

LOAD DATA INTO HDFS

```
./hdfs dfs -copyFromLocal '/Users/nikit/Ashita/airports.csv' '/'
```

```
./hdfs dfs -copyFromLocal '/Users/nikit/Ashita/carriers.csv' '/'
```

```
./hdfs dfs -copyFromLocal '/Users/nikit/Ashita/flight.csv' '/'
```

HIVE COMMANDS:

```
create schema dataset_Airline;  
use dataset_Airline;
```

```
create external table dataset_Airline.onTimePerf (Year INT,  
    Month INT,  
    DayofMonth INT,  
    DayOfWeek INT,  
    DepTime INT,  
    CRSDepTime INT,  
    ArrTime INT,  
    CRSArrTime INT,  
    UniqueCarrier String,  
    FlightNum INT,  
    TailNum String,  
    ActualElapsedTime INT,  
    CRSElapsedTime INT,  
    AirTime INT,  
    ArrDelay INT,  
    DepDelay INT,  
    Origin String,  
    Dest String,  
    Distance INT,  
    TaxiIn INT,  
    TaxiOut INT,  
    Cancelled INT,  
    CancellationCode String,  
    Diverted String,  
    CarrierDelay INT,  
    WeatherDelay INT,
```

```
NASDelay INT,  
SecurityDelay INT,  
LateAircraftDelay INT  
)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
```

```
LOAD DATA INPATH '/flight.csv' OVERWRITE INTO TABLE  
dataset_Airline.onTimePerf;
```

```
create schema airline;  
use airline;
```

```
create table airline.onTimePerf (DayofMonth INT,  
DayOfWeek INT,  
DepTime INT,  
CRSDepTime INT,  
ArrTime INT,  
CRSArrTime INT,  
UniqueCarrier String,  
FlightNum INT,  
TailNum String,  
ActualElapsedTime INT,  
CRSElapsedTime INT,  
AirTime INT,  
ArrDelay INT,  
DepDelay INT,  
Origin String,  
Dest String,  
Distance INT,  
TaxiIn INT,  
TaxiOut INT,  
Cancelled INT,  
CancellationCode String,  
Diverted String,  
CarrierDelay INT,  
WeatherDelay INT,  
NASDelay INT,  
SecurityDelay INT,
```

```
LateAircraftDelay INT
)
PARTITIONED BY (Year Int, Month Int) ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
```

```
SET hive.exec.dynamic.partition = true;
SET hive.exec.dynamic.partition.mode = nonstrict;
```

```
INSERT OVERWRITE TABLE airline.onTimePerf PARTITION(Year,
Month) Select DayofMonth ,
    DayOfWeek ,
    DepTime ,
    CRSDepTime ,
    ArrTime ,
    CRSArrTime,
    UniqueCarrier,
    FlightNum ,
    TailNum ,
    ActualElapsedTime,
    CRSElapsedTime ,
    AirTime ,
    ArrDelay ,
    DepDelay ,
    Origin ,
    Dest ,
    Distance,
    TaxiIn,
    TaxiOut,
    Cancelled,
    CancellationCode,
    Diverted,
    CarrierDelay,
    WeatherDelay,
    NASDelay,
    SecurityDelay,
    LateAircraftDelay , Year, Month FROM stg_airline.onTimePerf;
```

```
create external table airline.airports (Iata String, airport String, city String,  
state String, country String, lat String, longi String) ROW FORMAT  
DELIMITED FIELDS TERMINATED BY ',' ;
```

```
LOAD DATA INPATH '/airports.csv' OVERWRITE INTO TABLE  
airline.airports;
```

```
create external table airline.carriers (Code String, Description String) ROW  
FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

```
LOAD DATA INPATH '/carriers.csv' OVERWRITE INTO TABLE  
airline.carriers;
```

```
drop schema if exists airline cascade;
```

HIVE ANALYSIS 1: FLIGHTS THAT STARTED LATE BUT REACHED ON TIME

```
INSERT OVERWRITE DIRECTORY '/Output/HiveOutput/Analysis1'  
select  
Year,Month,DayofMonth,Origin,Dest,AirTime,Distance,TaxiIn,TaxiOut  
from ontimeperf where DepTime>CRSDepTime and  
ArrTime<=CRSArrTime;
```

HIVE ANALYSIS 2: FLIGHTS THAT TRAVEL LESS THAN 1000 MILES

```
INSERT OVERWRITE DIRECTORY '/Output/HiveOutput/Analysis2.0'  
select count(*) from onTimePerf where Distance > 1000;
```

```
INSERT OVERWRITE DIRECTORY '/Output/HiveOutput/Analysis2.1'  
select count(*) from onTimePerf where Distance < 1000;
```

HIVE ANALYSIS 3: COUNT OF FLIGHTS FOR EACH CARRIER

```
INSERT OVERWRITE DIRECTORY '/Output/HiveOutput/Analysis3'
```

```
Select carriers.description, uniqCount.countCancelled,  
uniqCount.countCarrier from  
(Select UniqueCarrier, sum(cancelled) as countCancelled, count(*) as  
countCarrier from onTimePerf group by UniqueCarrier)  
AS uniqCount, carriers  
where carriers.code = uniqCount.UniqueCarrier;
```

HIVE ANALYSIS 4: FLIGHTS THAT TOOK MORE THAN 15 MINS TO TAXIIN AND TAXOUT

```
select count(*) from ontimeperf where TaxiIn+ TaxiOut>15;
```