Business Intelligence Automation

Digital Minds
TECHNOLOGIES INC.
Value Delivered

"Meeting of the intellectual minds where technology and creativity come together"

Agenda

WHY AUTOMATION TESTING?

> **Fast:** Compared to manual testing, automated tools run tests significantly faster.

> **Repeatable:** You can test how the software reacts under repeated execution of the same operations.

> **Programmable:** You can program sophisticated tests that bring out hidden information from the application.

> **Comprehensive:** You can build a suite of tests that covers every feature in your application.

> **Reliable:** Tests perform precisely the same operations each time they are run, thereby eliminating the potential chances of occurring human error.

> **Reusable:** You can reuse tests on different versions of an application, even if the user interface changes.

> **Better Quality Software:** Because you can run more tests in less time with fewer resources.

> **Cost Reduction:** As the number of resources for regression test are reduced.

Overview
BI Automation Testing

Normally testing of Data Warehouse or Business Intelligence (DW/BI) applications is different from the testing of traditional web based applications as it involves multiple systems and hence requires a data-centric testing approach.

The Typical Challenges Faced While Testing DW/BI Implementations Include

Volume of data, variety and complexity

Data anomalies from different data sources

Possibility of data loss during data integration process and handshaking mechanism between sources

Time consuming activity

No audit/log trails, reusability resulting into high cost of quality

Specialized skills required to execute data validation and verification process

To ensure data completeness, accuracy, consistency, security and reliability throughout the life cycle, it is important to test all these aspects at each data entry point in the BI architecture and not just at the end through reports or dashboards.

Test Strategy and defining the scope

The key areas to focus are

Scope of testing: Describe testing techniques and types to be used

Test environment set up

Test Data Availability - it is recommended to have production like data covering all/critical business scenarios

Data quality and performance acceptance criteria.

Factors to consider for defining the test scope

Size of project

Complexity of project

Budget for project

Time scope for project

Number of staff

Validating a typical BI scenario

Data Collection

The primary aim is to ensure that all of the data is extracted that needs to be loaded in the target. The key areas this phase should focus on are:

Validating and the availability of the data sources from which this data needs to be extracted.

Data profiling to detect any data issues early in the cycle

Data Integration

Once the data is transformed, thorough testing needs to be executed to ensure underlying data complies with the expected transformation logic. Key areas this phase should focus on are:

Validating the Data Model

Reviewing the Data Dictionary

Validating the Source to Target Mapping

Data Storage

The data loads can be one time, incrementally or in real-time. Key areas this phase should focus on are:

Validating data loads based on time

Performance and Scalability

Parallel Execution and Precedence

Validating the Archival and Purge Policy

Verifying error logging, exception handling and recovery from failure points.

Data Presentation

This is the final step of the testing cycle and has the privilege of having a graphical interface to test the data. Key areas this phase should focus on are:

Validating the Report Model.

Report layout validation as per mockups and data validation as per business requirements

