

410253 (C): Business Intelligence
Group A
Assignment No.: 1

Title of the Assignment:

Import the legacy data from different sources such as (Excel, Sql Server, Oracle etc.) and load in the target system. (You can download sample database such as Adventure works, Northwind, foodmart etc.).

Objective of the Assignment:

To introduce the concepts and components of Business Intelligence (BI).

Outcome:

1. Apply basic principles of elective subjects to problem solving and modeling.
2. Use tools and techniques in the area of software development to build mini projects.

Pre-requisites:

1. Basics of dataset extensions.
2. Concept of data import.

Contents for Theory:

1. Legacy Data
2. Sources of Legacy Data
3. How to import legacy data step by step.

Theory:

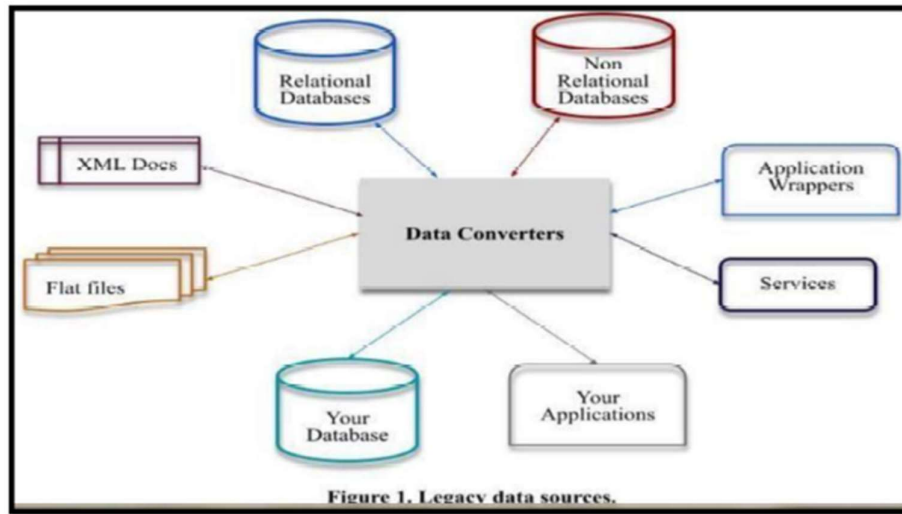
1. What is Legacy Data?

Legacy data, according to Business Dictionary, is "information maintained in an old or out-of-date formator computer system that is consequently challenging to access or handle."

2. Sources of Legacy Data

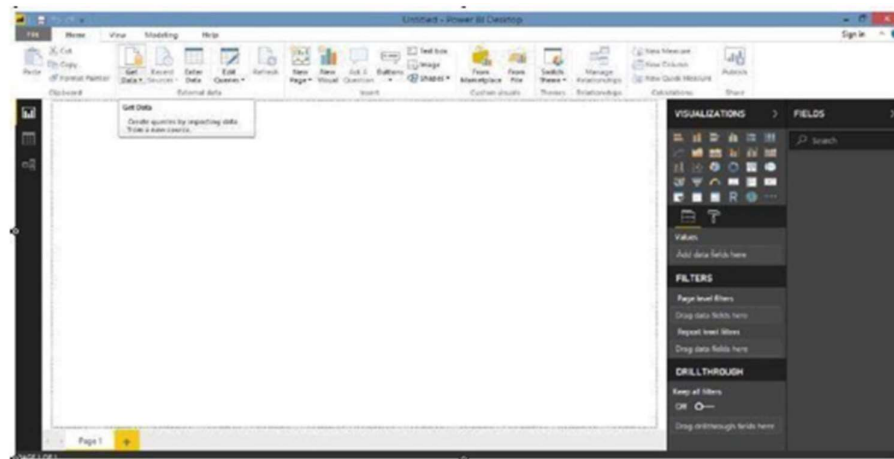
Where does legacy data come from? Virtually everywhere. Figure 1 indicates that there are many sources from which you may obtain legacy data. This includes existing databases, often relational, although non- RDBs such as hierarchical, network, object, XML, object/relational databases, and NoSQL databases. Files, such as XML documents or "flat files" such as configuration files and comma-delimited text files, are also common sources of legacy data. Software, including legacy applications that have been wrapped (perhaps via CORBA) and legacy services such as web services or CICS transactions, can also provide access to existing

information. The point to be made is that there is often far more to gaining access to legacy data than simply writing an SQL query against an existing relational database.

