**Pricing Analytics:**

In this Pricing analytics case we will identify some of the trends, Identifying the price changes. highest grosser of the year. And various other types of insights can be driven from this data. Hive uses the functionality of Hadoop and is suitable to perform the same on Big data.

**Problem Statement:**

* To identify how many people will be accommodated in the flights, what was the highest number of people with the lowest amount of revenue.
* Identifying the highest revenue generation year and quarter.

**Data:**

The data set contain the revenue generate per seat and the total number of bookings received.

Below are the headings in the data

* Year: The year is specified
* Quarter: The quarter of the years
* Average revenue per seat: Average amount of the revenue generated
* Total no. of booked seats: Seat booked per quarter

hive (niit)> create table airlines(yr int, qtr int, rev\_seat double, booked\_seats bigint)

> row format delimited

> fields terminated by ','

> stored as textfile

> tblproperties("skip.header.line.count"="1");

OK

Time taken: 0.157 seconds

hive (niit)> load data local inpath '/home/hduser/airlines.csv' overwrite into table airlines;

Loading data to table niit.airlines

Table niit.airlines stats: [numFiles=1, numRows=0, totalSize=1820, rawDataSize=0]

OK

Time taken: 0.111 seconds

hive (niit)> select \* from airlines;

OK

airlines.yr airlines.qtr airlines.rev\_seat airlines.booked\_seats

1995 1 296.9 46561

1995 2 296.8 37443

1995 3 287.51 34128

1995 4 287.78 30388

1996 1 283.97 47808

1996 2 275.78 43020

1996 3 269.49 38952

1996 4 278.33 37443

1997 1 283.4 35067

1997 2 289.44 46565

1997 3 282.27 38886

------

To get distinct years:

hive (niit)> select distinct(yr) from airlines;

Query ID = hduser\_20171224134106\_bc263144-42b6-4082-8a81-c8a4f95a7e40

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1514094803278\_0014, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0014/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0014

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-12-24 13:41:10,088 Stage-1 map = 0%, reduce = 0%

2017-12-24 13:41:14,245 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.94 sec

2017-12-24 13:41:18,337 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.2 sec

MapReduce Total cumulative CPU time: 2 seconds 200 msec

Ended Job = job\_1514094803278\_0014

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.2 sec HDFS Read: 8452 HDFS Write: 105 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 200 msec

OK

yr

1995

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013

2014

2015

Time taken: 12.709 seconds, Fetched: 21 row(s)

For each year we have 4 quarters so 21X4 we have 84 records:

1. Find out total revenue/sales generated each year.

We have data from 1995 to 2015. I want to know sales figure for each year.

hive (niit)> select yr, cast(sum(rev\_seat\*booked\_seats) as bigint) from airlines group by yr;

Query ID = hduser\_20171224134635\_1df63661-abe0-469e-8fc9-0ba2b16d525d

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1514094803278\_0015, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0015/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0015

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-12-24 13:46:39,379 Stage-1 map = 0%, reduce = 0%

2017-12-24 13:46:42,579 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.77 sec

2017-12-24 13:46:47,705 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.12 sec

MapReduce Total cumulative CPU time: 4 seconds 120 msec

Ended Job = job\_1514094803278\_0015

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.12 sec HDFS Read: 10889 HDFS Write: 294 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 120 msec

OK

yr \_c1

1995 43494243

1996 46358778

1997 45385236

1998 42035717

1999 48757714

2000 52342926

2001 55533779

2002 47499146

2003 49273210

2004 50631364

2005 46376786

2006 50437898

2007 57309216

2008 57653170

2009 46746446

2010 54861521

2011 51888286

2012 62199127

2013 66363208

2014 62624175

2015 62378990

Time taken: 13.772 seconds, Fetched: 21 row(s)