DATA WAREHOUSE AND BUSINESS INTELLIGENCE PROJECT

HASSAN WAQAR

19I-1895 (BS)DS-N

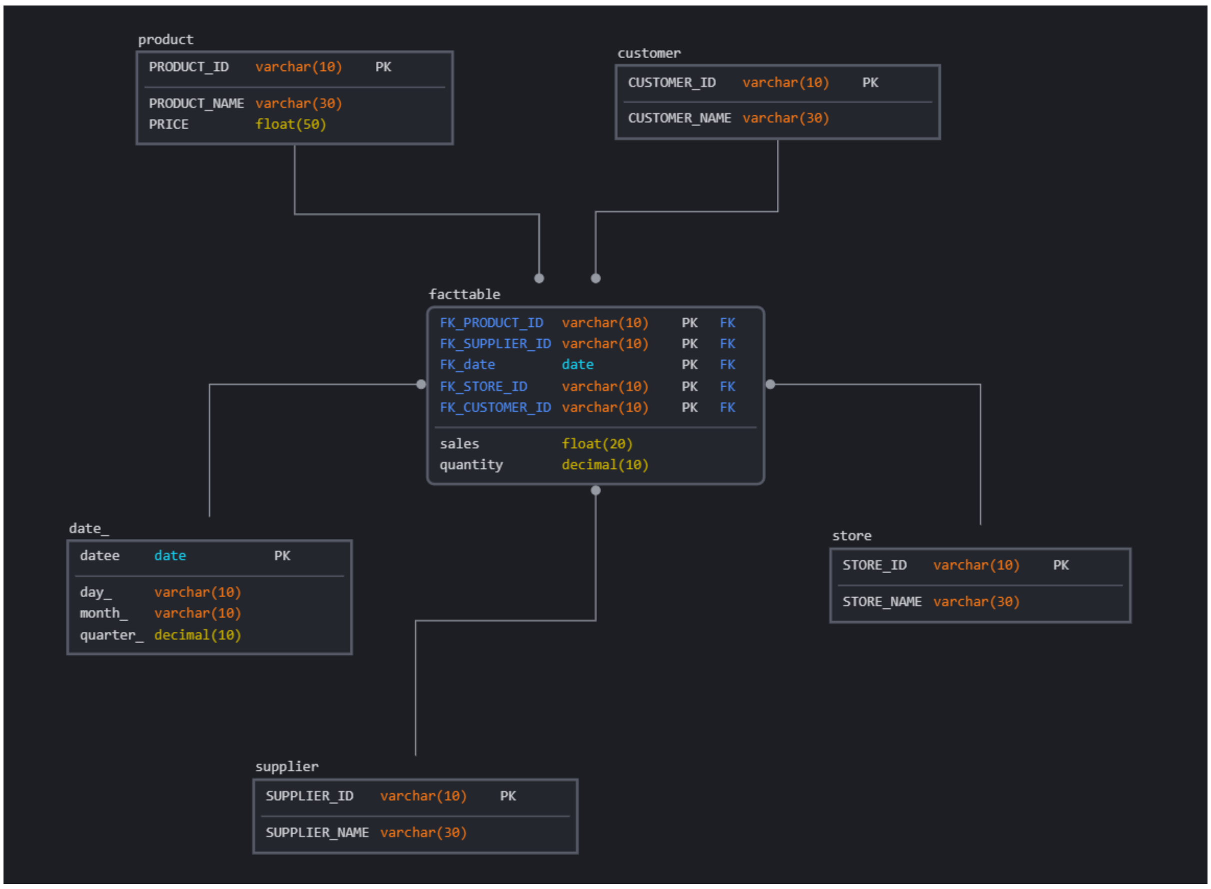
LEARNINGS FROM PROJECT

- **I learned Java for the first time.**
- **I got an in-depth view of how a Data Warehouse works**
- **I learned why Data Warehouses are essential in our world today**

MESH JOIN

Firstly, read master and transaction data in chunks of predefined size i.e., 10 and 50 respectively. We put transaction data in a multi valued hash map and then joining is done by comparing products ids from master data with the ones of the multi valued hash map. If the two product ids match, they'll be inserted into the Data Warehouse depending upon their dimensions.

