# DWH ETL testing

Data Warehouse Testing

Data Warehouse Testing is a testing method in which the data inside a data warehouse is tested for integrity, reliability, accuracy and consistency in order to comply with the company's data framework. The main purpose of data warehouse testing is to ensure that the integrated data inside the data warehouse is reliable enough for a company to make decisions on.
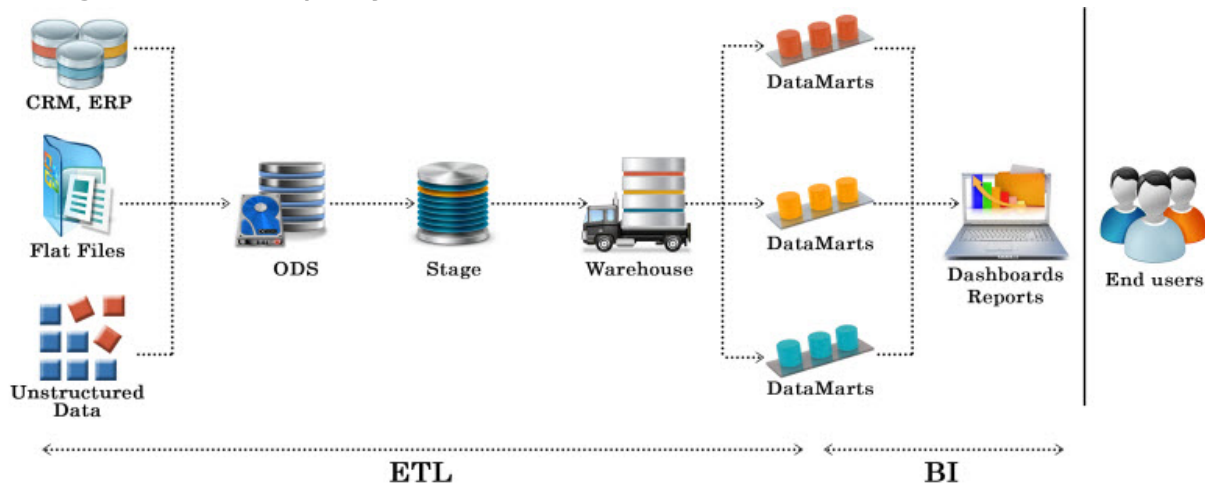
What is ETL?

ETL stands for Extract-Transform-Load and it is a process of how data is loaded from the source system to the data warehouse. Data is extracted from an OLTP database, transformed to match the data warehouse schema and loaded into the data warehouse database. Many data warehouses also incorporate data from non-OLTP systems such as text files, legacy systems and spreadsheets.
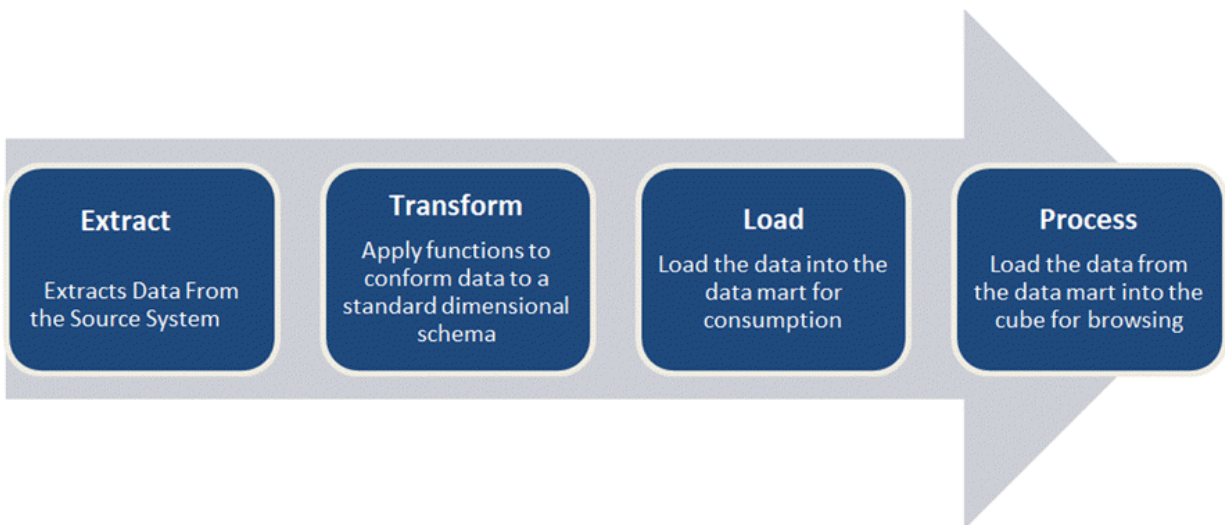
Let see how it works

For example, there is a retail store which has different departments like sales, marketing, logistics etc. Each of them is handling the customer information independently, and the way they store that data is quite different. The sales department have stored it by customer's name, while marketing department by customer id.

Now if they want to check the history of the customer and want to know what the different products he/she bought owing to different marketing campaigns; it would be very tedious.

