Data Warehouse Design Report

1. Introduction

The purpose of this Data Warehouse (DW) design is to integrate and organize data from operational sources into a structured analytical environment that supports business intelligence and decision-making. The design follows **dimensional modeling** principles, enabling fast, flexible, and user-friendly data analysis.

2. Methodology

The design process followed these steps:

1. **Requirement Analysis** – Identified key business processes, measures, and analytical needs.

2. **Data Source Review** – Assessed the transactional datasets to determine which entities and attributes to include in the DW.

3. **Schema Design** – Developed a **star schema** for efficient query performance and clear analytical structure.

3. Schema Overview

The **Star Schema** was selected because it:

- Simplifies analytical queries

- Supports OLAP operations such as slicing, dicing, roll-up, and drill-down

- Improves aggregation performance

Fact Table:

- **FactSales**

  o Measures: SalesAmount, Quantity, Profit

  o Keys: Foreign keys referencing dimension tables

1	sale_id	time_id	product_id	customer_id	store_id	quantity_sold	unit_price	sales_amount	discount	cost_amount	payment_method	transaction_id
2	1	20240101	1	1	1	2	500.00	1000.00	0.00	600.00	Credit Card	TXN001
3	2	20240215	2	2	1	1	1200.00	1200.00	50.00	800.00	Cash	TXN002
4	3	20240410	3	3	2	5	15.00	75.00	0.00	25.00	Mobile Money	TXN003
5	4	20240720	4	1	2	3	45.00	135.00	10.00	60.00	Debit Card	TXN004

Dimension Tables:

1. **DimProduct**

1	product_id	product_sku	product_name	category	sub_category	brand	unit	cost_price	list_price
2	1	ELEC001	Smartphone X	Electronics	Phones	TechBrand	piece	300.00	500.00
3	2	ELEC002	Laptop Pro	Electronics	Computers	CompTech	piece	800.00	1200.00
4	3	CLOT001	T-Shirt Cotton	Clothing	Tops	FashionCo	piece	5.00	15.00
5	4	CLOT002	Jeans Slim Fit	Clothing	Bottoms	DenimWorld	piece	20.00	45.00

2. **DimCustomer**

1	customer_id	first_name	last_name	gender	birth_date	age_group	email	phone	region	city
2	1	Alice	Johnson	Female	1990-05-14	25-34	alice@example.com	555-1234	Central	Nairobi
3	2	Bob	Smith	Male	1985-09-23	35-44	bob@example.com	555-5678	Coast	Mombasa
4	3	Carol	Adams	Female	2000-03-10	18-24	carol@example.com	555-8765	Rift Valley	Nakuru

3. **DimTime**

1	time_id	date	day	month	quarter	year	weekday	is_weekend
2	20240101	2024-01-01	1	1	1	2024	1	0
3	20240215	2024-02-15	15	2	1	2024	4	0
4	20240410	2024-04-10	10	4	2	2024	3	0
5	20240720	2024-07-20	20	7	3	2024	6	1

4. **DimStore**

| 1	store_id	store_name	region	city	store_type
2	1	Downtown Store	Central	Nairobi	Retail
3	2	Coastal Mall	Coast	Mombasa	Retail
4	3	Quick Mart	Nyanza	Kisumu	Retail
5	4	Kamakis	Rift	Eldoret	Retail

Why star schema?

Star schema denormalizes attributes into wide dimension tables for fast, simple joins against a single fact table — ideal for typical OLAP queries (aggregations, rollups, slicing). It reduces join complexity and improves query performance for reporting and BI (at the acceptable cost of some redundancy), so it's preferred over snowflake for read-heavy analytics workloads.

4. Advantages of the Design

- **Scalability** – Can easily incorporate additional data sources and dimensions.
- **Performance** – Star schema optimizes complex analytical queries.
- **Flexibility** – Supports multiple analytical methods with the same structure.
- **Data Consistency** – Provides a single, reliable source of truth.

