**Assignment 9.2**

**Question:** Explain the working of below pig script commands with an example.

* LOAD
* STORE
* DUMP
* FOREACH
* FILTER
* GROUP BY
* ORDER BY
* DESCRIBE

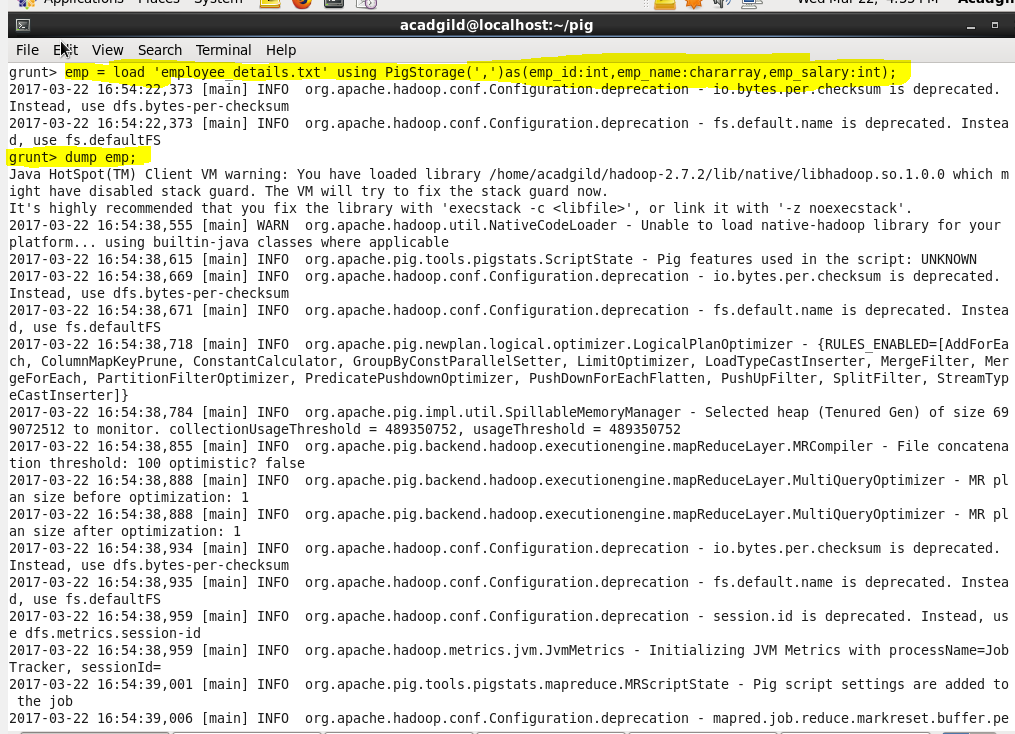
***LOAD:***

* We can load data into Pig from the file system (HDFS/ Local) using ***LOAD*** operator of Pig Latin.

**Syntax:**

***Relation\_name = LOAD ‘Input file Path’ USING function as schema;***

* The load statement consists of two parts divided by the “=” operator. On the left-hand side, we need to mention the name of the relation where we want to store the data, and on the right-hand side, we have to define how we store the data. Given below is the syntax of the Load operator.