End to End Testing

Validating a typical BI scenario

End-to-End Testing of Data Stream Throughout the Data Lifecycle

Test Entry Point   Test Entry Point   Test Entry Point   Test Entry Point

Source Data   ELT Process   Data Warehouse   ELT Process   Data Mart   ELT Process   BI,Analytics & Reporting
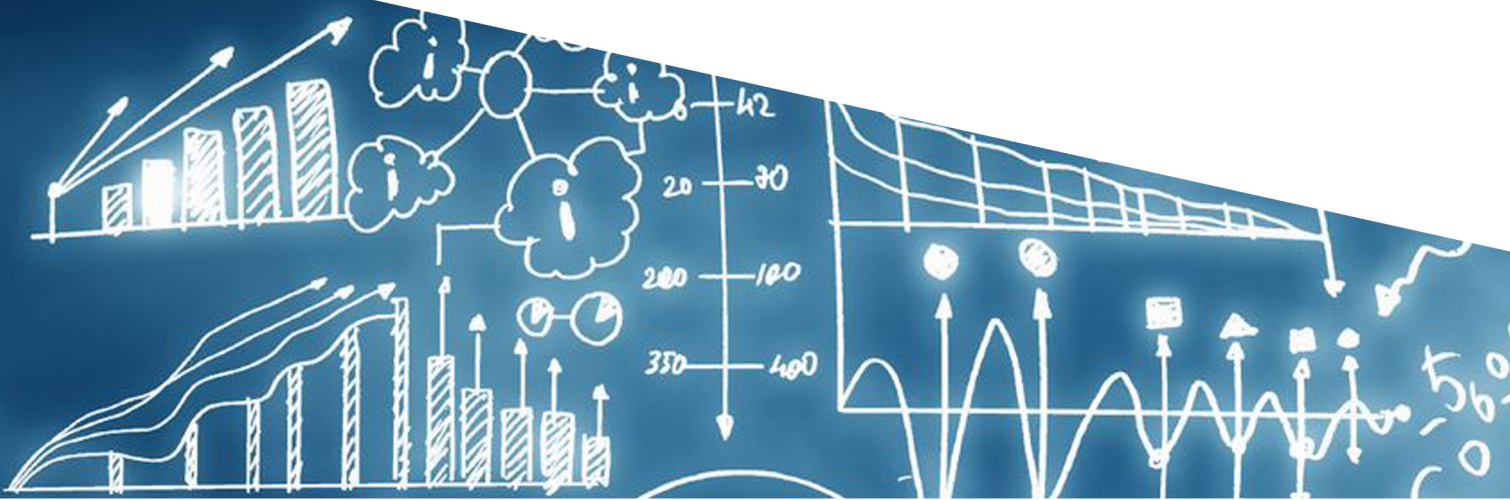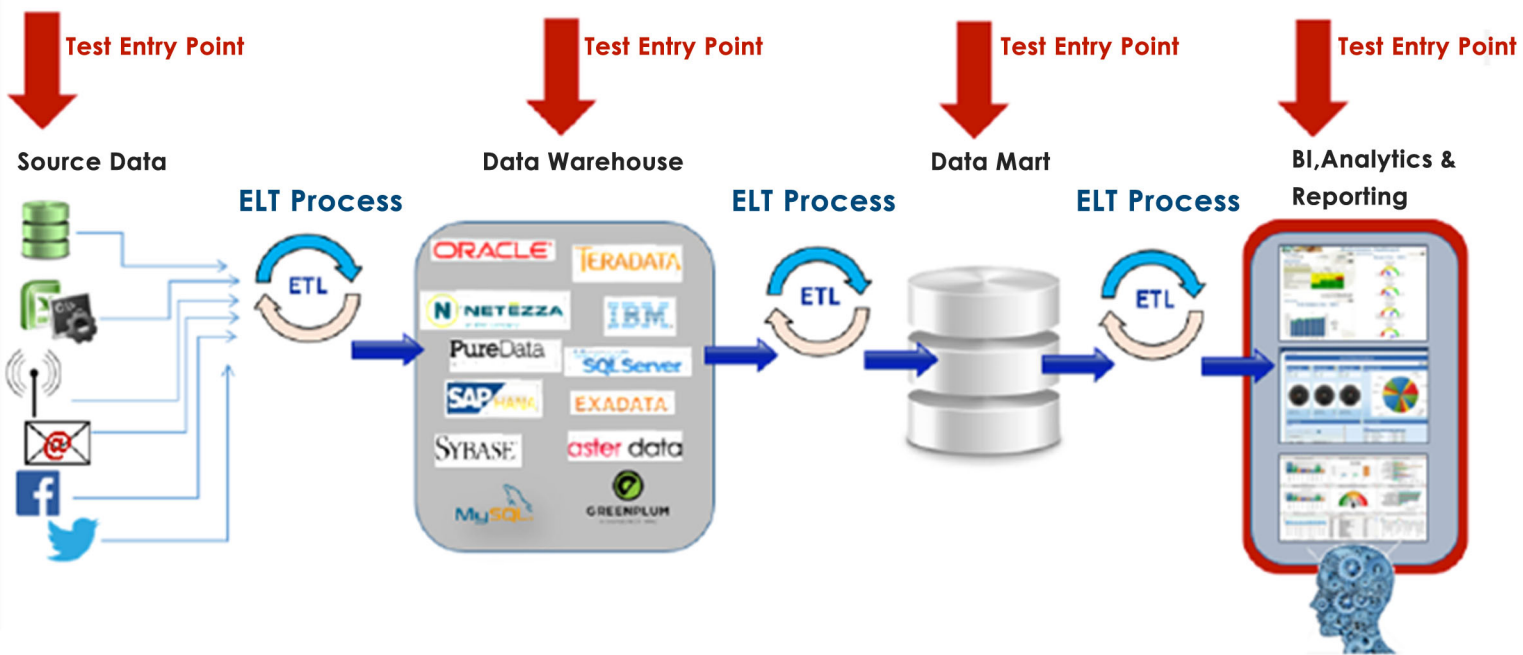
System Testing
BI Test Approach

