



Test Data Management Trends



TABLE OF CONTENTS

1.	Introduction	1
2.	Challenges in Test Data Management	2
3.	Emerging Trends in Test Data Management	3
4.	Conclusion	4



Introduction

As we ride the digital transformation wave, the volume of data that is being managed has risen exponentially. Achieving the right test coverage to meet the different permutations and combinations of your consumer profile while ensuring quality, security and agility have become critical in this competitive digital market. The characteristics of test data requirements have also had to evolve at a rapid pace to meet the software testing demands. IT needs test data to be available, flexible, reusable, linkable, and customizable for delivering at pace.

Test data management (TDM), also known as software test data management, is the planning, designing, storing, provisioning, and managing of test data to be used in testing software or products. This paper will look into some of the challenges in managing your test data sets and view the current trends in TDM that address these challenges.



Challenges in Test Data Management

Understanding your Test data requirements before it is too late :

A frequent observation is the absence of a test data management strategy. Data availability is mainly based on assumptions that production data would suffice subsetting, and hand cranking can be done to meet specific scenarios. This is usually driven by previous experience and the common "what could be better than production?". However, times have changed, and getting the required data subset from production is not as easy as it used to be. Data must ensure coverage, compliance, controllability, and customizability. Settling for anything else may end up in a billion-dollar software glitch payout or non-compliance lawsuit due to lack of test coverage.

Managing heterogeneous data sources :

Data source types are being created at a rapid pace to meet the demand for speed to market. Test data needs to talk to your applications as well as across applications. Your data journey in an end-to-end flow would involve a combination of data types/logic/permutation and in large volumes. With the various types of test data that must be managed, one tool may not answer all your data needs.

Organizations tend to have a narrow approach to meeting their test data requirements. This tends to lead to settlements (production subsets) or workarounds (excel sheets, stubs), which adversely impact the referential integrity of the data provisioned to test environments for functional and non-functional testing.

Centralized teams can sometimes deviate and slow you down:

Organizations tackle their test data problem by transferring the test data management function ownership with a centralized team independent of Agile Sprint and DevOps teams. This does work in some cases with the right approach and team size. However, we usually see elongated data provisioning cycles that are SLA-driven due to the large volume of test data requests. We also know a misalignment with the Agile & DevOps teams that cannot benefit from continuous integration and testing practices.

Test data platform complexity and lack of expertise:

Most data management platforms available in the market require specialized training and skilled resources. The scarcity of test data management expertise within the software testing community is another challenge. Also, the TDM tools available in the market do not provide all the required capabilities. As a result, organizations have multiple tools for masking, data virtualization, and synthetic data generation.

Poor orchestration of test data with automation tests:

The onset of Agile and DevOps, the glue that integrates test data with the scripts and continuous integration and deployment tools, is not always applied. Also, the test data consumption rates associated with automation scripts are high as one may need to execute the same scripts across multiple browsers and device combinations. Many organizations cannot realize the real benefits of test automation, as data provisioning is a process that requires manual intervention. The integration of test data provisioning is often carried out using custom scripts or solutions and often restricts the ability to provide data as part of the DevOps pipeline. Most data management tools available in the market do not offer plugins or APIs that support integrating tools with automation, performance engineering, or DevOps.

Adherence to data regulations :

Implementing the General Data Protection Regulation was an example of how secure and prepared (or lack there of) we were in handling data. It exposed the void in test data management practices and raised its relevance in the IT industry manifolds. Organizations need to also comply with other regulatory standards like Gramm-Leach-Bliley Financial Act (GLBA), Payment Card Industry Data Security Standard (PCI-DSS) when dealing with test data. Organizations still adopt workarounds and loopholes to manage their data which is at risk of non-compliance mainly due to the initial operational costs for building an effective TDM solution. Organizations need to have control of the data they are handling. The ongoing discussion with social media platforms on their handling of consumer data is a clear sign of a new slew of compliance rules to come.





Emerging Trends in Test Data Management

In this section, we will take a few of the emerging approaches to implement TDM solutions. These approaches have proven effective in addressing the challenges mentioned in the previous section.

Principle-based TDM practices

Test data management needs to be viewed as an entity in your quality engineering vision. By placing it as an entity, you ensure that it is given due relevance and percolates to the bottom layers of your organization. Below is a snapshot of how the TDM Entity based framework would reside in your QE strategy.

