

Case Study Building Data Warehouse

Coffee Shop Sales Data Warehouse and Data Mart

Created by Nabila Ijtihatil Ismail

Business Intelligence Bootcamp batch 9

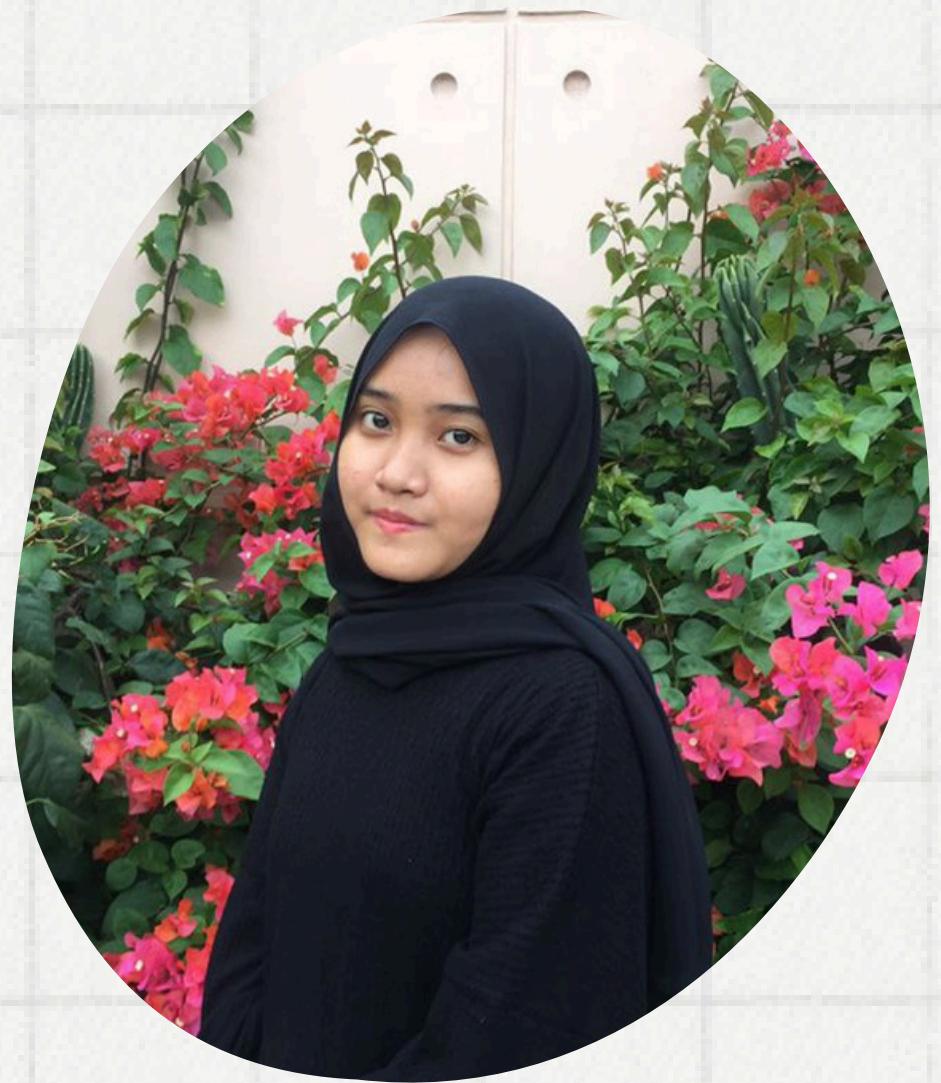
Table of Content

- INTRODUCTION
- TOOLS
- PROBLEM
- BUSINESS CASE
- COFFEE SHOP'S DATA MART
- ERD
- CONCLUSION

INTRODUCTION

Hello! I'm a student in Business Intelligence BootCamp batch 9 in Dibimbing.id. I will share my mini project that analyze Coffee Shop business using a tool called PostgreSQL.

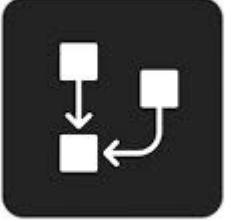
I will arrange an Entity-Relationship Diagram (ERD) for the data warehouse and make the data mart for each purpose. The goal of this analysis is to offer strategy recommendations that can increase Coffee Shop's sales.