System Testing - ETL

Extraction

Verify extraction jobs / routines moving data to the Staging Area to ensure coverage of all DW Source

Verify whether all source extracts are present as per requirements applicable for each load.

Validate data loaded to the staging area by comparing row counts & check sums

Transformation & Load

Verify transformation rules/components and ETL job specs against mapping rules against the business requirements

Perform data reconciliation in the Target Layer through data validation checks to

Compare Record counts & Check sums	Check unique values of key fields, truncations and filed boundaries	Compare range/distribution of values b/w source & target data
Validate match/merge, cleansing & de-duplication for Master Data	Validate derivations, aggregations etc.	Validate parent-child relationship

Review data load procedures in ETL jobs and batches to

Verify DB connections of source/ target table accesses by ETL jobs and load procedures	Verify sequencing and scheduling of batches

Test Refresh/Initial and Incremental/delta Load procedures through separate testing and data reconciliation checks

Test Error processing by checking rejected records, error tables and error logs for ETL jobs.

Test Restart and Recovery of ETL system.

Verify the performance of ETL Workflows & jobs against requirements

System Testing – BI Test Approach

System Testing - Reports

Reports Testing

Review Semantic Model of reports/Cubes

- Correct Business views & Names
- Relationships & Mappings
- Business Logic
- DB Connections & Security settings

Validate Report Data through data reconciliation

- Report data contents against DB views.
- Linkages of reports to sources
- Base and derived metrics
- Dimensions and attribute hierarchies
- Data outputs from prompts, drilling, roll-ups, sorting, export functions and other special features

- UI features - Report layout , format etc.
- Data Values, Headers, Footers, Column Labels and other information presented
- Ad-hoc query results
- Scheduling and dispatching of reports

Verify the non-functional aspects of the reports against requirements

- Performance against requirements of report response time
- Security and accessibility against user roles and entitlements

System Integration Testing

Test overall flow of data across upstream and downstream applications

Test touch points between applications

Perform End -to-End Testing of business critical reports . (Sources->staging Area -> EDW/Data Marts->Reports). Run reports in the same sequence and with the same dependencies as in Production

Test reports across different platforms, language/time zones as per requirements

System Testing – BI Test Approach

Data Model & DB Testing

Review Conceptual and Logical Data Model of DM/EDW against Business Requirements

Verify the following:

- Dimensions & Facts
- Relationships
- Aggregates
- Schema Definitions

Review Physical Data Model of DM/EDW against Business Requirements

Verify the following:

- Primary keys
- Foreign Keys & other constraints
- Data types & default values
- Indexes for optimal Performance

