SSIS Exploration Document

Inventory and Sales Management System
Project business intelligence

Avishay Aknin

CONTENT

1 Document Management

1.1 Project Scope & Objective

The purpose of the project is to provide Data Warehouse Solution Creation from AdventureWorks2019 Database for AvishayDB, A company that supplies products and conducts business transactions.
This project aims to establish a comprehensive BI solution transferring data from the AdventureWorks2019 Database for SSIS_DB. The solution will contain summarized data tables, with a focus on sales data, employee records, customer information, product details.
Moreover, the project will solve the following issues:

- Providing a unified approach to organized and well-suited data for significant decision-making scenarios.

- Integrating data from different sources to be further used by reporting activities. The data is updated in real time, which makes it conveniently used for automated actions.

1.2 Project Content

In this project, we will build a Data Mart that will include information about many aspects of a real-world business order detailing from end-to-end perspective.

1. Data Cleaning and Preparation: Before the analysis starts, we filter our data via data cleaning method and preparation to verify their quality and consistency.

2. Main summary tables that will be built for the company's demands:

    - **Sales_Fact_table**– Contains information about all the orders, including dates, pricing and quantity per customer from the transactions of the transactional database.

    - **Stores_Dim**– Information about the company's Stores**.**

    - **Customer_Dim**– Information about the company's customers.

    - **Products_dim**- Information about the products that are sold by the company.

    - **Calendar_dim** - Information about date of transaction .

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

- ✔ **Sales Transaction orders Department:**

  The Orders Department will focus on the sales-related data, when the order took place, identified by the customer that made the order, the store that activated the order. This department will contain a full description of the purchase including the productID, its price and the total amount per product in the order.
  The department gathers information every time that a new order enters to the system and help the user to follow over the new transactions.

- ✔ **Store Department:**

  Within this department, stores are tasked with initiating orders for specific customers, with each order being documented in the Sales_Fact_table table. Relevant stores details are housed in the Stores_Dim table. Each store is distinguished by a unique identifier (StoreId), establishing a direct connection to the Sales_Fact_table table for efficient order management.

- ✔ **Customer Department:**

  A customer can make a new order, all the data about the customer will be stored inside the dimCustomers table – including an identifier (CustomerID) that is connected to the Sales_Fact_table table.

- ✔ **Production Department:**

  A table that contains all the manufactured product in the company, every product has a description about it. It includes an identifier (ProductID) which is connected to Sales_Fact_table Table and will take part inside the transaction process.

2 Technical Specifications

2.1 Prerequisites – the involved systems during the process

System / Process	Explanation
SQL Server	Operational DB – tables – data (SQL Files)
SSIS	ETL processes using SSIS in Visual Studio
Data Refreshing	Refreshing processes through an attribute of Employees in SSMS

