Multidimensional Challenges and the Impact of Test Data Management

To deliver benefits such as reducing test windows, improving test quality, optimizing data processing/storage costs and maintaining complex data model/relationships, TDM must be embedded equally within an IT organization’s QA and development teams.

Executive Summary

Managing costs without compromising quality is the latest mantra of IT organizations. End-to-end test data management (TDM), as against mere provisioning of data for test execution, has far-reaching potential and can help rein in operational and processing costs.

Over the past few years organizations have focused more on test data management, realizing its potential impact on multiple aspects of the quality assurance (QA) process, such as testing window reduction, improvement in quality of testing, optimizing data processing and storage costs and maintaining complex data models and relationships.

To deliver these multidimensional impacts, TDM must be embedded within quality assurance and development teams, alike. Moreover, TDM must also intersect with other disciplines such as release management, environment management and data security (see Figure 1). TDM acts as a conduit between development and QA teams. It also needs to interface with the disciplines shown in the diagram to deliver the larger impacts discussed in the prior paragraph. These interfaces with the various disciplines increase the complexity that must be overcome.

Why is TDM so pertinent? Our survey of the top 250 lines of business served by us makes clear why:

- 80% of IT organizations acknowledge that they either have no tools to query/retrieve the right data or require improvements to their in-house tools.
- 61% of IT organizations do not have any TDM tools.
- 50% of IT organizations do not have a data reusability strategy in place.
- 40% of IT organizations face data integrity issues in their test environments.
- 38% of IT organizations have sensitive production data that can be accessed by in-house testing or QA teams.

This white paper explores various dimensions of potential benefits organizations can derive from effective TDM and what is needed for QA teams to leverage these elements to achieve critical multidimensional benefits.
Unleash the Potential of Test Data Management

By maximizing the potential of TDM, IT organizations can address the above challenges and deliver several valuable benefits:

- **Increased quality of test conditions and the leveraging of data integration tools.** Based on our survey, nearly two-thirds of IT organizations do not have any TDM tools. Many organizations face a common challenge of ensuring that the test environment replicates the exact patterns of production. Some data integration and data profiling and extraction tools are built with features that can help identify data patterns in the production stage. Such tools can be used to capture and mimic similar patterns in the test environments. This can help elevate test data quality to the next level. The tools result in increased business confidence regarding test coverage and, eventually, in higher quality applications.

- **Optimized storage and processing costs.** Typically, the size of a test environment increases as additional projects are brought online. This results in increasing storage costs. Bulky test environments also add indirectly to processing expenses such as archival and batch processing costs, because larger volumes of data are processed.

- **An efficient TDM framework needs to consider not only making the data available to testers but also must minimize the volumes of data in a test environment.** A test data strategy needs to consider a long-term pipeline of activities and arrive at an optimal test bed that does not compromise data availability. In addition, the test data strategy must be supplemented by an optimal archival, refresh or restore plan.

- **A shorter batch process window.** This is a natural extension of the resulting impact of the previous challenge. Not only does it cost more to process unnecessary data, but also the batch process itself takes longer than needed. Today’s TDM approach tries to minimize the batch window.

- **Improved compliance to data protection policies and regulatory requirements.** Our survey also revealed that over one-third of IT organizations let in-house testing teams access sensitive production data for their testing needs. IT organizations require test data architects to devise a solution to comply with their data protection policy, while simultaneously retaining production-like quality test data. Persistent and dynamic data masking solutions are highly sought-after, especially in the context of the dynamic environment of regulations. This is more pronounced in domains such as healthcare, banking and financial services and insurance where data privacy is a highly sensitive concern.

Also, a small percentage of IT organizations quoted data privacy as a roadblock for offshoring. Data masking solutions can enable...
offshore test execution in compliance with the policy and relevant regulatory requirements. The key here is to ensure data integrity of the masked data, without exposing sensitive data. There are also instances in which a synthetic data generation solution has been leveraged to overcome data privacy concerns, paving the way for offshoring of TDM.

- **Better technical support to black-box testing teams for knowledge management.** Nearly half of IT organizations surveyed face data integrity issues in their test environments. Test data teams can act as a bridge between the white-box development teams and black-box functional testing teams. While the focus of functional testing teams is to treat the application under test as a black box, the setup of the prerequisite test data requires a deep understanding of the underlying data structure and flow. Thanks to test data management tools, the knowledge about numerous business rules and the interactions among interfaces can be captured in a more elaborate and robust manner. This can lead to enhanced knowledge quality, better retention and easier maintenance.

- **Improved test coverage.** It is commonly assumed that production data, though present in large volumes, only covers about 30% of possible functionalities or scenarios. To maximize test coverage, testers need to condition the data to ensure testing of all possible scenarios. Here, a synthetic data generation solution can play a crucial role in creating the much-needed variety of data with desired distribution patterns. It can also be extended to improve coverage of performance testing by generating volumes that are sometimes many times bigger than the current production size. Nearly two-thirds of IT organizations surveyed have synthetic data generation needs.

Also, as part of migration/upgrade scenarios, when large volumes of data are moved from source to target systems, it is only a small portion of target data that is typically tested manually with the help of scripts and queries. The low coverage of testing leads to the risk of untested data at the target. Organizations can leverage validation-related products, either those in-house or the ones available commercially, to ensure 100% coverage.

