

NAME:- AFNAN ATTAR PRN:- F19112003 CLASS:- BE
 COMP-II SUBJECT:- LP VI [BI] ASSIGNMENT :- 01

q1) What do you understand by BI?

Ans. 1. BI stands for business intelligence.

2. It is a technology driven method that helps users analyze and visualize business data from thousands of data sources and share insights across their organizations.

3. It also provides other information that may help corporate executives, business managers and other users make business decision.

4. Common functions of BI include but are not limited to data mining, process mining, reporting, online analytical processing, prescriptive analysis.

q2) What is Power BI?

Ans. 1. Power BI is a Business Analytics solution or a Business Intelligence and Data Visualization tool.

2. It is developed by Microsoft in the year 2014.
 3. It enables users to convert data from thousands of data sources into interactive dashboards and BI reports.

4. It is a cloud-based self service BI tool that collects application, multiple software connectors and services used to organize raw data into important content.

q3) What are the essential applications that use Power BI?

Ans Microsoft's Power BI is mostly used by the following applications.

Page No.	
Date	

- 1) Business & Data Analyst
- 4) Data Scientists
- 2) Project Management Office (PMO)
- 5) IT Team / Professionals
- 3) Developers & Database Admin.
- 6) Consumers for final report.

Q4) In how many formats Power BI is available in the market?

Ans It is available in 3 formats:-

- 1) Power BI desktop:- For desktop computers.
- 2) Power BI service:- Online SaaS.
- 3) Mobile Power BI apps:- For Android/iOS devices.

Q5) What do you mean by the term Power BI Desktop?

- Ans 1. Power BI Desktop is a free business intelligence app.
- 2. It is used to install on the computer system to work with Power BI services.
- 3. It offers shaping, modelling, data exploration, and report creation using high interactive visualization.
- 4. It allows us to save all our work to a file to publish reports and data to the Power BI site so that we can share it to others too.

NAME:- AFNAN ATTAR PRN:- F19112003 CLASS:- BE COMP II
 SUBJECT:- LPVI [BI] ASSIGNMENT NO. :- 02

Q1) Explain how ETL works?

Ans let us see what happens in each step of the process:-

I) Extraction:-

1. During data extraction, raw data is copied or exported from source location to a staging area.
2. Data management teams can extract data from various data sources that can be structured or unstructured.
3. Sources include but are not limited to:-
 - SQL or NoSQL servers.
 - CRM and ERP systems
 - Flat Files, Emails, Web Pages.

II) Transform:-

1. In this process data is transformed and consolidated for its intended analytical use case.
2. It may involve:-
 - Filtering, cleaning, de-duplicating, validating and authenticating data.
 - Performing calculations, translations or summarization based on raw data.
 - Conducting audits to ensure data quality & compliance.
 - Removing, encrypting, or protecting data governed by industrial or governmental regulators.
 - Formatting the data into tables or joined tables to match the schema of the target warehouse.

III) Load:-

1. In this step the transformed data is moved from staging area into a target data warehouse.
2. It involves initial loading of all data followed by periodic loading of incremental data changes, and, less often, full refreshers to erase and replace data in the warehouse.

Q2) What are the benefits and challenges of ETL?

- Ans 1. ETL solutions improve quality by performing data cleansing prior to loading the data to a different repository.
2. A time-consuming batch operation, ETL is recommended for creating smaller target repository more often than not that require less frequent updating while other methods such as:-
 - i) Extract, Load, Transform [ETL]
 - ii) Change Data Capture [CDC]
 - iii) Data Visualizationare used to integrate increasingly larger volumes of data that changes or real-time data streams.

NAME:- AFNAN ATTAR PRN:- F19112003 CLASS:- BE COMP II
 SUBJECT:- BI ASSIGNMENT No.: 03

q1) What do you understand by cube?

