

**RCW**

Report Content Writer
www.reportcontentwriter.com.

AUTHOR: Rachel Agheyisi

EMAIL: rachel@reportcontentwriter.com.

TWITTER: www.twitter.com/rachel_rcw.

Business Intelligence and ETL

Business intelligence (BI) is a tech-based tool that allows decision makers process information to gain insights that drive enterprise performance. However, the path from information to knowledge is not always linear. Typically, a series of steps, including collection, organization, evaluation, and transformation precede the data analysis and reporting that provide insight.

Issues relating to data access – the so-called ETL components – have significant implications for practically every phase of a BI project. ETL stands for **extraction, transformation, and loading** of data. The implementation of ETL depends on factors such the type of data, number of data sources, data volume, and the preferred mode of data transportation. This article highlights ETL in the context of a BI project.

E is for Extraction

Extraction involves identifying the relevant data sets and pulling them from one or more operational and transactional source systems. The size or volume of the data is a function of business operation. It is not unusual for businesses to generate huge megabytes of data from multiple sources on a daily basis. The choice of extraction method depends on the specifications of the data source systems and the target warehouse environment.

One of the challenges of the extraction phase is making sense of possible differences in the formats of the data source systems. Examples of commonly used formats include flat files, relational databases, and non-relational databases. The variety of formats necessitates the parsing of extracted data to determine their cohesiveness and conformity with underlying structures.

The data extraction method may be full or incremental. The full extraction method means that all currently available data are pulled completely from the source systems. This method assumes that the available data set is, indeed, complete. The incremental method involves the extraction of only the portion of data that has changed since a marked historical period of event. However, in order to utilize the incremental method, the data warehouse has to incorporate change capture techniques as part of the extraction process.

Other considerations in the extraction phase include whether or not to incorporate data transformation as part of the process, whether to extract the data directly from the source systems, or extract them from a staged area outside of the original source systems. None of these considerations is trivial.

T is for Transformation

The transformation phase of ETL involves preparing the extracted data for loading to the target database or databases. Some data may require little or no transformation, while other data sets will need to be restructured, cleaned, and organized before transportation. Typical types of data transformation include:

- Re-mapping
- Sorting
- Filtering
- Cleaning
- Aggregation
- Pivoting
- Recoding
- Standardization

SQL functionality makes several of these transformations relatively routine in most BI environment.

L is for Loading

Experts consider data loading from the source systems a simple task in the ETL process, assuming a successful completion of the extraction phase. Depending on the extraction method, data loading may overwrite existing data, be cumulative, or involve a blend of both. Each requires specific constraints, triggers, and audits to ensure the preservation of the integrity of the source data.

The most direct loading route involves moving the extracted data from a source system directly to a target database in a data warehouse. Other less straightforward transportation routes involve an intermediary (pre-warehouse) staging database or an additional (post-warehouse) database.

Tools of the Trade

Although relatively easy to compartmentalize, the actual implementation of ETL varies from enterprise to enterprise. In practically all cases, ETL is resource-intensive. This explains the evolution of ETL tools, which are intended to facilitate the processes.

Practically, all the top BI brands include ETL tools. High profile examples include:

- Microsoft's SQL Server Integration Services (SSIS)
- Oracle's Warehouse Builder (OWB)
- IBM's Manager WebSphere DataStage
- SAP- Business Object's Data Integrator

- MicroStrategy's MicroStrategy 9

Conclusion

The foregoing are highlights of the ETL processes in a BI project. Any enterprise seriously considering BI deployment needs to commit significant amount of time to shop around. White papers will prove to be helpful research resource for the quick education of current and prospective BI users. Most vendors post these reports on their websites. However, less focus on the flashy side of BI will prove beneficial in identifying what is ultimately useful for the enterprise.

About the Author: Rachel Agheyisi is a business writer who specializes in white papers and case studies that address regulatory compliance and business intelligence issues. Related articles are available on her website at www.reportcontentwriter.com.