Data transformation Test

It is a process of converting the data, cleansing the data into a required business format. Validating that data is transformed correctly based on business rules. Business rules can be the most complex part of testing in ETL application with significant transformation logic.

Between SOURCE and TARGET: - Test should make sure that the data type of each column of each table is as per the column mapping document. If no specific details are mentioned in STM document about the tables schema etc, then test should make sure on the below cases :

- The data type of the source column and destination column are same.
- The destination column length is equal to or greater that the source column length.
- Validation should be done that all data specified get extracted.

Test should include the check to see that the transformation and cleansing process are working correctly. The following types of data transformation activities take place in staging.

1. Data Cleansing
2. Data Merging
3. Data Scrubbing
4. Data Aggregation

Data Cleansing

It is a process of changing inconsistence, inaccurate data and removing unwanted data.

The ultimate goal of data cleaning is to improve the organization confidence in their data. List types of data error that needs to be addressed such as:

a. Making first character as Capital letter.
b. Decoding data
c. Rounding the decimal data
d. Missing data—removing records which contains NULL's.
e. Data that contains unwanted junk such as an apostrophic or a comma or extra spaces.
f. Telephone numbers in the wrong format

Data Merging:

It is process of integrating the data from multiple operational sources into a single output pipeline.

Data Scrubbing:

It is process of deriving new attributes to meet their warehouse requirement.

Example: - Fact Table Column “SOLD_QTY”,”SALES_AMOUNT

Data Aggregation:

It is process of calculating the summarized from details data. The following aggregation functions can be used to calculate the aggregates.

SUM (), COUNT (), AVG (), MAX (), MIN ()

1. Validate correct processing of ETL generated fields such as surrogate keys.
2. Testing slowly changing dimensions (TYPE1, TYPE2, TYPE3 Dimension).
3. Setup data scenarios that test referential integrity between tables (Dimension & Fact).