

Test Data Management – Better Test Data for Better Testing

By Ankur Mathur, <Author 2>

1. Abstract

Technology innovations are increasingly intensifying the pressure on IT departments globally. To keep a pace with that, testing functions are getting more focused and agile. For clients who need to transform their testing function and build high-quality software and applications without large overhead costs, Test Data Management is the way to go.

Test data management is the practice of applying a structured and centralized approach to the management of test data at an enterprise level, in order to reduce cost whilst increasing efficiency and quality in development and testing. It is, therefore, perhaps no surprise that more and more organizations are looking to implement TDM solutions that provide a systematic approach to creating and provisioning "fit-for-purpose" test data, on demand, throughout the SDLC.

This paper points out the importance of test data management in current world, the challenges faced, benefits and Capgemini's approach for managing test data.

2. Background of Test Data Management:

Test data is an essential component of the software testing process. To accurately predict testing outcomes, QA needs reliable sources of well-defined and consistent data as input for the testing process. To achieve the right level of test coverage, testers require extensive test data with multiple instances and sufficient variations. Organizations often struggle with the increasingly large volume and configurations of data required for testing, which are magnified by the need to maintain multiple versions of the data in the test environment. The research shows that, in 16% of cases, organizations create test data as they go. Tests based on this type of data are difficult to repeat and even more difficult to automate. The increase in the number of instances of "on the fly" test data generation from 5% in 2012 to the current 16% can be attributed to ever-increasing application complexity. As IT systems become more intricate, so does the challenge to accurately prepare the test data. Solving the dilemma of increased complexity versus test data quality requires extensive knowledge of the data needed for testing.

Test Data management in earlier days:

- Historically Test Data requirements was fulfilled in adhoc manner
- Developers are leveraged to find test data for testers
- Testers used to randomly select the test data and use it on trial and error basis
- Un-reliable test data was used for testing the newly developed applications
- Test data with same attributes was created for the test cases with any complexity
- Knowledge gap of testers was a major hindrance for quality of test data
- No predefined set of test data was prepared; Random data was created and utilized for testing

■ All the figures are taken from the World Quality Report 2013/14/15

3. What is Test Data management?

Test Data Management is a crucial factor in providing a consistent and standardized approach to testing projects. It ensures the availability of the right kind of data, in right amount and at right time in the non-production environment. Test Data Management is currently moving from the secondary needs to the basic needs for an organization.

Test data preparation and utilization in the testing process eats up major portion of the Testing timeline. Test Data Management caters to the Testing teams for the test data minimizing the efforts and cost overhead.

Test Data Management utilizes the optimum data from the production masks the data and stores it on the Testing database. This data can be utilized by the Client's testing team along with the customers to the Client. TDM not only delivers the data but also ensures that the security of the live data is not compromised. It plays a major role during testing of applications in the heterogeneous architecture. Organizations have either started the setup of the TDM services or are in the process of starting it.

4. Need of an Established TDM Organization

Complex IT environment and Organization with number of inter related and technically diverse application; have emphasized the need to have a credible test data. Organizations now consider lack of test data a major impediment in the execution of effective testing.

According to the world quality report (WQR) 2014-15 as shown in the below image, Organizations without TDM teams are facing the following problems in relation to the creation and maintenance of test data. Setting up the Test Data Management process bring efficiencies to test data setup, improve the quality and productivity for developing the applications, which in turn would help in reducing the cost of development and improving time to market.

Exhibit : Challenges organizations face to control and retain test data



43%

of respondents found it challenging to Manage the size of test data set



41%

of the respondent found it difficult to maintain the right test data set versions with different test versions



41%

of the organization researched were not complying with data security and data privacy regulations for test data



37%

of respondent found it difficult to create and retaining useful copies from production data



35%

Of the respondent found it difficult in creating and maintaining test data which are not copies from production data



32%

of large programs lacked test data for complex integration testing across systems and organizations

5. Capgemini Point of View (POV) on TDM

Test Data Management is currently moving from the secondary needs to the basic needs for an organization. There are scenarios wherein the TDM helps in resolving the issue with minimum investment following are few examples that are explained

Currently the IT infrastructure is complexly inter-connected environment, with Mergers, Acquisitions and Technologies becoming outdated have added to this complexity. There are scenarios where it's mandatory for organizations to develop a service/application, which bring the legacy database and the new database in play. Data from this multiple databases or file systems interact with each other to generate a required output. Testing or developing any application/service in this kind of inter-connected environment leads to heavy investment in efforts, storage systems. This is the when the Test Data Management Process comes in the picture and a proper implementation of TDM helps in maintaining the cost and delivering the solution in time.

6. Typical TDM Challenges

In recent days most of the organizations are facing the challenges for acquisition of the Test Data, With Capgemini TDM team's experience of implementing and managing testing services for large organizations, following challenges were identified:

Data not masked revealing the live data.

Data required for testing from a huge volume increases time to market of the new services.

Data not consistent across different platforms, databases.

Inability to work with production data due to size / security concerns.

Multiple testing teams utilizing the same set of data..

Timely availability of quality test data that matches business rules.

Unavailability of the same set of test data for Regression testing.

Test data aging is complex & time consuming.

External interface test data is limited & takes time to setup.

Financial balancing requires dedicated test environment & controlled test data.

High volume of data required for non-functional testing.

Conversion dry runs & data quality scans require full production volume data sets.