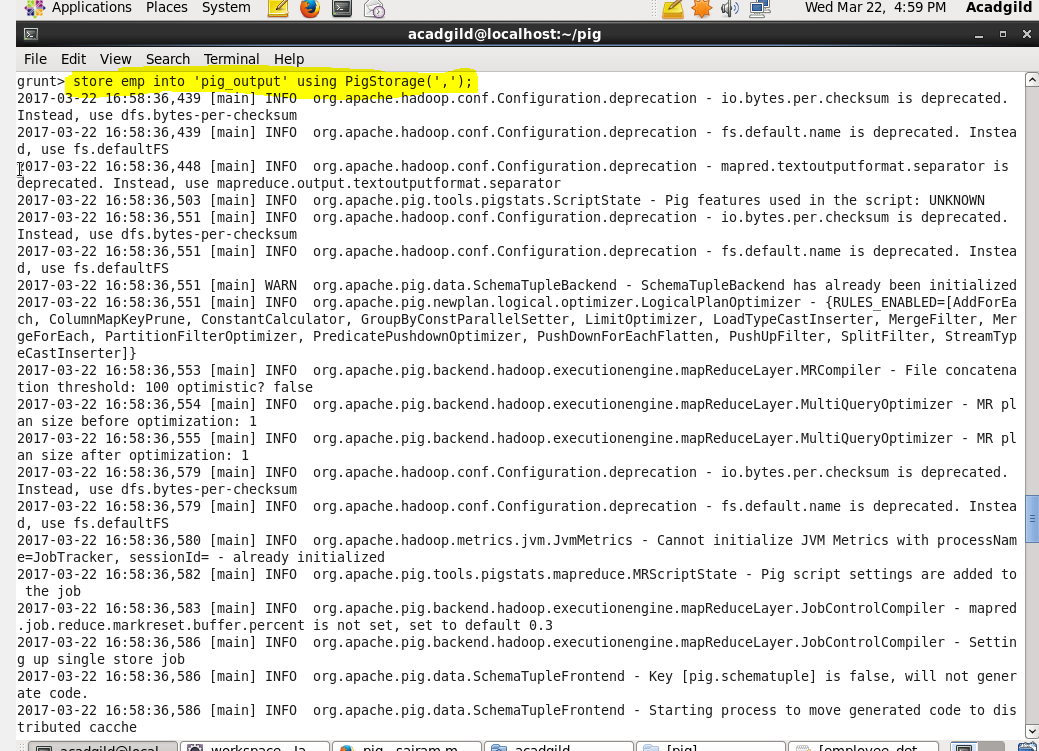


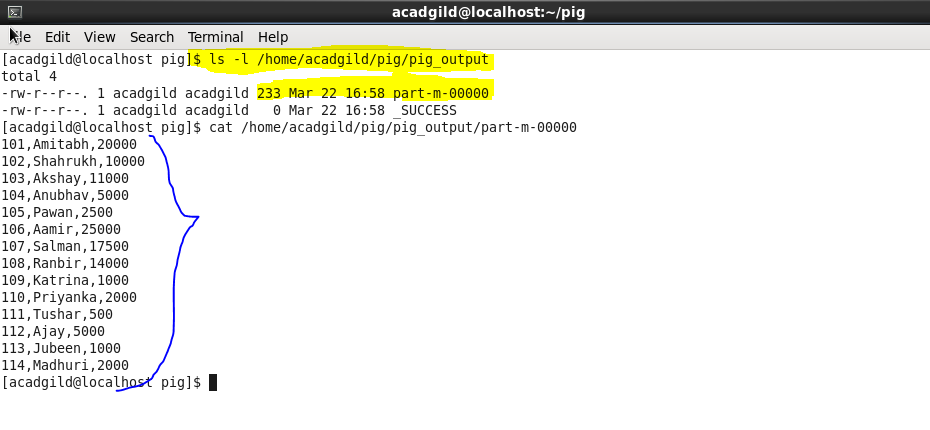
***STORE:***

* We can store the loaded data in the file system using the ***store*** operator.