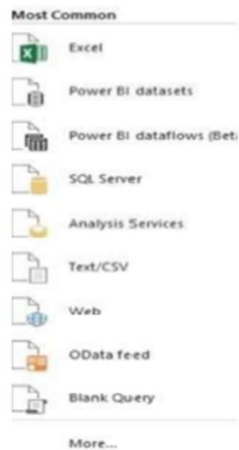


How to import legacy data step by step:

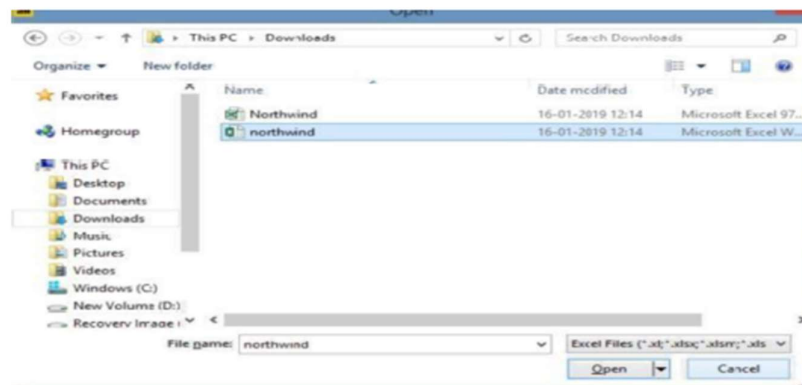
Step 1: Open Power BI



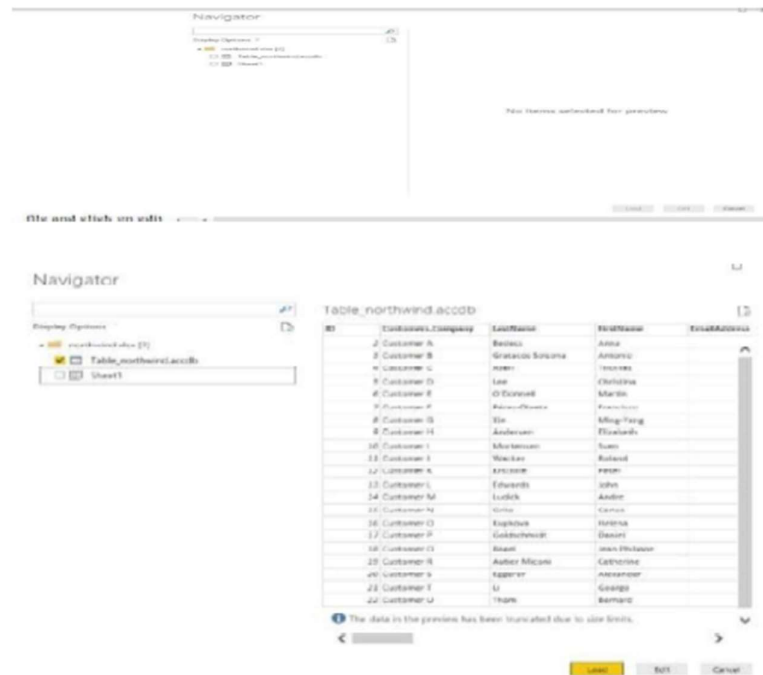
Step 2: Click on Get data following list will be displayed -> select Excel



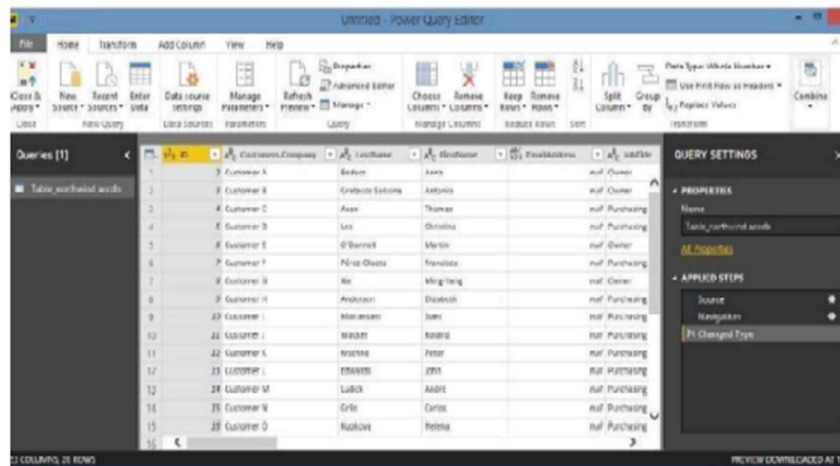
Step 3: Select required file and click on Open, Navigator screen appears



