BI System Specification Document

Date: 22/01/2022

Version: 1.00

Fares Saffouri

Content

1. General

    1.1. Project Objective (including information consumers)

    1.2. Project Contents

2. Preparation of a work plan, distribution of tasks and schedules for execution (Gantt)

3. Technical Specification

    3.1. Prerequisites - A comprehensive list of systems including access methods.

    3.2. Solution Architecture and Flow Chart (HLD)

4. Functional Specification:

A detailed end-to-end description of each process in the project, including screenshots of key stages.

    4.1. Creation of final Source to Target and ERD models.

    4.2. Detailed description of all ETL processes.

    4.3. Description of data tables in the Data Mart, including history tables.

    4.4. Specification of measures, tables, and filters for reports in Power BI

1. General

1.1. Project Objective

The purpose of the project is to provide a Full-Scale BI Solution Creation from PriorityERP Database for HP company.

HP develops and provides a wide variety of hardware components, as well as software and related services to consumers.

This project aims to establish a comprehensive BI solution leveraging data from the PriorityERP system for HP. The solution will encompass summarized data tables, with a focus on sales data, alongside customer information, product details, stores, dates.

Furthermore, the project will facilitate the visual representation of data through dashboards and a reporting system designed to serve the company's sales managers effectively.

Hewlett Packard, or HP, is an American company founded more than 80 years ago that manufactures desktop computers, laptops, monitors, printers, and accessories. Since the 1990s, it has been the world's largest PC manufacturer.

By examining customer data, we could assist HP in better understanding customer preferences, purchasing behaviors. This could lead to more targeted marketing campaigns. Additionally, knowing which products the best-selling, and their categories are allowing for optimizing production quantities for each model.

1.2. Project Contents

In this project, we will build a Data Mart that will contain information about sales data.

1. Data Cleaning and Preparation: Prior to analysis, we will need to perform thorough data cleaning and preparation to ensure their quality and consistency.

2. Main summary tables to be built for the company's needs

• FactSales – Information about all the orders, which product in which quantity. Data loading process for this table will be incremental.

• DimProducts – Information about the products sold by the company. Additionally, a history table (DimProductsHistory) will be included to track changes in products over time using Slowly Changing Dimensions (SCD) Type 2,4.

• DimCustomers – Information about the company's customers

• DimEmployees – Information about the company's employees.

• DimStores – Information about the company's stores.

• DimDate – A table of dates for data analysis over time.(Added in the power BI).

3. The project will contain measures that will contribute to the achievement of the project's goal:

• Sales Department:

The Sales Department will focus on sales-related data, seller performance, sales by countries, sales type, and compare the results with the previous year to gain a better understanding of goal achievement.

• Customer Department:

The company's customers will be interested in data related to their purchases and overall experience. They will have access to information such as which customers purchased the most, total units sold by country, number of customers of each country and more.

• Dashboard:

The company's dashboard will include a high level of details about the company, including sales difference by month and year, and the growth number of customers by month.

2. Gantt

❖ Gantt link: gantt

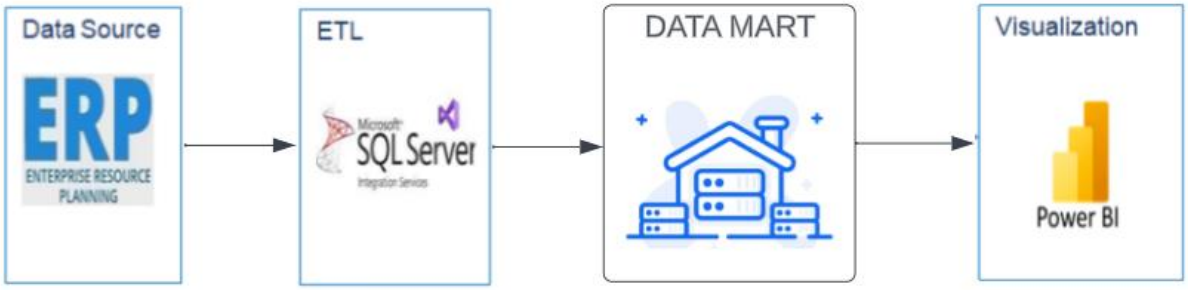
3. Technical Specification

3.1. Prerequisites

- SQL Server: ERP system in the operational DB (PriorityERP).
- SSIS: ETL processes using SSIS in Visual Studio
- Data refresh processes through the definition of JOBS in SSMS
- Power BI: Creating reports and dashboards using Power BI

3.2. Solution Architecture

HLD:



Data collection and exploration from the ERP system will be performed in SQL Server. The data will undergo an ETL process for organization and arrangement into a Data Warehouse using SSIS. Finally, the presentation of measures in reports and visuals will be presented in Power BI.

The report for the Customer Department consists of:

- The top 3 selling by product category.
- The total units sold by Store.
- Number of customers in each Country.
- Top 10 Customers by sales.
- Total Sales by year.
- Average Sale per customer.
- Total orders and total sold units by customers.

The report for the Sales Department consists of:

- Total sales by country.
- percentage of sales type.
- Top 10 salesperson by sales.
- Percentage of sales difference by year.
- Percentage of sale price category by total orders.
- Average sale for order.
- Average units for order.

The Dashboard consists of:

- Percentage of Orders Sales Difference by Year.
- Percentage of Units Sales Difference by Year.
- Percentage growth Of Customers by Month.
- Percentage of Sales Difference by Month.

4. Functional Specification

4.1. Creation of final Source to Target and ERD models.

4.1.1 Source to Target

- ❖ Source To Target link: S2T

The data tables that will be used from the OLTP database consist of a total of 11 tables as seen in the S2T.

4.1.2 ERD model

The ERD of the tables, attached link: ERD

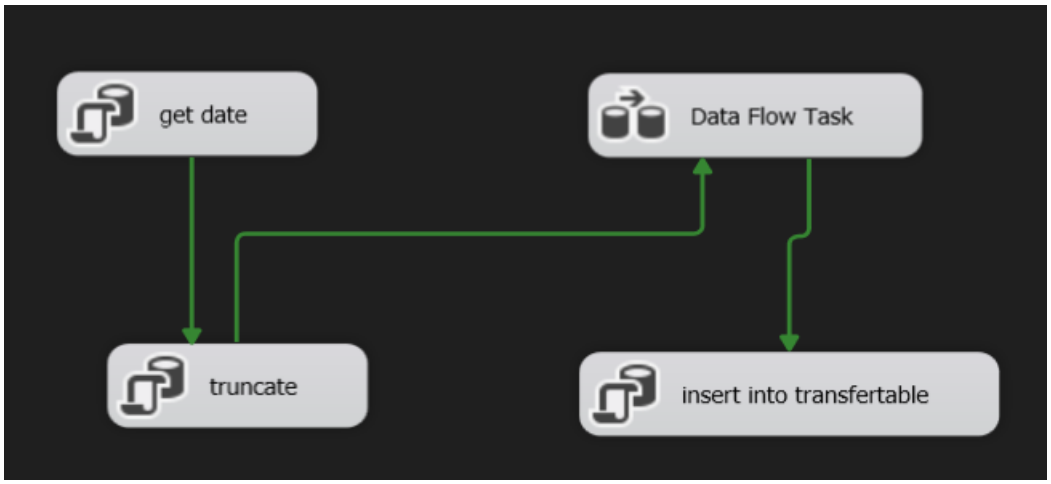
4.2. Detailed description of all ETL processes.

In SSIS I have 6 solutions:

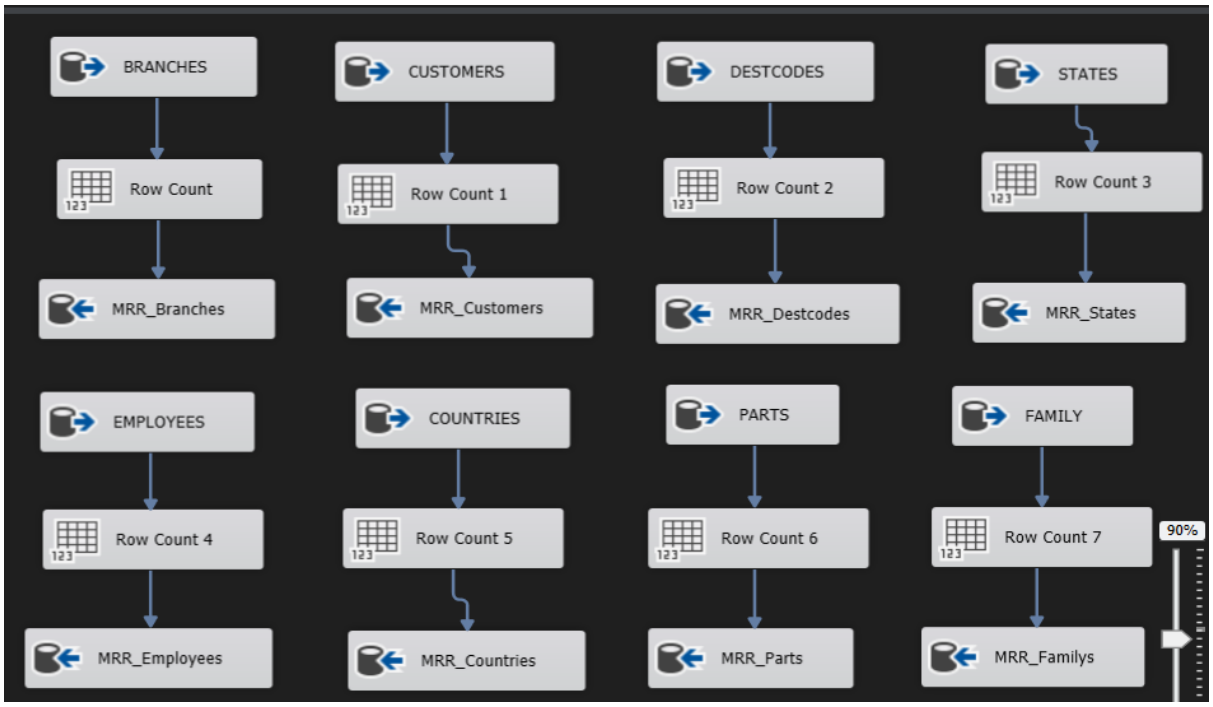
Mrr: in this solution I have 3 packages.