**Syntax:**

***STORE Relation\_name INTO ‘required\_directory\_path’ [USING Function];***

****

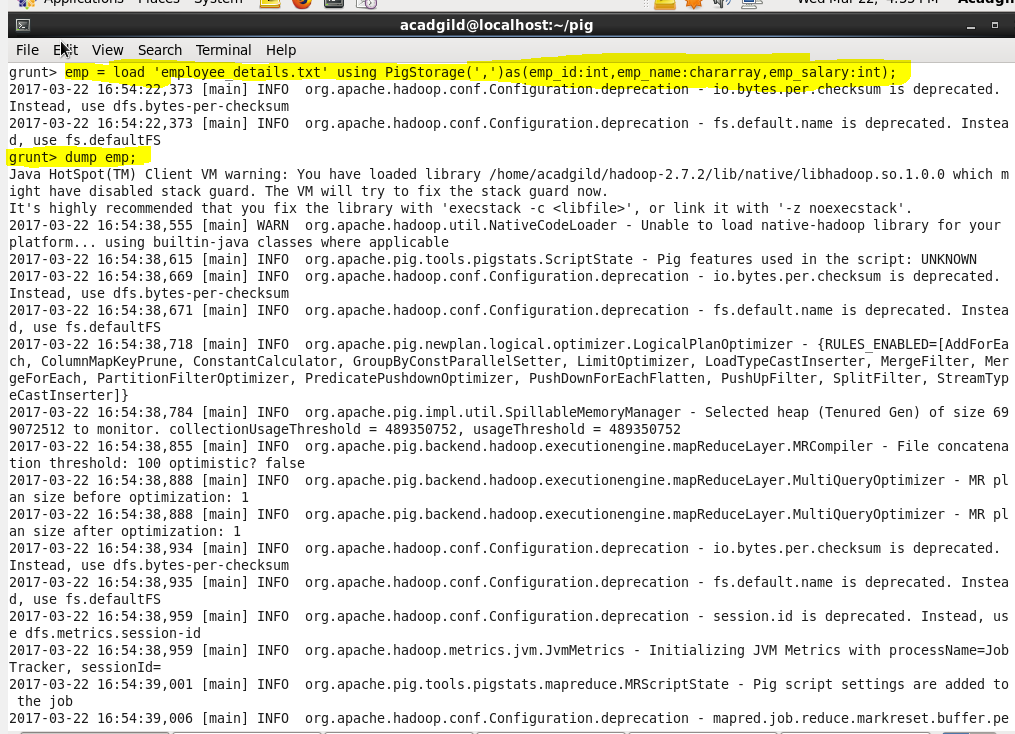