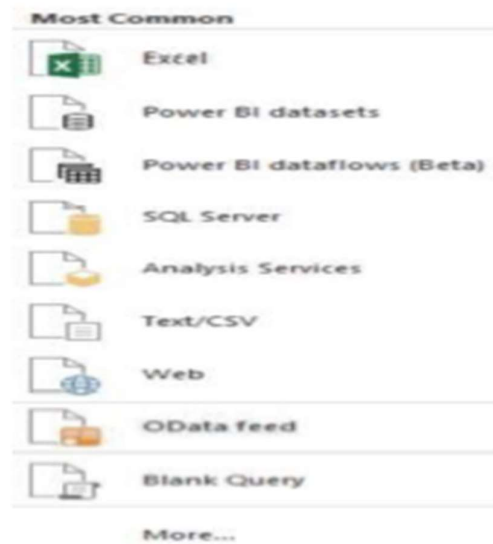
Step 4: Select file and click on edit



Step 5: Power query editor appears

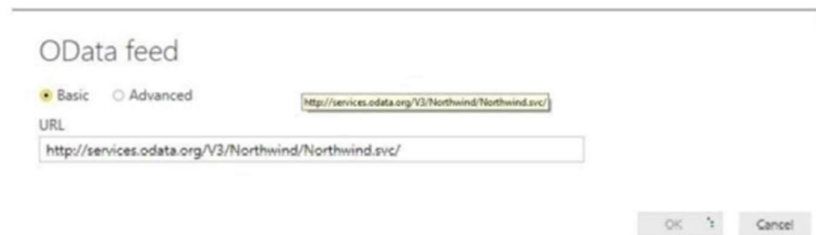


Step 6: Again, go to Get Data and select OData feed:

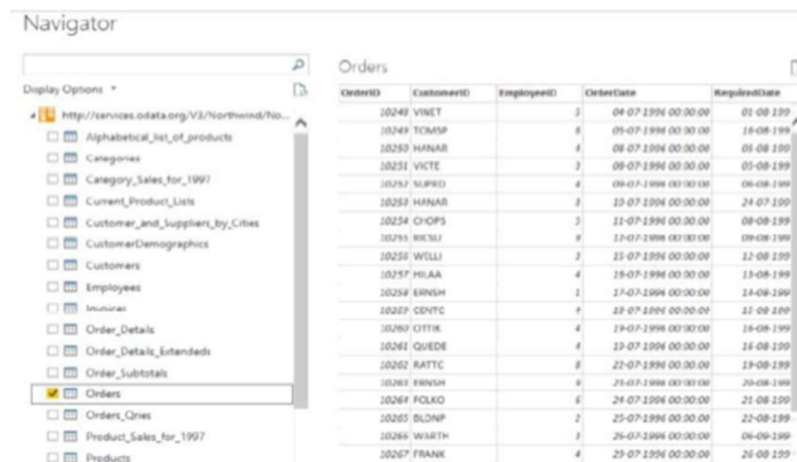


Step 7: Paste URL as

<http://services.odata.org/V3/Northwind/Northwind.svc/> Click on ok



Step 8: Select orders table And click on edit Note: If you just want to see preview you can just click on table name without clicking on checkbox Click on edit to view table.



Untitled - Power Query Editor

File Home Add Column View Help

Get Data Data Source Settings Manage Parameters Refresh Preview Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Data Type: Whole Number Use First Row as Headers Replace Values Combine

	OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate	ShipDate
1	10248	VINET	5	04-07-1996 00:00:00	01-08-1996 00:00:00	16-07-1996 00:00:00
2	10249	TOMSP	6	05-07-1996 00:00:00	16-08-1996 00:00:00	17-07-1996 00:00:00
3	10250	HANAR	4	08-07-1996 00:00:00	05-09-1996 00:00:00	12-07-1996 00:00:00
4	10251	VICTE	3	08-07-1996 00:00:00	05-08-1996 00:00:00	15-07-1996 00:00:00
5	10252	SUPRO	4	09-07-1996 00:00:00	08-08-1996 00:00:00	12-07-1996 00:00:00
6	10253	HANAR	3	10-07-1996 00:00:00	24-07-1996 00:00:00	16-07-1996 00:00:00
7	10254	CHOPS	5	11-07-1996 00:00:00	08-08-1996 00:00:00	21-07-1996 00:00:00
8	10255	RICSU	3	12-07-1996 00:00:00	09-08-1996 00:00:00	17-07-1996 00:00:00
9	10256	WELLI	3	15-07-1996 00:00:00	12-08-1996 00:00:00	17-07-1996 00:00:00
10	10257	MILAA	4	28-07-1996 00:00:00	13-08-1996 00:00:00	22-07-1996 00:00:00
11	10258	ERIND	1	27-07-1996 00:00:00	14-08-1996 00:00:00	23-07-1996 00:00:00
12	10259	CENTC	4	18-07-1996 00:00:00	15-08-1996 00:00:00	25-07-1996 00:00:00
13	10260	OTTIR	4	29-07-1996 00:00:00	14-08-1996 00:00:00	29-07-1996 00:00:00
14	10261	QUJED	4	19-07-1996 00:00:00	16-08-1996 00:00:00	03-08-1996 00:00:00
15	10262	RATTC	8	22-07-1996 00:00:00	19-08-1996 00:00:00	25-07-1996 00:00:00

QUERY SETTINGS

PROPERTIES

Name

OrderID

Applied Steps

Source

Navigation

PREVIEW DOWNLOADED AT 1

Conclusion:

In this way we import the Legacy datasets using the Power BI Tool.

Assignment No.: 2

Title of the Assignment:

Perform the Extraction Transformation and Loading (ETL) process to construct the database in the Sqlserver.

Objective of the Assignment:

To introduce the concepts and components of Business Intelligence (BI).

Outcome:

1. Apply basic principles of elective subjects to problem solving and modeling.
2. Use tools and techniques in the area of software development to build mini projects.