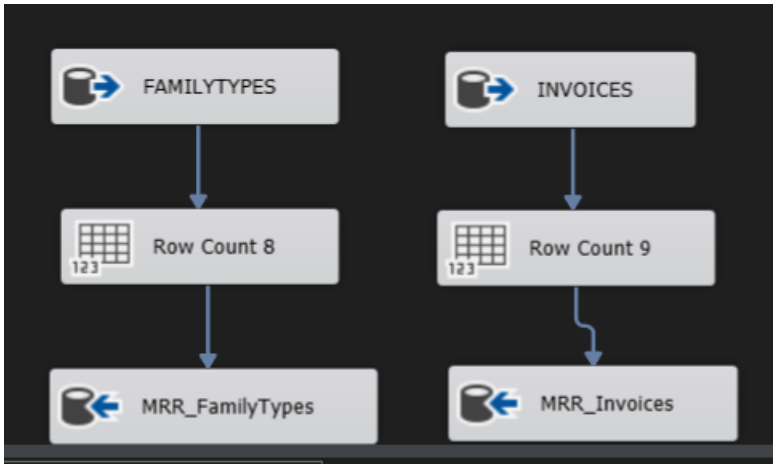
MrrDims: this package for getting relevant columns from different tables to create each dim table.

In the control flow:



In the data flow:

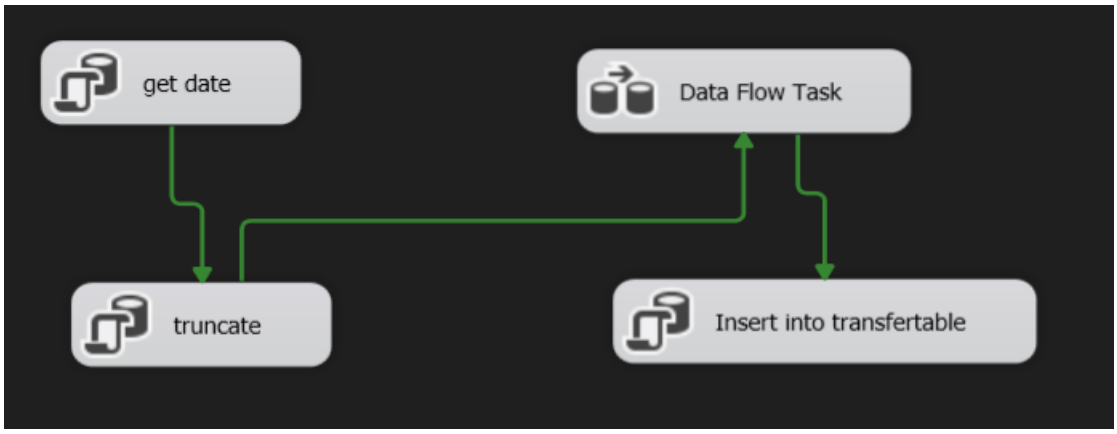




MrrSales:

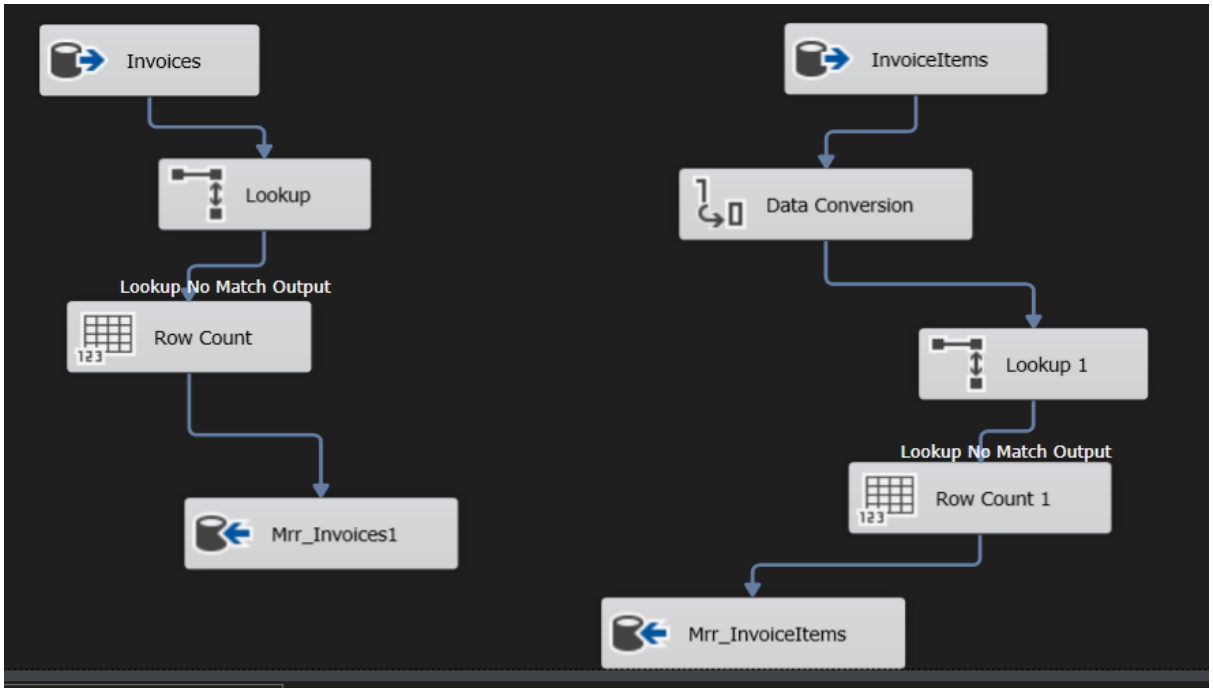
This package for getting columns from the two sales tables to create the fact sales table.

In the control flow:



In the data flow:

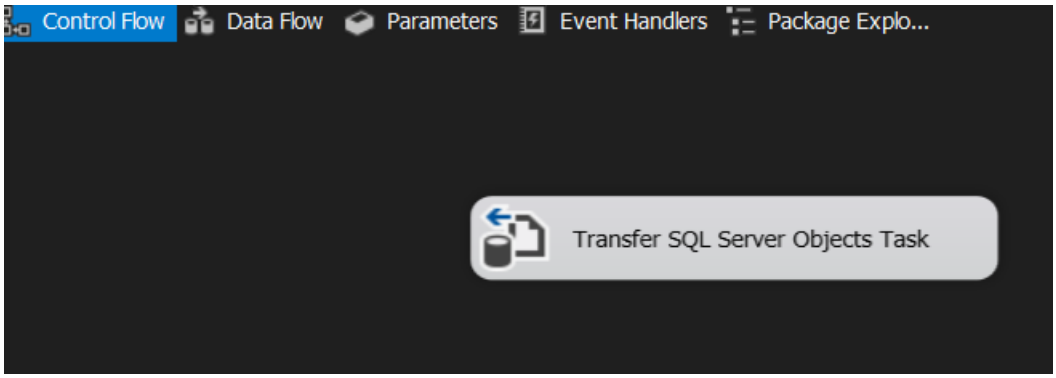
The use of lookup for incremental load:



Transfer db package:

This package is for copying all the data from the DB to another DB (Developing) after the creation of the data mart.

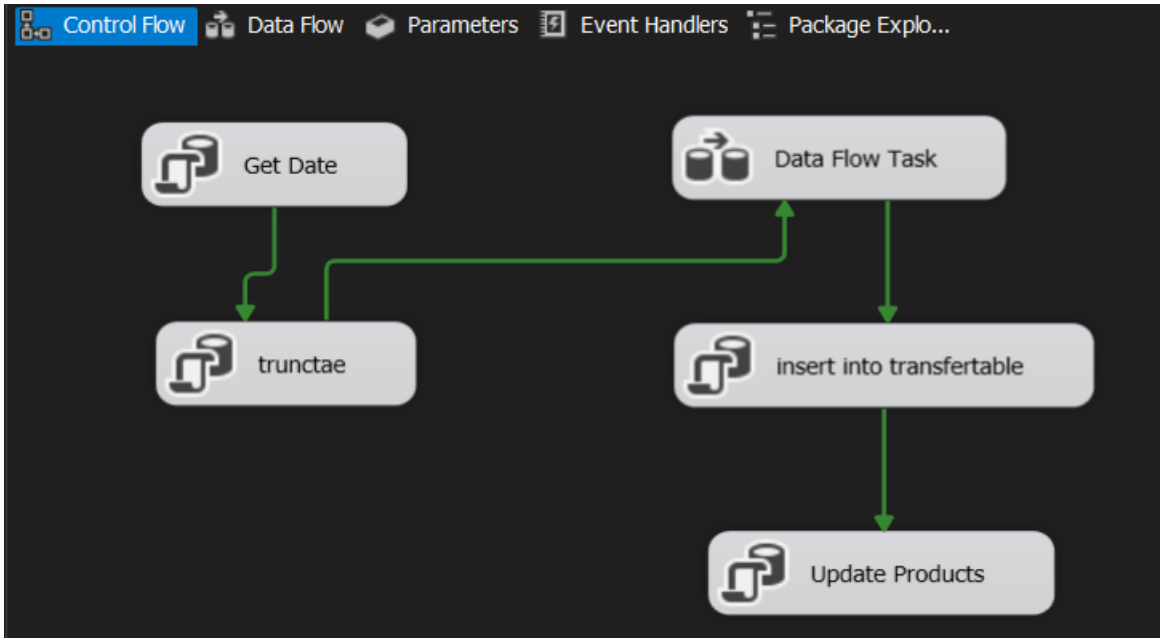
In the control flow:



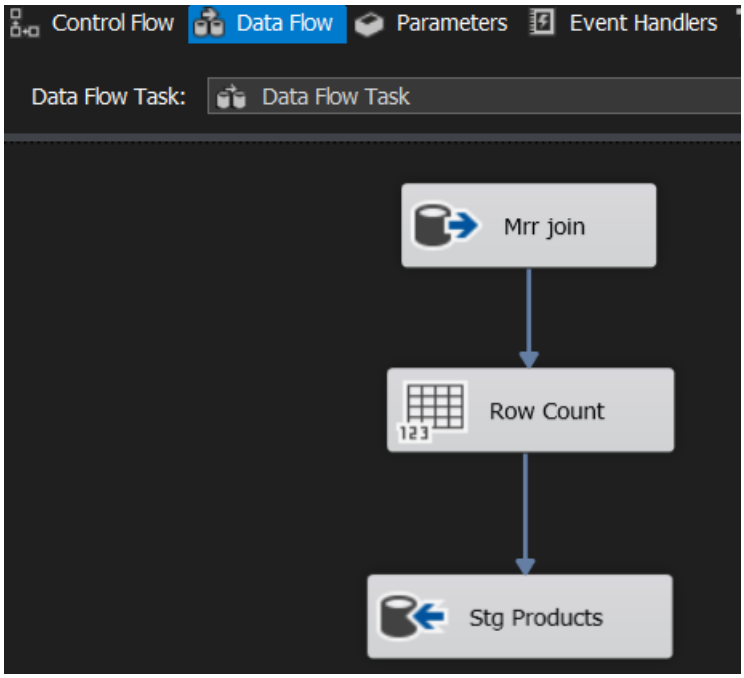
Products: in this solution I have 3 packages.