The solution is to use a [datawarehouse](#) to store information from different sources in a uniform structure using ETL. ETL can transform

dissimilar data sets into an unified structure.Later use BI tools to derive meaningful insights and reports from this data.

The following diagram in this ETL testing tutorial gives you the ROAD MAP of the ETL Testing process flow and various ETL testing concepts:

Extract	Transform	Load	Process
Extracts Data From the Source System	Apply functions to conform data to a standard dimensional schema	Load the data into the data mart for consumption	Load the data from the data mart into the cube for browsing

1. **Extract**
   - Extract relevant data
2. **Transform**
   - Transform data to DW (Data Warehouse) format
   - Build keys - A key is one or more data attributes that uniquely identify an entity. Various types of keys are primary key, alternate key, foreign key, composite key, surrogate key. The datawarehouse owns these keys and never allows any other entity to assign them.
   - Cleansing of data :After the data is extracted, it will move into the next phase, of cleaning and conforming of data. Cleaning does the omission in the data as well as identifying and fixing the errors. Conforming means resolving the conflicts between those data's that is incompatible, so that they can be used in an enterprise data warehouse. In addition to these, this system creates meta-data that is used to diagnose source system problems and improves data quality.
3. **Load**
   - Load data into DW (Data Warehouse)

- Build aggregates - Creating an aggregate is summarizing and storing data which is available in [fact table](#) in order to improve the performance of end-user queries.

What is ETL Testing?

ETL testing is done to ensure that the data that has been loaded from a source to the destination after business transformation is accurate. It also involves the verification of data at various middle stages that are being used between source and destination. ETL stands for Extract-Transform-Load.
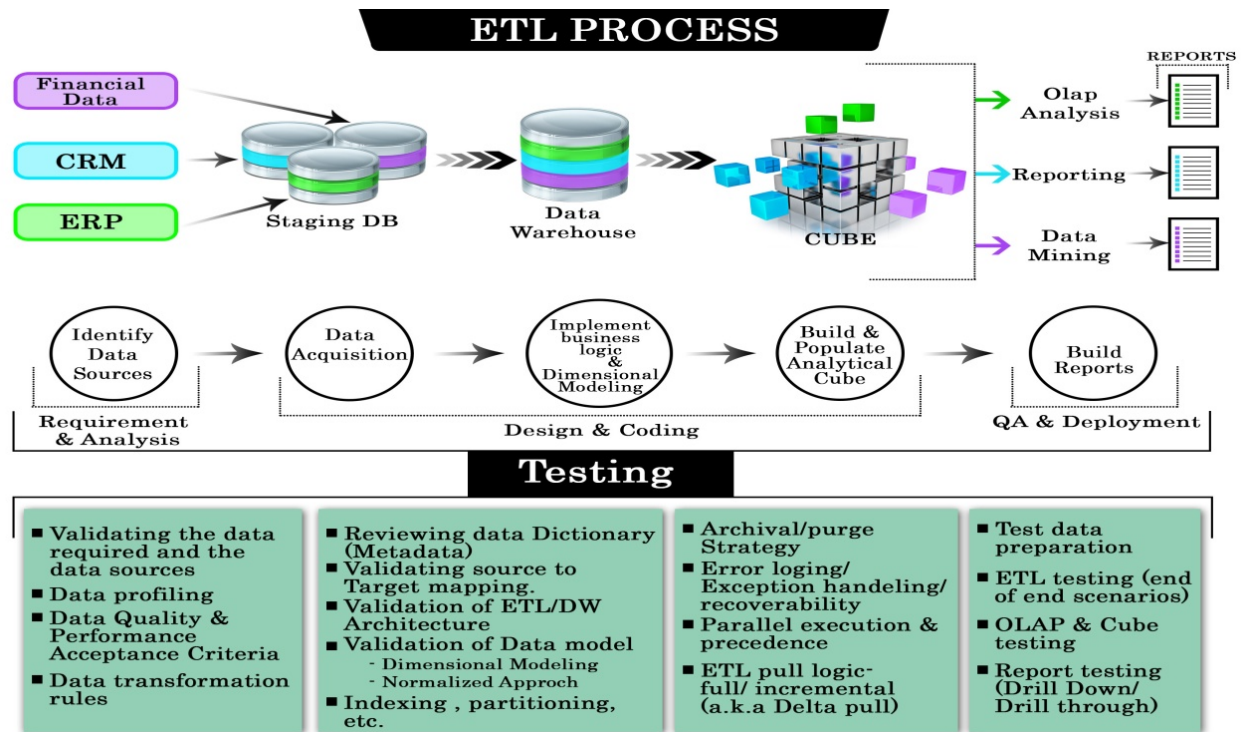
ETL Testing Process

