Data Mart

BI System Specifications Document

Date: 01.07.2024
Version: 1.00
Written By: Tomer

Western Digital

1. General

1.1 Project Goals

The purpose of this project is to provide a full BI Solution creation form Priority ERP Database for Western Digital stores around the world.

Western Digital is known for designing, developing, manufacturing and marketing storage products. The products sale throw retails stores and online by Amazon.

The solution will encompass summarized data tables, with a focus on sales in Western Digital stores alongside information on the customers and products sales over time.

The Bi solution includes dashboards and reports to assist management, department heads, and sales managers gaining insight into performance of products, stores and customers to get better strategic decisions.

1.2 Project Content

This project will contain Data mart with information about sales data.

1. Data cleaning and preparation: prior to analysis, we will perform data cleaning and preparation to ensure their quality and consistency.
2. Main summary tables to be built for the company need:

   Fact_Sales – Information about all the generated invoices for customers. Data loading will be incremental.
   Dim_Products - information about company products.
   Dim_Customers - information about the company customers.
   Dim_Stores - information about the company retails and online.

   History Management Table:

   TransferTable – Log table for information about all the updates of the tables.
   Dim_Products_History - information about tracking the historical changes in products data over time. Using Slowly Changing Dimensions (SCD) Type 4.

Dim_Customers_History - information about tracking the historical changes in customers data over time. Using Slowly Changing Dimensions (SCD) Type 4.

The ERD of tables shown in attached link: ERD

Source To Target document in this link: S2T

3. The reports will include data visualization that will support the project goals in the following way:

   - **Sales Analysis:**

     The sales analysis will contain data about the sales by date, product and store. The reports will help to identify sales trends of seasonality and products. spot top performance products and stores, understanding the different behavior between online store to retail store.

   - **Customers Analysis:**

     The customers analysis will contain data about customers by date, store, country, product, and category. The report will help to understand the customer behavior, like where they shop, if they returned customers.

   - **Executive Dashboard:**

     The dashboard will contain key visuals from the last two reports. The dashboard will show a wider perspective on the data.

2. Gantt

Gantt Link

3. Technical Specification

3.1 Prerequisites

SQL Server – ERP system in the operational DB (PriorityERP)
SSIS – ETL processes using SSIS in Visual Studio
Power BI – Creating dashboards using Power BI

3.2 Solution Architecture

3.2.1 High Level Design:

The ETL process, which includes arranging the data into Data Mart will be performed in SQL Server using SSIS. After Data Mart creation, report will be created using Power BI.

3.2.2 Power BI Reports:

3.2.2.1 The report for the sales department will consist of:

- Total revenue
- Average monthly revenue
- Total orders
- Average revenue per order
- Revenue / Order / Unit by store type (online vs retails) and by year, quarter and month
- Top 5 stores by Revenue / Order / Unit
- Revenue / Order / Unit by category
- Top 5 products by Revenue / Order / Unit

3.2.2.2 The report for customer department will consist of:

- Total number of customers
- Average revenue per customer
- Average orders per customer
- Revenue per customer by year, quarter and month
- Customers grouped by number of orders
- Avg number of orders and revenue per customer by country
- Top 5 selling products by number of customers

3.2.2.3 The executive dashboard will consist of:

- MTD Revenue
- MTD orders
- MTD unit
- LY vs YTD revenue
- Revenue YoY growth by year, quarter and month

- Revenue MoM growth by year, quarter and month
- Top country by Revenue / Order / Unit
- Top state by Revenue / Order / Unit
- Top store by Revenue / Order / Unit
- Top product by Revenue / Order / Unit

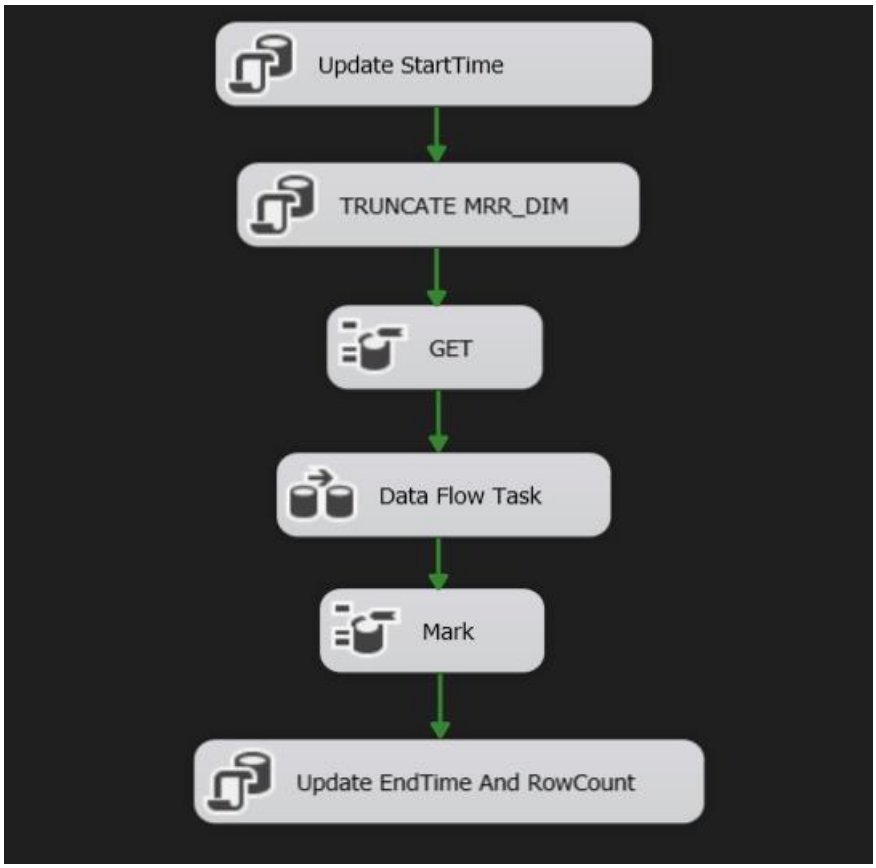
4. Functional Specification

4.1 ETL processes

1. MRR_Dim package:

Mirror tables are truncate using stored procedure and data is loaded from PriorityERP database to MRR tables.

