

White Paper



Test Data Management

Enabling reliable testing through realistic test data

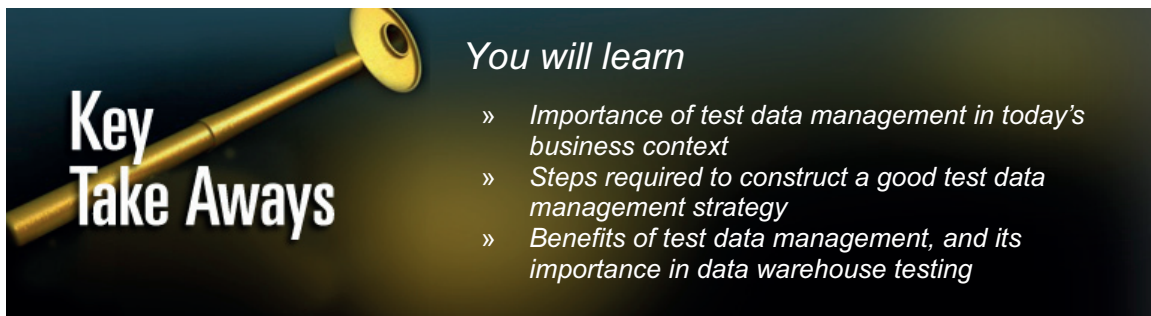
K. Krishna Murthy, Janardhana S Channagiri

Abstract

The ever increasing complexity and scale of IT applications has made quality and reliability of paramount importance. However, delivering the same through testing is becoming a challenge. Test teams not only have to follow exact test methodologies, but also ensure accuracy of test data. They also need to ensure tests correctly reflect production situations, both functionally and technically.

A well rounded Test Data Management (TDM) strategy can help. A well implemented TDM strategy can rapidly reduce inefficiencies, help extract greater value from expensive data and make validated test data available in an organized, secure, consistent and controlled manner. By using the right test data and test data management strategy, QA teams can get one step closer to reliable testing.

This paper addresses the importance of test data management in today's world, steps involved in the construction of a good test data management strategy, and the benefits associated with one.



High quality and reliability of enterprise applications and systems has become a must in today's wireless and digital age. Testing plays a major role in ensuring that enterprise applications meet those standards. However, studies clearly show that, up to 60 percent* of application development and testing time is devoted to data-related tasks, making it cumbersome and time consuming¹. For organizations to speed up testing, and control costs, they need well captured test data requirements and the right Test Data Management (TDM) strategy. Without these in place, the entire process of following the exact methodologies of test planning, scripting and automation will be lost.

Further, the lack of the right test data and a TDM strategy poses a major challenge from an end-user perspective. In a number of cases, system functionalities are not adequately tested, during continuous enhancements, due to required test data not being available or created. This becomes a major challenge, especially when gaining the confidence of end users on the quality of systems and applications delivered.

It has been proven that a comprehensive TDM strategy, which accurately reflects end-to-end business processes, is the winning solution for enterprises. By using the right TDM strategy, QA organizations are able to provide a consistent and standardized approach to testing applications and data from multiple databases and file types. With the right test data and TDM strategy in place, implementers of applications, systems and Data Warehouses (DW) can detect defects early in the process, leading to cheaper and easier ways to fix them. Hence, having the right TDM strategy is essential for QA organizations to improve quality, accelerate release cycles of systems and cut costs aggressively.

A good Test Data Management (TDM) strategy covers the following steps:

- **Cloning of production data on test environments (also referred to as production data migration to test environment):** This helps in defining the different test environments in which the production data needs to be cloned or migrated. The size of test environments is determined by the purpose and strategic usage of those environments. This step also involves the process of making a decision of how much production data needs to be cloned and into which test environment. For example: A final pre-production environment would require full cloning, whereas a system test environment would need only 60% of the data. Defining of periodicity of migration is also done during this stage. This depends on the amount of change in production data, and amount of application changes normally implemented in a half yearly cycle, typically six or nine months for major releases.
- **Creating additional data on test environments:** Depending on the type of test environment and the type of testing required (performance, functional or data warehouse), new data may have to be created, beyond the migrated production data. This strategy should depend on:
 - Exceptions under which projects can create new data
 - Type of tooling required for data creation
 - Necessary approval mechanisms to enable the same
- **Cleaning up test environments post testing completion:** In order to test all needed scenarios, test data may have to be modified in certain cases, leading to the existence of altered test data at the end of project testing. A clear methodology and approach needs to be in place, with set guidelines, to ensure how and when the cleaning up of test data needs to be done, post testing completion. Based on the prescribed data integrity guidelines, there will be instances when the altered test data cannot be, or should not be, cleaned up.