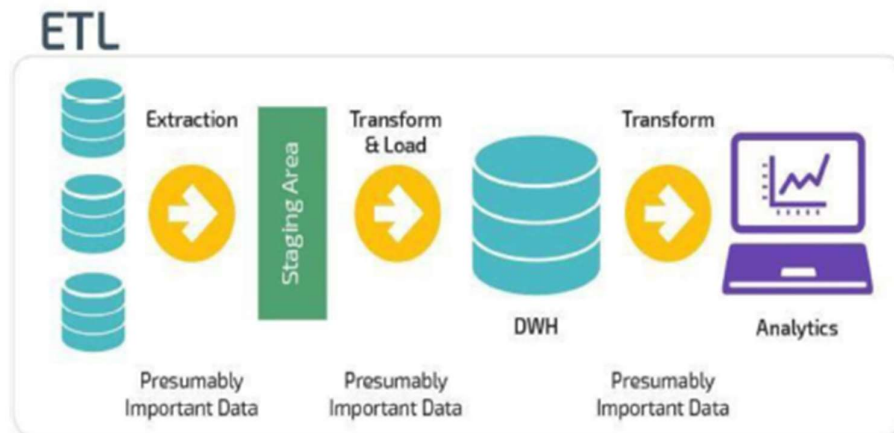
Pre-requisites:

1. Basic of ETL Tools.
2. Concept of Sql Server.

Theory:

ETL (Extract, Transform and Load):

ETL is a process in Data Warehousing and it stands for Extract, Transform and Load.



It is a process, in which an ETL tool extracts the data from various data source systems, transforms it in the staging area and then finally, loads it into the Data Warehouse system.

Extraction:

1. Identify the Data Sources: The first step in the ETL process is to identify the data sources. This may include files, databases, or other data repositories.

2. Extract the Data: Once the data sources are identified, we need to extract the data from them. This may involve writing queries to extract the relevant data or using tools such as SSIS to extract data from files or databases.

3. Validate the Data: After extracting the data, it's important to validate it to ensure that it's accurate and complete. This may involve performing data profiling or data quality checks.

Transformation:

1. Clean and Transform the Data: The next step in the ETL process is to clean and transform the data. This may involve removing duplicates, fixing invalid data, or converting data types. We can use tools such as SSIS or SQL scripts to perform these transformations.

2. Map the Data: Once the data is cleaned and transformed, we need to map the data to the appropriate tables and columns in the database. This may involve creating a data mapping document or using a tool such as SSIS to perform the mapping.

Loading:

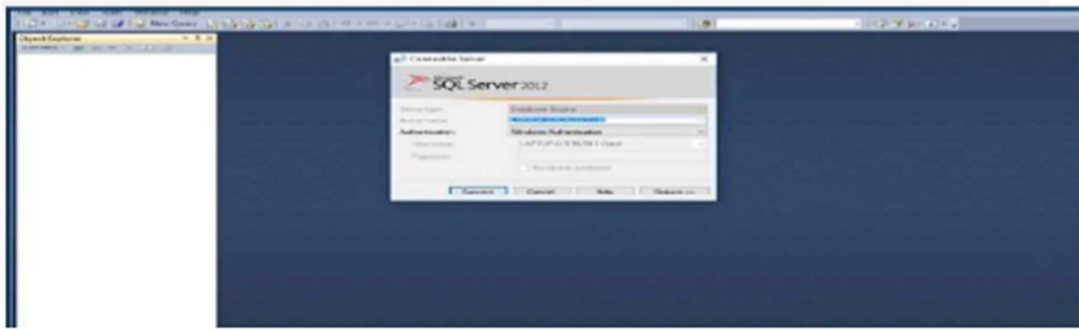
1. Create the Database: Before loading the data, we need to create the database and the appropriate tables. This can be done using SQL Server Management Studio or a SQL script.

2. Load the Data: Once the database and tables are created, we can load the data into the database. This may involve using tools such as SSIS or writing SQL scripts to insert the data into the appropriate tables.

3. Validate the Data: After loading the data, it's important to validate it to ensure that it was loaded correctly. This may involve performing data profiling or data quality checks to ensure that the data is accurate and complete.

- Perform the Extraction Transformation and Loading (ETL) process to construct the database in the SQL server.
- Software requirements: SQL SERVER 2012 FULL VERSION (SQLServer2012SPI-FullSlipstream-ENU-x86)
- Steps to install SQL SERVER 2012 FULL VERSION (SQL Server2012SPI-FullSlipstream-ENU-x86) are given in my previous post.

Step 1: Open SQL Server Management Studio to restore backup file.



Step 2: Right click on Databases Restore Database

Step 3: Click on towards end of device box

Step 4: Click on Add Select path of backup files

Step 5: Select both files at a time

Step 6: Click ok and in select backup devices window Add both files of Adventure Works

Step 7: Open SQL Server Data Tools Select File New Project Business Intelligence Integration Services Project & give appropriate project name.

