: How does DSS data differ from operational data?

by Daniel J Power

Editor, DSSResources.com

Decision support data used for analytics and data-driven DSS is related to past actions and intentions. The data is a historical record and the scale of data captured is often high volume. It may be qualitative and semi-structured. Stored data is used by both data-driven DSS and transaction processing systems (TPS). The data organized specifically for decision support differs from the data organized as a record of business transactions. Both data-driven DSS and TPS create information from the stored data that can be used by people.

In general, decision support data records business activities and a summary of transactions; operating data is a detailed record of a company's daily business transactions. Decision support data is organized to help provide tactical and strategic business meaning to operating data and relevant external data. So operating data and DSS data serve different purposes. The difference in purpose means that the data formats and structures will likely differ.

Decision support data may be traditional quantitative data sourced from internal business systems, non-traditional human sourced data like comments, text messages, or documents, and finally machine-generated data from sensors or logs may support decision making.

Managers and systems analysts must understand that decision support data and operating data usually differ in terms of six major factors: data structures, time span, summarization of data, data volatility, data dimensions, and metadata. Let's explore these differences.

Data Structures

A data structure refers to how data is organized. Often operational data is stored in a relational database management system. These systems have data structures organized in tables that have been carefully designed to avoid anomalies in the data when transactions like updating, adding records
and deleting records occur. Normalization is the process of reducing a complex data structure into its simplest, most stable structure. The process involves removing redundant attributes, keys, and relationships from a conceptual data model. Operational data is normalized.

In general, in an operational data store or transaction database, both the software and the hardware are optimized to support transactions about the daily operations of a company. In order to provide effective and efficient update performance, transaction systems store data in many small tables, each with a minimum number of fields. Thus, a simple sales transaction might need to have data elements recorded in five or more different tables. For example, we may add or update records in an invoice table, an invoice line table, a discount table, a store table, and a department table. In general, we do not add or update individual records in a DSS data store. The data is a static record of actions and occurrences.

Although this structural approach of creating many small tables is effective in a transaction database, it is not appropriate for decision support data. Queries will tend to be slow and in many cases tables will need to be joined to complete a query. For example, to create an invoice to mail to a customer from the operating database all of the tables may need to be joined. Decision support data
stores rarely include the details of each operating transaction. DSS data may include transaction summaries. In general, DSS data are integrated from multiple operational databases and DSS data are sometimes aggregated and summarized in the database to support predefined decision support queries. If it will speed up routine queries, Decision support data may have redundancies in the data structure.

Time Span

Operating data is current and it shows the current status of business transactions. DSS data is a snapshot of the operational data at given points in time. Therefore, DSS data is an historic time series of operating data. We are storing multiple "time slices" of operating data. Inmon (1992) says the DSS data is "time-variant". This characteristic is analogous to putting a time stamp on DSS data when it is loaded in the database or data store.

Summarization of data

Data loaded from operational systems is often summarized in a decision support database. Some DSS databases consist exclusively of summarized or what is often called derived data. For example, rather than storing each of 10,000 sales transactions for a given retail store on a given day, a DSS database may contain
the total number of units sold and the total sales dollars generated during that given day. Decision support data might be collected to monitor total dollar sales for each store or unit sales for each type of product. The purpose of the summaries may be to establish and evaluate sales trends or product

sales comparisons that will serve decision needs. We may want to ask questions like: What are sales trends for product X? Should we discontinue a product? Has advertising been effective as measured by sales changes? All of these questions can be answered using summarized data. Operational data is not summarized within the database.

Data Volatility

Only two kinds of operations should occur in a traditional decision support data store, loading of data and accessing data. We can add data in batches but there is no online updating and changing of data. So decision support data is non-volatile. Operating data is volatile. Operating data changes when new transactions occur.

Data Dimensions

Decision support data has multiple dimensions. This is probably the most distinguishing characteristic of DSS data. From a manager's and a DSS analyst's point of view, DSS data are always related in many different ways. For example, when we analyze product sales to a specific customer during a given span of time, we are likely to ask multi-dimensional questions. We may want to ask, "How many products of type X were sold to customer Y during the most recent six months?" DSS data can and will be examined from multiple dimensions, for example, product, region, and year. The ability to analyze, extract, and present data as information in meaningful ways is one of the major differences between DSS data and operational data. In contrast to decision support data, operational data has only one dimension, i.e., the time when the transaction occurred.

Metadata

In a Data-Driven DSS it is especially important to develop and maintain metadata about the DSS data. Data about the data in the DSS data store can help managers and analysts understand what questions can be asked using the systems. Also, because DSS data usually comes from many sources, creating metadata is especially important. Also, we need integration of DSS data that comes from different
sources. A data dictionary provides a reference about how we have combined data
from various data sources. Metadata is defined as "data about the data" in a DSS database. Metadata is semantic information associated with a given variable. This type of data usually includes business definitions of the data and clear, accurate descriptions of data types, potential values, the original
source system, data formats, and other characteristics. DSS data should have more complex

metadata associated with it in the storage system.

Conclusions

When we build a strategic data-driven DSS we are organizing and summarizing data in multiple dimensions for fast retrieval and for ad hoc analysis. The primary goal of these systems is to help managers transform data into information and knowledge. We should not compromise and use data structured for tactical decision support or operational transaction processing. The goal is to build a DSS data store with accurate data that is appropriately organized to support routine and non-routine decision support queries.

References

Inmon, W.H., "What is a Data Warehouse?", PRISM, vol. 1, No. 1, 1995.

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

Power, D., Decision Support Systems: Concepts and Resources for Managers, Greenwood/Quorum Books, 2002.

Power, D. J. Decision Support Systems Hyperbook. Cedar Falls, IA: DSSResources.COM, HTML version, 2000, accessed on (today's date) at URL http://dssresources.com/subscriber/password/dssbookhypertext.

Rob, P. and C. Coronell. Database Systems: Design, Implementation, and Management, Course Technology, 1997.

UNECE Statistics Wikis, "Classification of Types of Big Data," 2013 at URL http://www1.unece.org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data

Author: Daniel Power
Last update: 2016-08-07 06:54