--**make new directory**

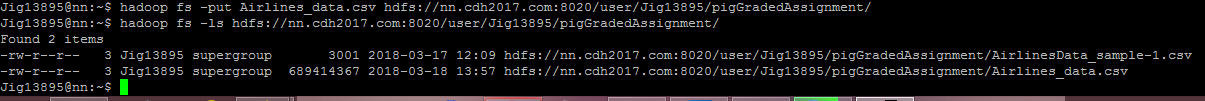
hadoop fs -mkdir pigGradedAssignment

--**move the input file to hdfs**

hadoop fs -put AirlinesData\_sample-1.csv hdfs://nn.cdh2017.com:8020/user/Jig13895/pigGradedAssignment/



hadoop fs -put Airlines\_data.csv hdfs://nn.cdh2017.com:8020/user/Jig13895/pigGradedAssignment/



**--register the necessary jar**

register '/home/data/piggybank.jar';

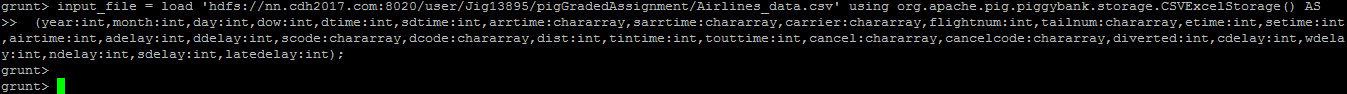


register '/home/Jig13895/minutes.jar';

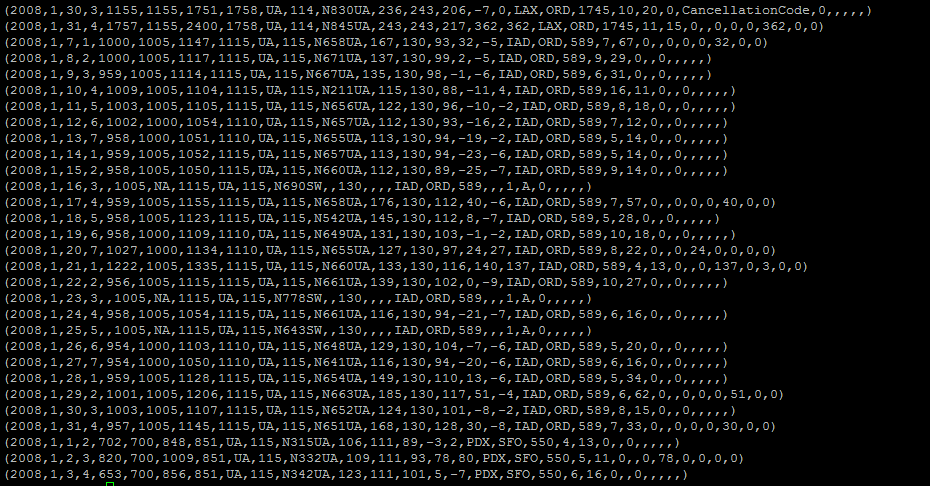


**--load input file**

input\_file = load 'hdfs://nn.cdh2017.com:8020/user/Jig13895/pigGradedAssignment/ Airlines\_data.csv ' using org.apache.pig.piggybank.storage.CSVExcelStorage() AS (year:int,month:int,day:int,dow:int,dtime:int,sdtime:int,arrtime:chararray,sarrtime:chararray,carrier:chararray,flightnum:int,tailnum:chararray,etime:int,setime:int,



**--dump input file**



**--filter records by the given year**

filtered\_out = FILTER input\_file BY year == 2008;

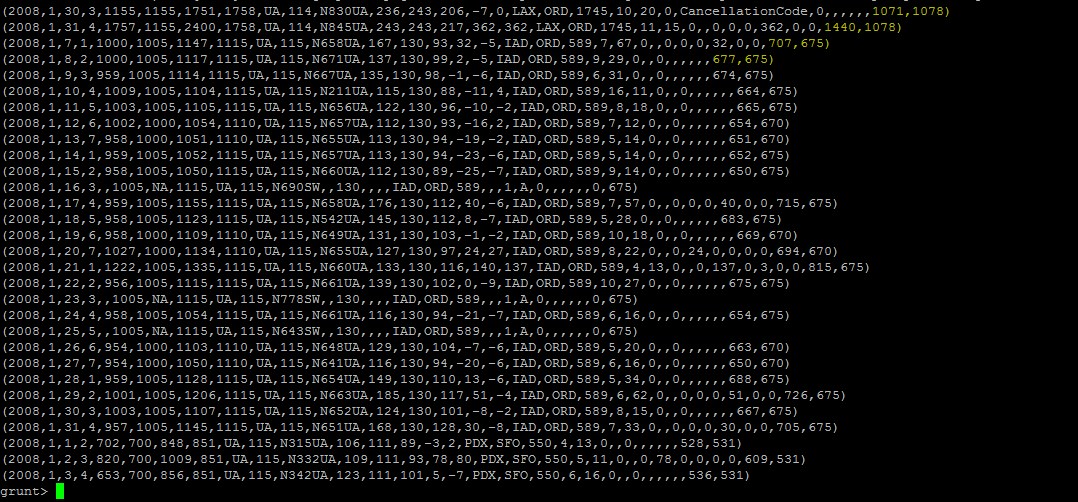


**--transformed input file to convert local time to minutes to compute delay**

ip\_file\_modified= FOREACH filtered\_out generate $0..$28, GetMinutes(arrtime), GetMinutes(sarrtime);



dump ip\_file\_modified;

