To do :

1. Identify three data verification and three data discovery questions for a data warehouse based on this data. Explain giving more context to these questions in terms of how/when managers will need answers to these questions. - [Done](http://drive.google.com/open?id=0BxmZWRHq0gCkM0RQSHpRLTVaVWc)

2. Develop a Star Schema (to be discussed after midterm) with the objective of answering the above sets of questions. When the Star Schema is converted to a “Data Mart”, some of the tables will be different from those based in the Transactional DB. Implementing the Star Schema again involves writing SQL queries to transfer data from the Transactional Database to the Data Mart.

Dimensions:

Sales - Vivian

Store - Fiona

Time - Keith

Weather - Fiona

Item - Vivian

Fact table:

Line Total - WIP

Submit:

4. One page explanation for B.1 (3 points)

Find this at:

<https://docs.google.com/document/d/12e3YEFdmnfBSuLkH6ac0MZOJTQUW-tdC0Xgs4XZ4Ing/edit>

5. Star schema using WB (5 points) [(ERD Picture)](https://drive.google.com/file/d/0B4K8j8tJGDA1Mml3X3dheEZjU3c/view?usp=sharing)

6. SQL code to create the data warehouse from the transactional DB in part A. Similar to Part A you can organize your SQL code as EXTRACT, TRANSFORM and LOAD segments. (However, you do NOT need to write SQL code to extract data from the star schema to answer the verification/discovery questions.) (7 points)

Submitting data (if any)

Write very clearly how I can access the additional raw data you may have used and which I did not provide you. You do not need to submit the intermediate data you create since I will use your SQL queries to do so myself.

use db\_vphan;

#Star Schema

#dim\_weather: (weather\_key (PK), event\_id, tmax,tmin,tavg,depart,dewpoint,wetbulb,heat,cool,sunrise,sunset,codesum,snowfall,preciptotal,stnpressure,sealevel,resultspeed,resultdir,avgspeed,weather\_station\_id)

#dim\_store: (store\_key (PK), store\_id, store\_address, store\_postal\_code, store\_city, store\_station\_id)

#dim\_item: (item\_key (PK), item\_id, item\_name, item\_description, item\_type, item\_unit\_price, item\_store\_id, item\_inventory\_id, item\_inventory\_quantity, item\_department\_id)

#dim\_department: (department\_key (PK), department\_id, department\_name)

#dim\_date: (date\_key (PK), date, date\_year, date\_month, date\_day, date\_year\_week, date\_week, date\_quarter, date\_weekday, date\_month\_name, date\_day\_name)

#fact\_sale: (sale\_key, weather\_key, item\_key, department\_key, store\_key, sale\_quantity, sale\_total, sale\_month, sale\_quarter, sale\_year)

#create dim\_weather table

drop table if exists dim\_weather;

Create table dim\_weather as

select \*, station\_id as weather\_station\_id from `weather event`;

alter table dim\_weather

add `weather\_key` int auto\_increment primary key first;

select \* from dim\_weather;

#create dim\_store table

drop table if exists dim\_store;

create table dim\_store as

select store\_id,address,postal\_code,city,station\_id as store\_station\_id from store;

alter table dim\_store add `store\_key` int AUTO\_INCREMENT PRIMARY KEY first;

#create dim\_item table

drop table if exists dim\_item;

create table dim\_item as

select i.\*, i.department\_id as item\_department\_id, v.inventory\_id as item\_inventory\_id,

v.quantity as item\_inventory\_quantity, v.store\_id as item\_store\_id

from item i, inventory v

where i.item\_id=v.item\_id;

Alter table dim\_item

add column item\_key int not null auto\_increment first,

add primary key (`item\_key`);

select \* from dim\_item;

#create dim\_department table

drop table if exists dim\_department;

create table dim\_department as

select \* from department;

Alter table dim\_department

add column department\_key int not null auto\_increment first,

add primary key (`department\_key`);

select \* from dim\_department;

#create dim\_date time dimension

drop table if exists T;

create table T as select event\_id from db\_vphan.`weather event`;

#find time span

select min(date), max(date) from `weather event`;

SET @d0 = "2012-01-01";

SET @d1 = "2014-10-31";

SET @date = date\_sub(@d0, interval 1 day);

#set up the dim\_date time dimension table

DROP TABLE IF EXISTS dim\_date;

CREATE TABLE `dim\_date` (

`date\_key` varchar(45) NOT NULL,

`date` date DEFAULT NULL,

`date\_year` smallint DEFAULT NULL,

`date\_month` smallint DEFAULT NULL,

`date\_day` smallint DEFAULT NULL,

`date\_year\_week` smallint DEFAULT NULL,

`date\_week` smallint DEFAULT NULL,

`date\_quarter` smallint DEFAULT NULL,

`date\_weekday` smallint DEFAULT NULL,

`date\_month\_name` char(10) DEFAULT NULL,

`date\_day\_name` char(10) DEFAULT NULL,

PRIMARY KEY (`date\_key`)

);

#populate the table with dates

INSERT INTO dim\_date

SELECT @date := date\_add(@date, interval 1 day) as `date`,

date\_format(@date, "%Y%m%d") as date\_key,

year(@date) as date\_year,

month(@date) as date\_month,

day(@date) as date\_day,

date\_format(@date, "%x") as date\_year\_week,

week(@date, 3) as date\_week,

quarter(@date) as date\_quarter,

weekday(@date)+1 as date\_weekday,

monthname(@date) as date\_month\_name,

dayname(@date) as date\_day\_name

FROM T

WHERE date\_add(@date, interval 1 day) <= @d1

ORDER BY date;

select \* from dim\_date;

drop table T;

#Truncate the dim\_date to the weekly level from the above daily level

drop table if exists tempdate;

create table tempdate as

select min(date) as weekbegin, date\_year, date\_quarter, date\_month, date\_week from dim\_date

group by date\_year, date\_quarter, date\_month, date\_week

order by weekbegin;

select \* from tempdate;

#create fact\_sale table

create table fact\_sale as

select w.event\_id as weather\_key, i.item\_id as item\_key,

d.department\_id as department\_key, s.store\_id as store\_key, sum(line\_units) as sale\_quantity,

sum(line\_totals) as sale\_total, date\_month as sale\_month, date\_quarter as sale\_quarter, date\_year as sale\_year

from `weather event` w, station st, item i, department d, store s, line\_item l, inventory v, sale sa, tempdate as t

where w.station\_id = st.station\_id

and st.station\_id = s.station\_id

and s.store\_id = v.store\_id

and v.item\_id = i.item\_id

and i.department\_id = i.department\_id

and i.item\_id = l.item\_id

and l.sale\_id = sa.sale\_id

and sa.store\_id = s.store\_id

and t.weekbegin <= sale\_date

and date\_add(weekbegin, interval 7 day) > sale\_date

group by weather\_key, item\_key, department\_key, store\_key, date\_year, date\_quarter, date\_month

order by date\_year, date\_quarter, date\_month;