Step 8: Right click on Connection Managers in solution explorer and click on New Connection Manager. Add the SSIS connection manager window.

Step 9: Select OLEDB Connection Manager and Click on Ad

Step 10: Configure OLE DB Connection Manager window appears Click on New

Step 11: Select Server name (as per your machine) from drop down and database name and click on Test connection. If the test connection succeeded, click on OK.

Step 12: Click on OK. Connection is added to connection manager

Step 13: Drag and drop Data Flow Task in Control Flow tab

Step 14: Drag OLE DB Source from Other Sources and drop into Data Flow tab

Step 15: Double click on OLE DB source -> OLE DB Source Editor appears->click on New to add connection manager.

Select [Sales].[Store] table from drop down ok

Step 16: Drag ole db destination in data flow tab and connect both

Step 17: Double click on OLE DB destination Click on New to run the query to get [OLE DB Destination] in Name of the table or the view. Click on OK.

Step 18: Click on Start

Step 19: Go to SQL Server Management Studio In database tab Adventure works Right click on [dbo].[OLE DB Destination]Script Table as SELECT To New Query Editor Window

Step 20: Execute the following query to get output.

```
USE [AdventureWorks2012]GO
```

```
SELECT [BusinessEntityID]
```

```
,[Name]
```

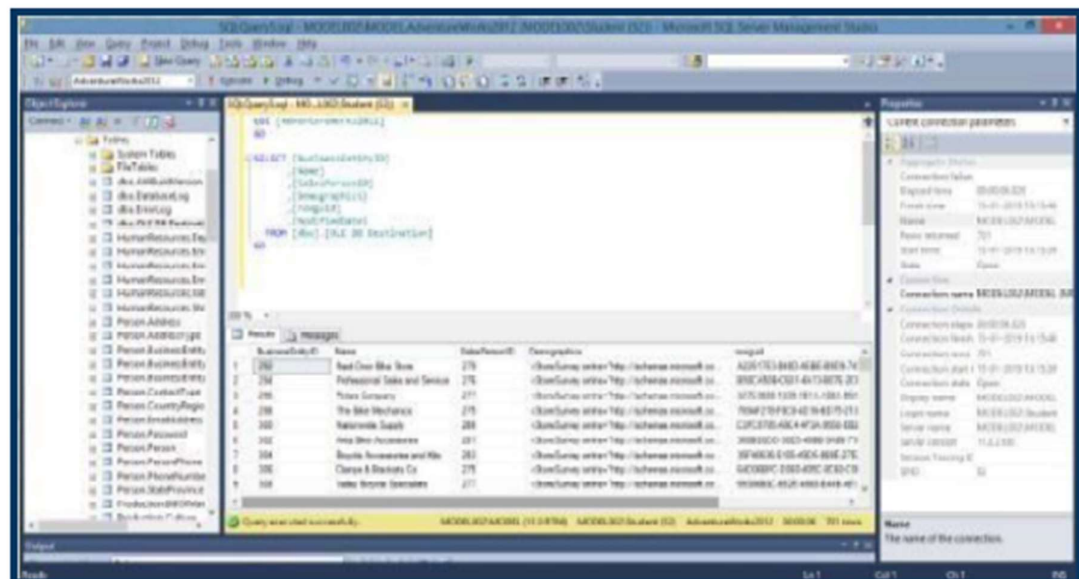
```
,[SalesPersonID]
```

```
,[Demographics]
```

```
[rowguid]
```

```
[ModifiedDate]
```

```
FROM [dbo].[OLE DB Destination]GO
```



Assignment No.: 3

Title of the Assignment:

Create the cube with suitable dimension and fact tables based on ROLAP, MOLAP and HOLAP model.

Objective of the Assignment:

To introduce the concepts and components of Business Intelligence (BI).

Outcome:

1. Apply basic principles of elective subjects to problem solving and modeling.
2. Use tools and techniques in the area of software development to build mini projects.

Pre-requisites:

1. Basic of OLAP.
2. Concept of Multi-Dimensional Cube.

Theory:

1. What is Fact Table?

In Business Intelligence (BI), a Fact Table is a table that stores quantitative data or facts about a business process or activity. It is a central table in a data warehouse that provides a snapshot of a business at a specific point in time.

For example – A Fact Table in a retail business might contain sales data for each transaction, with dimensions such as date, product, store, and customer. Analysts can use the Fact Table to analyse trends and patterns in sales, such as which products are selling the most, which stores are performing well, and which customers are buying the most.

2. What is a ROLAP, MOLAP and HOLAP model:

ROLAP, MOLAP, and HOLAP are three types of models used in Business Intelligence (BI) for organizing and analysing data:

a. ROLAP (Relational Online Analytical Processing):

In this model, data is stored in a relational database, and the analysis is performed by joining multiple tables. ROLAP allows for complex queries and is good for handling large amounts of data, but it may be slower due to the need for frequent joins.

b. MOLAP (Multidimensional Online Analytical Processing):

In this model, data is stored in a multidimensional database, which is optimized for fast query performance. MOLAP is good for analysing data in multiple dimensions, such as time, geography, and product, but may be limited in its ability to handle large amounts of data.

c. HOLAP (Hybrid Online Analytical Processing):

This model combines elements of both ROLAP and MOLAP. It stores data in both a relational and multidimensional database, allowing for efficient analysis of both large amounts of data and complex queries. HOLAP is a good compromise between the other two models, offering both speed and flexibility.

