Assignment 6.3

1. LOAD statement is used to loads data from the file system.

LOAD ‘data’ [Using function] AS schema.

1. DESCRIBE – Return the schema of the relation

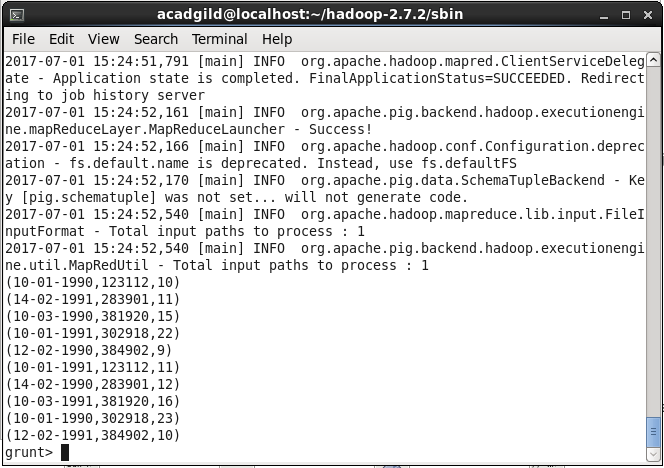
DESCRIBE relationname;

1. DUMP – used to display the result on the screen.

DUMP relationname;

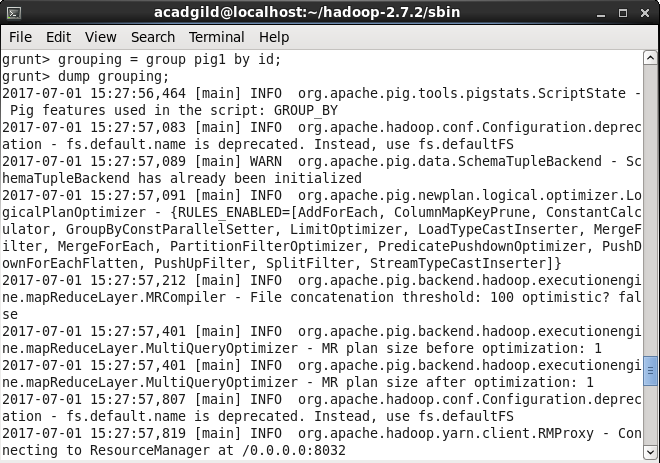


DUMP result is shown below from the loaded file “file1”.

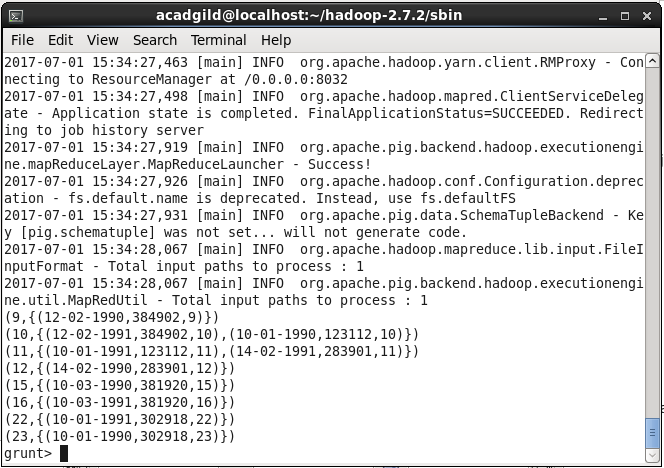


1. GROUP – groups the data in one or more relations.

GROUP relationname BY attributename.



Grouping result is shown below based on the id.



1. FOREACH – generates data transformation based on columns of data.

FOREACH groupname { generate….};



Foreach output displays the unique number of ids count.