2.2 Source to Target and ERD models

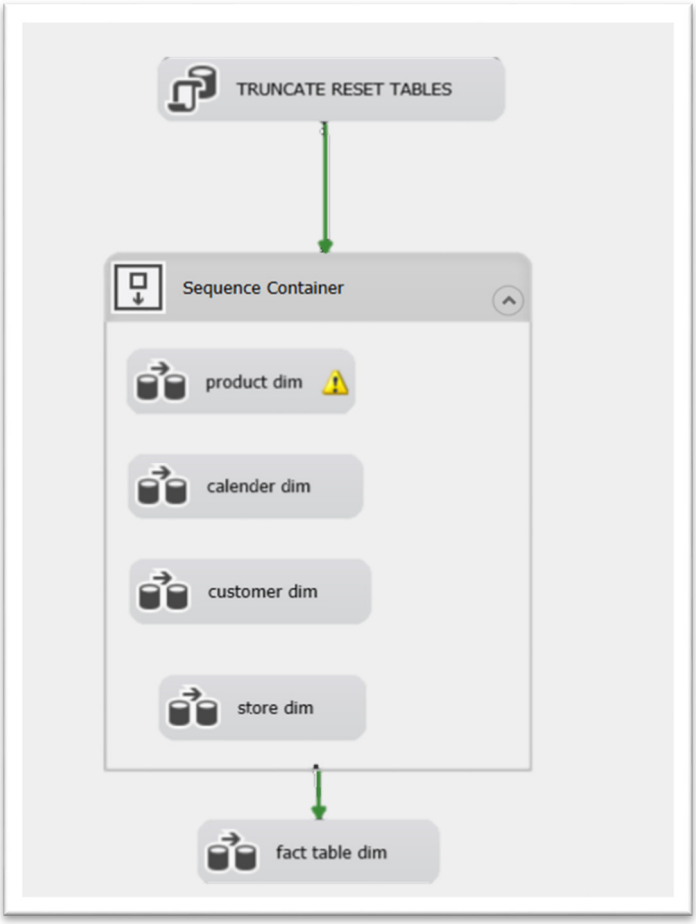
2.3 Detailed description of ETL Process

1. Master Package-

Stage One: Data Deletion
The initial phase involves clearing existing data from dimension and FACT tables to establish a clean environment for subsequent operations. This is achieved by connecting to the database, executing SQL tasks to truncate or delete data from the tables, and ensuring seamless interaction between the SSIS package and the database.

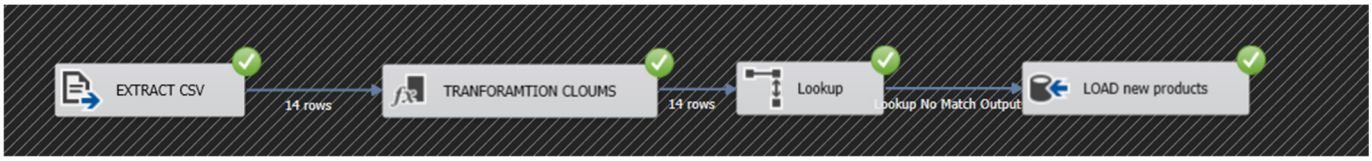
Stage Two: Data Manipulation
In this stage, data manipulation occurs, focusing on loading data into dimension tables. A container is introduced for organizational clarity, within which "Data Flow Tasks" are created for each dimension. These tasks extract data from source views and load it into corresponding dimension tables, with field mappings ensuring data alignment.

Stage Three: Updating the FACT Table
The final stage revolves around updating the FACT table with processed data from the dimensions. A "Data Flow Task" is employed to transfer data from source views containing all dimensions and the FACT table. Field mappings are established to integrate data seamlessly into the FACT table, maintaining data integrity.

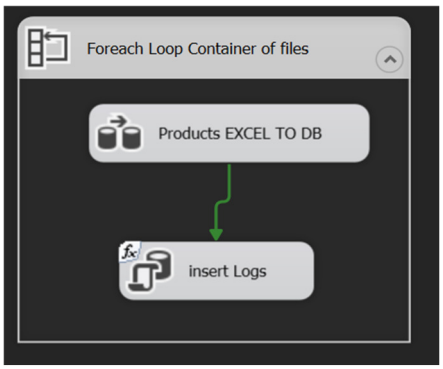


2. CSV TO DB -
Product table – insert just the new products

Loading the csv files: Retailers products(such as spring , tnuva) into Products tables.
The load action will load ONLY new records that did not appear in the 'products' table before to avoid a case that we load duplicated data. All the new records will be counted.



In this loop there is an exchange of the files that contain items of products in Excel files that belong to each company - such as Castro, Tnuva, etc.



In addition, I built documentation of the logs of all the files, if there are new details there will be documentation in the database "LogTable"

	id	status	file_name	records_insert	log_date
1	1	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	3	2024-03-18 22:01:23.903
2	2	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	6	2024-03-18 22:01:24.097
3	3	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	1	2024-03-18 22:01:24.290

3.