STAR SCHEMA



product
PRODUCT_ID varchar(10) PK
PRODUCT_NAME varchar(30)
PRICE float(50)
customer
CUSTOMER_ID varchar(10) PK
CUSTOMER_NAME varchar(30)
facttable
FK_PRODUCT_ID varchar(10) PK FK
FK_SUPPLIER_ID varchar(10) PK FK
FK_date date PK FK
FK_STORE_ID varchar(10) PK FK
FK_CUSTOMER_ID varchar(10) PK FK
sales float(20)
quantity decimal(10)
date_
datee date PK
day_ varchar(10)
month_ varchar(10)
quarter_ decimal(10)
store
STORE_ID varchar(10) PK
STORE_NAME varchar(30)
supplier
SUPPLIER_ID varchar(10) PK
SUPPLIER_NAME varchar(30)

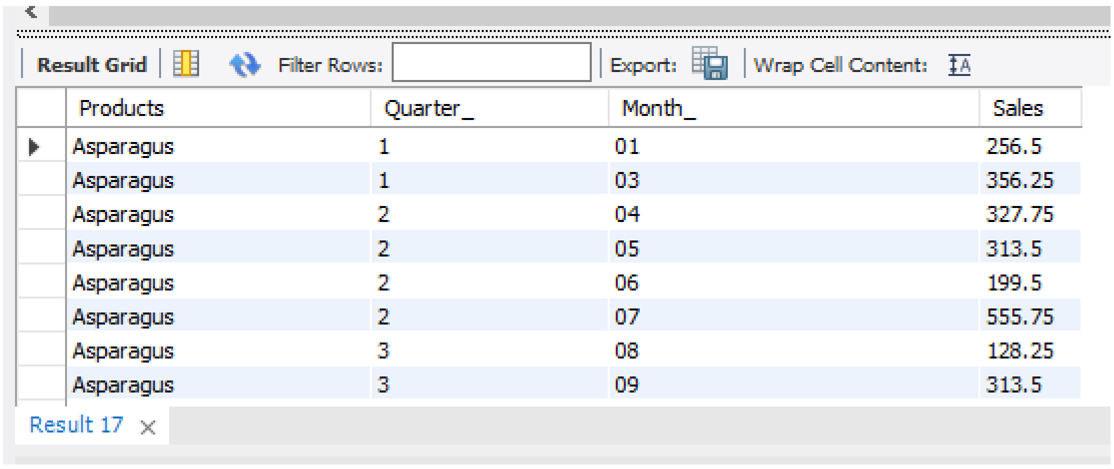
QUERIES

Q1

select PRODUCT_NAME as Products, quarter_ as Quarter_, month_ as Month_, sum(sales) as Sales

from facttable FACTTABLE join date_ on (FACTTABLE.FK_date = datee) join product PRODUCTS on (PRODUCTS.PRODUCT_ID = FACTTABLE.FK_PRODUCT_ID)

group by FACTTABLE.FK_PRODUCT_ID, quarter_, month_;


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Products	Quarter_	Month_	Sales
Asparagus	1	01	256.5
Asparagus	1	03	356.25
Asparagus	2	04	327.75
Asparagus	2	05	313.5
Asparagus	2	06	199.5
Asparagus	2	07	555.75
Asparagus	3	08	128.25
Asparagus	3	09	313.5

