

Test Data Manager

AI-Driven, Compliant Data for Accelerated, High-Fidelity Releases

At a Glance

TDM provides a comprehensive solution for the automated provisioning, masking, and generation of high-quality synthetic data across diverse enterprise environments. By streamlining data management workflows, the platform enables organizations to accelerate software delivery cycles while ensuring strict adherence to global data privacy and compliance standards. With expanded capabilities to support agentic data generation and database virtualization, TDM allows teams to leverage AI for data generation for complex schemas and instant, storage-efficient data provisioning.

Key Benefits

- Improve application quality.
- Deliver data faster.
- Increase testing efficiency.
- Reduce infrastructure costs.
- Ensure zero-trust compliance.
- Accelerate GenAI initiatives.

Key Features

- **PII Discovery and Scalable Data Masking:** Automatically identifies sensitive information and applies consistent, format-preserving masking that scales horizontally to process billions of rows across relational, NoSQL, and mainframe databases.
- **High-Fidelity Synthetic Data Generation:** Creates production-ready synthetic data that maintains statistical integrity and complex relationships without using real sensitive information.
- **Database Virtualization Engine (DAVE):** Provisions instant, storage-efficient virtual database clones that allow developers to snapshot, test, and rollback environments in seconds.

Overview

The reliability of modern AI-driven and data-intensive applications depends on continuous validation against high-fidelity datasets that accurately mirror complex, real-world production environments. As teams transition to Agentic AI and DataOps workflows, traditional manual data provisioning has become a primary bottleneck, struggling to keep pace with real-time data lakes and non-deterministic AI testing requirements.

Broadcom[®] Test Data Manager (TDM) solves these challenges by automating the entire data lifecycle; delivering secure, fit-for-purpose datasets through database virtualization and AI enabled synthetic generation. By eliminating data wait times and ensuring utmost privacy compliance, TDM shortens release cycles from weeks to days, allowing organizations to scale their AI and cloud initiatives with absolute confidence.

Business Challenges

Organizations are under relentless pressure to deliver feature-rich, AI-enabled applications at breakneck speed without compromising on reliability. Despite the shift toward modern architectures, many teams remain tethered to legacy data practices and manual workflows, which creates several critical issues:

- **Inadequate coverage:** Modern applications, especially those utilizing LLMs and generative AI, require diverse datasets to validate unpredictable outputs. Production data is not only expensive to replicate across massive cloud data lakes, but it often lacks the “what-if” scenarios, negative edge cases, and synthetic variations needed to stress-test AI models, leading to costly failures in production.
- **Data bottlenecks:** The sheer scale of modern data environments makes locating, subsetting, and copying data a slow, error-prone process. Developers and data scientists frequently face data starvation, where testing and model training are stalled by manual provisioning, directly threatening sprint commitments and slowing down DataOps pipelines.
- **Zero-trust and compliance friction:** With the rise of global AI regulations and strict privacy mandates like GDPR, using raw production data is no longer viable. Manually masking sensitive PII across interconnected, heterogeneous platforms is insufficient; without automated, persistent masking and agentic data generation, organizations face heightened exposure risks and the inability to prove compliance during rapid release cycles.
- **Operational dependency:** In many enterprises, data provisioning remains a highly specialized, centralized task reserved for DBAs or TDM engineers. This creates a chronic request-and-wait culture where developers are sidelined for days waiting for a specific dataset. Without a transition to an API-first, self-service model, the lack of data democratization prevents teams from achieving the true speed of modern DevOps, regardless of how fast their code is written.

These challenges create a fundamental disconnect between the speed of development and the availability of secure, high-fidelity data, making it difficult to scale automation and meet the demands of a data-first economy.

Key Features (cont.)

- **AI-Enhanced Data Provisioning:** Leverage your own AI models to build comprehensive data models and generate high fidelity data.
- **PII Audit and Compliance:** Generates detailed audit reports and visual heat maps that identify sensitive data risks, providing an immutable trail of compliance for regulations such as GDPR, HIPAA, and CCPA.
- **Self-Service Data Provisioning:** Empowers testers to independently find, reserve, and provision specific datasets through a centralized portal, eliminating DBA bottlenecks.
- **Mainframe-Native Data Management:** Executes complex data engineering and masking directly on z/OS for DB2, IMS, and VSAM, ensuring sensitive data never leaves its secure native environment.

Solution Overview

TDM addresses modern development challenges by providing a unified, AI augmented platform for the discovery, creation, security, and instant provisioning of high-fidelity test data. The platform moves beyond a reliance on raw production copies by allowing teams to model exact data requirements and leverage agentic data generation to bridge coverage gaps with rich synthetic datasets. Through DAVE and self-service portals, teams can instantly locate, reserve, and provision complete, versioned environments, enabling multiple teams to work in parallel without data conflicts.

