Assignment no 9 – Mayur Borse

1. **#Data Lake : -**

   - A centralized repository that allows storing raw and unstructured data at scale. It provides a flexible and cost-effective solution for big data storage and analytics.

2. **#Data Warehouse : -**

   - A large, centralized repository that aggregates and stores structured data from different sources. It is designed for efficient querying and analysis to support business intelligence.

3. **#Data Wrangling : -**

   - The process of cleaning, structuring, and enriching raw data into a usable format for better decision-making. It involves tasks like cleaning, transforming, and combining data from different sources.

4. **#Data Governance : -**

   - The framework of policies, processes, and standards to ensure high data quality, integrity, and security across an organization.

5. **#ETL (Extract, Transform, Load) : -**

   - A data integration process that involves extracting data from source systems, transforming it into a desired format, and loading it into a target data warehouse or database.

6. #Data Integration : -

- The process of combining data from different sources to provide a unified view. It ensures that data from diverse systems can work together seamlessly.

7. #Data Quality : -

- The measure of accuracy, completeness, and reliability of data. Data quality management involves strategies and processes to maintain high-quality data.

8. #Data Mining : -

- The process of discovering patterns, trends, and insights from large datasets using statistical and machine learning techniques.

9. #Data Mart : -

- A subset of a data warehouse that is focused on a specific business function or department. It contains a smaller, more specialized set of data for targeted analysis.

10. #Master Data Management (MDM) : -

- The process of managing and organizing key business data, such as customer information or product data, to ensure consistency and accuracy across an organization.

11. #Big Data : -

- Refers to large and complex datasets that traditional data processing applications may struggle to handle. Big data technologies enable the storage, processing, and analysis of such datasets.

12. #Data Mining : -

- The process of discovering patterns, trends, and insights from large datasets using statistical and machine learning techniques.

13. #Data Mart : -

- A subset of a data warehouse that is focused on a specific business function or department. It contains a smaller, more specialized set of data for targeted analysis.

14. #Master Data Management (MDM) : -

- The process of managing and organizing key business data, such as customer information or product data, to ensure consistency and accuracy across an organization.

15. #Big Data : -

- Refers to large and complex datasets that traditional data processing applications may struggle to handle. Big data technologies enable the storage, processing, and analysis of such datasets.

16. #Data Mining : -

- The process of discovering patterns, trends, and insights from large datasets using statistical and machine learning techniques.

17. #Predictive Analytics : -

- The use of statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data.

18. #NoSQL Database : -

- A type of database that provides a flexible and scalable approach to storing and managing unstructured or semi-structured data. It is suitable for big data and real-time applications.

19. #Data Masking : -

- The technique of protecting sensitive information by replacing, encrypting, or scrambling original data while maintaining its usability for testing or analysis.

20. #Data Scientist : -

- A professional who analyzes and interprets complex data sets to inform business decision-making. Data scientists use a combination of statistical analysis, machine learning, and domain knowledge.