Key Benefits of ETL in Financial IT Systems

Sandeep Rachapudi

Sandeep.Rachapudi@gmail.com

Abstract

The Extract, Transform, Load (ETL) framework is a cornerstone of data management. Every financial system relies on ETL for data acquisition, data processing, data analysis, and data retrieval. As organizations increasingly rely on data to drive decision-making and compliance, understanding the benefits of ETL becomes critical. This paper explores the evolution of ETL, its key benefits, and its role in enhancing financial operations.

Keywords: Data warehouse, ETL, Data Staging, Operational Data Source, Financial Systems, Data management, slowly changing dimension

Introduction

Most businesses rely on multiple streams of data from different sources and formats. If there is no centralized source for all this information, then it becomes that much harder to make informed decisions for financial institutions. That's why organizations have turned to ETL to provide the context needed to understand their data and set priorities that are supported by the statistics they value. Data from the ETL process is often stored in a data warehouse. Slowly Changing Dimension (SCD) is used for storing both current and historical data over time in a data warehouse. SCD is implemented as one of the most critical ETL tasks in tracking the history of dimension records.
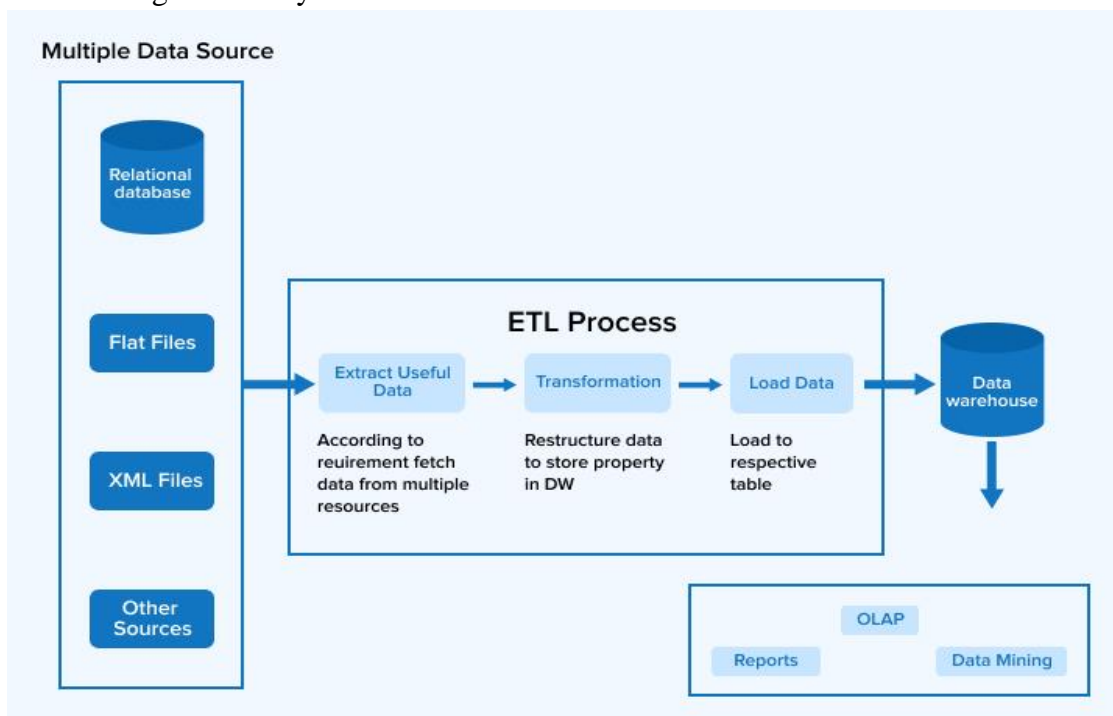

Fig. 1.0. [1]: Representation of the ETL Framework

Problem

The main challenge associated with financial institutions is collecting meaningful data and improving data quality. A centralized data source to perform data analytics is the key for making data-driven decisions.

Solution

Implementation of the Extract, Transform, Load (ETL) framework typically consists of several key components and stages that facilitate the process of moving data from source systems to a target data warehouse or data lake.