StgProducts: in this package I load to the StgProducts table the data from SQL Command (join relevant Mrr tables)

In the control flow:

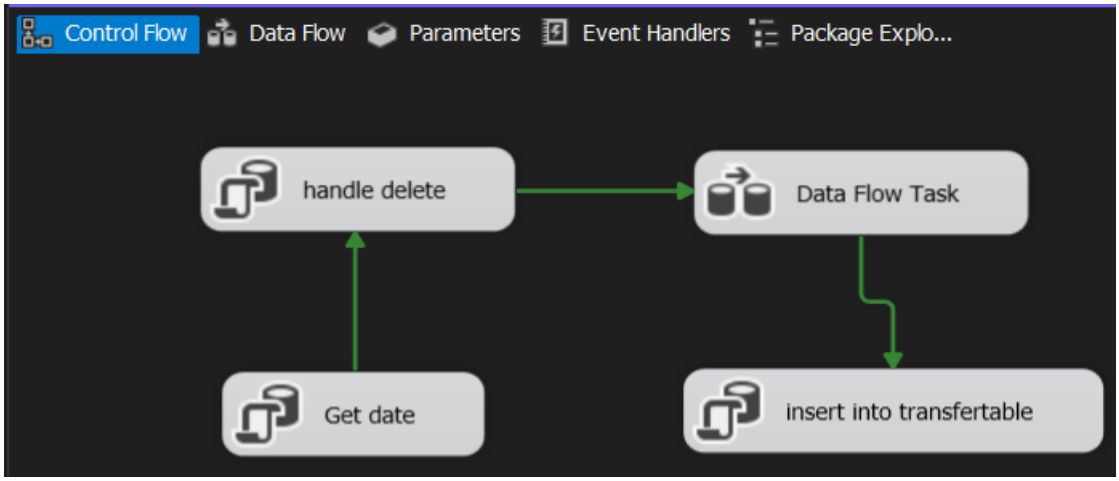


In the data flow:

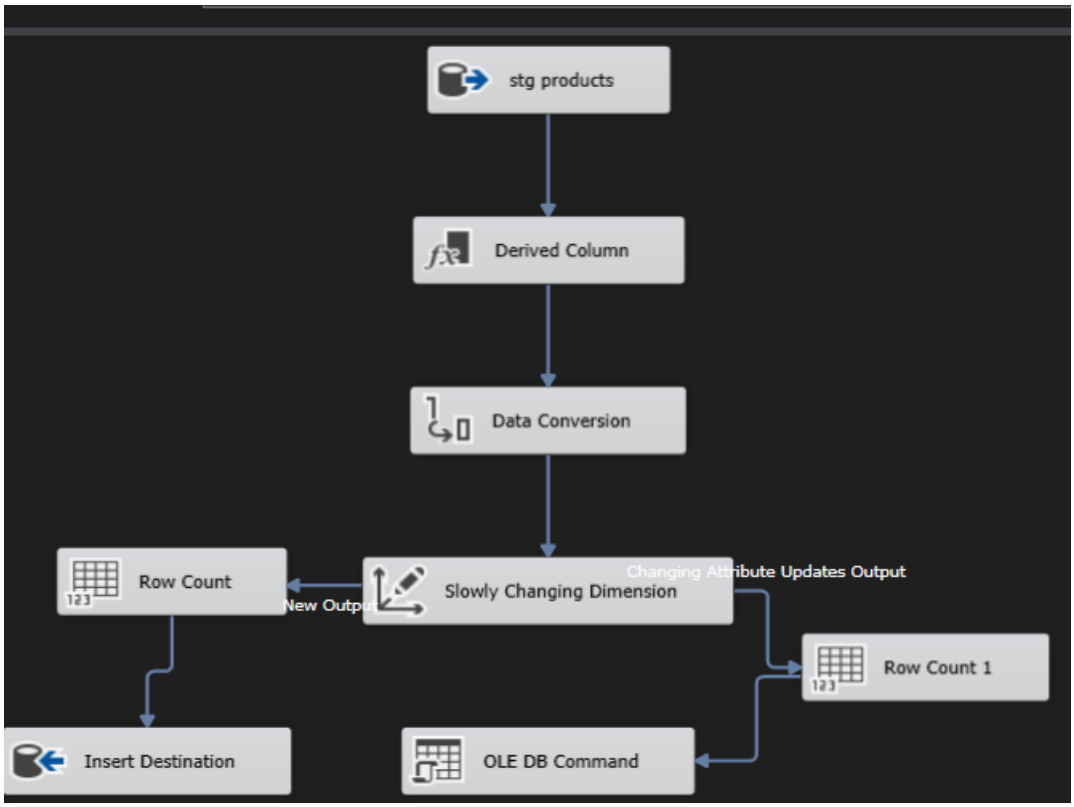


DimProducts: in this package I have loaded the data from the Stg table of products to the DimProducts table.

In the control flow:



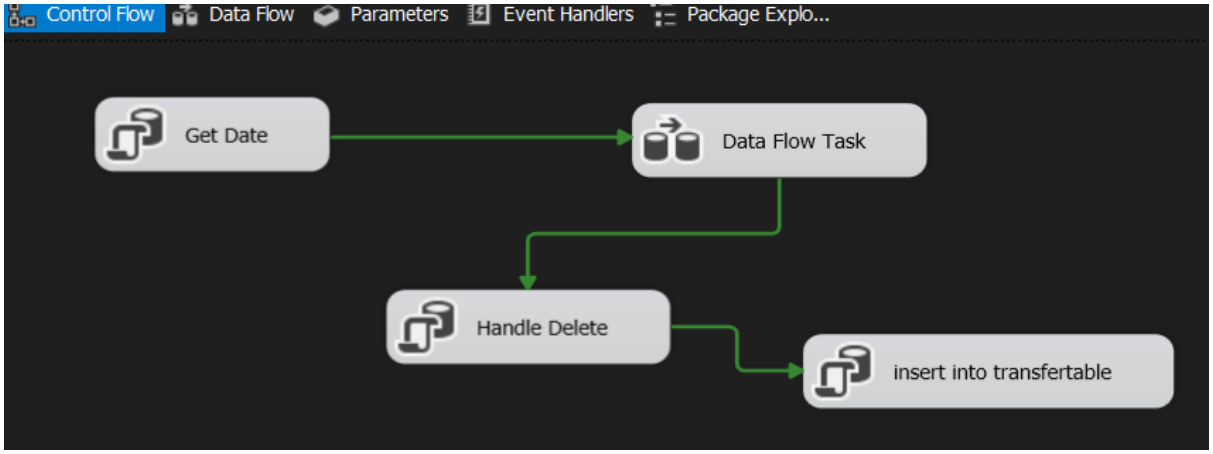
In the data flow:



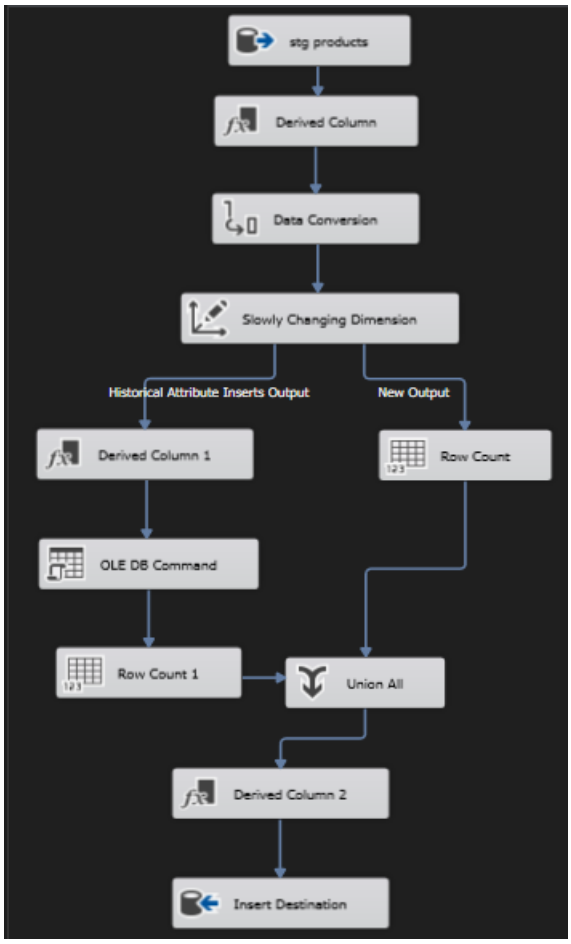
Derived column for replacing NULL values with UnKnown.
Data Conversion to change the types from Varchar to Nvarchar and to minimize the length.

DimProductsHistory package: in this package I have inserted the changes of products in the history table of the products.

In the control flow:



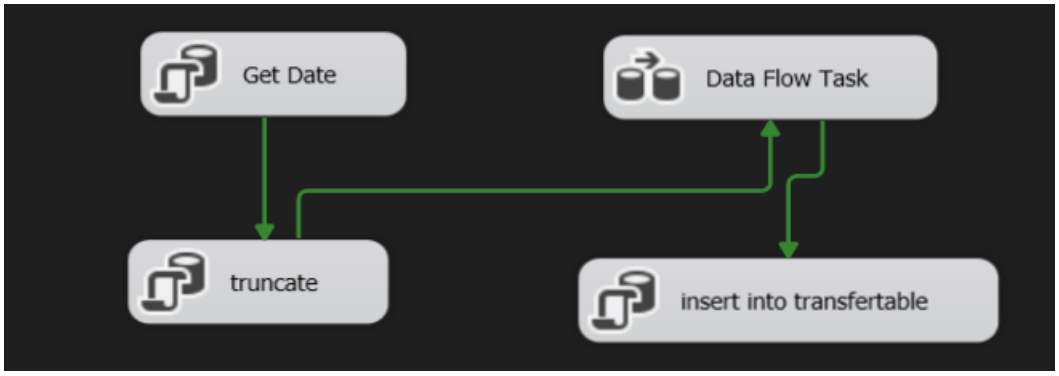
In the data flow:



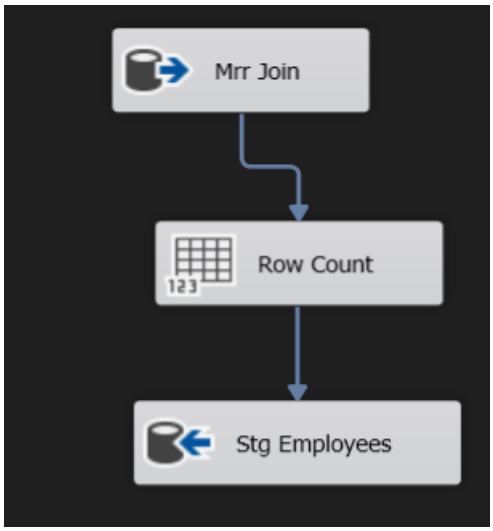
Employees: in this solution I have 2 packages.

