Data modeler, ETL architect and BI developer: Charmain Lebelo
Business driver: Farai Matyukira
Database administrator: Clarissa Meyer

Business Requirements

1. Which bike brand is most popular at specific times of the year?
2. Sales per day
3. Which store has most sales?
4. Which staff is responsible for most sales?

Business process

The sales of different bike brands at Bike Store shops.

Grain

A single row represents the total sales revenue and quantity sold for a specific brand, from a specific store, made by a specific staff member, on a specific day.

Dimensions

Dim_BikeBrands

- brand_id (PK)
- brand_name
- TimeOfYear (PK)
- total_sales

Dim_DailySales

- sales_date (PK)
- total_sales

Dim_StaffSales

- staff_id (PK)
- staff_name
- total_sales

Dim_StoreSales

- store_id (PK)
- store_name
- total_sales

Fact

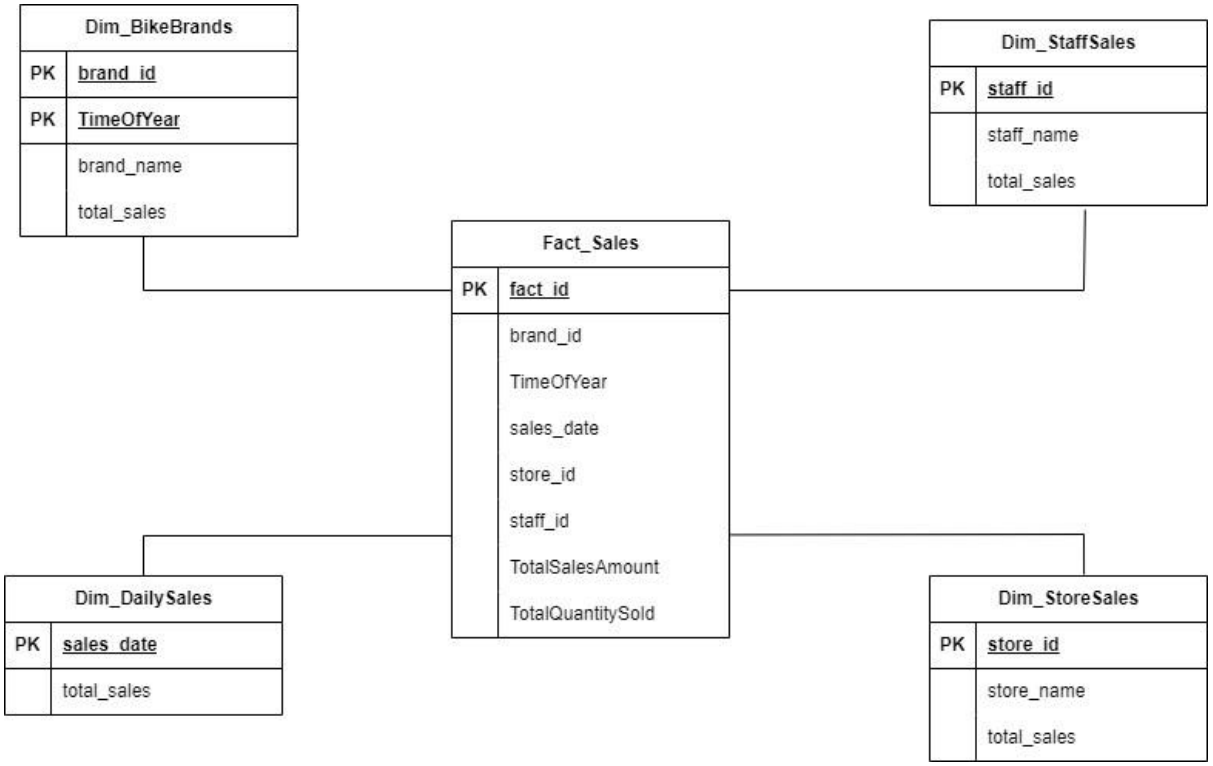
Fact_Sales

- fact_id (PK) • brand_id
- TimeOfYear
- sales_date
- store_id
- staff_id
- TotalSalesAmount • TotalQuantitySold

