: What is ETL software and how is it related to DSS?

Data warehousing practitioners have all experienced the frustration of creating a large data store from existing operating data bases built using a variety of data models and technologies. ETL software attempts to make that task easier. ETL is an acronym for extract, transform and load. This type of software is also called data extraction software and it is sometimes referred to as "pipeline software" (cf., Gray and Watson, 1998).

In managing databases, ETL refers to a software package with three primary functions. EXTRACT reads data from a specified database and writes a subset of data to a file. TRANSFORM changes the new data set using rules or lookup tables. Finally, LOAD writes the data set to a database intended to support reporting, queries, drill down and other decision support solutions.

ETL software can be used to create a temporary data set for a specific decision support purpose or it may be used to create and refresh a more permanent data mart or data warehouse.

Why is ETL software needed? Because in many situations the data that managers want to analyze is in diverse databases and in diverse formats. ETL software can get existing data into a form and format that is useful in a specific decision support task. Also, in some situations ETL software can be used to improve data quality and data integration. According to a Business Objects press release, ETL is "a critical building block to a successful business intelligence deployment".

The ETL task is a complex problem in most companies and for most decision support projects. The goal is to extract comparable data that is consistent and then "clean up" the data and structure it for rapid queries (cf. Anahory and Murray, 1997 p. 22-26).

Although I have moved data sets and recoded fields and populated new databases with existing data, I'm not an "ETL person". I know how important the extract, transform and load tasks are, but they also seem a bit tedious and boring from my perspective. You can get more information on ETL at the web sites of vendors. Three vendors are Ascential (www.ascentialsoftware.com) with its DataStage XE product, IRI (http://www.iri.com) and Acta Technology (www.acta.com). On July 9, 2002, Business Objects announced it would acquire Acta Technology.

So ETL software helps prepare data from diverse data sources for use in a data-driven DSS. ETL begins the data warehouse/data-driven DSS development process. ETL software does NOT however insure the quality of DSS data -- it is only a tool that needs to be used intelligently.

Readers interested in creating a DSS data store should check Larry English's White Paper titled "Ten Mistakes to Avoid If Your Data Warehouse Is To Deliver Quality Information" that is available at DSSResources.COM.

References

Anahory, S. and D. Murray, Data Warehousing in the Real World: A Practical Guide for Building Decision Support Systems, Reading, MA: Addison-Wesley, 1997.

Business Objects Press Release, "Business Objects to Acquire Acta Technology, Inc.", July 09, 2002.

Gray, P. and Watson, H.J., Decision Support in the Data Warehouse, Upper Saddle River, NJ; Prentice-Hall PTR, 1998.

The above response is from Power, D., What is ETL software and how is it related to DSS? DSS News, Vol. 3, No. 16, August 4, 2002.

Author: Daniel Power
Last update: 2005-08-06 22:02