***Abstract*:This study focuses on the deep understanding, analysis, and appreciation of relative effectiveness of the two different data management techniques-ETL and ELT in data warehousing. Paper discusses and compares various aspects like resources, time, volume and type of data etc. so that deploying an appropriate technique saves valuable data resources and development cycles.**

***Keywords:*** ETL, ELT, DM, BI, Data Provision, Data Warehouse.

I. Introduction

*A. Background*

It is very crucial to decide over the architecture to be employed in a project as three quarters of the funds are attributed to the design and development of the data warehouse. Implementation cost is very less when compared to design and development, it is vital decision to choose the right technique because a faulty decision can be costly and may not reach the objectives of the project.

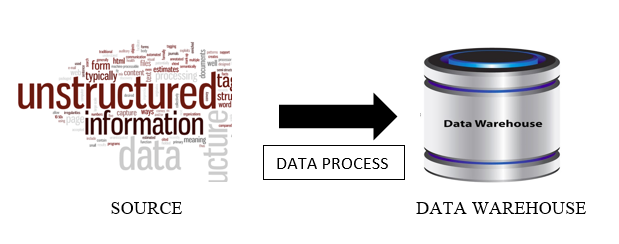
*B. Brief Overview on DM*

DM-Data Management is not a technology, it is a regulatory procedure through which the necessary data is obtained, validated, stored, secured, and processed. By this information openness, unwavering quality, and convenience is guaranteed.

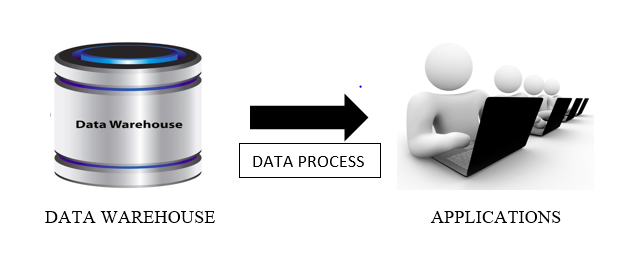
Data management relates with overall data lifecycle, from the data generation to its dissolution, and the controlled betterment of data to and from each level within its lifecycle. Data management helps to minimize the risks and costs of regulatory disagreement, legal issues, and security breaches. It also provides access to relevant data whenever required, without any uncertainty, thereby helps to be precise without any misperception.

Data management is essential in two areas:

* Processing the data while moving unstructured data from source to Data Warehouse.

Fig. 1 Data from source to warehouse

* Transforming the data into user requirement while moving data from Data Warehouse to other application.

 Fig. 2 Data from warehouse to different applications

Above two figures show the flow of data from source to data warehouse and from data warehouse to applications. Data is processed between these two processes and the Data Process which is depicted in the above two figures includes three main stages as listed above:

Extract – The process by which data is obtained from the data source.

Transform – The process to mutate the source data into a format relevant to the solution.

Load – Loading of data into the target.

II. Overview on ETL and ELT

*A. ETL*

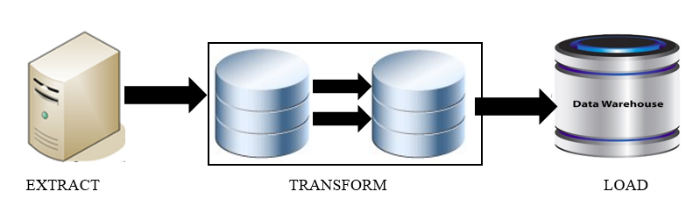
Extract, Transform, Load is the conventional and most widely used method for developing a Data Warehouse. It is used to pull out data from one or more operational systems and placing it into a data warehouse. The below figure elucidates the stages in ETL.

Fig. 3 ETL process

With reference to the above figure, we can say that, the data is first *extracted* from different data sources using a data extraction tool. Then, a series of *transformation* patterns are applied on this extracted data to make it into the target format. Then, this data is *loaded* into the data warehouse. Let us look into each action in the process in detail.