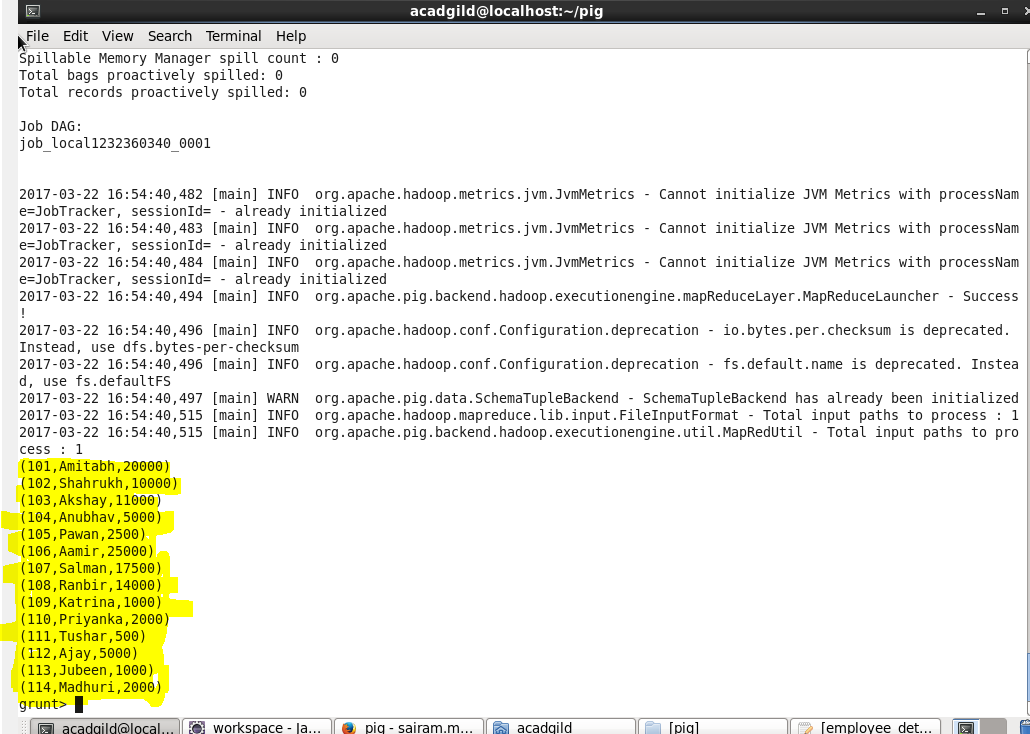
***DUMP:***

* The *Dump* operator is used to run the Pig Latin statements and display the results on the screen.

