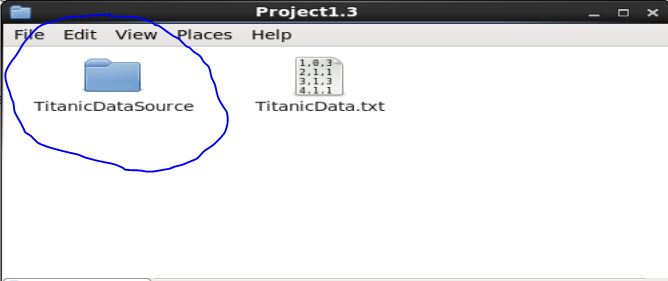
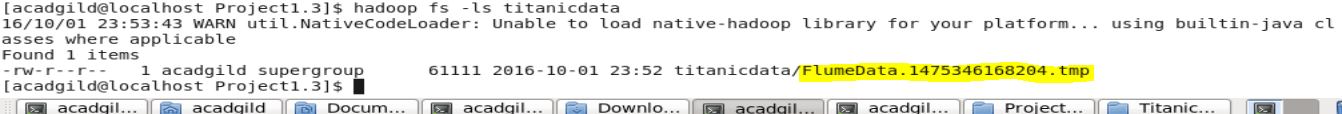
Project 2

Creating directory "titanicdata” to store the data in HDFS which is going to act as a directory while loading data using flume.

Created "TitanicDataSource" folder on local drive which will act as a source for transferring data using flume.





**Problem 1:**

Created the below Pig script and ran using following command in MapReduce mode:

**C:\Users\RAKESH\Desktop\Screenshot\Project1.3\S13.JPG**

Script:

rawdata = LOAD '/user/acadgild/titanicdata' USING PigStorage(',') AS (passengerid: int,survived: int,pclass: int,name: chararray,sex: chararray,age: int,sibsp: int,parch: int,ticket: chararray,fare: float,cabin: chararray,embarked: chararray);

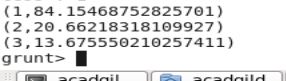
generate\_class = FOREACH rawdata GENERATE pclass,fare;

group\_class = GROUP generate\_class by pclass;

avg\_fare = FOREACH group\_class GENERATE group as pclass, AVG(generate\_class.fare);

dump avg\_fare;

Output:

****

**Problem 2:**

Created the below Pig script and ran using following command in MapReduce mode:

C:\Users\RAKESH\Desktop\Screenshot\Project1.3\S14.JPG

Script:

rawdata = LOAD '/user/acadgild/titanicdata' USING PigStorage(',') AS (passengerid: int,survived: int,pclass: int,name: chararray,sex: chararray,age: int,sibsp: int,parch: int,ticket: chararray,fare: float,cabin: chararray,embarked: chararray);

generate\_filter = FILTER rawdata by survived == 0 AND embarked == 'S';

generate\_pclass = FOREACH generate\_filter GENERATE pclass;

group\_pclasss = GROUP generate\_pclass by pclass;

count\_people = FOREACH group\_pclasss GENERATE group as pclass, COUNT(generate\_pclass.pclass);

dump count\_people;

Output:



**Problem 3:**

Created the below Pig script and ran using following command in MapReduce mode:

C:\Users\RAKESH\Desktop\Screenshot\Project1.3\S15.JPG

Script:

rawdata = LOAD '/user/acadgild/titanicdata' USING PigStorage(',') AS (passengerid: int,survived: int,pclass: int,name: chararray,sex: chararray,age: int,sibsp: int,parch: int,ticket: chararray,fare: float,cabin: chararray,embarked: chararray);

generate\_died = FILTER rawdata by survived == 1;

generate\_class = FOREACH generate\_died GENERATE pclass,sex;

filter\_male = FILTER generate\_class by sex == 'male';

group\_male = GROUP filter\_male by pclass;

count\_male = FOREACH group\_male GENERATE group as pclass,COUNT(filter\_male.sex) AS m;

filter\_female = FILTER generate\_class by sex == 'female';

group\_female = GROUP filter\_female by pclass;

count\_female = FOREACH group\_female GENERATE group as pclass,COUNT(filter\_female.sex) AS f;

join\_count = JOIN count\_male by pclass, count\_female by pclass;

final\_count = FOREACH join\_count GENERATE count\_male::pclass,m,f;

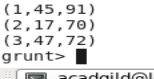
dump final\_count;

Output:

(1,45,91)

(2,17,70)

(3,47,72)

****