COURSE NAME:
DATA WAREHOUSING & DATA MINING

LECTURE 10
TOPICS TO BE COVERED:

- **Testing Data Warehouse**

TESTING DATA WAREHOUSE

- Testing the data warehouse and business intelligence system is critical to success. Without testing, the data warehouse could produce incorrect answers and quickly lose the faith of the business intelligence users. Effective testing requires putting together the right processes, people and technology and deploying them in productive ways.

Testing Data Warehouse

- Testing is undoubtedly an essential part of DW life-cycle but it received a few attention with respect to other design phases.

DW Testing Specificities:

– Software testing is predominantly focused on program code, while DW testing is directed at data and information.

– DW testing focuses on the correctness and usefulness of the information delivered to users

– Differently from generic software systems, DW testing involves a huge data volume, which significantly impacts performance and productivity.

– DW systems are aimed at supporting any views of data, so the possible number of use scenarios is practically infinite and only few of them are known from the beginning.

– It is almost impossible to predict all the possible types of errors that will be encountered in real operational data.

WHAT & HOW IS TESTED

- ***Data quality:*** *entails an accurate check on the correctness* of the data loaded by ETL procedures and accessed by front-end tools.
- ***Design quality:*** *implies verifying that user requirements are* well expressed by the conceptual and by the logical schema.
  - Conceptual schema
  - Logical schema
  - ETL procedures
  - Database
  - Front-end

WHAT & HOW IS TESTED

- ***Functional test:*** *it verifies that the item is compliant with its specified* business requirements.
- ***Usability test:*** *it evaluates the item by letting users interact with it, in* order to verify that the item is easy to use and comprehensible.
- ***Performance test****: it checks that the item performance is satisfactory* under typical workload conditions.
- ***Stress test****: it shows how well the item performs with peak loads of data* and very heavy workloads.
- ***Recovery test****: it checks how well an item is able to recover from* crashes, hardware failures and other similar problems.
- ***Security test****: it checks that the item protects data and maintains* functionality as intended.
- ***Regression test****: It checks that the item still functions correctly after a* change has occurred.

DATA WAREHOUSE TESTING RESPONSIBILITIES

Who should be involved with testing? The right team is essential to success:

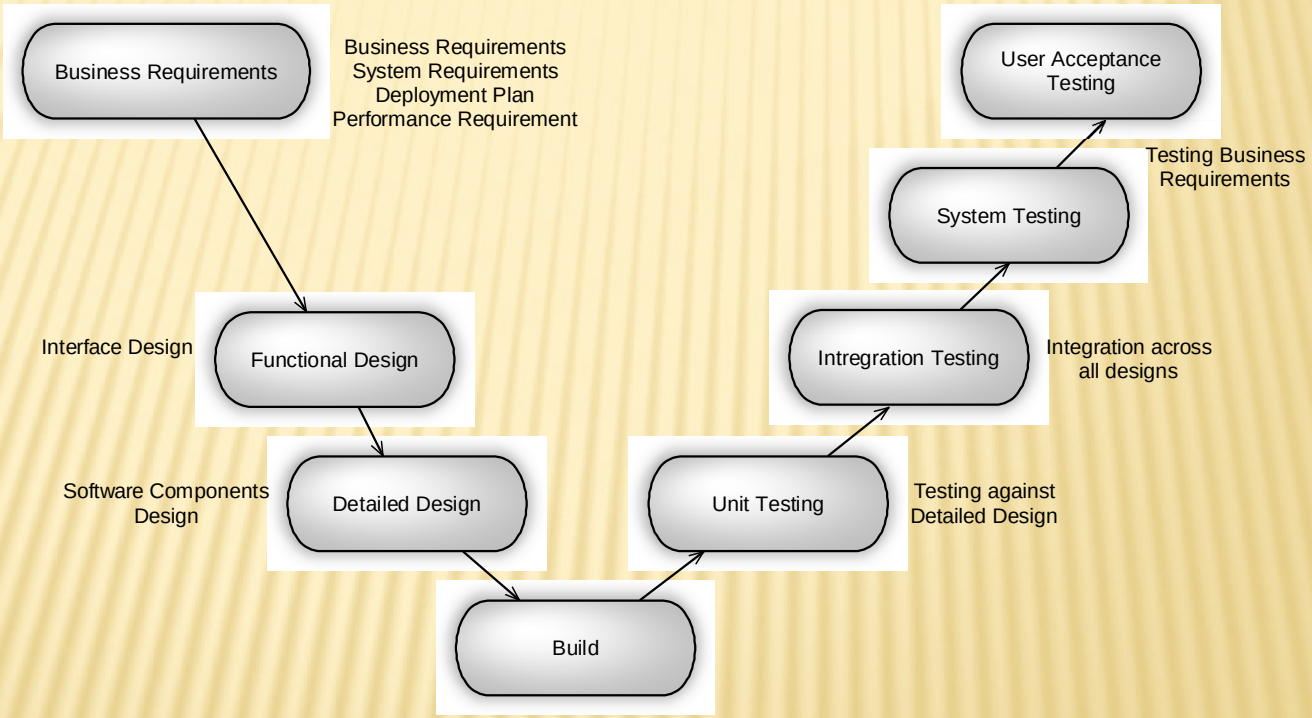
- Business Analysts gather and document requirements
- QA Testers develop and execute test plans and test scripts
- Infrastructure people set up test environments
- Developers perform unit tests of their deliverables
- DBAs test for performance and stress
- Business Users perform functional tests including User Acceptance Tests (UAT)

CHALLENGES OF DATA WAREHOUSE TESTING

- Data selection from multiple source systems and analysis that follows pose great challenge.
- Volume and the complexity of the data.
- Inconsistent and redundant data in a data warehouse.
- Inconsistent and Inaccurate reports.
- Non-availability of History data.

TESTING METHODOLOGY

- Use of Traceability to enable full test coverage of Business Requirements
- In depth review of Test Cases
- Manipulation of Test Data to ensure full test coverage
- Provision of appropriate tools to speed the process of Test Execution & Evaluation
- Regression Testing


Business Requirements
Business Requirements
System Requirements
Deployment Plan
Performance Requirement
Interface Design
Functional Design
Software Components
Design
Detailed Design
Build
Unit Testing
Testing against
Detailed Design
Intregration Testing
Integration across
all designs
System Testing
Testing Business
Requirements
User Acceptance
Testing

TESTING TYPES

The following are types of Testing performed for Data warehousing projects.

- Unit Testing.
- Integration Testing.
- Technical Shakedown Testing.
- System Testing.
- Operation readiness Testing
- User Acceptance Testing.

DATA WAREHOUSE TUNING

DATA WAREHOUSE TUNING

- Data warehouse tuning is an activity of making a database application run more quickly.“More Quickly “ usually means higher throughput, though it may means lower response time for time critical application.

DATAWAREHOUSE TUNING

- Aggregate (strategic) targeting:
  - Aggregates flow up from a wide selection of data, and then
  - Targeted decisions flow down
- Examples:
  - Riding the wave of clothing fads
  - Tracking delays for frequent-flyer customers

DATA WAREHOUSE WORKLOAD

- Broad
  - **Aggregate queries over ranges of values, e.g., find the total sales by region and quarter.**
- Deep
  - **Queries that require precise individualized information, e.g., which frequent flyers have been delayed several times in the last month?**
- Dynamic (vs. Static)
  - **Queries that require up-to-date information, e.g. which nodes have the highest traffic now?**

TUNING KNOBS

- Indexes
- Materialized views
- Approximation