ACCELARIO

Tech Highlight | March 2023

Reinventing TDM for DevOps

Introduction

Test data management is not a new practice; in fact, most enterprise QA organizations include a centralized TDM team, responsible for providing test datasets for various environments. Legacy implementations of this practice combine tools that subset, mask, and synthesize test data with a ticket request-based process of data preparation, anonymization, data delivery, and data refreshes. Given the centralized and manual nature of this process, software teams that aspire to increase delivery velocity and, specifically teams that adopt automated CI/CD and DevOps frameworks, often find that the legacy TDM solutions fail to support their needs.

Modern DevOps teams can't afford to wait days and sometimes weeks before a dataset is delivered into their pipeline; they need a solution that can deliver datasets in minutes, in an automated API-based fashion that integrates quickly and easily into the CI/CD pipelines.

Test Data Management solutions aim to provide test datasets based on production-quality data into QA environments. In addition, the delivered datasets must comply with data privacy regulations to ensure that sensitive data isn't copied into non-production systems. The test datasets are typically carved out of very large production data sources in a relatively complex process to ensure that data objects are copied over together with their entity relationships - otherwise the business process integrity will be compromised. For example, if a test requires customer account records, the carving process needs to ensure that all relevant fields related to the customer object such as customer transactions or contacts are also copied over, so they will be part of the dataset during the test,leading to a recursive like process per record. Because the source databases are typically very large, this subset process is used to shorten the time and space it takes to deliver these dataset copies.

Traditional TDM solutions fail to support the modern DevOps data needs

Traditional TDM practices rely on a ticket-driven request-fulfill model in which different data consumers submit requests for datasets that are served in a queue-like fashion. Each request requires a lengthy data preparation process, which typically includes:

1. Mapping datasets to test requirements.
2. Recursively searches for all relevant entity objects.
3. Applying subsetting algorithms so the data will support business process integrity.
4. Identifying sensitive data fields and applying anonymization while maintaining referential integrity.
5. Delivering the size-reduced DB copies to the data consuming teams.

This complex data preparation process is needed because traditional TDM tools don't offer a practical way to copy the entire production database, due to its typical large size. This limitation leads to a solution that is:

- Time consuming - most TDM ticket requests get served in a matter of days or even weeks, which is entirely too long to meet the response times demanded of modern DevOps teams
- Complex - the required data preparation process, including both the entity copy and the sensitive data anonymization, lead to a complex process that delays the initial delivery and also limits flexibility in refreshing the test datasets
- **DBA and SME centric** - the need to map the business process with the carved out data requires tight collaboration between DBAs and application SMEs. This creates a high dependency on central teams, limiting the ability to deliver datasets in a self-service manner
- **Storage intensive** - even with reduced size DBs, the need to deliver multiple copies of the DBs can lead to an increased storage footprint, increasing storage costs dramatically
- **Fragile** - the high dependency of the dataset on accurate mapping to the business processes as well as the entity relationships and referential integrity lead to a very fragile process; potential data quality issues surface with every small change or refresh to the dataset
- **Expensive** - the high-touch nature of the solution results in a very expensive operation in headcount, tools, and storage
- **Offers partial test coverage** - finally, even with a very well thought out mapping of business processes to datasets, the delivered testing datasets offer only partial representation of the production data, reducing test coverage and leading to quality risk.

As a result, many of the DevOps benefits related to release velocity and software quality remain constrained by slow, stale, and low-quality test data. To address modern challenges, DevOps teams must expand beyond legacy TDM practices to automate database delivery to CI/CD pipelines via unified APIs or a self-service process.

The Modern DevOps TDM Alternative

It is time for DevOps, DBAs, and infrastructure teams to consider an alternative approach to legacy TDM, one in which test database copies can be delivered in minutes without wasting time on complex upfront data preparation processes or incurring high storage costs. This approach is possible with the advancement in virtual database technologies. Virtual databases are full read/write replicas of source databases created in minutes, which take up only a fraction of the storage space consumed by the original source database. The ability to deliver full copies of source databases into lower environments eliminates the biggest bottleneck of the traditional TDM approach, the upfront data preparation process. With this modern approach, business processes don't need to be mapped to datasets, and recursively searching for all related entity records of an object can be eliminated, since the entire database with all record dependencies is copied over, ensuring that all test cases have the data they need to complete their scenarios.
The use of virtual databases as TDM enablers result in a faster solution that can be fully automated and delivered in a self-service fashion.

> By delivering full read/write virtual copies of databases, Accelario solves the biggest TDM challenge: delivering production quality data in a fast, compliant and storage-efficient way

Deliver Full-size, Production-quality, Compliant Test Data in Minutes Not Weeks

Accelario delivers a TDM solution that operates at the speed of DevOps. With Accelario, test data delivery can be controlled via APIs or through a self-service web portal, delivered on premise or across multi clouds. Test data can be continuously refreshed, cloned, versioned, and rolled back in time as needed and at scale. This opens up dev's and QA's ability to perform more exhaustive test scenarios without being blocked by shared database conflicts or delayed by lengthy data preparation-centric processes. CISOs and data administrators can ensure that all test data is free from sensitive information, as Accelario has integrated data privacy search and masking capabilities as part of its TDM platform.