System Integration Testing
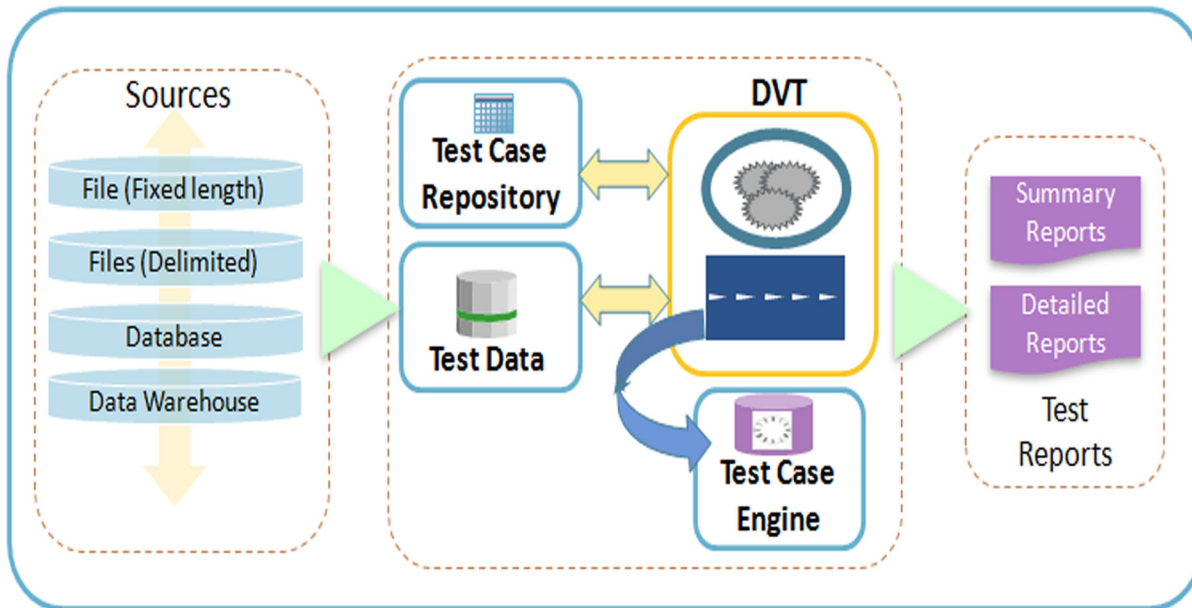
Test overall flow of data across upstream and downstream applications

Test touch points between applications

Perform End -to-End Testing of business critical reports . (Sources->staging Area -> EDW/Data Marts-> Reports). Run reports in the same sequence and with the same dependencies as in Production

Test reports across different platforms, language/time zones as per requirements

Tools - Data Validation Tool

Sources
- File (Fixed length)
- Files (Delimited)
- Database
- Data Warehouse

Test Case Repository

Test Data

DVT

Test Case Engine

Summary Reports

Detailed Reports

Test Reports

Business rule validation of 2M records
<20 mins in DB
<30 mins for Excel
<40 mins for flat files

Comparison of 1M records(100 cols) - (55% data mismatch)
<28 mins in DB
<14 mins for Excel
<8 mins for flat files

~25-35%
Reduction in Test Cycle Time

~35-45%
Reduction in Manual Effort

Performs Business Rule Validation on individual Tables/files

Compares source and target data

Has Multi-DB connectivity to DBs such as Oracle, DB2, SQL Server, Teradata, Postgres, Netezza and connectivity to files.

Has predefined set of validations and an enhanced query builder to create test steps.

