Name: AKULA SHARATHCHANDRA

Batch: Data Engineering

Date:17/01/2024

QUESTION: Explain data Engineering and data warehousing concept?

SOLUTION:

DATA ENGINEERING :They build and scale the platforms that enable data collection, processing and storage for data science/business analytics use.

ETL : Extract transform load Architecture of ETL

DATA CLASSIFICATION:

1.Raw Data: Unprocessed data in format used on source e.g JSON

2.Processed data Raw data with schema applied

3.Cooked data Processed data that has been summarized.

- **BIG DATA PROPERTIES(4 V's):**

Volume: How much data you have.

Velocity: How fast data is getting to you

Variety: How different your data is

Veracity: How reliable your data is

Data Processing Methods:

1. Stream Processing

(Data->Storage->Analytics->Insight) 2.Batch

processing (Live data Processing)

DataBase Storage:

1. SQL

2. No sql

DATA WAREHOUSE: A Data Warehouse is a comprehensive, subject-oriented, and integrated repository of data that, being time-variant and non-volatile, provides a consistent representation of business conditions for effective management decision-making at a specific moment.

1) **Features of Data Warehouse:-**

i) Subject-oriented: It mainly focuses on modeling and analysis of data for decision makers, not on daily operations or transaction processing.

ii) Integrated: Constructed by integrating multiple, heterogeneous data sources like relational databases, flat files, on-line transaction records.

iii) Time-variant: Stores historical data, allowing users to analyze changes and trends over time for better decision-making.

iv) Nonvolatile: No updates are allowed. Once the data enters into the data warehouse, they are never removed.

2) **Need for Decision Support Systems (DSS) in Business**:Quick Decision Making: Essential in competitive markets.
Information Systems: Necessary for diverse information and decision-making needs.

3) Structured and Unstructured Components of DSS:

Structured: Components directly aiding decisions.
Unstructured: Components requiring human interaction, like intuition.

4) **DSS Architectural Styles**:

OLTP (Online Transaction Processing): Traditional operational systems.OLTP databases access the data in the form of operations like -inserting,updating,deleting etc as rdbms provides/access bulk amount of data at a time.

OLAP (Online Analytical Processing): Used by Data Warehouses for analytical queries.

5) **Operational Database (OLTP):**Optimized for fast transaction processing.Involves operations like inserting, deleting, and updating data.
ATM is an example of OLTP.

6) **OLAP**:Methodology for providing users access to large data for analytical queries.

Cube Structure: Allows multi-dimensional analysis for insights.

7) **OLAP Architecture**:ROLAP (Relational OLAP): Analyzes data stored in relational databases.MOLAP (Multi-dimensional OLAP): Utilizes cube structures for analysis.

8) **Data Warehouse Components**:Data Store, Extraction, Filtering, Query Tool, Presentation Tool.

9) **Data Mart**:Subsets of data warehouse, often oriented to specific business lines or teams.

10) Data Warehouse Applications:

Information Processing: It allows for data to be processed through querying,basic statistical analysis, reporting, and visualization.
Analytical Processing: supports advanced analysis through basic OLAP operations such as slice-and-dice, drill down, drill up, and pivoting.
Data Mining: Focuses on knowledge discovery from hidden patterns.