

# **ETL BASICS – FAQ'S**

## **1. What are the advantages of ETL?**

The process of ETL is required so that data from different heterogeneous sources can be combined and brought into one common source. The advantage of having the process of ETL is that, as data from different sources can be brought together, highly complex and user friendly reports can be generated for easy decision making.

## **2. What are the steps involved in ETL process?**

The steps involved are defining the source; target, creating the mapping, creating the session, and creating the workflow.

## **3. Give the methods of data extraction.**

The extraction process can be done either by hand coded method or by using tools. Tool based extraction have a well-defined approach with a better documentation and it also makes the extraction process easier by a simple click, drag and drop features that are more user-friendly to the programmers.

## **4. What is Data Staging?**

Data staging is used in cleansing, transforming, and integrating the data.

## **5. What are the importance of Metadata?**

- Metadata establishes the context of warehouse data.
- Metadata facilitates the analysis process.
- Metadata are a form of audit trail for data transformation.
- Metadata improve or maintain data quality.

## **6. What is ETL Testing Considerations?**

**UNIT Testing** – Ensures that each component within the system successfully performs the individually responsibility when executed individually.

**INTEGRATION Testing** –Ensures seamless run of the entire process within an application or specific stage with an eye on the details of each of the steps /modules, capturing the responses as the data moves across the system.

## **7. What is Exception handling?**

Exception handling is the process of handling these exceptions without corrupting any existing committed data and terminating the ETL process gracefully. During ETL process exceptions occur either due to incorrect data or due to infrastructure issues.

## **8. What is ETL Job Scheduling?**

ETL Job Scheduling is an operational process which is required to determine the sequence and time of execution of the various data flows (Jobs/Mapping).

## **9. What is ETL Job Monitoring?**

ETL Monitoring takes many aspects of the process into consideration. Resources outside the scope of the ETL system such as hardware and infrastructure administration and usage as well as source and target environments, play crucial parts in the overall efficiency of the ETL system.

## **10. What are the criteria for tool identification?**

The Source System Platform and Database, Date Size, Built in Functionalities and Cost of the tool.