Extract: The extract component is integrated to various data source systems such as databases, APIs, flat files, etc. A lot of market data for financial institutions is obtained in the form of data feeds, which are flat files that are in various formats. Data from investment research firms like Morningstar and Bloomberg is often obtained in the form of feed files.

Market data feeds are real-time streams of financial information and data. These streams relate to various financial instruments such as stocks, bonds, commodities, currencies, and more. Through market data feeds, traders and investors gain up-to-the-minute information about
the financial markets, including:

- Price Quotes
- Trade Volumes
- Bid and Ask Prices
- Historical Data
- Other relevant statistics

Financial institutions consume multiple market data feeds on a recurring basis. These are feeds that are obtained at varying frequencies depending on the type of feed. The quarterly feeds contain consolidated revenue results of various firms, whereas the daily feeds have price data.

Data extraction can be done in full extraction mode or incremental extraction mode based on the type of feed and the data provider.

Transform: Once the data is extracted, it is transformed to make it meaningful for consumption by the financial institution. Most of the business logic is applied during transformation. Database stored procedures are often used for applying transformation logic. The various steps involved in transformation are:

- **Data Cleansing:** Remove duplicates, correct errors, and standardize formats to ensure quality.
- **Data Enrichment:** Enhance data with additional information (e.g., lookup tables).
- **Data Aggregation:** Summarize data as needed for reporting (e.g., calculating totals or averages).
- **Data Transformation:** Apply business rules and logic to transform data into the desired format.
- **Deriving:** Generate new values using available values (e.g. Total from price and quantity).
- **Sorting:** Sort data according to some specific value for fast retrieval.
- **Joining:** Integrate data elements of multiple sources and generate default values (lookup & merge) while maintaining data quality.

Load: Once the data transformation is complete, the next stage is loading the data into its destination tables. Based on the ETL Job process, data can be loaded into a data warehouse for performing data analytics, reporting [3], and data mining. The load phase marks the end of ETL process, but this is the phase where meaningful data is available in a centralized location for making data driven decisions.

ETL effectively improves your data quality, which can be utilized by BI tools for different purposes, such as understanding customer trends, optimizing operational efficiency, increasing user satisfaction, enhancing business performance, and much more.

Data load can be done to a data lake [4], which can store data in varying formats. The data load can be done either in an incremental fashion or as a full override of existing data. Maintenance of historical data along with current data is done by implementing SCD. Slowly Changing Dimensions (SCD) are a critical concept in data warehousing and business intelligence. They refer to the methods used to manage and track changes in dimension data over time.
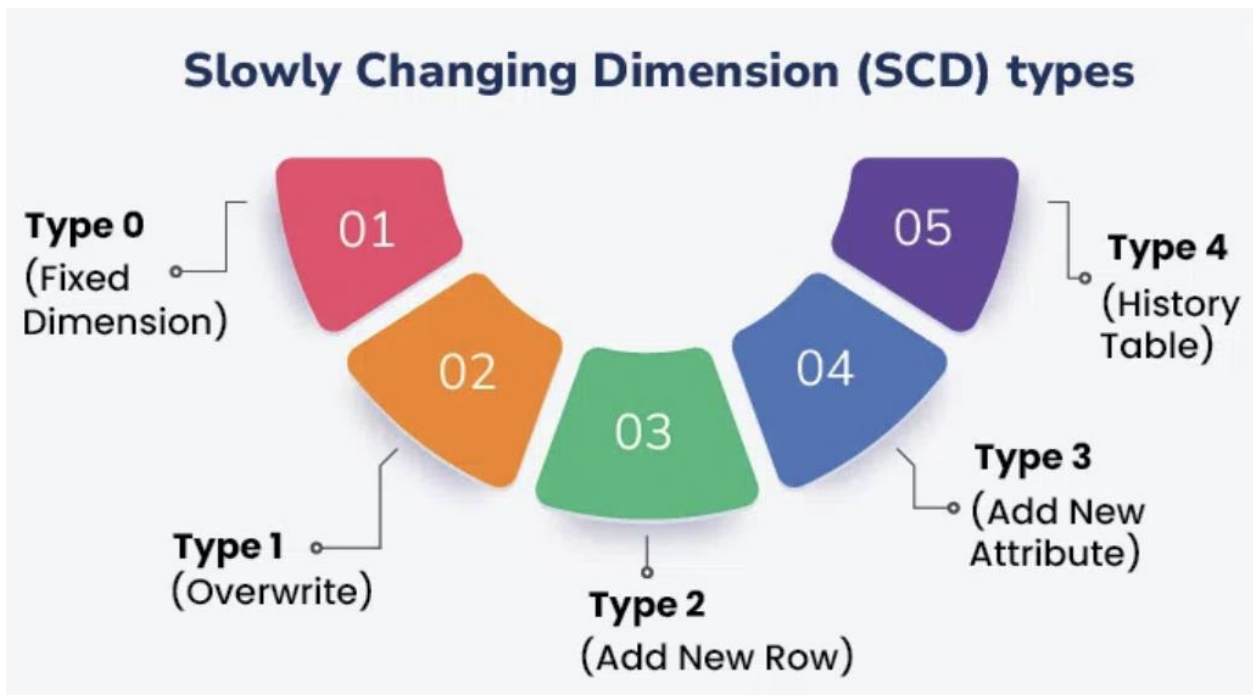

