CSE 453 Assignment

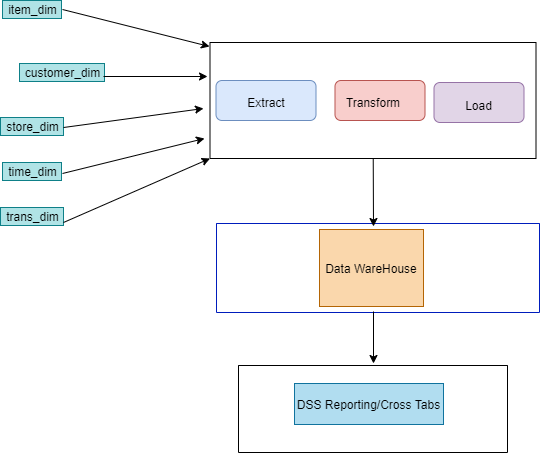
Name: Jainta Paul

ID: 1605022

Section: A

Level 4 Term 1

**TASK 1(Design Architecture)**

****

A Datawarehouse is a single data repository where records from various data sources in integrated for business purpose. The first step to generate data warehouse is the **“ETL”** which includes Extract, Transform , and Load. The steps are described below:

1)Extract: From the given csv file, 5 csv files were created according to the requirements.

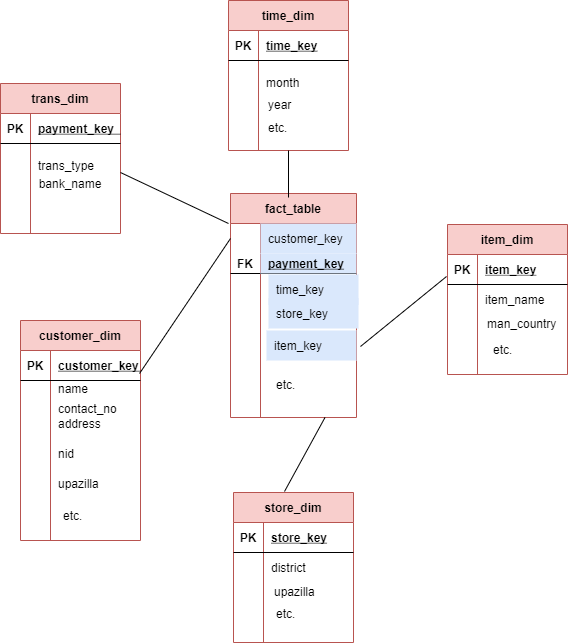
2)Transform: Some preprocessing were required because some of the cells contained null values. The null / absent values were replaced with some default values in order to maintain consistency.

3) The main csv file was decomposed into 5 csv files and those were loaded into 5 different data tables. These are **fact\_table,trans\_dim,customer\_dim,item\_dim,time\_dim,store\_dim**.

**TASK 2(Design Architecture)**

Here there is a “**fact\_table**” and 5 other dimension tables. The 5 dimension tables each has one primary key. The total 5 primary keys are used as foreign keys in the “**fact\_table**”. Apart from these 5 fields, the “fact\_table” has 4 additional fields.

The data of the superstore database will be collected to the DW using **source driven method.**



**TASK 3(Implementation of Star Schema)**

The start schema was implemented in PostgreSQL using Pgadmin using the following queries:

1)Fact Table Creation:

CREATE TABLE Public."fact\_table"(

payment\_key varchar(1000),

coustomer\_key varchar(1000) ,

time\_key varchar(1000) ,

item\_key varchar(1000),

store\_key varchar(1000),

quantity int,

unit varchar(1000),

unit\_price NUMERIC(30) ,

total\_price NUMERIC(30),

PRIMARY KEY (payment\_key,coustomer\_key,time\_key,item\_key,store\_key));

COPY Public."fact\_table" FROM 'F:\dw-assgnment-datasets\fact\_table.csv' DELIMITER ',' CSV HEADER;

1)Dimension Table Creation: Dimension tables were created using the same approach. An example is given below:

CREATE TABLE Public."item\_dim"(

item\_key varchar(1000) PRIMARY KEY,

item\_name varchar(1000),

description varchar(1000),

unit\_price NUMERIC(30),

man\_country varchar(1000),

supplier varchar(1000),

stock\_quantity varchar(1000),

unit varchar(1000) );