**Syntax:**

***DUMP Relation\_name;***

******

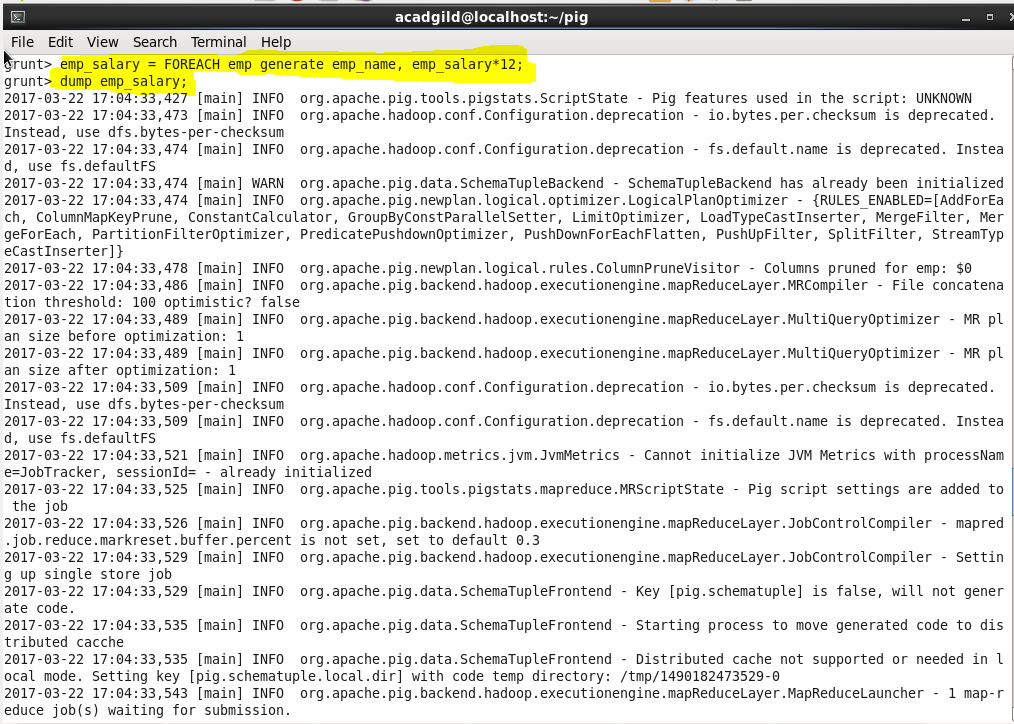
******

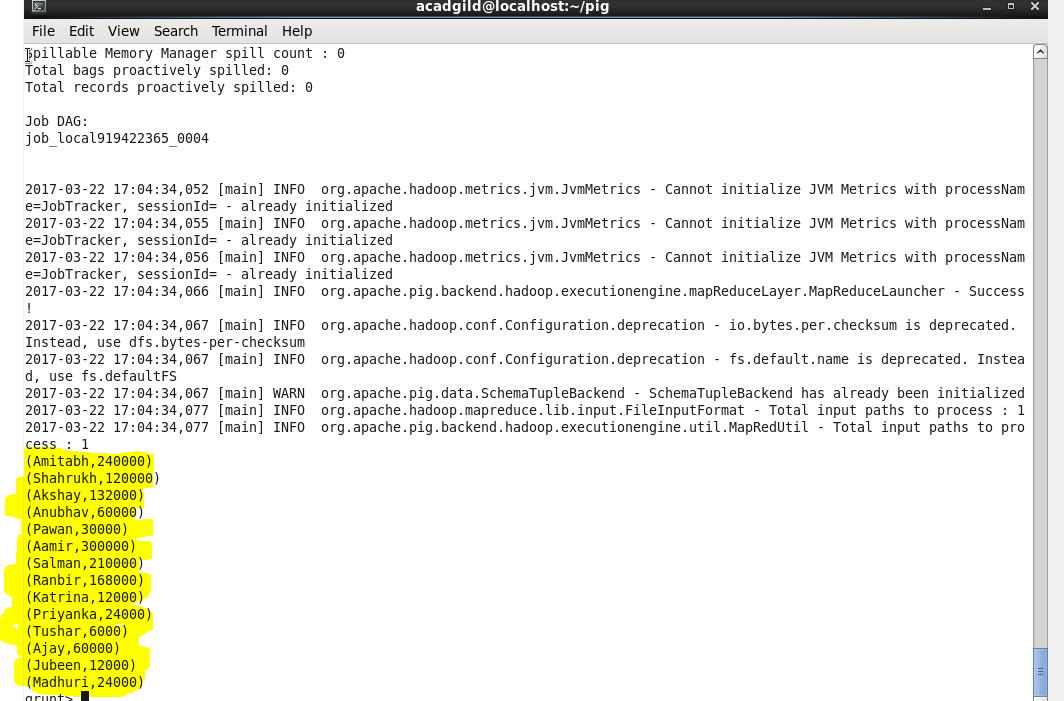
***FOREACH:***

The *FOREACH* operator is used to generate specified data transformations based on the column data.