5. Diagram

Generated by python (schema_diagram.ipynb)
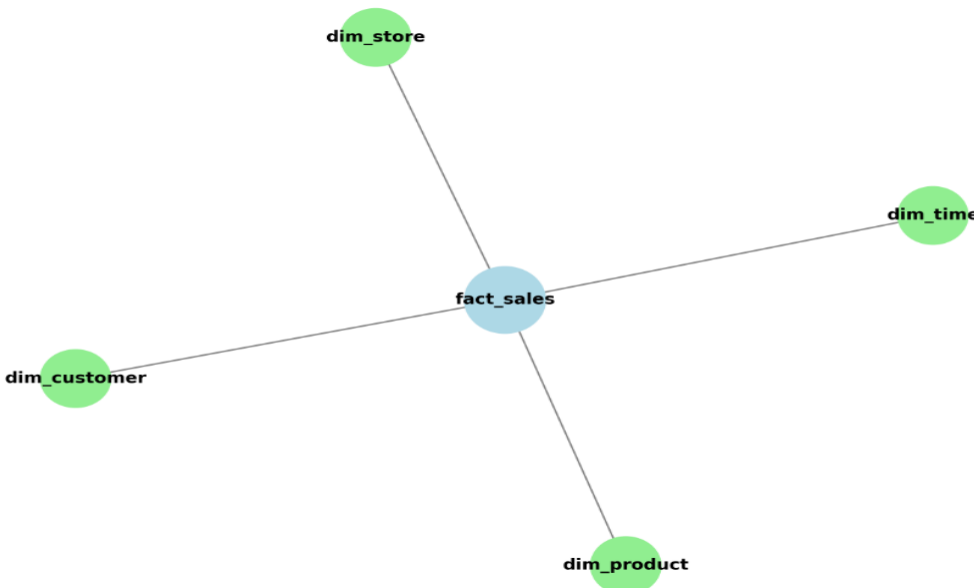
Drawn in Drawio
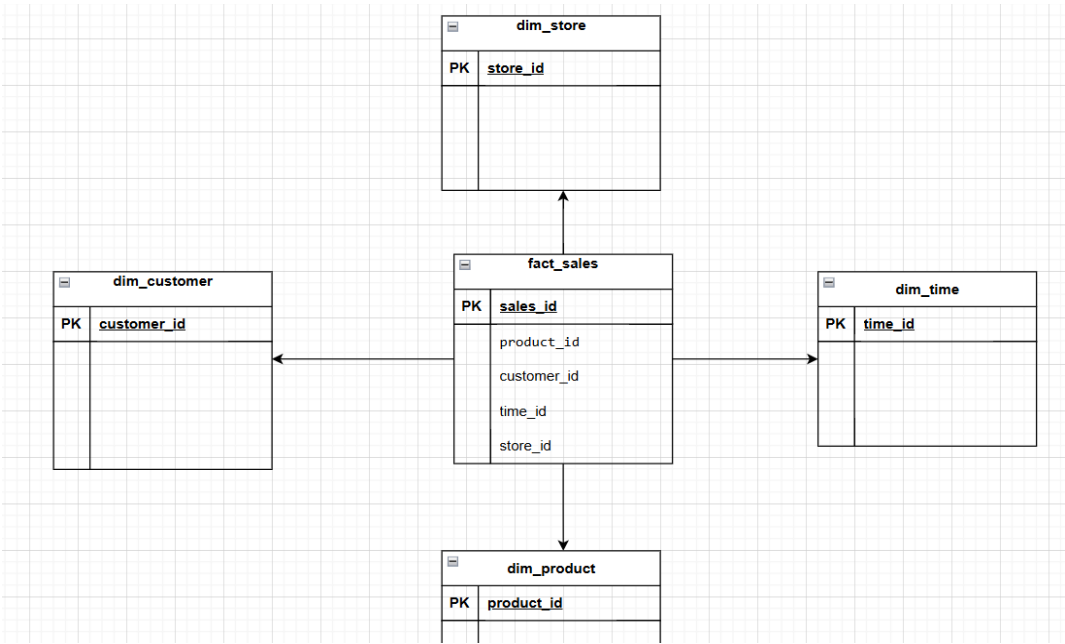
6. Conclusion

The proposed Data Warehouse design offers a structured, high-performance environment for analytical processing. Its dimensional modeling approach ensures that business users and analysts can derive insights quickly and reliably.