



ETL Process

ETL Process

THIS PRESENTATION WILL EXPLORE THE ESSENTIAL COMPONENTS AND STEPS INVOLVED IN THE ETL PROCESS.



What is ETL?

What is ETL?

▶ **Extract, Transform and Load:-**

- ▶ In simple terms, it's a Data process where we take data from one or multiple sources, apply some changes to it and store it in a new place.
- ▶ Let's understand this with a example.

What is ETL?

- ▶ Let's say we have some **Customer data** in a CRM system and some **Sales Data** in Sales system.
- ▶ With the use of ETL, we can generate a analysis of which customers are purchasing the product regularly.
- ▶ This is to show that, **ELT** is **Great for reporting and Analysis.**

What is ELT?

- ▶ Let us assume that **ELT is not existing**. To do this Analysis we would have to go to the CRM and Sales systems, take all required data, do the combinations and filtering and cleaning and then do the analysis.
- ▶ **That's a lot of Work!!**
- ▶ Now doing this once seems okay. But doing all of these processes on a daily basis..... **Will lead to your resignation :(**

What is ETL?

- ▶ With ETL, what companies do is they create a **process** which periodically **Extract** the data from their business systems, apply some business rules to **Transform** the data, and then **Load** it onto a place that contains such transformed data (like a Data warehouse).
- ▶ This makes the **tracking** and **analysis** of data a lot **more easier** and **efficient!**

ETL Operations

ETL Operations

▶ Data Cleansing:-

- ▶ It involves **fixing or removing** incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
- ▶ This process makes sure that only **best-quality data** is **loaded** into the data warehouse, which is **necessary** for **accurate business analytics** and **decision making**.

ETL Operations

► Keys and IDs:-

- The **transformation** phase of ETL makes the **use of keys** and **IDs** to ensure **data integrity** and **consistency**.
- When using data from multiple sources, **unique identifiers (IDs)** are used to **match records** across different datasets. **Keys** like **primary keys** and **foreign keys** are helpful for **linking related data** and ensuring that **duplicate records** are removed.

ETL Operations

- ▶ **Combining Data:** This process involves **merging data** from various sources **into a single dataset**. It is crucial for **joining information** from **different systems**, such as **transactional databases**.
- ▶ **Merging Data:** **Formatting** and **aligning** the **combined data** by resolving conflicts, removing duplicates, and ensuring consistency for **seamless integration** this process is known as **Merging**.

ETL Tools

ETL Tools

- ▶ SQL:- This is the **simplest platform** to perform ETL and is **recommended to students** for learning ETL process.
- ▶ Many companies **prefer to create** their own **ETL platform** for **efficient processing**.
 - ▶ Like IBM uses **IBM Data Stage**, Oracle uses **Data Integrator** created by them, Amazon uses **AWS**, Azure has **Azure Data factory** product and Google uses **Google Cloud** to perform ETL process.

ETL Tools

- ▶ Now we will see a simple example of ETL in SQL.
 - ▶ `SELECT * FROM sales_raw;`



sale_id	product_name	sales_amount	cost_price	sale_timestamp
1	Dark Choco	500	300	2025-02-03 14:30:00
1	Dark Choco	500	300	2025-02-03 14:30:00
2	Milk Choco	600	400	2025-02-03 15:00:00

Output (transformed_sales) Table


sale_id	product_name	sales_amount	cost_price	profit	sale_date
1	Dark Choco	500	300	200	2025-02-03
2	Milk Choco	600	400	200	2025-02-03

```
CREATE TEMP TABLE
transformed_sales AS
SELECT
DISTINCT sale_id,
product_name,
sales_amount,
cost_price,
(sales_amount cost_price) AS
profit,
DATE(sale_timestamp) AS
sale_date
FROM
sales_raw;
```


Output (sales_data) Table

sale_id	product_name	sales_amount	cost_price	profit	sale_date
1	Dark Choco	500	300	200	2025-02-03
2	Milk Choco	600	400	200	2025-02-03

```
CREATE TABLE IF NOT EXISTS sales_data (  
  sale_id INTEGER PRIMARY KEY,  
  product_name TEXT,  
  sales_amount NUMERIC,  
  cost_price NUMERIC, profit  
  NUMERIC, sale_date  
  DATE);  
SELECT *  
INSERT INTO sales_data  
FROM transformed_sales;
```



Made By:-
Nrusinha Doiphode
Harsh Deshpande
Prasad Sahotre
Sonal Gorde