The platform's modern, scalable architecture supports massive enterprise data volumes by utilizing horizontally auto-scaling masking containers to process billions of rows across heterogeneous sources in just a few hours. This ensures consistent, repeatable test environments across mainframe, cloud, and distributed systems while protecting sensitive information through automated PII discovery and in-motion masking. By integrating seamlessly into DevOps toolchains through APIs and frameworks like Zowe, TDM accelerates the delivery of compliant, high-quality software.

Key Features

- **PII Discovery and Scalable Data Masking:** Automatically identifies sensitive information and applies consistent, format-preserving masking that scales horizontally to process billions of rows across relational, NoSQL, and mainframe databases.
- **High-Fidelity Synthetic Data Generation:** Creates production-ready synthetic data that maintains statistical integrity and complex relationships without using real sensitive information.
- **DAVE:** Provisions instant, storage-efficient virtual database clones that allow developers to snapshot, test, and rollback environments in seconds.
- **AI-Enhanced Data Provisioning:** Leverage your own AI models (BYOA) to build comprehensive data models and generate high fidelity data.
- **PII Audit and Compliance:** Generates detailed audit reports and visual heat maps that identify sensitive data risks, providing an immutable trail of compliance for regulations such as GDPR, HIPAA, and CCPA.
- **Self-Service Data Provisioning:** Empowers testers to independently find, reserve, and provision specific datasets through a centralized portal, eliminating DBA bottlenecks.
- **Mainframe-Native Data Management:** Executes complex data engineering and masking directly on z/OS for DB2, IMS, and VSAM, ensuring sensitive data never leaves its secure native environment.

Key Benefits

- **Improve application quality:** Ensure comprehensive coverage by using high-fidelity, targeted data that accounts for real-world, negative, and non-deterministic AI edge cases, significantly reducing escaped defects.
- **Deliver data faster:** Accelerate DevOps and CI/CD cadences by providing developers and testers with instant, self-service access to production-quality data via the Database Virtualization Engine (DAVE).
- **Increase testing efficiency:** Eliminate environment contention and manual bottlenecks by automating the end-to-end data lifecycle, enabling parallel testing and shortening release cycles from weeks to days.
- **Reduce infrastructure costs:** Drastically lower storage and MIPS expenditures by replacing massive physical data clones with virtualized, storage-efficient datasets that scale across mainframe and cloud.
- **Ensure zero-trust compliance:** Protect sensitive PII with automated, persistent masking and generate immutable audit reports to demonstrate effortless compliance with evolving mandates like GDPR, HIPAA, and CCPA.
- **Accelerate GenAI initiatives:** Leverage AI augmented workflows to autonomously discover complex relationships and generate synthetic data rules, reducing manual configuration efforts by up to 90% for modern AI workloads.

Critical Differentiators

TDM delivers unique capabilities that address the most complex enterprise data requirements through advanced automation, mainframe integration, and AI innovation:

- **Massive horizontal scalability:** Secure terabytes of data at enterprise scale by leveraging Kubernetes to horizontally scale masking operations, enabling organizations to process billions of rows in hours through automated resource management.
- **AI-driven data intelligence:** Utilize *Bring Your Own AI* (BYOAI) models to automate synthetic rule generation and perform adaptive subsetting, reducing manual configuration efforts through advanced machine learning and natural language interfaces.
- **Storage-efficient data virtualization:** Use the DAVE to create multiple, production-like virtual environments instantly, accelerating defect reproduction and AI model testing without increasing physical storage requirements.

- **Broad-spectrum data management:** Seamlessly orchestrate data discovery, privacy, and synthesis across your entire ecosystem with unified capabilities to scan, mask, audit, and generate data across diverse environments; from legacy mainframes and relational databases to modern NoSQL clusters and unstructured file systems.
- **Advanced NIST-compliant privacy:** Implement unified PII discovery and NIST-compliant Format-Preserving Encryption (FPE) to ensure sensitive fields remain functionally useful for testing while staying structurally accurate and fully protected across all environments.
- **Mainframe-native management:** Perform data management tasks directly on the mainframe using native engines, eliminating the need for slow, high-risk ETL processes and ensuring sensitive z/OS data never leaves its secure native platform.
- **Dynamic self-service provisioning:** Empower teams with a centralized portal providing on-demand access to data environments and reducing data contention through automated, reusable reservation workflows.

Related Products

TDM is part of a comprehensive Broadcom DevOps portfolio that helps remove typical constraints across the delivery pipeline:

- **Agile Requirements Designer:** Provides a complete solution for requirements definition, model-based test case design, and automation, defining models for every test case. TDM provides aligned test data for cases that require it.
- **Service Virtualization:** Automatically creates virtual services to simulate dependent systems for end-to-end development and testing. TDM generates the data needed to develop realistic request-response pairs and scenarios.

Summary

High-quality software demands more than just automated tests; it requires high-fidelity data. Our TDM eliminates the industry's most persistent delivery bottleneck by providing an agile, secure, and reusable framework to discover, generate, and provision data on demand. By unifying AI-augmented synthesis, agentic data generation, and database virtualization with seamless DevOps integration, it empowers teams to maximize coverage and shrink test cycles while maintaining rigorous compliance and cost efficiency.