: Is a Data Warehouse a DSS? What is a star schema?

No, a data warehouse is not a Decision Support System. A data warehouse is however usually the "driver" and dominant component for a Data-driven DSS. A data warehouse is an organized collection of large amounts of structured data. It is a database designed and intended to support decision making in organizations. It is batch updated and structured for rapid online queries and managerial summaries of its contents. According to Bill Inmon (1993), who is often called the "father" of data warehousing, "a data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data". Ralph Kimball (1996), another data warehousing expert, says "A data warehouse is a copy of transaction data specifically structured for query and analysis". A data warehouse is often the component that stores data for a Data-driven DSS.

When a data warehouse is included as a component in a Data-driven DSS, a DSS analyst or data modeler needs to develop a schema or structure for the database and identify analytic software and end user presentation software to complete the DSS architecture and design. The DSS components need to be linked in an architecture that provides appropriate performance and scalability. In some Data-driven DSS designs, a second multidimensional database management system (MDBMS) will be included and populated by a data warehouse built using a relational database management system (RDBMS). The MDBMS will provide data for on-line analytical processing (OLAP). It is common to build a data warehouse using an RDBMS from Oracle or IBM and then use query and reporting and analytical software from a vendor like Brio or Business Objects as part of the overall Data-driven DSS design. What some vendors call "business intelligence software" often provides the analytics and user interface functionality for a Data-driven DSS built with a data warehouse component.

In a data warehouse built using a RDBMS the most common data model is called the star schema. A related model is called a snowflake schema. A star schema is organized around a central fact table that is joined to some dimension tables using foreign key references. The fact table contains data like price, discount values, number of units sold, and dollar value of sales. The fact table usually has some summarized and aggregated data and it is usually very large in terms of both fields and records. The basic premise of a star schema is that information can be classified into two groups: facts and dimensions. Facts are the core data elements one is analyzing. For example, units of individual items sold are facts, while dimensions are attributes about the facts. The star schema has also been called a star-join schema, data cube, data list, grid file, and multidimensional schema. The name star schema comes from the pattern formed by the entities and relationships when they are represented as an entity-relationship diagram. Metaphorically, the results of a specific business activity are at the center of the star schema database and are surrounded by dimensional tables with data on the people, places, and things that come together to perform the business activity. These dimensional tables are the points of the star.

How does a snowflake schema differ from a star schema? A snowflake schema is an expansion and extension of a star schema to additional secondary dimensional tables. In a star schema each dimension is typically stored in one table; the snowflake design principle expands a dimension and creates tables for each level of a dimensional hierarchy. For example, a Region dimension may contain the levels Street, City, State and Country. In a star schema, all these attributes would be stored in one table, in a snowflake schema one would expand the schema and a designer might add city and state secondary tables.

Well as the above discussion indicates creating the data model for a Data-driven DSS is a complex task. Whether DSS data is stored in a flat file, a hierarchical or multidimensional database or a

relational database management system, a large, well-organized database of business facts provides the functionality for a Data-driven DSS. A data warehouse is only part of such a system, but when it is used the data component is the "driver" for decision support.

References

Inmon, W. H. "What is a Data Warehouse", PRISM Newsletter, Center for the Application of Information Technology, Washington University in St. Louis, vol. 1, no. 1, 1993.

Kimball, R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, 1996.

The above response is from Power, D., Is a Data Warehouse a DSS? What is a star schema? How does a snowflake schema differ from a star schema? DSS News, Vol. 3, No. 4, February 17, 2002.

Author: Daniel Power
Last update: 2005-08-06 21:57