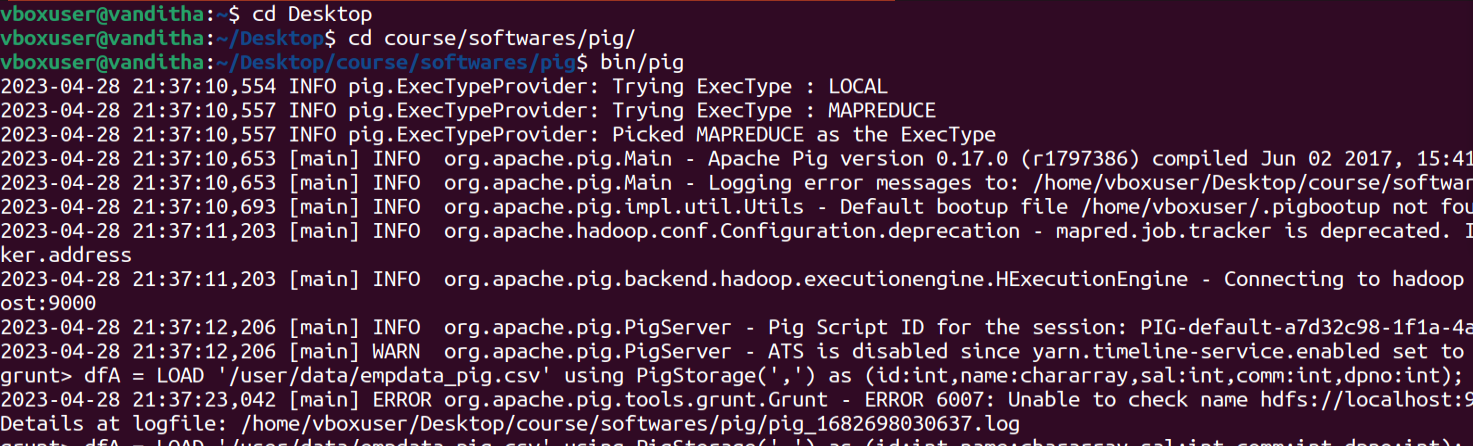
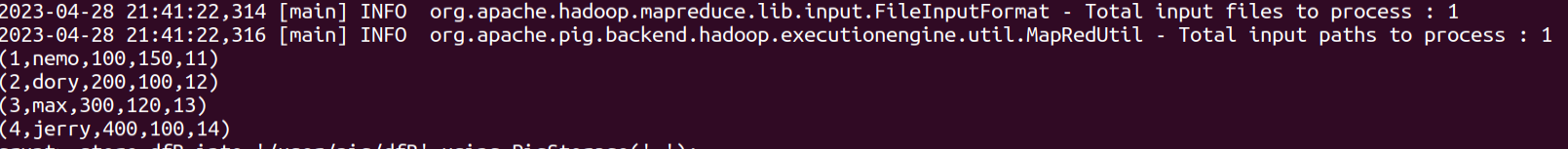
**PIG SCRIPTING CODES:**

Load Data: loading the csv file from HDFS to pig with a given schema

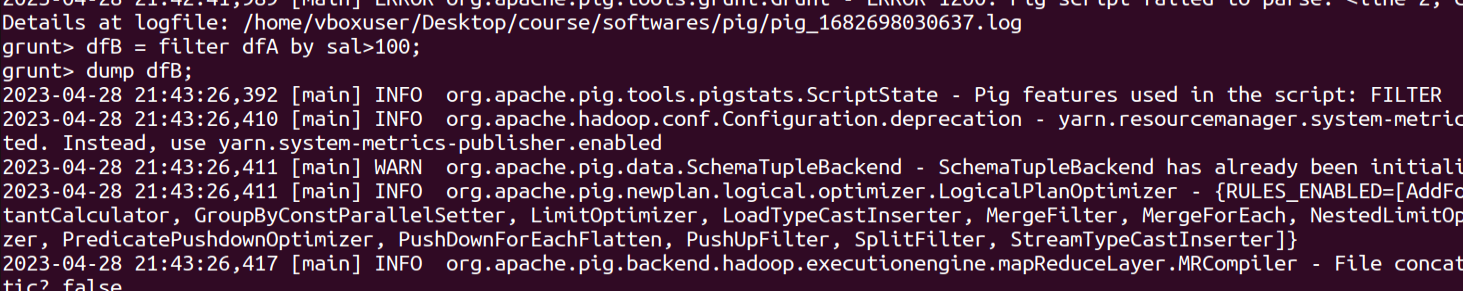
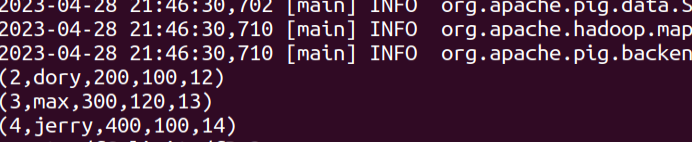
Dump the loaded data dfA. 



