

BUSINESS INTELLIGENCE

with Jackie

Back to Basics



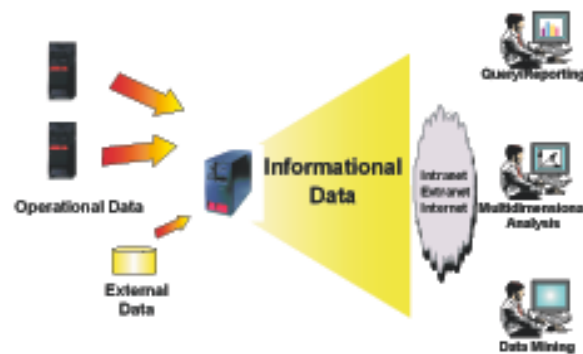
Jackie Jansen

We all know of companies that have tried to implement query tools and found that the end users weren't as happy as IT expected. In many cases the end users are trying to query production files. Now F001 might mean something to a programmer but it doesn't (and shouldn't!) mean anything to a business analyst.

The starting point in a data warehouse environment is extracting the data that you need out of your operational system and copying it into an informational database i.e. your data warehouse. Over time your data warehouse will contain a copy of all the data your end users query on a regular basis. Your data warehouse will continue to grow. It will probably start fairly small with sales data for example. You may then add financial data, next operational data, and then data for human resources etc. Your warehouse will often contain data from outside of your production system. Many companies have important information such as budget and forecast data stored only in users spreadsheets. This information should be regularly uploaded into the data warehouse. There are also a lot of external databases that a company can purchase. These might include census data, demographic data, retail data etc.

When you designed the files for your production system you took into account rules like; never duplicate a customer's name and address in multiple files in case they get out of sync, don't put 3 prices

breaks in 1 product record in case the company introduces a 4th level. Just as you learned how to design for production or online transaction processing systems (OLTP), you now need to learn new skills when it comes to designing your data warehouse.



So far I have oversimplified matters when I talked about copying your data from production into your warehouse. This is actually a whole topic unto itself. This stage includes extracting changed data from your production system, transforming and cleansing that data and inserting it into newly designed file layouts.

There are 3 standard ways of accessing your informational data. Today the most common method is using SQL based query and reporting tools. Most of these tools interface with DB2/400 via ODBC. According to the Gartner Group the next category, called multidimensional analysis, will overtake the use of SQL-like query tools by 2001. You may have

heard the terms OLAP and/or "cube". These are basically synonyms for a multidimensional database. Products such as DB2 OLAP are used to build multidimensional data marts. Another method of accessing your data is using data mining tools. These tools are very different from the ones we have just discussed. Typically data mining tools are used by a lot fewer people in an organization than query or OLAP tools. If the fit is good the pay back for data mining can be extraordinary. Some common data mining algorithms are used for customer segmentation and market basket analysis.

Often these various access methods don't go against the data warehouse itself but rather against a data mart. A data mart can be as simple as a summary table or it can be as complex as a full star schema or snowflake design covering one major informational area, i.e. sales. A multidimensional database is a specific type of data mart.

This has been a very quick review of a data warehouse architecture. If you wish to discuss this further please don't hesitate to contact me at jjansen@ca.ibm.com.

T&G

Jackie Jansen is a Certified Consulting IT Specialist. She currently works in the IBM Americas Business Intelligence Solutions Centre. Jackie is a frequent speaker at AS/400 Technical Conferences and User Group meetings. Contact her at jjansen@ca.ibm.com.