TOOLS



I will use PostgreSQL as data warehouse tool. This is also includes create the dimensions of the table and schema that will be used.



DrawSQL

I will use DrawSQL to visualize Entity–Relationship Diagram (ERD). Makes the ERD model is simple, attractive and the relationship is clear with this tool.

Problem

01

Using the provided data, select a business use case that you and your colleagues will work on to build a data warehouse

02

Create an appropriate data warehouse schema. Also, don't forget to structure the necessary fact and dimension tables using a star schema or snowflake schema!

03

The number of tables created will also be considered; create approximately minimum 3 tables for the data mart.

04

Implement it in the form of SQL for the data warehouse and data mart that have been designed



Business Case

Star Corporation has successfully developed 3 Coffee Shop where located in different areas. But one of them has the quantity of product sold is higher than the other store. So, **the team wants to proactively boost sales to maintain and expand their store.** In this case, I want to analyze the product, store, and salesperson performance to determine the best strategy for increasing sales.

Coffee Shop's Data

The Sales SQL dataset consists of **1 fact table**, sales fact, and **4 dimension tables** there are store dimension, product dimension, sales name dimension and time dimension.

Sales Fact Table

```
SELECT * FROM public.fact_sales
```

sale_id integer	store_id integer	sales_name_id integer	time_id integer	product_id integer	quantity integer	price integer
13	3	2	7	10	4	6
12	2	7	31	5	3	9
1	2	4	4	3	2	6
2	3	9	22	7	2	6
3	1	9	30	9	4	5
4	1	10	18	6	5	11
5	2	2	27	6	1	12
6	2	3	22	8	1	15
7	2	1	25	8	5	11
8	1	2	27	2	5	7
9	2	7	5	10	2	5
10	1	1	19	2	3	12
11	1	10	17	10	5	6
14	3	3	26	4	5	11
15	1	10	23	1	2	11
16	1	10	15	10	3	9
17	1	3	20	10	1	7
18	1	6	2	8	1	14

The “**sales_fact**” table, with columns like “**sale_id**”, “**store_id**”, “**sales_name_id**”, “**time_id**”, “**product_id**”, “**quantity**”, and “**price**” is designed to know the detail for the product sales transaction. The type of those columns is “**integer**” which means that the data is numeric.

sale_id is the “**primary key**” from this table. The primary key must be a unique id. Furthermore, there is a “**foreign key**” used to connect with other reference tables. **store_id**, **sales_name_id**, **time_id**, **product_id** is the foreign key.

Store Dimension Table

```
SELECT * FROM public.dim_store  
ORDER BY store_id ASC
```



store_id [PK] integer	store_name character varying (100)	city character varying (50)	state character varying (50)	country character varying (50)
1	Starbucks Central	New York	NY	USA
2	Downtown Starbucks	Los Angeles	CA	USA
3	Starbucks Corner	Chicago	IL	USA

The “**dim_store**” table, with columns like “**store_id**”, “**store_name**”, “**city**”, “**state**”, and “**country**” is designed to know the store information about different store locations. This table is crucial for managing data related to each store.

The type of those columns are “**integer**” and “**character varying**” which means that the data is numeric (only for **store_id**) and character.

Product Dimension Table

```
SELECT * FROM public.dim_product  
ORDER BY product_id ASC
```



product_id [PK] integer	product_name character varying (100)	category character varying (50)
1	Caffè Americano	Coffee
2	Caramel Macchiato	Coffee
3	Iced Green Tea Latte	Tea
4	Java Chip Frappuccino	Coffee
5	Croissant	Bakery
6	Pumpkin Spice Latte	Coffee
7	Blueberry Muffin	Bakery
8	Chai Tea Latte	Tea
9	Flat White	Coffee
10	Mocha	Coffee

The “dim_product” table, with columns like “product_id”, “product_name”, and “category” is designed to give information about the products offered by the store. The types of those columns are “integer” and “character varying”.