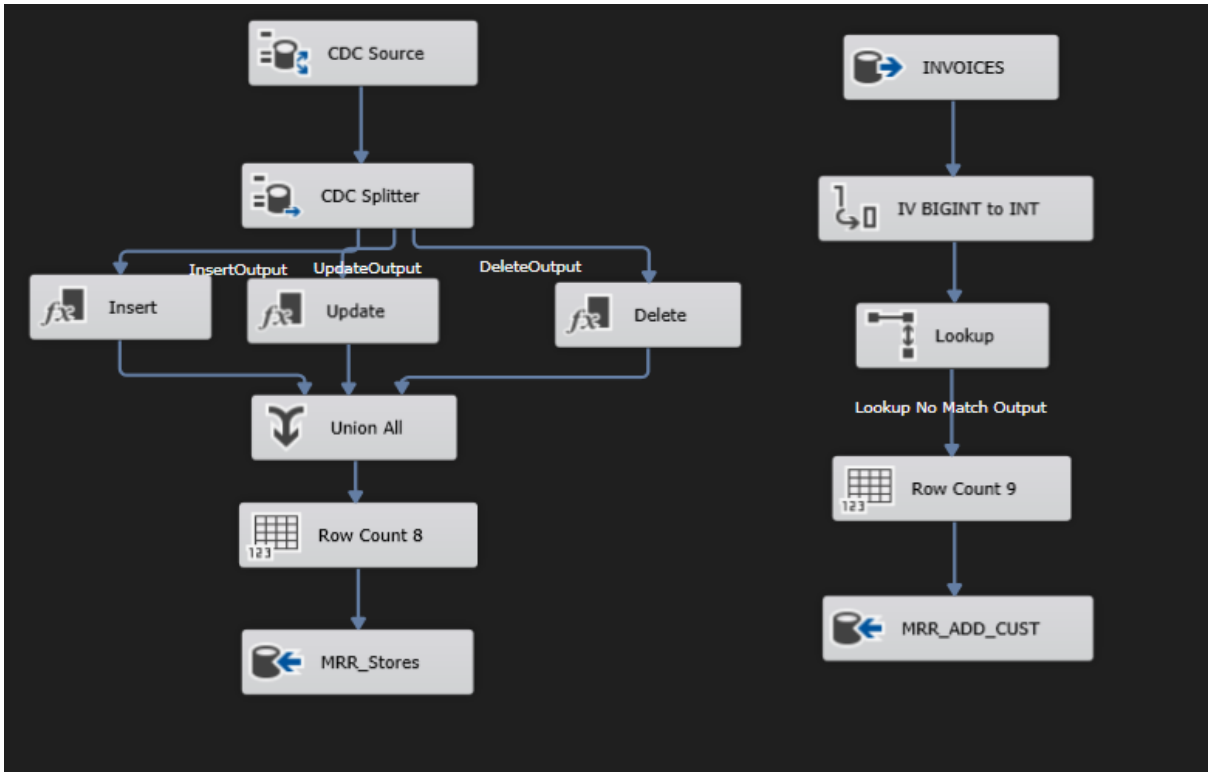
In the Control Flow:



Update StartTime - insert new row into transfer table with the name of the package and the current time and date.
Update EndTime And RowCount - is Updating the inserted row in transfer table with the number of rows changed in the data flow and the end time.
GET – Read from table cdc_states the timestamp of the last time data was update in the Stores table.
Mark – Update the timestamp for the next CDC.
Data Flow Task – Transfer OLTP table to MRR tables.

In the Data Flow:

Row Count - Count the rows that copied from OLTP tables to MRR tables.

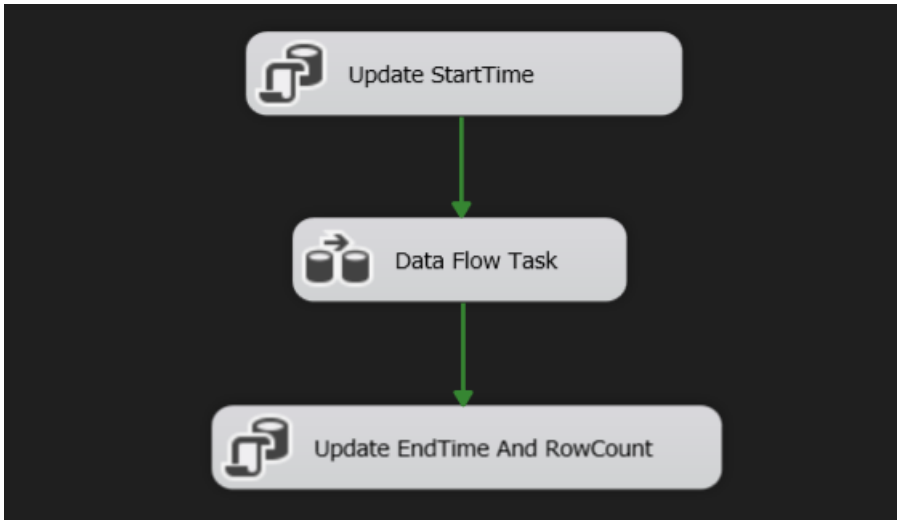


Because of the CDC Process on Stores Table in the PriorityERP database, CDC Source read the added columns and CDC splitter split it to derived columns insert with Status = 1, update with Status = 2, delete with Status = 3.

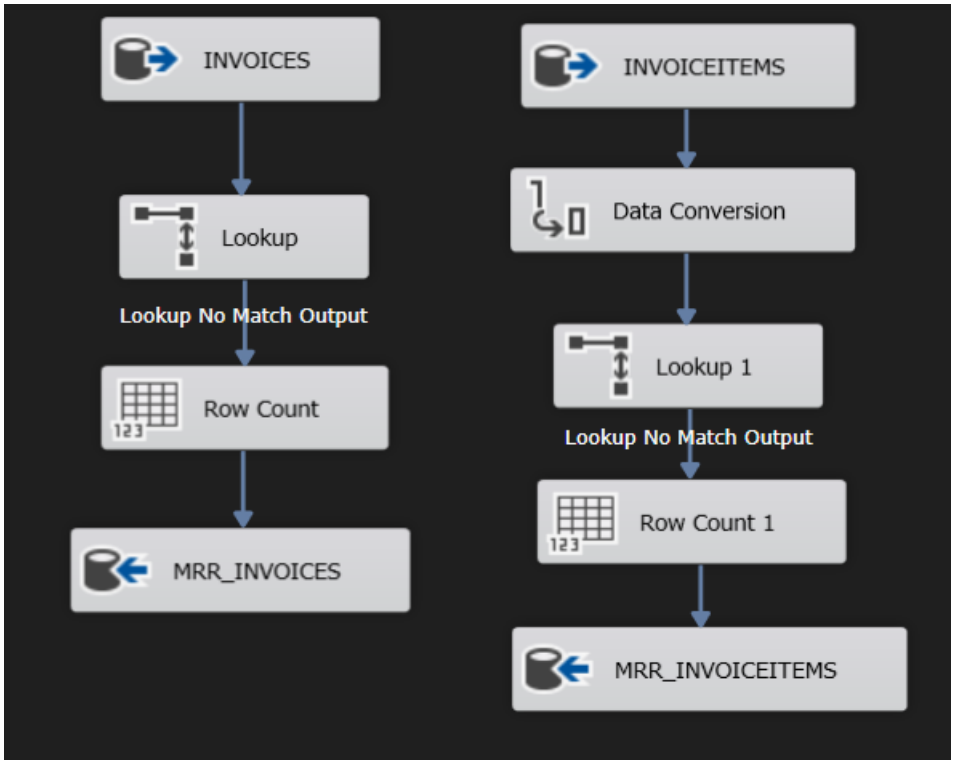
Lookup compare the CustomerID on INVOICES with CustomerID on CUSTOMERS Table and add the not match to MRR_ADD_CUST.

