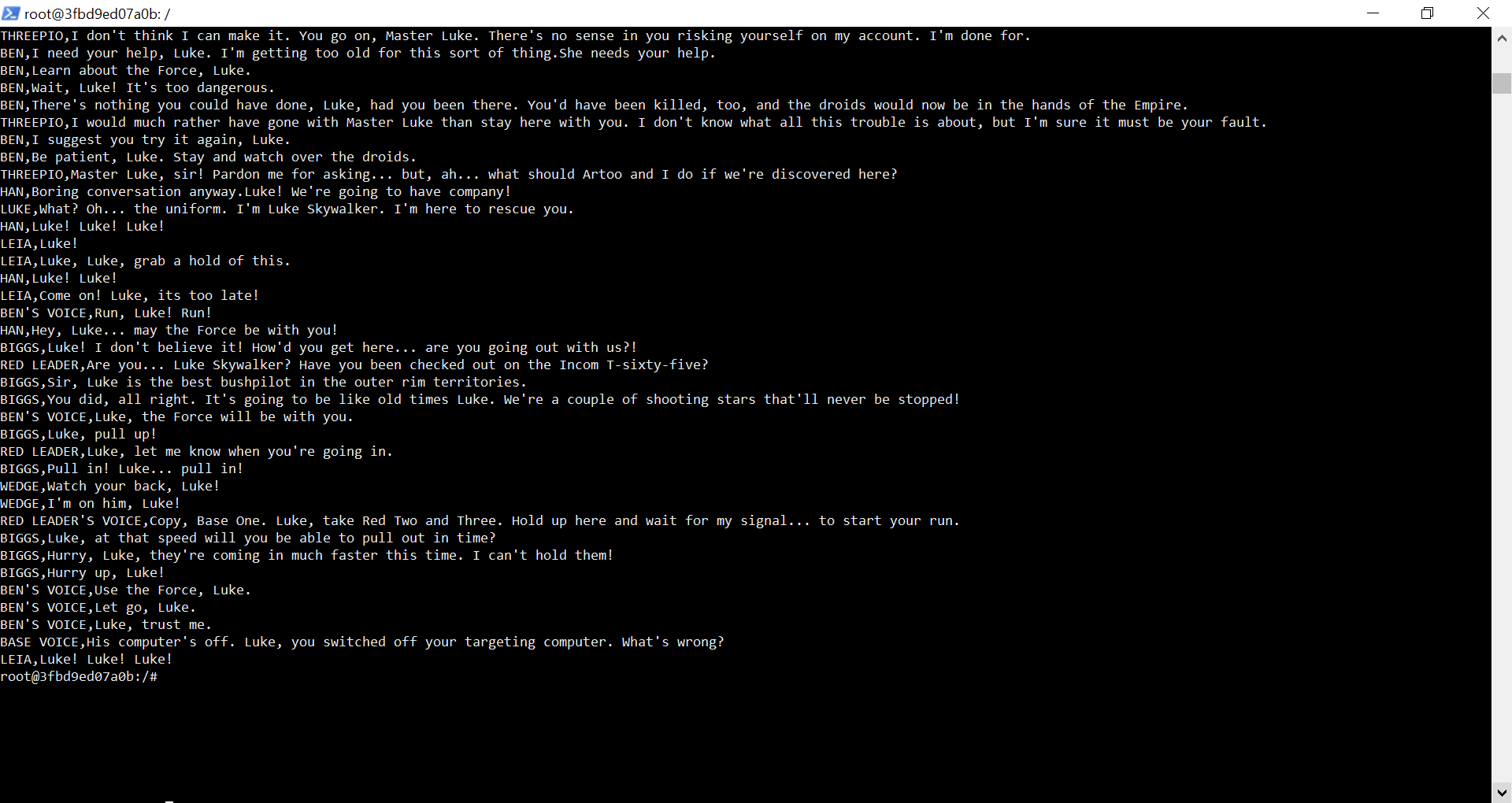


INSERT OVERWRITE DIRECTORY '/user/Project\_module4/result\_hive2.txt'

