hdfs dfs -mkdir -p rawdata/airline/flights

hdfs dfs -mkdir -p rawdata/airline/planeinfo

hdfs dfs -mkdir -p rawdata/airline/airport

hdfs dfs -mkdir -p rawdata/airline/carriers

A screenshot of a computer

Description automatically generated

CREATE DATABASE airline\_db;

USE airline\_db;

A screenshot of a computer

Description automatically generated

A white background with black text

Description automatically generated

CREATE EXTERNAL TABLE Airports\_stg (

    IATA STRING,

    Name STRING,

    City STRING,

    State STRING,

    Country STRING,

    Lat DECIMAL(9, 6),

    Long DECIMAL(9, 6)

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/mavricbdhjan002/rawdata/airline/airport'

tblproperties(

"skip.header.line.count"="1"

);

A screenshot of a computer

Description automatically generated

CREATE EXTERNAL TABLE Carrier\_stg (

    Code STRING,

    Description STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/mavricbdhjan002/rawdata/airline/carriers'

tblproperties(

"skip.header.line.count"="1"

);

A white background with black text

Description automatically generated

CREATE EXTERNAL TABLE PlaneInfo\_stg (

    tailNum STRING,

    type STRING,

    manufacturer STRING,

    issue\_date STRING,

    model STRING,

    status STRING,

    aircraft\_type STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/mavricbdhjan002/rawdata/airline/planeinfo'

tblproperties(

"skip.header.line.count"="1"

);

A white background with black text

Description automatically generated

CREATE EXTERNAL TABLE Flights\_stg (

    Year INT,

    Month INT,

    DayofMonth INT,

    DayOfWeek INT,

    DepTime INT,

    CRSDepTime INT,

    ArrTime INT,

    CRSArrTime INT,

    UniqueCarrier STRING,

    FlightNum STRING,

    TailNum STRING,

    ActualElapsedTime INT,

    CRSElapsedTime INT,

    AirTime INT,

    ArrDelay INT,

    DepDelay INT,

    Origin STRING,

    Dest STRING,

    Distance INT,

    TaxiIn STRING,

    TaxiOut STRING,

    Cancelled STRING,

    CancellationCode STRING,

    Diverted STRING,

    CarrierDelay INT,

    WeatherDelay INT,

    NASDelay INT,

    SecurityDelay INT,

    LateAircraftDelay INT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/mavricbdhjan002/rawdata/airline/flights'

tblproperties(

"skip.header.line.count"="1"

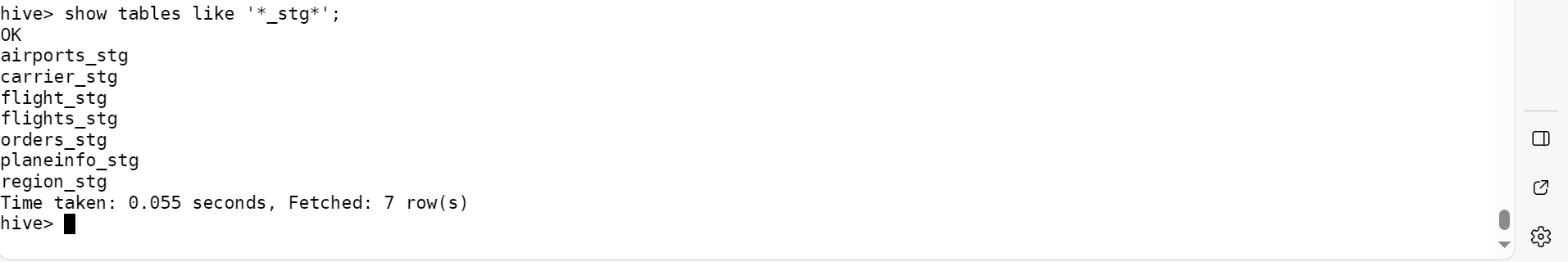
);

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



LOAD DATA INPATH 'rawdata/airline/airport/airports.csv' OVERWRITE INTO TABLE Airports\_stg;

LOAD DATA INPATH '/rawdata/airline/carriers/carriers.csv' OVERWRITE INTO TABLE Carrier\_stg;

LOAD DATA INPATH 'rawdata/airline/plane\_info/plane-data.csv' OVERWRITE INTO TABLE PlaneInfo\_stg;

LOAD DATA INPATH 'AirlinesOntimeAnalysis/2003.csv' OVERWRITE INTO TABLE Flights\_stg;

LOAD DATA INPATH 'AirlinesOntimeAnalysis/2004.csv' OVERWRITE INTO TABLE Flights\_stg;

LOAD DATA INPATH 'AirlinesOntimeAnalysis/2005.csv' OVERWRITE INTO TABLE Flights\_stg;