Test case can be stored in the tool's repository and scheduled for later execution.
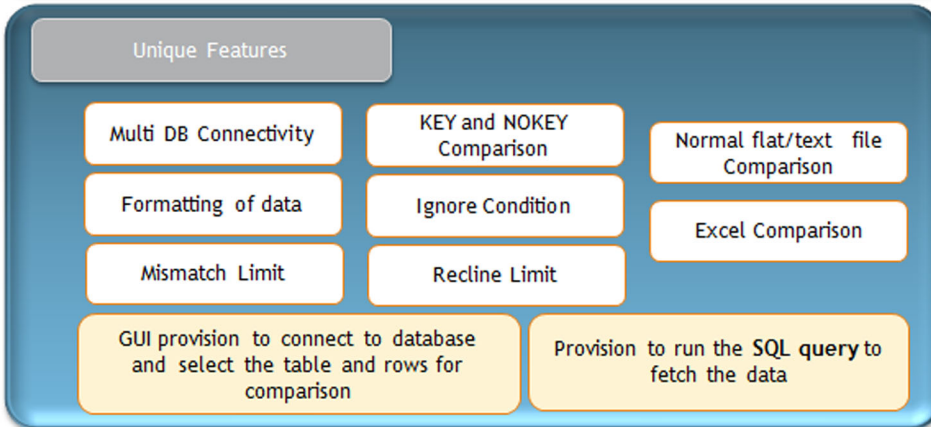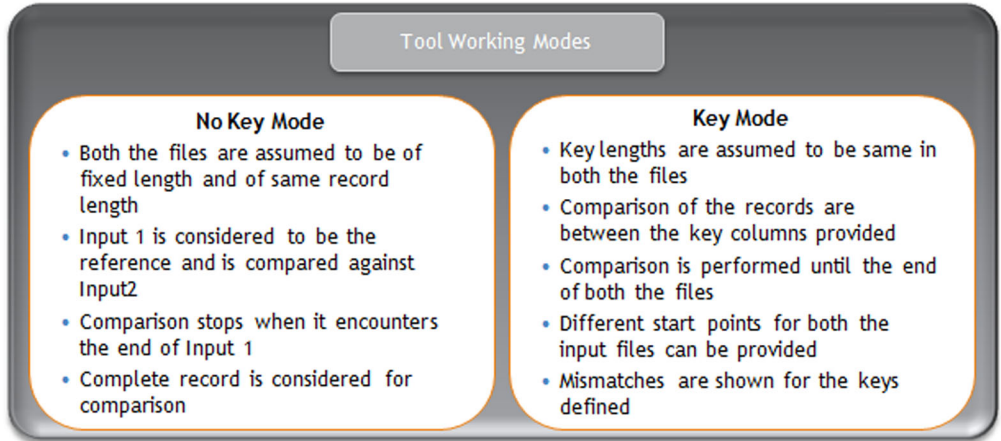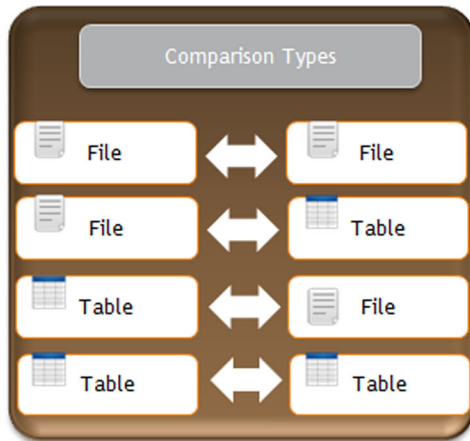

