

Progressing the journey from ETL to ELT with UniConnect



Extract > Load > Transform

A Percipient Technology White Paper

Author: DK Sharma
Americas Business Head
Updated Nov 2017

ETL has been able to hold on its own in environments where data grew at a few hundreds of megabytes per day.

ETL (Extract from source systems, translate into target data formats in a staging area and Load in the target system) as the enabler of data analysis struggles when confronted with requirements of unstructured, voluminous and volatile data. It is our submission that it has outlived its usefulness in the dynamic environment of ever growing data that organizations find themselves dealing with day in and day out. ETL has been able to hold on its own in environments where data grew at a few hundreds of megabytes per day.

But, when incremental data is coming at the rates of terabytes per day and increasingly in newer and disparate forms, organizations using ETL tools to integrate data for analytic purposes find themselves in a vicious cycle of endless design and test of newer schemas for their staging environments without being able to focus on the end game, which is to generate significant insights that this data could provide

ELT models allow faster turnarounds in use of data, greater flexibility and agility in incorporating new sources and changing definitions,

With commodity hardware and storage now providing potentially endless compute capability, either on premise or in the cloud, ELT (Extract and Load as per source format with some basic filtering if required, and translate on demand or translate at target) models can provide a much needed relief from the quicksand that ETL users find themselves sinking into. This take greater significance when enterprises are faced with an avalanche of ever increasing data of interest to their internal clients and applications. ELT models allow faster turnarounds in use of data, greater flexibility and agility in incorporating new sources and changing definitions, all done in a highly cost efficient manner on low cost hardware and using open source technologies.

However the flexibility afforded by ELT models comes at the cost of time taken at the target, either during data lake creation or dynamic translation at the time of use and in terms of additional expertise and skills required to manage this translation on the fly. Technologies from hardware vendors such as Intel are now allowing organizations to allay the concerns associated with high latency performance of downstream / on-the-fly

translations. However, there is a need for software solutions that can take advantage of this compute power.

Uniconnect provides in-built support for a rich library of functions that can operate on the underlying raw data in high speed memory, thereby enabling the T in ELT to happen at lightening speeds.

UniConnect from Percipient enables data transformation on-the-fly, making full use of the available compute resources over a cluster of processing nodes. This is achieved using highly reliable and resilient open source resource management tools such as YARN. Uniconnect provides in-built support for a rich library of functions that can operate on the underlying raw data in high speed memory, thereby enabling the T in ELT to happen at lightening speeds. End users can also register custom functions and routines build to take care of their unique requirements into UniConnect's library of supported functions and easily extend this power of dynamic translation using Uniconnect. And the users execute these translations via simple SQLs.

Moreover, with the ability to join data in-memory created by legacy ETLs that are loaded into traditional data marts with data translated on the fly, Uniconnect enables the best of both ETL & ELT.

