

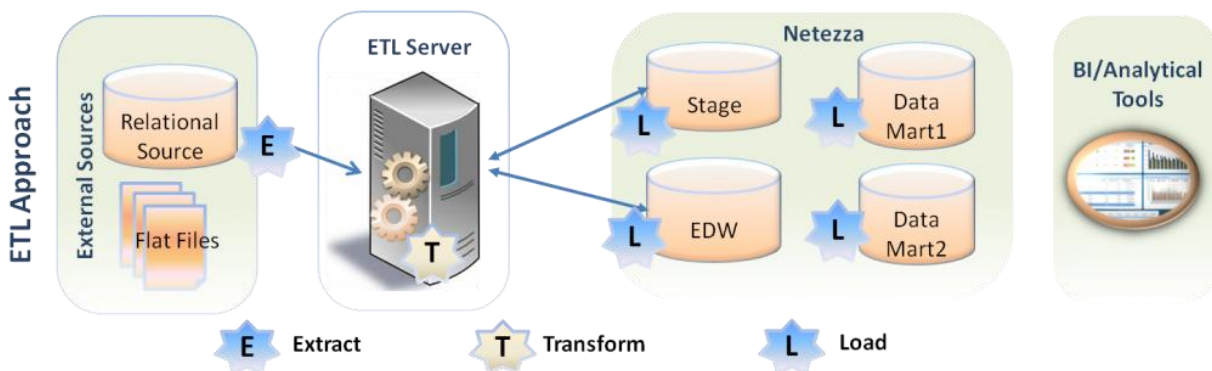
ETL vs. ELT: Frictionless data integration

As the volumes and varieties of structured, complex, semi-structured and unstructured data both external and internal to the enterprise are growing at an astronomical pace, more efficient ways to source, move, enrich and publish this data becomes critical to the business and enterprise architecture. Ensuring the veracity of your data is also crucial to valid analysis and reporting.

ETL

Traditional data integration and ETL tools are becoming an inhibitor to timely availability of high-value data to the business and will not be able to scale effectively with these ever-growing volumes of data.

In addition to the costs and challenges of attempting to scale these environments as data volumes grow, the issue of data latency due to the intermediate systems required for these platforms becomes a bigger and bigger threat to the enterprise.

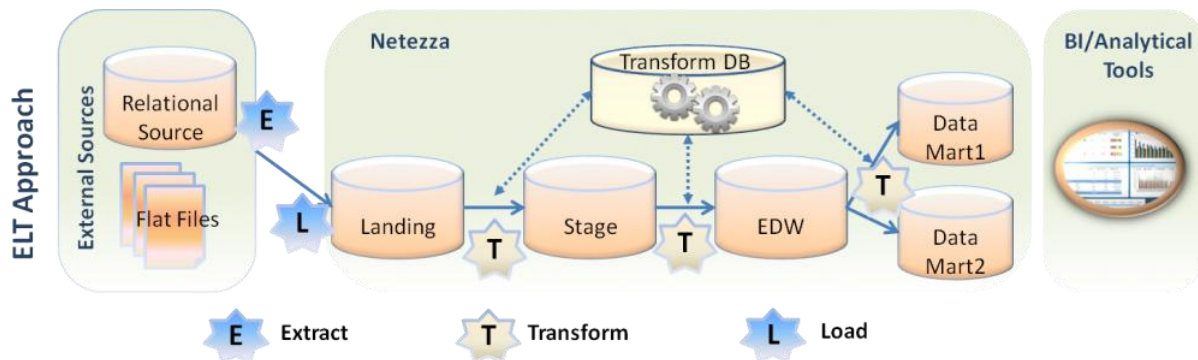


Another issue with traditional ETL platforms is that they cannot easily take advantage of the rich array of analytical functions available with today's data warehouse appliances. They also offer only a limited SQL-pushdown capability which requires additional software to be purchased and manual changes to be implemented.

ELT

An approach to eliminating this multi-stop data movement is with a frictionless approach to data integration. By removing the middleman platforms and loading source data directly into the target environments, data becomes available significantly sooner.

