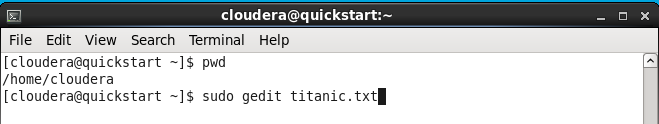
**Project : Titanic Data Analysis**

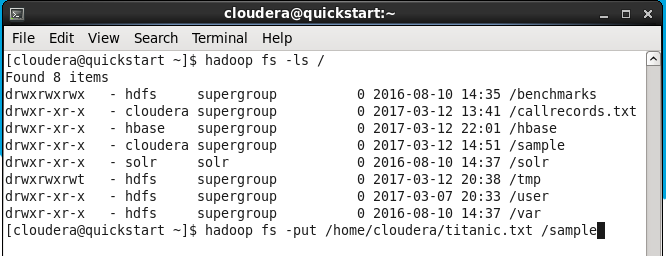
Q1. In this problem statement we will find the average fare of each class.

**Create a text file**

**-Copy Data Sets in titanic.txt**



-**Copy titanic.txt into HDFS**



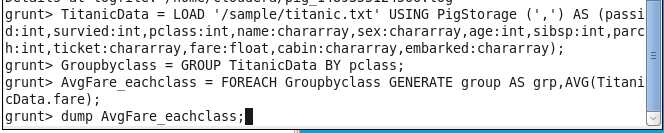
**-Start grunt shell with pig command**

-Load titanic.txt

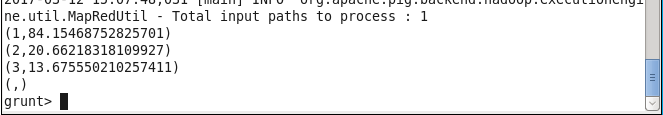
-second group by pclass

-third generate data to find the average fare of each class

-to check output use dump command



**-output**

-its showing first column is pclass and second column is avg fare of each class

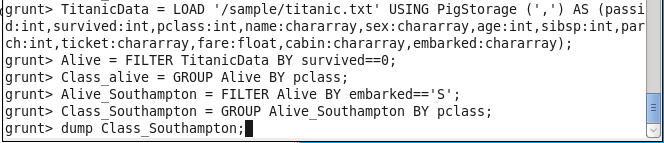
**2.In this problem statement we will find the number of people alive in each class and are embarked in Southampton**.

**-**filter data to see no.of people alive and with the help of group to see no.of people alive in each class

**-**second check no. of people alive in each and embarked in Southampton

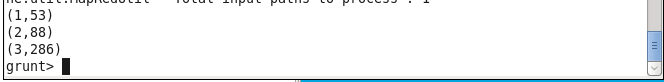
-with the help of group to see no. of people alive in Southampton in each class

-to check output use dump command



**-output**

-its showing first column is class and second column is no. of people alive from Southampton



**3.In this problem statement we will find out number of male and female people died in each class.**

**Check Femal who died in each class**

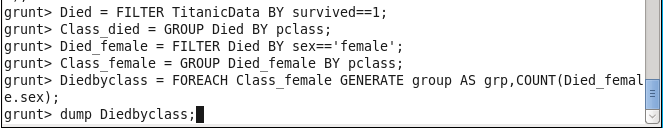
-filter data who died

- group by class

-filter data by females and group by class

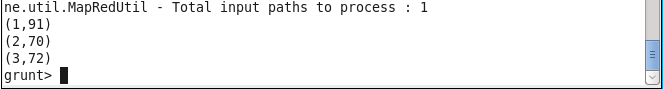
-generate data of females who died and count from each class

-to check use dump



**-output**

-its showing first column is class and second column is no. of females died

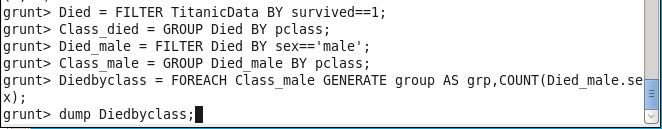


**Check males who died in each class**

-filter data by males and group by class

-generate data of males who died and count from each class

-to check use dump



-**output**

-its showing first column is class and second column is no. of males died