7. Capgemini TDM implementation approach:

Capgemini's Test Data Management (TDM) implementation approach begins with the assessment using Quality Blueprint™. QBP approach can set an organization on the path to achieve a higher level of TDM maturity. Quality Blueprint™ is more than just a methodology or framework—it includes automated tools and reports which bring rigor to your assessment, peer benchmark and roadmap. It uses a continuously updated repository which contains real-world data from hundreds of test assessments for financial and healthcare services firms. Following are the phases for the implementation of the TDM process.

Initiation

- TDM Maturity Assessment using Quality Blueprint™ to identify gaps and areas of improvement in terms of Quality, Coverage and Maturity
- Accelerated Solutions Environment (ASE) for Stakeholder alignment and future state definition

Pilot and Setup

- Establish TDM Governance, process and framework, metrics, metadata initial setup, pilot selection and solution design.

Rollout and Enablement

- Enable data creation at enterprise level for other Business Application Services area
- Track metrics and incrementally develop the metadata values for reusability and storage of complex queries

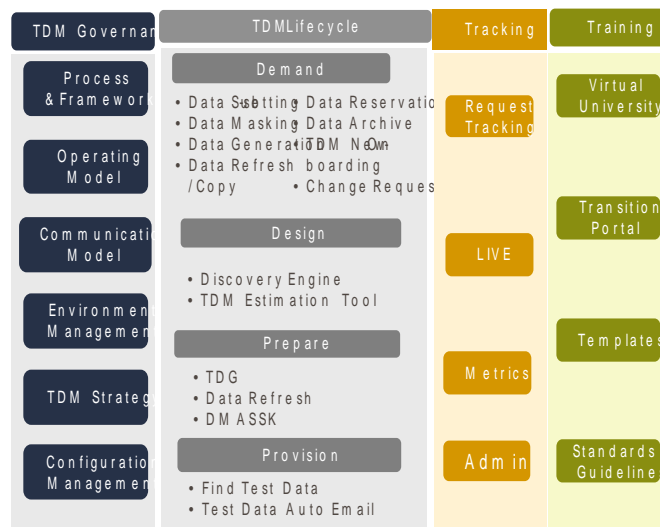
Operations

- Operations phase deals with real-time activities of Centralized TDM by Client Team with Capgemini support and ongoing Improvement tracking SLAs and Metrics

8. Capgemini's Test Data Management Solution

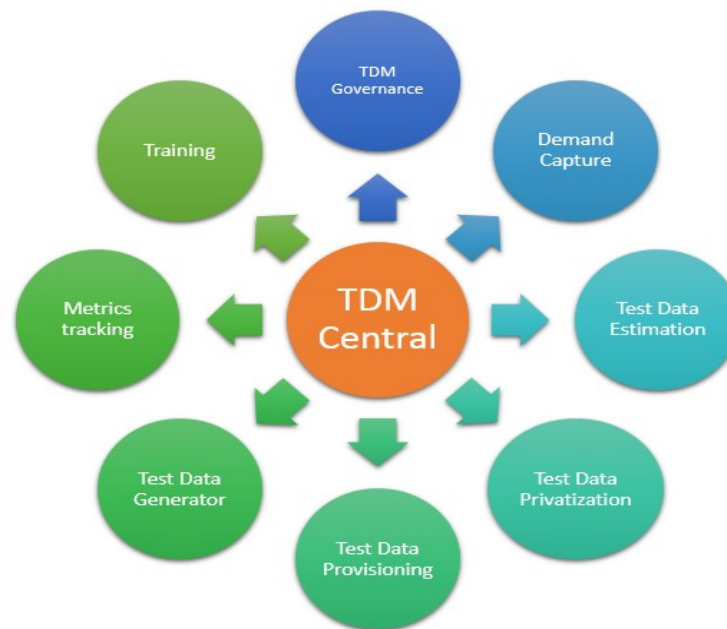
Capgemini implements the tested and proven concept for Test Data Management through its solution known as **Test Data Management Central (TDM Central)**. Following snapshot gives the functional view of the TDM Central, The snapshot depicts the four major pillars to establish the Test Data Management Central; it contains TDM Governance TDM Lifecycle, Tracking and Training.

Exhibit : Capgemini Test Data Management Solution



TDM Central is a one stop portal of all of test data requirements. The portal provides a single view of the dashboard, a self-service portal for testers, test data requests – on boarding, change management and refresh management, data reservation reporting across management and business units. This will enable Client's to have a central tracking mechanism for all Test Data Management requirements. It is an integrated End to End Test Data Management Framework to build efficiencies throughout Test Data Management Life Cycle.

Exhibit : TDM Central



9. Key takeaways:

Centralized TDM helps in reducing the time to market for new applications, new feature rollouts resulting from process standardization, optimize of the QA process framework to meet specific business requirements.

Benefits of setting up the Centralized Test Data Management:

- 25% effort is involved in TDM as compared to the test data activities for testing.
- Centralized TDM governance and processes –co-ordination between the testing teams.
- Industry process standardized (industrialization).
- Consistent data masking across the entire application landscape.
- Issues of Data security compliance (HIPPA, PHI, PCI and PII) are addressed due to data masking in a Test Data Management activity.
- Increase in the self service capabilities in test data area of an organization.
- Reporting to management.
- Usage of Tools, automation reduces manual intervention thus providing the quality data.
- Training to client resources.