StgEmployees:

In the control flow:

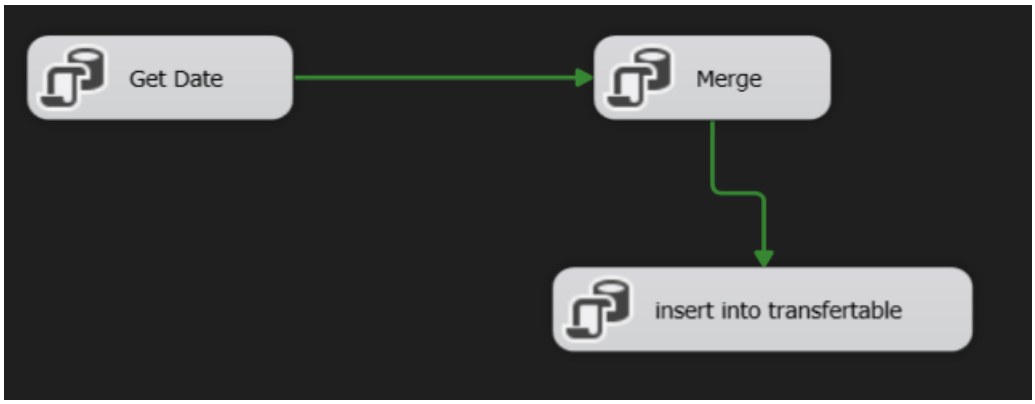


In the data flow:



DimEmplyees:

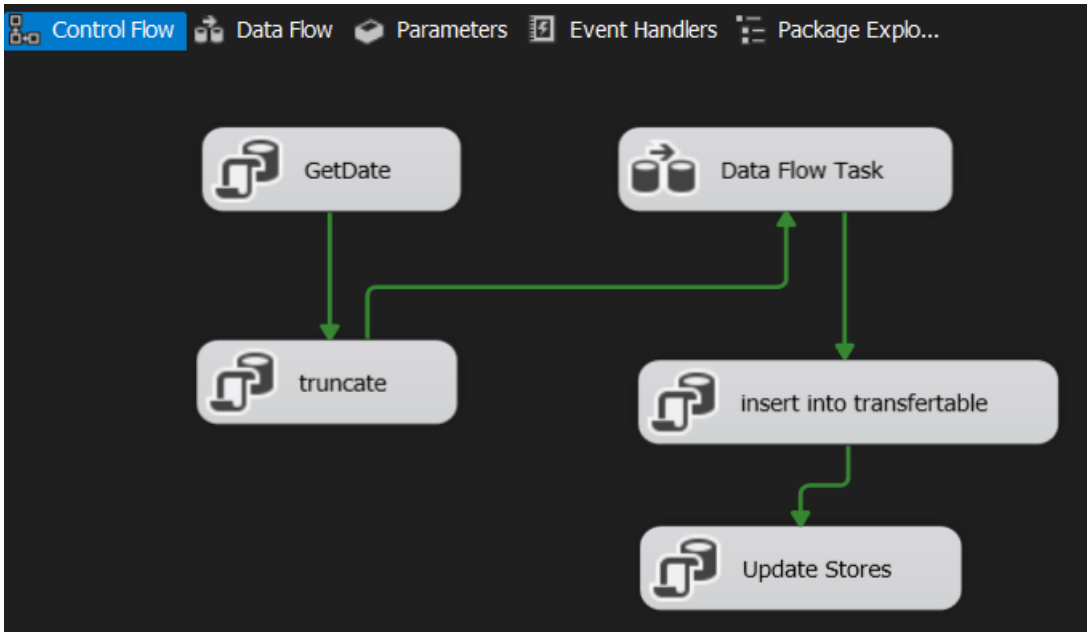
In the control flow:



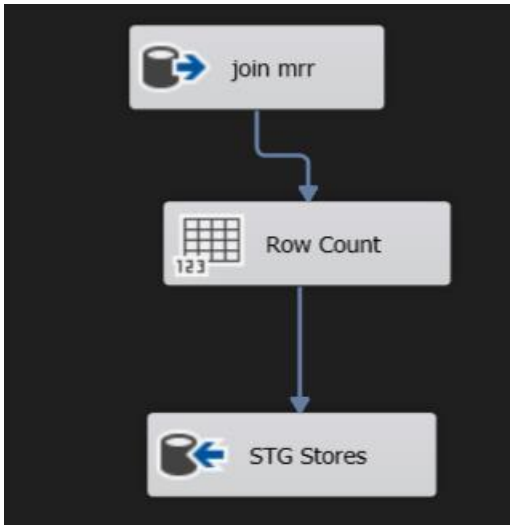
Stores: in this solution I have 2 packages.

StgStores:

In the control flow:

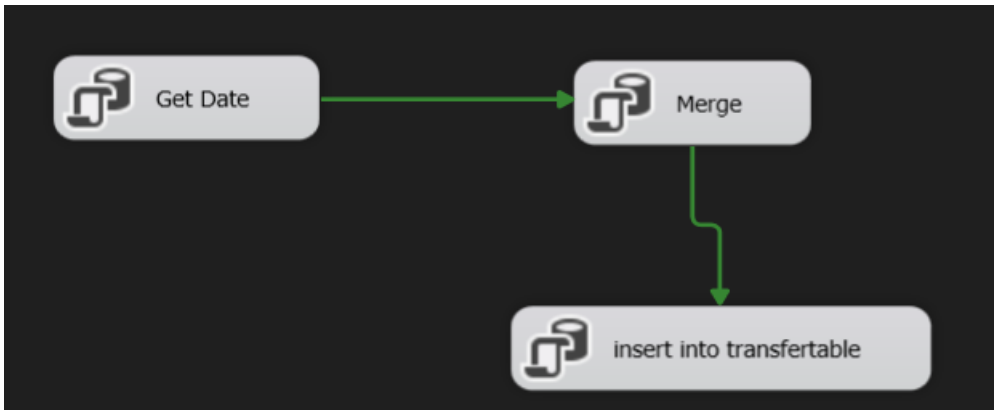


In the data flow:



DimStores:

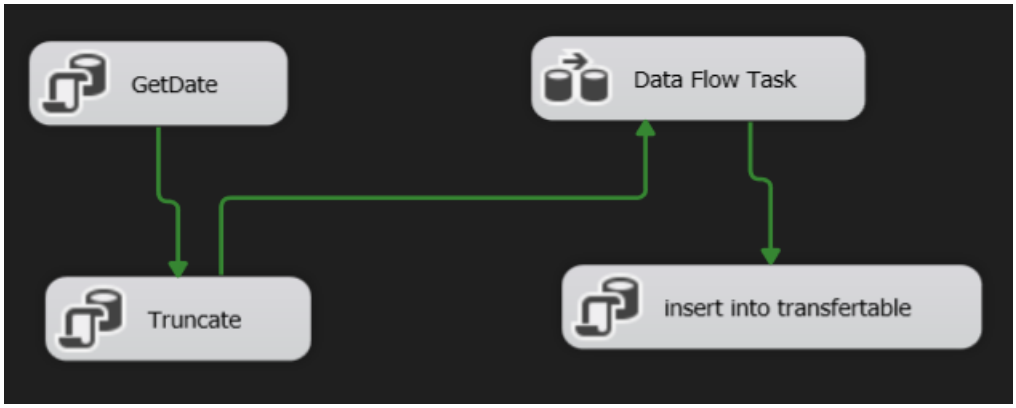
In the control flow:



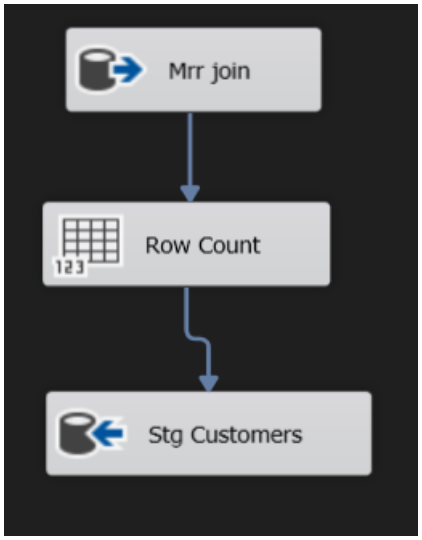
Customers: in this solution I have 2 packages.

