

UNIT-5

Applications of Data Warehousing and Data Mining in Government:

⊗ Introduction: Data warehousing is a collection of tools and techniques using which more knowledge can be driven out from a large amount of data. This helps with the decision-making process and improving information resources. Data warehouse is basically a database of unique data structures that allows relatively quick and easy performance of complex queries over a large amount of data.

Features / Characteristics of Data Warehouse:

- i) Subject Oriented: A data warehouse is subject-oriented. It provides useful data about a subject instead of the company's ongoing operations, and these subjects can be customers, suppliers, marketing etc.
- ii) Time-Variant: The different data present in the data warehouse provides information for a specific period. Historical data is kept in a data warehouse. For example, one can retrieve data from 3 months, 6 months, 12 months, or even older from a data warehouse.
- iii) Integrated: A data warehouse is built by joining data from heterogeneous sources, such as social databases, excel documents etc. It is constructed by integrating data from heterogeneous sources such as relational databases, flat files etc. This integration enhances the effective analysis of data.
- iv) Non-Volatile: It means, once data entered into the warehouse cannot be changed. The data resided in data warehouse is permanent.

Advantages of Data Warehouse:

- i) Delivers enhanced business intelligence: By having access to information from various sources from a single platform, decision makers will no longer need to rely on limited data, and can be applied to a business processes.
- ii) Saves time: Executives can query the database themselves with little to no IT support, saving more time and money.

iii) Enhances data quality and consistency: A data warehouse converts data from multiple sources into a consistent format. This will lead to more accurate data, which will become the basis for solid decisions.

iv) Generates a high Return on Investment (ROI): Companies experience higher revenues and cost savings than those that haven't invested in a data warehouse.

v) Streamlines the flow of information: Data warehousing facilitates the flow of information through a network connecting all related or non-related parties.

Applications of Data Warehousing:

- | | |
|--------------------------|-------------------------|
| → Financial services | → Analytical Processing |
| → Banking services | → Data Mining |
| → Consumer goods | → Real Life |
| → Retail sectors | → Various Industries |
| → Information Processing | → Decision making |

Database vs. Data Warehouse:

Data Warehouse (OLAP)	Operational Database (OLTP)
i) Online Analytical Processing.	i) Online Transactional Processing.
ii) The number of users is in hundreds.	ii) The number of users is in thousand.
iii) It provides summarized and multidimensional view of data.	iii) It provides detailed and flat relational view of data.
iv) The database size is from 100 GB to 100TB.	iv) The database size is from 100MB to 100GB.
v) It contains historical data.	v) It contains current data.

⊗. Data Mining:

Data mining refers to extracting knowledge from large amounts of data. The data sources can include databases, data warehouse, web etc. Data mining refers to the analysis of data. It is the computer-supported process of analyzing huge sets of data that have either been compiled by computer systems or have been downloaded into the computer. In data mining process, the computer analyzes data and extract useful information from it. Data mining aims to enable business organizations to view business behaviours, trends, relationships that allow the business to make data-driven decisions.

Applications of Data Mining:

- 1) Data Mining in Healthcare: Data mining in healthcare has excellent potential to improve the health system. It uses data and analytics for better insights and to identify best practices that will enhance health care services and reduce costs. Data Mining can be used to forecast patients in each category.
- 2) Data Mining in Education: Education data mining is a newly emerging field, concerned with developing techniques that explore knowledge from the data generated from educational environments. An organization can use data mining to make precise decisions and also to predict the results of the student.
- 3) Data Mining in Manufacturing Engineering: Knowledge is the best asset possessed by a manufacturing company. Data mining tools can be beneficial to find patterns in a complex manufacturing process. Data Mining can be used in system-level designing to obtain the relationships between product architecture, product portfolio, and data needs of customers.

4) Data Mining in CRM (Customer Relationship Management):

It is all about obtaining and holding customers, also enhancing customer loyalty and implementing customer-oriented strategies. To get a descent relationship with the customer, a business organization needs to collect data and analyze the data.

5) Data Mining in Fraud Detection:

Billions of dollars are lost to the action of frauds. An ideal fraud detection system should protect data of all the users. A model is constructed using data, and the technique is made to identify whether the document is frauded or not.

⑥. National Data Warehouses:

The National data warehouse allows researchers and policy makers to view historical data sets and extract data across different groups. The large number of data warehouses can be identified from the existing data resources within the center government ministries. In Nepal, big corporate organizations like Data Center of Singh durbar, Agriculture Bank Development, Nepal Telecom, Nepal Stock Exchange etc. have been seen to make extensive use of data warehouse.

Census Data: Census data is an official count or survey, especially of a population. A census is the procedure of systematically acquiring and recording information about the members of given population. The term is used mostly in connection with national population and housing censuses; other common censuses include agriculture, business, and traffic censuses. The Census is also an important economic tool. At national level census information is used to plan the provision of health care, education, employment, transport etc. It is used to help determine where to build new schools, roads, health care facilities, child-care etc.

Prices of Essential Commodities: Commodities are extremely important as they are essential factors in the production of other goods. A commodities future price is determined primarily by the supply and demand for the commodity in the market. Since commodities are traded on exchanges, their prices are not set by a single individual or entity. Data warehouse help and analysts to find out the lacking problem of commodities, destruction, storing and stocking process.

⊗. Other areas for data warehouse and data mining:

Agriculture: The Agricultural Census performed by the Ministry of Agriculture, Government of Nepal, compiles a large number of agricultural parameters at the national level. District-wise agricultural production area and yield of crops is compiled; this can be built into a data warehouse for analysis, mining and forecasting. Data on agricultural inputs such as seeds and fertilizers can also be effectively analyzed in a data warehouse. Land-use pattern statistics can also be analyzed in a warehousing environment. Other data such as watershed details and also agricultural credit data can be effectively used for analysis. Thus there is substantial scope for application of data warehousing and data mining techniques in Agricultural sector.

Rural Development: Data on individuals below poverty line (BPL survey) can be built into a data warehouse. Drinking water census data (from Drinking Water Mission) can be effectively utilized by OLAP and data mining technologies. Monitoring and analysis of progress made on implementation of rural development program, can also be made using OLAP and data mining techniques.

Health: Community needs assessment data, immunization data, data from national programs on controlling blindness, etc. can all be used for data warehousing implementation, OLAP and data mining applications.

Planning: At the planning commission, data warehouses can be built for state plan data on all sectors: labor, energy, education, trade and industry, five year plan etc.

Education: The Sixth All India Educational Survey data has been converted into a data warehouse (with about 36GB of data). Various types of analytical queries and reports can be answered.

⊗ Data warehousing vs. Data Mining:

Data Warehousing	Data Mining
i) Data warehousing is the process of compiling and organizing data into one common database.	i) Data Mining is the process of extracting meaningful data from the database.
ii) A data warehouse is a database used to store data.	ii) Data Mining is actually the analysis of data.
iii) It is a process of transforming data into information and making it available to users for analysis.	iii) Data mining is a logical process that is used to search through large amount of data in order to find useful data.
iv) Data warehousing is a process which needs to occur before any data mining can take place.	iv) Data mining can only be done once data warehousing is complete.



**If my notes really helped
you, then you can support
me on esewa for my
hardwork.**

Esewa ID: 9806470952