**Data validation**

means checking the accuracy and quality of source data before using, importing or otherwise processing data.

ETL Test Scenarios and Test Cases:

|  |  |
| --- | --- |
| **ETL Test Scenario** | **ETL Test Cases** |
| **Mapping doc validation** | We will verify the mapping document whether the ETL information provided or not. Log change should maintain in every mapping doc. |
| **Validation** | * We will validate the target and source table structure with the corresponding mapping doc. * The data type of source and target table should be the same. * Length of the data type of both source and target should be the same. * We will verify the data field type, and formats which are specified. * The length of the source data type should not be less than the length of the target data type. |
| **Constraint Validation** | The constraint should be defined for a specific table as per our expectation. |
| **Data Consistency Issues** | * Data Type and length of a particular attribute may vary in files or tables through the semantic definition. * Misuse of integrity constraint. |
| **Completeness Issues** | * Here, we have to be ensure that all the expected data is loaded into the target table. * In this scenario, record counts will be compared between source and target. * We will check the rejected records. * Data should not be truncated in the column of the truncated table. * Will check the boundary value analysis. * We will compare unique values of critical fields between the data loaded in warehouse and source data. |
| **Correctness Issues** | * This scenario is used to correct the data, which is misspelled or inaccurately recorded. * To correct the data, that is null, non-unique, and out of range. |
| **Transformation** | * This scenario is used to check the transformation. |
| **Data Quality** | * This scenario is used to check the number and validate it. * Data Check: This scenario will follow the date format, and it should be same for all the records. * Precision check * Data check * Null check |
| **Null Validate** | * This scenario will verify the null values, where "Not null" values are specified for a specific column. |
| **Duplicate Check** | * In this scenario, we will check the validation of unique key, primary key, and any other column should be unique as per business requirement having any duplicate rows. * We will check if any duplicate values exist in any column which is extracted from multiple column sources and combine them into one column. * As per the client requirements, we need to ensure that there are no duplicates in a combination of multiple columns with target only. |
| **Date Validation** | * Date values are using many areas in development to know the row creation date. * Identify the existing records as per the ETL development perspective. * Sometimes on the date values, the updates and inserts are generated. |
| **Data Cleanness** | * The unnecessary column should be removed before loading into the staging area. |

Responsibility of ETL tester

**ETL tester is responsible for validating the data sources, applying transformation logic, and loading the data in the target table, extraction of data.**

**The responsibilities of ETL tester are:**

**Verify the table in the source system. It involves the following types of operation:**

* **Count Check**
* **Data Type check**
* **Reconcile records with source data**
* **Ensure no spam data is loaded**
* **Remove duplicate data**
* **Check all the keys are in place**

**Apply Transformation Logic**

**Transformation logic is applied before loading the data. It involves the following operations:**

* **Transformation logic is applied before and after checking the record of the count.**
* **Validation of data flow from the staging area to the intermediate table.**
* **Check the data threshold validation; for example, the age value should not be more than 100.**
* **Check the surrogate key**

**Data Loading**

**Data is loaded from the staging area to the target systems. It involves the following operations:**

**We will check if the aggregate values and calculated measures loaded in the fact table.**

* **During the loading of the data, we will check the modeling views based on the target table.**
* **We will check, if the CDC has been applied to the incremental load table.**
* **Check the data dimension table and review the history of the table.**
* **Check the reports of BI which are based on the loaded fact and dimension table as per the expected results.**

Testing of ETL Tools

**ETL testers are required to test the test cases and tools as well. It involves the following operations:**

* **Test the ETL tool and its functions**
* **Test the ETL Data Warehouse system**
* **Create, design and execute the test cases and test plan**
* **Test the flat file data transfer**

Advantages of ETL Testing

**Benefits of ETL testing are given below:**

1. **ETL testing can extract or receive data from any data sources at the same time.**
2. **ETL can load the data from heterogeneous sources to a single generalized (frequent)\ different target at the same time.**
3. **ETL can be able to load different types of the goal at the same time.**
4. **ETL can be able to extract required business data from various sources and can be needed load business data into the different target as the desired format.**
5. **ETL can perform any data transformation according to the business.**

Disadvantages of ETL Testing

**Disadvantages of ETL testing are given below:**

1. **One of the main disadvantages of ETL testing is that we must be a data-oriented developer or database analyst to use it.**
2. **When we need a fast response, it is not ideal for real-time or on-demand access.**
3. **ETL testing will take months to put on any place.**
4. **It is challenging to keep the data in the changing requirement.**

**heading:**

**ETL testers are required to test the test cases and tools as well. It involves the following operation:**

* **Test the ETL tool and its function**
* **Test the ETL Data Warehouse system**
* **Create, design and execute the test cases and test plan**
* **Test the flat file data transfer**

Future Scope of ETL Testing

**The scope of ETL testing is very bright. ETL tools like Informatica PowerCenter, Oracle Data Integrator, Microsoft SQL server integrated service, SAS, IBM infosphere information server, etc. all are in huge demand in the industry because of its demand. The scope of ETL testing will increase in the future.**

Conclusion

**ETL testing is a type of business testing in which developers, business analyst, end-users, and DBAs are involved. ETL testing requires the knowledge of SDLC and ETL policies, and the tester should know how to write the SQL queries. Many businesses consider ETL as a challenge, but the fact is that it is beneficial for the business. It is essential to protect the data from loss, and it is necessary to update the data to meet the requirements of the market.**