

View Point



Enhancing Test Effectiveness through Test Data Management

Harish Krishnankutty, Ravi Illimattathil

Abstract

Managing test data in non-production environments is essential since it helps ensure proper validation of business functionality. Ongoing management of test data or rather the lack of it, often causes conflicts among test teams. Some of the typical challenges encountered include lack of specific data requirement for validating particular functionality, test teams lacking necessary skills and access to create back-end data, not knowing who is responsible for creation of particular data, considerable co-ordination effort due to conflicting data requirements, etc. In this point of view we take a closer look at these challenges and how they can be addressed through the use of an effective test data management approach.

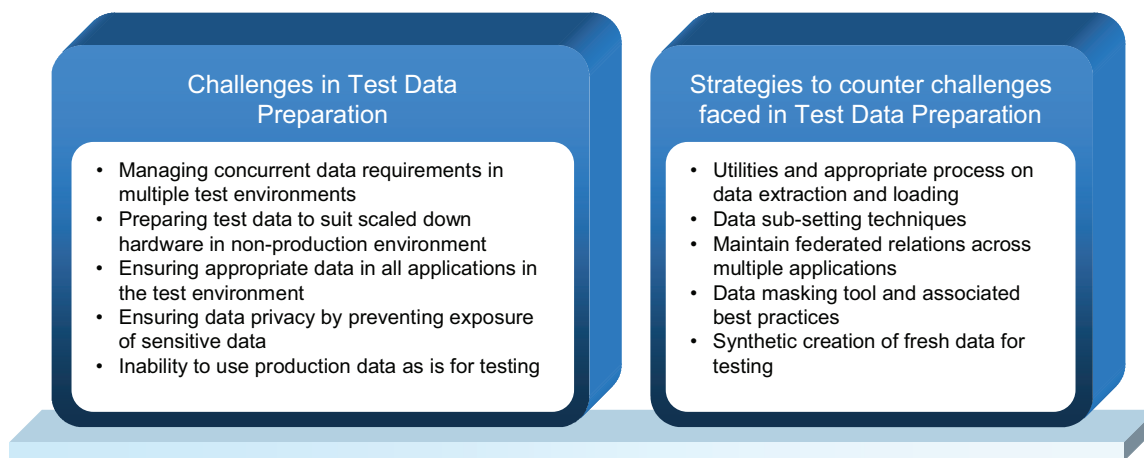
Background

Today, most enterprises are becoming increasingly aware of the value that independent validation brings to their business critical applications. Independent QA teams are brought in to focus on application accuracy, reliability and performance to minimize downtime and disruptions to business. For the QA teams, comprehensive test coverage and early detection of defects are key success factors. A comprehensive test data management strategy is vital in ensuring reliable and comprehensive testing¹. While a comprehensive test data strategy is a good starting point, the success of the testing efforts depends largely on the effectiveness of practices employed in the ongoing management of test data.

Managing test data across the testing lifecycle

An enterprise application typically consists of several interrelated systems hosted on multiple heterogeneous technology platforms. When one of the applications undergoes a change, or when a new system is added to the application landscape, extensive validation and end-to-end testing of business functions is required. Typically, a new functional verification environment would be created and the interrelated applications would be made available in the new environment. Most of the reference data and to some extent some transaction data would also be required in the newly created environment. In such scenarios, testing teams need to address the requirements in both stages of the test data management lifecycle, namely (a) Test Data Setup (Preparation) and (b) Ongoing Test Data Administration.

Today, there are several tools, well-established techniques and easy-to-adopt best practices available to address these challenges from a test data setup standpoint. QA teams having a thorough understanding of data requirements would be able to provide appropriate solutions during the test data setup phase using these tools and techniques



However, the same is not true of ongoing test data administration.

Test Data Administration – Understanding the challenges

Ongoing administration of data in test environments continues to be a major challenge for QA teams even when test data setup has been completed successfully. Teams traditionally face challenges on three major fronts:

- Managing reference data
- Streamlining the data provisioning process and
- Ensuring data quality in test environments

Let us now look at each of these points in greater detail.

¹ Read POV on – “Test Data Management – enabling reliable testing through realistic test data” (<http://infosys.com/offering/IT-services/independent-validation-testing-services/white-papers/pages/index.aspx>)

Managing reference data

It is essential to have well organized reference data which can be used as a consistent source for refreshing the test environments on a need basis. The data in every critical application should be sufficient to validate all the functionalities. Also, care must be taken to maintain the federated relations that exist across applications, in order to validate end-to-end functional flows.

