



HOW ETL DESIGNING CAN SUPPORT EFFICIENCY ON DATA MOVEMENT

TECHNICAL PAPER (English Language)

TABLE OF CONTENTS

1 OVERVIEW 3

2 KEY ELEMENTS OF AN EFFICIENT ETL 4

3 ETL FOUNDATIONS APPLIED 5

4 CHALLENGES ON ETL ADOPTION 6

5 SUMMARY..... 7

6 HOW CAN WE SUPPORT YOUR DATA INTEGRATION PROJECT 8

1 OVERVIEW

Extract, transform, and load (ETL) is a term used to describe the movement and transformation of data between systems involving high data volumes and complex business rules. Enterprise databases typically merge data from many different sources that have a plethora of formats and purposes.

ETL software brings all the data together in a standard, homogeneous environment. ETL processes are especially important in today's growing big data world. Collecting and storing terabytes of data is useless unless it can be leveraged in a meaningful way. ETL provides data quality and profiling to ensure the trustworthiness of data, and transforms it so it can be used for business intelligence.

Automated ETL tools are widely used in data integration, data migration, and master data management projects and are critical for data warehouses, business intelligence systems, and big data platforms because they can be used to retrieve data from operational systems and process it for further analysis by reporting and analytics tools.

The reliability and timeliness of the entire business intelligence platform depend on ETL processes.

2 KEY ELEMENTS OF AN EFFICIENT ETL

- **Reusable components** that can be automated to perform data movements jobs on a regular basis

- **Performance** supporting massive parallel processing for large data volumes

- **Hardware** - The reduced data in the data warehouse requires less storage memory

- **Administration** – Reduced data also requires less security overhead

- Designed to be very **efficient**, scalable, and maintainable

3 ETL FOUNDATIONS APPLIED

EXTRACT

The first part of an ETL process is to extract the data from the source system. In this stage the data is converted into a single format to prepare it for the transformation stage. Extracting data correctly is critical since it is the foundation for the rest of the ETL process, and, if not done correctly, can result in failure of the entire project. Because most ETL tools consolidate data from multiple sources/systems, it can be a challenge to integrate data that is often in disparate formats.

TRANSFORM

This stage of extract, transform, load applies a series of rules or functions to the extracted data to transform it into the finished product that will be loaded into the destination. This involves cleaning, applying business rules, checking for data integrity, etc. Some data requires no transformation, but often one or more transformations may be required to meet the business and technical needs of destination, including joining data, transposing data, disaggregation, lookup, and simple or complex validations.

LOAD

The load phase of an ETL application moves the cleaned and transformed data into the destination, usually a data warehouse, data mart, or operational data store. This process varies widely depending on the requirements of the organization. Some data warehouses update existing information with the extracted data on a daily, weekly, or monthly basis. Other data warehouses may add new data in an historical form at regular intervals—for example, hourly. The timing and scope for replacing or appending data are choices made depending on the time and resources available, as well as business needs.

4 CHALLENGES ON ETL ADOPTION

The challenges to implementing reliable ETL processes in today's world of massive and complex amounts of data are many.

DATA VOLUMES - Thanks to the availability of massive amounts of data today, data volumes are growing exponentially. And while some business systems need only incremental updates, others require a complete reload each time. ETL tools must scale for large amounts of both structured and unstructured (complex) data.

DATA SPEED - Businesses today need to be connected at all times to enable real-time business insight and decisions and to share the same information both externally and internally. As business intelligence analysis moves toward real time, data warehouses and data marts need to be refreshed more often and more quickly. This requires real-time processing as well as batch processing.

DISPARATE SOURCES - As information systems become more complex, the number of sources from which information must be extracted are growing. ETL software must have flexibility and connectivity to a wide range of systems, databases, files, and web services.

DIVERSE TARGETS - Business intelligence systems and data warehouses, marts, and stores all have different structures that require a breadth of data transformation capabilities. Transformations involved in ETL processes can be highly complex. Data needs to be aggregated, parsed, computed, statistically processed, and more. Business intelligence-specific transformations are also required, such as slowly changing dimensions. Often data integration projects deal with multiple data sources and therefore need to handle issue of having multiple keys in order to make sense of the combined data.

5 SUMMARY

Issues such as the data, technology, and people also influence ETL processes. Data issues include quality of data, similarity of source and target structures, kinds of data dependencies, how meta data is used, and complexity of data relationships.

Technology issues include volume and frequency of load, hardware and memory robustness, interoperability of source and target platforms, and scheduling.

Finally, people issues include the level of technology with which management feels comfortable, amount of in-house expertise, and who will support the ETL tools.

These factors influence the approach to loading the data warehouse, as well as the cost of the solution in terms of labor and products and its ease of development and maintenance. Before investing in a solution, it is important for organizations to do a thorough ETL tool comparison in order to ensure that they are getting the functionality, scalability, and performance needed at a reasonable price.

6 HOW CAN WE SUPPORT YOUR DATA INTEGRATION PROJECT

Axiate ETL specialists can work side by side with you to plan and execute any or all of the following objectives:

- Build an enterprise data warehouse, departmental data mart, or operations data source
- Accommodate assorted data sources or volumes
- Leverage flexible batch windows that include scheduled batch, near-real-time integration, and real-time integration
- Create tools for dimension table management, lookup caching, and data profiling
- Load data incrementally with the advanced change data capture functionality
- Design advanced transformations for efficiently processing complex data
- Robust parallel processing engine to handle large volumes quickly and efficiently
- Create sophisticated job scheduling and management capabilities
- Design easy-to-use interfaces enabling business users productivity