

“A synthetic data generation tool which generates non-production data that reliably mimics an organization’s actual data so that systems and applications developers can perform rigorous and valid systems tests.”

Test Data Manager

Test data generation, Sensitive Data identification

Enterprises are increasingly becoming data-driven thanks to their digital transformation initiatives. Delivering these business-critical applications on time and with high quality requires a reliable and repeatable testing process to accelerate time-to-market and reduce costs. However, compliance requirements for privacy laws such as GDPR, CCPA, and HIPAA limit the usage of PII/PHI data in test environments. Preparing data for testing is highly labor-intensive and can consume more than 50 percent of your overall testing time and resources. Even if the test data were prepared manually, human errors and low data volumes became significant limitations. Data-driven testing challenges can cause spiraling costs and pose a substantial risk to the quality of the resulting application. It is simply not cost-effective to clone large production databases for testing environments in addition to increasing compliance risks.

The legacy test data management tools helped generate fake data based on manually defined rules such as ranges, distributions, and correlation. However, defining these rules is a tedious and error-prone process. The modern test data management uses a Machine Learning-based approach to learn these rules automatically and generate realistic looking test data referred to as Synthetic Data.

DataOps TDM uses Artificial intelligence and Machine Learning based synthetic data generation algorithms to automatically detect PII elements and generate realistic looking synthetic data. This is done by using a subset of original data or user-defined sample data to generate a large amount of synthetic data which reliably mimics the actual data behavior but does not reveal any information regarding real data.



“Data privacy laws such as GDPR, CCPA, and HIPAA limit the usage of PII/PHI data in the test environments. DataOps TDM helps meet these compliance requirements by using AI to generate realistic-looking synthetic data while reducing tedious manual work.



Benefits



Automatically learns data patterns



Reduced cost of test data preparation



Compliance with Data Privacy requirements



Avoids the need for manual rule creation, thus saving time



Superior Quality of generated synthetic data



Improved Testing effectiveness with realistic data

Features



Connectivity and data formats

Ability to generate synthetic data in multiple file formats and automatically load it to your data sources on-demand.



AI-based algorithm to automatically learn data patterns and correlations

DataOps TDM automatically trains and generates a Machine Learning based model for generating test data that is statistically similar to the sample data.



Reduces storage requirements

Reduces the cost of data storage for databases used in testing environments by providing the ability to generate data on demand.



Supports data privacy and security

Generates fake sensitive data with built-in functions for commonly used data types (for example, social security numbers, phone numbers, and so on) and Supports custom selection for any data type, enabling compliance with corporate data security policies.



Efficiently generates large data volumes

DataOps TDM can generate millions of records in test data within a few minutes. The generation options can be easily extended using Python snippets.

Selected List Of Supported Data Sources

ORACLE®



teradata.



Parquet

Email us at contact@datagaps.com for a price quote, pilot or to know more about how we can help you solve your enterprise data testing issues.