- **Optimized data provisioning time: TDM accelerators.** More than two-thirds of the IT organizations surveyed acknowledge that they have no tools to either query or retrieve the right data. To capture the business rules and understand the information data structures more efficiently, accelerators can be set up that readily contain this information for packages such as SAP or Siebel. Accelerators are also available for regulatory compliances such as Payment Card Industry’s Data Security Standards (PCI DSS) and Protected Health Information (PHI). These accelerators not only cut down the configuration time and effort, but also improve the consistency of how these rules are applied.

- **Maximized utilization of knowledge and skills.** A small percentage of IT organizations surveyed have centralized their test data team services. As such, the benefits of a managed services model can be applied to TDM. TDM requires skills in niche areas such as extract-transform-load (ETL) and online analytical processing (OLAP)/online transaction processing (OLTP) which may not be necessarily available with functional testing teams. Having such high-skilled resources functioning as TDM specialists can help achieve greater efficiency.

Thanks to test data management tools, the knowledge about numerous business rules and the interactions among interfaces can be captured in a more elaborate and robust manner. This can lead to enhanced knowledge quality, better retention and easier maintenance.
Managed Services TDM in Action

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Improvements and Changes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Multibillion Banking &amp; Financial Services Organization in the UK</td>
<td>Manually intensive test data provisioning. Lack of a formal data request tracking mechanism. Lack of consolidated knowledge on data conditions.</td>
<td>Centralized team with internal ownership of knowledge management. API-based automation (data extraction and data generation). Centralized data request tracking: SLAs based on the turnaround time and data suitability.</td>
</tr>
<tr>
<td>A U.S.-Based Retail Conglomerate</td>
<td>Complex application design. Sensitive production data used in test environment. Full load of data each time, making it longer than required.</td>
<td>Deployed the TD Maxim™ framework³ and configured complex rules removing manual efforts. TD Maxim used for applying filters to get just enough data.</td>
</tr>
<tr>
<td>A Leading U.S.-Based Healthcare Organization</td>
<td>Sensitive protected health information (PHI) was unprotected in the test environment. Inconsistent and effort-intensive methods for generating test data.</td>
<td>Deployed the TD Maxim framework with necessary masking algorithms and policies. The framework was also utilized for synthetic data generation.</td>
</tr>
<tr>
<td>A Global Biopharmaceuticals Giant</td>
<td>A large-scale program facing huge schedule slippage due to unavailability of test data. Simple stubbing ruled out because of the need for high-quality test data.</td>
<td>Deployed the TD Maxim framework for synthetic data generation, which ensures data integrity rules across all interfaces.</td>
</tr>
</tbody>
</table>

Figure 2

Business Outcomes: Case in Point

Figure 2 highlights engagements where we helped IT organizations embrace, deploy and benefit from TDM delivered via a managed services approach.

Looking Ahead

A comprehensive and end-to-end TDM strategy that addresses the multidimensional challenges IT organizations face can deliver multifaceted benefits to enterprises across industries. It can help reduce operational and processing costs, while simultaneously compressing the time needed to achieve the cost savings. It can also eliminate the risk of exposing sensitive data to nonproduction users while also improving test coverage.

TDM, in our experience, furthers better knowledge retention and management with QA organizations, and higher data reuse. Applying a managed services model to TDM provides key benefits to IT organizations, including wider access to niche skills and optimal resource utilization. TDM has a far wider potential than mere provisioning of data for test execution. The advent of big vendors such as Informatica and IBM from the data integration space has only advanced TDM’s ability to deliver large-scale benefits, while shrinking processing windows.
Footnotes

1 Test data management is the discipline of eliciting/understanding the data requirements for testing and delivering relevant test data, meeting technical requirements ahead of the test execution timeline.

2 Our survey was conducted with nearly 250 lines of business that the company serves. Survey participants included a mix of program managers and test managers.

3 TD Maxim is a customizable test data solution framework engineered by us in partnership with Informatica.

About the Author

Gopinath Mandala leads Cognizant’s QE&A Test Data Management Center of Excellence in India. He has 17-plus years of IT experience spanning application design and development, test environment management, test process and automation consulting, program management and enterprise CMMI assessments. Gopi has mentored and transformed test data management processes for large enterprises by leveraging multiple avenues such as efficient test data request provisioning systems, data reuse strategy and test data provisioning tools. He currently leads Cognizant’s Test Data Management Center of Excellence. Gopi can be reached at Gopinath.Mandala@cognizant.com.

Footnotes

1 Test data management is the discipline of eliciting/understanding the data requirements for testing and delivering relevant test data, meeting technical requirements ahead of the test execution timeline.

2 Our survey was conducted with nearly 250 lines of business that the company serves. Survey participants included a mix of program managers and test managers.

3 TD Maxim is a customizable test data solution framework engineered by us in partnership with Informatica.

About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process outsourcing services, dedicated to helping the world’s leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 75 development and delivery centers worldwide and approximately 178,600 employees as of March 31, 2014, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world. Visit us online at www.cognizant.com or follow us on Twitter: Cognizant.