Clyde Joshua Delgado BS in Computer Science III

**A Data Warehouse**

A data warehouse is also a database but is different from the organization's operational database. This data warehouse works independently from organization’s operational database. Unlike in the operational database, updating is infrequent in a data warehouse. Historical or past data is stored in these databases which is needed for the organization to analyze its business through data collected from the past. A data warehouse contributes to organizing, understanding, and using of their data to make strategic decisions for businesses involved by analysis of consolidated historical data.

A data warehouse is separated from the operational database for several reasons. First, an operational database is made for tasks and workloads such as searching records, indexing, etc. On the other hand, data warehouse queries are often complex and they present a general form of data. Secondly, Operational databases support concurrent processing of several operations. Concurrency control and recovery are needed in operational databases for ensured robustness and consistency. Data warehouses are not design for this since they only hold past data. Lastly, an operational database query allows to modify data on databases. Queries on data warehouses should only allow reading of data.

 We can say that operational databases holds current data while a data warehouse keeps historical data for future analysis.

There are key features for a data warehouse. It is subject oriented since it gives information about a subject instead of the organization's ongoing operations. Subjects can be defined as their products, customers, suppliers, sales, revenue, etc. It focuses more on showing and analysis of data for decision making.

A data warehouse is constructed by integrated data from heterogeneous sources such as relational databases, flat files, etc. Integrated data helps in its analysis. Data warehouses are also time variant. Collected data in the warehouse is identified with certain time periods which provides information from a historical point of view. Lastly, it is non-volatile where the previous data is not deleted if fresh data is appended. Since it is separate from the operational database, frequent changes is not reflected in the data warehouse.

Some examples of businesses and services that use data warehouses are Financial services, Banks, Consumer goods, Retail sectors and Controlled manufacturing.

**Data Warehousing**

Data warehousing is the construction of thee data warehouse and at the same time using it. Like what was said before, A data warehouse is integrated specifically from sources that support analytic reporting, structured ad hoc queries and decision making. The use of the data warehouse involves data cleaning, data integration, and data consolidations.

Decision support technologies use the data available in a data warehouse. These technologies help executives to use the warehouse optimally and effectively. The information gathered and analyzed in a warehouse can be used in tuning production strategiesby arranging and managing the product portfolios by comparison of quarterly or yearly sales. Another domain is Customer Analysis. It is the analysis of the customer's buying preferences, buying time, budget cycles, etc.

 Lastly, it helps in customer relationship management, and making environmental corrections which contributes to user experience. The information also allows us to analyze business operations.

To integrate heterogeneous databases, there are two approaches. First, the query-driven approach which is the traditional approach when integrating heterogeneous databases. This approach was used to build wrappers and integrators on top of multiple heterogeneous databases. These integrators are also known as mediators. Secondly, the update-driven approach which is followed on today's data warehouse systems rather than the earlier approach. Here the information from multiple heterogeneous sources are integrated in advance and are stored in a warehouse. This information is available for direct querying and analysis.

The functions of data warehouse tools and utilities are Data Extraction which is the gathering data from multiple heterogeneous sources. Data Cleaning is the finding and correcting the errors in data. Data Transformation is the conversion of data from legacy format to warehouse format. Data Loading is Sorting, summarizing, consolidating, checking integrity, and building indices and partitions. Lastly, Refreshing is the updating from data sources to the warehouse.

**Data Mining**

Data mining the retrieving of hidden predictive information from large data warehouses. Data mining is new technology which is a powerful tool in these data warehouses and has great potential in helping companies focus in their data warehouse’s most important information. Future trends and behaviors are predicted by data mining tools which allows businesses to make tactical decisions. Data mining tools can answer business questions that were formerly too inconvenient to resolve. They look for hidden patterns in data warehouses, finding predictive information that lies outside their expectations. Data in warehouses are more solid than opinions in some cases.

Companies have already collected and refined tons of data. Data mining techniques can be implemented on existing software and hardware platforms. This is to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought online. If implemented on powerful client/server or parallel processing computers, data mining tools can analyze massive data warehouses giving answers to questions such as, "Which advertisement is more effective in advertising products through number of clicks and why?"

Data mining techniques are results of a long process of research and product development. This evolution started when business data was first stored on computers. It continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time through online websites and more. Evolutionary process is taken by Data mining beyond retrospective data access and navigation to prospective and proactive information delivery.

Data mining is ready for application in the business community if it can collect massive data from customers, has powerful multiprocessor computers and effective data mining algorithms for their data warehouses.

Data mining can predict these certain events through modeling data. It is the act of building a model in a known situation and then applying trying to apply it to unknown future situations.

Sources:

<https://www.tutorialspoint.com/dwh/dwh_tutorial.pdf>

Pages 1-5

<http://www.thearling.com/text/dmwhite/dmwhite.htm>

An introduction to data mining