**Syntax:**

***Relation\_name2 = FOREACH Relation\_name1 GENERATE (required data);***

******

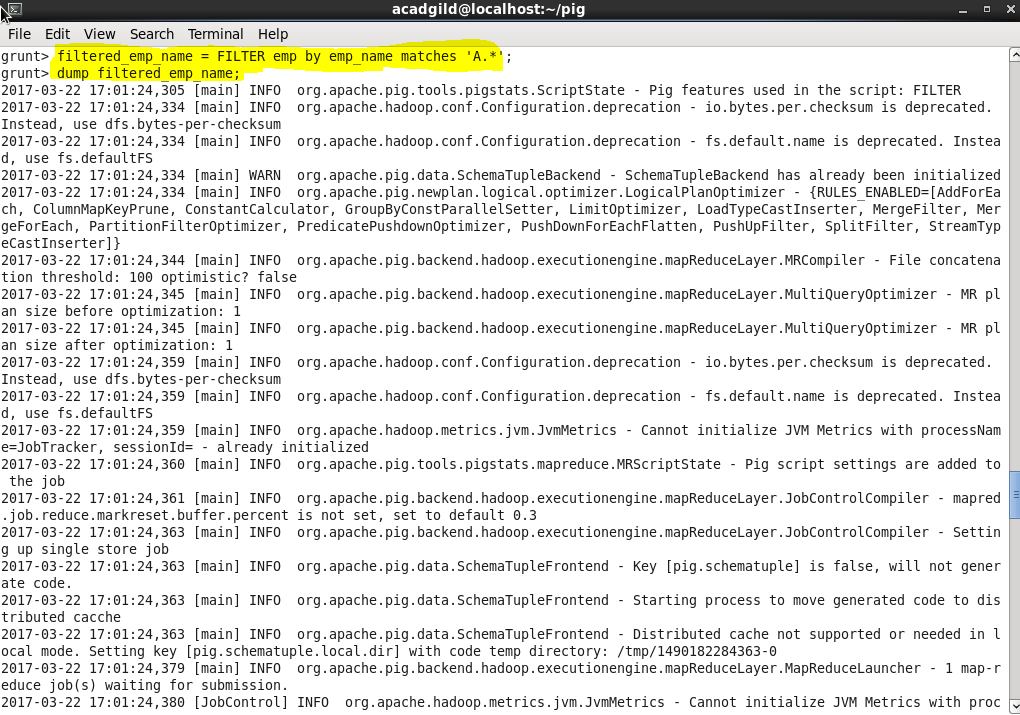
******

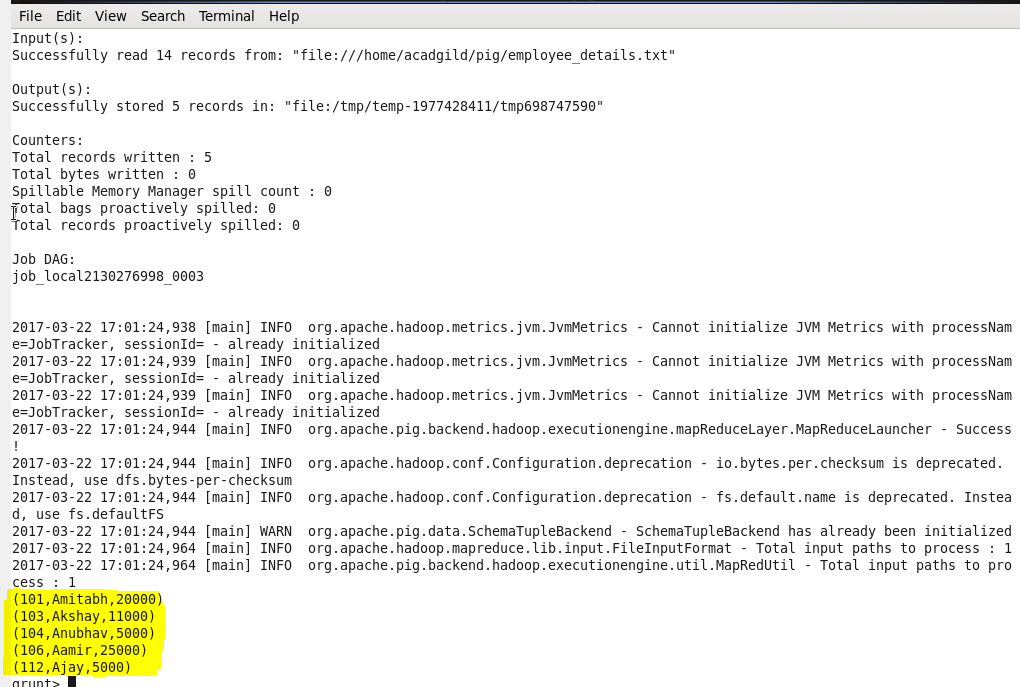
***FILTER:***

The *FILTER* operator is used to select the required tuples from a relation based on a condition.