Aggregate (by row):

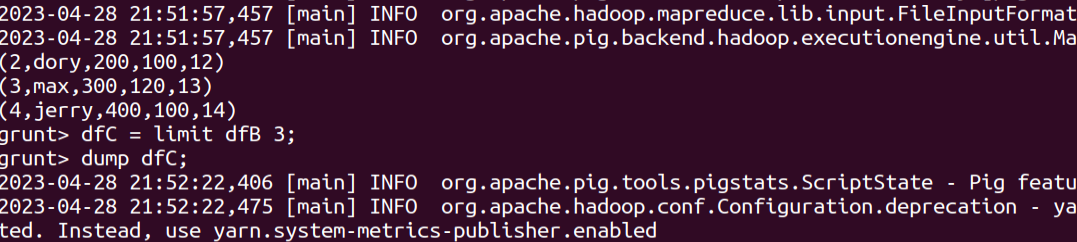
Filter the contents of dfA by condition salary>100.

Dump dfB;

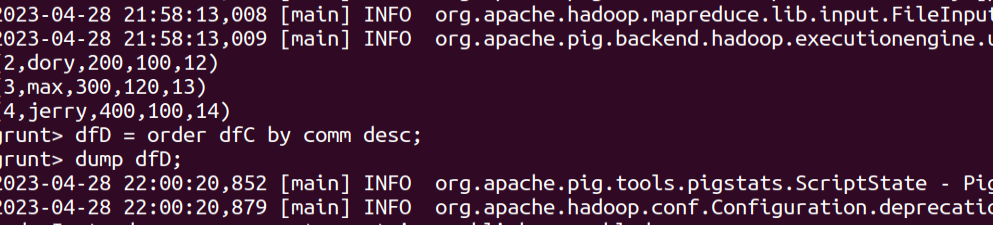
Get the first 3 rows of dfC using limit operation.

Dump dfC.

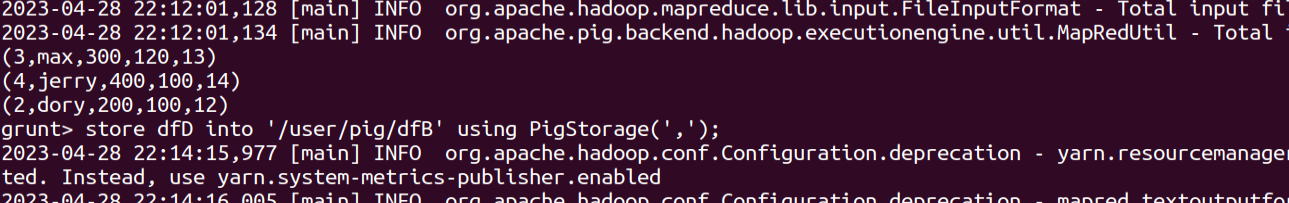


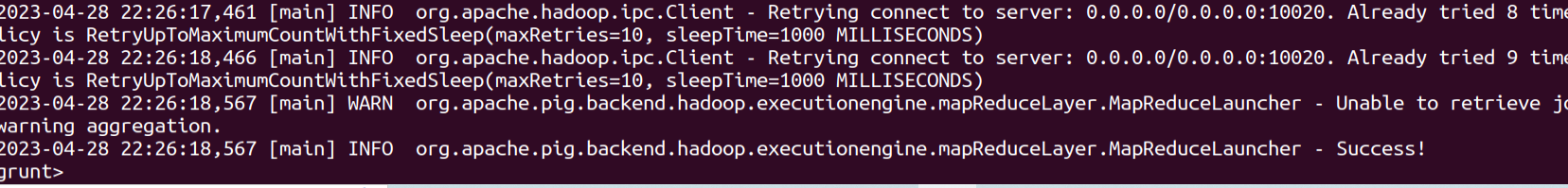
dfD contains all Sorted rows from dfC in descending order

