**Titanic data analysis**

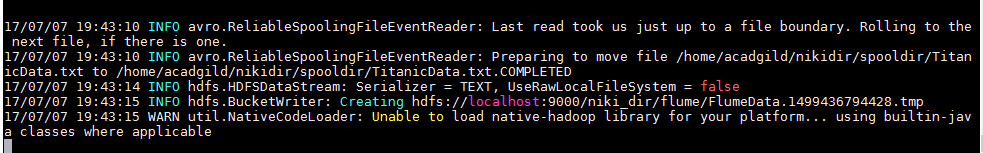
* **Flume**

Command to run flume:

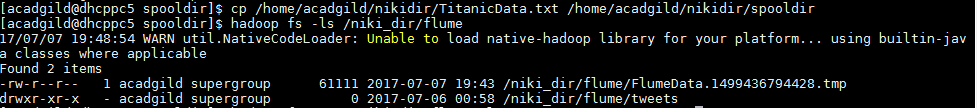
flume-ng agent -n TitanicAgent -c conf -f /usr/local/flume/conf/titanic\_flume.conf

Copy the data file spooldir mentioned in conf file:

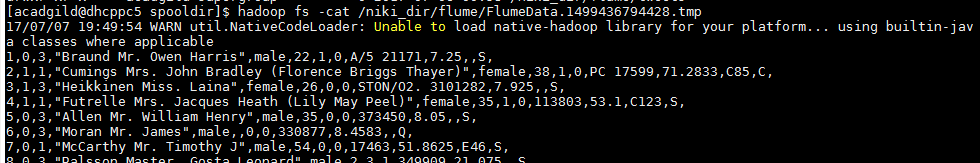
cp /home/acadgild/nikidir/TitanicData.txt /home/acadgild/nikidir/spooldir



Check the file in hdfs:



Check the content:



* **Hive:**

Creating table:

create table titanic\_data

(

passenger\_Id int,

is\_survived int,

pclass int,

name string,

sex string,

age int,

sib\_sp int,

parch int,

ticket string,

fare float,

cabin string,

embarked char(1))

row format delimited

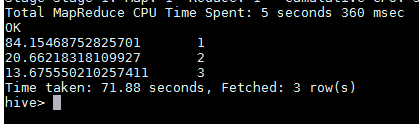
fields terminated by ',';

Inserting data:

LOAD DATA INPATH 'hdfs://localhost:9000/niki\_dir/flume' OVERWRITE INTO TABLE titanic\_data;

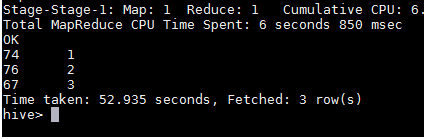
1. In this problem statement, we will find the average fare of each class.

select AVG(fare),pclass from titanic\_data group by pclass;



1. In this problem statement, we will find the number of people alive in each class and embarked at Southampton.

select count(\*), pclass from titanic\_data where embarked='S' and is\_survived=1 group by pclass;



1. In this problem statement, we will find out number of males and females who died in each class.

select pclass, sex, count(\*) from titanic\_data where is\_survived=0 group by pclass,sex;