1) Extracting the data**:** An apt and proper extraction is a key to success in developing data warehouses and data marts. Data extraction is the process of retrieving data from data sources for further data processing.

During this, the required data is recognized and taken out from different sources, including database systems and applications. There will be cases where in, more data should be extracted than necessary and relevant data is identified at later point of time when desired data could not be identified before.

All the data obtained from different source systems is changed into one structured data warehouse format which is to be sent for transformation processing.

2)Transforming the data: It is the process of altering data from source format to target database format. At this stage different set of rules are applied to transform the data from the source to the destination. The following are the tasks involved in this step.

* Applying business rules for joining data from different sources, generating keys, sorting and deriving new calculated values.
* Cleaning – ensures quality of data in the data warehouse. Examples include, mapping NULL to 0 or MALE to ‘M’, validate address fields etc.
* Filtering – selecting only specific columns to load.
* Applying data validations.

3) Loading the data**:** Once the data is transformed into a structure consistent with the target data warehouse requirements, data is ready for loading into the data warehouse.

The first step consists of populating the tables in the data warehouse and then verifying if the data is ready to use. After loading the data, referential integrity between data sets is verified to ensure that all data relate to particular data in other tables.

ETL process requires a separate tool that transforms the data before loading into target. There are many well-known commercial tools of ETL. Informatica, Datastage, IBM Optim, SQL Server Integration Services, Data Integrator are some of those.

*B. ELT*

This is an alternative to ETL process whose purpose is also transferring raw data from a source system to a target system. The below figure illustrates how the stages are different from the already discussed ETL process.

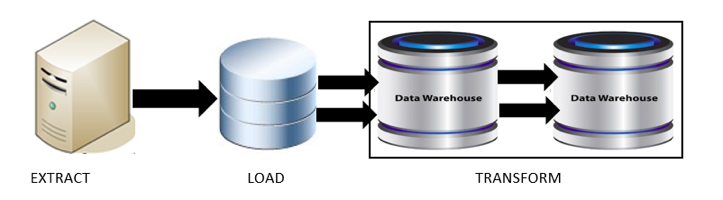
**

Fig. 4 ELT process

In contrast to ETL where data transformation is performed at an intermediate staging area before loading into the server, ELT permits the raw data from different source systems to be loaded directly into the target and then transformation is applied. This ability is very useful when handling large data sets required for business intelligence and when the target is a high-end data engine, such as a data appliance, Hadoop cluster.

Except the order, the three actions, Extract, Transform and Load are same as described in ETL. As the data transformation is performed at the tat system, ELT does not need any separate transformation tools. Instead, it needs a powerful target system that can transform the data after loading. Cloud installation, Hadoop cluster, Data appliance are some of the ELT tools.

III. Analyzing Aspects

Here are the various aspects that are taken into account to analyze the effectiveness of ETL and ELT techniques:

* Volume of data
* Environment
* Source and Target Database
* Type of data
* Resource
* Tools availability
* Cost
* Throughput and Functionality

In depth analysis of each aspect is discussed below:

*A. Volume of data*

Data volume often determines whether a DWA should develop a ETL or ELT technique. In general, ETL processes the information row-by-row and so the data integration into third party tools would work well. ETL can balance the workload and share the workload with the RDBMS when dealing with Teradata. But tools like HADOOP which processes through ELT technique can handle Big Data when compared to ETL which need better processing system for high volume of data.

*B. Environment*

The target system is built before the data processing solution in a typical Date Warehousing system implementation. As a result, the target system is optimized to perform the tasks it was meant to and not much else. This is mainly because either data processing requirements were not considered at the time of system creation or the requirements did not evolve enough to build out the system.

In ETL process, transformation takes place before loading the data into the target environment, by which target environment is not effected. Where as in ELT process, transformation takes place in target side which may affect the processes which are already running in that environment. So it is always better to consider all the processes before design and develop.