2. MRR_Sales package:

In the Control Flow:



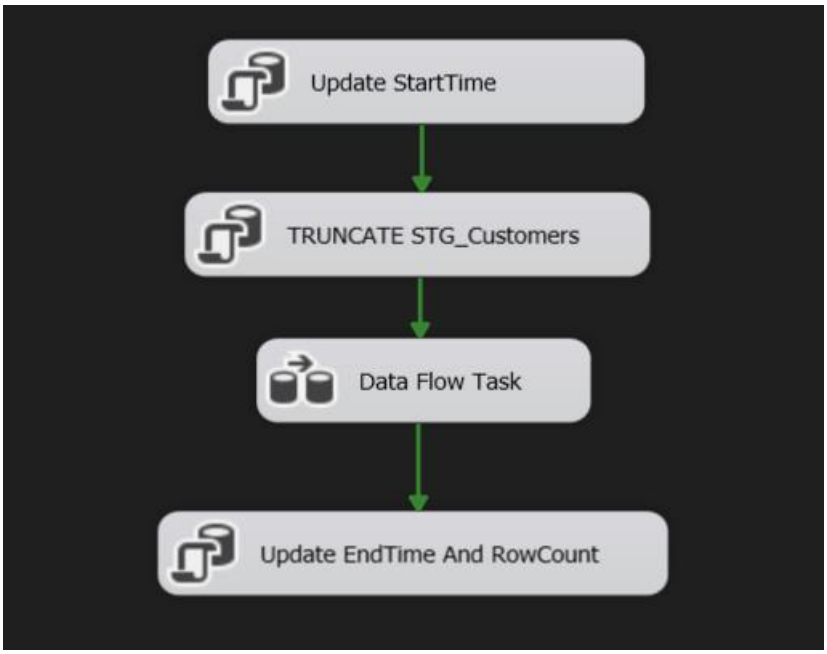
In the Data Flow:



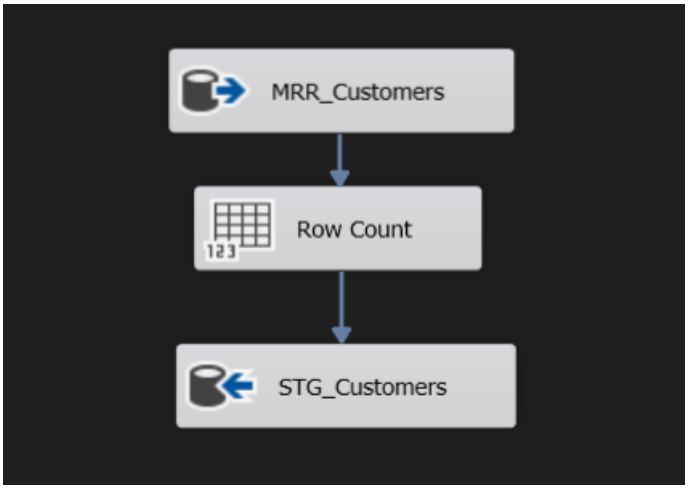
Data Conversion – convert the SaleID from the OLTP to the type of FACT_Sales.
Lookup - Compare OLTP INVOICES table with FACT_Sales by SaleID and adding only the not match Invoices.
The same for INVOICEITEMS.

3. **STG_Customers**

In the Control Flow:



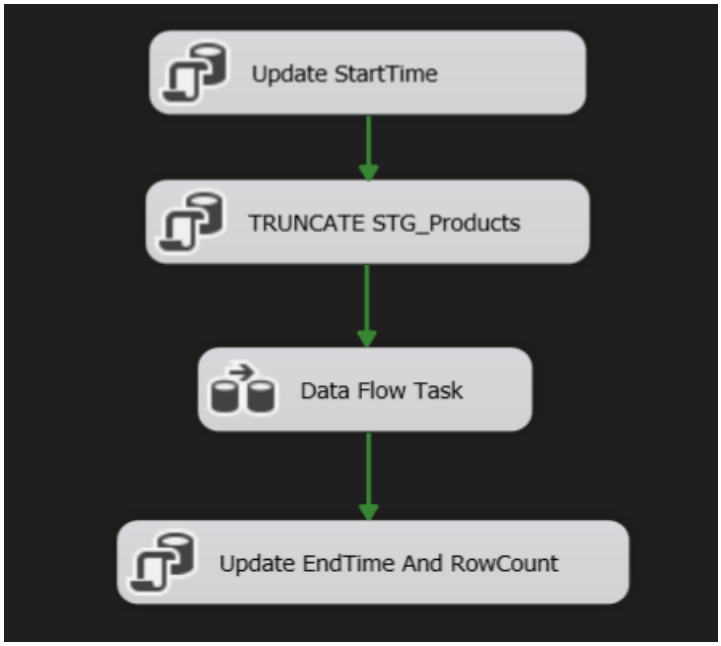
In the Data Flow:



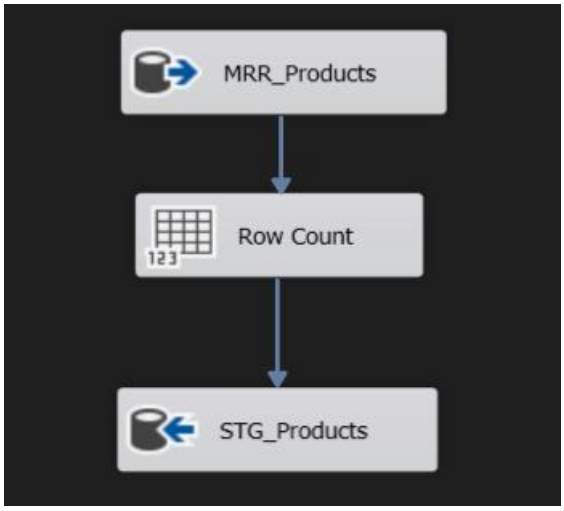
MRR_Customers is a SQL command that JOIN all the MRR tables with columns for the DIM_Customers, and also Union, and JOIN for MRR_ADD_CUST.

4. **STG_Products**

In the Control Flow:



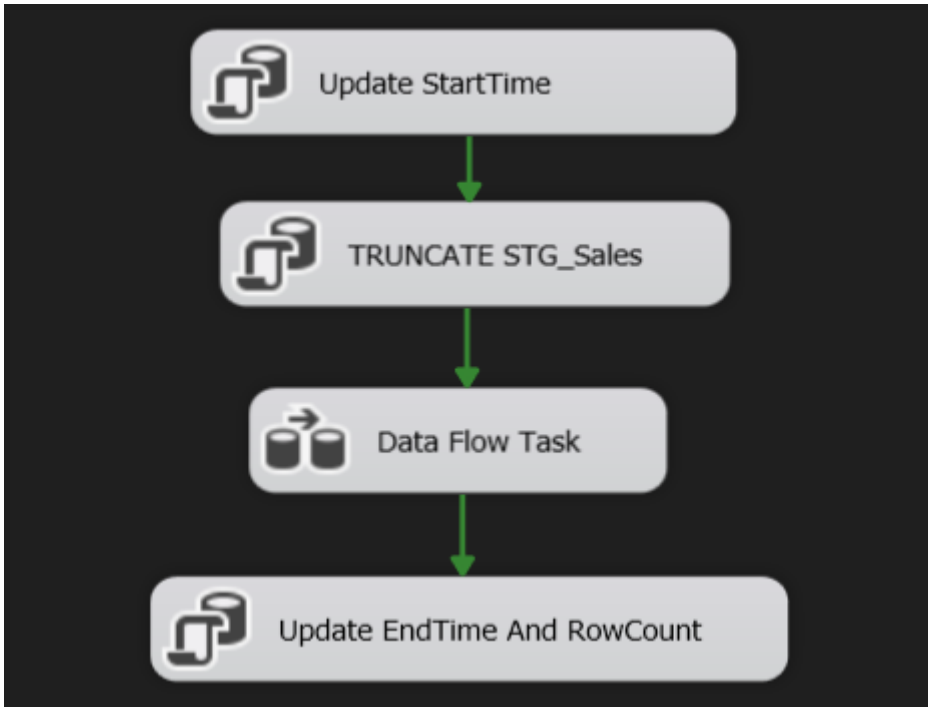
In the Data Flow:



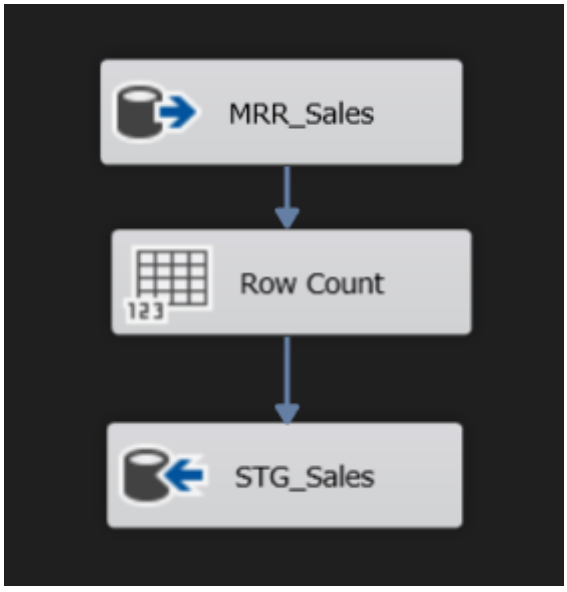
MRR_Products is a SQL command that JOIN all the MRR tables with columns for the DIM_Products.

5. **STG_Sales**

In the Control Flow:



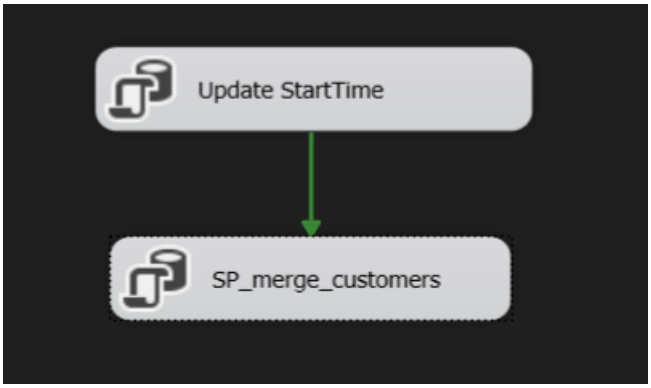
In the Data Flow:



MRR_Sales is a SQL command that JOIN all the MRR tables with columns for the FACT_Sales.

6. **DIM_Customers**

In the Control Flow:

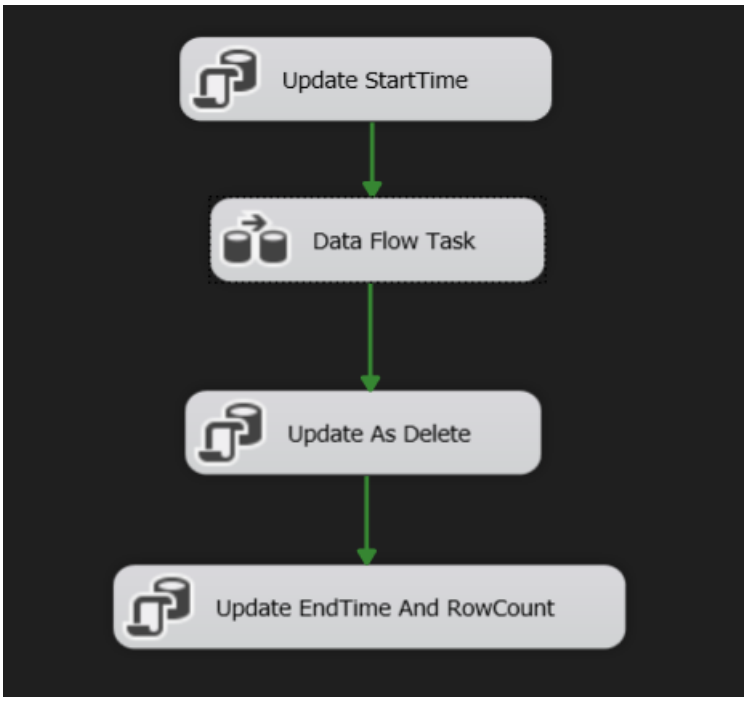


SP_merge_customers – is stored procedure that make conversion to data types of the STG_Customers to the data types of the DIM_Customers, and merge them.
When Match and the fields are not the same, we update.
When not match by Target we insert.
When not match by Source we update is active columns to 0.

Also, there is update to the Transfer_Table with @@ROWCOUNT of the rows that was affected.

