Task: 04– by Bytewise Team

# **What is ETL in detail?**

What is ETL? The way I like to break it down is to first define what this acronym means and then highlight its importance and its benefits.

* E in ETL stands for Extract. When you do ETL, you have data from a variety of different sources and the goal is to have them all together. When do you have it? You move to the next step.
* T stands for Transform. When data is altogether. We do processes of decoupling, de -normalizing and combining, and many more. In short, you transform unorganized data into organized and meaningful data and develop different relationships.
* L stands for Load in ETL. Finally when you have transformed and gained a new view of your data. You want to load that new curated data into data sources like data warehouses.

Now you know what ETL means, The next obvious question is why it is so important.

* First, let’s talk about the context ( suppose you have a dataset related to your own application ) and consolidation. Having all the at one place give you a huge advantage to perform analysis, reporting, etc. and continuously update it and make more accurate insights.

* Secondly, ETL increases productivity. what ETL accomplishes, think that manually. You need more manpower and time to do this process manually. ETL is a repetitive process you feed it data and it gives you an analysis of your data.

* The last one is accuracy. When you did the whole process. You will probably build nice fancy charts/reports using BI tools. All the data from different sources is already curative.

# **What is ELT in detail?**

ELT stands for Extract, Load, and Transform, and it refers to a data integration process in which data is extracted from source systems, loaded into a target system (such as a data warehouse), and then transformed into a format suitable for analysis.

Here I have broken down ELT's importance into steps:

* In ELT, the data is first loaded into the target system and then transformed in place using the target system's processing power. This means that the transformation step can be done more quickly and efficiently, as it can take advantage of the target system's processing power and storage capabilities.

* ELT can also help reduce the complexity and cost of data integration, as it eliminates the need for a separate transformation platform or tool, and can leverage existing processing power and storage capacity in the target system.

* ELT is particularly useful for big data environments, where the volume of data is too large to be processed outside of the target system. Additionally, ELT can help reduce data latency, as data can be processed and transformed more quickly and efficiently.

# **What is 3 Tier Architecture in Data Engineering?**

The 3-tier architecture is a widely used approach in data engineering to design scalable and efficient data processing systems. Here 3 Tiers mean three layers that are: the Client layer, the Business layer, and the data layer. Do note here, people also use different names, when recalling these layers. So, do not get confused. Let’s understand them.

* **Client Layer:** This layer is a place where the user types or interacts. In simple words, this is an interface for our customers/ users.

**E.g** you are using a booking application to book a specific hotel room in advance. Here you are using a client machine/ app interface. Which will pass your query to the next layer and help you connect with the database. This will help to decide which rooms are available and you can book and which are not.

* **Business Layer:** It’s the 2nd layer, this layer supports the interface of an application. All the queries by users are answered by this layer. It’s also known as Application Server.

**E.g** If you have a query like. which rooms are available tomorrow so you can book at a specific hotel. This question' answer will be provided by this layer as it supports an interface to manage users by resolving queries. Every action/input user takes on the interface generates a query which will pass to the 2nd layer, which answers it.

This layer helps this architecture make it more scalable, secure, and efficient.

* **Data Layer:** It’s the 3rd layer. Which represents the data sources. Let’s understand it with an example.

**E.g**: When you input all personal info for booking a hotel including your bank account details using the application Interface. This generates the query, which is passed down to the Business layer. Here 2nd layer checks the query, and in case of any error resolves it by instructing the user. Then transform the query into low-level/machine language. Which is then passed down to the 3rd layer and stored in a database or data source.

When data is stored, the Business layer then takes particular information/data from the Data layer related to the query and displays it to the interface for the user.

Business layer work as an intermediate between two layers.

# **ETL TOOLS (any 3)**

Azure Data Factory

Apache Airflow

Informatica

AWS Glue