International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 4, 2022

Available at www.ijsred.com

A Study on ETL and its Applications in the field of Finance

Sreelakshmi[1], Hasifa A S[2], Kavitha S N[3]

[1](Information Science and Engineering, R V College of Engineering, Bengaluru, India, sreelakshmi.is18@rvce.edu.in)
[2](Information Science and Engineering, R V College of Engineering, Bengaluru, India, hasifaas.is18@rvce.edu.in)
[3](Information Science and Engineering, R V College of Engineering, Bengaluru, India, kavithasn@rvce.edu.in)

----------**********----------

Abstract:

ETL (expands to extraction, transformation, and load) tools play a vital role in the data integration strategies and are used for the management of data. They allow businesses to gather the data from multiple sources and then consolidate it into a single and centralized location. They are primarily programmed to clean the data that is captured from different sources and combine it. After all the data is collected by the ETL tools and stored, the data can then be used for further analysis. It also makes it feasible for a variety of data to work together. Therefore, an ETL process in finance is crucial and forms a base of financial data analysis. Since the data is generated in large numbers and variations, the importance of ETL is growing by the day. ETL tools are designed to handle massive amounts of data. This study first defines and understands what exactly the ETL process is and then discusses what makes it important in business analytics especially in the financial domain.

Keywords — **ETL, data warehouse, business intelligence, finance markets.**

----------**********----------

I. INTRODUCTION

To control the flow of funds, banks, financial organizations, and businesses in general require cutting-edge tools and technology. Data plays a part in financial management as well. The world generates a massive amount of financial data. As a result, it is crucial for organizations to use the most up-to-date technologies and procedures for managing financial data. To ensure smooth storage and transfer of financial data, most institutions and businesses eliminate obsolete financial tools and processes. Because the volume of data has expanded dramatically, businesses all over the world have realized that they would require the most up-to-date financial management or data management technologies. Simultaneously, there is a requirement for real-time data handling. To smooth real-time data flows, financial services will require a more modern data infrastructure. The latest ETL tools are proving to be quite beneficial to the finance sector. The use of ETL solutions for financial management will only grow in the future years.

Fig. 1 describes the ETL process as a collection of data pipeline techniques. It extracts data in its original form from its resources, processes and aggregates it (transforms), and stores it in a data warehouse (loads) or database where it can be

studied. The entire ETL procedure gives the organization's information structure. Instead of trying to perform procedures to obtain relevant data at each level, It allows us to spend more time studying unique questions and gaining fresh insights.

An ETL pipeline; also known as a data pipeline, is the system that allows ETL activities to take place. Data pipelines are a set of tools and actions for transporting data from one system to another, where it might be stored and managed differently.

An ETL process gathers and modifies different types of data before delivering it to a data warehouse. Data can also be migrated amongst a variety of sources, destinations, and analytic tools using ETL.

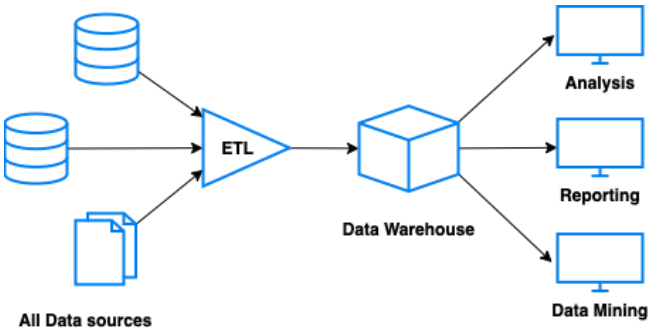

Fig. 1 ETL Process

A. Extract

The ETL process's "Extract" stage focuses on gathering data from all the data sources. The rows and columns of the analytics database will be created from this data. As these were the key sources of information for enterprises, extraction entailed collecting data from Excel files and Relational Management Database Systems. With the rise of Software as a Service (SaaS) applications, major chunks of organizations now discover useful data in the programs themselves. Data extraction nowadays is largely about using APIs or webhooks to get data from an application's storage like stock price, interest rate etc. in the financial domain.

B. Transform

The "Transform" phase of the ETL process alters (transforms) the information obtained during the extraction stage before saving it to the analytic database. It involves applying business rules to it to ensure its consistency and quality. It aims at data integrity to ensure that data sent to the target is consistent, clean and enriched. There are multiple types of transformations broadly falling into two categories:

1) ***Data Cleaning:*** Data Cleaning involves removing or correcting any suspicious data. For example, removing missing values, removing outliers, changing data types etc.

2) ***Data Enriching:*** This process involves addition of any new information to the already collected information by applying business rules, calculations etc. For example, joining various sources, calculating any values from other values etc.

C. Load

"Loading" after the transformation process involves storing the data to any target data source like Relational DBMS, NoSQL database like MongoDB, data warehouse etc. A given schema depending on the business requirement is created and meta data is formed. When the data is loaded for the first time, we have an initial load wherein the schema and meta-data is created and further loads just require storing of the data according to the schema.

The ETL process should be designed to ensure that the data pipelines and analytics add value to the business and suit all the requirements. Following are the benefits of a well-engineered process:

1) ***Information clarity:*** Data is cleansed and merged across sources during ETL transformations before being saved in the database, where it may

subsequently be analyzed. These actions enable us to work with clear data and decipher ambiguous, raw data.

2) ***Information completeness:*** All of the business sources that are relevant to operations are collected in one location (the destination - data warehouse / database) in a well-designed ETL pipeline. There are no missing puzzle pieces because all of the information is complete.

3) ***Information quality:*** Data is validated throughout the extraction phase, and data is corrected or discarded during the transformation process. This ensures that data quality is always checked before it is processed, boosting confidence in the analysis and providing a way to leverage data for data-driven decision making and business intelligence.