Fig. 2.0 [2]: Types of SCD

Type 0: Type 0 refers to dimensions that never change. You can think of these as mapping tables in your data warehouse that will always remain the same, such as states, zip codes, and county codes. In addition to mapping tables, other pieces of data like social security number and date of birth are considered type 0 dimensions.

Type 1: Type 1 refers to data that is overwritten by new data without keeping a historical record of that old piece of data. With this type, there is no way to keep track of changes over time. Many companies use this type of dimension accidentally, not realizing that they can never get the old values back. When implementing this dimension, make sure you do not need to track the trends in that data column over time.

Type 2: Type 2 dimensions are always created as a new record. If a detail in the data changes, a new row will be added to the table with a new primary key. However, the natural key would remain the same in order to map a record change to one another. Type 2 dimensions are the most common approach to tracking historical records.

Type 3: Type 3 dimensions track changes in a row by adding a new column. Instead of adding a new row with a new primary key like with type 2 dimensions, the primary key remains the same and an additional column is appended. This is good when the primary key needs to remain unique and only have one record for each natural key.

Type 4: Type 4 dimensions exist as records in two different tables—a current record table and a historical record table. All of the records that are active in a given moment will be in one table, and then all of the records considered historical will exist in a separate history table. This is a great way of keeping track of records that have many changes over time.

Conclusion

ETL is critical in financial IT systems for ensuring data integrity, regulatory compliance, and effective decision-making, ultimately supporting the organization's financial health and strategic objectives.

As ETL technology evolved, both data types and data sources increased exponentially. Cloud technology emerged to create vast databases. ETL tools have also become more sophisticated and can work with modern data bases. They can convert data from legacy data formats to modern data formats.

Effective usage of ETL can provide a competitive edge over competitors when you uncover new relationships in data, often pinpointing features that can stand out from the rest of the market. Gaining traction is easy when you know your product's popular features, along with figuring out where you can bring in improvements to your product. Organizations need to adopt or enhance their ETL strategies to remain competitive in a data-driven landscape.

References

1. "ETL Process". https://www.tatvasoft.com/blog/etl-process-extract-transform-load/ (Accessed Nov 5th 2023)
2. "slowly changing dimensions". https://www.geeksforgeeks.org/slowly-changing-dimensions/ (Accessed Nov 10th 2023)
3. H Mokhamad, Y Erwin, R Jack, Awan S Awan, I Benie. "Extract transform load process in banking reporting system". https://doi.org/10.1016/j.mex.2021.101260 (Accessed Oct 25th 2023)
4. "What is ETL?". https://aws.amazon.com/what-is-etl/ (Accessed Oct 15th 2023)