Objective:

The objective of this study is to provide an informative overview about different ETL testing approaches, different testing methods and also different automated testing tools which has been adopted by different organizations. It will provide the complete knowledge about the ETL testing. It includes the fundamentals, performance, advantage, disadvantage, comparative study of both testing methods and tools and how it evolved, the working, and the technical specifications. The present study will be conducted with the following objectives.

Objective of ETL testing is to show incorrectness and when error is detected ETL testing is considered to successful. The purpose of ETL testing is to detect error/faults/bugs, not to convert them in the software. Aim of ETL testing is to produce a product that is economic, errorless, useful and safe for people. The main goal of ETL testing is to find bugs and fixed them as early as possible.

1. Survey of ETL testing techniques in software engineering.
2. Study of different mythology of ETL Testing.
3. To study about the different testing techniques.
4. To compare the different testing methods and their aspects.
5. To study about the different automated testing tools for different environment.
6. To analyze the performance of different automated testing tools, that has been used for same environment.
7. To design and develop a methodology prioritization of test cases.

ETL Testing Techniques:

Verify that data is transformed correctly according to various business requirements and rules. Apart from this main ETL testing methods other testing methods like integration testing and user acceptance testing is also carried out to make sure everything is smooth and reliable.
Material & Methods:

- To perform the research following Tools, software, hardware components are used to support research assignment.
- The data warehousing & RDBMS concepts
- The dimensional and relational modeling concept
- Oracle 10g SQL/PL SQL, Oracle Package, Store Procedure, Functions, Sequences, Cursors, Help from DBA team
- Microsoft SQL Server 2008, Help from SQL Server DBA team
- Unix and Windows Operating System and Help from Unix administrator also need graphical tool to access the unix environment.
- The Informatica Power center 9.0.1 ETL tool
- The Lesson learned documents of past projects

Organization process assets available to support the project

ETL Testing Process:

- Validating test estimation
- Business and requirement understanding
- Test planning based on the inputs from test estimation and business requirement
- Designing test cases and test scenarios from all the available inputs
- Once all the test cases are ready and are approved, testing team proceed to perform pre-execution check and test data preparation for testing
- Lastly execution is performed till exit criteria are met
- Upon successful completion summary report is prepared and closure process is done.

It is necessary to define test strategy which should be mutually accepted by stakeholders before starting actual testing. A well defined test strategy will make sure that correct approach has been followed meeting the testing aspiration. ETL testing might require writing SQL statements extensively by testing team or may be tailoring the SQL provided by development team. In any case testing team must be aware of the results they are trying to get using those SQL statements.

ETL Testing Challenges:

ETL testing is quite different from conventional testing. There are many challenges we faced while performing data warehouse testing. Here is the list of few ETL testing challenges I experienced on my project:
- Incompatible and duplicate data.
– Loss of data during ETL process.
– Unavailability of inclusive test bed.
– Testers have no privileges to execute ETL jobs by their own.
– Volume and complexity of data is very huge.
– Fault in business process and procedures.
– Trouble acquiring and building test data.
– Missing business flow information.

Data is important for businesses to make the critical business decisions. ETL testing plays a significant role validating and ensuring that the business information is exact, consistent and reliable. Also, it minimizes hazard of data loss in production.