Result 17

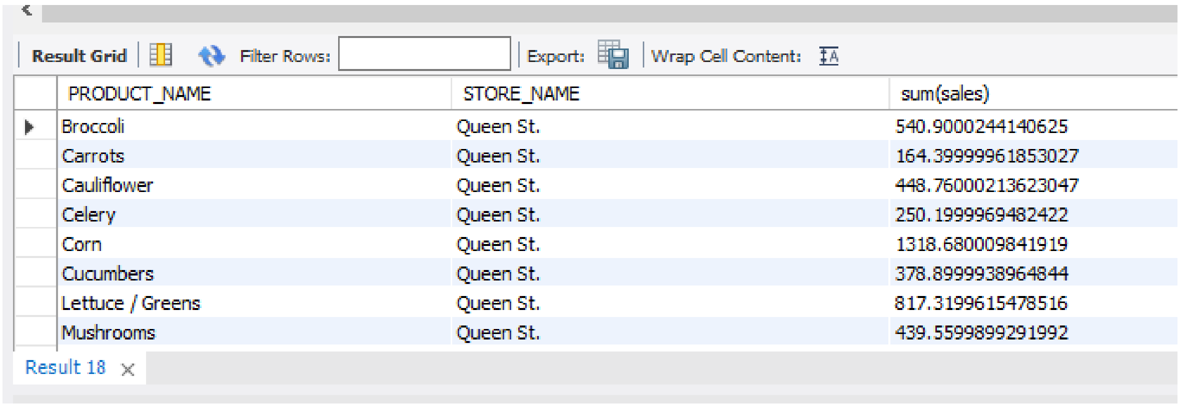
Q2

select PRODUCT_NAME, STORE_NAME, sum(sales) from facttable FACTTABLE

join product PRODUCTS on (FACTTABLE.FK_PRODUCT_ID = PRODUCTS.PRODUCT_ID)

join store STORE on (STORE.STORE_ID = FACTTABLE.FK_STORE_ID)

group by FACTTABLE.FK_PRODUCT_ID, STORE_ID;


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

PRODUCT_NAME	STORE_NAME	sum(sales)
Broccoli	Queen St.	540.9000244140625
Carrots	Queen St.	164.39999961853027
Cauliflower	Queen St.	448.76000213623047
Celery	Queen St.	250.1999969482422
Corn	Queen St.	1318.680009841919
Cucumbers	Queen St.	378.8999938964844
Lettuce / Greens	Queen St.	817.3199615478516
Mushrooms	Queen St.	439.5599899291992

Result 18

Q3

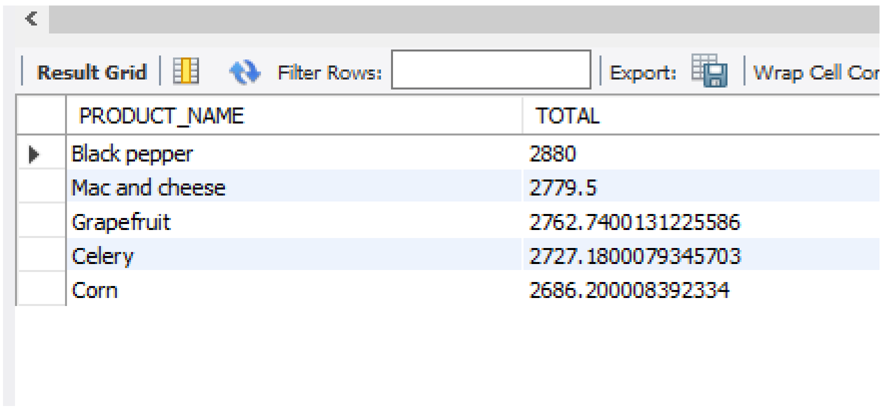
select PRODUCT_NAME, sum(sales) as TOTAL from facttable FACTTABLE

join date_ on (FACTTABLE.FK_date = datee)

join product PRODUCTS on (PRODUCTS.PRODUCT_ID = FACTTABLE.FK_PRODUCT_ID) where dayname(datee) in ("Sunday", "Saturday")

group by product_name, dayname(datee)

order by TOTAL desc limit 5;



Result Grid | Filter Rows: | Export: | Wrap Cell Co

PRODUCT_NAME	TOTAL
Black pepper	2880
Mac and cheese	2779.5
Grapefruit	2762.7400131225586
Celery	2727.1800079345703
Corn	2686.200008392334

