1. 스키마 파일

CREATE TABLE airdelay(

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 COMMENT 'A = carrier, B = weather, C = NAS, D=security' ,

Diverted INT COMMENT '1 = yes, 0 = no',

CarrierDelay STRING,

WeatherDelay STRING,

NASDelay STRING,

SecurityDelay STRING,

LateAircraftDelay STRING)

Partitioned by (DelayYear INT)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

2. 데이터 로드: '1999~2002.csv파일

LOAD DATA INPATH ‘1999.csv’

OVERWRITE INTO TABLE airdelay

PARTITON(delayYear=’1999’);

LOAD DATA INPATH ‘2000.csv’

OVERWRITE INTO TABLE airdelay

PARTITON(delayYear=’2000’);

LOAD DATA INPATH ‘2001.csv’

OVERWRITE INTO TABLE airdelay

PARTITON(delayYear=’2001’);

LOAD DATA INPATH ‘2002.csv’

OVERWRITE INTO TABLE airdelay

PARTITON(delayYear=’2002’);