**Extract Walmart Data**

#Insert given Walmart Data (Weather, key, train)

#Create and Load Key

create table if not exists `key` (

store\_nbr int(2),

station\_nbr int(2),

primary key (store\_nbr)

);

LOAD DATA LOCAL INFILE "C:/Users/Keith/Desktop/key.csv"

into TABLE `key`

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

ignore 1 lines;

#Create and load Train

create table if not exists `train` (

`Date` DATE ,

`store\_nbr` INT,

`item\_nbr`INT,

`units` INT);

LOAD DATA LOCAL INFILE "C:/Users/Keith/Desktop/train.csv"

into TABLE `train`

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

ignore 1 lines;

select \* from train;

#Create and load Weather

create table if not exists`Weatherimport` (

`station\_nbr` INT ,

`Date` DATE,

`tmax` VARCHAR(45) ,

`tmin` VARCHAR(45) ,

`tavg` VARCHAR(45) ,

`depart` VARCHAR(45) ,

`dewpoint` VARCHAR(45) ,

`wetbulb` VARCHAR(45) ,

`heat` VARCHAR(2) ,

`cool` VARCHAR(2) ,

`sunrise` VARCHAR(4) ,

`sunset` VARCHAR(4) ,

`codesum` VARCHAR(45),

`snowfall` VARCHAR(45) ,

`preciptotal` VARCHAR(45) ,

`stnpressure` VARCHAR(45) ,

`sealevel` VARCHAR(45) ,

`resultspeed` VARCHAR(45) ,

`resultdir` VARCHAR(45),

`avgspeed` VARCHAR(45));

LOAD DATA LOCAL INFILE "C:/Users/Keith/Desktop/weather.csv"

into TABLE `weatherimport`

FIELDS TERMINATED BY ',' ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

**Load**

#Insert station

insert into station

(station\_id)

select distinct(station\_nbr)

from `key`;

#postal code, address, city generated from external data

create table Storetemp(

`store\_id` INT(11) NOT NULL COMMENT '',

`address` VARCHAR(45) NULL DEFAULT NULL COMMENT '',

`postal\_code` VARCHAR(45) NULL DEFAULT NULL COMMENT '',

`city` VARCHAR(45) NULL DEFAULT NULL COMMENT '',

`station\_id` INT(11) NOT NULL COMMENT '');

LOAD DATA LOCAL INFILE "C:/Users/Keith/Desktop/store.csv"

into TABLE `Storetemp`

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

ignore 1 lines;

insert into store

(address,postal\_code,city,station\_id)

select storetemp.address, storetemp.postal\_code, storetemp.city, `key`.station\_nbr

from storetemp, `key`

where `key`.store\_nbr=storetemp.store\_id AND `key`.station\_nbr=storetemp.station\_id;

select \* from store;

#Insert weather table

drop table if exists storetemp;

select \* from weatherimport;

#insert values into weather event

insert into `weather event`

(date,tmax,tmin,tavg,depart,dewpoint,wetbulb,heat,cool,sunrise,sunset,codesum,snowfall,preciptotal,stnpressure,sealevel,resultspeed,resultdir,avgspeed,station\_id)

select date,tmax,tmin,tavg,depart,dewpoint,wetbulb,heat,cool,sunrise,sunset,codesum,snowfall,preciptotal,stnpressure,sealevel,resultspeed,resultdir,avgspeed,station\_nbr

from `weatherimport`;

select \* from `weather event`;

#delete table after insert

DROP TABLE IF EXISTS weatherimport;

#name, description, type of items and departments are extracted from external data

CREATE TABLE itemtemp (

`item\_id` INT NOT NULL COMMENT '',

`name` VARCHAR(45) NULL COMMENT '',

`description` VARCHAR(45) NULL COMMENT '',

`type` VARCHAR(45) NULL COMMENT '',

`department\_id` INT NOT NULL COMMENT '');

LOAD DATA LOCAL INFILE "C:/Users/Vivian/Desktop/item.csv"

into TABLE `itemtemp`

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

ignore 1 lines;

CREATE TABLE departmenttemp(

`department\_id` INT NOT NULL COMMENT '',

`department\_name` VARCHAR(45) NULL COMMENT '');

LOAD DATA LOCAL INFILE "C:/Users/Vivian/Desktop/department.csv"

into TABLE `departmenttemp`

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

ignore 1 lines;

#load data into department table

insert into department (department\_id, department\_name)

select department\_id, department\_name

from departmenttemp;

drop table if exists departmenttemp;

#load data into item table

insert into item (item\_name, item\_description, item\_type, item\_unit\_price, department\_id)

select name, description, type, ROUND((RAND() \* (50-1))+1), department\_id

from itemtemp;

Create table inventorytemp(

`inventory\_id` INT NOT NULL AUTO\_INCREMENT COMMENT '',

`item\_id` INT NOT NULL COMMENT '',

`store\_id` INT NOT NULL COMMENT '',

`last\_update` TIMESTAMP NOT NULL COMMENT '',

`quantity` INT NULL COMMENT '',

PRIMARY KEY (`inventory\_id`) COMMENT '');

insert into inventorytemp (item\_id, store\_id, quantity)

select item\_nbr, store\_nbr, ROUND((RAND() \* (100-1))+1)

from train

limit 4884;

insert into inventory (item\_id, store\_id, last\_update, quantity)

select item\_id, store\_id, last\_update, quantity

from inventorytemp;

drop table inventorytemp;

#create temporary sale table

CREATE TABLE saletemp (

`sale\_id` INT NOT NULL AUTO\_INCREMENT COMMENT '',

`sale\_date` DATETIME NULL COMMENT '',

`sale\_total` VARCHAR(45) NULL COMMENT '',

`store\_id` INT NOT NULL COMMENT '',

PRIMARY KEY (`sale\_id`) COMMENT '');

#load data into saletemp table

insert into saletemp (sale\_date, store\_id)

select date, store\_nbr

from train

group by date, store\_nbr

order by date;

#create temporary line\_item table

CREATE TABLE lineitemtemp (

`line\_id` INT NOT NULL AUTO\_INCREMENT COMMENT '',

`sale\_id` INT NOT NULL COMMENT '',

`item\_id` INT NOT NULL COMMENT '',

`line\_units` INT NULL COMMENT '',

PRIMARY KEY (`line\_id`) COMMENT '');

#transform data into lineitemtemp table

insert into lineitemtemp (sale\_id, item\_id, line\_units)

select sale\_id, item\_nbr, units

from train t

inner join sale s

on t.store\_nbr = s.store\_id

group by t.date, store\_nbr, item\_nbr

having t.date in (select date(sale\_date) from sale)

and units > 0;

#load data into line\_item table

insert into line\_item (sale\_id, item\_id, line\_units)

select sale\_id, item\_id, line\_units

from lineitemtemp;

#calculate and insert sale totals into sale table

insert into sale (sale\_total)

select sum(item\_unit\_price \* line\_units)

from item i, line\_item l, sale s

where i.item\_id = l.item\_id

and l.sale\_id = s.sale\_id

group by sale\_id

order by sale\_id;