//first import the file from local directory to hdfs directory

hdfs dfs -put titanic.txt internship/titanic.txt

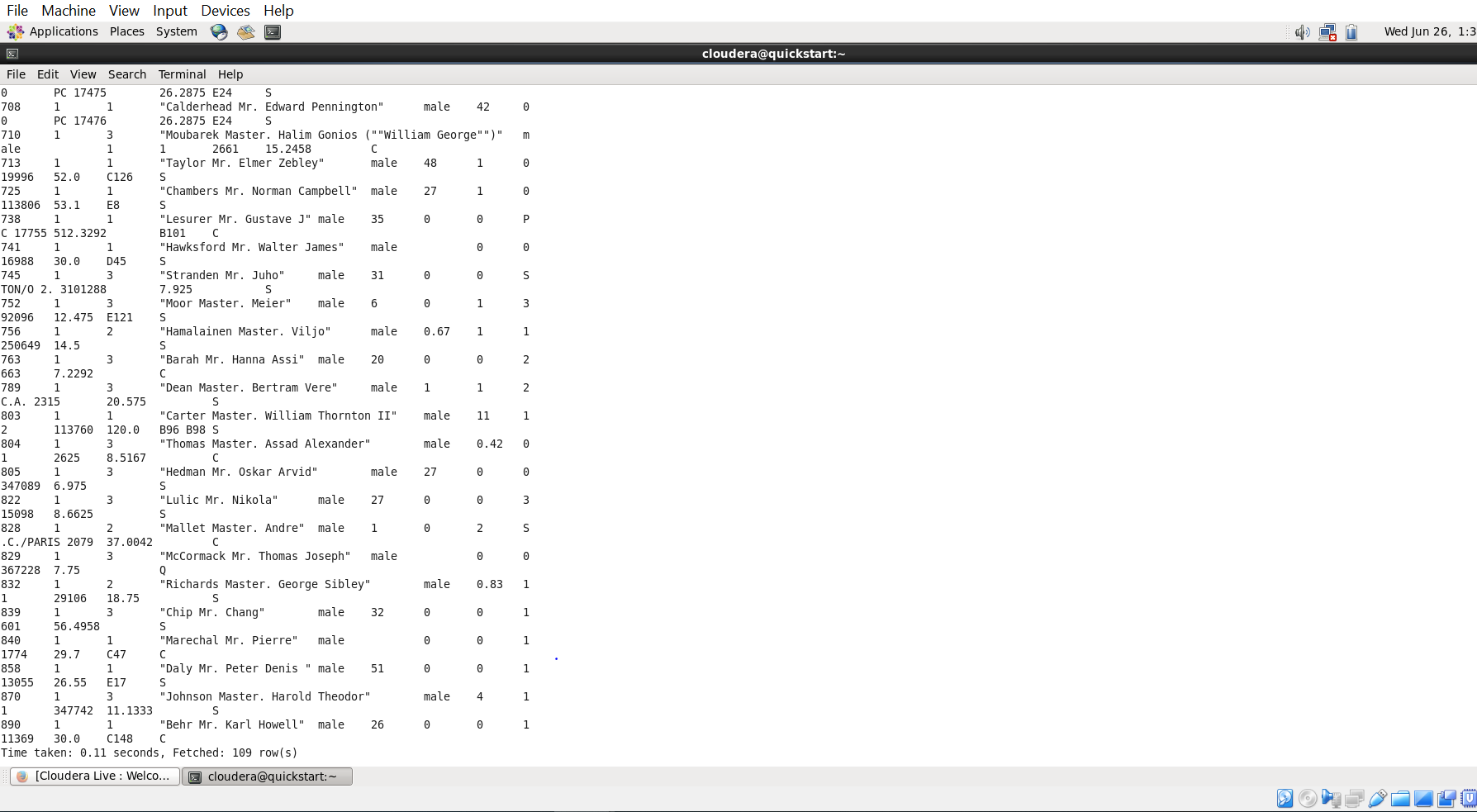
//now create table in hive

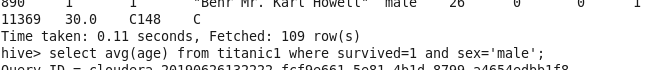
create external table titanic1

(userid int,survived int class int,name string,sex string,age string,sibsp int,parch int,tkt string,fare float,cabin string,embarked chararray) row format delimited fields terminated by ",";