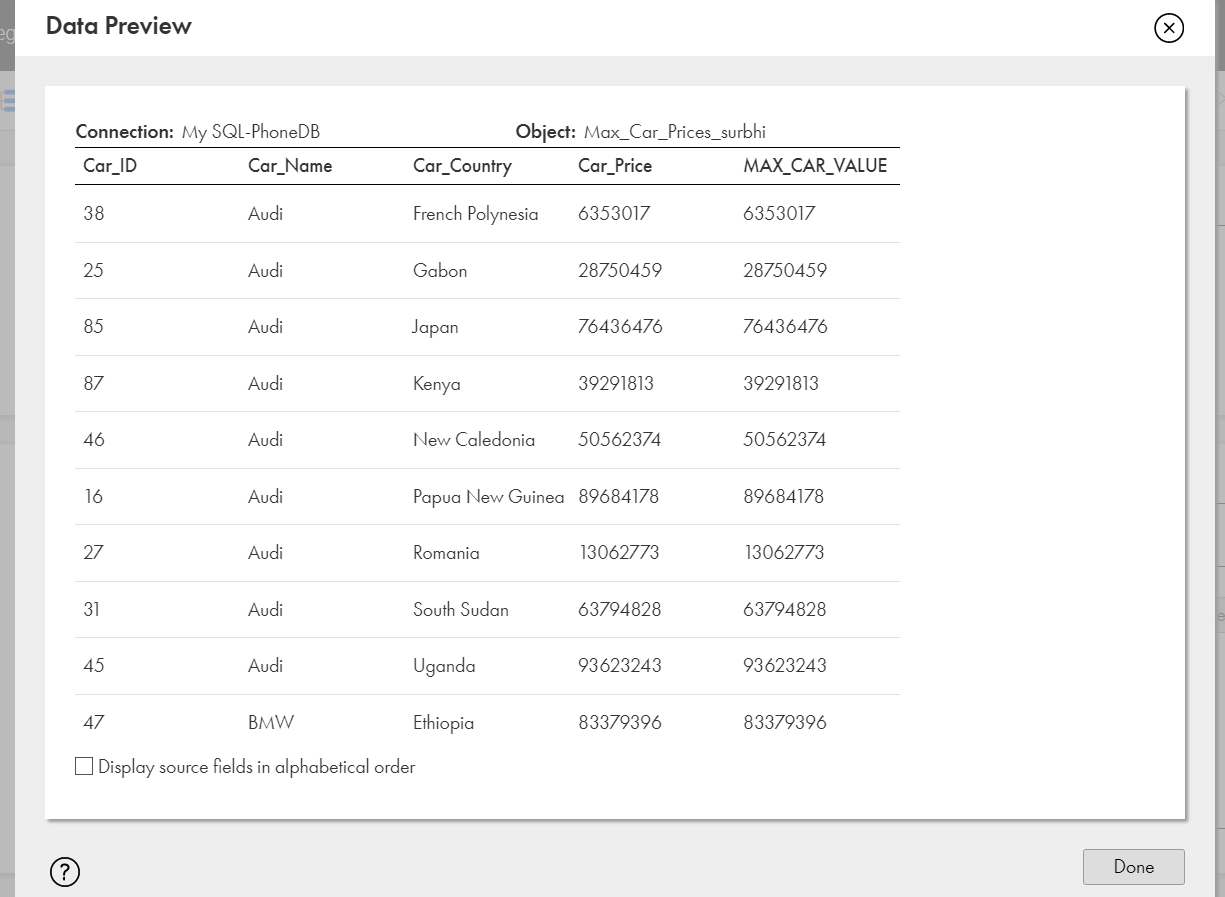
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

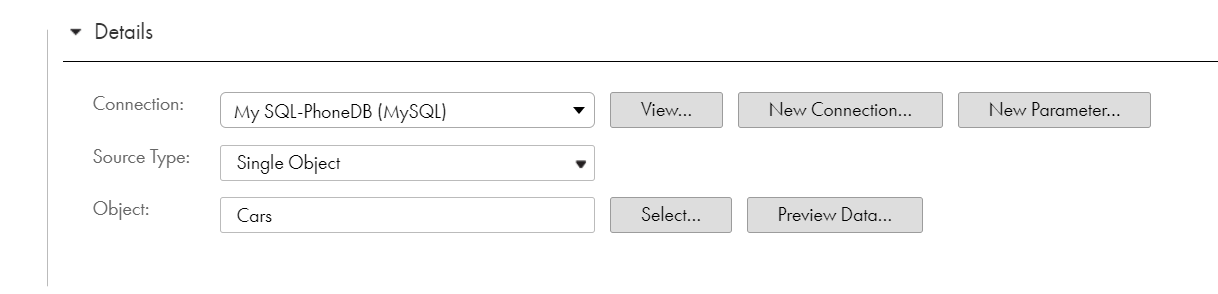
> select \* FROM file2 where instr(conv,’LUKE’)!=0