Q4

select FK_PRODUCT_ID as PID,

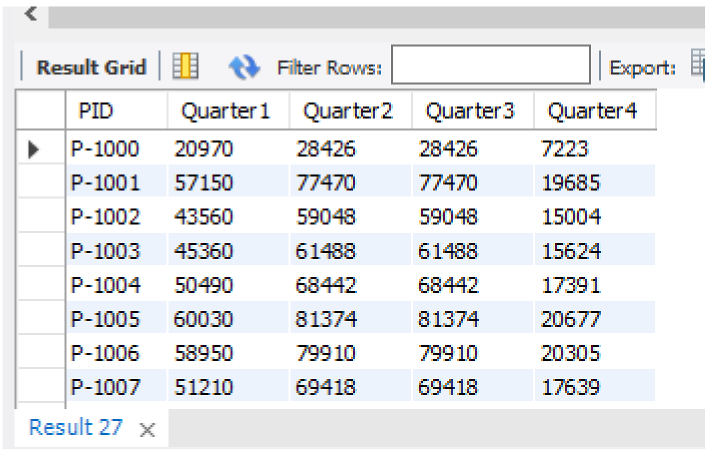
sum(case when quarter_ = 1 then quantity else 0 end) as Quarter1,

sum(case when quarter_ = 2 then quantity else 0 end) as Quarter2,

sum(case when quarter_ = 3 then quantity else 0 end) as Quarter3,

sum(case when quarter_ = 4 then quantity else 0 end) as Quarter4

from facttable natural join date_ group by FK_PRODUCT_ID;



Result Grid | Filter Rows: | Export:

PID	Quarter1	Quarter2	Quarter3	Quarter4
P-1000	20970	28426	28426	7223
P-1001	57150	77470	77470	19685
P-1002	43560	59048	59048	15004
P-1003	45360	61488	61488	15624
P-1004	50490	68442	68442	17391
P-1005	60030	81374	81374	20677
P-1006	58950	79910	79910	20305
P-1007	51210	69418	69418	17639

Result 27

Q5

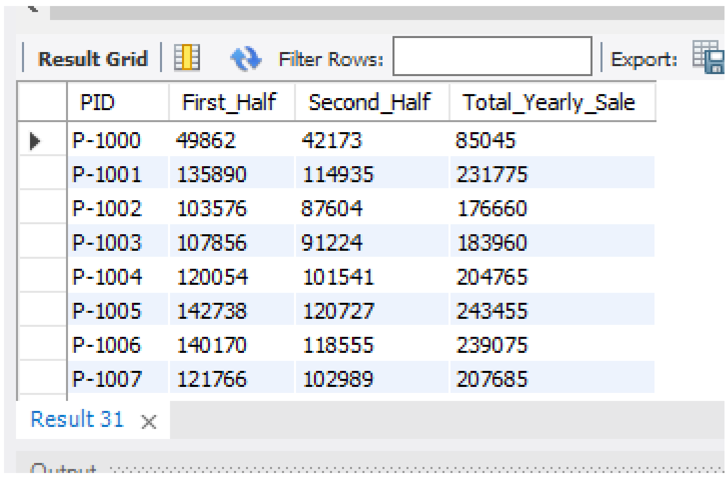
select FK_PRODUCT_ID as PID,

sum(case when month_ >= 6 then quantity else 0 end) as First_Half,

sum(case when month_ <= 6 then quantity else 0 end) as Second_Half,

sum(case when month_ <= 12 then quantity else 0 end) as Total_Yearly_Sale

from facttable natural join date_ group by FK_PRODUCT_ID;

