

The Need for Data Warehousing

In an interesting book, “Blink: The Power of Thinking Without Thinking”, Malcolm Gladwell talks about the theory of thin slices - how our brain, when overwhelmed with enormity or complexity of information to be analyzed for decision making, depends on thin slices of key information. Well, even the best of us with superior Ventromedial Prefrontal Cortex (the part of brain believed to play role in decision making), could use some help from technology to get those thin slices of key information right!

Fortunately, there are *Business Intelligence (BI)* technologies and applications for gathering, storing, analyzing, and providing access to data, and abstracting the key information to help make business decisions.

Data Warehousing is an important part and in most cases the foundation of BI architecture. I assume you have a good working definition of a data warehouse in your mind. Hence, I will not go into what a data warehouse is and is not. Let me highlight what you need a data warehouse for:

Data Integration

Even if you are a small Credit Union, I bet your enterprise data flows through and lives in a variety of in-house and external systems. You want to ask questions that represent those slices of key information (referred to as *Key Performance Indicators* or KPIs) such as - What is the member profitability or member value attrition? Oh, by the way, you want to be able to analyze it across all products by location, time and channel. You realize that all the required data is probably there but not *integrated* and organized in a way for you to get the answers easily.

Perhaps your IT staff has been providing the reports you need every time through a series of manual and automated steps of stripping or extracted the data from one source, sorting / merging with data from other sources, manually scrubbing and enriching the data and then running reports against it. You wonder there ought to be a better and reliable way of doing this.

Data Warehouse serves not only as a repository for historical data but also as an excellent data integration platform. The data in the data warehouse is integrated, subject oriented, time-variant and non-volatile to enable you to get a 360° view of your organization.

Advanced Reporting & Analysis

The data warehouse is designed specifically to support querying, reporting and analysis tasks. The data model is flattened (denormalized) and structured by subject areas to make it easier for users to get even complex summarized information with a relatively simple query and perform multi-dimensional analysis. This has two powerful benefits – multi-level trend analysis and end-user empowerment.

Multi-level trend analysis provides the ability to analyze key trends at every level across several different dimensions, e.g., Organization, Product, Location, Channel and Time, and hierarchies within them. Most reporting, data analysis, and visualization tools take advantage of the underlying data model to provide powerful capabilities such as drill-down, roll-up, drill-across and various ways of slicing and dicing data.

The flattened data model makes it much easier for users to understand the data and write queries rather than work with potentially several hundreds of tables and write long queries with complex table joins and clauses.

Knowledge Discovery and Decision Support

Knowledge discovery and data mining (KDD) is the automatic extraction of non-obvious hidden knowledge from large volumes of data. For example, *Classification* models could be used to classify members into low, medium and high lifetime value. Instead of coming up with a one-size-fits-all product, the membership can be divided into different clusters based on member profile using *Clustering* models, and products could be customized for each cluster. *Affinity groupings* could be used to identify better product bundling strategies.

These KDD applications use various statistical and data mining techniques and rely on subject oriented, summarized, cleansed and “de-noised” data which a well designed data warehouse can readily provide.

The data warehouse also enables an *Executive Information System (EIS)*. Executives typically could not be expected to sift through several different reports trying to get a holistic picture of the organization’s performance and make decisions. They need the KPIs delivered to them.

Some of these KPIs may require cross product or cross departmental analysis, which may be too manually intensive, if not impossible, to perform on raw data from operational systems. This is especially relevant to relationship marketing and profitability analysis. The data in data warehouse is already prepared and structured to support this kind of analysis.

Performance

Finally, the performance of transactional systems and query response time make the case for a data warehouse. The transactional systems are meant to do just that – perform transactions efficiently – and hence, are designed to optimize frequent database reads and writes. The data warehouse, on the other hand, is designed to optimize frequent complex querying and analysis. Some of the ad-hoc queries and interactive analysis, which could be performed in few seconds to minutes on a data warehouse could take a heavy toll on the transactional systems and literally drag their performance down.

Holding historical data in transactional systems for longer period of time could also interfere with their performance. Hence, the historical data needs to find its place in the data warehouse.

Rajiv Maheshwari, Director Solutions, CS Solutions Inc.