Sales Name Dimension Table

```
SELECT * FROM public.dim_sales_name  
ORDER BY sales_name_id ASC
```



sales_name_id [PK] integer	sales_name character varying (100)	sales_age integer	sales_gender character varying (10)
1	John Doe	35	Male
2	Jane Smith	28	Female
3	Michael Johnson	45	Male
4	Emily Brown	32	Female
5	Daniel Williams	40	Male
6	Jessica Davis	26	Female
7	David Martinez	38	Male
8	Sophia Wilson	29	Female
9	James Anderson	41	Male
10	Olivia Taylor	34	Female

The “**dim_sales_name**” table, with columns like “**sales_name_id**”, “**sales_name**”, “**sales_age**” and “**sales_gender**” is designed to know the detail of the salesperson. The types of those columns are “**integer**” and “**character varying**”.

Time Dimension Table

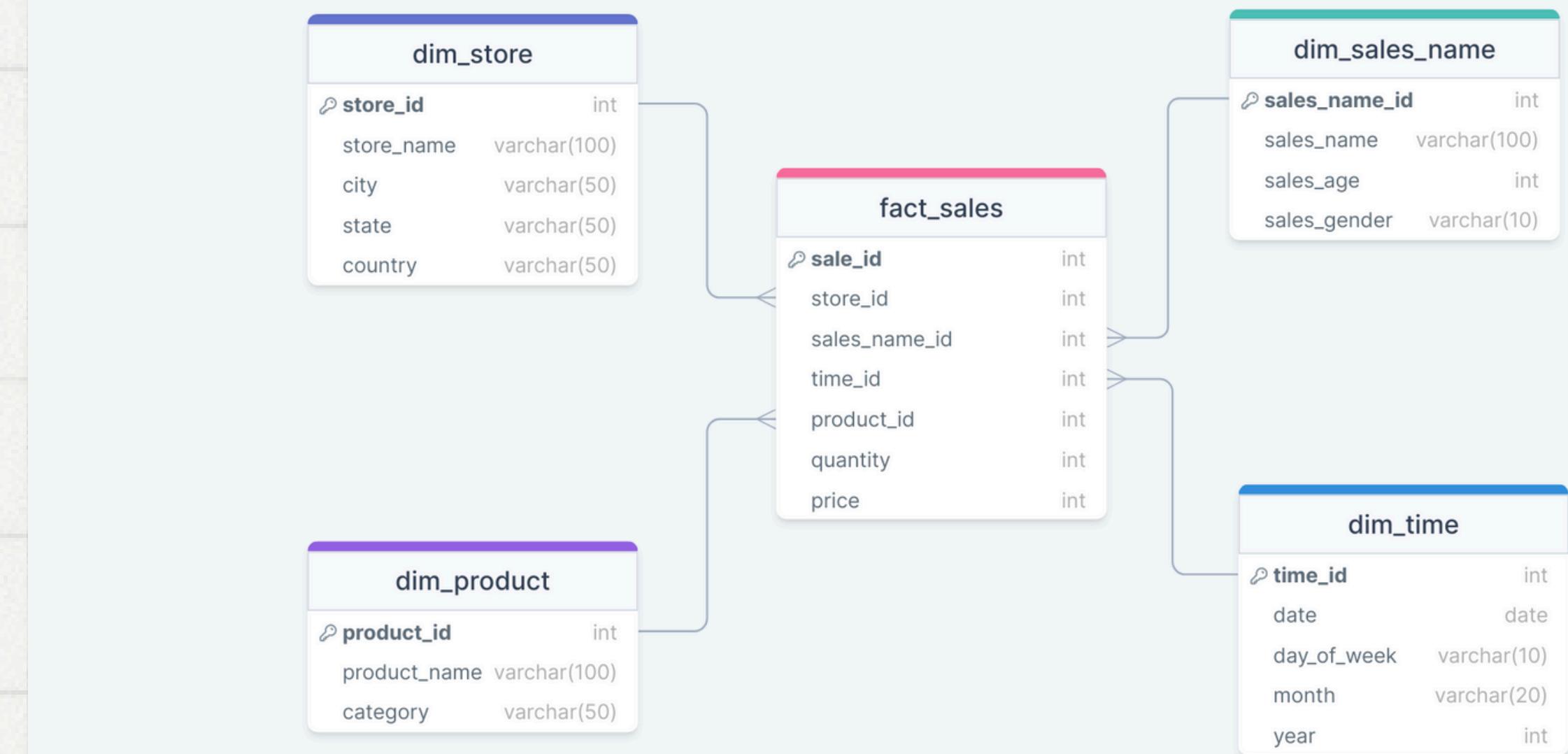
```
SELECT * FROM public.dim_time  
ORDER BY time_id ASC
```