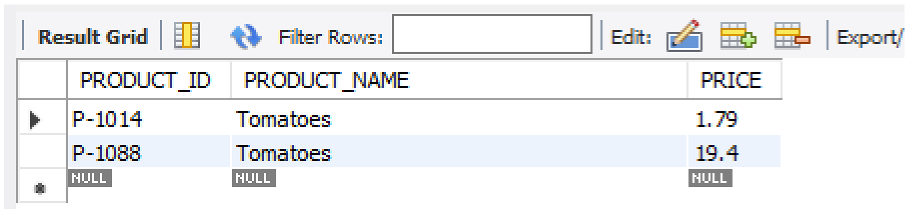


PID	First_Half	Second_Half	Total_Yearly_Sale
P-1000	49862	42173	85045
P-1001	135890	114935	231775
P-1002	103576	87604	176660
P-1003	107856	91224	183960
P-1004	120054	101541	204765
P-1005	142738	120727	243455
P-1006	140170	118555	239075
P-1007	121766	102989	207685

Result 31

Q6

select * from product where (PRODUCT_NAME = "Tomatoes");



PRODUCT_ID	PRODUCT_NAME	PRICE
P-1014	Tomatoes	1.79
P-1088	Tomatoes	19.4
NULL	NULL	NULL

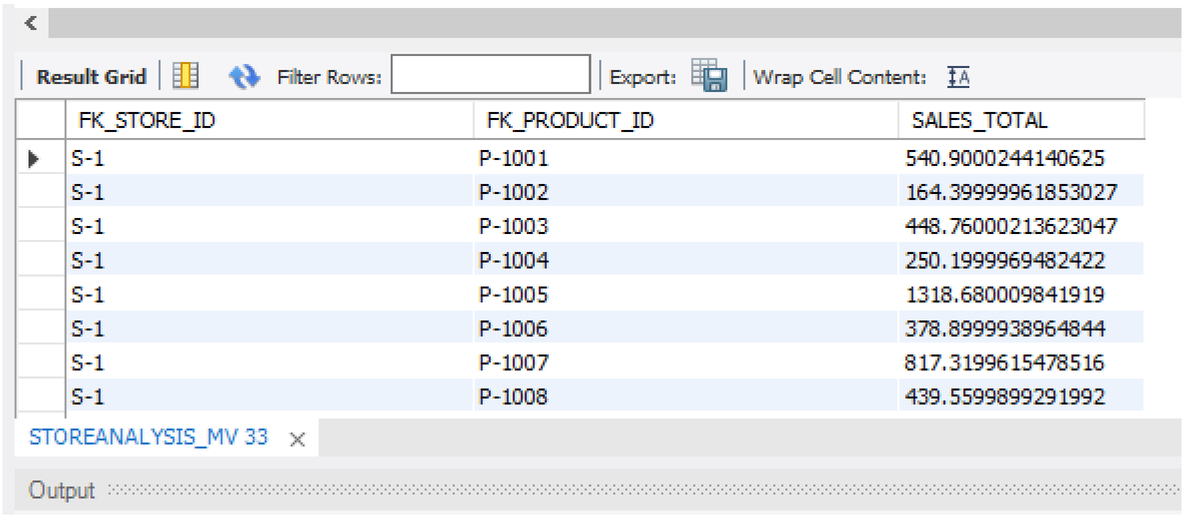
Q7

DROP TABLE if exists STOREANALYSIS_MV;

create view STOREANALYSIS_MV AS SELECT FK_STORE_ID, FK_PRODUCT_ID, sum(sales) as SALES_TOTAL from facttable inner join store on store.STORE_ID = FK_STORE_ID

inner join product on product.PRODUCT_ID = FK_PRODUCT_ID group by STORE_ID, PRODUCT_ID;

```
select * from STOREANALYSIS_MV;
```



FK_STORE_ID	FK_PRODUCT_ID	SALES_TOTAL
S-1	P-1001	540.9000244140625
S-1	P-1002	164.39999961853027
S-1	P-1003	448.76000213623047
S-1	P-1004	250.1999969482422
S-1	P-1005	1318.680009841919
S-1	P-1006	378.8999938964844
S-1	P-1007	817.3199615478516
S-1	P-1008	439.5599899291992