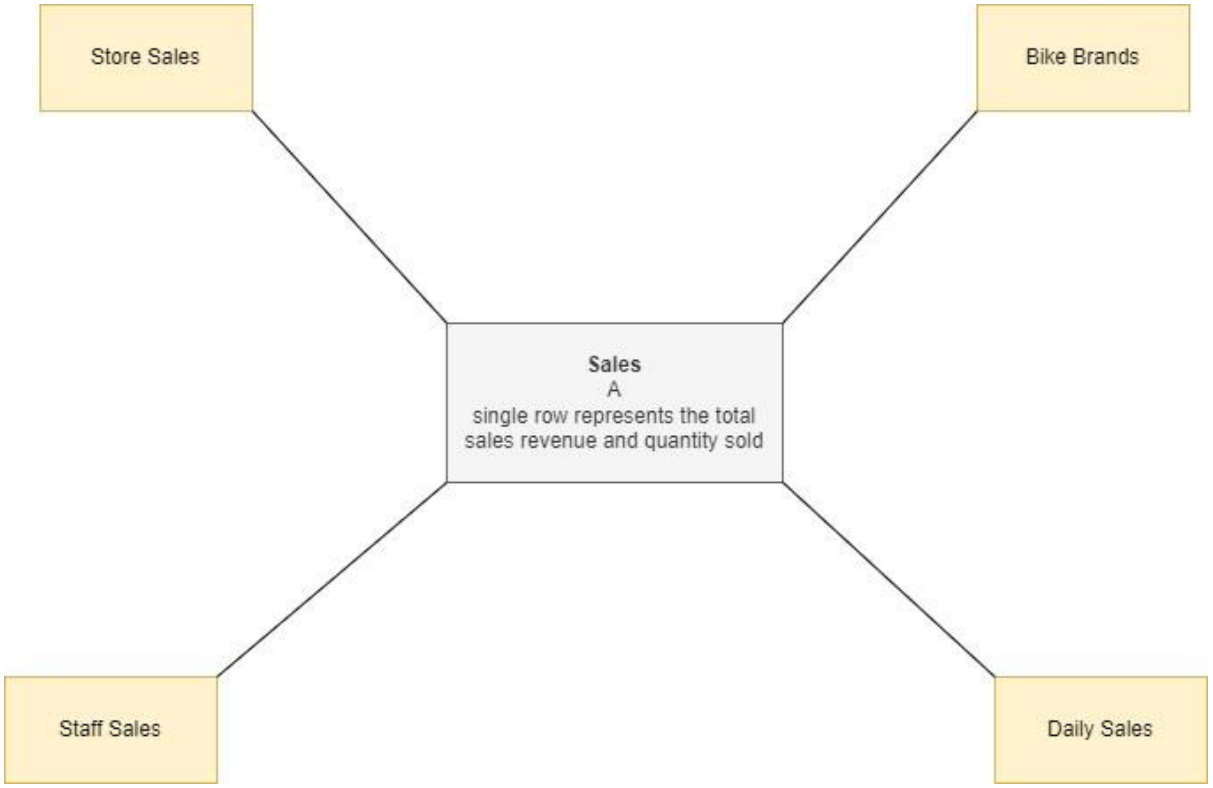
Matrix

			Base Dimension Name	Dim_BikeBrands	Dim_DailySales	Dim_StaffSales	Dim_StoreSales
			Dimension Type	SCD 1	SCD 0	SCD 2	SCD 2
Business Process	Fact Table	Fact Table Type	Granularity	Per brand, per time of the year	Daily	Per staff member	Per store
Sales Revenue	Fact_Sales	Transactional	X	X	X	X	X
Quantity Sold	Fact_Sales	Transactional	X	X	X	X	X

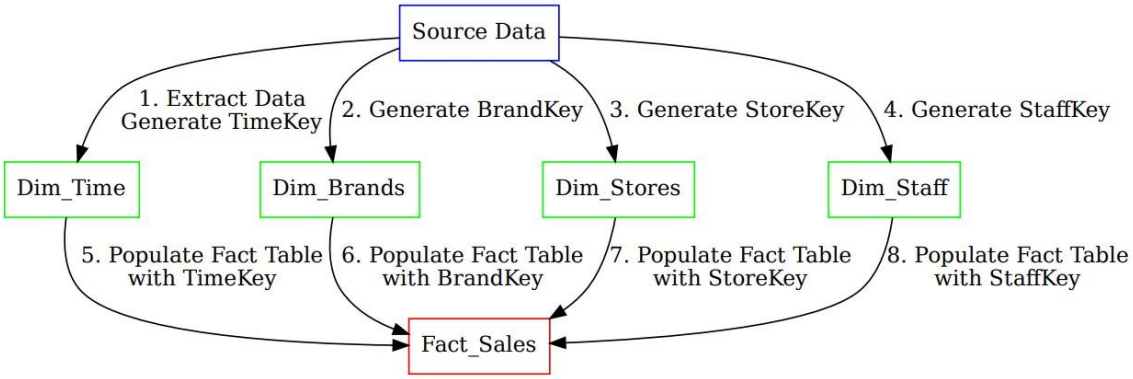
Star Schema



High-level model



Pipeline



- **Source Data**

  This is your original data source from where all the raw data is extracted.

- **Dim_Time**

  The raw data is used to generate a unique TimeKey for each day in the Dim_Time dimension table.

- **Dim_Brands**

  Similarly, unique BrandKey values are generated for each brand in the Dim_Brands dimension table.

- **Dim_Stores**

  Unique StoreKey values are generated for each store in the Dim_Stores dimension table.

- **Dim_Staff**

  Unique StaffKey values are generated for each staff member in the Dim_Staff dimension table.

- **Fact_Sales**

  Finally, the Fact_Sales table is populated. The surrogate keys from the dimension tables (TimeKey, BrandKey, StoreKey, StaffKey) and the facts (total sales amount and total quantity sold) from the source data are inserted into the Fact_Sales table, establishing the relations through foreign keys.

  The arrows represent the flow of data, with annotations describing each step in the process. This pipeline helps in maintaining data integrity and enables efficient data analysis in the data warehouse.

An error that may occur is that a Dimension table may not have a record corresponding to the Natural ID being looked up.