*C. Source and Target Database*

In ETL, Data Overhead is reduced as only the desired data is loaded into target and also simplifies the security regime. On the same side, the targeted data cannot be used for any future requirements as the target has only designed for the current requirement. This leads to re-design and development of ETL routines which affects the cost and time.

On the other side, in ELT all the extracted source data is loaded into target before transformation, which can be used for future requirements.

*D. Type of data*

Data provision activity can be taken place on any type of data such as production environment, development environment, QE & A Environment etc. If the data is being moved from production environment to any other non-production environment it might cause Data breach. Domains like banking, Retail, insurance has huge amount of sensitive data in the Production environment.

ETL transforms the data before moving the data into target environment. Production Data can be masked on while. Even though the data can be masked after loading the data using ELT process, client doesn’t accept the ELT approach in such cases as they are afraid of data breach.

*E. Resource*

Human Resource: ETL has much trained resources in market for the tools like Ab Initio, Informatica and many other grid tools. ELT tools such as Hadoop, Data Appliance are currently emerging, so there are limited human resources and need to be rigorously trained. If the existing target is powerful system and can handle high level transformation on the data, the existing resource will be able to perform the ELT task effectively. In this case ELT takes over ETL.

Software Infrastructure: ETL requires a new infrastructure to install the tool whereas ELT doesn’t require any new environment as we do the transformation on existing target side, so target needs to be powerful enough to handle it.

*F. Tools availability*

There are good number of tools available to implement ETL process which allows the user to identify the most appropriate tool for their purpose. Informatica, Datastage, IBM Optim, SQL Server Integration Services, Data Integrator are few notable tools in ETL side. Being an emerging technology, ELT has limited Tools available such as Data appliance, Hadoop cluster, cloud installation.

*G. Cost*

ETL system or licensing tools require extra cost to build it. And each feature like ‘Connection to Non-Relational Databases’ require more licensing costs. As ETL need to be installed on separate hardware and software it requires separate cost. ELT tools like HADOOP are free licensed tools and doesn’t need any other hardware, as the transformation is done on existing target side.

*H. Throughput and Functionality*

A typical ELT process gives more than 2X times throughput than ETL process, as the transformation doesn’t take place on the flow. ETL process has less throughput as the data is processed row-by-row and is transformed on the flow before loading into the target.

As ELT is an emerging process, it provides low level transformation functionalities, where ETL is a traditional and has prolific number of tools which provides high level functionality compared to ELT.

IV. Conclusion

Data Management and its techniques are evolving over the course of time, utilizing ETL without a doubt will limit the ability to support this change. In the very best scenario unnecessary expense would be incurred and at the same time leaves at terrible risk. Considering the pro, cons of each technique and analyzing various aspects clearly depicts that it is the decision of the individuals to decide upon based on their project requirement. Some new emerging hybrid techniques like TELT, ETLT and even TETLT are making attempts to overcome the drawbacks, but in fact they seem to increase the complexity.

Acknowledgement

This work is supported by Dr. Atif Farid Mohammad who made us aware of all the basic concepts essential for digging deep into our study. This work is also supported by TDM CoE team of “Cognizant Technology Solutions” India.

Reference

* <http://www.iri.com/blog/data-transformation2/etl-vs-elt-we-posit-you-judge/>
* <http://searchdatamanagement.techtarget.com/definition/data-management>
* <https://www.informatica.com/services-and-training/glossary-of-terms/data-management-definition.html#fbid=EU8Js82h81a>
* <http://www.dataacademy.com/files/ETL-vs-ELT-White-Paper.pdf>
* <http://richrelevance.com/downloads/RichRelevance%20Tech%20Note%20-%20ETL%20vs%20ELT.pdf>
* <https://www.ironsidegroup.com/2015/03/01/etl-vs-elt-whats-the-big-difference/>