COPY Public."item\_dim" FROM 'F:\dw-assgnment-datasets\item\_dim.csv' DELIMITER ',' CSV HEADER;

**TASK 4(Cross-Tabs)**

The cross-tabs were performed using the following queries:

1)This is the cross-tab by joining fact\_table and trans\_dim

select bank\_name,trans\_type,quantity into fact\_table\_trans\_dim

from fact\_table f,trans\_dim t

where f.payment\_key=t.payment\_key;

select bank\_name, sum(quantity)

from fact\_table\_trans\_dim

group by bank\_name;

select trans\_type, sum(quantity)

from fact\_table\_trans\_dim

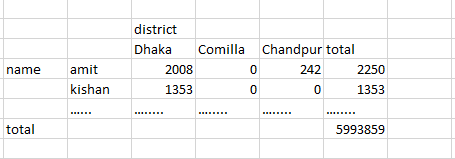
group by trans\_type;

select bank\_name,trans\_type, sum(quantity)

from fact\_table\_trans\_dim

group by bank\_name,trans\_type;

Output:



2) This is the cross-tab by joining fact\_table and customer\_dim

select name,district,quantity into fact\_table\_customer\_dim

from fact\_table f,customer\_dim c

where f.coustomer\_key=c.coustomer\_key;

select name, sum(quantity)

from fact\_table\_customer\_dim

group by name;

select district, sum(quantity)

from fact\_table\_customer\_dim

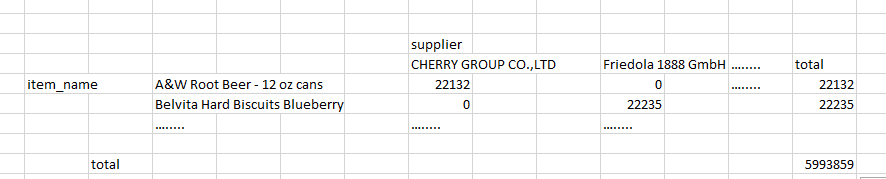
group by district;

select name,district, sum(quantity)

from fact\_table\_customer\_dim

group by name,district;

Output:



3) This is the cross-tab by joining fact\_table and item\_dim

select item\_name,supplier,quantity into fact\_table\_item\_dim

from fact\_table f,item\_dim i

where f.item\_key=i.item\_key;

select supplier, sum(quantity)

from fact\_table\_item\_dim

group by supplier;

select item\_name, sum(quantity)

from fact\_table\_item\_dim

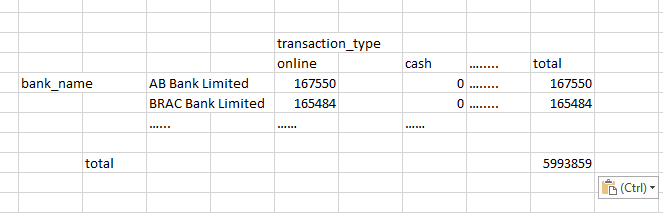
group by item\_name;

select item\_name,supplier, sum(quantity)

from fact\_table\_item\_dim

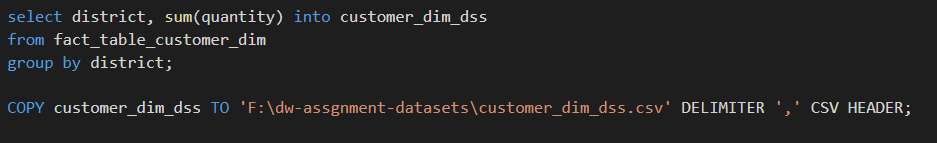
group by item\_name,supplier;

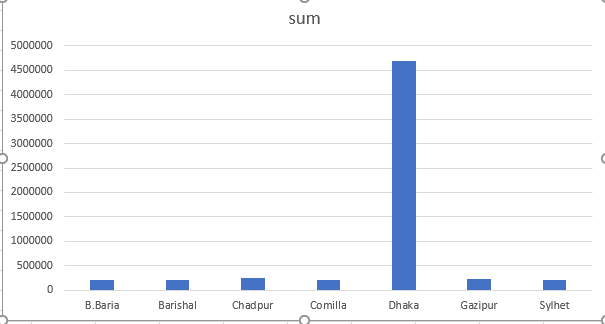
Output:



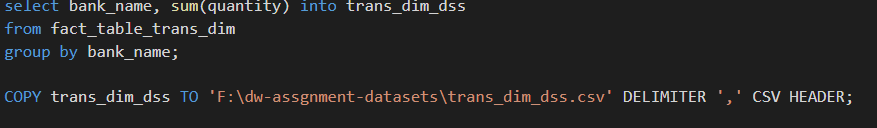
**TASK 5(DSS Reports)**

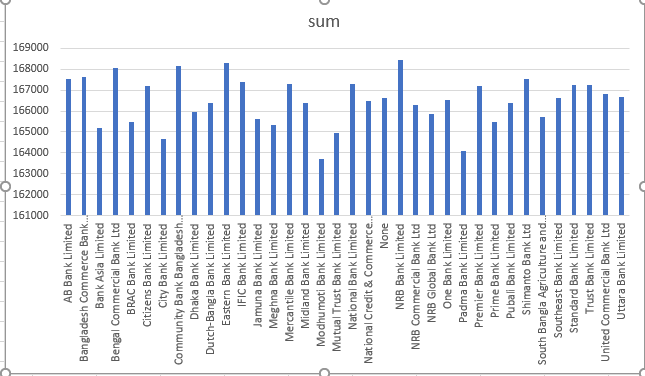
1)Bar chart of customer\_dim against district



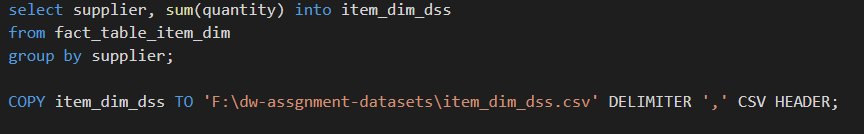


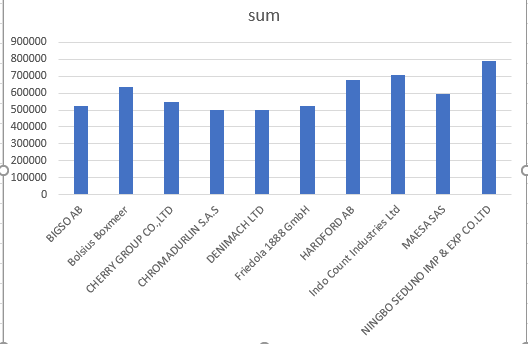
2) Bar chart of trans\_dim against bank\_name



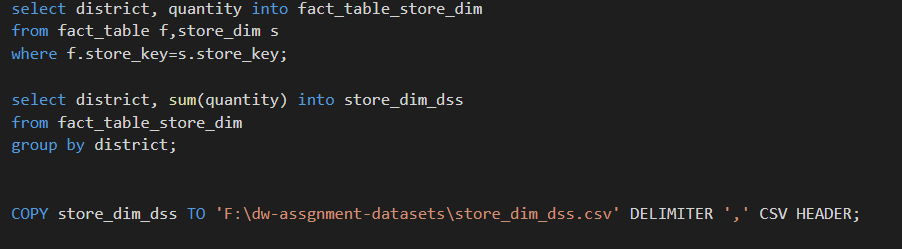


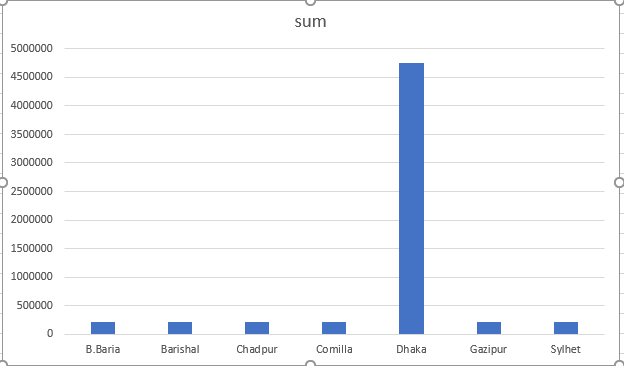
3) Bar chart of item\_dim against supplier





4) Bar chart of store\_dim against district





5) Bar chart of time\_dim against month