Similar to other Testing Process, ETL also go through different phases. The different phases of ETL testing process is as follows
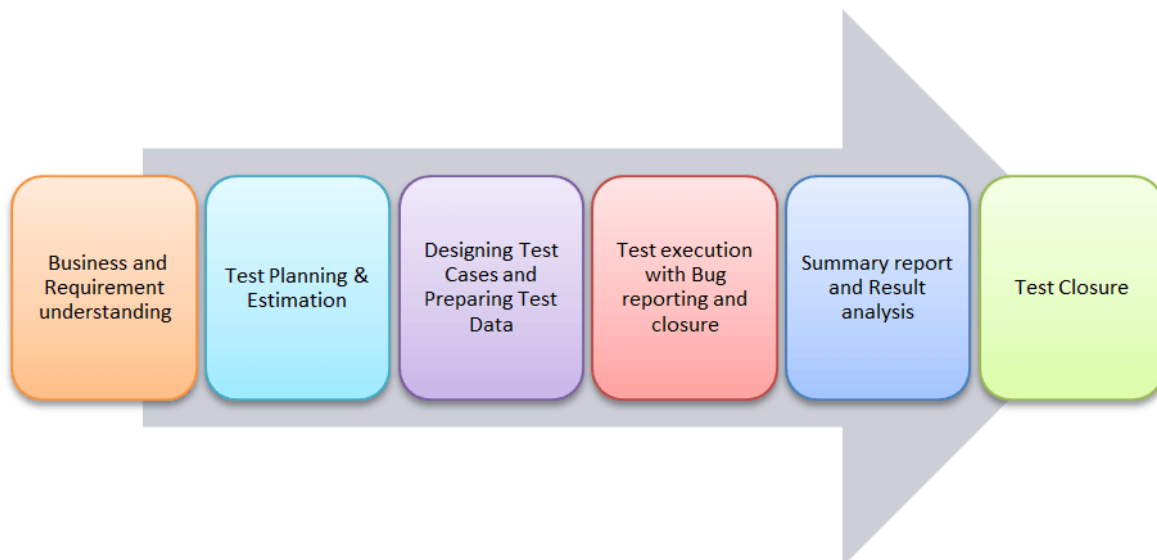
ETL testing is performed in five stages
1. Identifying data sources and requirements
2. Data acquisition
3. Implement business logics and dimensional Modelling
4. Build and populate data
5. Build Reports

ETL PROCESS

■ Validating the data required and the data sources ■ Data profiling ■ Data Quality & Performance Acceptance Criteria ■ Data transformation rules	■ Reviewing data Dictionary (Metadata) ■ Validating source to Target mapping. ■ Validation of ETL/DW Architecture ■ Validation of Data model 　- Dimensional Modeling 　- Normalized Approch ■ Indexing , partitioning, etc.	■ Archival/purge Strategy ■ Error loging/ Exception handeling/ recoverability ■ Parallel execution & precedence ■ ETL pull logic- full/ incremental (a.k.a Delta pull)	■ Test data preparation ■ ETL testing (end of end scenarios) ■ OLAP & Cube testing ■ Report testing (Drill Down/ Drill through)

Types of ETL Testing

Types Of Testing	Testing Process
Source to Target Testing (Validation Testing)	Such type of testing is carried out to validate whether the data values transformed are the expected data values.
Data Migration Testing	Such type of ETL testing can be automatically generated, saving substantial test development time. This type of testing checks whether the data extracted from an older application or repository are exactly same as the data in a repository or new application.
Metadata Testing	Metadata testing includes testing of

	data type check, data length check and index/constraint check.
Data Completeness Testing	To verify that all the expected data is loaded in target from the source, data completeness testing is done. Some o the tests that can be run are compare and validate counts, aggregates and actual data between the source and target for columns with simple transformation or no transformation.
Data Accuracy Testing	This testing is done to ensure that the data is accurately loaded and transformed as expected.
Data Transformation Testing	Testing data transformation is done as in many cases it cannot be achieved by writing one source SQL query and comparing the output with the target. Multiple SQL queries may need to be run for each row to verify the transformation rules.
Data Quality Testing	Data Quality Tests includes syntax and reference tests. In order to avoid any error due to date or order number during business process Data Quality testing is done. Syntax Tests: It will report dirty data, based on invalid characters, character pattern, incorrect upper or lower case order etc. Reference Tests: It will check the data according to the data model. For example: Customer ID Data quality testing includes number check, date check, precision check, data check ,

	null check etc.
Incremental ETL testing	This testing is done to check the data integrity of old and new data with the addition of new data. Incremental testing verifies that the inserts and updates are getting processed as expected during incremental ETL process.
GUI/Navigation Testing	This testing is done to check the navigation or GUI aspects of the front end reports.

How to Create ETL Test Case

ETL testing is a concept which can be applied to different tools and databases in information management industry. **The objective of ETL testing is to assure that the data that has been loaded from a source to destination after business transformation is accurate.** It also involves the verification of data at various middle stages that are being used between source and destination.

While performing ETL testing, two documents that will always be used by an ETL tester are

1. **ETL mapping sheets :**An ETL mapping sheets contain all the information of source and destination tables including each and every column and their look-up in reference tables. An ETL testers need to be comfortable with SQL queries as ETL testing may involve writing big queries with multiple joins to validate data at any stage of ETL. ETL mapping sheets provide a significant help while writing queries for data verification.
2. **DB Schema of Source, Target:** It should be kept handy to verify any detail in mapping sheets.