¹ source - http://www.compuware.com/resources/mainframe/Fileaid_clientserver.pdf

- **Managing data requirements of conflicting projects:** IT managers always grapple with conflicting project priorities for the test environment. In order to avoid delays and soaring costs, a clear guideline and mechanism needs to be in place to decide which project should go for testing first, based on data dimension parameters.
- **Controlling provisioning of data to required projects:** A smart data provisioning mechanism can solve this problem effectively. The control mechanism should be applicable for both newly created data and production migrated data. The mechanism can be as simple as blocking a set of customers for a specific project or as complex as sampling based on multiple dimensions.
- **Data privacy and protection:** As soon as it is decided that data needs to be cloned, concerns regarding data privacy need to be addressed. The mechanism, tooling and the stage of cloning process, at which depersonalization would be implemented, should be clearly defined. There are several tools available in the market; however the scale and volume of test data should determine the mechanism to be adopted. In the scenario of testing being outsourced or offshored, it is essential that the organization, along with service provider, jointly decide the data privacy mechanism, rather than treating it as an afterthought or individual priority.

With the right TDM strategy in place, test teams can easily analyze, compare and learn from past data errors. TDM helps to:

- Create “right-sized” test databases that accurately reflect end-to-end business processes
- Correct defects early in the development process, when they are cheapest and easiest to fix; improving quality and accelerating release cycles.
- Execute comprehensive non-functional testing to obtain end-user confidence on production performance and scalability

Why TDM in Data Warehouse Testing?

Data loaded into an organization’s Data Warehouse (DW) comes from multiple sources and dimensions. The analysis extracted from this Data Warehouse, supports an organization’s strategic planning, decision making and performance monitoring systems. Hence, in Data Warehouses, the cost of defects is compounded because of the additional business costs of using incorrect data to make mission critical business or strategic planning decisions. Due to this, there is an important need for testing Data Warehouses from a data completeness, data transformation, quality, security, performance, scalability, integration and user acceptance or confidence standpoint. These tests are essential for the successful implementation or enhancement of an organization’s Data Warehouse.

In order to enforce the best practices of testing for Data Warehouse it is important to use the correct test data and environment. The data used for every regression test cycle must be correct, clean and secure. In recent years, Data Warehouse implementations have increased in size, ranging up to terabytes of data. This means that the test data required for such an initiative will have to be significantly large. Unless there is a way to create or generate large amounts of test data, and at the same time manage it well, doing any kind of testing will not add value, or guarantee the overall quality and success of a Data Warehouse deployment. Not having a proper TDM strategy in place, will not only result in exorbitant testing costs, but will also make the entire process of testing very cumbersome. The process will lack quality, and this will reflect in the low confidence a business shows in the applications/systems.

Test Data Management helps create an environment that can simulate and enable the proper and secure cloning of production data. The TDM strategy helps in maintaining secure and well integrated test data quality, over repeated test cycles. This enhances overall quality of data warehouses testing, bolstering business confidence in making critical decisions using the models generated from the data.

From an overall business impact perspective, a good TDM strategy ensures a faster, better and cost effective product or solution for the client. Furthermore, it gives the client IT management control over budgets by keeping system capacity low.

To summarize, the increasing need to handle higher test data volumes, with a wider range of data types, at various geographical locations, has made TDM a key component to ensuring high system and application quality and reliability. Advanced development techniques, powerful analysis tools and reduced development times, demand a TDM solution that helps rapidly reduces inefficiencies, extract greater value from expensive data and make validated data available in an organized, secure, consistent and controlled manner. It enhances system performance and availability, making reliable applications, from reliable testing and realistic test data, a reality.

About Authors:

[K. Krishna Murthy \(Krishnamurthy@infosys.com\)](mailto:Krishnamurthy@infosys.com) is a Business Manager at Infosys Ltd. He has over 20 years of testing experience. He has managed and directed large test organizations in Finance, Insurance, Healthcare, Hospitality and Manufacturing sectors.

[Janardhana S Channagiri \(Janardhana_c@infosys.com\)](mailto:Janardhana_c@infosys.com) is a Group Project Manager at Infosys Ltd. He has 14 years of IT experience across different spheres of IT. He has led many large testing programs, especially in the Banking sector.



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.