Entity	Attribute	Guiding Principal
Test Data Management	Available	Test data must be available / provisioned on-demand to support any test requirement
	Compliant	Test data must be compliant with data and compliance standard
	Customizable	Capability to modify test data to create data for edge case scenarios
	Coherent	Test data can be integrated into automated tests and utilized during execution either statically or dynamically
	Sustainable	Test data patterns are recorded to create reusable test data sets. Any data that is consumed or obsolete is archived, removed, or refreshed

The organization must ask themselves whether their TDM solutions ensure quality test data by adhering to the attributes and their guiding principles.

Implementing intelligent data generation

New trends in test data generation involve the use of synthetic data generation tools. These tools can generate test data in multiple formats based on the data structure provided to them. As the name suggests, the data is synthetic and does not contain any confidential information, making our test data more compliant with various industry standards.

Standard tools can generate data that can be random, goal-oriented, or pathwise. In addition, test data generation tools have also evolved during the Agile and DevOps transformation journey to intelligent automated solutions that identify patterns in your production data and can automatically generate synthetic data copying the same structure as production and maintaining the referential integrity in the synthetic information as well.

Present-day intelligent data generation tools are modular and communicate via external APIs. This enables them to integrate with automation frameworks as well as DevOps

pipelines. These solutions are adopted across different data types (structured, unstructured), formats (flat file, XML, pdf, etc.) and can generate data permutations. In addition, the user-friendly interface and drag and drop capabilities of these tools make them easy to learn and removes any specific skill requirements like writing complex Jobs for extraction and masking of data.

Centralized teams offering de-centralized capabilities

Centralized TDM teams working as shared services would ideally be the best answer for having a focused strategy for Test data. However, effective centralization is rarely seen. The current emerging trend for TDM support is a mix of centralized team and data-on-demand (DoD) portals. The centralized team will be a team of experts focusing on strategic and innovative data creation and supporting any issues with the DoD portals.

The data-on-demand portals may be internal web portals to specify the data to be created/extracted. The engine behind the on-demand portal may be intelligent test data generators or production subsets, or test data lakes. Certain organizations have gone further and given their test team the flexibility to create customized test data via the DoD portals.

Building prefabricated data sets

Test data should be sustainable; though the test data may be burnt, new test data in the same pattern may be required in future testing phases. Hence the approach for reusing your test data is very critical to have an effective TDM solution. An established practice in this regard is the creation of prefabricated test data suites which are usually based on domains that are relatively stable in the data structure and format. The key advantages of this approach are the time that it saves for test data creation and the identification of any domain-based errors in the system.

The most recent trend has been using Virtual Assistants to interact with the users to extract data from prefabricated test data sets after specifying the requirements. This approach cuts down any delays in data provisioning and allows the testing team to focus on what they are good at. Prefabricated data may not expose all your problems; hence it is always safer to go with a combination of prefabricated and internal data sets.



Implementing data virtualization solutions

The testing team may require multiple versions of test data to be available to test different application versions. If the data masking approach is practiced, one may store terra bytes of data, resulting in massive data storage requirements. Also, refreshing test data from production and keeping multiple copies updates is a time and effort-intensive activity.

By applying data virtualization capabilities, one can create numerous virtual clones of a database and can use them in the same manner as a physical database. The virtual copies can be continually synchronized with the source database and quickly provisioned to a test or development environment. As the data is stored in a compressed manner, data storage costs less, and it is easy to roll back the database to a previous state. Some of the pioneers in this space include Delphix, Actifio, NetApp, amongst others.

Conclusion

Quality engineering cannot be delivered without an effective test data solution and strategy. The right solution for your organization may be a combination of multiple tools and approaches. The evolution of TDM solutions is making rapid strides for matching pace with the speed of delivery that agile demands. These solutions are actively used for generating bulk data to support the testing of AI/ML engines. To conclude, an effective TDM solution will ensure high-quality data is available, compliant, coherent, sustainable, and customisable.



George Ukkuru

Technology Principal - QE & Automation
Marlabs



About Marlabs

Marlabs designs and develops advanced digital solutions that help its clients improve business outcomes swiftly and precisely. It succeeds by harnessing the power of the Digital Collective™, which brings together design-led digital innovation with human experience, composable digital platforms, and a collaborative ecosystem of first-class technology partners and innovators.

Marlabs is headquartered in New Jersey, with offices in the US, Germany, Brazil and India. Its 2500+ global workforce includes highly experienced technology, platform, and industry specialists from the world's leading technical universities.

Marlabs Inc.(Global Headquarters) One Corporate Place South, 3rd Floor, Piscataway, NJ - 08854-6116,
Tel: +1 (732) 694 1000 Fax: +1 (732) 465 0100, Email: contact@marlabs.com

