CREATE EXTERNAL TABLE delay\_flights (

Id int,

Year int,

Month int,

DayofMonth int,

DayOfWeek int,

DepTime int,

CRSDepTime int,

ArrTime double,

CRSArrTime int,

UniqueCarrier string,

FlightNum int,

TailNum string,

ActualElapsedTime double,

CRSElapsedTime int,

AirTime double,

ArrDelay double,

DepDelay int,

Origin string,

Dest string,

Distance int,

TaxiIn double,

TaxiOut int,

Cancelled int,

CancellationCode string,

Diverted int,

CarrierDelay double,

WeatherDelay double,

NASDelay double,

SecurityDelay double,

LateAircraftDelay double)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LOCATION “s3://delayedflight-hive/”;

CREATE TABLE delay\_flights(

Year int,

ArrDelay double,

CarrierDelay double,

WeatherDelay double,

NASDelay double,

SecurityDelay double,

LateAircraftDelay double)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

STORED AS

INPUTFORMAT

'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'

OUTPUTFORMAT

'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'

LOCATION 's3://hive-new-assi/'

TBLPROPERTIES (

"s3select.format" = "csv"

);

\*Year wise carrier delay from 2003-2010\*/

SET s3select.filter=true;

SELECT Year, avg((CarrierDelay /ArrDelay)\*100) AS CarrierDelay FROM delay\_flights GROUP BY Year;

/\*Year wise NAS delay from 2003-2010\*/

SELECT Year, avg((NASDelay /ArrDelay)\*100) FROM delay\_flights GROUP BY Year;

/\*Year wise Weather delay from 2003-2010\*/

SELECT Year, avg((WeatherDelay /ArrDelay)\*100) FROM delay\_flights GROUP BY Year;

/\*Year wise late aircraft delay from 2003-2010\*/

SELECT Year, avg((LateAircraftDelay /ArrDelay)\*100) FROM delay\_flights GROUP BY Year;

/\*Year wise security delay from 2003-2010\*/

SELECT Year, avg((SecurityDelay /ArrDelay)\*100) FROM delay\_flights GROUP BY Year;