**Homework 3**

1) Describe which part of this system should be tested.

1. Data sources, Internal Bank Data -> Level 1 (Staging Layer).

Check how data are extracted to Staging Layer and verify that all data are correctly stored.

1. Level 1 (Staging Layer) -> Level 2 (Data Integration Layer).

Check how data are transformed and schema formed.

1. Level 2 (Data Integration Layer) -> Level 3 (Access Layer).

Ensure all data marts are properly loaded with data in a format that can be used for reporting and analysis.

1. Level 3 (Access Layer) -> Report and analytical tools.

Test the report applicability to business, integration of BI elements, the readability of reports.

2) Provide tests to check ETL process.

* Structure Validation
* Validating Mapping document
* Validate Constraints
* Data Consistency check
* Data Completeness Validation
* Data Correctness Validation
* Data Transform Validation
* Data Quality Validation
* Null Validation
* Duplicate Validation
* Full Data Validation Minus Query
* Data Cleaning

3) Provide tests to check reports based on data in DWH.

* Data aggregation and calculation rules
* Data filtering
* Data formats
* Data types
* Data value sorting
* Drill up and drill down display correct values
* Data export
* Graphed data, Dashboard etc.
* Report data match DWH/Data mart
* Report field default values
* Report formats comply with requirements
* And match more

4) Provide tests to check performance of the system.

Data warehouse must ensure the scalability of the system with increasing loads. With this, there should not be any degradation in the performance while executing the queries, with anticipated results in specific time frames.

**Below are samples for Performance and Scalability Testing:**

* Do the performance testing by loading production volumes of data and ensure that the time frames are not missed.
* Validate the performance of each query with bulk data. Test the performance by using simple joins and multiple joins.
* Load double (or) triple to the volumes of data expected to calculate the capacity of the system approximately.
* Test by running jobs for all the listed reports at the same time.