Collins Iregi

CIT-223-041/2020

Question 1

a,

1. Primary function: TPS process high volume , routine transactions while DSS analyze large datasets to support decision making
2. data focus: TPS uses current operational data(account balances, transactions) while DSS use historical and aggregate data(trends and forecasts)
3. user focus: tps is geared towards tellers, customer service representatives and back-office staff while dss is used by managers, analysts, loan officers and executives
4. Response times: TPS is real-time or near real-time while DSS is flexible with response times varying based on complexity
5. User interface: TPS has simple user friendly interfaces for quick data entry and retrieval while DSS offer more complex interfaces with data analysis and visualization tools

b,

1. Librarians   
   library managers  
   students and readers
2. **Librarians** require data on patron reader demographics, borrowing history and feedback as well as inventory data for physical and digital resources, their usage statistics and circulation data. They use Library management systems to manage resource cataloging, resource circulation and patron accounts   
   **Library managers** need access to budget data(expenses and revenue), staff performance data and facility data(maintenance records, energy consumption). They use financial management software to track budgets expenses and revenue and human resource management systems to manage employee data, payroll and benefits

**Students and readers** need bibliographical data to access library catalogs, databases and materials. This is managed by library catalogs and search interfaces

c,

A data warehouse is a large, centralized repository of data that stores integrated information from throughout an organization for reporting and data analysis, whereas a data mart is a subset of a data warehouse that is tailored to meet the needs of a specific business unit or team.

Question 2

a,

**Early Era (1960s-1980s):** Hierarchical and Network Databases were used, but they had limited BI capabilities due to their complex structures.

**Relational Databases (1980s-2000s):** The emergence of SQL and relational databases like Oracle, DB2, and MySQL revolutionized data access and manipulation. Data warehouses also emerged for BI analysis.

**OLAP and OLTP (1990s-2000s):** OLAP databases optimized for data analysis and OLTP databases for high-volume transactions were introduced. Data mining techniques also emerged for extracting insights from large datasets.

**Big Data Era (2000s-Present):** NoSQL databases like MongoDB and Cassandra, cloud databases, and in-memory computing emerged to handle big data, enabling real-time analytics and complex data exploration.

**Current Technologies and Trends:** Machine Learning and AI are integrated into databases for automated data analysis. Data lakes store all types of data. There’s increased focus on data governance and security. Self-service BI tools are now available for business users.

b,

1. Confidence Factor: In data mining, confidence is a measure of the reliability of an association rule. It helps understand the strength of the association between items.
2. Support Count: Support is the relative frequency of an item set in a dataset. It helps identify common item sets.

c,

1. Data Cleaning and Preprocessing: This involves identifying noisy data points, correcting them if possible, or removing them if necessary.
2. Data Transformation: Techniques like normalization, standardization, binning, and smoothing are used to reduce the influence of outliers and make data easier to analyze.
3. Robust Data Mining Algorithms: Choosing the right algorithm and using ensemble methods can help mitigate the impact of noise and improve overall accuracy in data analysis.