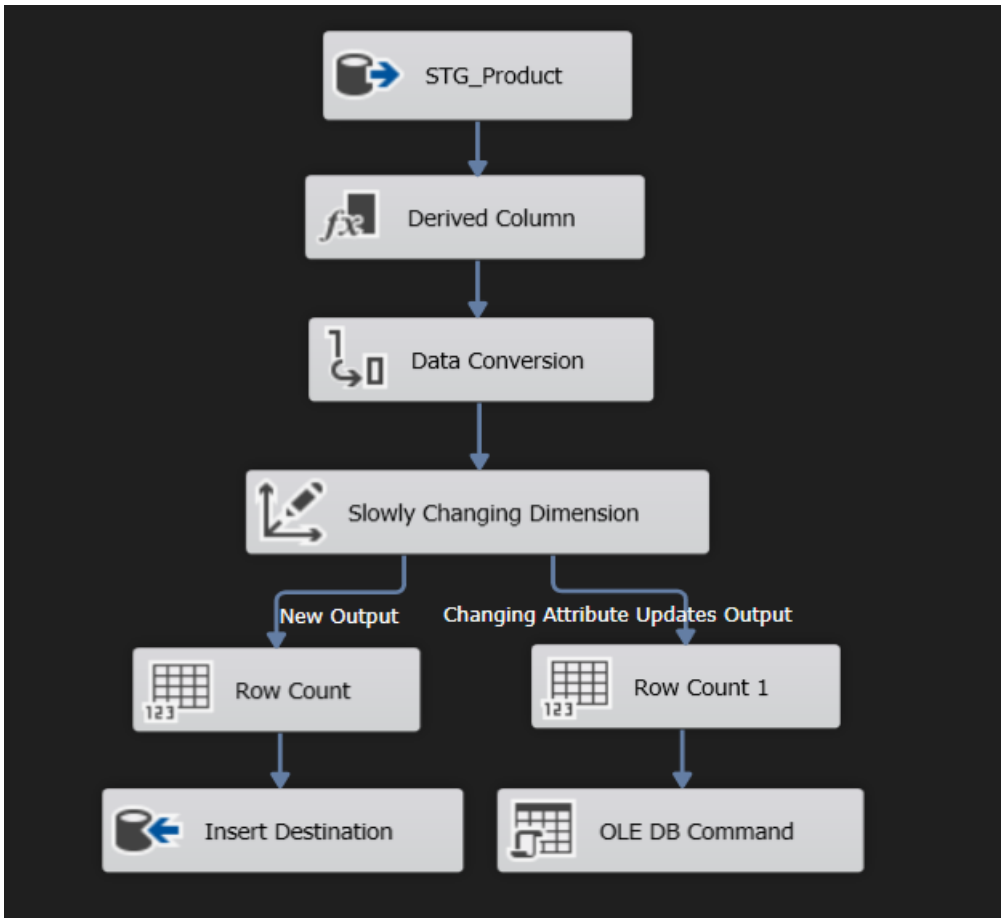
7. DIM_Product

In the Control Flow:



Update As Delete – Contain stored procedure that update delete products in the OLTP database as isActive = 0 in the DIM_Products.

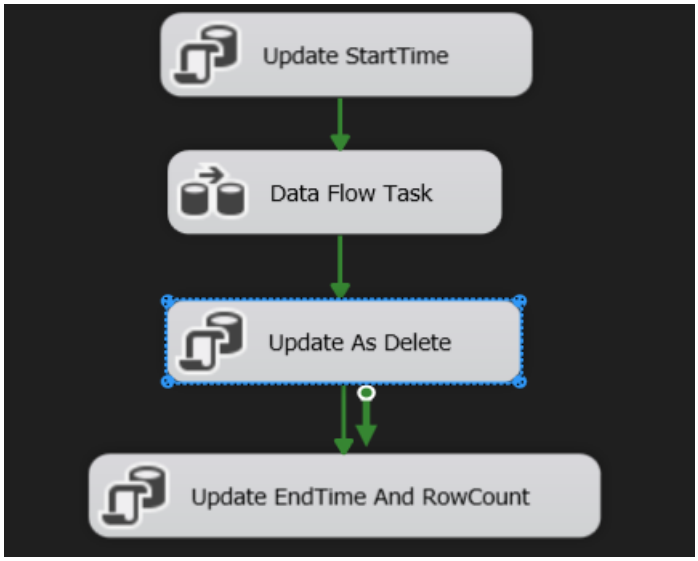
In the Data flow:



Derived column - Add isActive column with Default True value.
Data conversion – Convert the data types to the data type of the DIM_Products table.
Slowly Changing Dimensions – Track changes in data update if there is a change and insert if there is a new record.

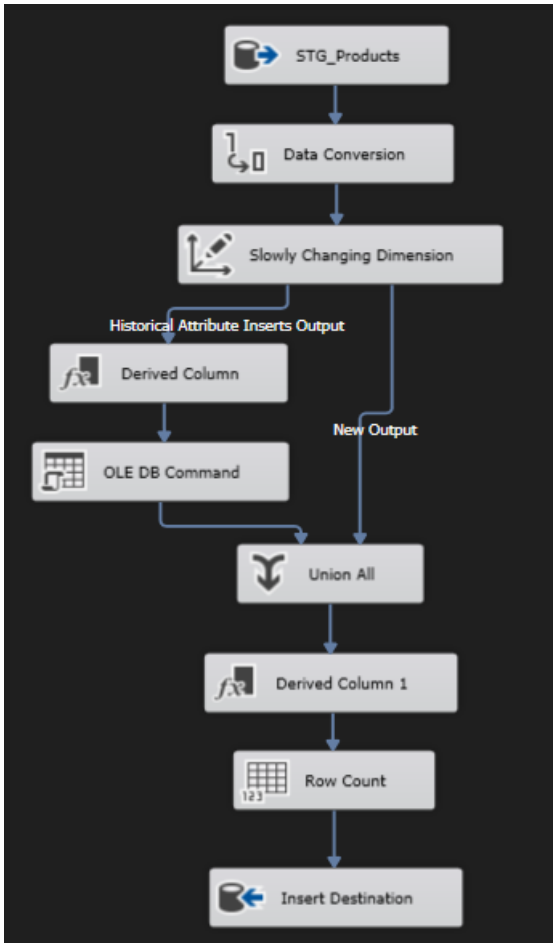
8. DIM_Products_History

In the Control Flow:



Update As Delete – Contain stored procedure that update EndTime with timestamp to products that was deleted in the OLTP database.

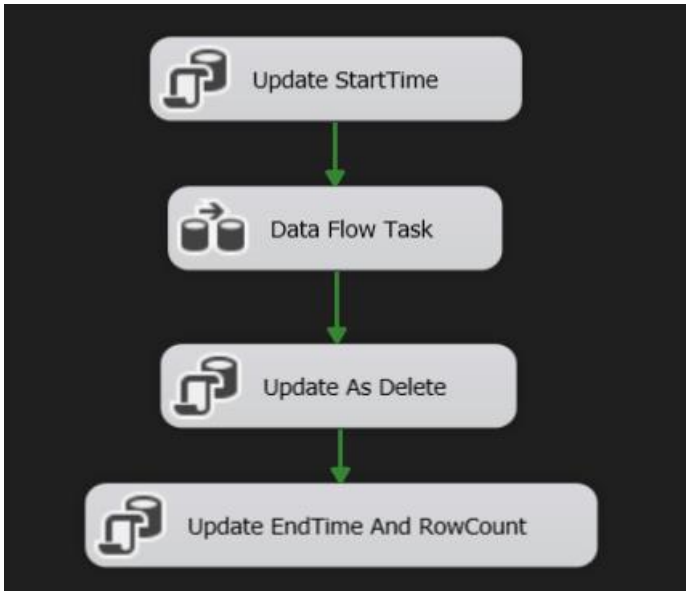
In the Data Flow:



Slowly Changing Dimensions - Track changes in data, every row get StartDate with timestamp and if there is change, the SCD add new row to the DIM_Products_History and update the EndDate with timestamp for the changed row.