3. Create the cube with a suitable dimension and fact tables based on OLAP?

Step 1: Creating Data Warehouse:

Let us execute our T-SQL Script to create a data warehouse with fact tables, dimensions and populate them with appropriate test values.

Download the T-SQL Script attached with this article for creating of Sales Data Warehouse or download from this article “Create First Data Warehouse” and run it in you SQL Server.

Downloading “Data_WareHouse_SQLScript.zip” from the article <https://www.codeproject.com/Articles/652108/Create-First-Data-WareHouse>





After downloading the extract file in the folder.

Follow the given steps to run the query in SSMS (SQL Server Management Studio).

1. Open SQL Server Management Studio 2012
2. Connect Database Engine

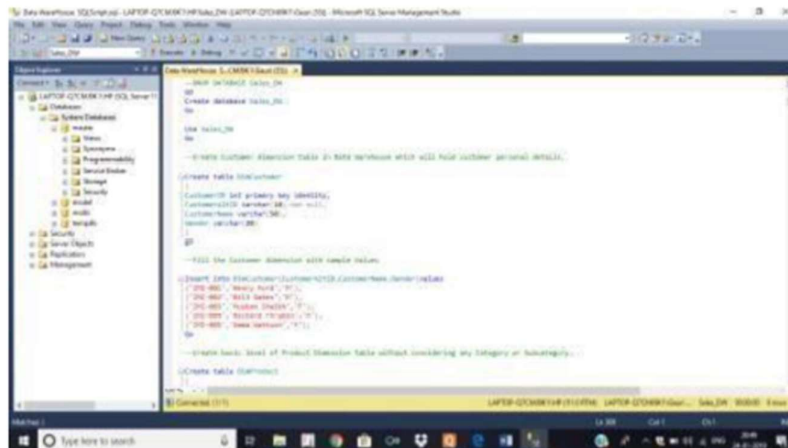


Password for sa: admin123 (as given during installation) Click Connect.

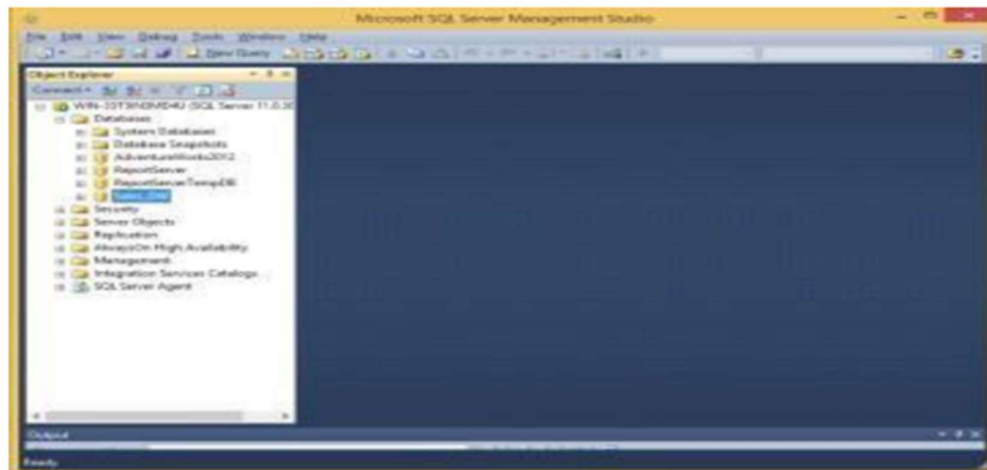
3. Open New Query editor
4. Copy paste Scripts given below in various steps in new query editor window one by one
5. To run the given SQL Script, press F5
6. It will create and populate "Sales_DW" database on your SQL Server OR.