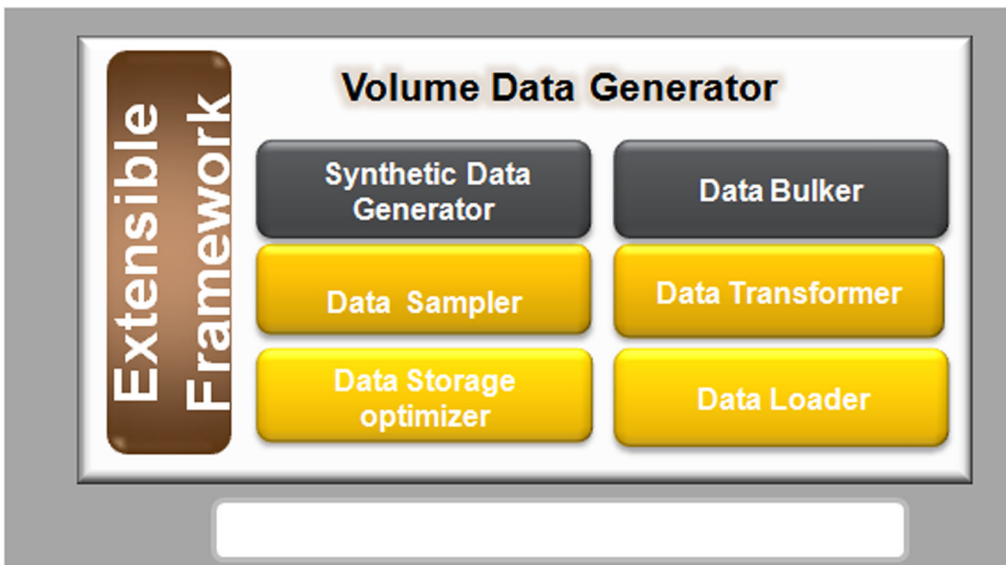## Validation of

DWH-BI Testing → **Extracted data in Staging DB/File** | **Transformed Data in target DB** | **Data in End-user Reports / Queries**

Tools - Data Comparison Tool

Comparison Types

File	↔	File
File	↔	Table
Table	↔	File
Table	↔	Table

Tool Working Modes

No Key Mode

- Both the files are assumed to be of fixed length and of same record length
- Input 1 is considered to be the reference and is compared against Input2
- Comparison stops when it encounters the end of Input 1
- Complete record is considered for comparison

Key Mode

- Key lengths are assumed to be same in both the files
- Comparison of the records are between the key columns provided
- Comparison is performed until the end of both the files
- Different start points for both the input files can be provided
- Mismatches are shown for the keys defined

Available Platforms

UNIX	Mainframe	Windows

Unique Features

- Multi DB Connectivity
- KEY and NOKEY Comparison
- Normal flat/text file Comparison
- Formatting of data
- Ignore Condition
- Excel Comparison
- Mismatch Limit
- Recline Limit

GUI provision to connect to database and select the table and rows for comparison

Provision to run the **SQL query** to fetch the data

Value Propositions

- Cost Effective
- Easy Customization
- Support for tool
- Tested and used in various projects

Tools - Volume Data Generator

Volume Data Generator

Extensible Framework

- Synthetic Data Generator
- Data Bulker
- Data Sampler
- Data Transformer
- Data Storage optimizer
- Data Loader

- Creates High volume test data either from scratch or from a small volume of seed data.
- Data Sampler can help create a small sample of seed data from any source
- Rich set of in-built data transformation techniques.
- Pluggable Interface for adding custom data transformation technique
- Data storage optimizer to save disk space.
- Framework for easy loading of data in target environment.
- Creates data for heterogeneous databases and platforms.
- Supports resume operation from point of failure.

Recommended Best Practices in BI Testing

▶ Shift Left approach – Involve the testing team from Requirements Analysis phase itself to understand Test requirements, define test strategy and start test planning

▶ Environment set up verification and Test Data Creation with positive and negative test scenarios

▶ Data profiling in Dev and QA environment to ensure Data quality in target system

▶ Detailed Data Reconciliation checks

▶ Benchmarking performance of ETL jobs and verifying the actual performance statistics against these benchmarks

▶ Prepare Source to target mapping (STM) documents of ETLs used in Data Migration and get these reviewed and approved by BA and development teams to remove the ambiguity and to confirm the functionality of custom ETLs.

▶ Verification of functional validity of the report data by comparing against raw data from application screens

▶ Following a Risk Based Test Strategy to ensure that all critical mappings , frequently run tests, tests with high possibility of failures etc. are tested on priority.

▶ Following change management control for all the testing artifacts

Risks & Mitigations - BI Automation Initiatives

S. No	Risk Description	Mitigation Plan
1	Requirement Definition	• Document the detailed requirement and sign-off. • Identify critical ETL jobs & workflows and reports to be tested for End-to-End testing upfront
2	Skill Set	• DMT has pool of talented testers under assurancepractice
3	Non-availability of reports / Dashboards for testing as per the test schedule.	• Identify the external system impact during release planning andinform to the customer Release manager for timely readiness • Constant follow-up by test lead for external system readiness.
4	Availability of Test Data (Content)	• Identify test data requirements in beginning of test planning cycle and inform stakeholders for timely readiness
5	Testing for data utility	• E2E test strategy is aligned not only for data flow check, but also for correctness and completeness of data against the requirements
6	Access Control Process	• Working with DBAs and BAs ensuring getting access to the required DB schemas, source extracts, ETL repositories and Reports with appropriate privileges.