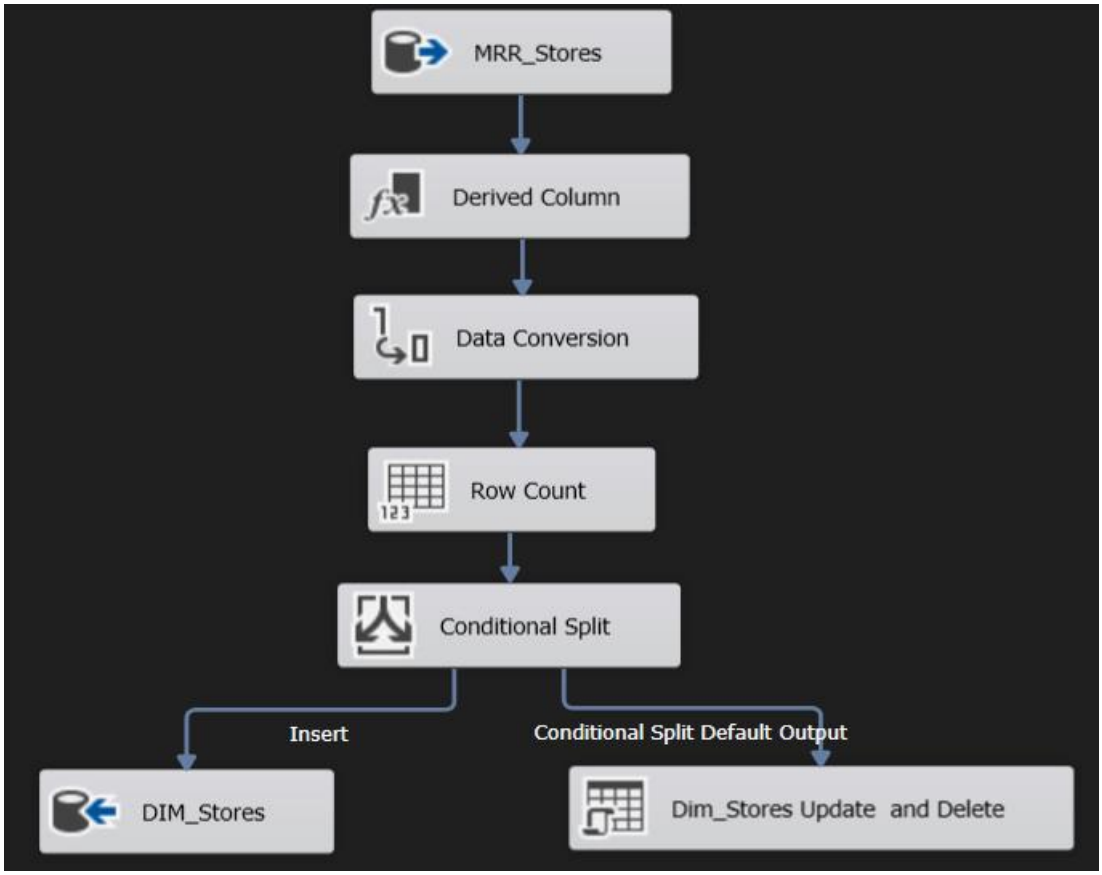
9. DIM_Stores

In the Control Flow:



Update As Delete – Contain stored procedure that update isActive = 0 if StoreID from DIM_Stores do not exist in MRR_Stores.

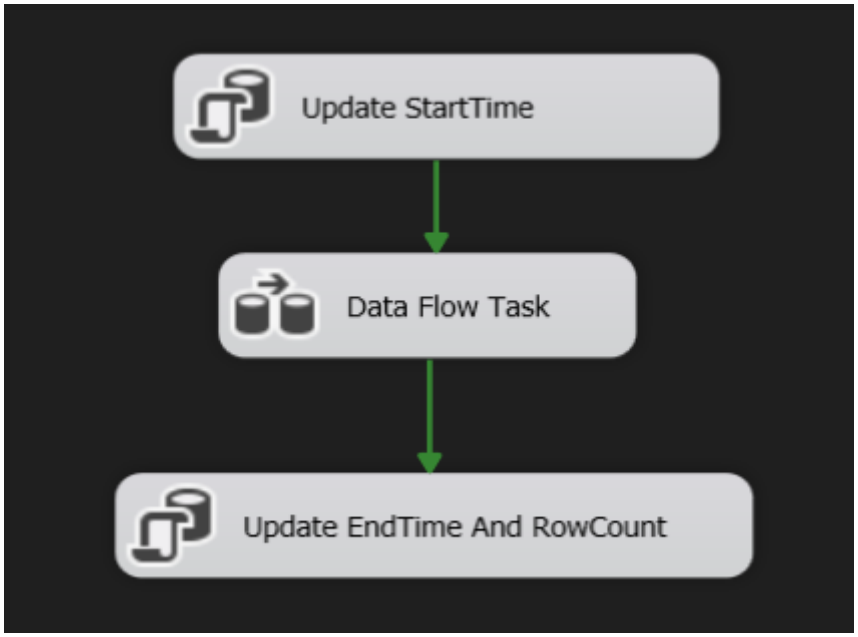
In the Data Flow:



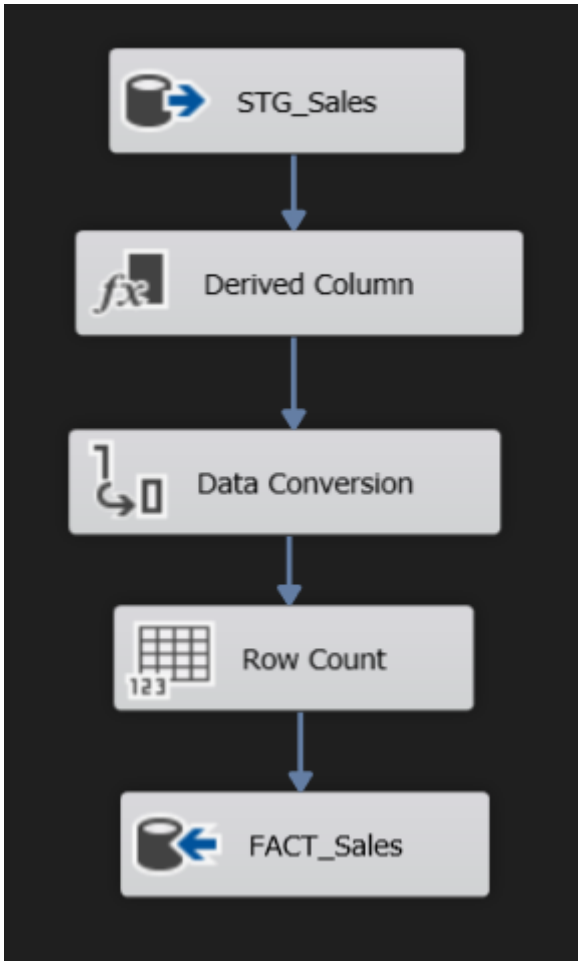
Derived column - Add isActive column with Default True value.
Data conversion – Convert the data types to the data type of the DIM_Stores table.
Conditinal Split – Check if the CDC Status Column is equal to 1
If equal insert the record to DIM_Stores table.
If not equal go to stored procedure that merge DIM_Stores with MRR_Strores when match update the changes.

10. FACT_Sales

In the Control Flow:



In the Data Flow:



Derived column - Add calculated columns of SubTotal, Tax and Total .

Data conversion – Convert the data types to the data type of the FACT_Sales table.