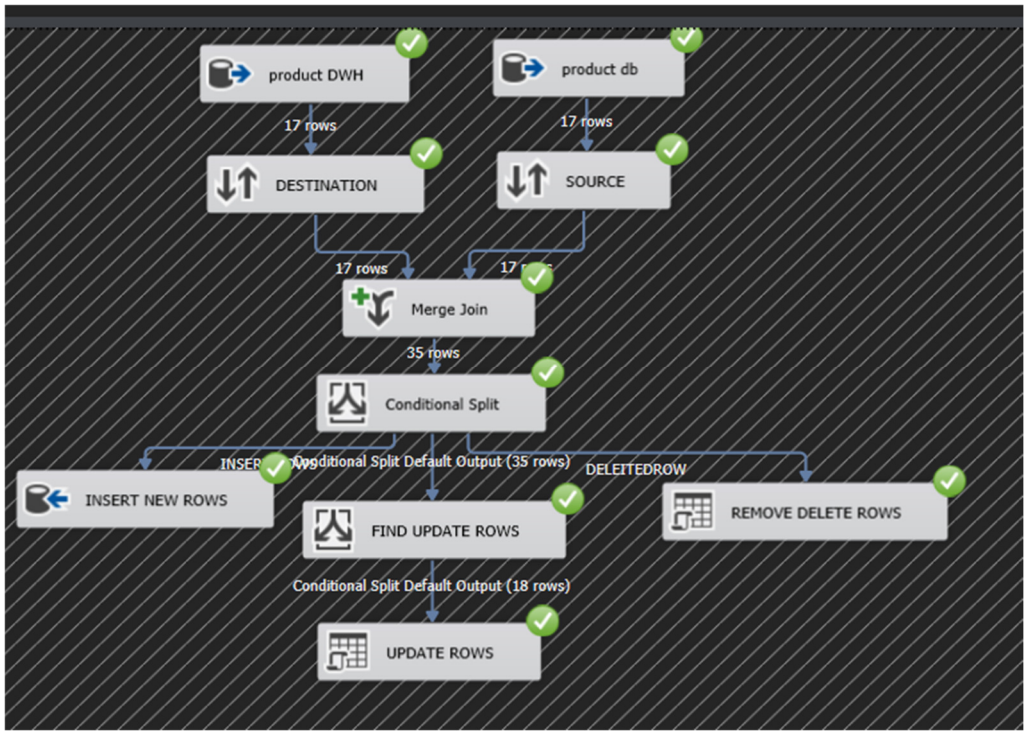
DB TO DWH -
Product table – insert, delete, update just the relevant products rows

Packeage aim:
Ensure synchronization of data between a database (db) table and a data warehouse (dwh) table efficiently and accurately using SQL Server Integration Services (SSIS).

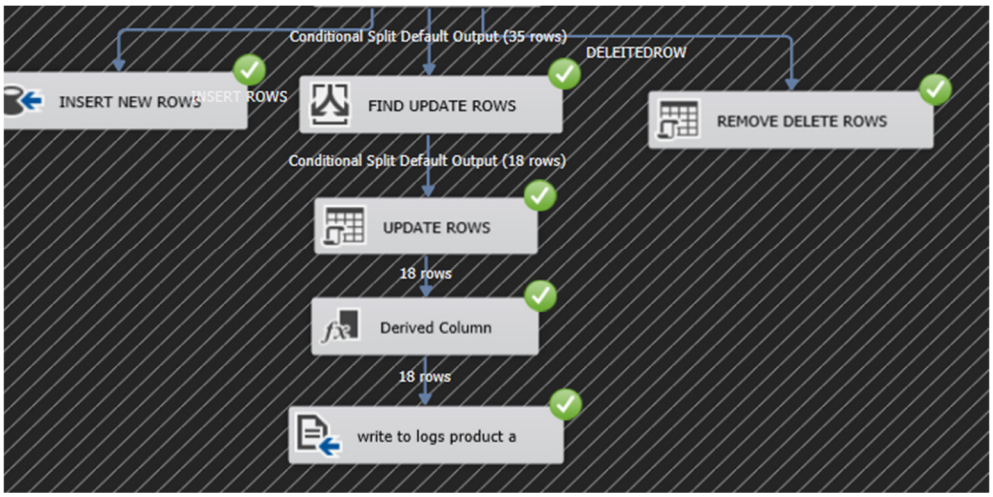
Levels Explained:

- Database and Data Warehouse Setup:

- Create tables in both the database (db) and data warehouse (dwh).
- Data Comparison and Sorting: Employ a Merge Join tool to compare data between the db and dwh tables. Utilize a Conditional Split tool to categorize changes such as new data, updates, or deletions of data.
- Detailed Data Analysis: Conduct in-depth analysis of existing data to identify any discrepancies between the db and dwh tables.
- Verify if the data in both tables aligns accurately.



