CSE 453 Assignment

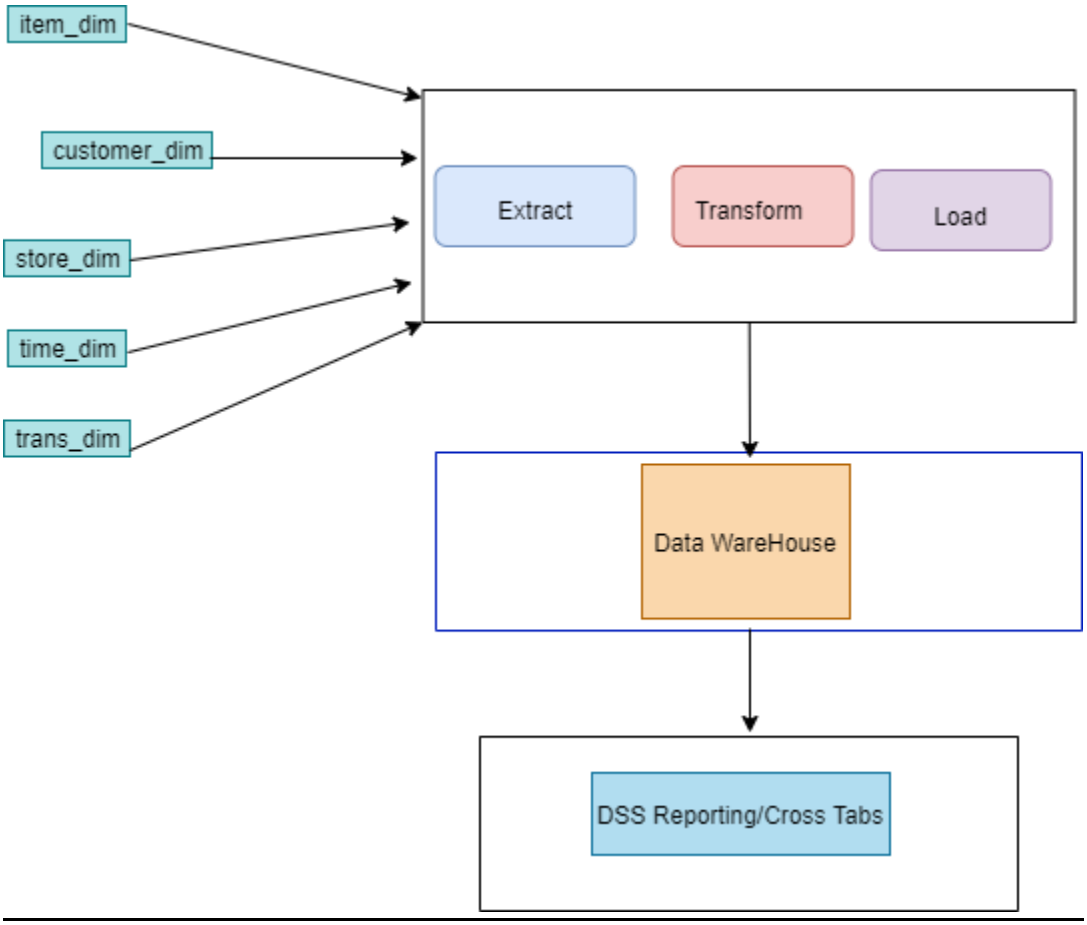
Name: Jainta Paul

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.