7. Go to the extracted sql file and double click on it.

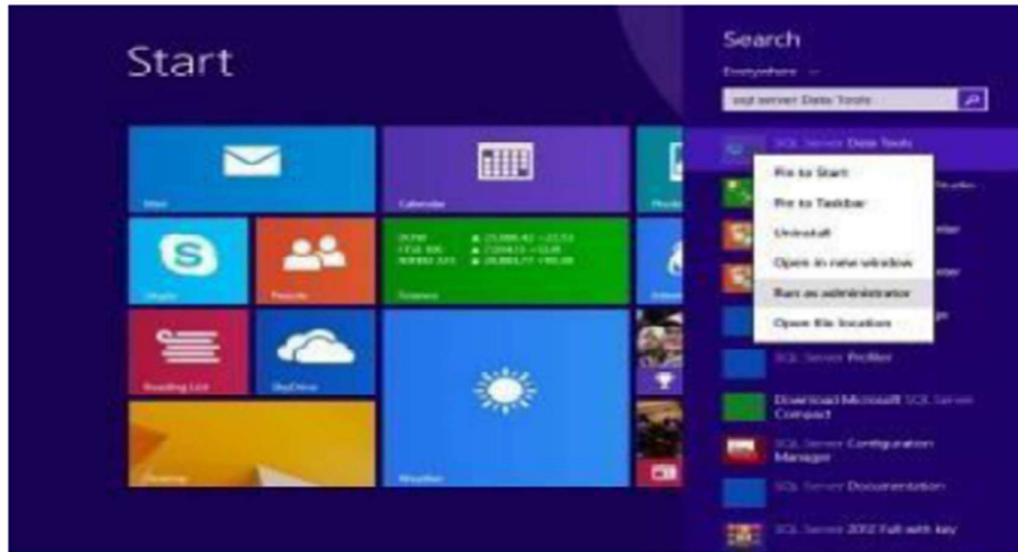
8. New Sql Query Editor will be opened containing the Sales_DW Database.



9. Click on execute or press F5 by selecting the query one by one or directly click on Execute. After completing execution save and close SQL Server Management studio & reopen to see Sales_DW in Databases Tab.



Step 2: Start SSDT environment and create New Data Source Go to Sql Server Data Tools --> Right click and run as administrator.



- Click on File -> New -> Project
- In Business Intelligence -> Analysis Services Multidimensional and Data Miningmodels - appropriate project name -> click OK
- Right click on Data Sources in solution explorer -> New Data Source Data Source Wizard appears
- Click on New
- Select Server Name - select Use SQL Server Authentication - Select or enter a database name (Sales_DW) Note: Password for sa : admin123 (as given during installation of SQL 2012 fullversion)
- Click Next
- Select Inherit -> Next
- Click Finish
- Sales_DW.ds gets created under Data Sources in Solution Explore

Step 3: Creating New Data Source View In Solution explorer right click on Data Source View -> Select New Data Source View.

- Click Next
- select FactProductSales(dbo) from Available objects and put in Includes Objects by clicking
- Click Next
- Click Finish

- Sales DW.dsv appears in Data Source Views in Solution Explorer.

Step 4: Creating new cube:

- Right click on Cubes - New Cube
- Select Use existing tables in Select Creation Method - Next
- In Select Measure Group Tables - Select FactProductSales - Click Next
- In Select Measures - check all measures -> Next
- In Select New Dimensions -> Check all Dimensions – Next
- Click on Finish
- Sales_DW.cube is created

Step 5: Dimension Modification

- In dimension tab -> Double Click Dim Product.dim
- Drag and Drop Product Name from Table in Data Source View and Add in AttributePane at left side

Step 6: Creating Attribute Hierarchy in Date Dimension:

- Double click On Dim Date dimension -> Drag and Drop Fields from Table shown in Data
- Source View to Attributes-> Drag and Drop attributes from leftmost pane of attributes to middle pane of Hierarchy.
- Drag fields in sequence from Attributes to Hierarchy window (Year, Quarter Name, Month, Name, Week of the Month, Full Date UK)

Step 7: Deploy Cube:

- Right click on Project name – Properties
- This window appears
- Do following changes and click on Apply & ok
- Right click on project name -> Deploy
- Deployment successful
- To process cube right click on Sales_DW.cube -> Process
- Click run
- Browse the cube for analysis in solution explorer

Conclusion: In this way we successfully implement cube with suitable dimension and fact tables based on ROLAP, MOLAP and HOLAP model.

Assignment No.: 4

Title of the Assignment:

Import the data warehouse data in Microsoft Excel and create the Pivot table and Pivot Chart.

Objective of the Assignment:

To introduce the concepts and components of Business Intelligence (BI).

Outcome:

1. Apply basic principles of elective subjects to problem solving and modeling.
2. Use tools and techniques in the area of software development to build mini projects.

Pre-requisites:

1. Basic of Google Sheets.
2. Concept of Table, Chart.

Contents for Theory:

1. What is a Data Warehouse?
2. What is Pivot Table and Pivot Chart?
3. Steps for Creating a Pivot Table in Google Sheets.
4. Steps for Creating a Pivot Chart in Google Sheets.

Theory:

1. What is a Data Warehouse?

A data warehouse is a centralized repository of integrated and transformed data from multiple sources within an organization. It is designed to support business intelligence (BI) activities, such as data analysis, reporting, and decision-making.

2. What is Pivot Table and Pivot Chart?

A pivot table is a powerful tool in spreadsheet software (such as Google Sheets or Microsoft Excel) that allows you to summarize and analyse large datasets by grouping and summarizing data in different ways. Pivot tables allow you to quickly create tables that show a summary of data based on specific criteria or dimensions. For example, you can use a pivot table to summarize sales data by region or by product category. A pivot chart is a graphical representation of the data in a pivot table. Pivot charts allow you to visualize the summarized data in a way that is easy to understand and interpret. They can be created based on the data in a pivot table, and can be customized in a variety of ways to better represent the data being

analysed. Pivot charts are especially useful when dealing with large amounts of data, as they can help identify patterns and trends that might not be immediately obvious from the raw data.

