

Accelario Synthetic Data Module

Efficiently augment testing and AI/ML training data with a fully automated, privacy-proof synthetic data portal

The Challenge: Delivering synthetic data that meets testing and AI/ML training data requirements while complying with stringent privacy regulations

Increased delivery velocity is a central tenet of agile methodologies and microservice practices. To keep up, agile QA testers need fast, independent access to testing data that covers every scenario and enables full testing coverage and scalability performance. At the same time, data analysts are struggling to find high-quality data that maintain real-world characteristics to use with their machine learning algorithms. Privacy is also a major concern, as any data copied from production for the use of 3rd party development teams must be devoid of identifiable information in order to comply with stringent regulations.

Synthetic data extends a powerful solution for the data needs of

both testers and analysts. With synthetic data, testers benefit from highly flexible data that can be manipulated into any data type and at any scale to potentially cover all of their test requirements. Analysts can use synthetic data that maintains real-world statistics to more effectively perform their analysis processes. What's more, the data is completely fabricated so that privacy issues are eliminated and lower environments can be provisioned for the use of 3rd party development teams.

However, the reality is that many QA professionals and analysts are still using spreadsheets to create their synthetic data. This highly tedious, manual process of generating

synthetic data is inefficient, ineffective and substantially slows down delivery velocity. Moreover, it typically produces insufficient data sets and inadequate data randomization that do not cover all scenarios and data types. This outdated process reduces testing accuracy for QA professionals and delivers suboptimal statistical distribution accuracy for analyst training purposes.

QA and dev teams are challenged to quickly provision the synthetic testing data required to satisfy agile requirements for agility and fast, high-quality deliveries. In many cases QA professionals are spending more time manually generating synthetic data than on actual testing.

The Solution: A simple-to-use synthetic data portal that empowers teams to independently design their own data for any testing or training requirement

Accelario's Synthetic Data module enables QA and analyst teams to instantly -- and without DBA support -- create synthetic databases that meet any testing and training requirements while conforming to privacy regulations. Synthetic data can be based on production metadata or can be

created from scratch, including metadata. Accelario learns statistical rules from the source so that the target synthetic database is always statistically correct, regardless of the size of the source. Test coverage is ensured by generating extreme data cases and bad path

scenarios, while scalability performance needs are achieved by easily and quickly generating target databases that are much larger than the source. Full support for Restful APIs enables seamless CI/CD data pipeline integration.

The Accelario Advantage

Accelario's advanced solution simplifies the entire process of provisioning and manipulating synthetic data. Simple to use, the module is a self-service portal that significantly reduces reliance on DBAs, speeding up the entire development process. Highly flexible, synthetic data can be generated to conform to any testing or training requirement.

› Highly flexible and customizable

Teams easily design their own synthetic data sets from scratch, or customize synthetic data based on production metadata to ensure that all testing and training requirements are met.

› Fast and scalable

An advanced data generation engine produces tens of TBs of synthetic data 10x faster than anyone else to ensure full testing coverage and scaling performance.

› Accuracy you can rely on

Production statistic rules are stringently maintained to ensure that synthetic databases are statistically correct even when the synthetic database is exponentially larger than the original data source.

› Fully automated testing

Full support for Restful APIs enables seamless integration with the CI/CD data pipeline so that test data is no longer an obstacle to fully automated testing.

› Privacy compliant

Synthetic data is completely fabricated, so that even when provisioning data to less secure environments privacy compliance is a nonissue.

› Optimized test coverage

Ability to generate extreme data cases and bad path scenarios to ensure full test coverage.

› Self-service

QA and analyst teams can provision their own synthetic data at the click of a button.