analytical reporting, structured and/or ad hoc queries, and decision making.

Data warehousing involves data cleaning, data integration, and data consolidations. To integrate heterogeneous databases, we have the following two approaches −

- Query Driven Approach
- Update Driven Approach

Query-Driven Approach

This is the traditional approach to integrate heterogeneous databases. This approach is used to build wrappers and integrators on top of multiple heterogeneous databases. These integrators are also known as mediators.

Process of Query Driven Approach

- When a query is issued to a client side, a metadata dictionary translates the query into the queries, appropriate for the individual heterogeneous site involved.
- Now these queries are mapped and sent to the local query processor.
- The results from heterogeneous sites are integrated into a global answer set.

Disadvantages

This approach has the following disadvantages −

- The Query Driven Approach needs complex integration and filtering processes.
- It is very inefficient and very expensive for frequent queries.

Preview from Notesale.co.uk
Page 14 of 20

warehouses constructed by such preprocessing are valuable sources of high quality data for OLAP and data mining as well.

- Available information processing infrastructure surrounding data warehouses − Information processing infrastructure refers to accessing, integration, consolidation, and transformation of multiple heterogeneous databases, web-accessing and service facilities, reporting and OLAP analysis tools.
- OLAP−based exploratory data analysis − Exploratory data analysis is required for effective data mining. OLAM provides facility for data mining on various subset of data and at different levels of abstraction.
- Online selection of data mining functions − Integrating OLAP with multiple data mining functions and online analytical mining provide users with the flexibility to select desired data mining functions and swap data mining tasks dynamically.

Preview from Notesale.co.uk
Page 17 of 20

Data Mining - Terminologies

Data Mining

Data mining is defined as extracting the information from a huge set of data. In other words we can say that data mining is mining

involves transformations to correct the wrong data. Data cleaning is performed as a data preprocessing step while preparing the data for a data warehouse.

Data Selection

Data Selection is the process where data relevant to the analysis task are retrieved from the database. Sometimes data transformation and consolidation are performed before the data selection process.

Clusters

Cluster refers to a group of similar kind of objects. Cluster analysis refers to forming group of objects that are very similar to each other but are highly different from the objects in other clusters.

Data Transformation

In this step, data is transformed or consolidated into forms appropriate for mining, by performing summary or aggregation operations.

Preview from Notesale.co.uk
Page 20 of 20