11. CopyDB_To_Production

In the Control Flow:

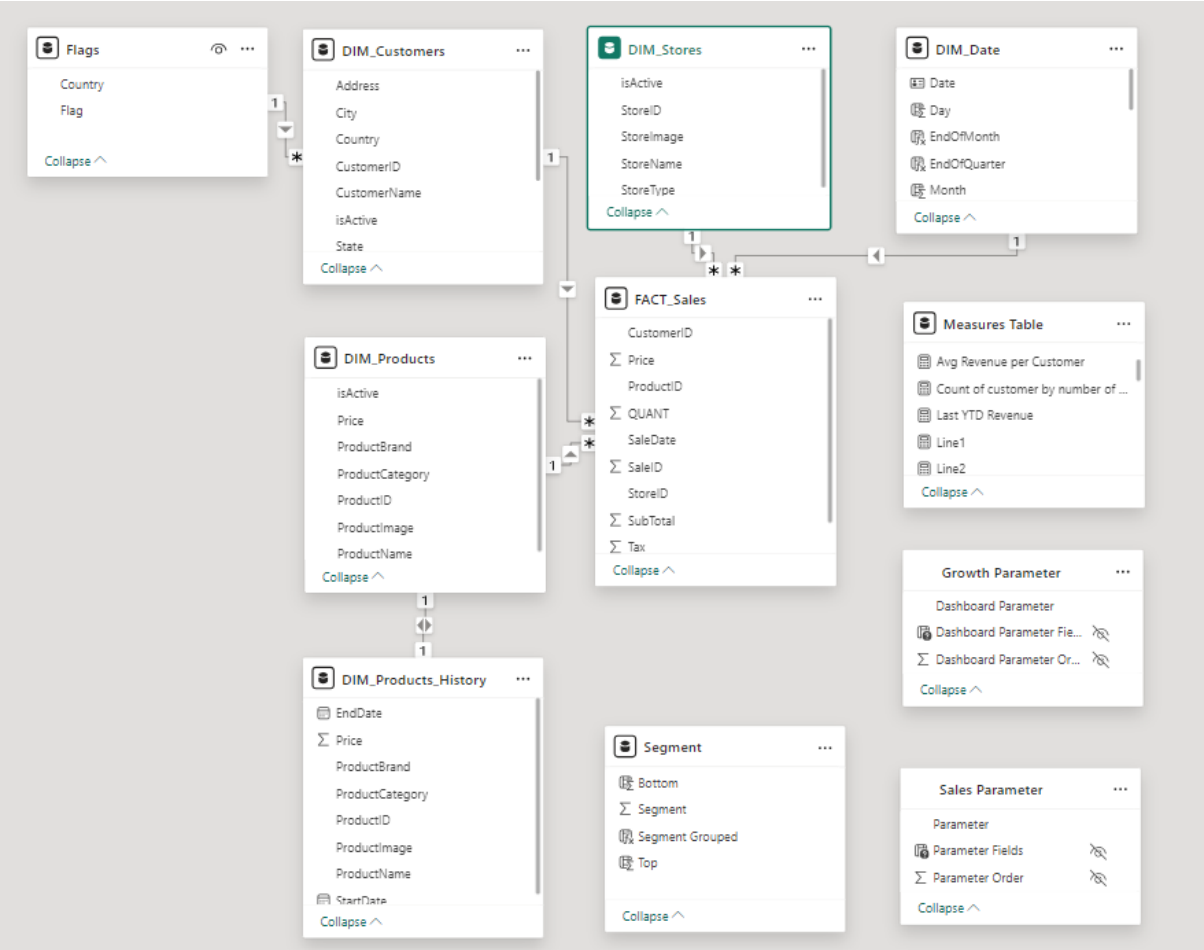
Copy Dev to Prod Task – Drop the production database and then copy the Dev database to production.

12. Deploy To Production

Deploy all packages to production

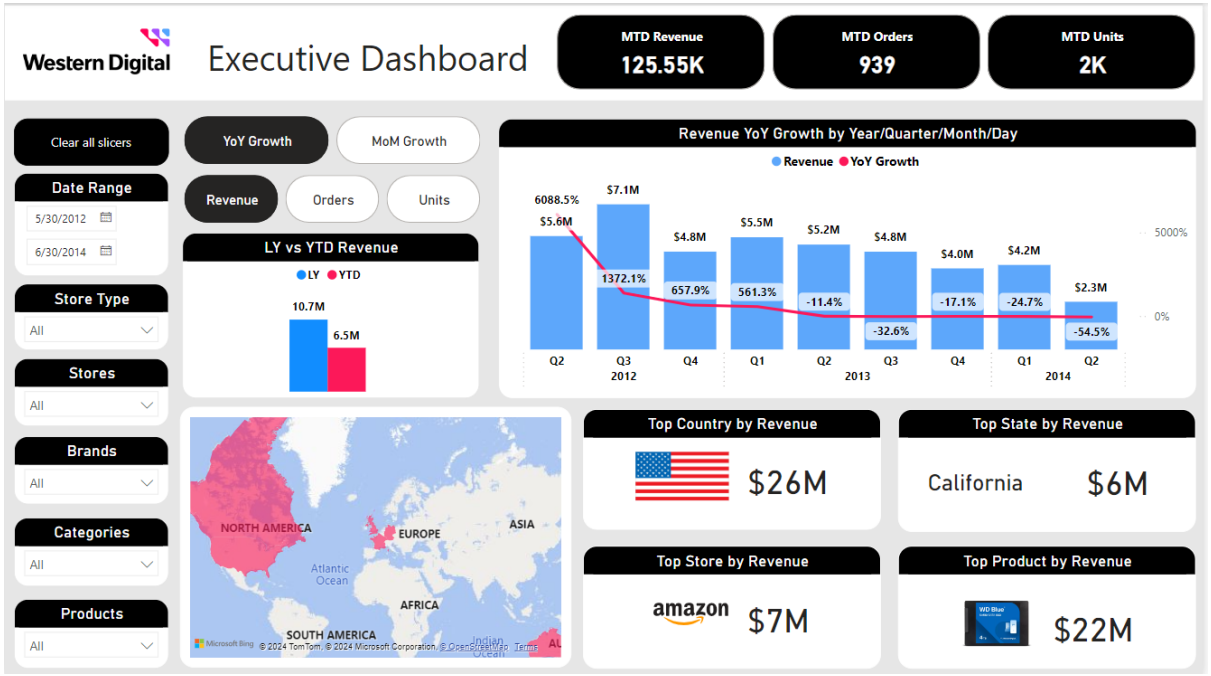
4.2 Visualization in Power BI

4.2.1 The reports created using Power BI Desktop the model includes Dim Stores, Dim Customers, Dim Product, Dim Products History and Fact Sales.
Dim Date - for showing visuals over time.
Dim Flags - flag image URL for the countries in Dim Customers
Segment Table – create to divided customer by number of orders
Sales Parameters – contains parameter for Revenue / Orders / Unit
Growth Parameters – contains parameter for YoY growth and MoM growth
Measure Table – contain all the measure that was created



