**The data set contains information about passengers who boarded Titanic ship. It contains data**

**points like:**

** Passenger’s age**

** Their native place**

** Details of who survived**

** Fare details of various travel classes**

** Number of casualties from various classes etc.**

**https://drive.google.com/file/d/0ByJLBTmJojjzNmV0dk1EMmwwQ1U/view?usp=sharing**

**DATA SET DESCRIPTION**

**Column 1 : PassengerId**

**Column 2 : Survived (survived=0 & died=1)**

**Column 3 : Pclass**

**Column 4 : Name**

**Column 5 : Sex**

**Column 6 : Age**

**Column 7 : SibSp**

**Column 8 : Parch**

**Column 9 : Ticket**

**Column 10 : Fare**

**Column 11 : Cabin**

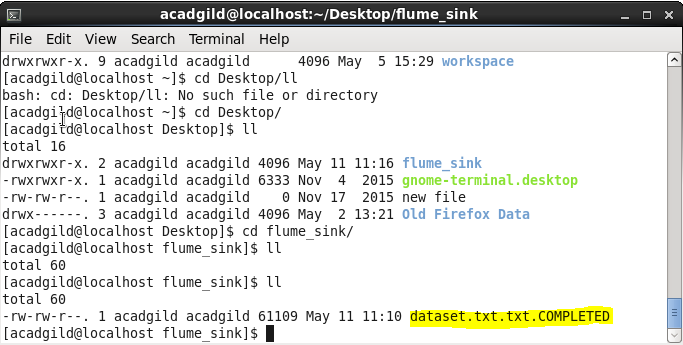
**Column 12 : Embarked**

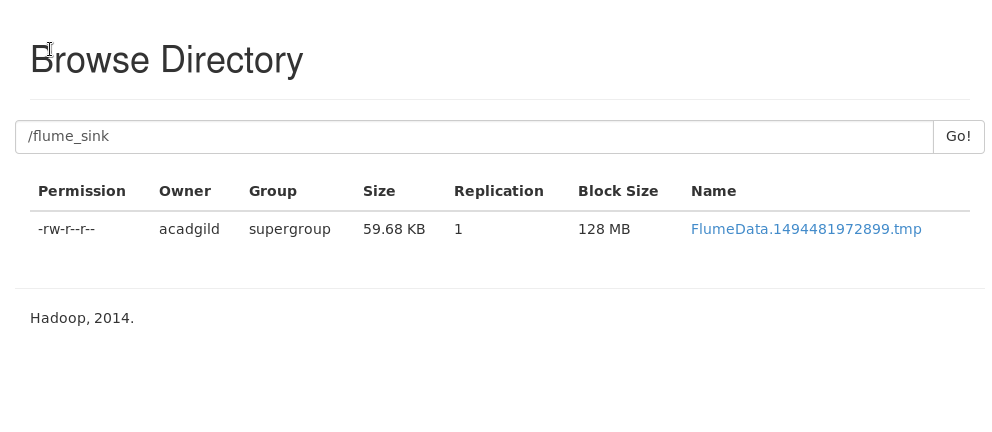
**You can use any of the technologies like Map Reduce, Pig or Hive of your choice.**

**Note:You need to copy the data set into HDFS using flume and send the screen shot of that with the**

**project solution**

***SOLUTION:***





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

CREATE TABLE data(

PassengerId varchar(50),

Survived varchar(50),

Pclass varchar(50),

Name varchar(50),

Sex varchar(50),

Age varchar(50) ,

SibSp varchar(50) ,

Parch varchar(50),

Ticket varchar(50),

Fare varchar(50),

Cabin varchar(50),

Embarked varchar(50))

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

hive> load data inpath '/flume\_sink/FlumeData.1494481972899' overwrite into table data;

hive> select Pclass,AVG(Fare) from data group by Pclass;

**OUTPUT:**

1 84.15468749999992

2 20.66218315217391

3 13.675550101832997

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

**are embarked in Southampton.**

hive> create view viewa as

select \* from data where Embarked=="S" AND Survived==0;

hive> select Pclass,COUNT(\*) from viewa group by Pclass;

**OUTPUT:**

1 53

2 88

3 286

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

**each class.**

hive> select Pclass,COUNT(\*) from data

where Survived==1 AND Sex="male"

group by Pclass;

**OUTPUT:**

1 45

2 17

3 47

hive>select Pclass,COUNT(\*) from data

where Survived==1 AND Sex="female"

group by Pclass;

**OUTPUT:**

1 91

2 70

3 72