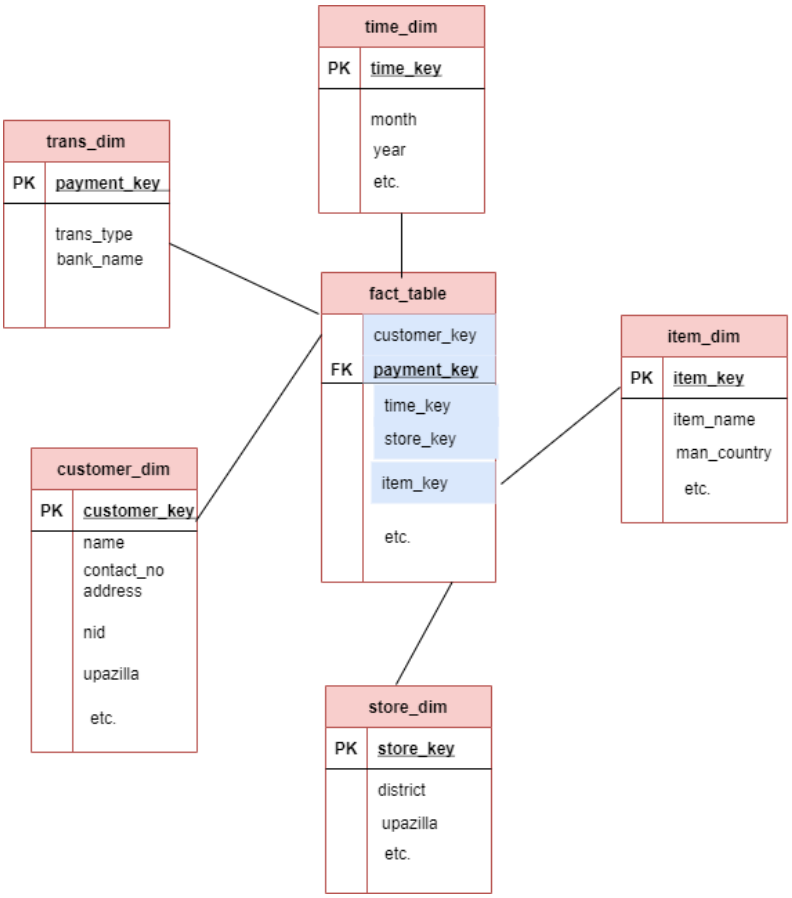
2)Transform: Some preprocessing were required because some of the cells contained null values. The null / absent values were replaced with some default values in order to maintain consistency.

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:

		district			
		Dhaka	Comilla	Chandpur	total
name	amit	2008	0	242	2250
	kishan	1353	0	0	1353
	.....	.....	.....	.....	.....
total					5993859

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:

			supplier				
			CHERRY GROUP CO.,LTD	Friedola 1888 GmbH	.....	total	
item_name		A&W Root Beer - 12 oz cans	22132	0	.....	22132	
		Belvita Hard Biscuits Blueberry	0	22235		22235	
		.....	.....	.....			
	total					5993859	

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:

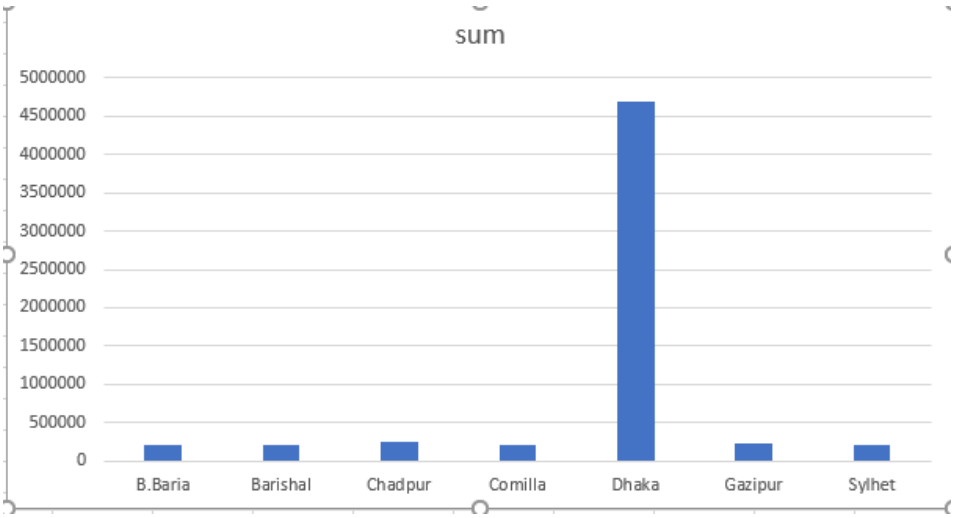
			transaction_type			
			online	cash	.....	total
bank_name		AB Bank Limited	167550	0	.....	167550
		BRAC Bank Limited	165484	0	.....	165484
		.....	.....	.....		
	total					5993859

