



## Assignment No - 1

Title :- Import the legacy data from different sources such as (Excel, sqlserver, oracle) and load in target system.

### Theory :-

Power BI :-

Microsoft Power BI is a BI platform that provides non-technical business users with tools for aggregating, analyzing, visualizing and sharing data. Power BI's user interface is fairly intuitive for users familiar with Excel, and its deep integration with other Microsoft products makes it a versatile self-service tool that requires little upfront training.

### Common Uses of Power BI :-

Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into data model and create charts or graphs to provide visuals of data. All of this can be shared with other Power BI users within organization.

Power BI can also provide executive dashboards for administrators or managers, giving management more insight into how departments are doing.

### Who uses Power BI?

Though Power BI is a self-service BI tool that brings data analytics to employees, it's mostly used by data analysts and BI professionals who create data models before disseminating reports throughout the organization.

## Key features of power BI :-

Microsoft has added a number of data analytics features to power BI since its inception, and continues to do so. Some of the most important features are following

- 1) Artificial Intelligence :- Users can access image recognition and text analytics in power BI, create ML models using automated ML capabilities and integrate with Azure ML.
- 2) Quick Insights :- This feature allows user to create subsets of data and automatically apply analytics that information.
- 3) customization :- This feature allows developers to change the appearance of default visualization and reporting tools and import new tools into the platform.
- 4) APIs for Integration :- This feature provides developers with sample code and application program interfaces.
- 5) Self-service data prep :- Using power query, business analysts can ingest, transform, integrate and enrich big data into the power BI web service.

## Power BI components :-

The different components of power BI are meant to let users create and share business insights in a way that fits with their role

- power query - a data mashup and transformation tool.
- power pivot - a memory tabular data modeling tool.
- power view - a data visualization tool.
- power map - a 3D geospatial data visualization tool.
- power Q&A - a natural language question & answering engine.



### How to use power BI :-

Power BI Desktop is where analysts and other users can create data connections, data models and reports. The Power BI service is where those reports can be shared, so other users can view and interact with them.

To build a Power BI report, take the following steps:-

- connect the data sources.
- Query the data to create reports based on user needs
- Publish the reports to the Power BI Service
- Share the report, so cloud and mobile users can see and interact with it.
- Add permissions to give colleagues the ability to edit reports or create dashboards or limit their ability to edit.

### Conclusion :-

We successfully import the legacy data from different sources and loaded it in target system

## Assignment No - 2

Title :- perform the Extraction transformation and Loading (ETL) process to construct the database in the sqlserver

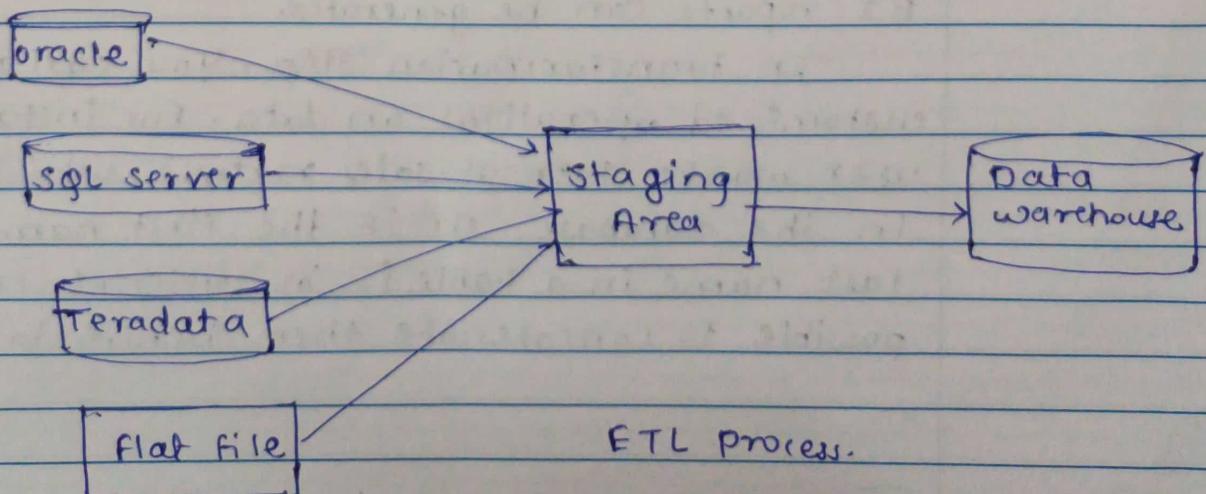
Theory :-

ETL is a process that extracts the data from different source systems, then transforms the data and finally loads the data into the Data Warehouse system. full form of ETL is Extract, Transform and Load

It's tempting to think a creating a data warehouse is simply extracting data from multiple source and loading into database of Data warehouse. This is far from the truth and requires a complex ETL process.

In order to maintain its value as a tool for decision-makers, Data warehouse system needs to change with business changes.

ETL is 3-step process :-





### Step 1) Extraction :-

In this step of ETL architecture, data is extracted from the source system into the staging area. Transformation if any are done in staging area so that performance of source system is not degraded.

#### Three Data extraction methods :-

- 1) Full Extraction
- 2) Partial Extraction - without update notification
- 3) Partial Extraction - with update notification

Some Validations are done using Extraction.

### Step 2) Transformation :-

Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleaned, mapped and transformed. In fact, this is the key step where ETL process add value and changes data such that insightful BI reports can be generated.

In transformation step, you can perform customized operations on data. For instance, if the user wants sum-of-sales revenue which is not in the database. Or if the first name and the last name in a table is in different columns. It is possible to concatenate them before loading.



### Step 3) Loading :-

Loading data into target datawarehouse database is the last of ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in relative short period.