- Ans 1.
1. An OLAP cube is a data structure that overcomes the limitations of relational databases by providing rapid analysis of data.
 2. Cubes can display and sum large amounts of data while also providing users with searchable access to any data points.
 3. This way, the data can be rolled up, sliced and diced as needed to handle the most/widest variety of questions that are relevant to a user's state of interest.
 4. IT developers with a working knowledge of OLAP cubes can create management packs to define their own extensible and customizable OLAP cubes that are built on the data warehouse infrastructure.
 5. These cubes are stored in SQL Services Analysis Services [SSAS].

q2) Explain about MOLAP?

- Ans 1.
1. Classif form of OLAP is known as MDLAP and often called as OLAP.
 2. Simple databases structure such as time, period, product, location, etc. are used.
 3. Functioning of each and every dimensions or data structure is defined by one or more hierarchies.

q3) Explain about ROLAP?

- Ans 1.
1. Functioning of ROLAP occurs simultaneously with

relational databases.

2. Data and tables are stored as relational tables.
3. To hold new information or data new tables are created.
4. Functioning of ROLAP depends upon specialized schema design.

Q4. What is hybrid Explain about ROLAP Hybrid?

- Ans 1. When a database developer uses Hybrid OLAP it divides the data between relational and specialized storage.
2. In some particular modifications a HOLAP database may store huge amounts of data in its relational tables.
 3. Specialized data storage is used to store data which is less detailed and more aggregate.

Q5. Explain difference between MOLAP and ROLAP?

ROLAP

MOLAP

- | | |
|--|---|
| 1. Stands for relational online analytical processing. | Stands for multidimensional online analytical processing. |
| 2. Used for large data volumes. | Used for limited data volumes. |
| 3. Access is slow. | Access is fast. |
| 4. Data stored in relational tables. | Data stored in multidimensional cubes. |
| 5. Data is fetched from data warehouse. | Data is fetched from MDDBS database. |
| 6. In ROLAP, complicated SQL queries are used. | Here sparse matrix is used. |
| 7. Static multidimensional view of data is created. | Dynamic multidimensional view of data is created. |

NAME:- AFNAN ATTAR PRN: F19112003 CLASS:- BE COMP II
 SUBJECT:- BI ASSIGNMENT NO.: 4

Q1) What is pivot table?

- Ans 1. A pivot table is a tool that allows us to summarize and analyze data in spreadsheet or database.
2. You can use pivot tables to group data, calculate, average or totals, and create charts and graphs.
 3. For typical data entry and storage, data usually appears in flat types, meaning that they consists of only columns and rows.
 4. Excel pivot tables include the features to directly query an online analytical processing [OLAP] server for retrieving data instead of getting data from an excel spreadsheet.

Q2) What are some ways to rearrange data within a pivot table?

Ans There are few ways to rearrange data within a pivot table including:-

- 1) Change the order of fields that we are using.
- 2) Change the way data is being summarized.
- 3) Change the way the data is being filtered.

Q3) What is a page field in context with pivot tables?

Ans 1. A page field is a field in pivot table that you can use to filter the data in the table.

2. For example, if you have a pivot table that contains data on different countries, you could use a page field to filter the table so the only data on one particular country is shown.

Q4) What are the advantages of using a pivot chart over a regular chart?

- Ans 1. Pivot charts offer a number of advantages over regular charts, chief among them being that they are far more flexible and customizable.
2. With a regular chart, you are limited to the data that is already present in the chart.
 3. With a pivot chart we can easily change the data that is being used to generate the chart, without having to recreate the entire chart from scratch.
 4. This makes pivot charts more versatile and useful for data analytics.

Q5) How is a pivot table different from a summary table?

- Ans 1. A pivot table is a type of summary table that allows you to recognize and summarize data in a flexible way.
2. With a pivot table, we can choose which columns and rows to include in your summary, and you can also choose how to summarize data.
 3. For example, you could use a pivot table to sum data by month, or you could use a pivot table to calculate the average price of a product by category.
 4. A summary table, on the other hand, is a static table that shows a summary of data from a larger data set.
 5. Summary tables are typically used to show things like totals and averages, and they cannot do customization the way pivot tables can.

