The new data warehousing technology

Science, Computer Science

Data warehousing is a new technology that provides the users with the tools to store the summarized information from multiple, heterogeneous databases in a single repository. The data warehouse is also seen as a data arrangement and analysis technology, adopting an “ update” approach. A well-defined and properly implemented data warehouse can be an invaluable competitive tool. It provides tools to satisfy the information needs of the employees at all organizational levels — not just for complex data queries, but as a general facility for getting quick, accurate and often insightful information. A data warehouse is designed so that its users can recognize the information they want and access that information they want and access that information using simple tools. Data warehouse users make requests and are delivered information “ products” that are created from the components and parts stored in the warehouse.

A data warehouse is a typical blending of technologies, including relational and multidimensional databases, client/server architecture, extraction/ transformation programs, graphical user interfaces and more. A few definitions of data warehouse are as follows: According to Vidette Poe, “ Data warehouse is a read-only analytical database that is used as the foundation of a decision support system”. Amy Helen Johnson defines data warehouse as “ a database that collects business information from many sources in the enterprise covering all aspects of the company’s processes, products, and customers”. W. H. Inmon, who is considered to be the Father of Data Warehousing, has given the following definition: “ A Data Warehouse is a subject-oriented, integrated, time-varying, non-volatile collection of data in support of the management’s decision-making process”. Subject-Driven A

data warehouse is organized around major subjects and contains only the information necessary for the decision support processing. It is not organized according to the application (e. g., a data warehousing for the bank would be organized by the customer, deposit/advances, interest rate and not by different products).

A data warehouse is always a physically separate data store. The relative data is transformed from the application data. As such, data warehousing does not require processing of transactions, recovery, etc. The data is not updated or changed either after the data enters data warehouse. Data is only loaded, refreshed and accessed for queries.

Time-varying data in the data warehouse is collected from the corporate data archives and could be 3 to 10 years old or even older. The data provide historical perspective and is used for comparisons, trends, and forecasting. Integrated While constructing the data warehouse, multiple, heterogeneous sources such as relational databases, flat files, and OLTP files are utilized and data collected from them is integrated. Data cleaning and data integration techniques are applied to maintain consistency in naming conventions, measures of variables, encoding structure and physical attributes.