gradeup

Data Warehousing and Data Mining

Content:-

Data Warehouse: Data warehouse is the collection of information as well as support system . However, a clear distinction exists. Traditional database are transactional (relational, object-oriented, network, or hierarchical). Data warehouse have the distinguishing characteristic that they are mainly intended for decision-support applications.

Data warehouse as a subject-oriented, non- volatile,integrated , time-variant collection of data in support of management's decision. Data warehouses provide access to data for knowledge discovery,complex analysisand decision making. They support high performance demands on a organization's data and information. Several type of applications - OLPA ,DDS and data mining applications - are supports.

OLAP:- OLAP stands for Online Analytical Processing. This term is used to describe the analysis of complex data from the data warehouse. In the hands of skill knowledge workers, OLAP tools use distributed computing capabilities for analysis that require more storage and processing power than can be economically and efficiently located on an individual desktop.

gradeup

Gradeup UGC NET Super Subscription
Access to all Structured Courses & Test Series

DSS:- DSS stands for decision-support systems. It is also known as ESI(executive information systems)(not to be confused with enterprise integration systems) support an organization’s leading decision makers with higher level data for complex and important decisions. Data mining knowledge is used for knowledge discovery , the process of searching data for unanticipated new knowledge.

Traditional databases support **Online transaction processing(OLTP)**, which includes insertions, deletions and updates , while also supporting information query requirements. Traditional relational databases are optimized to process queries that may touch a small part of database and transactions that deal with insertions or updates of a few tuples per relation process. Thus they cannot be optimized for DSS ,OLAP or data mining. By contrast, data warehouses are designed precisely to support efficient extraction, processing and presentation for analytic and decision-making purposes. In comparison to traditional database, data warehouses generally contain very large amount of data from multiple sources that may include database from different data models and sometimes files acquired from independent systems and platforms.

gradeup

Characteristics of Data Warehouse:

The multidimensional data model is good fit for OLAP and decision-support technologies. In contrast to multi databases, which provide access to disjoint and usually heterogeneous databases, a data warehouse is frequently a store of integrated data from multiple sources, processed for storage in a multidimensional model.

Gradeup UGC NET Super Subscription
Access to all Structured Courses & Test Series

PHOTO - 1

Data warehouse more generally as a collection of decision support technologies, aimed at enabling the knowledge worker (executive , manager ,analyst) to make better and faster decisions. Diagram gives an overall view of the conceptual structure of a data warehouse. It shows the entire data warehousing process , which includes possible cleaning and reformatting of data before loading it into the warehouse. This process is presently handled by tools known in industry as ETL(extraction, transformation and loading) tools . At the back end process ,OLAP, data mining and DSS may generate new relevant information such as rules;

Data Modelling for Data warehouse :- Multidimensional models take advantages of inherit relationships in data to populate data in multidimensional matrices called data cubes. (these may be called hyper cubes if they have more than three dimensions) . For data that lends itself to dimensional formatting , query performance in multi dimensional matrices can be much better than in the relational data model. Three dimensional example in corporate data warehouse are the corporation’s fiscal periods, products and regions .

Photo -2

A standard spreadsheet is a two-dimensional matrix. One example would be a spread sheet of regional sales by product for a particular time period. Products could be shown as rows , with sales revenues for each region comprising the column. (diagram 2 shows two dimensional organization) . Adding the time dimension , such as an organization’s fiscal quarters , would produce a three dimensional matrix, which could be represented using a data cube .

gradeup

Gradeup UGC NET Super Subscription
Access to all Structured Courses & Test Series

Photo–3

Diagram 3 shows a three-dimensional data cube that organizes product sales data by fiscal quarters and a sale regions . Each cell could contain data for specific product , specific fiscal quarter and specific region . By including additional dimensions , a data hyper cube could be produced , although more than three dimensions cannot easily visualized for graphically presented. The data can be queried directly in any combination of dimensions, bypassing complex data queries. Tools exits for viewing data according to the user dimensions.

Changing from one dimension hierarchy(orientation) to another is easily accomplished in data cube with a technique called pivoting (also called rotation) .

Multi dimensional models lend themselves readily to hierarchical views in what is known as roll-up display and drill down display . **Roll-up display** moves up the hierarchy , grouping into larger units along a dimensions. A **drill down** display provides the opposite capability, furnishing a finergrained view , perhaps disaggregating country sales by region and then regional sales by sub region and also breaking up products by style .

The multidimensional storage model involves two types of tables : dimensional tables and facts tables. A dimension table consists of tuples of attribute of the dimension. A fact table can be thought of as having tuples , one per a recorded fact. This fact contains some measured or observed variable(s) and identify it(them) with pointer dimension tables. The fact table contains the data, and the dimensions identify each tuple in that data. Diagram 4 contains an example of fact table that can be viewed from the prespective of multi dimension table.

Gradeup UGC NET Super Subscription
Access to all Structured Courses & Test Series

Photo -4

Two common multidimensional schemas are the star schema and the snowflake schema . The **star schema**consists of a fact table with a single table for each dimension (diagram 4). The **Snowflake** schema is variation on the star schema in which the dimensional table from a star schema are organized into a hierarchy by normalization them (diagram 5).

Photo -5

Building a Datawarehouse:-

Following steps are involved for the acquisition of data for the warehouse:-

The data must be extracted from multiple, heterogeneous sources, for example , databases or other data feeds such as those containing financial market data or environment data.

- Data must be formatted for consistency within the warehouse. Names, meanings and domains of data from unrelated sources must be reconciled.
- The data must be cleaned to ensure the validity. Data cleaning is an involved and complex process that has been identified as a largest labor -demanding component of data warehouse construction. For input data, cleaning must occur before the data is loaded into the warehouse. There is nothing about cleaning data that is specific to data warehousing and that could not be applied to a host database.
- Data may have to be converted from relational, object-oriented or legacy data-bases(network/or hierarchical) to a multidimensional model.

Gradeup UGC NET Super Subscription
Access to all Structured Courses & Test Series

Data Mining:-

Process of exacting the useful data from the large set of raw data is known as **Data Mining.** Analysing different data patterns in large batches of data using one or more software. It has applications in multiple fields like research and science. With the help of data mining the business can learn more about their customers and develop more effective strategies related to various business function. Data mining involves effective warehousing and data collection as well as computer processing. Another name of Data mining is knowledge Discovery in Data (KDD).

Data Mining Key features are:-

1. Likely Outcomes based on prediction
2. Decision oriented information is created.
3. Based on the trend and behaviour automatic pattern prediction is done.
4. For analysis of data it focuses on large data set and databases.

Data Sized:- More data is required to be processed and maintained for creation of power systems .

Query Complexity:- More power system is required for querying or processing more complex queries and large number of query.

Uses:-

1. Mining techniques are useful in many research projects including cybernetics, mathematics, marketing and genetics.
2. With the help of data mining, retailer could manage and use point-of-sale records of customer purchases to send the targeted promotions based on an individual purchase history.

Gradeup UGC NET Super Subscription
Access to all Structured Courses & Test Series

gradeup
SUPER

Gradeup UGC NET
Super Superscription

Features:

1. 7+ Structured Courses for UGC NET Exam
2. 200+ Mock Tests for UGC NET & MHSET Exams
3. Separate Batches in Hindi & English
4. Mock Tests are available in Hindi & English
5. Available on Mobile & Desktop

Gradeup Super Subscription, Enroll Now

gradeup.co