(Ctrl)

TASK 5(DSS Reports)

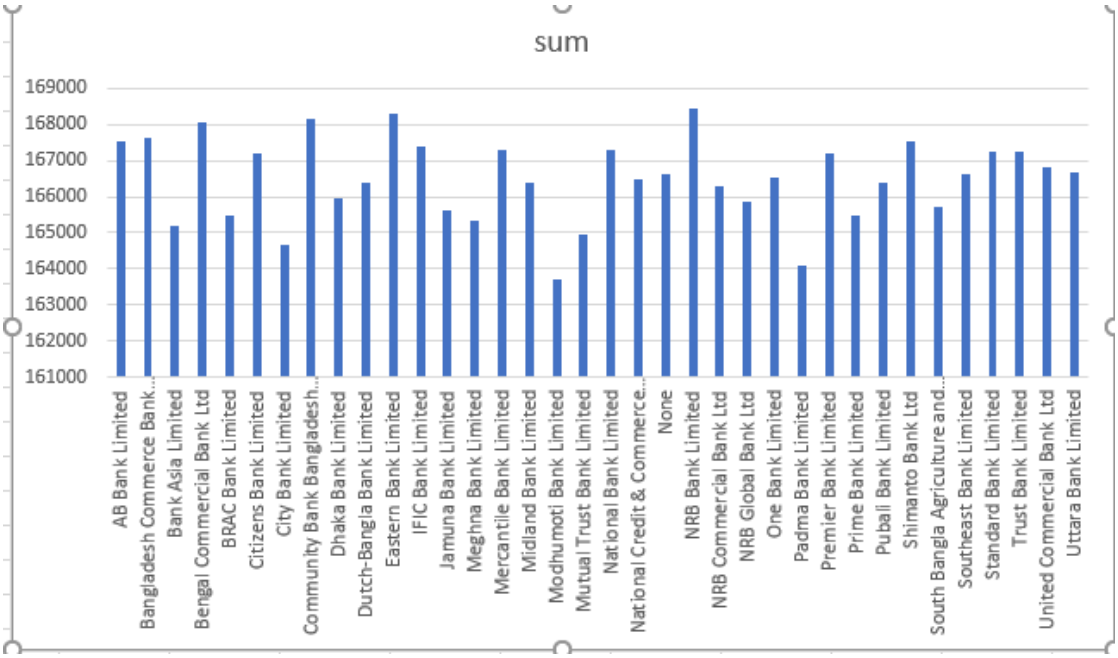
1)Bar chart of customer_dim against district

```
select district, sum(quantity) into customer_dim_dss
from fact_table_customer_dim
group by district;

COPY customer_dim_dss TO 'F:\dw-assgnment-datasets\customer_dim_dss.csv' DELIMITER ',' CSV HEADER;
```