StgCustomers:

In the control flow:

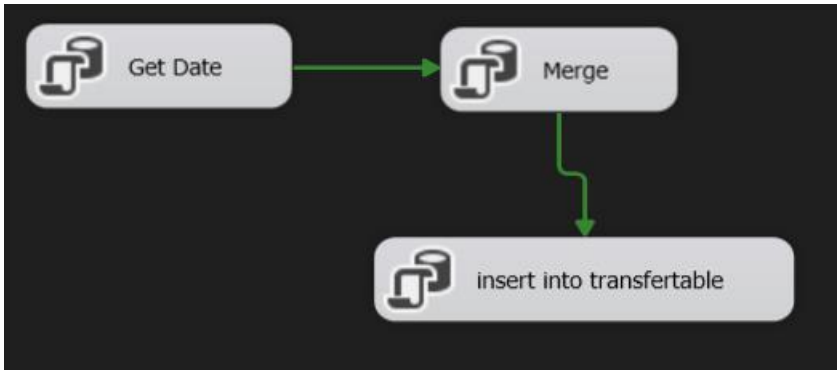


In the data flow:



DimCustomers:

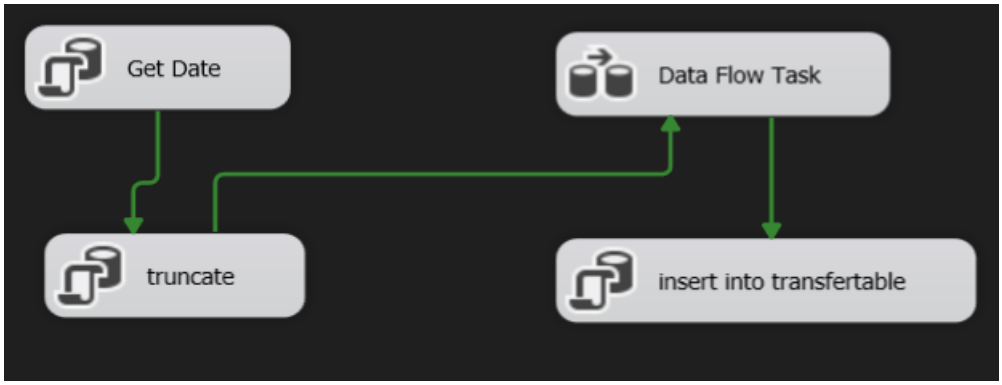
In the control flow:



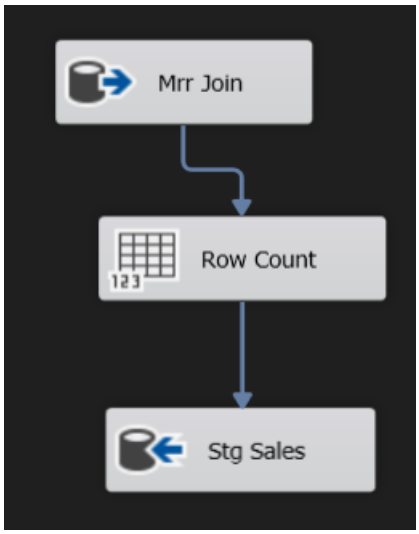
FactSales: in this solution I have 2 packages.

StgSales:

In the control flow:

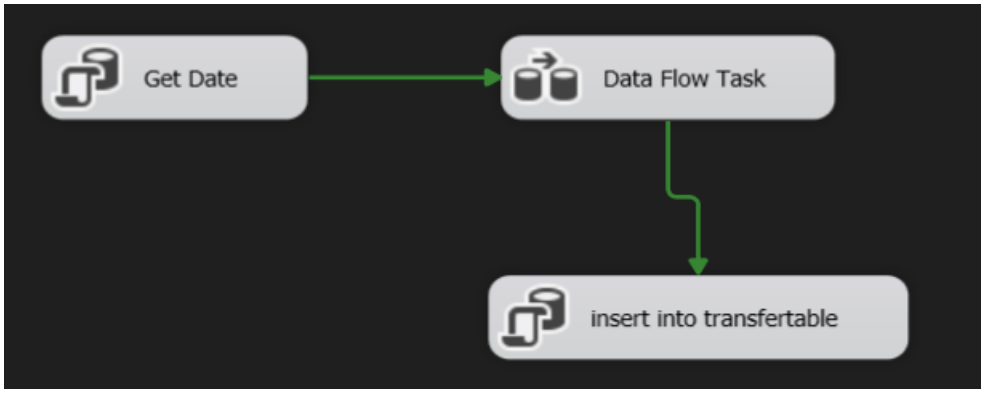


In the data flow:

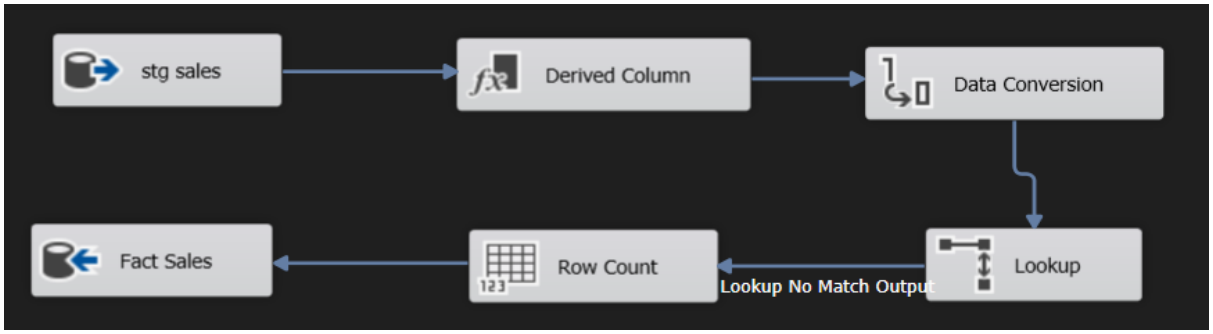


FactSales:

In the control flow:



In the data flow:

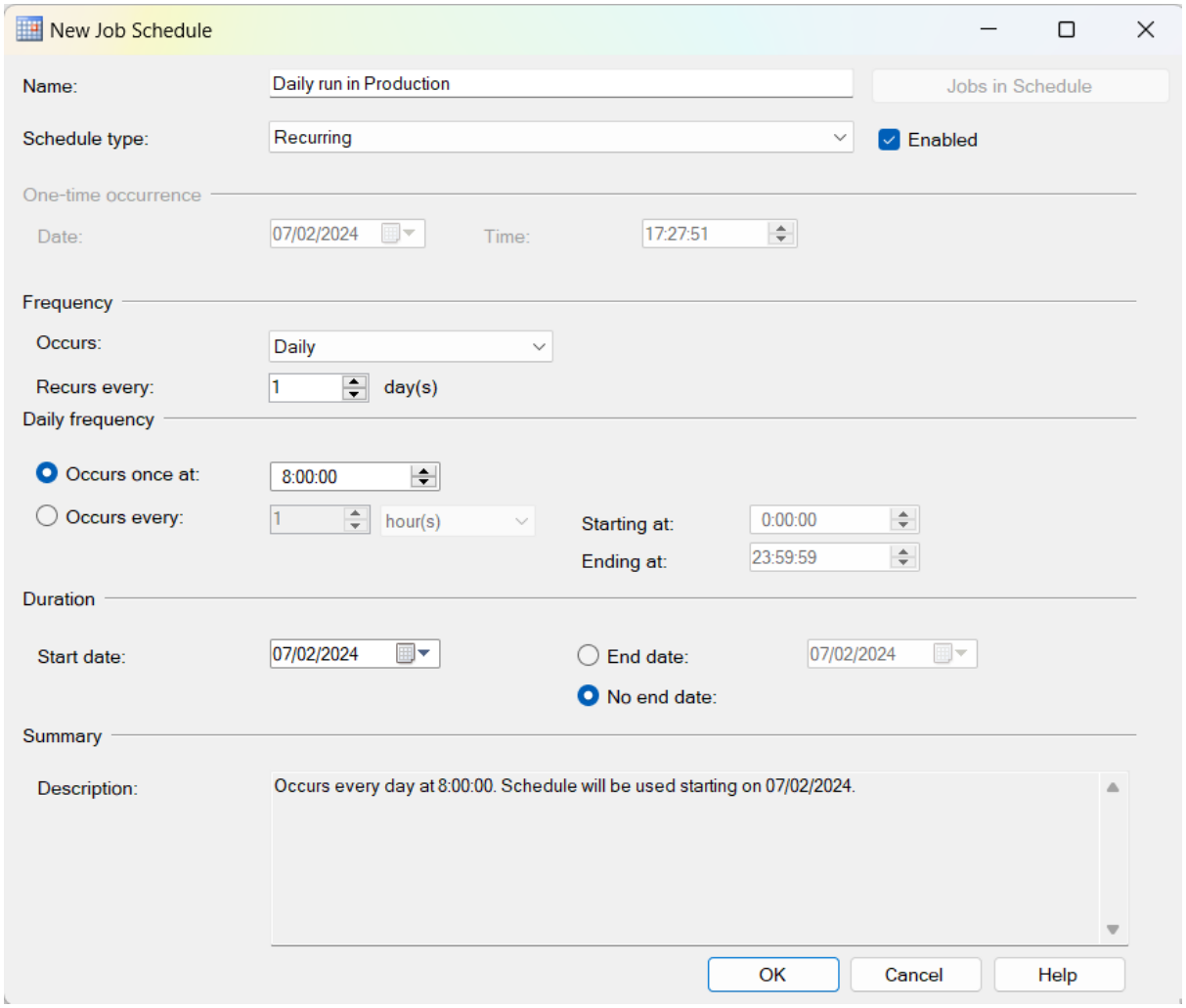


Derived column for calculating the Total.
Data conversion for changing the types from varchar to Date and from float to Decimal.

Next step is to deploy the project to SQL Server, create a job by the following steps:

Date	Step ID	Server	Job Name	Step Name	Notifications	Message
05/02/2024 23:17:...		LAPTOP-ULQRU4KB	E2E_HP			The job succeede
05/02/2024 23:...	13	LAPTOP-ULQRU4KB	E2E_HP	Fact Sales		Executed as user:
05/02/2024 23:...	12	LAPTOP-ULQRU4KB	E2E_HP	DIM Products History		Executed as user:
05/02/2024 23:...	11	LAPTOP-ULQRU4KB	E2E_HP	DIM Customers		Executed as user:
05/02/2024 23:...	10	LAPTOP-ULQRU4KB	E2E_HP	DIM Stores		Executed as user:
05/02/2024 23:...	9	LAPTOP-ULQRU4KB	E2E_HP	DIM Employees		Executed as user:
05/02/2024 23:...	8	LAPTOP-ULQRU4KB	E2E_HP	DIM Products		Executed as user:
05/02/2024 23:...	7	LAPTOP-ULQRU4KB	E2E_HP	STG Sales		Executed as user:
05/02/2024 23:...	6	LAPTOP-ULQRU4KB	E2E_HP	STG Customers		Executed as user:
05/02/2024 23:...	5	LAPTOP-ULQRU4KB	E2E_HP	STG Stores		Executed as user:
05/02/2024 23:...	4	LAPTOP-ULQRU4KB	E2E_HP	STG Employees		Executed as user:
05/02/2024 23:...	3	LAPTOP-ULQRU4KB	E2E_HP	STG Products		Executed as user:
05/02/2024 23:...	2	LAPTOP-ULQRU4KB	E2E_HP	MrrSales		Executed as user:
05/02/2024 23:...	1	LAPTOP-ULQRU4KB	E2E_HP	MrrDims		Executed as user:

The job will run in the production environment every day:



4.3. Description of data tables in the Data Warehouse (DWH):

As seen in the S2T: here is some screenshots of every table in the DataMart.