time_id [PK] integer	date date	day_of_week character varying (10)	month character varying (20)	year integer
1	2024-01-01	Monday	January	2024
2	2024-01-02	Tuesday	January	2024
3	2024-01-03	Wednesday	January	2024
4	2024-01-04	Thursday	January	2024
5	2024-01-05	Friday	January	2024
6	2024-01-06	Saturday	January	2024
7	2024-01-07	Sunday	January	2024
8	2024-01-08	Monday	January	2024
9	2024-01-09	Tuesday	January	2024
10	2024-01-10	Wednesday	January	2024
11	2024-01-11	Thursday	January	2024
12	2024-01-12	Friday	January	2024
13	2024-01-13	Saturday	January	2024
14	2024-01-14	Sunday	January	2024
15	2024-01-15	Monday	January	2024
16	2024-01-16	Tuesday	January	2024
17	2024-01-17	Wednesday	January	2024
18	2024-01-18	Thursday	January	2024

The “dim_time” table, with columns like “time_id”, “date”, “day_of_week”, “month”, and “year” is designed to give store information related to time. This table is essential to make a time-based reporting. The types of those columns are “integer”, “date” and “character varying”.

Here is the ERD of the sales dataset



The ERD uses **star schema**. Star schema is a model which the fact table directly connected to the dimension table. The structure looks like a star, where the dimension table is around the fact table.

Foreign key from fact table connected to the dimension table. For example, `store_id` connects to the `store` dimension table, and `time_id` connects to the `time` dimension table. This makes the **performance query faster** and **simpler**.

COFFEE SHOP'S DATA MART

DATA MART 1: PRODUCT ANALYSIS

★ this data mart is designed to **analyze sales data for each product**. It enables the business to gain insight into product sales and determine the right strategy for each product.

DATA MART 2: STORE PERFORMANCE ANALYSIS

★ this data mart is designed to **analyze total revenue from each store**. It enables the business to determine the right marketing strategy to increase sales in each store.

DATA MART 3: SALESPERSON ANALYSIS

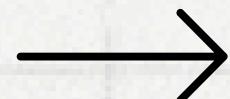
★ this data mart is designed to **analyze salesperson performance**. It enables the business to make good decisions for employee development training and give appreciation to the best employees.

DATA MART 1: PRODUCT ANALYSIS

Top Selling Products

```
create view top_selling_products as
select
    dp.product_name,
    dp.category,
    sum(s.quantity) as total_quantity_sold
from fact_sales s
join dim_product dp on s.product_id = dp.product_id
group by dp.product_name, dp.category
order by total_quantity_sold desc;
```

Top-selling products table is created to track **how many pieces were sold of each product**, with columns including **product_name**, **category** and **total_quantity_sold**.



```
SELECT * FROM public.top_selling_products
```

product_name	category	total_quantity_sold
Iced Green Tea Latte	Tea	332
Croissant	Bakery	316
Flat White	Coffee	315
Blueberry Muffin	Bakery	314
Pumpkin Spice Latte	Coffee	308
Caramel Macchiato	Coffee	296
Java Chip Frappuccino	Coffee	294
Caffè Americano	Coffee	289
Mocha	Coffee	271
Chai Tea Latte	Tea	266

The table shows the **total quantity sold** for each product. The **Iced Green Tea Latte** stands out at **the top** sold with a total quantity sold 332 pcs. Meanwhile, **Chai Tea Latte** is in the last position with total quantity sold 266 pcs.