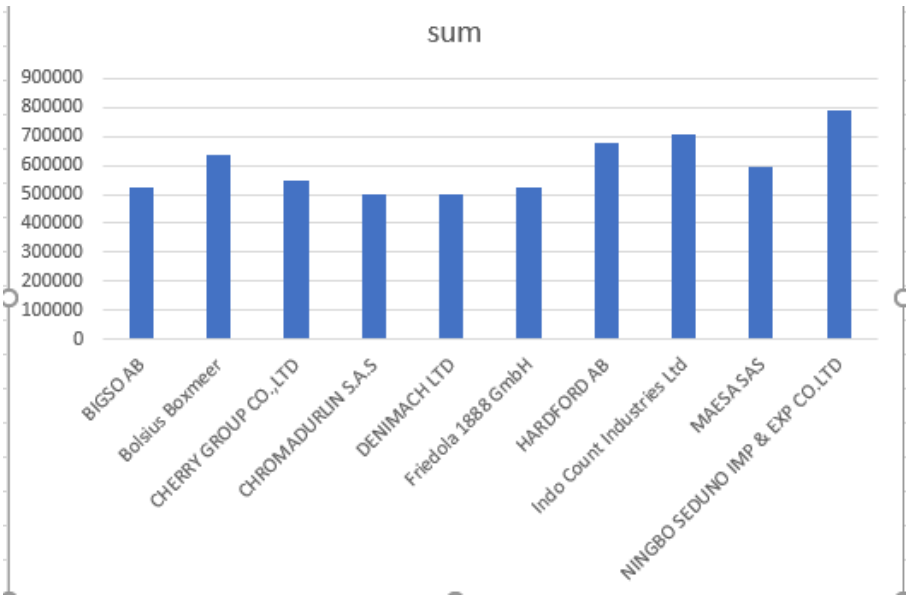
2) Bar chart of trans_dim against bank_name

```
select bank_name, sum(quantity) into trans_dim_dss
from fact_table_trans_dim
group by bank_name;

COPY trans_dim_dss TO 'F:\dw-assgnment-datasets\trans_dim_dss.csv' DELIMITER ',' CSV HEADER;
```



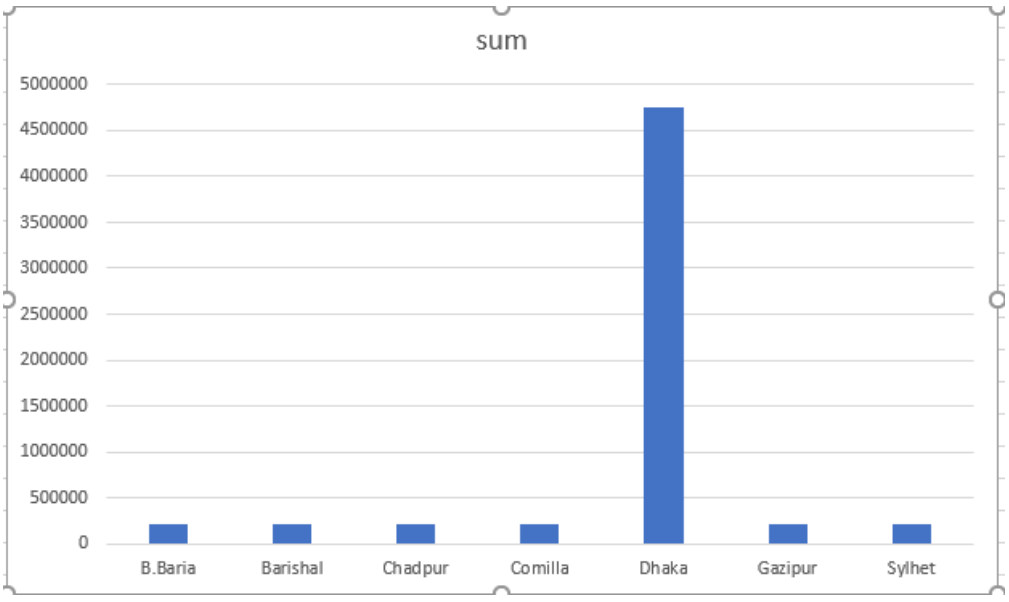
3) Bar chart of item_dim against supplier

```
select supplier, sum(quantity) into item_dim_dss
from fact_table_item_dim
group by supplier;

COPY item_dim_dss TO 'F:\dw-assgnment-datasets\item_dim_dss.csv' DELIMITER ',' CSV HEADER;
```