**Syntax:**

***Relation\_name2 = FILTER Relation\_name1 BY (condition);***

******

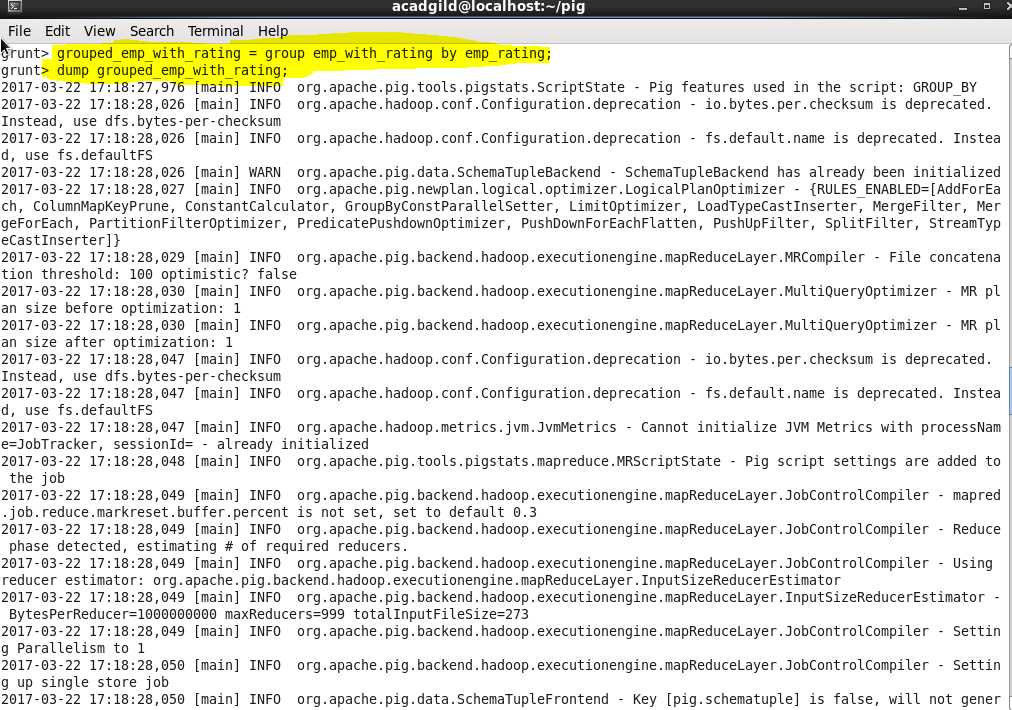
******

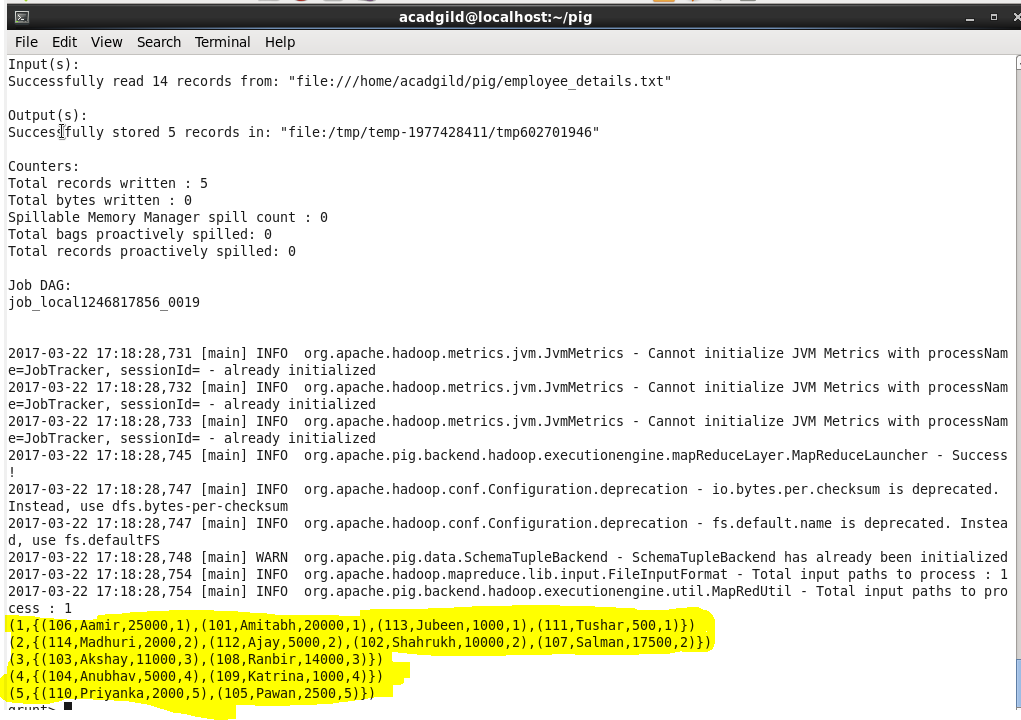
***GROUP BY:***

The *GROUP* operator is used to group the data in one or more relations. It collects the data having the same key.