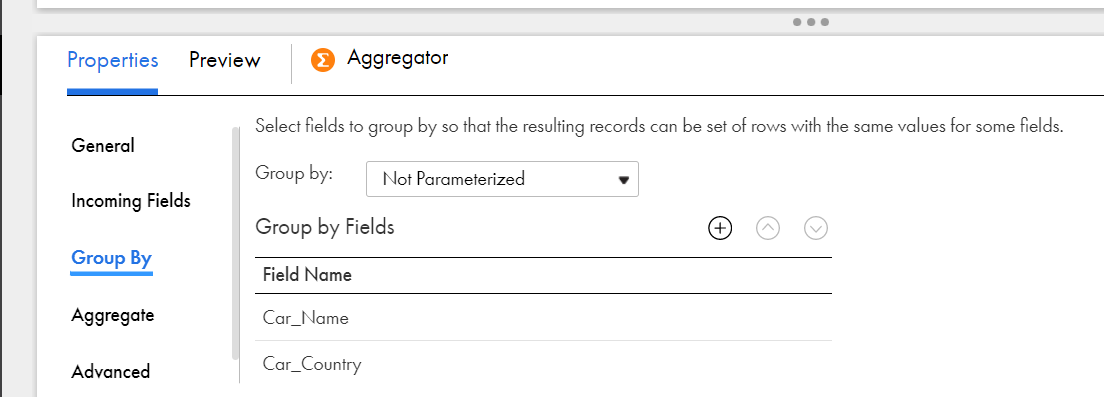


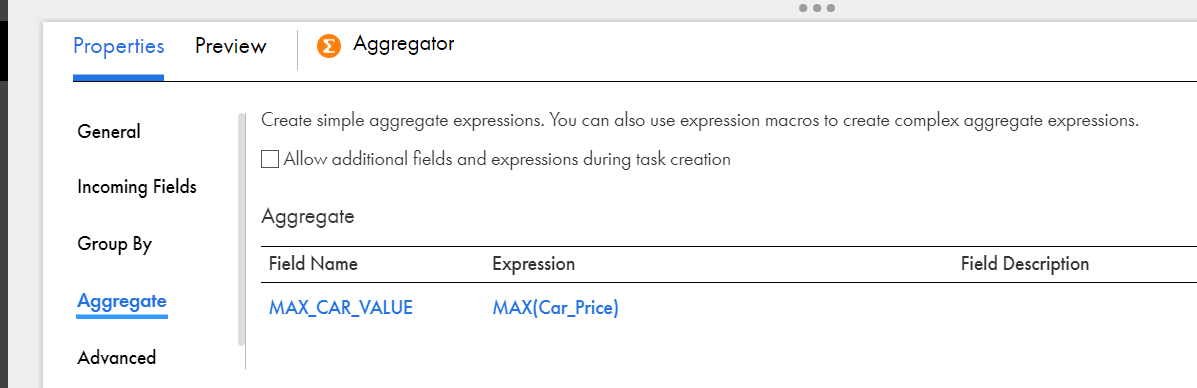


Informatica :-

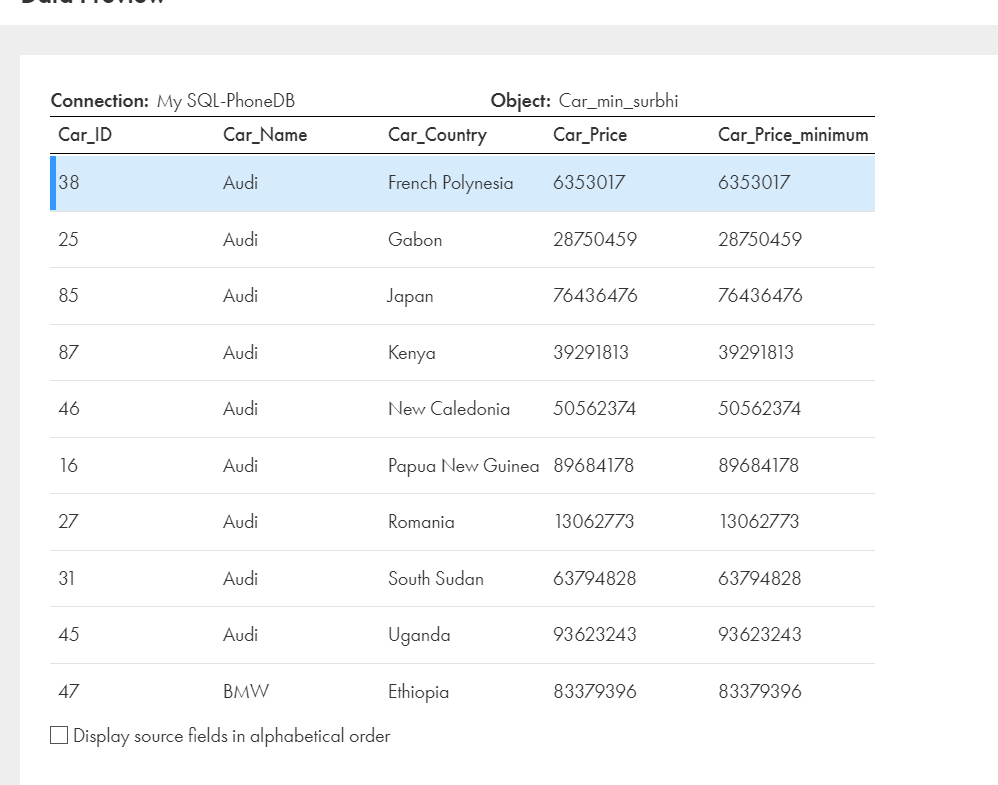








Informatica 2



<https://docs.google.com/spreadsheets/d/1GN_1udS4395TkhK0YAJZZej664GnQzarwyUI8H71Ci8/edit#gid=495919687>

https://github.com/training-support/FST-M4