Streamlining the data provisioning process

- **Ensuring data specificity in complex scenarios:** Specific data may be required for validating particular application functionality. Today, requesting for the data and having the data provisioned in test environment is a tedious process.
- **High co-ordination efforts:** It is often challenging to identify the appropriate contacts who have the required access rights and skills to make specific data available for testing. The dependence on individuals and the absence of a workflow based system, further add to this challenge.
- **Inconsistent data request formats and unclear requirements:** In some organizations, test data is managed by a dedicated team of data experts. These experts struggle to process test data requests due to unclear data requirements being passed to them by the testing teams.
- **Lack of traceability between data and test scenarios:** Significant effort is spent in searching for the appropriate data (to check if it already exists) as the test cases are not mapped to the data.
- **Test data overstepping:** It is common to assign specific data for validation of certain functionality. However in an integrated test environment there is always a strong likelihood of this data getting updated by execution of another scenario.
- **Test data depletion:** Test data gets used up in a particular execution cycle and hence may not be available for subsequent cycles of the same test.
- **Ineffective reuse of test data assets:** Test data management queries and techniques are seldom reused, resulting in effort wastage during re-creation of test data.
- **Lack of visibility in to test data management activities:** The test data management tasks are largely limited to the test data teams and are often loosely integrated to overall test management tools and plans.

Ensuring Data quality in test environments

Due to several factors, such as execution of negative scenarios, application errors etc, there is a strong possibility that as testing progresses, some inconsistencies may creep into the test environment. A set of data quality analysis techniques would help uncover these issues in time and hence ensure high quality of test execution.

An Effective Test Data Administration Approach

The challenges in test data administration call for an approach that would reduce the test data management efforts considerably by ensuring a smooth, integrated and well defined workflow. Let us now discuss such an approach and a process (see figure 1) that would help address the aforementioned challenges.

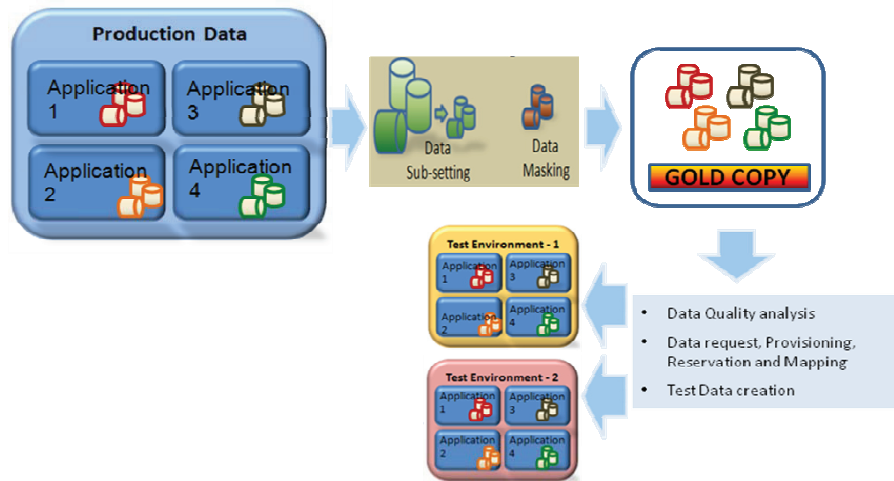


Figure 1: Effective Test Data Administration Approach

Maintain a reference database

It may not be a good idea to use data preparation activities to extract data as and when needed. Instead, a reference repository of test data referred to as a “Gold Copy” can be maintained. As and when the need arises, the test environments can be refreshed using this gold copy. A gold copy can be a generic database like an Oracle or SQL server, where different schemas can be maintained for different applications. Data can then be extracted from production environment after applying appropriate masking/sub-setting methods. In case of extremely sensitive data, fresh data can be created using the gold copy data base.

Implementing a workflow based request provisioning system

A workflow based request provisioning system provides a common interface for placing the test data requests and data management activities. This ensures request management, automated provisioning, mapping data requests to test cases and optional reservation of data. Internally the solution follows a modular structure for organizing data management artifacts so that re-use of techniques can be maximized. Well defined data templates ensure consistent format for data requests. The solution should maintain federated data relations and should ensure they are adhered to while provisioning. Further, it is important for the solution to provide adequate visibility in to test data management activities, through a centralized test management tool.

Conduct periodic Data Quality Analysis

Periodic data quality analysis of test environments is essential as it helps ensure high quality of data and hence high quality of application validation. Several data inconsistencies such as metadata issues, business rule inconsistencies and pattern mismatches can be uncovered by analyzing error prone candidates in test environments. Certain statistical analysis techniques can be used to discover broad level inconsistencies and outliers.

Conclusion

The effectiveness of test data management is critical in ensuring the success of independent validation efforts. QA teams having a thorough understanding of data requirements and tools and techniques in test data preparation would be able to ensure test data is appropriately set-up in the test environments. The challenges in the ongoing administration of test data can be addressed through a data provisioning workflow, focused management and data quality checks in test environments. Effective test data management through well defined processes and appropriate usage of tools helps in achieving better test coverage, optimizing test efforts and consequently lowering cost of quality significantly.



For more information, contact askus@infosys.com

About Infosys

Many of the world's most successful organizations rely on Infosys to deliver measurable business value. Infosys provides business consulting, technology, engineering and outsourcing services to help clients in over 30 countries build tomorrow's enterprise.

For more information about Infosys (NASDAQ:INFY), visit www.infosys.com.