In case of load failure, recover mechanism should be configured to restart from the point of failure without data integrity loss.

#### Types of loading :-

- 1) Initial load -
- 2) Incremental Load -
- 3) Full Refresh -

#### Conclusion :-

We learned most important process of Extraction Transformation and loading to construct the database in the sqlserver in this practical.

## Assignment No - 3

Title :-

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

Theory :-

Multidimensional cubes and transactional databases are two very different things. From experience I have seen even veteran DBAs avoid the subject of cubes completely because it is too much of an unknown area for them. This article assumes you already have a data warehouse and uses Adventureworks as an example. Following these steps should put you on the road to decent SQL Server business intelligence solution based on a read-optimized OLAP cube.

Getting started -

Every OLAP cube is created with an analysis solution. To create one, open SQL Server Data Tools or BIDS and create a new Analysis Services multidimensional and Data mining project.

This will open a blank solution and will create the basic folders needed in Solution Explorer. Something that is quite handy to know about SQL Server Analysis projects is that the order of creation or the workflow follows the order of the folders in the Solution Explorer.



So let's start by creating a new data source. To do this, right click on Data source and click "New Data source". This will open a wizard that will allow you to make a normal database connection to your data warehouse.

Having completed this step you now have a source of data for your OLAP cube. Next you have to create a data source view and specify which data you want to use for your cube. The wizard will ask you which connection you would like to use and which fact.

It is important to have correct foreign key relations in your data warehouse beforehand because SQL Server Data Tools is able to create a decent view diagram using existing relations in the DW. If you have not done this you will have to explicitly specify relations.

You are nearly done. The basic configuration is complete but your cube only exists as a blueprint at this stage. In order to create it on your Analysis Services server you first need to tell Data Tools where your server is located.

To create the cube and process the data from the data warehouse to the new cube you need to click Build > Process. Next a "Process cube" screen will appear. At this screen you can leave everything as default and click "Run..." and if you have no configuration errors your cube will process.

Once you have processed the cube for the first time you can also bring up the same screen by opening SQL Server Management Studio and connecting to your Analysis server and right clicking on cube. There



you can choose "Explore".

on the screen you can verify the data in the cube and also generate basic queries to use in your reports. The reports also have this query designer window if you choose an analysis source instead of normal SQL server database source. The queries created are in MDX format which is a whole other language.

You can get a feel for it on the MSDN website. However, the subject of creating SSRS reports based on MDX queries will be covered in later article

Conclusion :-

In this practical we created the cube with suitable dimension and fact tables based on ROLAP, MOLAP, & HOLAP model.



## Assignment No - 4

Title :- Import the data warehouse data in Microsoft Excel and create the Pivot table & Pivot.

Theory :-

Steps :-

- 1) Open Excel 2013 (Professional)  
Go to Data tab → Get External Data → From other sources  
→ From Data connection Wizard
- 2) In Data connection Wizard → Select Microsoft SQL Server → Click on Next.
- 3) In Connect to Database server provide server name  
provide password for sa account as given during installation of SQL server. Click on Next.
- 4) In Selected Database and Table → Select Sales-DW  
(already created in SQL) → Check all dimensions and import relationships between selected tables.
- 5) In save data connection files browse path and click on finish.
- 6) In import data select Pivot chart and click on ok
- 7) In fields put salesDatekey in filters, fullDateUK in axis and sum of productActual cost in values.

8) In Insert Tab → go to Pivot table.

9) click on choose connection to select existing connection with sales-DW and click on open

→ after connection click with shortcut F5 to refresh  
connection & it will show data

By following all the above steps, we

successfully created the pivot table and pivot chart.



## Assignment No - 5

Title :- perform the data classification using classification algorithm or perform the data clustering using clustering algorithm

Theory :-

Classification -

Machine learning applies automatic data-driven learning methods to obtain accurate predictions from observations with previous data.

Data automated classification is one of the aim of machine learning. We can consider three types -

- supervised classification - We have a dataset that we will training data where each datum is associated with a label. In the training stage, we build a model with this dataset using the labels, that helps us to assess the correct or incorrect classification of an image while building a model.
- Unsupervised classification - The dataset comes without labels and the data is classified using their inner structure.
- semi-supervised classification - We can apply it when some data comes with labels, but not all of them. These algorithms can be considered a variant of the supervised classification with a strategy to overcome the lack of labels for part of the data.



### k-means algorithm :-

K-means is an unsupervised classification algorithm, also called clusterization, that groups objects into k groups based on their characteristics. The grouping is done minimizing the sum of the distances between each object and the group or cluster centroid.

Algorithm has three steps:-

- 1) Initialization - once the number of groups k, has been chosen, k centroids are established in data space, for instance, choosing them randomly.
- 2) Assignment of objects to the centroids - each object of the data is assigned to its nearest centroid
- 3) Centroids update - The position of centroid of each group is updated taking as the new centroid the average position of object belonging to said group

### \* Clustering algorithms -

Clustering algorithms can be classified into two main categories linear clustering algorithms & Non-linear.

- Linear clustering algorithm -
  - k-means
  - Fuzzy c-means
  - Gaussian (EM)
  - Hierarchical
  - Quality threshold.



- Non-linear clustering algorithm -
  - MST based clustering algorithm
  - kernel k-means clustering algorithm.
  - Density-based clustering algorithm.

- Application of clustering algorithm -

- Recommender system
- Anomaly detection
- Human genetic clustering.
- search result grouping
- Grouping of shopping items.
- crime analysis
- climatology.

Conclusion :-

- classification - predicting target class for test dataset from the trained modeled from the training dataset.
- clustering - Using different similarity measure to place the all similar items in a group.