TRANSFORMATION

Transformation is in Informatica are the object which creates, modifies or passes data to the defined target structure (tables, files or any other target)

The purpose of the transformation in Informatica is to modify the source data as per the requirement of target system.

*Active Transformation

Number of rows changes during passes from source to target

*Passive Transformation

Number of rows don't changes during passes from source to target

1. Filter transformation – Active transformation

Filter transformation can be used to filter the rows in mapping that meet the condition

2. Aggregator Transformation – Active transformation

These transformation is useful to perform the calculation such as sum, average etc(mainly to perform calculations on multiple rows and group)

3. Expression transformation – Passive transformation

These can be use to calculate the values in a single row before writing to the target

4. Router transformation – Active transformation

It is similar to filter transformation the only difference is filter transformation drop the data that do not meet the condition

Where as router transformation as an option to capture the data that do not meet the condition. It is useful to test the multiple condition.

5. Lookup transformation – Passive transformation

It is used to lookup data in a flat file, tables or views. Lookup transformation is used to get a value from the lookup table based on the source value

6. Sorter transformation – Active transformation

It is used to sort data from a source transformation either in ascending order (or) in descending order into the target table.

7. Joiner transformation – Active transformation

It used to join two sources coming from to different location or from same location

TEST STRATERGY FOR ETL TESTING

There will be some standard test for DWH that should be carried as part of testing for every DWH project

There are:

1. **Data completeness**: Ensures that all expected data is loaded.

2. **Data transformation**: Ensures that all data is transformed correctly according to business rules.

3. **Data quality**: Ensures that the ETL application correctly rejects, subtitles, default values and reports invalid data.

4. **Initial load/full load testing**

5. **Incremental load testing**

6. **BI testing: Testing BI reports in data warehouse testing**

7. **Job dependency testing**

ETL (INFORMATICA POWER CENTRE ETL TOOL)

1.Power centre designer

2.Power centre workflow manager

3.Power centre workflow monitor

1.Power centre designer (Designer)

From designer window we(develop) will designer the mapping(program)

2.Workflow manager window

We can run the work flow(the mapping which has been designed in the designer window)

3.Workflow monitor

From workflow monitor we can check the status of the work flow(mapping) which has views from the workflow manager window.

*Work flow status:

1.Running(executory)
2.succeded
3.Failed(fail)
4.STOP
5.ABORT
6.WAITING

- For work flow monitor window we cannot the describe the mapping.
- can we seen work flow from work flow monitor window?

Ans. Yes, but that work flow should have been seen atleast ones from work flow manager window.(Restart)

What is mapping:-

A mapping represents the data flow from source system to target system

NOTE- From designer window we cannot run/execute the mapping(ETL program)

What is work flow:-

A work flow is a set of instruction that test the informatica server how to execute the task.

What is session:-

A session is a set of instruction that describe how and when to move data from sources to target

What is mapplet:-

A mapplet creates a set of transformation which is reusable.

NOTE:- from the work flow manager window we cannot check the status of the work flows.

From the work flow manager we cannot design the mapping

Only runs

1.DATA COMPLITENESS TEST

Are designed to verify that all the expected data loads into the data warehouse.This includes summary detailed tests to verify that all records, all fields and the full contents of each field are loaded strategies to consider includes.

- Records counts must be compared between the source data and target data

2.DATA TRANSFORMATION TEST

Created by developer

- 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 values can be the most complex part of testing as ETL application with significant logic.
- Validation should be done that all the data specified gets executed.
- Test should include the check to see that the transformation and cleansing process are working correctly. The following types of data transformation activities takes pace in staging.

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

1.Data Cleansing

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

- The ultimate goal of data cleansing is to improve the organisations confidence in their data.
- Below is the list of data errors which needs to be cleaned before loading the data into DWH.

*matching first character as a capital
*Decoding data
*Rounding the decimal data
*missing data- removing records which contain null's

2.Data Merging

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

3.Data Scrubbing

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

4.Data Aggregation

It is a process of calculating th e summarise from detail data

- the following aggregation functions can be used to calculate the aggregation
  1.sum 2.count 3.average 4.max 5. Min
- validate correct processing of ETL generated fields such as surrogate key
- Testing slowly change dimension (SCD type1.type2,type3)