4) ***Information velocity:*** When new data is added to the sources or current data is modified, ETL processes can be designed to activate the complete ETL pipeline. As a result, we have control over the data's 'freshness' as well as the speed with which we make judgments based on external signals.

II. RELATED WORK

A web based framework model architecture to represent the process of extraction of information from single or multiple sources of data and apply transformation business logic to it and finally load that information to the destination for data analytics is proposed in [1]. They come up with this idea against the traditional method of doing this by hard coding or writing huge modules of code to perform this which requires a lot of time and manual work.

A Model-Driven Development Approach for the development of ETL processes is addressed in "Application of ETL Tools in Business Intelligence" by Nitin Anand [2]. The work aims to illustrate that this methodology proves to be efficient in organizing all the elements of the architecture to carry out the building and automation of all the steps of an ETL process. This work is based on considering the fact that a properly structured ETL plays a crucial role in Business Analytics and this process requires enormous input and work [3].

ETL methodologies, the challenges in the process of testing ETL, and different methods of ETL testing ideas are discussed in [4]. They illustrate how data is extracted from diverse sources, cleaned, customized, reformatted, and integrated software techniques called ETL tools are used to load the data into a data warehouse. One of the difficult jobs in building a warehouse is to develop the ETL process.

The study [5] identifies and assesses the current methods used to implement existing ETL solutions after conducting a thorough review of 97 papers in the literature. It was discovered that the most common method used to develop ETL solutions is conceptual modeling, such as UML, BPMN, and MDA. However, cutting-edge methods like robots, artificial intelligence, and machine learning are either under-utilized or not used at all while creating ETL solutions.

[6] uses Universiti Teknologi Malaysia (UTM) as a case study to get an idea and analyze the relation of data integration with ETL process. In order to improve the management of higher education, this article also highlights the importance of data integration and ETL in the business intelligence framework and briefs out its applications.

III. APPLICATIONS

A. Delivering a single point-of-view

Business analytics most of the time demand a single perspective for a cluster of a variety of data. This indeed requires handling various datasets that requires considerable time and effort and still can yield discrepancies and latency. With the help of ETLs, we integrate different sources of data into one single curated form which makes it ready to use for reporting and analysis.

B. Providing historical context

Aggregation of data enables to produce a long-term view or a broader picture of data collected over a period of time from various platforms. This also allows us to compare older information with recent new data.

C. Improving efficiency and productivity

Building automated ETLs enables the process of extraction, cleaning, integration and aggregation to be done more efficiently and productively ruling out manual intervention. This results in people being able to focus more on creativity and advancements rather than in manually carrying out tedious tasks of ETL.

D. Market Analysis

ETL is one of the crucial steps being employed in business intelligence. It is an information technology procedure that allows data from numerous sources to be gathered in one curated view in order to instigate the discovery of market insights through the programmatic analysis of data.

E. Research

Research methodologies always involve an ample amount of data scraping and data analysis in which ETL again forms a crucial part. Various methodologies to build innovative ETLs to facilitate extraction of data insights are being studied and researched to generate more advancements in the field of finance and technology.

IV. CONCLUSION

In today's world, markets compete at a global level which requires managing massive amounts of data. ETL techniques are used to efficiently perform this, making it easier for financial teams to manage, handle, and examine data to transform it into business intelligence. Most of the time, data is recorded in both structured and unstructured format. Although, it may be adequately evaluated with the help of ETL technologies. All the cutting-edge data management solutions were created with the goal of analyzing large amounts of data. The majority of today's financial ETL technologies are scalable. In other words, all the ETL applications discussed above form a crucial part of business analytics as they form the basis of data warehousing and data analytics that ultimately lead to making more informed decisions in less time.

ACKNOWLEDGMENT

We would like to thank the Department of Information Science and Engineering, R.V. College of Engineering for the constant support in the field of Research and Development. We are indebted to our mentor, Kavitha S N, who helped in the preparation of this thesis, for her hearty support, suggestions and invaluable advice throughout our project work.

REFERENCES

[1] V. Radha krishna, V. Sravan kiran and K. Ravi kiran, "Web based ETL component extended with loading and reporting facilitations a financial application tool," 2010 2nd International Conference on Software Technology and Engineering, 2010, pp. V1-419-V1-423, doi: 10.1109/ICSTE.2010.5608874.

[2] Nitin Anand, "Application of ETL tools in Business Intelligence", International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012.

[3] Fikri, N., Rida, M., Abghour, N. et al. An adaptive and real-time based architecture for financial data integration. J Big Data 6, 97 (2019).

[4] Vyas, Dr Sonali & Vaishnav, Pragya. A comparative study of various ETL processes and their testing techniques in data warehouses. Journal of Statistics and Management Systems. 20, (2017).

[5] Rodzi, Nur & Othman, Mohd & Mi Yusuf, Lizawati. Significance of data integration and ETL in business intelligence framework for higher education. 181-186. 10, (2015).

[6] Alain Berro, Imen Megdiche, and Olivier Teste, Graph-based ETL processes for warehousing statistical open data. In ICEIS 2015 - Proceedings of the 17th International Conference on Enterprise Information Systems, Volume 1, Barcelona, Spain, 27-30 April, 2015, pages 271 {278, 2015.

[7] Vassiliadis, Panos. A Survey of Extract-Transform-Load Technology.. International Journal of Data Warehousing and Mining. 5. 1-27, (2009).

[8] Q. Hanlin, J. Xianzhen and Z. Xianrong, "Research on Extract, Transform and Load(ETL) in Land and Resources Star Schema Data Warehouse," 2012 Fifth International Symposium on Computational Intelligence and Design, 2012, pp. 120-123, doi: 10.1109/ISCID.2012.38.