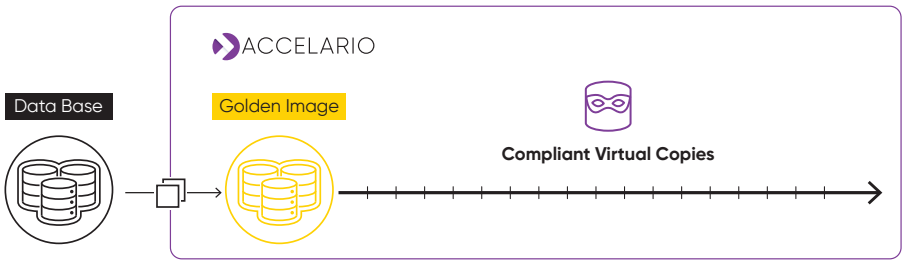

Diagram 1 | Many lightweight virtual replicas of test databases are created in minutes, with data privacy built in.

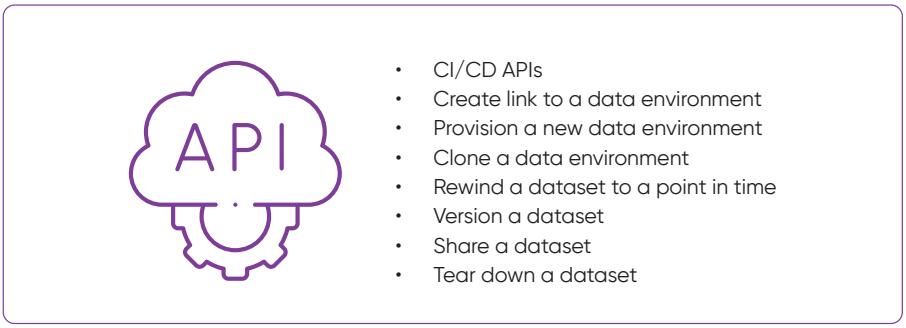

Diagram 2 | All test database operations can be automated via a RESTful API or through a web-based portal.

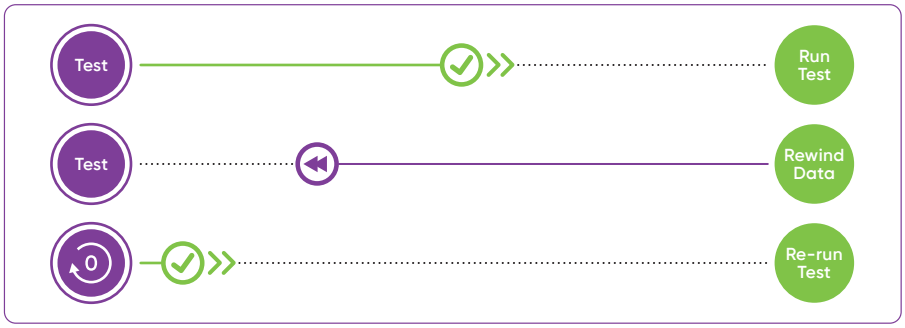

Diagram 3 | Test datasets can be rewound at the end of each test to provide the same data state in each test run.

By leveraging virtual databases that can deliver full read/write copies of the source databases in minutes, teams can leverage a TDM solution that offers significant benefits compared to the traditional legacy approaches to test data.

- **Fast test data delivery** - Accelario's TDM solution delivers vDB copies in minutes even if the source database exceeds TBs of storage
- **Privacy compliant data** - by integrating sensitive data search with anonymization and masking algorithms, Accelario ensures that test data is free from private or sensitive data
- **Self service** - Dev, QA, and UAT teams no longer have to rely on centralized TDM teams or DBAs to prepare and deliver test data. With Accelario, test data can be delivered via a self-service portal or as an integrated flow of the CI/CD pipeline via APIs

- **Storage efficient** - with Accelario, test data is stored within virtual DBs that don't take up additional storage as more copies are created, significantly reducing one of the highest cost factors in TDM, storage
- **Stable** - another key benefit of leveraging full copies of the source database and eliminating the need to subset and prepare data in advance is in the stability it provides to the ongoing use of the solution. When there is no risk to entity or referential integrity, there is also no risk that any future changes will break the environment
- **Full test coverage** - with the ability to deliver full copies of production DBs, teams no longer have to choose between data subsets; tests can now access the entire set of records in the database, ensuring 100% test coverage
- **Cost effective** - the combination of self service and storage efficiency results in the most cost-effective TDM solution in the market

Simply put, leveraging virtualized databases for TDM speeds time to market, reduces workloads on DBAs and SMEs, and increases the efficiency of the DevOps processes. Get in touch with our team at Accelario to learn more.

About **Accelario**

Data is one of the last real DevOps bottlenecks. Our mission is to provide organizations with fast access to high quality data, so that they can embrace agile processes that enable them to innovate quickly and deliver even faster.

Accelario is revolutionizing the way organizations move, access, use and manage data with a suite of simple-to-use, self-service solutions and modules that accelerate and simplify complex IT processes. Our flagship product–Accelario Continuous DataOps Platform– delivers access to critical data, in real-time and across any environment, while ensuring that regulatory and privacy compliance is maintained.

For more information or to schedule a demo of the **Accelario Platform for DevOps** go to **www.accelario.com**. Follow us on LinkedIn, Twitter, and Facebook.