NAME:-AFNAN ATTAR CLASS:- BE COMP II PRN:- F19II2003
 SUBJECT:- BI ASSIGNMENT No. :- 05

Q1) What do you mean by clustering?

- Ans 1. In machine learning, we often group examples as a first step to understand and a subject in machine learning system.
2. Grouping unlabelled examples is called clustering.
 3. As the examples are unlabeled, clustering relies on unsupervised machine learning.
 4. If the examples are labelled, then clustering becomes classification.
 5. Before group similar examples we need to first find similar examples.
 6. One can measure similarity between examples by combining the example's feature data into a metric, called a similarity measure.
 7. When each example is defined by one or two features, it's easy to measure similarity.
 8. As the number of features increases, creating a similarity measures becomes more complex.
 9. Clustering is used in:-
 - Market segmentation.
 - Social Network analysis.
 - Search result grouping.
 - Medical imaging.
 - Image segmentation..
 - Anomaly detection.

Q2) What is a hierarchical clustering algorithm?

- Ans 1. A hierarchical clustering method works via grouping data into a tree of clusters.
2. Hierarchical clustering begins by treating every data point as a separate cluster.
 3. Then, it repeatedly executes the steps:-

- i) Identify the 2 clusters which can be closest together.
- ii) Merge the 2 maximum comparable clusters. We need to continue these steps until all clusters are merged together.
4. In hierarchical clustering, the aim is to produce a hierarchical series of nested clusters.
5. A diagram called dendrogram graphically represents this hierarchy and is an inverted tree that describes the order in which factors are merged or clusters are broken up.
6. Hierarchical clustering is a method of cluster analysis in data mining that creates a hierarchical representations of the clusters in a dataset.

Q3) Explain the differences between classification and clustering?

Ans Classification

Clustering

- | | |
|---|--|
| 1. Approach to classify input instances on the basis of related class labels. | It is used to set the instances on their resemblance without class labels. |
| 2. It is a type of supervised learning method. | It is a kind of unsupervised learning method. |
| 3. It prefers a training dataset. | It does not prefer a training dataset. |
| 4. It is more complex to make compared to clustering. | It is less complex. |
| 5. We utilise labels for training data. | Here; labels are not preferred for training data. |
| 6. Example:- logistic regression. | Example:- K-Means. |

Q4) What is classification?

- Ans 1. Classification is the problem of identifying which of a set of categories an observation belongs to.
2. Examples are assigned to spam or non-spam if one is classifying emails.
 3. Often, the individual observations are analyzed into a set of quantifiable properties, known variously as explanatory variables or features.
 4. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier.
 5. Classification is often done with logistic regression or similar procedure, the properties of observations are termed explanatory variables and the categories to be predicted are known as outcomes, which are considered to be possible values of dependent variable.
 6. Classification has many applications, in some of these it is employed as a data mining procedure, while in others more detailed statistical modelling is undertaken:-
- Biological classification.
 - Document classification.
 - Biometric.
 - Credit scoring.
 - Computer Vision.
 - Geostatistics.

Q5 How to evaluate a classification model?

Ans Mostly a confusion matrix is used to evaluate a classification model.

		Predicted Value	
		Positive	Negative
Real Value	Positive	TP	FP
	Negative	FN	TN

Page No.	
Date	

- True Positive (TP) :- Predicted positive, Real positive.
- True Negative (TN) :- Predicted negative, real negative..
- False Positive (FP) :- Predicted positive, real negative.
- False Negative (FN) :- Predicted negative, real positive.

Evaluation metrics are also given as :-

- Precision = $\frac{TP}{TP + FP}$
- Recall = $\frac{TP}{TP + FN}$
- F1 Score = $\frac{2(Recall \times Precision)}{Recall + Precision}$
- Specificity = $\frac{TN}{TN + FP}$
- Fall-out = $\frac{FP}{TN + FP}$
- Miss Rate = $\frac{FN}{TP + FN}$

These metrics help us evaluate a classification model.