3. Steps for Creating a Pivot Table in Google Sheets.

1. Open a Google Sheets document with the data you want to use for the pivot table.
2. Select the range of data you want to use for the pivot table.
3. Click on the "Data" tab in the top menu, then click on "Pivot table."
4. In the "Create Pivot Table" dialog box, select the range of data you want to use for the pivot table and choose where you want to place the pivot table (in a new sheet or in the same sheet).
5. Click on "Create."
6. In the pivot table editor, drag and drop the columns you want to use for the pivot table into the "Rows," "Columns," and "Values" sections.
7. To add a filter to the pivot table, drag a column into the "Filter" section.
8. To customize the values in the pivot table, click on the drop-down menu in the "Values" section and choose the type of calculation you want to use (such as sum, count, or average).
9. Customize any additional options in the pivot table editor (such as sorting and formatting).
10. Click on "Update" to apply the changes and create the pivot table

4. Steps for Creating a Pivot Chart in Google Sheets.

1. Open a Google Sheets document with the data you want to use for the pivot chart.
2. Select the range of data you want to use for the pivot chart.
3. Click on the "Data" tab in the top menu, then click on "Pivot table."
4. In the "Create Pivot Table" dialog box, select the range of data you want to use for the pivot table and choose where you want to place the pivot table (in a new sheet or in the same sheet).
5. Click on "Create."
6. In the pivot table editor, drag and drop the columns you want to use for the pivot chart into the "Rows" and "Values" sections.
7. Click on the "Chart" tab in the pivot table editor.
8. Choose the type of chart you want to use for the pivot chart from the drop Down menu.

9. Customize the chart options (such as chart title, axis labels, and color) to Your liking.
10. Click on "Update" to apply the changes and create the pivot chart.

Conclusion: In this way we pivot table and pivot chart using Google spreadsheets Excel.

Assignment No.: 5

Title of the Assignment:

Perform the data classification using classification algorithm. Or perform the data clustering using a clustering algorithm.

Objective of the Assignment:

To introduce the concepts and components of Business Intelligence (BI).

Outcome:

1. Apply basic principles of elective subjects to problem solving and modeling.
2. Use tools and techniques in the area of software development to build mini projects.

Pre-requisites:

1. Basic of Tableau.

Contents for Theory:

1. What is Clustering and classification?
2. Clustering in Tableau:
3. Classification in Tableau:

Theory:

1. What is Clustering and classification?

Clustering and classification are two important techniques used in bioinformatics to analyse biological data. Clustering is the process of grouping similar objects or data points together based on their similarity or distance from each other. In bioinformatics, clustering is often used to group Genes or proteins based on their expression patterns or sequences. Clustering can help identify patterns and relationships between different genes or proteins, which can provide insights into their biological function and interactions. Classification, on the other hand, is the process of assigning a label or category to a new observation based on its features or characteristics. In bioinformatics, classifications often used to predict the function or activity of a new gene or protein based on its sequence or structure. Classification can help identify new drug targets or biomarkers for disease diagnosis and treatment. Both clustering and classification are important tools for analysing large and complex biological datasets and can provide valuable insights into the underlying biological processes.

2. Clustering in Tableau:

1. Connect to the data: Connect to the data set that you want to cluster in Tableau.
2. Drag and drop the data fields: Drag and drop the data fields into the view, and select the data points that you want to cluster.
3. Choose a clustering algorithm: Select a clustering algorithm from the analytics pane in Tableau. Tableau provides several built-in clustering algorithms, such as K-Means and Hierarchical Clustering.
4. Define the number of clusters: Define the number of clusters that you want to create. You can do this manually or let Tableau automatically determine the optimal number of clusters.
5. Analyse the clusters: Visualize the clusters and analyse them using Tableau's built- in visualizations and tools.

3. Classification in Tableau:

1. Connect to the data: Connect to the data set that you want to classify in Tableau.
2. Drag and drop the data fields: Drag and drop the data fields into the view, and select the target variable that you want to predict.
3. Choose a classification algorithm: Select a classification algorithm from the analytics pane in Tableau. Tableau provides several built-in classifications algorithms, such as Decision Trees and Random Forest.
4. Define the model parameters: Define the model parameters, such as the maximum tree depth or the number of trees to use in the forest.
5. Train the model: Train the model on a subset of the data using Tableau's built-in cross-validation functionality.
6. Evaluate the model: Evaluate the accuracy of the model using Tableau's built-in metrics, such as confusion matrix, precision, recall, and F1 score.
7. Predict the target variable: Use the trained model to predict the target variable for new data.
8. Visualize the results: Create visualizations to communicate the results of the classification analysis using Tableau's built-in visualizations tools.

Conclusion: In this way we implement classification and clustering using Tableau.