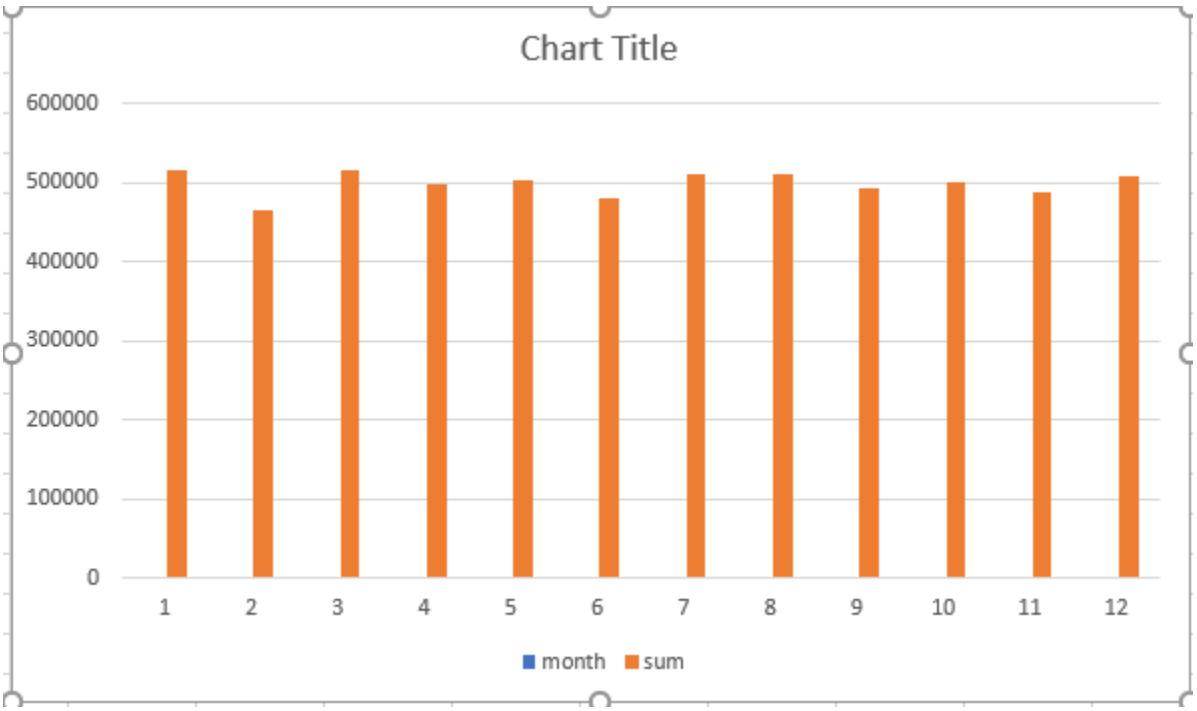
4) Bar chart of store_dim against district

```
select district, quantity into fact_table_store_dim
from fact_table f,store_dim s
where f.store_key=s.store_key;

select district, sum(quantity) into store_dim_dss
from fact_table_store_dim
group by district;


COPY store_dim_dss TO 'F:\dw-assgnment-datasets\store_dim_dss.csv' DELIMITER ',' CSV HEADER;
```



5) Bar chart of time_dim against month

```
select month, quantity into fact_table_time_dim
from fact_table f,time_dim t
where f.time_key=t.time_key;

select month, sum(quantity) into time_dim_dss
from fact_table_time_dim
group by month;

COPY time_dim_dss TO 'F:\dw-assgnment-datasets\time_dim_dss.csv' DELIMITER ',' CSV HEADER;
```



Cube Operations in SQl:

The cube operations SQL to get report data can be done as the following query:

```
select bank_name,trans_type, sum(quantity)
from fact_table_trans_dim
group by CUBE(bank_name,trans_dim);
```

Other 4 Cube operations can be done using this similar operations.