

Egypt University of Informatics

Computer and Information Systems

Big Data Engineering Course

**Technical Report**

**Data warehousing Design and Analytics Project**

Submitted by:

Nada Ashraf 22-101043

Aly Zaki 22-101096

Ahmed Waleed 22-101058

Omar Bayoumi 22-101022

Submission Date: 19th April 2025

**Introduction**

This project is a data warehousing-based project in which we implement concepts such as a dimensional modelling like a star schema to design and implement a data warehouse in which we can extract insightful information that will help in the strategic decision making of our company. This project is focused on designing a complete historical data warehouse in which we go through the processes of developing a data warehouse undergoing phases like ETL process, OLAP Cube for analytics, and finally presenting our insights by using Power BI to translate data into information.

**Project Overview**

This project focuses on the sales performance of a fictional bicycle retail company called BikeStores. The goal is to analyze the historical sales data to uncover trends and insights across time, store locations, product lines, etc. By conducting this, the company will be able to achieve a strategic decision that will aid it in future decision-making and development.

**Source Database Description**

Sample Database: BikeStores Database

Includes key tables such as:

* Orders, order items, products, categories, brands, customers, staff, and stores.
* It has two areas of focus: Sales and Production

**Source Database Entity-Relationship Diagram (ERD):**

A screenshot of a diagram

AI-generated content may be incorrect.

**Datawarehouse Business Area Focus:**

* Sales and Order

**Datawarehouse Business Goals and Objectives:**

* Analyze monthly sales trends to identify high-performing months.
* Determine which store generates the highest revenue.
* Identify top-selling bicycle products.
* Evaluate product categories by revenue and quantity sold.
* Identify top-performing salesperson staff based on total revenue.
* Identify which customers’ demographics generate the most revenue.

**4 – Step in Designing a DW:**

* **Subject**:
  + BikeStores Retail Sales
* **Grain**:
  + One row per product sold in a transaction per day per customer per staff per store.
  + For every product sold in a specific order, by a staff member, to a customer, in a store, on a specific day. (Transactional)
* **Dimensions**:
  + *Product Dimension:* Information related to products
  + *Store Dimension:* Information related to Stores.
  + *Date Dimension:* Hierarchy of Day, Month, and year.
  + *Customer Dimension:* Information related to Customer
  + *Category Dimension*: Information of product’s categories snow flaked from product dimension.
  + *Brand Dimension*: Information of product’s brand snow flaked from product dimension
  + *Staff Dimension*: Information related to salesperson that contributed to the sale.
* **Measures:** 
  + Quantity Sold
  + Sales Amount (unit price \* quantity \* (1-discount))
  + Gross amount
  + Discount

**Star Schema Design Visualization:**

A screenshot of a computer

AI-generated content may be incorrect.

**Methodology:**

This project follows a structured methodology aligned with the general data warehousing lifecycle. Incorporating ETL development, OLAP cube design, MDX querying, and Power BI dashboard presentation.

**Development Steps:**

1. Designing Data warehouse schema DDL
2. An ETL pipeline was developed using **SQL Server Integration Services (SSIS)**
3. OLAP Cube Development in **SQL Server Analysis Services (SSAS)**
4. **MDX** queries were developed in **SQL Server Management Studio (SSMS)**
5. **Power BI** Dashboard Creation from the cube developed.

**Data warehouse Schema in SSMS (DDL):**

A screenshot of a computer program

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer code

AI-generated content may be incorrect.

**SSIS ETL Process Documentation:**

**Tools Used**: Visual Studio 2019, SSIS

**Product Dimension ETL**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Customer Dimension ETL**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Store Dimension ETL**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Staff Dimension ETL**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Date Dimension ETL**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Date Dimension Explanation:**

Role Playing

* Date Dimension is executed in the form of role playing in which we had three views:
  + Order date
  + Required date
  + Shipped date
* We combined all these different dates into one physical date dimension and created a special date view for each one of the three views.
* Each view will appear in the fact table as a separate table with its own FK/PK referential integrity.

**Category and Brand Dimensions (Snow flaked from product)**

* These dimensions are snow flaked from the product to avoid redundancy since they are highly repetitive.

A screenshot of a computer

AI-generated content may be incorrect.

A diagram of a brand

AI-generated content may be incorrect.

**Sales Fact Table ETL**

A screenshot of a computer

AI-generated content may be incorrect.

**Fact Table Source SQL Command Line:**

A screenshot of a computer

AI-generated content may be incorrect.

**Role Playing View (Order Date) in Fact Table as Foreign Key:**

A screenshot of a computer

AI-generated content may be incorrect.

**Role Playing View (Required Date) in Fact Table as Foreign Key:**

A screenshot of a computer

AI-generated content may be incorrect.

**Role Playing View (Shipped Date) in Fact Table as Foreign Key:**

A screenshot of a computer

AI-generated content may be incorrect.

**Fact Table Destination Mappings:**

A screenshot of a computer

AI-generated content may be incorrect.

**Fact Table ETL Success Run:**

A screenshot of a computer

AI-generated content may be incorrect.

**Final ETL Process Success:**

A screenshot of a computer

AI-generated content may be incorrect.

**Verification of ETL Process**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**SSAS OLAP Cube Process Documentation:**

**Tools Used: SQL Server Analysis Services (SSAS)**

**OLAP Cube Setup:**

A screenshot of a computer program

AI-generated content may be incorrect.

**Product Dimension Attributes**

A screenshot of a computer

AI-generated content may be incorrect.

**Customer Dimension Attributes**

A screenshot of a computer

AI-generated content may be incorrect.