Furthermore, I have implemented a feature whereby all updated products are logged into a text file. This log file contains a comprehensive list of products that underwent updates on any given day.



This it the logs products.txt file

```
1X1,Milk          ,11,111     ,0       ,Update row ,19/03/2024 17:39:25
1X1,Milk          ,11,111     ,0       ,Update row ,19/03/2024 17:39:25
2X2,yogurt         ,7,222     ,0       ,Update row ,19/03/2024 17:39:25
2X2,yogurt         ,7,222     ,0       ,Update row ,19/03/2024 17:39:25
P044,coffe        ,12,111      ,0        ,Update row ,19/03/2024 17:39:25
P044,coffe        ,12,111      ,0        ,Update row ,19/03/2024 17:39:25
p342,underwear          ,17,V002     ,3        ,Update row ,19/03/2024 17:39:25
p342,underwear          ,17,V002     ,3        ,Update row ,19/03/2024 17:39:25
p342,underwear          ,17,V002     ,3        ,Update row ,19/03/2024 17:39:25
p999,socks        ,10,V002     ,3         ,Update row ,19/03/2024 17:39:25
p999,socks        ,10,V002     ,3         ,Update row ,19/03/2024 17:39:25
p999,socks        ,10,V002     ,3         ,Update row ,19/03/2024 17:39:25
P002,T-shirt      ,20,V002     ,3         ,Update row ,19/03/2024 17:39:25
P002,T-shirt      ,20,V002     ,3         ,Update row ,19/03/2024 17:39:25
P002,T-shirt      ,20,V002     ,3         ,Update row ,19/03/2024 17:39:25
P777,Pants        ,50,V002     ,3         ,Update row ,19/03/2024 17:39:25
P777,Pants        ,50,V002     ,3         ,Update row ,19/03/2024 17:39:25
P777,Pants        ,50,V002     ,3         ,Update row ,19/03/2024 17:39:25
```

The next step is to deploy the Project

The deployed path : "/SSISDB/Avishay_Integration_Services Project" Into SSMS.

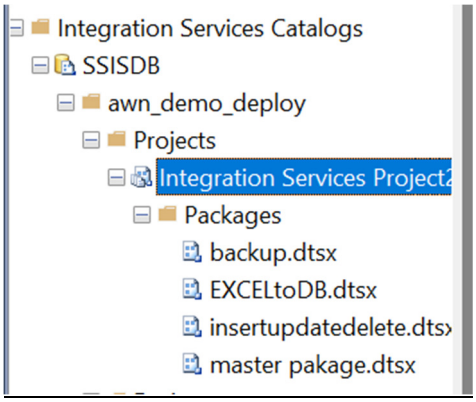
After that I create a job called master package which will call the next job in line until all the jobs will be done.

The jobs are separated into small parts because if an error will occur, we want to detect it in the specific job and be more efficient instead of loading massive amount of data each time.

Action	Result
Converting package backup.dtsx	Passed
Converting package EXCELtoDB.dtsx	Passed
Converting package insertupdatedelete.dtsx	Passed
Converting package master pakage.dtsx	Passed
Mapping package parameters	Passed
Creating project	Passed

Action	Result
Loading project	Passed
Connecting to destination server	Passed
Changing protection level	Passed
Deploying project	Passed