FactSales:

Order_ID	Order_Date	CustomerID	AgentID	TerritoryID	ProductID	Qty	Price	Discount	Total
49893	2013-02-28	29783	276	3	725	2	202.33	0.00	473.45
49893	2013-02-28	29783	276	3	762	6	469.79	0.00	3297.95
49893	2013-02-28	29783	276	3	843	2	15.00	0.00	35.10
49893	2013-02-28	29783	276	3	760	11	454.13	0.02	5727.81
49893	2013-02-28	29783	276	3	855	5	53.99	0.00	315.86
49894	2013-02-28	30096	277	2	780	5	1242.85	0.00	7270.68
49894	2013-02-28	30096	277	2	787	4	647.99	0.00	3032.61
49894	2013-02-28	30096	277	2	831	1	209.25	0.00	244.82
49894	2013-02-28	30096	277	2	783	8	1229.45	0.00	11507.73

DimStores:

Store_ID	Name	Location	IsActive	UpdateDate
292	Ivory	Tennessee	Y	2024-02-05 02:46:04.667
294	Ivory	California	Y	2024-02-05 02:46:04.667
296	Ivory	Minnesota	Y	2024-02-05 02:46:04.667
298	Ivory	Ohio	Y	2024-02-05 02:46:04.667
300	Ivory	New South Wales	Y	2024-02-05 02:46:04.667
302	Ivory	California	Y	2024-02-05 02:46:04.667

DimProducts:

ProductID	Product_Name	Sub_Category_Name	Category_Name	IsActive	UpdateDate
753	HP 250 G9 724K5EA	Laptop	PC	N	2024-02-05 01:45:45.757
779	HP 250 G9 6F1Z7EA	Laptop	PC	Y	2024-02-05 01:45:45.760
780	HP ProBook 440 G10 725H9EA	Laptop	PC	Y	2024-02-05 01:45:45.763
781	HP 200 G3 3VA74EA	Desktop	PC	Y	2024-02-05 01:45:45.763
782	HP All-in-One 27-cr0301nj	Desktop	PC	Y	2024-02-05 01:45:45.763
783	HP All-in-One 24-cr0235nj	Desktop	PC	Y	2024-02-05 01:45:45.763
784	HP 652 F6V25AE	Ink	Ink and Toners	Y	2024-02-05 01:45:45.763
793	HP 51639A	Ink	Ink and Toners	Y	2024-02-05 01:45:45.763

DimProductsHistory:

ProductID	Product_Name	Sub_Category_Name	Category_Name	startDate	endDate
798	HP LaserJet M110w	Combined	Printers	2024-02-05 22:48:55.000	NULL
797	HP M551N	Normal	Printers	2024-02-05 22:48:55.000	NULL
797	Road-550-W Yellow, 38	Road Bikes	Bikes	2024-02-05 01:55:20.000	2024-02-05 22:48:55.000
796	HP M551DN	Normal	Printers	2024-02-05 22:48:55.000	NULL
796	Road-250 Black, 58	Road Bikes	Bikes	2024-02-05 01:55:20.000	2024-02-05 22:48:55.000
795	Road-250 Black, 52	Road Bikes	Bikes	2024-02-05 01:55:20.000	2024-02-05 09:48:20.000
795	HP Sprocket 200	Normal	Printers	2024-02-05 09:48:20.000	NULL
794	HP 142A W1420A	Toner	Ink and Toners	2024-02-05 09:48:20.000	NULL
794	Road-250 Black, 48	Road Bikes	Bikes	2024-02-05 01:55:20.000	2024-02-05 09:48:20.000

DimEmployees:

Emp_ID	First_Name	Last_Name	Job_Title	Hire_Date	Phone_Number	Email_Address	Territory_Name	IsActive	UpdateDate
77777	NULL	NULL	Online	NULL	NULL	NULL	NULL	Y	2024-02-05 01:26:38.020
290	Ranjit	Varkey Chudukatil	Sales Representative	2012-05-30	1 (11) 500 555-0117	ranjit0@adventure-works.com	France	Y	2024-02-05 01:26:38.020
289	Jae	Pak	Sales Representative	2012-05-30	1 (11) 500 555-0145	jae0@adventure-works.com	United Kingdom	Y	2024-02-05 01:26:38.020
288	Rachel	Valdez	Sales Representative	2013-05-30	1 (11) 500 555-0140	rachel0@adventure-works.com	Germany	Y	2024-02-05 01:26:38.020
287	Amy	Alberts	European Sales Manager	2012-04-16	775-555-0164	amy0@adventure-works.com	NULL	Y	2024-02-05 01:26:38.020
286	Lynn	Tsoflias	Sales Representative	2013-05-30	1 (11) 500 555-0190	lynn0@adventure-works.com	Australia	Y	2024-02-05 01:26:38.020

DimCustomers:

CustomerID	Name	Address	City	Region	Country	StoreID	IsActive	UpdateDate
11000	Jon Yang	3761 N. 14th St	Rockhampton	Queensland	Australia	1995	Y	2024-02-05 01:29:07.313
11001	Eugene Huang	2243 W St.	Seaford	Victoria	Australia	1995	Y	2024-02-05 01:29:07.313
11002	Ruben Torres	5844 Linden Land	Hobart	Tasmania	Australia	1995	Y	2024-02-05 01:29:07.313
11003	Christy Zhu	1825 Village Pl.	North Ryde	New South Wales	Australia	1995	Y	2024-02-05 01:29:07.313
11004	Elizabeth Johnson	7553 Harness Circle	Wollongong	New South Wales	Australia	1995	Y	2024-02-05 01:29:07.313
11005	Julio Ruiz	7305 Humphrey Drive	East Brisbane	Queensland	Australia	1995	Y	2024-02-05 01:29:07.313
11006	Janet Alvarez	2612 Berry Dr	Matraville	New South Wales	Australia	1995	Y	2024-02-05 01:29:07.313

Transfertable:

This table counts the rows that inserted in every table from each stage:

PackageName	TableName	InsertDate	EndDate	RowsCount
FactSales	FactSales	2024-02-05 23:20:22.000	2024-02-05 23:20:22.890	0
DimProductsHistory	DimProductsHistory	2024-02-05 23:20:08.000	2024-02-05 23:20:08.667	0
DimCustomers	DimCustomers	2024-02-05 23:19:55.000	2024-02-05 23:19:55.993	0
DimStores	DimStores	2024-02-05 23:19:45.000	2024-02-05 23:19:44.860	0
DimEmployees	DimEmployees	2024-02-05 23:19:34.000	2024-02-05 23:19:34.200	0
DimProducts	DimProducts	2024-02-05 23:19:20.000	2024-02-05 23:19:20.770	0
StgSales	StgSales	2024-02-05 23:19:06.000	2024-02-05 23:19:06.370	0
StgCustomers	StgCustomers	2024-02-05 23:18:51.000	2024-02-05 23:18:51.790	19119
StgStores	StgStores	2024-02-05 23:18:38.000	2024-02-05 23:18:38.330	636
StgEmployees	StgEmployees	2024-02-05 23:18:23.000	2024-02-05 23:18:23.120	18
StgProducts	Stg_Products	2024-02-05 23:18:09.000	2024-02-05 23:18:08.813	241
Mrr_Sales	MRR_InvoiceItems	2024-02-05 23:17:54.000	2024-02-05 23:17:55.507	0
Mrr_Sales	MRR_Invoices1	2024-02-05 23:17:54.000	2024-02-05 23:17:55.503	0
Mrr_Dims	MRR_Invoices	2024-02-05 23:17:37.000	2024-02-05 23:17:39.237	31465
Mrr_Dims	MRR_Familys	2024-02-05 23:17:37.000	2024-02-05 23:17:39.233	38
Mrr_Dims	MRR_FamilyTypes	2024-02-05 23:17:37.000	2024-02-05 23:17:39.233	4

4.4. Power BI

➢ In Power BI, I have created several measures:

• Total Sales = SUM(FactSales[Total])

• Total Orders = DISTINCTCOUNT(FactSales[Order_ID])

• Total Units = SUM(FactSales[Qty])

• Number of Customers = DISTINCTCOUNT(FactSales[CustomerID])

• Avg sale per customer = DIVIDE('Measures Tbl'[Total Sales],'Measures Tbl'[Number of Customers],0)

• Online Sales = CALCULATE([Total Sales],DimStores[Store_ID]=1995)

• Average Sale Per Order = DIVIDE([Total Sales],[Total Orders],0)

• Average Units Per Order = DIVIDE([Total Units],[Total Orders],0)

• Total Sales by Emplyees = CALCULATE([Total Sales],FILTER('DimEmployees','DimEmployees'[Emp_ID]<>77777))

• Online Sales = IF(ISBLANK(CALCULATE([Total Sales],DimStores[Store_ID]=1995)),0,CALCULATE([Total Sales],DimStores[Store_ID]=1995))

• Direct Sale = CALCULATE([Total Sales],DimStores[Store_ID]<>1995)

• Online = DIVIDE([Online Sales],[Total Sales],0)

• Direct = DIVIDE([Direct Sale],[Total Sales],0)

• Price Category = SWITCH(TRUE(),'Sales Price Category'[Total Price Per Order]>0&&'Sales Price Category'[Total Price Per Order]<=500,"Low Price (1-500)",'Sales Price Category'[Total Price Per Order]>500&&'Sales Price Category'[Total Price Per Order]<=5000,"Medium Price (501-5000)",'Sales Price Category'[Total Price Per Order]>5000,"High Price (>5000)")

• All Orders = CALCULATE([Total Orders],ALL('Sales Price Category'))

• % price Category = [Total Orders]/[All Orders]

• YTD Sales = TOTALYTD([Total Sales],'Dim Date'[Date])

• Last Year Sales = CALCULATE([Total Sales],SAMEPERIODLASTYEAR('Dim Date'[Date]))

• % Sales Difference By Year = DIVIDE([YTD Sales]-[Last Year Sales],[Last Year Sales],0)

• YTD Orders = TOTALYTD([Total Orders],'Dim Date'[Date])

• Last Year Orders = CALCULATE([Total Orders],SAMEPERIODLASTYEAR('Dim Date'[Date]))

• % Orders Sales Difference By Year = DIVIDE([YTD Orders]-[Last Year Orders],[Last Year Orders],0)

