

Overview

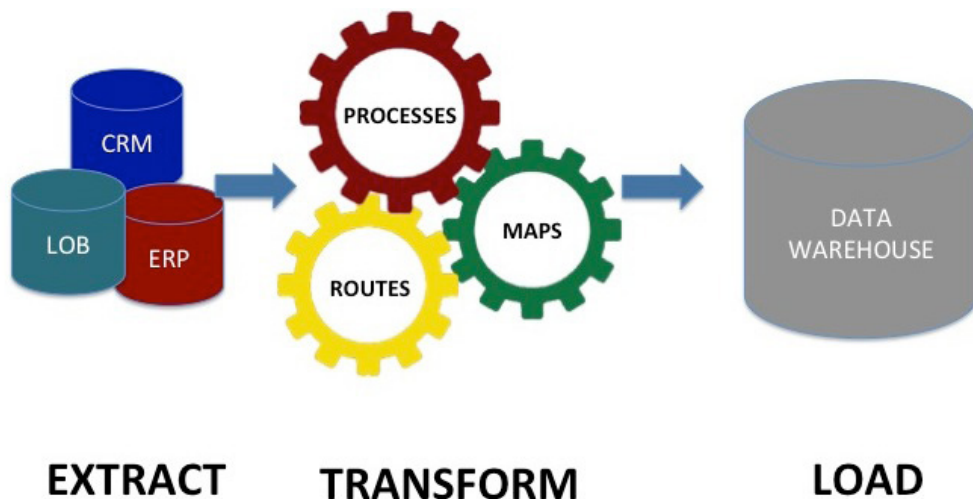
Extract, transform, and load (ETL) is a term used to describe the movement and transformation of data between systems involving high data volumes and complex business rules. Enterprise databases typically merge data from many different sources that have a plethora of formats and purposes. ETL software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

Benefits of ETL

- Efficiency - Reusable components that can be automated to perform data movements jobs on a regular basis
- Performance - Supports massive parallel processing for large data volumes
- Development - Because only relevant data is extracted and processed, development time is reduced and because only targeted data is loaded, the warehouse only contains relevant data
- 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



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.

Challenges of ETL

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.

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.



Centerprise for Smart, Powerful, and Scalable ETL

Centerprise is the ideal solution for extracting source data, transforming that data to suit your needs, and loading it into your data warehouse or database. With Centerprise ETL, you can elevate data accuracy and timeliness, so end users can be confident in the reliability and effectiveness of the warehouse.

With an interface designed for both technical and non-technical users and an agile data integration platform, Centerprise is scalable to meet even the most demanding data warehouse needs. Centerprise ETL offers data warehouse loading functionality, including the Slowly Changing Dimension (SCD) Transformation. It simplifies the end-to-end development, debugging, and maintenance process for intricate dataflow, transformation, and cleansing procedures. The powerful yet easy-to-use workflow environment and automation features significantly simplify complex data retrieval processes.

Centerprise ETL Benefits

- Build an enterprise data warehouse, departmental data mart, or operations data source
- Easily accommodate assorted data sources or volumes
- Leverage a flexible batch window that includes scheduled batch, near-real-time integration, and real-time integration
- Built-in tools for dimension table management, lookup caching, and data profiling
- Load data incrementally with the advanced change data capture functionality
- Built-in advanced transformations for efficiently processing complex data
- Robust parallel processing engine handles large volumes quickly and efficiently
- Sophisticated job scheduling and management capabilities
- Easy-to-use interface enables business users to be productive, lowering reliance on IT resources

Centerprise Solutions



Data Warehousing

High-performance data warehousing ETL features in a unified, intuitive environment.



CDC

Choose either a batch or real-time change data capture strategy for your particular requirements.



Data Mapping

Transforms advanced mapping, validating, and cleansing tasks into basic drag-and-drop or single-click commands.



ETL

Extract data from any source, transform it to suit your needs, and load it into your database or warehouse.



Data Conversion

Complex doesn't need to be complicated. Visual, code-free parsing, transforming, and loading of data from any source.



Data Integration

A single platform for complex, hierarchical integration the requires no coding.



Data Migration

Unique hierarchical data processing technologies automate and streamline data migration projects.



EDI

Full electronic data interchange functionality combined with Centerprise complex data mapping capabilities.



www.asteria.com

Contact us for more information or to request a free trial
sales@asteria.com 8888-77-ASTERIA

©2015 Astera Software Incorporated. All rights reserved. Astera and Centerprise are registered trademarks of Astera Software Incorporated in the United States and / or other countries. Other marks are the property of their respective owners.