**Syntax:**

***Group\_data = GROUP Relation\_name by column\_name;***

******

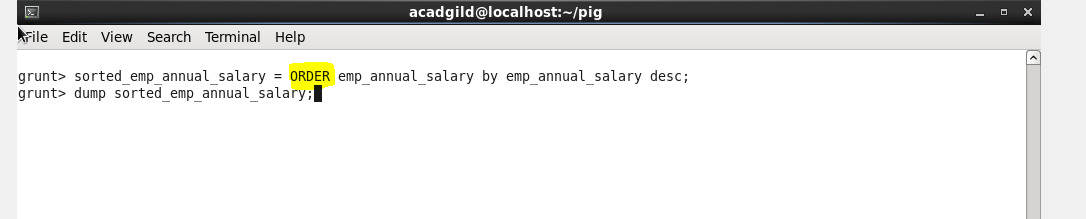
******

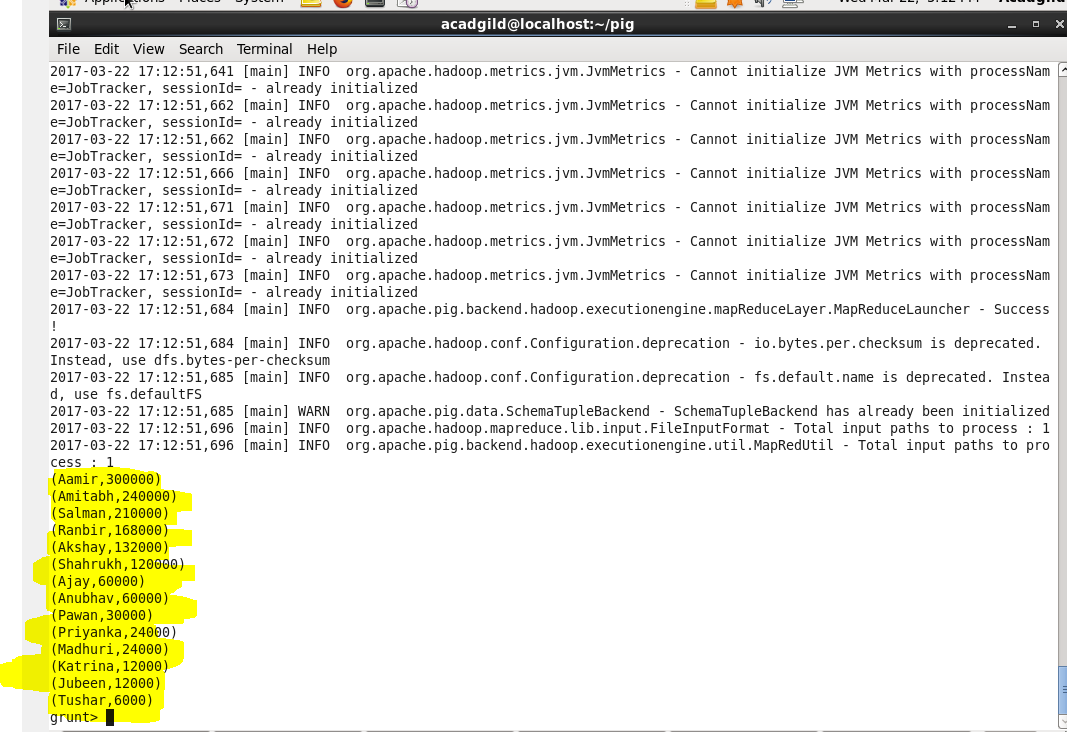
***ORDER BY:***

The *ORDER BY* operator is used to display the contents of a relation in a sorted order based on one or more fields.

**Syntax:**

***Relation\_name2 = ORDER Relation\_name1 BY (ASC/DESC);***

****

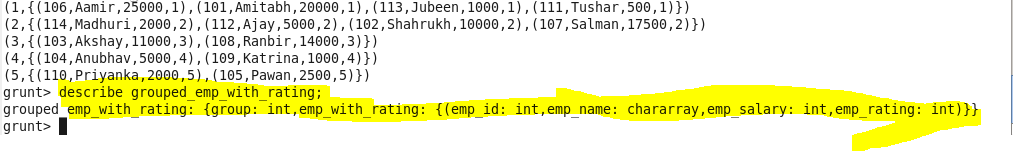
****

***DESCRIBE:***

The *DESCRIBE* operator is used to view the schema of a relation.

**Syntax:**

***DESCRIBE Relation\_name;***

******