

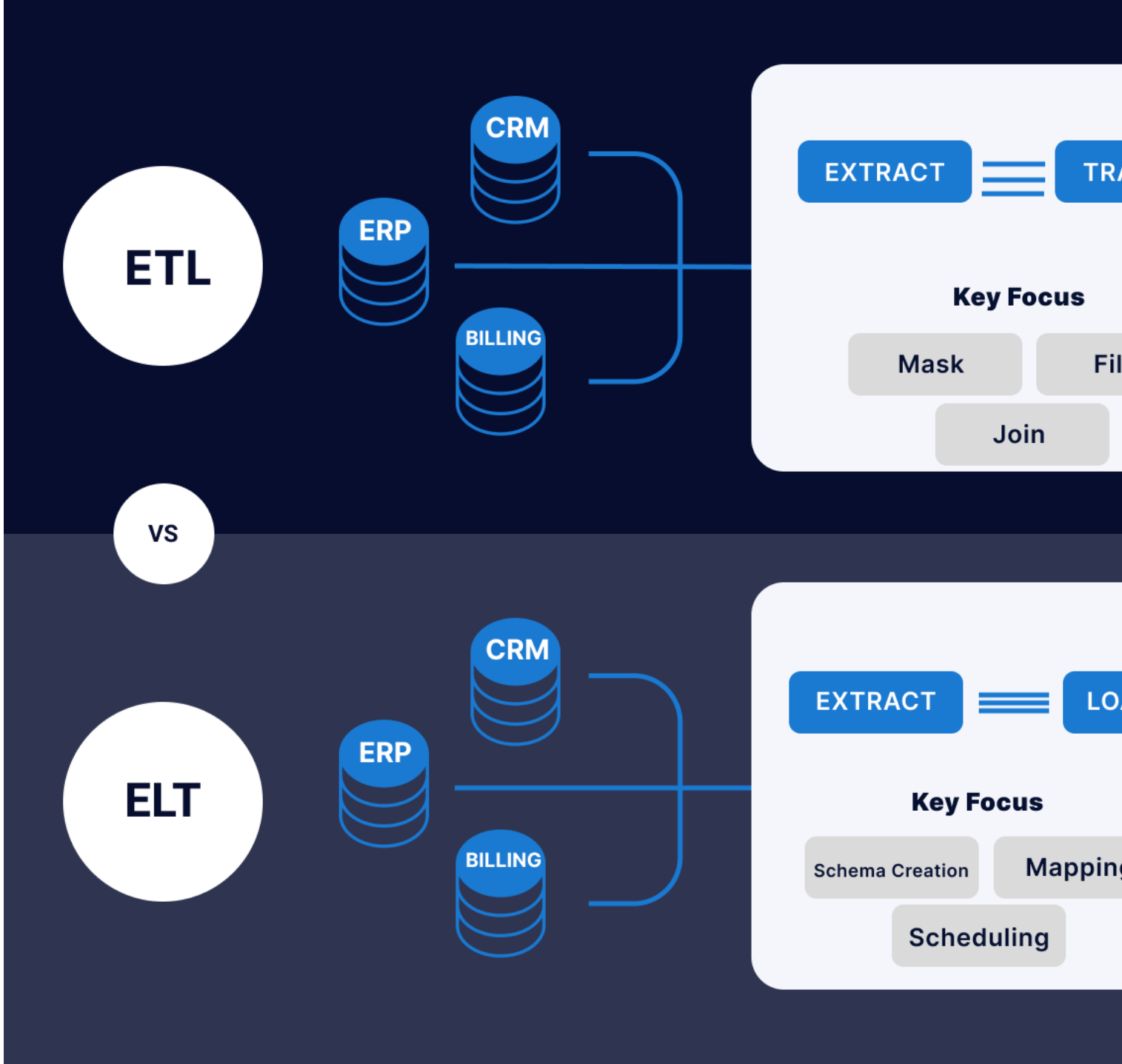
Unlock the True Potential of Your Data with ETL and ELT Pipeline

[AWS](#)[BEGINNER](#)[CLOUD COMPUTING](#)[DATA ANALYSIS](#)[DATA ENGINEERING](#)[DATA WAREHOUSE](#)

Introduction

This article will explain the difference between ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) when data transformation occurs. In ETL, data is extracted from multiple locations to meet the requirements of the target data file and then placed into the file. The transformation process occurs outside the target, a separate processing tool or system.

In ELT, data is extracted from multiple sources and placed into the target. The transformation process takes the data file using the processing power of the file. The transformation is performed after the data has been placed into the file.



Source: Striim

In summary, the main difference between ETL and ELT is the order of operations for data transformation and the location where the transformation occurs. ETL performs the transformation before loading data into the target file, while the ELT performs the transformation after loading data into the file.

Learning Objectives:

Here is some potential the article on ETL vs. ELT pipelines:

1. Understanding the difference between ETL and ELT pipelines, including when and where data transformations take place in each approach.
2. Comparing the pros and cons of ETL and ELT pipelines, including speed and data quality.
3. Becoming familiar with real-life examples of companies that have implemented ETL or ELT pipelines and understanding the factors that influenced their choice of approach.
4. Understanding the future trends and advancements in data integration, including the growing importance of cloud computing and real-time data processing.

Gaining a comprehensive overview of the data integration process, including the key steps in loading data. Evaluating the suitability of ETL and ELT pipelines for different types of data integration scenarios, factors

that should be taken choosing between these approaches.

This article was published as a part of the [Data Science Blogathon](#).

Table of Contents

1. Why is ELT outperforming ETL?
2. How to Determine the Best Approach – ETL or ELT?
3. Example of ETL and ELT Pipeline
4. Security and Data Governance
5. Future Trends and Advancements in the field of Data Integration
6. Real-life examples of companies using ETL or ELT pipelines
7. Conclusion

Why is ELT Outperforming ETL?

Here are some of the pros and cons of ETL and ELT pipelines:

Advantages of ETL pipelines:

1. ETL pipelines are widely adopted, providing a proven and well-understood solution for data integration.
2. ETL pipelines can handle large amounts of data, making batch processing and warehousing applications.
3. ETL pipelines provide data, making it easier to manage and access data.
4. ETL pipelines can clean and validate, improving data quality and making it easier for analysis and decision-making.

Disadvantages of ETL pipelines:

1. ETL pipelines can be slow and resource-intensive, making them less suitable for real-time data processing.
2. ETL pipelines can be complex and difficult to manage, requiring specialized technical skills.
3. ETL pipelines can be inflexible, making it difficult to respond to changing business requirements.
4. ETL pipelines can introduce latency in the data integration process, making it difficult to get real-time insights.

Advantages of ELT pipelines:

1. ELT pipelines are designed for real-time data processing, making applications insights and actions.
2. ELT pipelines can leverage the power of cloud computing, making it possible to of data in real-time.
3. ELT pipelines can integrate data from multiple sources and IoT devices, making it possible to gain a complete view of data.
4. ELT pipelines can be more flexible, changing business requirements.

Disadvantages of ELT pipelines:

1. ELT pipelines are a relatively new solution.
2. ELT pipelines can be more complex to set up and manage, requiring specialized technical skills.

3. ELT pipelines may provide a different level of data validation and pipelines, potentially leading to lower data quality.
4. ELT pipelines may not be suitable for batch processing, and data warehousing is optimized for real-time data processing.

Source: Forsys

How to Determine the Best Approach – ETL or ELT?

When choosing between ETL and ELT, it is specific requirements. For example, if the need to integrate data from multiple sources and perform complex transformations, ETL may be the better choice. ETL transformation capabilities and can handle complex data integration scenarios.

Some pointers to differentiate the two:

1. ETL focuses on data transformation before loading it into the target system, while ELT focuses on loading the data and performing the target system.
2. ETL can be more suitable for data transformation and integration from multiple sources.
3. ELT can be more efficient in terms of performance since the transformation step is a target system designed to handle large amounts of data.
4. ELT can be more suitable for data integration and must take advantage of the target system's resources and parallel processing capabilities.
5. ETL can be more scalable since the data transformation step is done using a separate ETL tool, which we can scale independently of the target system.

Example of ETL and ELT Pipeline

Here is an example of an ETL process using Python and the popular ETL tool Apache NiFi. The following code extracts data from a CSV file, performs, and loads the data into a MySQL database.

```
from nifi import ProcessGroup, ExtractText, ReplaceText, PutSQL # Extract data from a CSV file extract =
ExtractText(path="path/to/input.csv") replace = ReplaceText(search="old_value", replace="new_value") # Load
the data into a MySQL database load = PutSQL(connection_url="jdbc:mysql://host:port/database",
username="username", password="password", sql_select_query="INSERT INTO table (column1, column2) VALUES (?,
?)") # Create a ProcessGroup to link the processors together pg = ProcessGroup(processors=[extract, replace,
load]) # Run the ETL process pg.run()
```

Below is an example of an ELT process using Python and the SQLAlchemy library. The following code extracts data from a CSV file, load it into a PostgreSQL database and performs using SQL queries.

```
from sqlalchemy import create_engine import pandas # Connect to the PostgreSQL database engine =
create_engine("postgresql://username:password@host:port/database") # Load the data from a CSV file into a
Pandas DataFrame df = pd.read_csv("path/to/input.csv") # Load the data into the PostgreSQL database
df.to_sql("table_name", engine, if_exists="replace") # SQL queries with engine.connect() as con:
con.execute("UPDATE table_name SET column1 = column1 + 1") con.execute("DELETE FROM table_name WHERE column2
= 'value'")
```

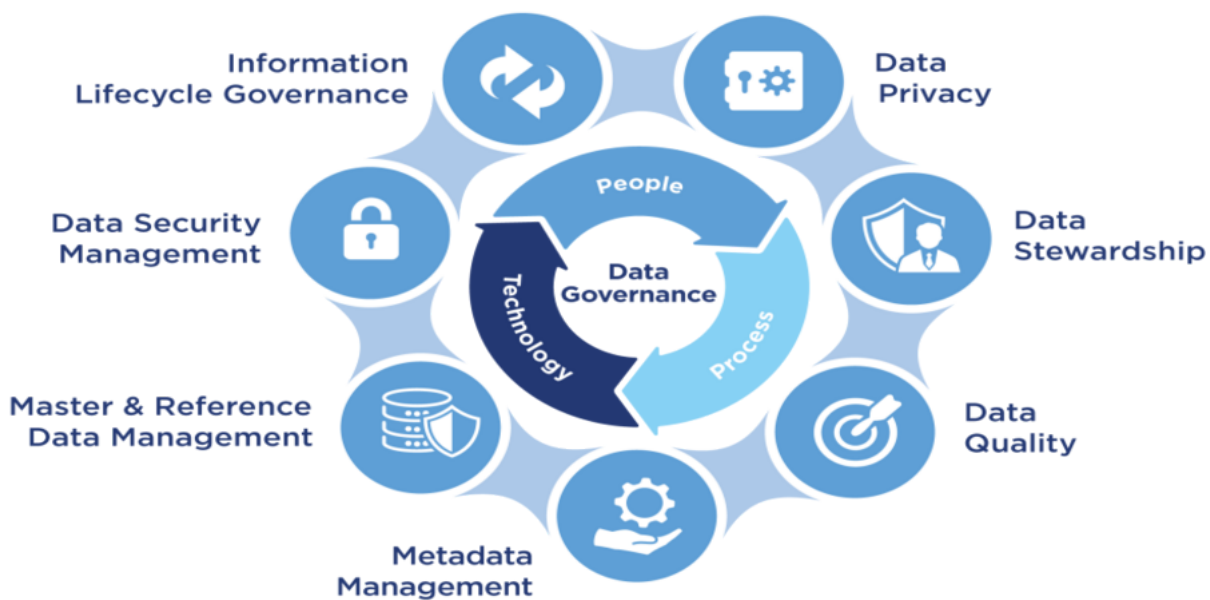
It's important to note that these examples are just ETL and ELT processes, and the process is much more complex in the real-world scenario. It may involve different libraries and tools.

Security and Data Governance

Security and data governance are critical factors when choosing between ETL and ELT. In ETL, data governance by the ETL tool can provide more granular control over the data. The ETL tool can be configured to implement security measures, encryption, data masking, and access controls. This can provide an added layer of security for sensitive data.

On the other hand, ELT relies on the target system for data governance. At the same time, the target system may provide robust security measures, but it has a different level of control than an ETL tool. Also, ELT loading data into the target system can create a security risk if the data contains sensitive information.

Data governance quality, lineage, cataloging, and compliance regulations. ETL can provide better data and allows better control over the data quality and lineage, making it easier to track data changes and better understand the data. On the other hand, ELT can be more challenging in data. It takes more work to track data changes and clearly understand the data.



Source: Analytics Vidhya

Future Trends and Advancements in the Field of Data Integration

Here are some future trends and advancements in the field of data integration:

1. **Automation and AI-driven data integration:** Automation and artificial intelligence will play a big role in the future of data integration, making the process faster, more efficient, and less prone to errors. [AI algorithms](#) can automatically identify patterns and anomalies in real-time data-driven decisions.
2. **Integration of Multi-cloud and Edge Computing:** With the rise of cloud increasingly relying on multiple cloud providers. Data integration solutions must seamlessly integrate data from multiple clouds and [IoT devices](#).
3. **Focus on Data Governance:** Data governance will become increasingly important. Data integration solutions must provide robust data governance capabilities, including privacy, security, and lineage.
4. **Greater Emphasis on Real-time Data Processing:** With the increasing demand for real-time insights and actions, data integration solutions will need to process data in real time, including IoT devices.
5. **Expansion of Low-code and No-code Platforms:** Low-code and no-code platforms will continue to grow in popularity, making it easier to create and manage data integration pipelines without requiring specialized technical skills.
6. **Increased Use of Blockchain Technology:** [Blockchain](#) technology can revolutionize data integration, sharing, and collaboration.
7. **The Emergence of Data Virtualization:** Data virtualization will become increasingly important, accessing and integrating data from multiple sources without moving the data physically.

These trends and advancements will play a crucial role in shaping the future of [data integration](#), making better data-driven decisions, improve their overall efficiency in the business landscape.

Source: www.remedi.com

Real-life Examples of ETL or ELT Pipelines

Here are some real-life examples of ETL or ELT pipelines:

1. **Walmart:** [Walmart](#) uses an ETL pipeline to extract data, including point-of-sale systems and suppliers, clean data, and load it into a central data warehouse. This allows the company to analyze sales data and decisions.
2. **Netflix:** [Netflix](#) uses an ELT pipeline to extract raw data, load it into a cloud-based data lake, and then use cloud-based analytics tools to analyze the data. Process massive amounts of data in real-time, making it possible to deliver personalized recommendations to users.
3. **Amazon:** [Amazon](#) uses ETL and ELT pipelines, depending on the specific. For example, ETL pipelines are used to process large amounts of data from multiple sources and load it into a data warehouse for analysis. On the other hand, ELT pipelines are used for real-time data processing and customer behavior to improve the shopping experience.
4. **Google:** [Google](#) uses ELT pipelines to extract data in real time, including data from search queries, advertisements, and other sources. This allows the company to gain insights into user behavior and make real-time decisions to improve the user experience.
5. **Spotify:** [Spotify](#) uses an ELT pipeline to extract, load, and of data generated by users, including listening habits and social location data. This allows the company to personalize the user experience and make data-driven decisions to improve the overall user experience.

Conclusion

In conclusion, ETL and ELT are common data integration approaches used to move from one system to another. ETL is the traditional approach, where data is the target system. ELT is a newer approach, where data is 1st loaded into the target system and then changed. ELT has benefits over ETL, including performance, more efficient use of resources, more real-time data integration, and governance.

ETL and ELT are not mutually exclusive and can be combined depending on ETL loading and ELT for real-time data integration. The choice between ETL and ELT should be based on specific requirements.

Key takeaways from this article:

1. ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) are integration approaches used to move from one system to another.
2. We discussed the benefits of ETL and ELT pipelines. As we have mentioned that ELT is more powerful

and useful than ETL.

3. We discussed how to choose between ETL and ELT. We must discuss and decide which pipeline best works.

4. We coded some examples of ETL and ELT pipelines using Python, in which we read data from a CSV file and put it into a [MySQL database](#) after performing. After that, we concluded the article.

The media shown in this article is not owned by Analytics Vidhya and is used at the Author's discretion.

Article Url - <https://www.analyticsvidhya.com/blog/2023/02/unlock-the-true-potential-of-your-data-with-etl-and-elt-pipeline/>



[Aditya Garg](#)