DATA MART 1: PRODUCT ANALYSIS

Best Product Category

```
create view best_product_category as
select
    dp.category,
    sum(s.quantity) as total_quantity_sold
from fact_sales s
join dim_product dp on s.product_id = dp.product_id
group by dp.category
order by total_quantity_sold desc;
```



```
SELECT * FROM public.best_product_category
```

	category character varying (50)	total_quantity_sold bigint
1	Coffee	1773
2	Bakery	630
3	Tea	598

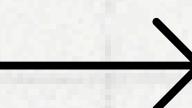
The **best product category** table is designed to **track the total quantity sold group by the product category**, with columns including **category** and **total quantity sold**.

The table shows that the **Coffee product in the 1st position with the total quantity sold is 1773 pcs**, then following by the **Bakery product 630 pcs** and the **Tea product in the last position with 598 pcs**.

DATA MART 2: STORE PERFORMANCE ANALYSIS

Monthly Sales

```
create view monthly_sales as
select
    dt.year,
    dt.month,
    ds.store_name,
    ds.city,
    sum(s.quantity) as total_quantity_sold,
    sum(quantity * price) as total_revenue
from fact_sales s
join dim_time dt on s.time_id = dt.time_id
join dim_store ds on s.store_id = ds.store_id
group by 1,2,3,4;
```



```
SELECT * FROM public.monthly_sales
```

year integer	month character varying (20)	store_name character varying (100)	city character varying (50)	total_quantity_sold bigint	total_revenue bigint
2024	January	Downtown Starbucks	Los Angeles	969	9767
2024	January	Starbucks Central	New York	952	9413
2024	January	Starbucks Corner	Chicago	1080	10474

The monthly sales table is designed to track and understand the total quantity sold and total revenue for each store in January 2024, with columns including year, month, store_name, city, total_quantity_sold and total_revenue.

The table shows that there is a difference that is not very significant across the store.

DATA MART 3: SALESPERSON ANALYSIS

Salesperson Performance

```
create view salesperson_performance as
select
    dsn.sales_name,
    sum(s.quantity) as total_quantity_sold
from fact_sales s
join dim_sales_name dsn on s.sales_name_id = dsn.sales_name_id
group by 1
order by 2 desc;
```



```
SELECT * FROM public.salesperson_performance
```

sales_name	total_quantity_sold
John Doe	367
James Anderson	331
Jane Smith	322
Emily Brown	310
Jessica Davis	302
Daniel Williams	299
David Martinez	280
Sophia Wilson	280
Michael Johnson	278
Olivia Taylor	232

The salesperson performance table is designed to analyze the performance of each salesperson based on the product that they sell.

From the table, John Doe can sell more products than others with 367 pcs products sold.

DATA MART 3: SALESPERSON ANALYSIS

Best Employee by Store

```
create view best_employee_by_store as
with RankedSales as (
    select
        dsn.sales_name,
        ds.store_name,
        sum(quantity) as total_quantity_sold,
        rank() over(partition by ds.store_name order by sum(quantity) desc) as rank_in_store
    from fact_sales s
    join dim_sales_name dsn on s.sales_name_id = dsn.sales_name_id
    join dim_store ds on s.store_id = ds.store_id
    group by 1,2)
select sales_name, store_name, total_quantity_sold, rank_in_store
from RankedSales
where rank_in_store = 1;
```

```
SELECT * FROM public.best_employee_by_store
```



sales_name	store_name	total_quantity_sold	rank_in_store
Emily Brown	Downtown Starbucks	130	1
Jessica Davis	Starbucks Central	120	1
John Doe	Starbucks Corner	158	1

The best employee by store table is designed to show the best employee in terms of quantity sold for each store.

From the table, we can know which salesperson that in the top position for each store. For Downtown Starbucks there is Emily Brown in the 1st position, then in the Starbucks Central there is Jessica Davis and in the Starbucks Corner there is John Doe.

Conclusion

Increasing sales requires more in-depth research that combines product, customer engagement, and how salesperson sell their products. Because different areas may have different promotion methods. We can see that John Doe is in the top position in the Starbucks Corner but not in the other store.



Thank you!



www.linkedin.com/in/nabila-ijtihatil-ismail/



nabila.ijtihatil@gmail.com