4.2.2 Executive Dashboard

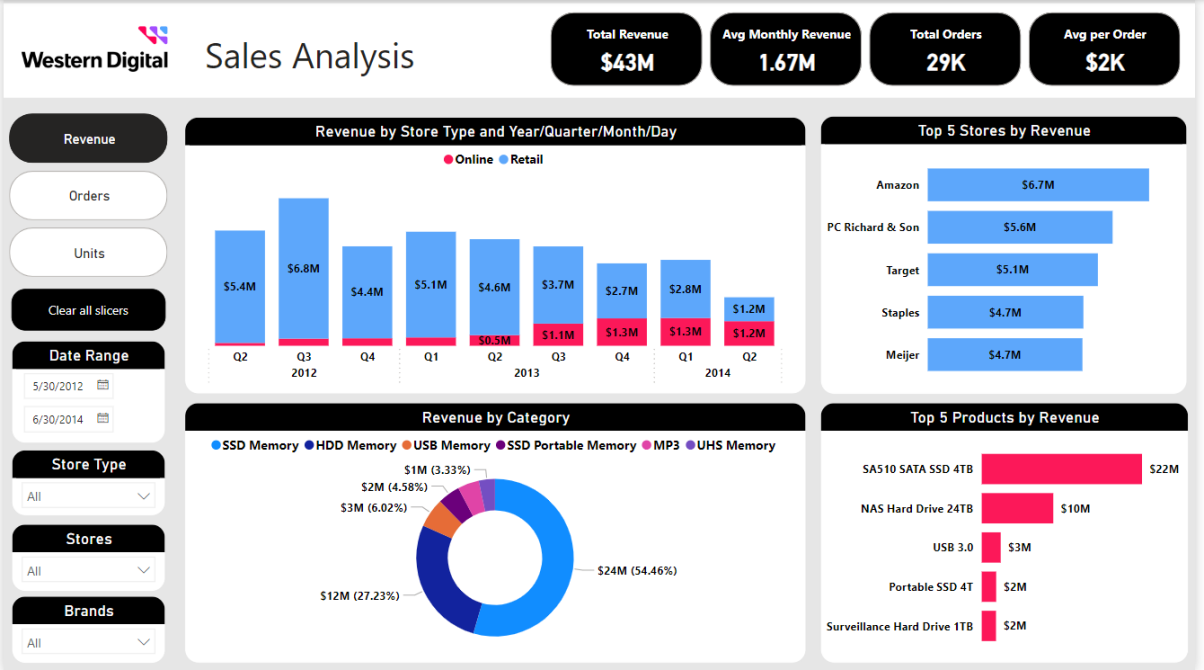
This report is for a broader look at the company's performance.



KPI cards - MTD by Revenue / Orders / Units show the performance of the current month.
LY vs YTD – Compare last year with the current year by Revenue / Orders / Units.
Growth by Year / Month / Day – Show in the bars the growth over time Revenue / Orders / Units, in the line show year vs last year / month vs last month by Revenue / Orders / Units.
Map – Help to drill down from countries throw states to cities for filtering the data.
Top cards – Show the top performing country, state, store and product by Revenue / Orders / Units.

4.2.3 Sales Analysis

This report is focused on the sale aspect of the company it aims to understand the sales performance.

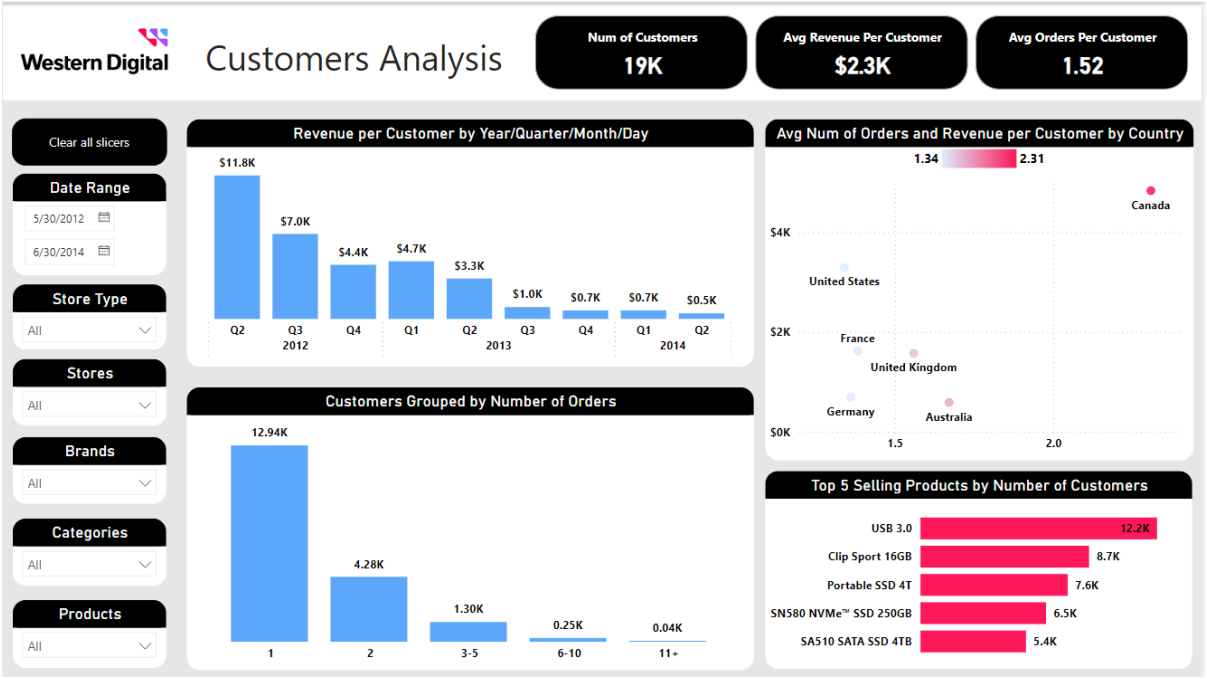


KPI cards – show the total revenue, average monthly revenue, number of total orders and average revenue per order.
Revenue / Orders / Units by Store Type and Year / Month / Day - show the change over time of revenue / orders / units in online vs retails stores.
Revenue / Orders / Units by Category - show the size in percentage and revenue / orders / units of the products categories.
Top 5 Stores – show the 5 top performing stores by revenue /orders / units.
Top 5 Products – show the 5 top performing products by revenue /orders / units.

4.2.4 Sales Analysis

This report is focused on the customers aspect of the company it aims to understand the customer behavior.

Western Digital



KPI cards – show total number of customers, average revenue per customer and average orders per customer.
Revenue per Customer by Year / Month / Day - show the change over time of the average revenue generated by customer.
Avg Num of Orders and Revenue per Customer by Country – show the average revenue per customer by country on the y axis, and the average numbers of orders per customer by country on the X axis.
Customers Grouped by Number of Orders – show the number of customers grouped by number of orders they made.
Top 5 Selling Products by Number of Customers – show top selling product by the amount of unique customer who bought them.