**Staff Dimension Attributes**

A screenshot of a computer

AI-generated content may be incorrect.

**Store Dimension Attributes**

A screenshot of a computer

AI-generated content may be incorrect.

**Date Dimension Attributes (Hierarchies)**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Finalized OLAP Cube:**

A screenshot of a computer screen

AI-generated content may be incorrect.

**Deployment Successful**

A screenshot of a computer

AI-generated content may be incorrect.

**OLAP Cube Successful Run:**

A screenshot of a computer

AI-generated content may be incorrect.

**OLAP Cube Successful Verification on SSMS:**

A screenshot of a computer

AI-generated content may be incorrect.

**MDX Queries on the OLAP Cube Analysis:**

* For each business objective, an MDX query is generated to aid in the analysis process.
* MDX queries are used on the OLAP Cube to generate multidimensional queries

**1. Monthly Sales Trends**

A screenshot of a computer

AI-generated content may be incorrect.

**2. Yearly Sales Trends**

A screenshot of a computer

AI-generated content may be incorrect.

**3. Top Sold Products generating revenue by Quantity Sold**

A screenshot of a computer

AI-generated content may be incorrect.

**4. Top Categories generating revenue by sales amount and quantity sold.**

A screenshot of a computer

AI-generated content may be incorrect.

**5. Top Brands generating revenue by sales amount and quantity sold**

A screenshot of a computer

AI-generated content may be incorrect.

**6. Top Stores generating revenue by sales amount.**

A screenshot of a computer

AI-generated content may be incorrect.

**7. Top performing salespersons that generate the highest revenue**

A screenshot of a computer

AI-generated content may be incorrect.

**8. Highest 3 states that generate revenue based on customer demographics**

A screenshot of a computer

AI-generated content may be incorrect.

**9. Top 5 Products in each brand that generates highest revenue**

A screenshot of a computer

AI-generated content may be incorrect.

**Power BI Visualization tool for Business Intelligence**

**Connection:** Live connection to Sales\_Analysis\_Cube from SSAS

**1. Sales amount by Year and Month:**

A graph showing a line

AI-generated content may be incorrect.

**Insights**:

* Consistent sales until a significant peak around **April 2018**, followed by a drastic drop.
* Crucial for identifying seasonal trends and demands

**2. Sales amount by Category Name:**

A graph of blue squares

AI-generated content may be incorrect.

**Insights**:

* **Mountain Bikes** lead in generating revenue, followed by **Road Bikes** and **Cruisers Bicycles.**
* Indicates high-end or specialized categories of bikes yield better profitability.

**3. Quantity Sold by Brand Name:**

A graph of blue squares

AI-generated content may be incorrect.

**Insights:**

* **Electra** and **Trek** are leading in the highest selling brands.
* Lower performing brands like Strider and Ritchey may need extra promotional support and evaluation

**4. Quantity Sold by Product Name**

A graph with blue lines

AI-generated content may be incorrect.

**Insights:**

* Top Selling products include **Electra Cruiser 1 (24-inch**) and **Electra Townie Original 7D EQ**
* Products under Electra brand dominate the list of products sold.
* Help in identifying which models are popular for future stockings and marketing campaigns.

**5. Sales Amount by Store Name**

A blue and orange pie chart

AI-generated content may be incorrect.

**Insights**:

* **Baldwin Bikes** accounts for 67.83% of the total sales.
* Indicates a strong and highly successful location
* May indicate where to add extra stores in which locations in a strategic expanding.

**6. Quantity Sold by State**

A blue and orange pie chart

AI-generated content may be incorrect.

**Insights:**

* **NY** contributes with the highest volume of sales, followed by **CA**, and lastly **TX**.

**7. Sales Amount by Salespersons (Staff)**

A graph with blue bars

AI-generated content may be incorrect.

Insights:

* **Marcelene Boyer** and **Venita Daniel** are the top performers in sales.
* Can be used for sales incentives, performance reviews, or assigning key territories.

**Powe BI Dashboard:**

A screenshot of a graph

AI-generated content may be incorrect.

**Conclusion**

This project successfully demonstrates the full life cycle of a business intelligence solution, starting from the extraction of transactional and operational data from the BikeStores database, through the transformation and loading of these data into a structured data warehouse star schema using SSIS. Followed by the creation of an analytical OLAP Cube via SSAS that finally aids with the analytical querying with MDX and visualization of insights by Power BI. This project showcases how data warehousing and OLAP technologies, combined with modern visualization tools, can significantly enhance sales performance analysis, strategic planning, and business decision-making.

**Workload Distribution**

|  |  |
| --- | --- |
| Data warehouse Initial Star Schema SQL DDL | All team members contributed |
| ETL Process | Nada Ashraf and Aly Zaki |
| Testing ETL Process | Ahmed Waleed and Omar Bayoumi |
| OLAP Cube | Ahmed Waleed and Omar Bayoumi |
| Testing OLAP Cube | Nada Ashraf and Aly Zaki |
| MDX Queries | 2 Queries per member |
| Power BI Visualization Charts | 2 Charts per member |
| Final Technical Report and PPT | All team members contributed |

**Resources:**

Sample Database:   
<https://github.com/sebsto/babelfish-demo/blob/main/sql/BikeStores%20Sample%20Database%20-%20create%20objects.sql>

**Tools:**

Microsoft SQL Server 2022: DDL Data warehouse

Visual Studio SSIS: ETL Process

Visual Studio SSAS: OLAP Cube

Microsoft SQL Server 2022: MDX Queries

Power BI: Visualizations and Charts