//load data into table

load data inpath "internship/titanic.txt" into table titanic1;



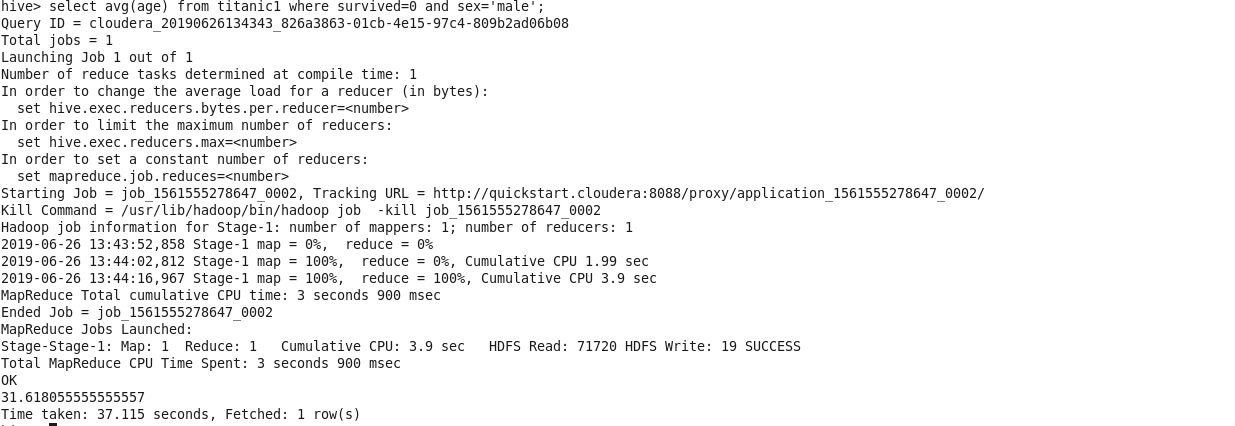


QUESTIONS:

1. AVG AGE OF MALE WHO DIED IN THE INCIDENT?

Query: select avg(age) from titanic1 where survived=0 and sex=”male”;

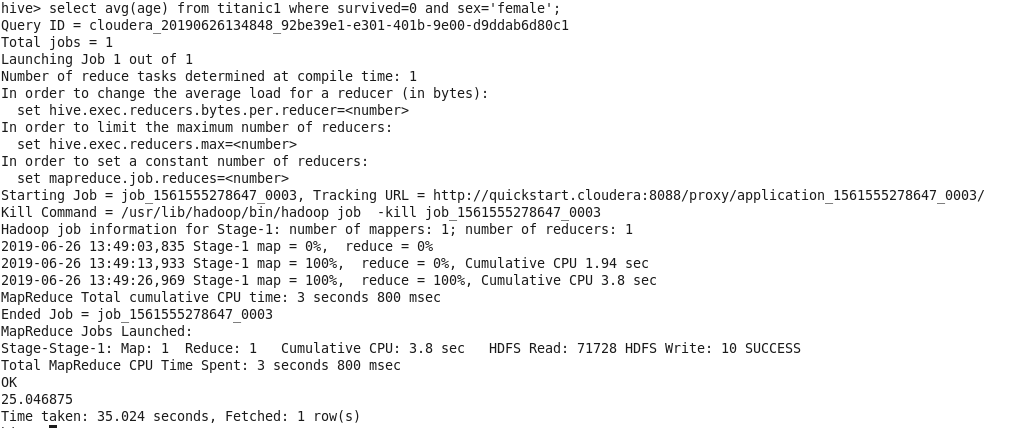
Ans: 31.618



1. AVG AGE OF FEMALE WHO DIED IN THE INCIDENT?

Query: select avg(age) from titanic1 where survived=0 and sex=”female”;

Ans:25.046



1. Number of people survived as per travelling class

Query : select count(userid) rom titanic1 where survived=1 group by class;

Ans:

Class1: 136

Class2 :87

Class3: 119