Risks & Mitigations - BI Automation Initiatives (Contd.)

S. No	Risk Description	Mitigation Plan
7	Lack of a standard methodology for testing and tools for automation	• DMT gives an Approach, Strategy and Road map for BI Testing and Automation
8	Managing the Inter-Dependencies of Data warehouse Projects, simultaneously running in the same environment	• Test Manager would request for separate test environment based on the criticality and the business needs of a project
9	Testing gets Impacted because of Shared tables being constantly modified by different development teams	• Well defined communication plan in place so that the testing gets data that is frozen and up to date.
10	Dependency of external vendor and hence delay in running the batch	• Define SLAs and adherence; Escalate (based on type of issue and Governance structure) if testing time get shortened
11	Historical data will normally be 1 time load. But incremental loads have impact on data versioning, and delaying the testing cycle.	• Decision on what data to be incrementally loaded has to be decided and test team should test only those incremental

VALIDATIONS APPROACH

QA VALIDATION FRAMEWORK

PROCESS VALIDATION

- Validate Weekly & Daily extracts
- Validate incremental load
- Validate Sequential and Parallel Processes
- Validate Archival & Purge strategy
- Validate Error logging
- Validate Rejection of Bad Records

DATA VALIDATION

- Validate Data Transformation
- Validate data between source and Target mappings
- Validate data masking/integrity
- Validate Data Aggregations
- Validate No Data loss from input to output

Review & Analyze current Framework ➡

Continuous Optimization

Identification of test scope effectively

- Thorough handshake between Dev and Testing on the impact analysis of the weekly code drop

- Scoping and prioritizing regression suite based on priority, defect distribution in the past iterations, code change, frequency of testing, impacted modules, no of requirements covered

Benefit: Reduction in the Cycle Time

Feedback from Product Leadership and Prod

- Adhoc testing is carried out by business analyst as a business user to unearth any data defects. These scenarios are added to the test suite.

- For any defect that occurs post testing, detailed root cause analysis is done and preventive and corrective actions are recorded

Benefit: Quality of Production environment improved

Forecasting Hurdles well ahead

- Top 10 issues in testing is published on a weekly scorecard to all stakeholders. Issues are assigned to respective owners and tracked to closure.

- All delivery dates are mapped against Prod and UAT dates. Gaps between actual VS estimated progress published with mitigation plans.

Benefit: Certainty on Delivery

Collaboration with Prod support effectively

- Testing team alarms production support team if time taken by any processes is longer or any manual intervention required. (to update run books)

- After release or refresh, pain areas in production environment are published by prod support team so that the root cause analysis is done by Dev and Testing team to put in fixes for next run.

Benefit: Improved Deployment process

Case study for one of the leading e-commerce company in the USA

Source Data	ETL	OLAP	BI

BI

- Legacy
- RDBMS
- External

Heterogeneous Sources

Staging Area → EDW → Models / Cubes / Data Marts → Reports / Analysis / Dashboards

Test Automation

Queries/Scripts, Validation Tools	Data Validation Tools (DVT)	Tools (DVT)	SQL, Script	Perf Test Tools	Tools (DVT)	Tools (DVT)	Functional Test Automation Tools(Eg. QTP)

| Data Profiling / Record Count / Data Structure | Data Comparison & Validation (Sources To Staging) | Data Comparison & Validation (Staging to EDW) | Business Rules / Aggregation / Summarization / Ref Data Test | Verification of Data Model | Analysis of Data & DB Performance | Data Reconciliation | Comparison & Validation of Report Data (DWH to Output) | Verification of Metadata Model / Report Layout / Security / BI Functions |

Reports
Test Summary, Test Detail Reports, Test Analysis, and Test Automation Dashboard

---

Digital Mind Technologies Corporate Headquaters

1111, North Plaza Drive, Suites 465

Schaumburg, IL - 60173

Desk: 1847 463 9300

Fax: 1800934 6716

www.digitalmindstech.com

info@digitalmindstech.com