• YTD Units = TOTALYTD([Total Units],'Dim Date'[Date])

• Last Year Units = CALCULATE([Total Units],SAMEPERIODLASTYEAR('Dim Date'[Date]))

• % Units Sales Difference By Year = DIVIDE([YTD Units]-[Last Year Units],[Last Year Units],0)

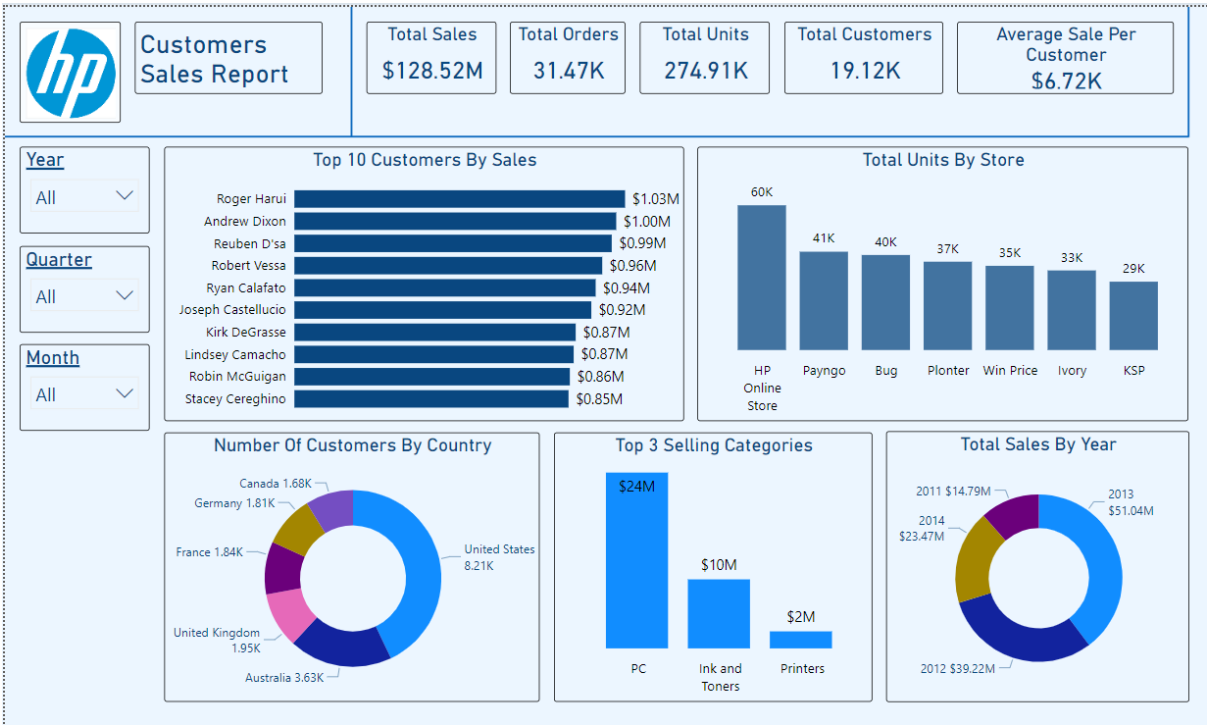
• Number Of Custmers Prev Month = CALCULATE([Number of Customers],DATEADD('Dim Date'[Date],-1,MONTH))

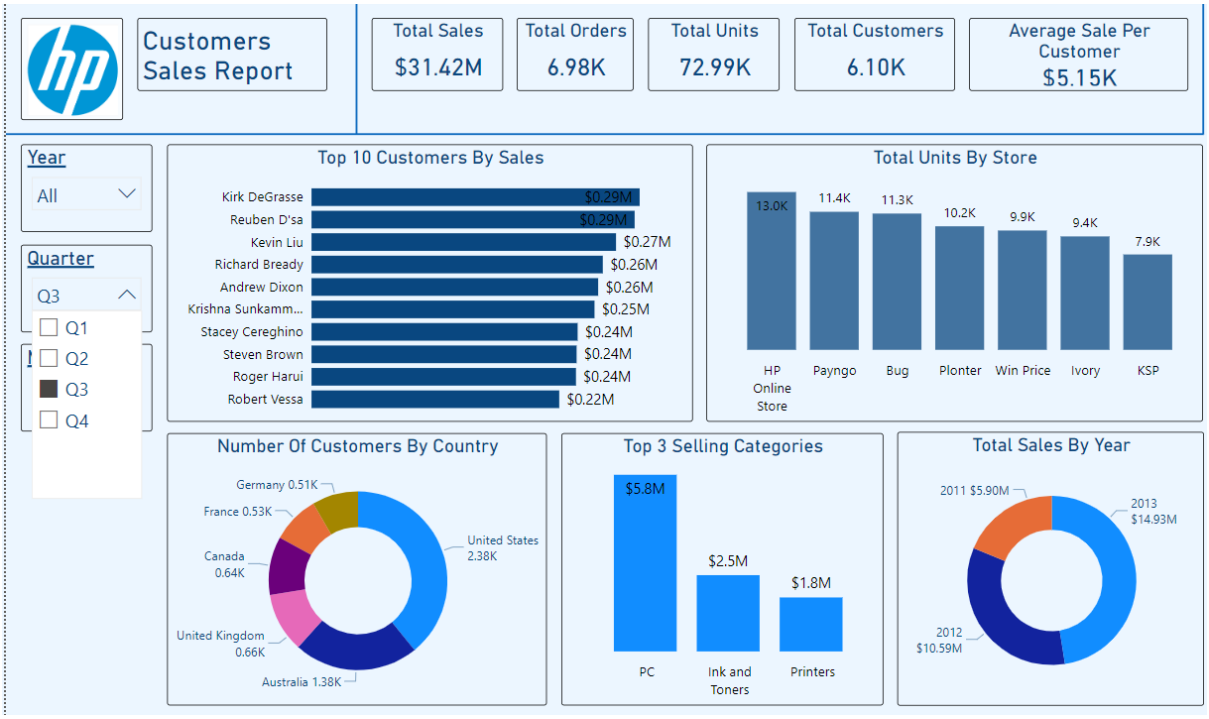
• % Growth Of Customers = DIVIDE([Number of Customers]-[Number Of Custmers Prev Month],[Number Of Custmers Prev Month],0)

• Total Sales Last Month = CALCULATE([Total Sales],DATEADD('Dim Date'[Date],-1,MONTH))

• % Sales Difference By Month = DIVIDE([Total Sales]-[Total Sales Last Month],[Total Sales Last Month],0)

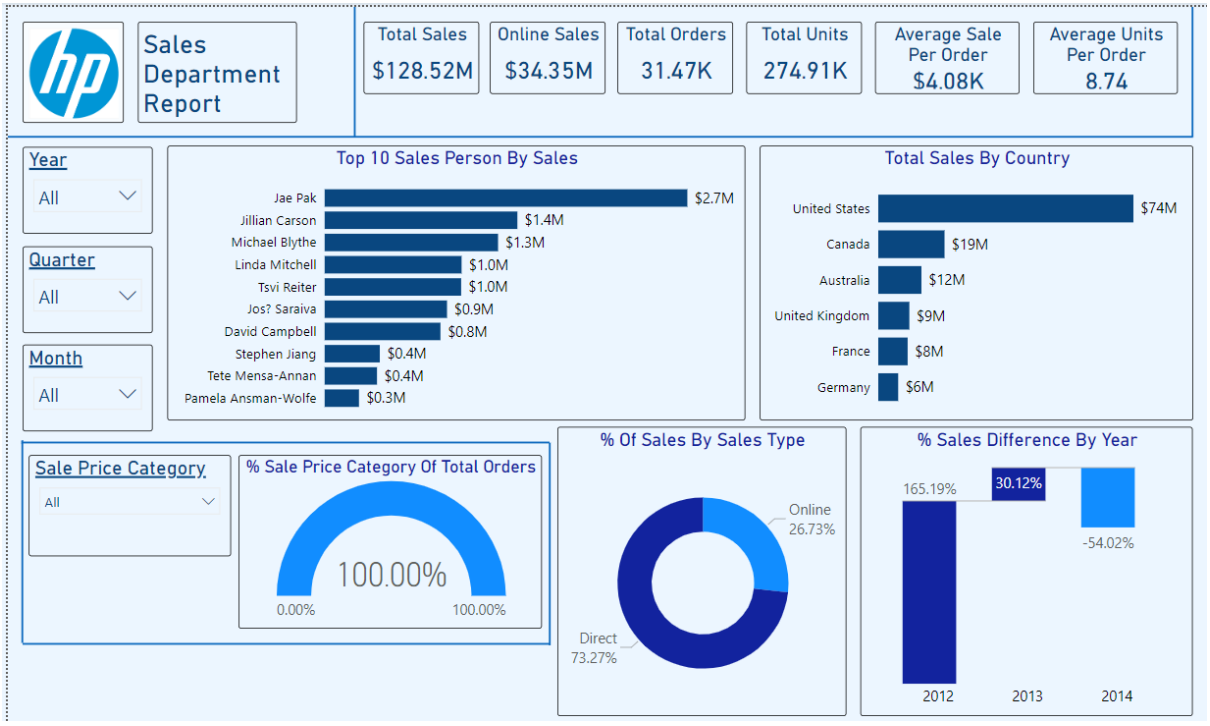
Customers Report:

