

## 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' '/'
```

## PIG COMMANDS:

### 1) LOAD FLIGHT.CSV DATA

```
flightData = LOAD '/flight.csv' USING PigStorage(',') AS (  
  Year:int,  
  Month:int,  
  DayofMonth:int,  
  DayOfWeek:int,  
  DepTime :int,  
  CRSDepTime:int,  
  ArrTime:int,  
  CRSArrTime:int,  
  UniqueCarrier:chararray,  
  FlightNum:int,  
  TailNum:chararray,  
  ActualElapsedTime:int,  
  CRSElapsedTime:int,  
  AirTime:int,  
  ArrDelay:int,  
  DepDelay:int,  
  Origin:chararray,  
  Dest:chararray,  
  Distance:int,  
  TaxiIn:int,  
  TaxiOut:int,  
  Cancelled:int,  
  CancellationCode :chararray,  
  Diverted:chararray,  
  CarrierDelay:int,  
  WeatherDelay:int,
```

```
NASDelay:int,  
SecurityDelay:int,  
LateAircraftDelay:int  
);
```

## **2)LOAD CARRIER.CSV DATA**

```
carrierData = LOAD '/carriers.csv' USING PigStorage(',') AS  
(code:chararray, description:chararray);
```

## **3) LOAD AIRPORTS.CSV DATA**

```
airportData = LOAD '/airports.csv' USING PigStorage(',') AS  
(Iata:chararray,airport:chararray, city:chararray, state:chararray,  
country:chararray,lat:chararray, longi:chararray);
```

## **4) SEQUENCE THE DATA**

```
seqFlightData = RANK flightData;
```

### **PIG ANALYSIS 1: NUMBER OF FLIGHTS ARRIVED BEFORE TIME**

```
filterFlightData = filter flightData by ArrTime<CRSArrTime;
```

```
beforeTimeFlight = foreach (GROUP filterFlightData ALL) GENERATE  
COUNT(filterFlightData);
```

```
store beforeTimeFlight INTO '/Output/PigOutput/Analysis1' USING  
PigStorage(',');
```

### **PIG ANALYSIS 2: NUMBER OF FLIGHTS BETWEEN 2 AIRPORTS**

```
filterFlightData = filter flightData by Origin matches 'BOS' and Dest  
matches 'DFW';
```

```
numberOfFlights = foreach (GROUP filterFlightData ALL) GENERATE  
COUNT(filterFlightData);
```

```
store numberOfFlights INTO '/Output/PigOutput/Analysis2' USING  
PigStorage(',');
```

### **PIG ANALYSIS 3: FLIGHTS BELONGING TO A PARTICULAR CARRIER**

```
joinFlightData = join seqFlightData by UniqueCarrier, carrierData by code;
```

```
filterFlightData = filter joinFlightData by description matches 'Southwest  
Airlines Co.';
```

```
store filterFlightData INTO '/Output/PigOutput/Analysis3' USING  
PigStorage(',');
```

## **PIG ANALYSIS 4: FLIGHTS THAT DEPARTED ON TIME**

```
onTimeFlight = filter seqFlightData by DepTime==CRSDepTime;

onTimeFlightCount = foreach (GROUP onTimeFlight ALL) GENERATE
COUNT(onTimeFlight.rank_flightData);

store onTimeFlightCount INTO '/Output/PigOutput/Analysis4' USING
PigStorage(',');
```

## **PIG ANALYSIS 4: FLIGHTS THAT ARE CANCELLED**

```
flightCancelled = filter seqFlightData by CancellationCode=='B';

store flightCancelled INTO '/Output/PigOutput/Analysis5' USING
PigStorage(',');
```

## **MOVE FILE TO LOCAL**

```
./hdfs dfs -copyToLocal '/Output/PigOutput/Analysis1'
'/Users/nikit/Desktop'

./hdfs dfs -copyToLocal '/Output/PigOutput/Analysis2'
'/Users/nikit/Desktop'

./hdfs dfs -copyToLocal '/Output/PigOutput/Analysis3'
'/Users/nikit/Desktop'

./hdfs dfs -copyToLocal '/Output/PigOutput/Analysis4'
'/Users/nikit/Desktop'

./hdfs dfs -copyToLocal '/Output/PigOutput/Analysis5'
'/Users/nikit/Desktop'
```