Friction-free ELT is achieved through direct data movement and 100% SQL pushdown – the benefits can be demonstrated quickly and easily.



As an example, with one of our Financial Services customers, by removing the Informatica work flow which was extracting data out of an Oracle database, processing some local transformations and then inserting or updating that data into a Netezza TwinFin in typical ETL fashion, we replaced that by using the power of frictionless data movement to extract from Oracle and load directly into Netezza and then take advantage of set-based parallel processing to perform the inserts and updates directly within the database, the end-to-end time to accomplish this job was reduced from over 6 hours to under 10 minutes, most of this time loading the data.

ELT in the LOGICAL DATA WAREHOUSE

By leveraging the power of your massively parallel processing data warehouse appliances and large-scale distributed Hadoop/HDFS systems to perform the necessary transformations in-place, the need for legacy ETL tools such as Informatica, IBM DataStage or Ab Initio goes away. In addition, the capability of leveraging MPP and distributed systems most effectively and efficiently is realized.

You can leverage Hadoop to process and identify the high values structured and semi-structured data you want in the warehouse as well as offload data no longer wanted in the warehouse but needed for access through the less expensive, higher-latency Hadoop environment. This helps you more readily move to a logical data warehouse model with the right data in the right platform at the right time.

TRACKING DATA LINEAGE

One problem with leveraging the ELT capabilities of MPP data warehouse appliances is that typical ETL tools or custom-coded SQL and scripts lose the ability to track data lineage once the data is in the database. A true complete ELT solution addresses this by ensuring all the details about where your data came from, what you did to it along the way, and where it ultimately ended up are captured and stored in a metadata repository.

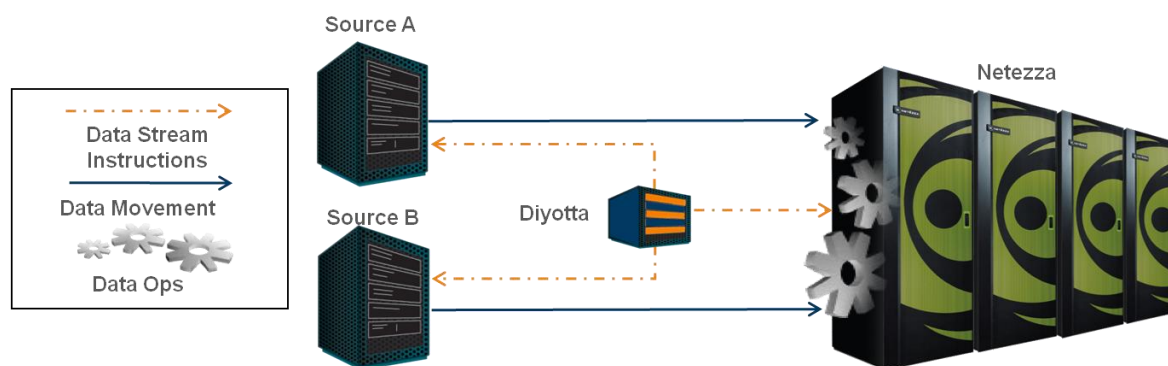
This allows you to perform comprehensive impact analysis across the data integration stream and provide the details necessary for compliance with federal regulations regarding data lineage.

FRICTIONLESS ELT for MPP

Having a friction-free data integration suite at your disposal that allows you to create the mappings, workflows, transformations, execution and monitoring of your ELT jobs in a collaborative graphical integrated development environment (GUI IDE) while taking advantage of true ELT with end-to-end data lineage capture and impact analysis allows you to move away from the risks inherent to the connecting flight paradigm with current ETL solutions to the benefits of a non-stop flight.

HOW DIYOTTA WORKS

- Objects are imported into the Metadata repository
- Data streams are designed in the Studio
- The Services Engine sends data stream instructions to source systems, pushing data to the target system
- Data Stream instructions are sent to the target, executing transforms in the database



For more information please visit us at www.diyotta.com