By Hemalatha Murugesan

Test Data Management – Left Shift for Business Agility

One of the biggest challenges most organizations face is ensuring that a product is 'fit for purpose' without compromising on compliance and quality. Project teams and the various stakeholders usually ensure that the requirements are captured with proper design and implementation coupled with a tight quality process. However, we still notice quite a few defects that become obvious once the product has been rolled out into production; too many of them can even result in a rollback of the product.

The fundamental problem is the absence of the right test data to review the functionality at various phases – design, unit testing, system integration, and user acceptance. After the recent economic downturn, unprecedented regulations were brought in across the globe to make an effective test data management (TDM) strategy a must for organizations. This will not only help streamline the processes but also ensure early releases are guaranteed to adhere to the regulations.

Introduction

Fundamental assurance for any product – be it a car, cookie, shirt, or any software application – is gained by getting it thoroughly tested with variants of test data for effective coverage. Often, we see advertisements luring customers to test drive a car, or offering newly launched cookies to get feedback on taste, price, size, etc. In these cases, the product is tested almost in a "live situation", enabling the organization to incorporate the feedback and remediate as necessary. The amount of extensive field testing or clinical trials that are undertaken for the launch of any drug have not only ensured extensive coverage across the globe over many years, but have also helped various international bodies to scrutinize the drugs.

However, in the case of software applications, the absence of the right test data leads to defects that are noticed much later in the testing cycle or post rollout into production, often causing serious issues.

The inevitability

The changing economic landscape has forced financial institutions to reduce costs while adhering to quality and compliance requirements. They are grappling with several challenges such as loss of trust, lack of confidence, increasing competition, terrorism, and growing compliance requirements. In addition, lower returns and constant scrutiny of their businesses have forced the banking industry to look for ways and means to reduce costs and speed up launches.

Some of the key drivers for an effective TDM at the organization level are:

- **Time** – The effort spent in ensuring TDM ranges anywhere between 12 % and 14 %. In some cases, where the applications are data intensive, the effort can exceed 21 %. Significant time and effort are spent in creating the test data manually or in mining it from production dumps to meet various test scenarios. More often, the test data needs to encompass various types of testing – functional, SIT, performance, security, compliance, configuration, etc. Also, the volume of test data to be provided can be huge in certain types of testing.

- **Compliance** – To bring in more soundness, accountability, and operational transparency across the globe, regulations have been stepped up to address systemic risks. Some of the important regulations are BASEL III, Dodd-Frank Act, Markets in Financial Instruments Directive (MiFID) II, Know your Customer (KYC), Gramm-Beach-Bliley Financial Act (GLBA), Payment Card Industry Data Security (PCS DS), SOX (Sarbanes Oxley Act), Australian Prudential Regulation Authority (APRA), New Zealand Privacy Act, Hongkong Personal Data Ordinance, Brown Vitter Bill (new one), Canadian Personal Information Protection and Electronic Documents Act, European Union Data Protection Directive 95/46/EC, FATCA, SEPA, Solvency II, and many more.

- **Standardization** – Standardization is a must-have for smooth coordination of testing. Testing ineffectiveness is an important issue to be dealt with. It can result from inadequate TDM processes, non-estimation of TDM effort, and non-budgeting for the effort. Different data form requests from various teams result in data usage conflicts as well as data contention. Moreover, absence of documented data models and test data repositories can lead to inconsistency, causing non-standardization.

- **Synchronization** – Different project teams request data which are not in sync in terms of formats, request types, etc., where the data is spread across various test environments, applications, and databases. Due to lack of coordination with upstream and downstream applications, the data needed is not synchronized, leading to errors.

- **Cost** – With banks trying to retain customers, new products are launched through various channels – social media, ATMs, online banking, etc. Customers look to utilize these products with ease of transformation which means increased testing with the right data for the product. Multiple production dumps increasing the storage across various environments with no rationalization of licenses and utilization details significantly increases the cost, ballooning operational expenditure (Opex).

TDM – the need?

There is a misconception that test data management and its governance are part of the IT function. Actually, it is a part of the entire organization. Data is the business driver and pivotal for survival and cost efficiency. Organizations are now accepting the fact that TDM and its infrastructure are no longer a cost drain but an important attribute for cost reduction and revenue generation.

Is TDM a functional or non-functional part of requirements?

Traditionally, functional requirements and, to some extent, non-functional requirements were captured from various stakeholders and the business. However, test data requirements were never captured or even discussed, leading to a lot of assumptions and resulting in poor coverage and production issues. Capturing the test data requirements both for functional and non-functional requirements will help organizations to left shift the whole project life cycle with a significant decrease in effort, time, and rework and an increase in quality.

What must organizations do?

Traditionally, organizations used to take a dump of the production and provision the data based on the need for the testing requirements. Today, these organizations must have a clear blueprint for managing their TDM systems for the future. They need to understand their data requirements and how the data is sourced and provisioned. By establishing a robust enterprise, the Test Data governance framework will help identify the various kinds of data usage, its consumption patterns, as well as ownership and accountability of the test data. Organizations can meet the various business requirements with customized golden copies made available to the testing teams through good data management models.