In ssms ssisdb in ssms -

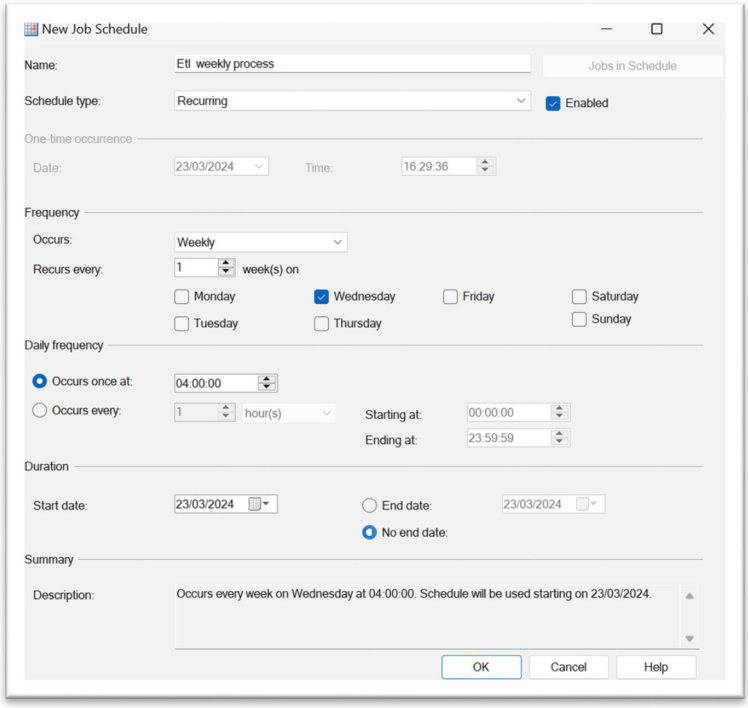


This is the chronological order of the jobs:

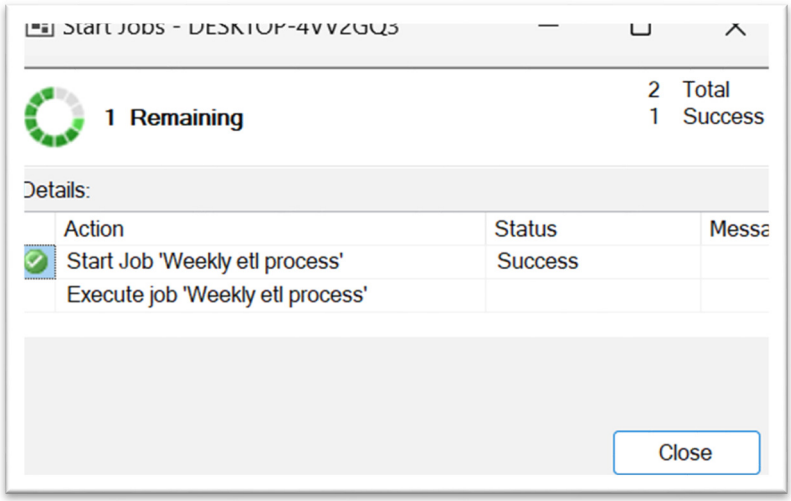
Master pakage -> EXCELtodb -> insertupdatedelete -> backup

Now create a new job - Etl weekly process

All the Extract-Transform-Load process will happen every Wendsday at 04:00 AM.



Check if this job work properly



After running the job we went to check that the tables would be updated in the middle of the log table.

In the LogTable table we have the following columns:

id: Unique identifier for each log entry.

status: Indicates the current status or outcome of the operation.

file name: Name of the file or source involved in the operation.

record insert: Specifies whether a new record was inserted during the operation.

log_date: Timestamp indicating when the log entry was created.

	id	status	file_name	records_insert	log_date
1	1	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	3	2024-03-18 22:01:23.903
2	2	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	6	2024-03-18 22:01:24.097
3	3	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	1	2024-03-18 22:01:24.290
4	4	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	0	2024-03-18 22:04:20.480
5	5	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	0	2024-03-18 22:04:20.640
6	6	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	0	2024-03-18 22:04:20.817
7	7	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:55:53.350
8	8	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:55:53.517
9	9	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:55:53.677
10	10	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:57:06.320
11	11	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:57:06.470
12	12	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:57:06.623
13	13	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:58:53.340
14	14	File loaded	C:\Users\מיכאל\Desktop\ssmsssispowerbi\databases...	100	2024-03-19 18:58:53.487