1. STORE – used to store the result into HDFS or local file system.

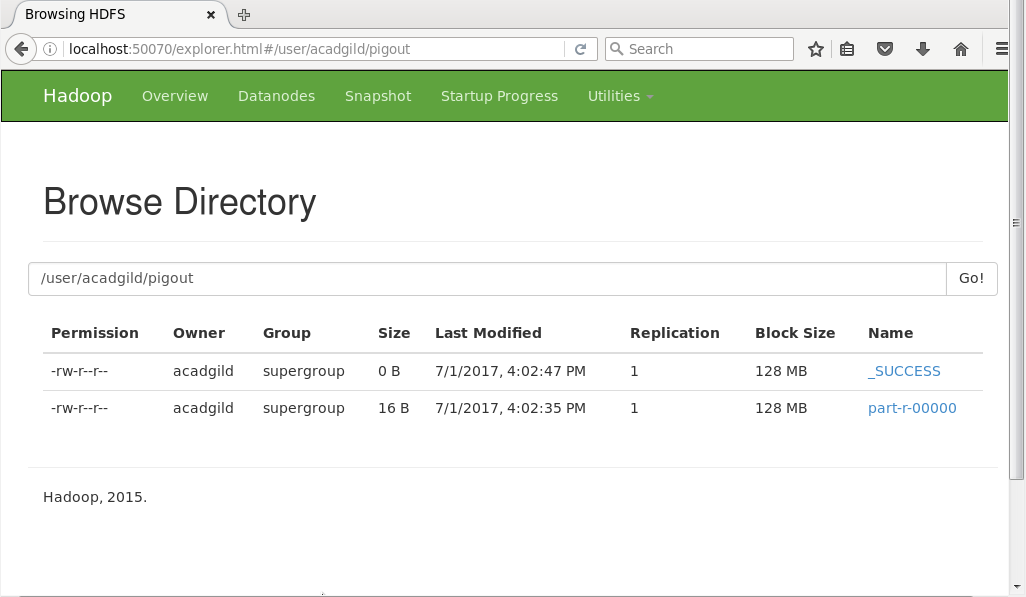
STORE relationname INTO ‘destination place’



Store command execution

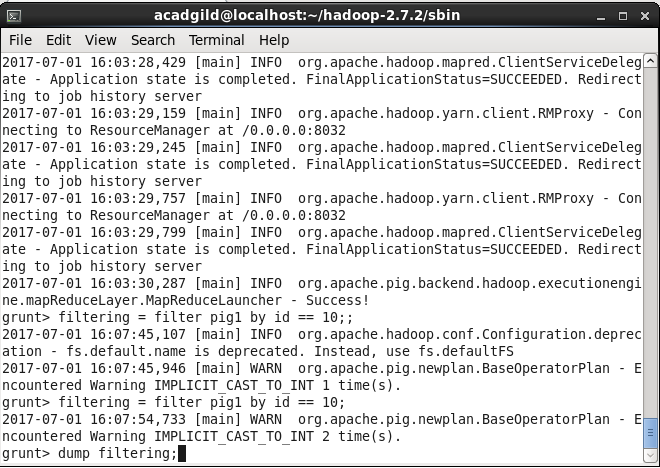


Unique count has been stored in the /user/acadgild/pigout

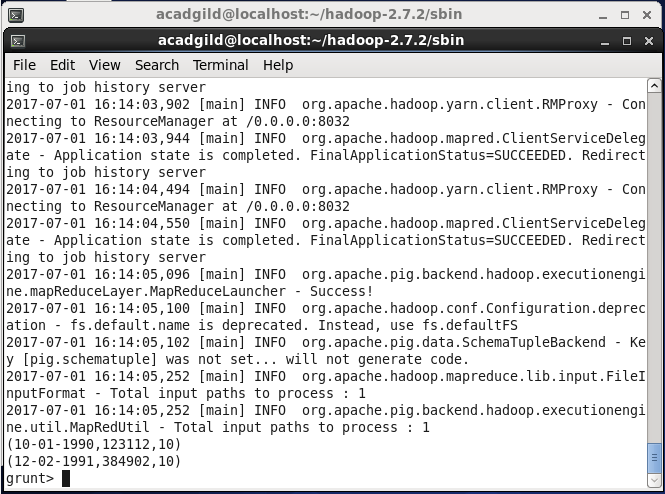


1. FILTER – Select tuples from a relation based on specified condition.

FILTER relationname BY condition.



Filtered output based on the id condition id == 10.



1. ORDER BY – ordering the data based on column condition.

ORDER relationname BY columnname [ASC/DESC]



Order by output based on the id by descending order.