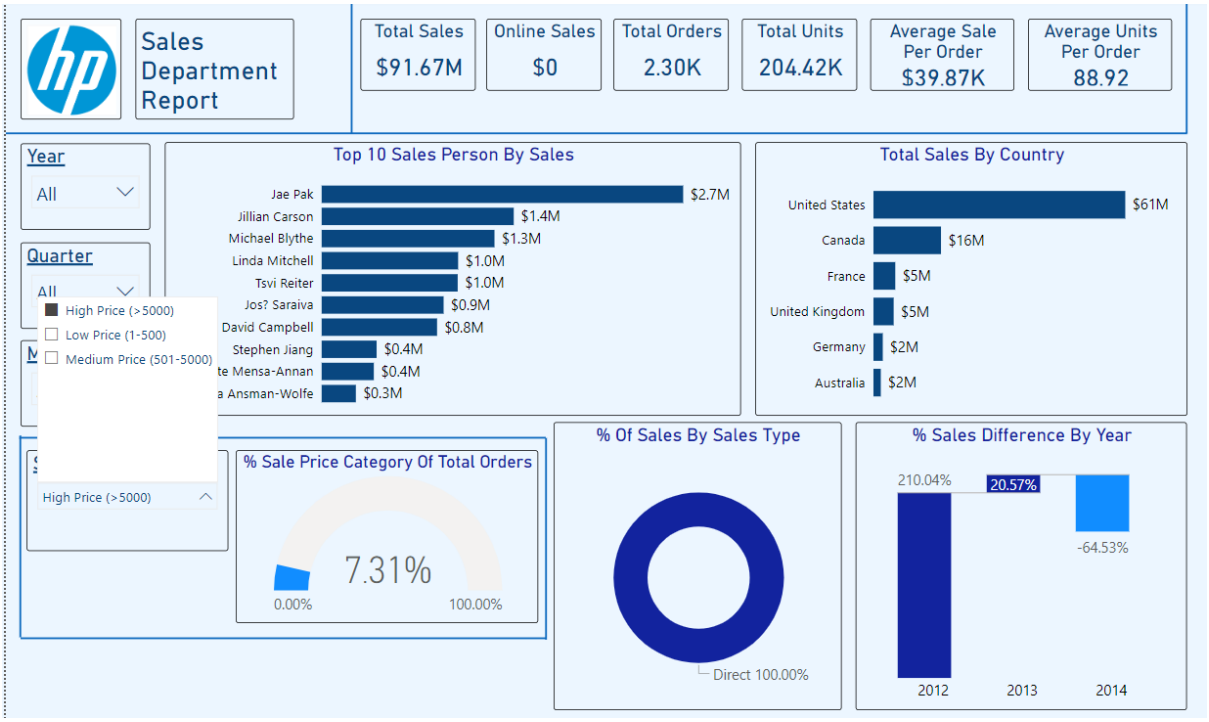




In the customers' report you can filter by year/quarter/month the top 10 customers by sales, total units sold by store, top 3 selling categories, number of customers by country and total sales by year.
And the 2 measures above will change accordingly (total customers and average sale per customer).

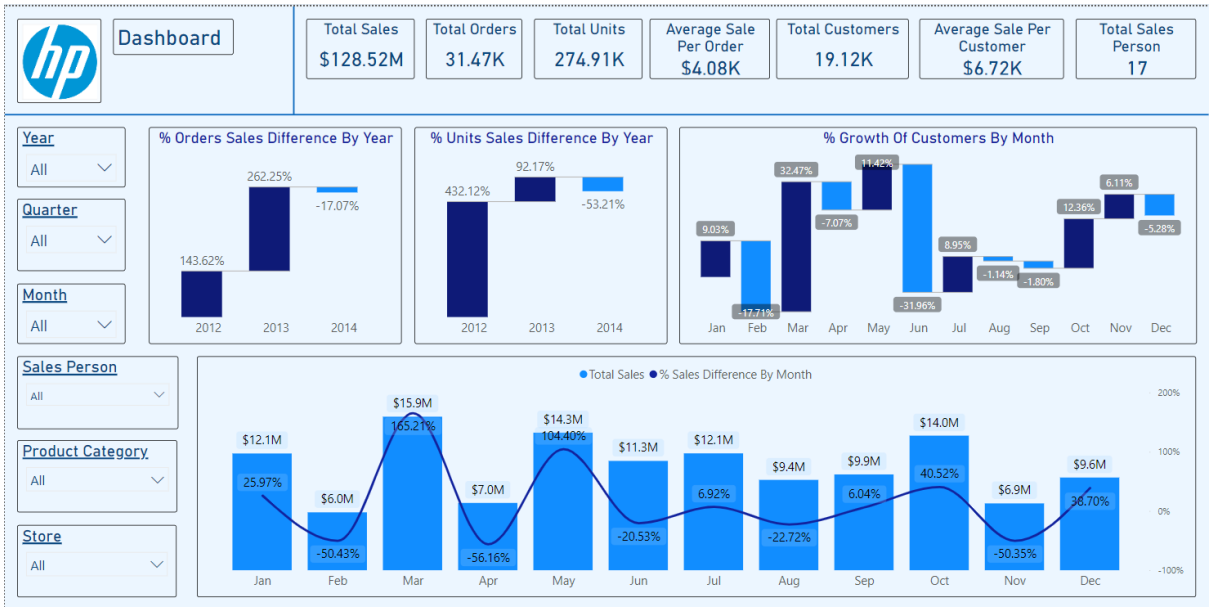
Sales Department Report:

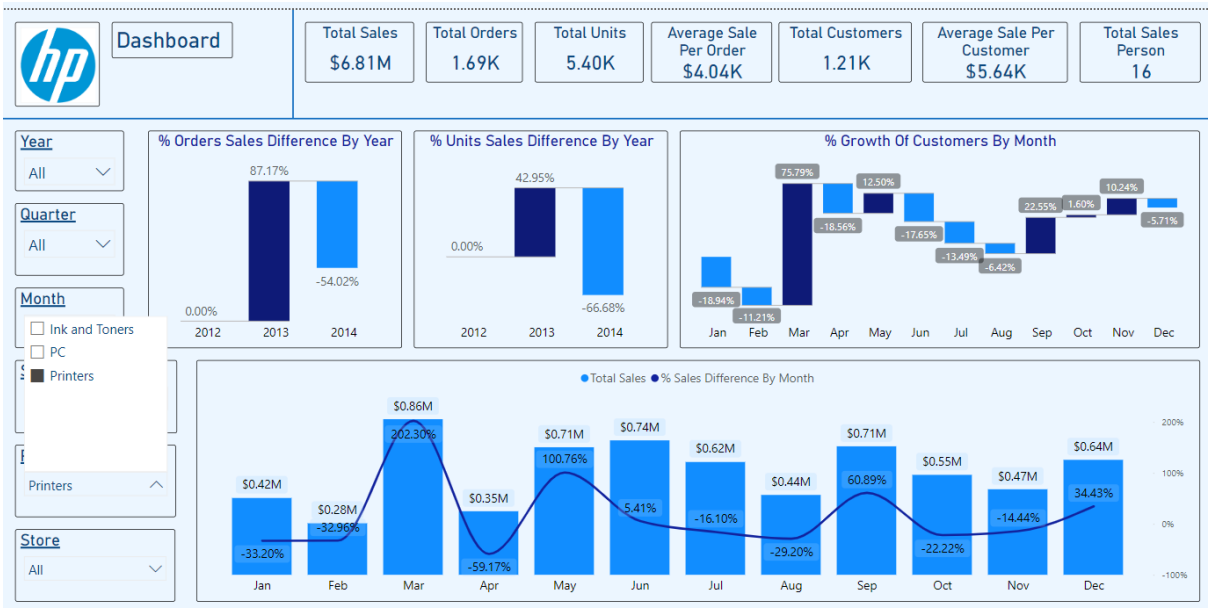




In the sales department report you can filter by year/quarter/month the top 10 salesperson by sales, total sales by country, percentage of sales by sale type, percentage of sales difference by year and percentage of sale price category by total orders.
And the 3 measures above will change accordingly (average sale per order, average units per order, and online sales amount).

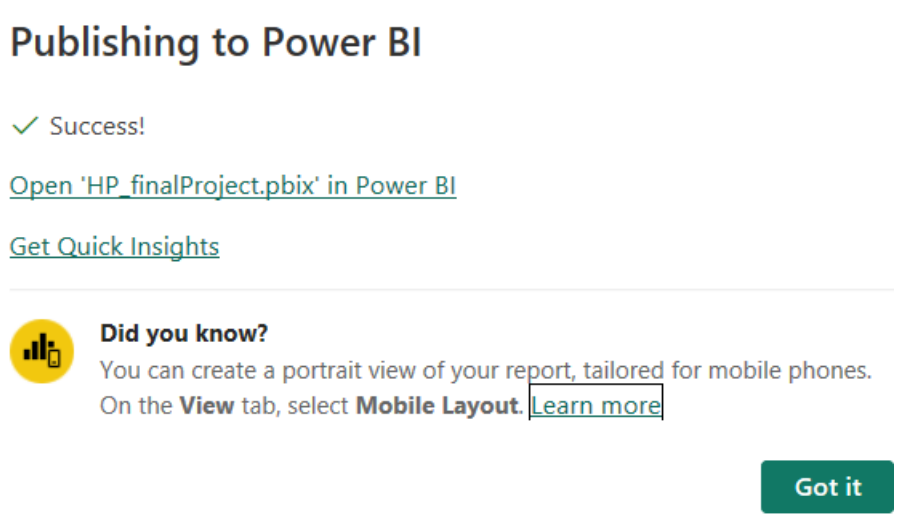
Dashboard:



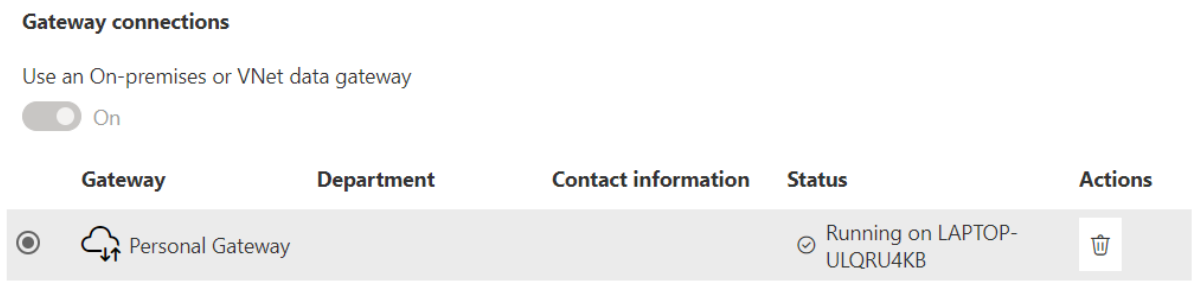


In the dashboard you can filter by sales person, product category, and store for analyzing the changing of sales every month, changing of the orders by year, changing of the units sold by year and the changing of the customers who have purchased every month.

Publish the reports:



Activate the gateway in power bi service:



Configuring a refresh schedule:

Refresh

Configure a refresh schedule

Define a data refresh schedule to import data from the data source into the semantic model.

On

Refresh frequency

Daily

Time zone

(UTC+02:00) Jerusalem

Time

9 00 AM

Add another time

Send refresh failure notifications to

Semantic model owner

These contacts:

F Fares Enter email addresses

Apply Discard

HP_finalProject	Semantic model	BI_HP	2/9/24, 10:51:51 PM	2/10/24, 9:00:00 AM	—	—

Then creating the app: