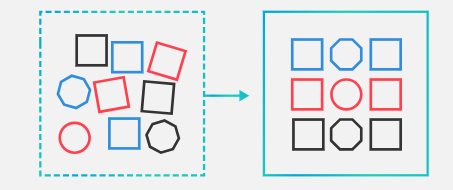
**APEXAIQ Assignments | 12/02/2025**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Data Transformation**

* It is the process of **changing** the **format**, **structure** and **values** of data such as Non-Normal data to Normal data.
* It converting data, cleaning data into a usable format for **analysis.**
* Data transformation is used when **data needs** to be converted to **match** that of the destination **(** target **)** system.



* **The process of data transformation can be:**
* **Constructive:**- where data is added, copied
* **Destructive:**- where records and fields are deleted
* **Aesthetic:-** where certain values are standardized
* **Structural:**- which includes **columns** being renamed, moved and combined.
* **How is data transformation used?**
* Data transformation works on the **extracting data from a source**.

Converting it into a **usable format** and then **delivering** the converted data to the **destination** system.

* Organizations perform data transformation to ensure the compatibility of data with other types while combining it with other information or migrating it into a dataset.
* To ensure the usability of the **extracted data** it must be **transformed** into the desired format by taking it **through a number of steps.**
* **Data Transformation Process:-**

**1. Discovery (first step)**

* The **first step** is to **identify** and **understand** data in its original **Source** **format** with the help of [data profiling](https://www.tibco.com/glossary/what-is-data-profiling) tools.
* **Data profiling** is a tool where **data** is examined and **analysed**, and summary statistics are generated is called data profiling.

**2. Mapping**

* The **transformation** is **planned** during the data mapping phase.
* In mapping it **determining** the **current structure** then mapping the **data** to **understand** at a **basic level**.

**3. Code Generation**

* The **code**, which is **required** to **run** the **transformation** **process**, is **created** **in this step** using a data transformation platform or **tool**.

**4. Execution (last step)**

* The **data** is finally **converted** into the **selected** **format** with the help of the **code**. The data is extracted from the source(s), which can vary from structured to streaming, telemetry to log files.

* The **transformed data** is then **sent** to the **destination** system which could be a **dataset** means **(** a structured collection of data organized and stored together for analysis or processing**. )**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* **Why do Organisation need Data transformation?**
* Dashboards and Reports which enable organizations to **achieve specific goals** and Runs on Data.
* Data allows Organization to visualize relationship.
* With help of Data Transformation , Organizations can collects **Facts** from Applications which they **track and predict.**

**Such Example:-**

We **collect raw data** on how many cars were sold last year. This data might be messy or scattered across different sources. We clean it by removing errors, organize it into a structured format, and aggregate it to get **useful insights** (e.g., **total sales per month** or region). Once the data is transformed into a clear and usable format, businesses can use it to analyze trends and make better decisions

**We make data to transform**



**Source**



**Organization**

**Destination**

**Tracking how many car Sold**

->>>>>>>>>>>



TATA

Selling a

**Predict** sense that how much **Profit** or

**Loss**

* **Reasons why Organizations should transform their data?**
* **Transformation** makes **disparate** means **(** formats of data in databases, or file formats **)** sets of data **successfully** with each other, which **makes** it **easier** to aggregate data for a thorough analysis.
* **Migration of data** is easier since the **source** format can be transformed into the **target** format.
* **Data transformation** helps in **consolidating** data means **(** bringing together data from various sources and assembling it within a single location **),** structured and unstructured.
* The process of transformation also allows for enrichment which enhances the quality of data.
  + **Benefits of Data Transformation**
* **Improved Data Quality** :- Transforming data ensures that it is properly formatted, validated and organized,
* Data transformation helps in organizing data and making it meaningful, which improves the overall quality of the data. This compatibility between systems provides valuable support for functions like analytics and machine learning.
* Data transformation not only helps organizations derive the maximum value from their data, but they also ensure that data can be managed in easier ways without feeling overwhelmed by the sheer amount of it all.
* Data holds the potential to directly affect an organization’s efficiencies and its bottom line. It plays a **crucial role** in understanding customer behaviour, internal processes, and industry trends.
* While every organization has the ability to collect an immense amount of data, the challenge is to ensure that this is usable.
* Data transformation processes **empower organizations** means **(** supporting accurate decision-making and organization growth **)** to reap the benefits offered by the data.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* **More accurate insights and predictions**
* The transformation process generates data models which are then converted to metrics, dashboards and reports which enable organizations to **achieve specific goals**.
* The metrics and key performance indicators help businesses quantify their efforts and analyze their progress.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* **Challenges of data transformation**
* Data transformation is considered essential by organizations due to all the benefits it has to offer.

**There are also a few challenges that come alongside :-**

1. **High cost of implementation**

The process of data transformation is an expensive one.

1. **Resource intensive**

Data transformation also needs expertise from data scientists, which can be expensive and divert attention from other task.

1. **Errors and inconsistency**

Without proper expertise, many issues can crop up **during** **transformation** which are likely to hamper the end results. Whether it is a poor transformation that results in flawed data, or a migration that fails and corrupts data, there are **risks**.