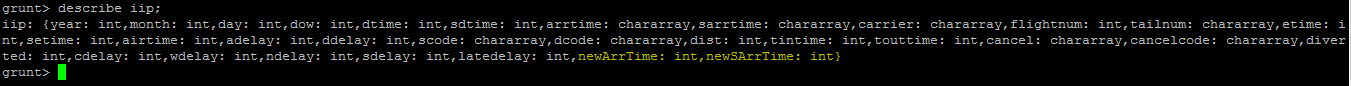


**--type conversion**

iip= FOREACH ip\_file\_modified generate $0..$28, (int)$29 as newArrTime,(int)$30 as newSArrTime;



describe iip;

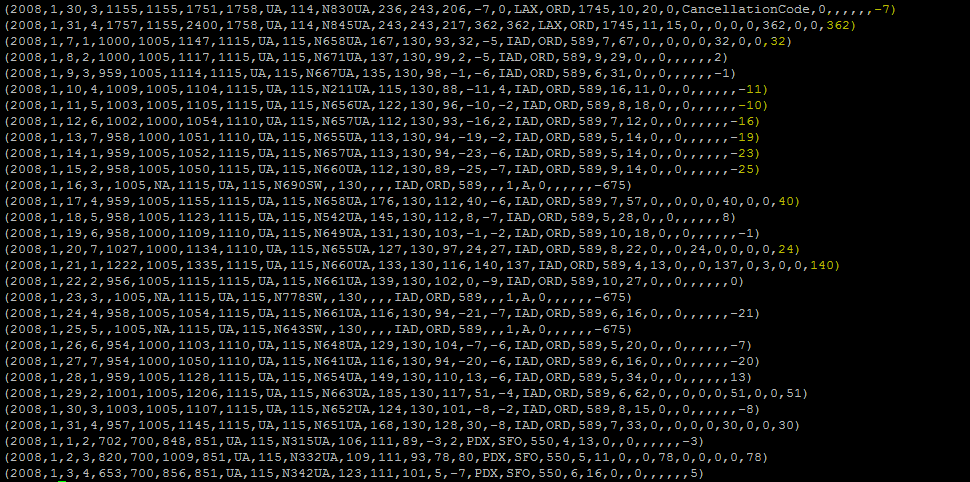


**--compute the delay**

modified\_out\_2= FOREACH iip generate $0..$28,(newArrTime-newSArrTime) as actualdelay;



dump modified\_out\_2;



**--group all flights by month**

grouped\_out\_1= GROUP modified\_out\_2 by month;

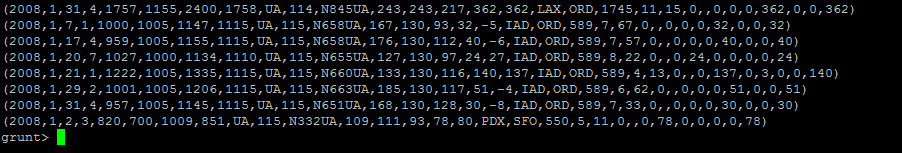


**--filter only delayed flights witm min >15**

delayed\_out = FILTER modified\_out\_2 by actualdelay>15;



dump delayed\_out;



**--group all delayed flights by month**

grouped\_out\_2= GROUP delayed\_out by month;

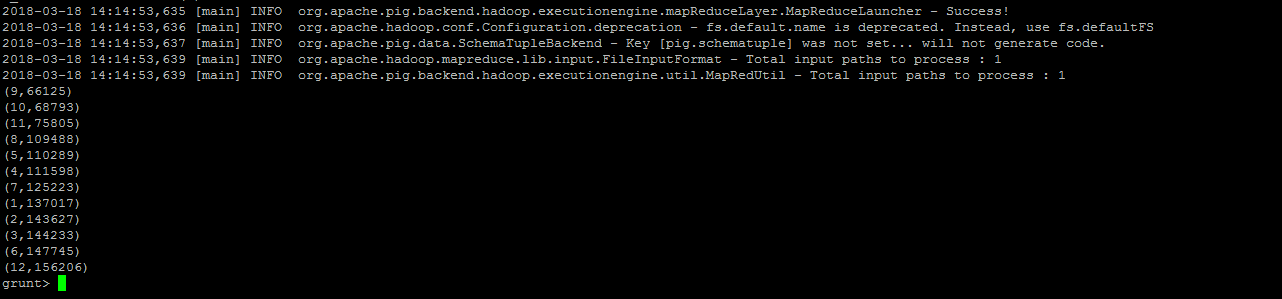


**--compute the total number of delayed flights for each month**

delayPerMonth= FOREACH grouped\_out\_2 generate group,COUNT(delayed\_out.flightnum);

result= order delayPerMonth by $1;

dump result;



**o/p:**

(9,66125)

(10,68793)

(11,75805)

(8,109488)

(5,110289)

(4,111598)

(7,125223)

(1,137017)

(2,143627)

(3,144233)

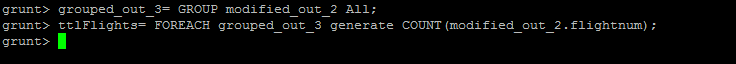
(6,147745)

(12,156206)

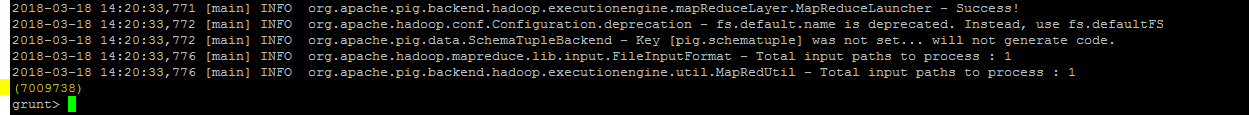
**-- compute the total number of flights**

grouped\_out\_3= GROUP modified\_out\_2 All;

ttlFlights= FOREACH grouped\_out\_3 generate COUNT(modified\_out\_2.flightnum);



dump ttlFlights;



**Total flights count -> o/p**: 7009738

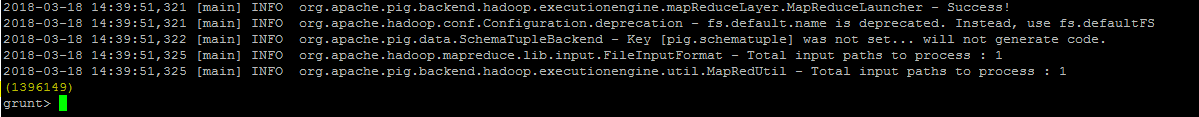
**-- compute the total number of delayed flights**

grouped\_out\_7= GROUP delayed\_out All;

ttlDelayedFlights= FOREACH grouped\_out\_7 generate COUNT(delayed\_out.flightnum);



dump ttlDelayedFlights;



**Total delayed flights count -> o/p**: 1396149

**•Fraction delay from the departure**

**--from the delayed out filter records where departure delay>15 min**

dept\_delay\_out= FILTER delayed\_out BY $15>15;

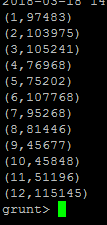
**--compute the total number of delayed flights for each month due to departure delay**

grouped\_out\_4= GROUP dept\_delay\_out by month;

dept\_delayPerMonth= FOREACH grouped\_out\_4 generate group,COUNT(dept\_delay\_out.flightnum);



dump dept\_delayPerMonth;



**•Fraction delay due to carrier**

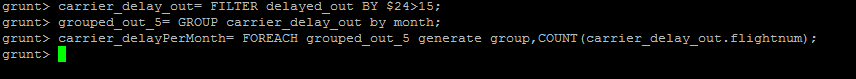
**--from the delayed out filter records where carrire delay>15 min**

carrier\_delay\_out= FILTER delayed\_out BY $24>15;

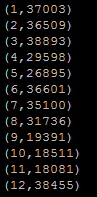
**--compute the total number of delayed flights for each month due to carrier delay**

grouped\_out\_5= GROUP carrier\_delay\_out by month;

carrier\_delayPerMonth= FOREACH grouped\_out\_5 generate group,COUNT(carrier\_delay\_out.flightnum);



dump carrier\_delayPerMonth;



**•Fraction delay due to weather**

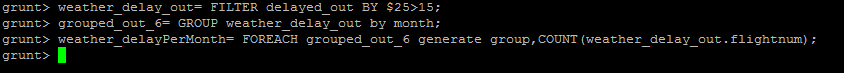
**--from the delayed out filter records where weather delay>15 min**

weather\_delay\_out= FILTER delayed\_out BY $25>15;

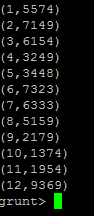
**--compute the total number of delayed flights for each month due to weather delay**

grouped\_out\_6= GROUP weather\_delay\_out by month;

weather\_delayPerMonth= FOREACH grouped\_out\_6 generate group,COUNT(weather\_delay\_out.flightnum);



dump weather\_delayPerMonth;



**Attached is the complete scripts **