**File in HDFS : Airlines\_data.csv  
Output folder name in HDFS : PIGAssignAll; PIGYearlyData**

airline = load 'Airlines\_data.csv' using PigStorage(',');  
airline\_2008 = filter airline by ((int)$0 == 2008);

delayarr = filter airline\_2008 by ((int)$14 >= 15);  
delaydepart = filter airline\_2008 by ((int)$15 >=15);  
delaycarrier = filter delayarr by NOT((int)$24 == 0);  
delayweather = filter delayarr by NOT((int)$25 == 0);  
delayNAS = filter delayarr by NOT((int)$26 == 0);  
delaysecurity = filter delayarr by NOT((int)$27 == 0);  
delaylateaircraft = filter delayarr by NOT((int)$28 == 0);

grpdepdelay = group delaydepart by $0;  
cntdepdelay = foreach grpdepdelay generate group as year, COUNT(delaydepart);

grparrdelay = group delayarr by $1;  
cntarrdelay = foreach grparrdelay generate group as month, COUNT(delayarr);

grpcardelay = group delaycarrier by $0;  
cntcardelay = foreach grpcardelay generate group as year, COUNT(delaycarrier);

grpweatherdelay = group delayweather by $0;  
cntweatherdelay = foreach grpweatherdelay generate group as year, COUNT(delayweather);

grpNASdelay = group delayNAS by $0;  
cntNASdelay = foreach grpNASdelay generate group as year, COUNT(delayNAS);

grpsecuritydelay = group delaysecurity by $0;  
cntsecuritydelay = foreach grpsecuritydelay generate group as year, COUNT(delaysecurity);

grplatedelay = group delaylateaircraft by $0;  
cntlatedelay = foreach grplatedelay generate group as year, COUNT(delaylateaircraft);

grp2008 = group airline\_2008 by $1;  
cnt2008 = foreach grp2008 generate group as month, COUNT(airline\_2008);

grpyear = group airline\_2008 by $0;  
cntyear = foreach grpyear generate group as year, COUNT(airline\_2008);

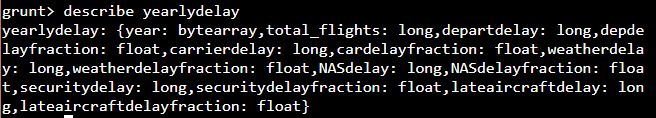
join\_cnt = JOIN cnt2008 by month, cntarrdelay by month;  
join\_yearly = JOIN cntyear by year, cntdepdelay by year, cntcardelay by year, cntweatherdelay by year, cntNASdelay by year, cntsecuritydelay by year, cntlatedelay by year;

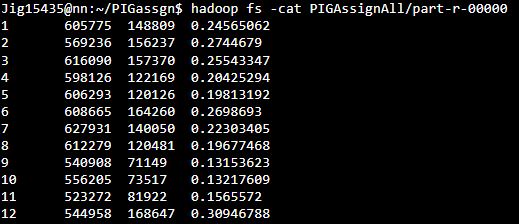
yearlydelay = foreach join\_yearly generate $0 as year, $1 as total\_flights,   
$3 as departdelay, ((float)$3 / (float)$1) as depdelayfraction,   
$5 as carrierdelay, ((float)$5 / (float)$1) as cardelayfraction,   
$7 as weatherdelay, ((float)$7 / (float)$1) as weatherdelayfraction,   
$9 as NASdelay, ((float)$9 / (float)$1) as NASdelayfraction,   
$11 as securitydelay, ((float)$11 /(float)$1) as securitydelayfraction,  
$13 as lateaircraftdelay, ((float)$13 / (float)$1) as lateaircraftdelayfraction;

delayreport = foreach join\_cnt generate $0 as month,$1 as total\_flights, $3 as arrivaldelayed, ((float)$3 / (float)$1) as arrdelayfraction;

store delayreport into 'PIGAssignAll' using PigStorage('\t');  
 store yearlydelay into 'PIGYearlyData' using PigStorage('\t');

**Description of ‘*delayreport’: ***

**Description of *‘yearlydelay’: ***

**Output of the above code: ****



**Column Names in order (Left to Right):**

**PIGAssignAll/part-r-00000:** Contains data of **‘delayreport’** relation

|  |  |  |
| --- | --- | --- |
| **Column#** | **Column Name** | **Description(per month in year 2008)** |
| 1 | Month | Month |
| 2 | total\_flights | Total number of flights |
| 3 | Arrivaldelayed | Number of arrival Delayed Flights |
| 4 | Arrdelayfraction | Arrival Delayed Flights Fraction |

**PIGYearlyData/part-r-00000**: Contains data of **‘yearlydelay’** relation.

|  |  |  |
| --- | --- | --- |
| **Column #** | **Column Name** | **Description** |
| 1 | year | Year |
| 2 | total\_flights | Total Flights in the Year |
| 3 | departdelay | Number of Departure Delayed Flights |
| 4 | depdelayfraction | Fraction of flights departure delayed |
| 5 | carrierdelay | Number of Carrier Delayed Flights |
| 6 | cardelayfraction | Fraction of flights carrier delayed |
| 7 | weatherdelay | Number of Weather Delayed Flights |
| 8 | weatherdelayfraction | Fraction of flights weather delayed |
| 9 | NASdelay | Number of NAS Delayed Flights |
| 10 | NASdelayfraction | Fraction of flights NAS delayed |
| 11 | securitydelay | Number of Security Delayed Flights |
| 12 | securitydelayfraction | Fraction of flights security delayed |
| 13 | lateaircraftdelay | Number of Late Aircraft Delayed Flights |
| 14 | lateaircraftdelayfraction | Fraction of flights delayed due to late aircraft |