Test data can be created by various means, and each has its own pros and cons depending on usage and business needs. It can be generated by using any one method or a combination of some or all methods. Careful analysis is required prior to laying down the strategy, as it will not only impact testing but will also influence cost and compliance. Some of the methods are:

- Production dump/cut, cloning, subset
- Manually created or synthetic data creation
- Extracted data from migration
- Data injection
- From previous releases
- By automated means
- Engineered data

More and more organizations are adopting synthetic data for testing, given the compliance and data security issues due to the consumption of production data. However, it is important to note that while usage of synthetic data solves many such problems, the data coverage is not the same as the coverage we get by utilizing production data. It is important

to evaluate the pluses and minuses of using synthetic data on a large scale while keeping the business risks in mind.

One efficient way to provision test data is through the creation of golden copies, by optimally sub-setting the production data. Base-lining of the golden copy helps to validate the delta quickly before requesting a refresh of the production data. Test data in the test environment will be provisioned from the golden copy on a needs basis, and this can also be scrambled data. Such mechanisms reduce the risk of legal exposure, because data privatization addresses the various compliance issues as well as restricting data theft.

Detailed analysis needs to be done before the sub-setting process is initiated, as it has a direct impact on the quality and quantity of data extracted. Best practice would be to do a sampling to confirm the rules created. Based on a careful analysis of releases, it would be advisable to create a reuse repository of known patterns of test data requests, the most common types of test data across various applications and testing types, and build meta-data definitions for on-demand test data, etc.

Creating a test data warehouse for frequent data provisioning requests will help the testing teams to utilize the on-demand test data quickly, resulting in huge savings in time and effort, as well as in disk storage. Traceability of test data back to the test cases to business requirements will help address coverage and the impact of the test data on the test case. Also, reservation of test data for project teams for a particular set of test cases/test types will reduce data overstepping and data contention.

Team Models

Previously, teams used to work in silos with no proper communication on the data provisioned, mechanism used, best practices, etc. Based on the organizational structure and its charter, businesses need to deploy centralized, federated, or hybrid models. Each of these models has its own pros and cons. Some of these may require a strong political will and senior management support for successive deployment and implementation of the TDM strategy.

With the explosion of data in the digital age, it is important that firms leverage automation and the use of tools. These TDM tools can be in-house utilities or any of commercial off-the-shelf (COTS) tools. A thorough feasibility analysis exercise would need to be done before finalizing the COTS tool, as it is fundamental to the success of the firm's TDM strategy. A proper governance structure with empowerment and authority is required for the implemented strategy to succeed. A clear RACI – (Responsible, Accountable, Communicated and Informed) matrix needs to be published and sent to stakeholders for effective channelization and coordination among teams.

Metrics

For the success of any strategy, it is important that it is measured periodically and a necessary course of correction is followed as soon as possible. Metrics form an important measure. Depending on the stakeholder(s), the reports and details can be published. Reporting can be at various levels – project, tactical, strategic, or even enterprise. Examples of indicative metrics would be % reduction in test data related to defects, or % of test data coverage achieved.

Data quality attributes mapped to TDM metrics could be:

- Timeliness – test data provisioning turnaround time/SLA adherence completeness
- Validity – test data efficiency, % data request complete with SLA missed/achieved
- Accuracy – test data defects reduction
- Consistency – test data reusability (% of data provisioned that could be used for one or more test data requests), test data productivity
- Integrity – improved end-to-end test data coverage

Conclusion – the way forward

In a highly volatile economic environment, firms across the globe are under tremendous cost and margin pressure. Having the right efficient Test Data Management framework is key to the firm's survival, together with reducing costs and legal exposure. With early involvement of TDM, organizations stand to gain significantly, thereby left shifting for early product launches. ■

> about the author

Hemalatha Murugesan is currently working as a Delivery Manager and leading the specialized testing services in the FSI unit at Infosys. She has been involved in setting up, pioneering, and evolving emerging testing services such as cloud testing, test data management, infrastructure testing, virtualization, and other upcoming specialized services at Infosys. Hemalatha has been instrumental in developing the Enterprise Performance Testing Solutions, which offer performance testing solutions and services, and has also set up the state-of-the-art performance testing lab at Infosys.

Hemalatha has co-authored a book named 'Software Testing: Effective Methods, Tools and Techniques' published by Tata-McGraw-Hill, 2003. She has over 16 years of experience and has published many papers in international conferences as well as conducted tutorials on topics such as testing as a main career stream, testing lifecycle, functional testing and UAT, ROI through performance testing. She has spoken at leading engineering colleges on the benefits of taking up testing as a career. She has presented a talk on performance testing through EduSat – reaching out to over 120 engineering colleges under VTU using the satellite medium. She was also on the technical review panel of the StepIn International conferences.