Dump dfD



Store the dfD data



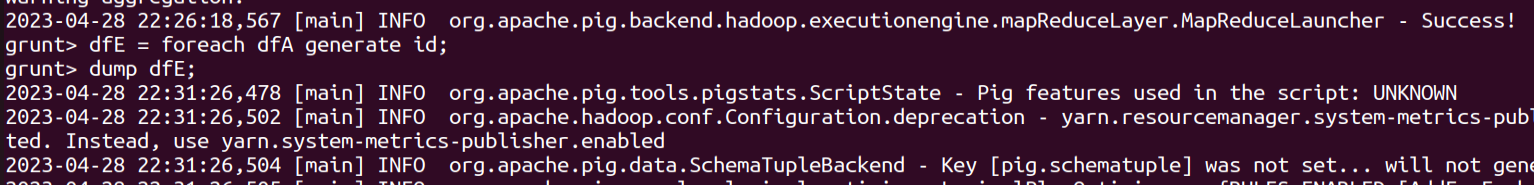


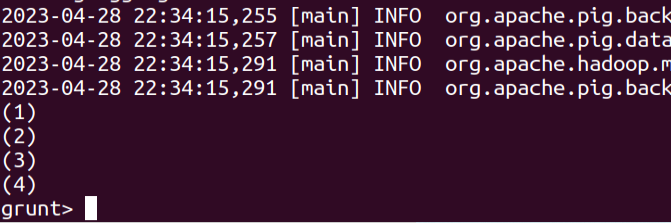
Transform (by column):

Select existing column

Select only the id column from dfA and load it into dfE

Dump dfE

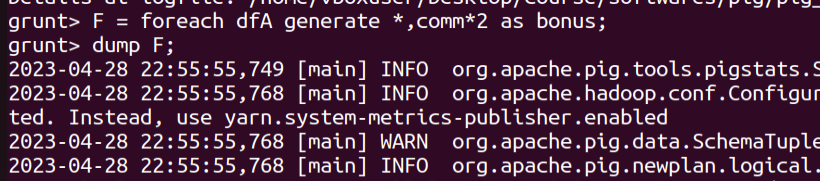


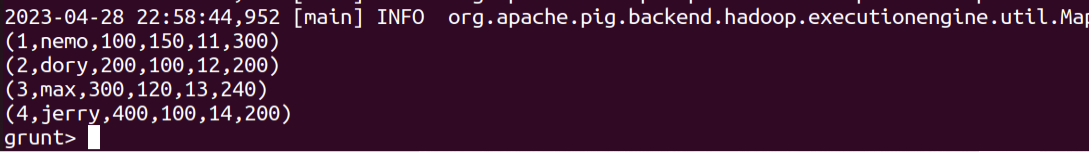


Create new column based on existing column

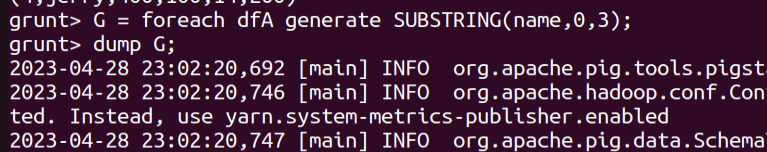
Create a table F where new columns are created based on some operation on existing table dfA

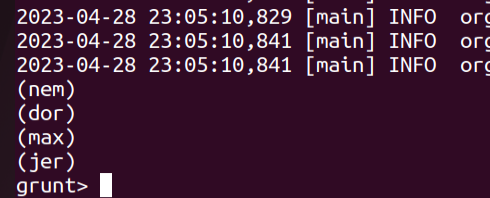
Dump F





Transform columns:





Store Data:

