Quality Evolution

How GenRocket Enables Synthetic Data Transformation

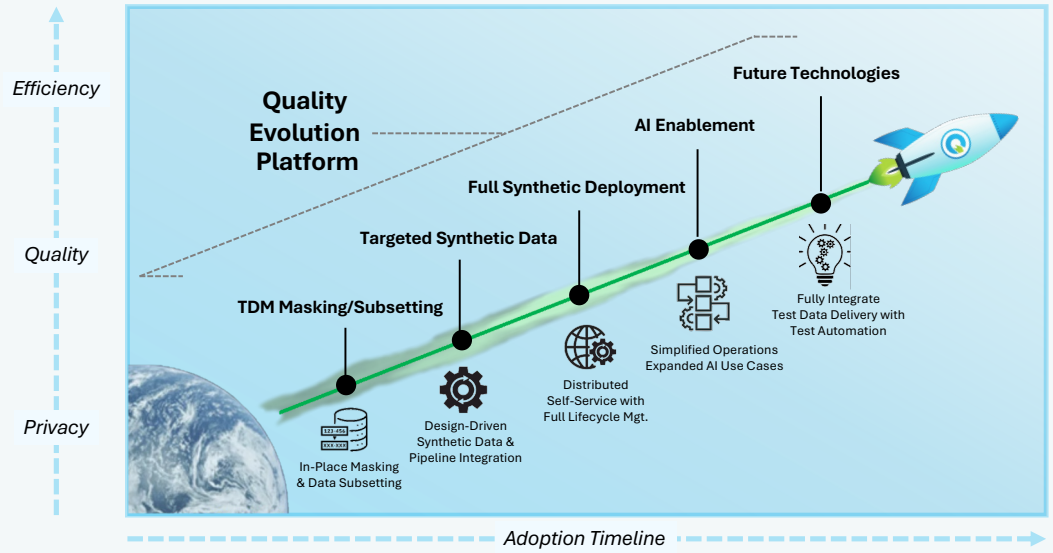
GenRocket

Quality Evolution is GenRocket's strategy for enabling our customers to transform from a reliance on production data to the design and delivery of synthetic data by deploying one integrated platform.

The GenRocket ***Quality Evolution Platform*** provides a bridge from the traditional approach to Test Data Management (TDM) to the industries most advanced platform for synthetic data generation. Additionally, the ***Quality Evolution Platform*** leverages AI technology to increase speed and simplicity.

In this presentation we will show how synthetic data transformation can be achieved to maximize data privacy, data quality, and operational efficiency.

How GenRocket Enables *Quality Evolution*



With GenRocket, you can start with familiar TDM processes like production data masking and subsetting. Then you can gradually transition to targeted synthetic data use cases as you follow an evolutionary graceful path to full synthetic data deployment.

GenRocket provides one integrated platform to enable this quality evolution journey. It includes a TDM bridge from production test data to the design and deployment of synthetic data for meeting any test data or training data requirement.

Following synthetic data transformation, the next step in this journey is to leverage advanced AI technology to streamline and automate data delivery for testing as well as generating high quality synthetic training data for AI-enabled applications.

Our platform is highly extensible and our roadmap anticipates exciting new technologies that will fully integrate synthetic data generation with test automation and development frameworks.

Quality Evolution

A TDM Bridge to the Future

Let’s start by focusing on the first phase of this quality evolution journey. That first phase is replacing a legacy TDM environment with GenRocket’s TDM Bridge to the future based on Design-Driven synthetic data generation.

The Bridge to Design-Driven Synthetic Data

PRIVACY – Absolute data security & privacy law compliance

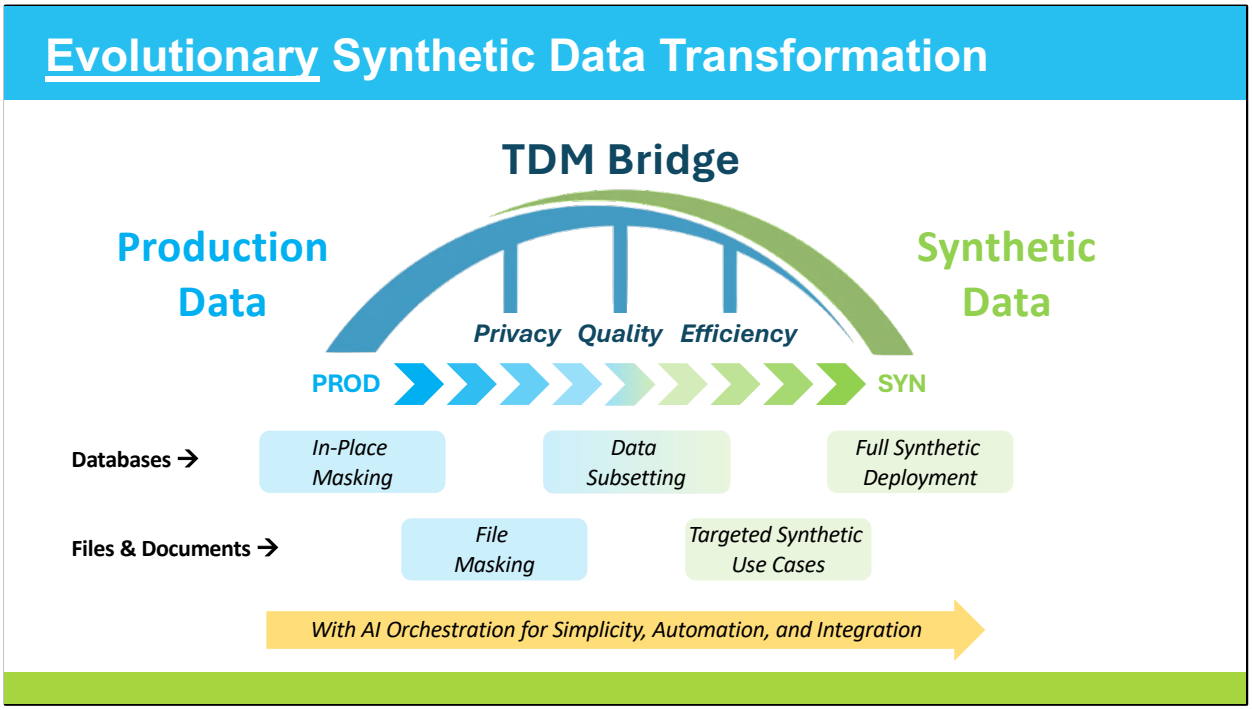
QUALITY – Expanded test coverage & accurate model training

EFFICIENCY – CI/CD pipeline integration & full test automation

Our goal is not to replace a traditional TDM platform with another traditional TDM environment. Instead, our TDM solution is meant to serve as a bridge to the future.

Using the GenRocket platform, Quality Engineering organizations can continue to provision test data using masked and subsetted production data just as they do today.

We also offer the technology and the best practices to gracefully evolve your current TDM environment to the next generation of Design-Driven synthetic data that will maximize data privacy, data quality, and operational efficiency.



Here's how our TDM bridge enable synthetic data transformation.

For masking full databases, we offer a high performance in-place masking solution. And for files and documents, we provide a flexible solution for performing advanced data masking operations for any kind of file.

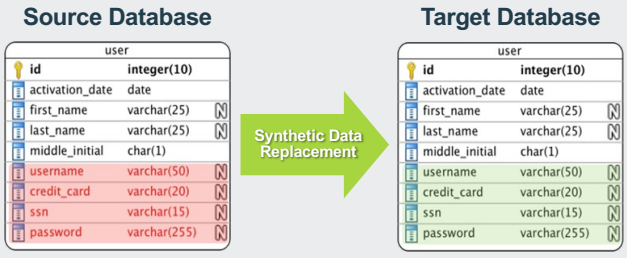
We also offer intelligent data subsetting that makes the provisioning of production data fast, flexible and cost-efficient.

Based on your adoption timeline, you can then introduce targeted synthetic data use cases using the same platform. Ultimately you can expand the system to full synthetic data deployment on a global scale.

GenRocket is also streamlining this transformation journey with AI orchestration that will simplify, automate, and integrate synthetic data generation with any testing tool or framework.

GenRocket's Advanced Data Masking Technology

Synthetic Data Replacement



Replace Sensitive Data with
Controlled & Conditioned Synthetic Data

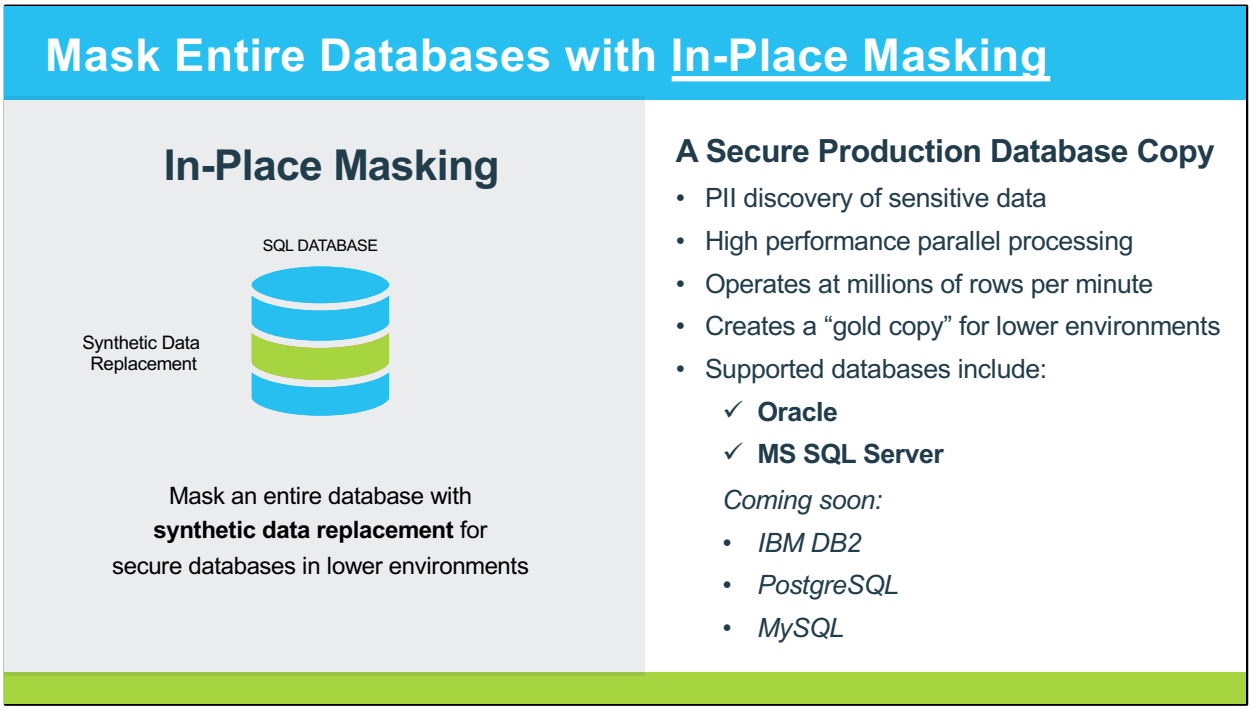
The Best of All Worlds:

- Cannot be reverse-engineered
- 100% privacy law compliant
- Full referential integrity
- Controlled data variety
- Fast & scalable performance

GenRocket uses the industry's most advanced form of data masking technology. It's based on a process called ***Synthetic Data Replacement*** (SDR). Using SDR, sensitive information is identified for replacement with synthetic data values that cannot be reverse-engineered and ensure 100% privacy law compliance.

We also ensure full referential integrity for all synthetically replaced key values across all related tables in the database environment. GenRocket holds the only US patent for maintaining referential integrity in a synthetic data platform.

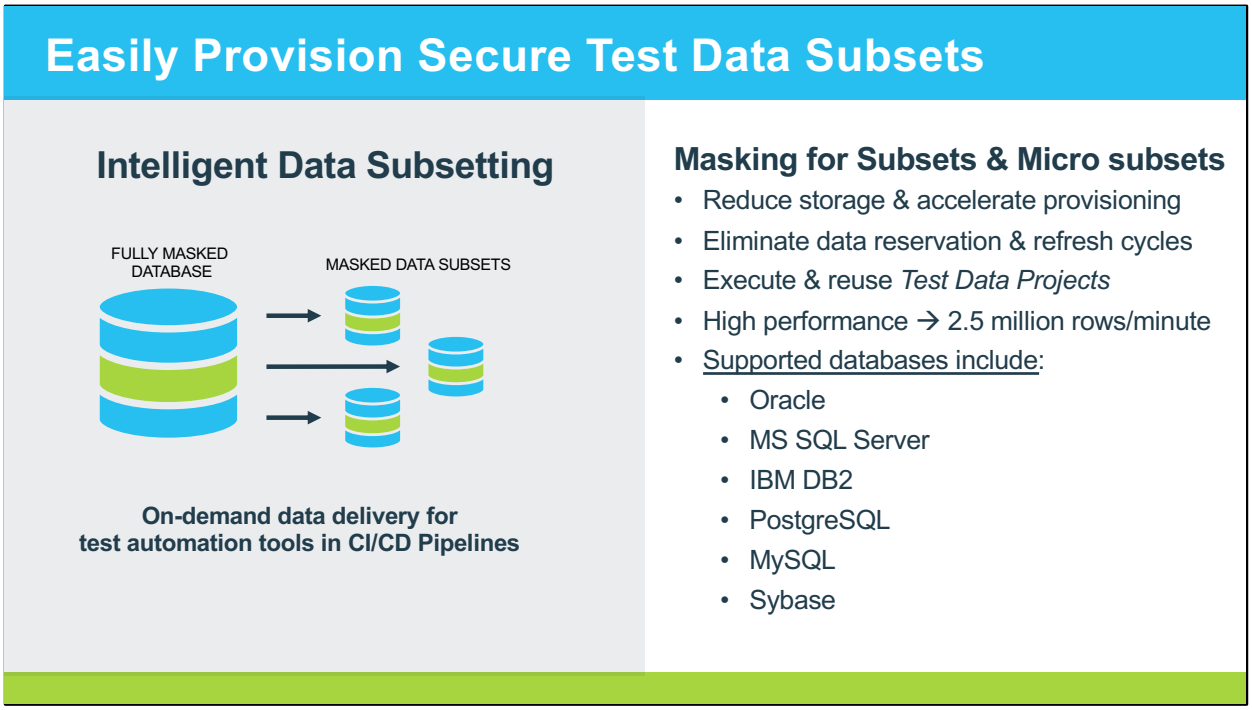
And with GenRocket, you control the volume, variety, and format of the data used for synthetic data replacement. Through the use of high performance parallel processing, we deliver fast and scalable performance for all masking operations.



GenRocket offers a comprehensive ***In-Place Masking*** solution. This allows you to mask an entire database, or multiple databases, using synthetic data replacement to provide a secure copy of one or more production databases in lower environments.

Our ***In-Place Masking*** solution also includes ***PII Discovery*** so that sensitive data can be quickly flagged for synthetic replacement during the masking process. And with the use of multiple processing threads, millions of rows of data can be masked per minute, allowing you to efficiently produce a gold copy of your production database to serve the needs of your development and testing operation.

The ***In-Place Masking*** solution currently supports Oracle and MS SQL server with support for IBM DB2, PostgreSQL, and MySQL coming soon.



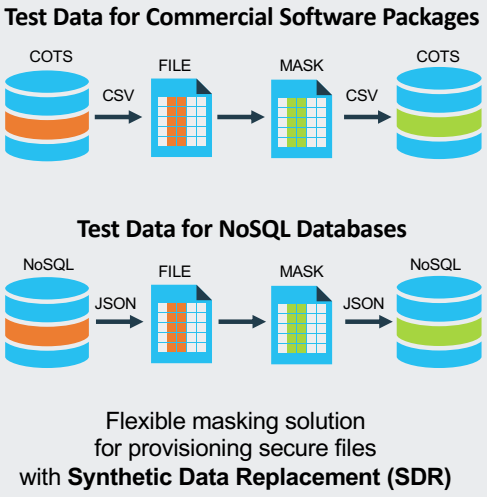
Once you have created a secure copy of your production database, you can quickly and easily provision secure test data subsets for fast and flexible data provisioning.

Using GenRocket's ***Intelligent Data Subsetting*** solution, secure production datasets are available for on-demand delivery through the use of executable and reusable Test Data Projects.

This reduces the amount of storage required for test data and accelerates the provisioning process. It also eliminates the need for the data reservation and refresh cycles commonly associated with traditional TDM. With GenRocket, a fresh copy of the data is delivered with the execution of each Test Data Project.

Our subsetting and masking solution supports a full range of SQL databases including Oracle, SQL server, IBM DB2, PostgreSQL, MySQL, and Sybase.

Perform File Masking for Any File Type



Ensure Data Privacy for Any File Type

- Selectively mask any column with SDR
- Support for COTS and NoSQL databases
- Database bulk load or import via API
- Supported File Types:
  - Any delimited file (e.g., CSV)
  - Any fixed file format (e.g., VSAM)
  - X12 EDI
  - JSON
  - ORC (Hadoop)

GenRocket offers a flexible file masking solution to ensure data privacy for any type of file. With GenRocket's file masking, test data files can be used for testing commercial software packages such as Guidewire, Workday, PeopleSoft and SAP.

File Masking also allows secure data for testing with NoSQL database environments. Like GenRocket's ***In-Place Masking*** and ***Intelligent Data Subsetting*** solutions, File Masking is performed using our advanced ***Synthetic Data Replacement*** technology.

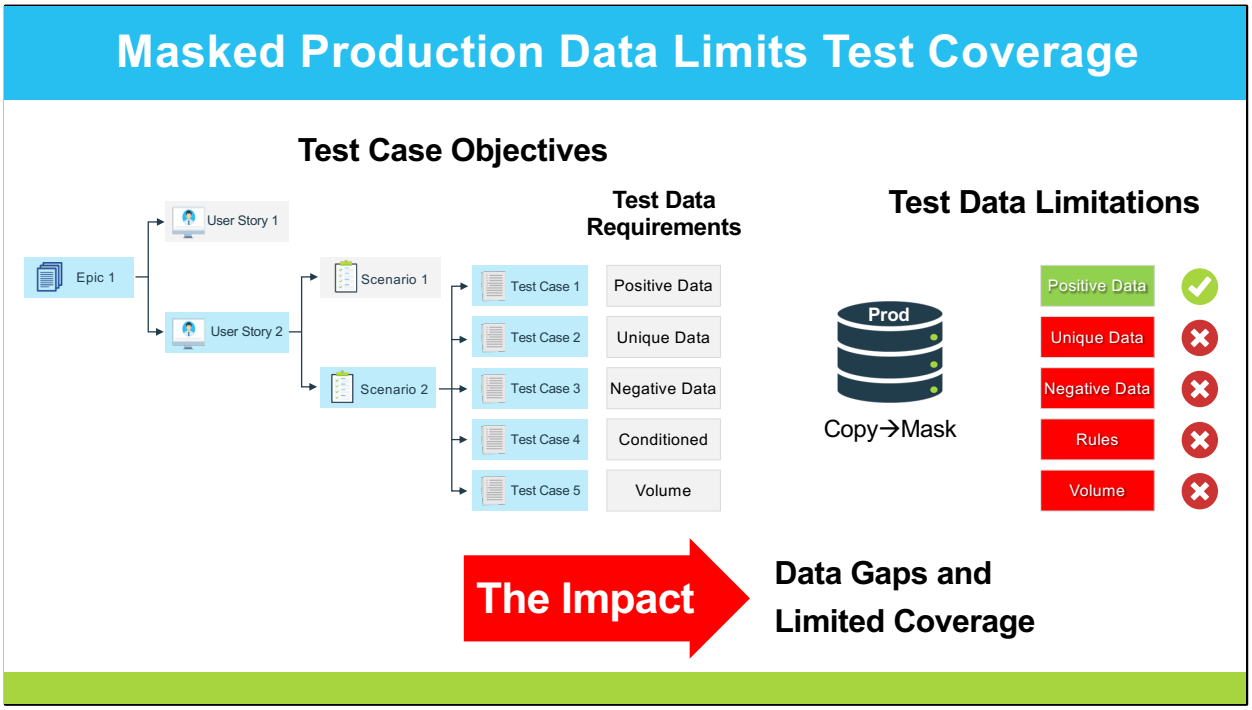
Once files are masked, the data can be used for bulk loading operations or imported via API into your test environment. File masking can be used with any delimited file and any fixed file format, as well as with X12 EDI, JSON, and ORC (Hadoop).

Quality Evolution

Realizing the Value of Synthetic Data Transformation

The TDM capabilities we have just described provide an evolutionary bridge from the use of production data to the generation of synthetic data.

Now let's have a closer look at the value of synthetic data transformation and GenRocket's unique ability to design synthetic data for any testing or AI model training requirement.

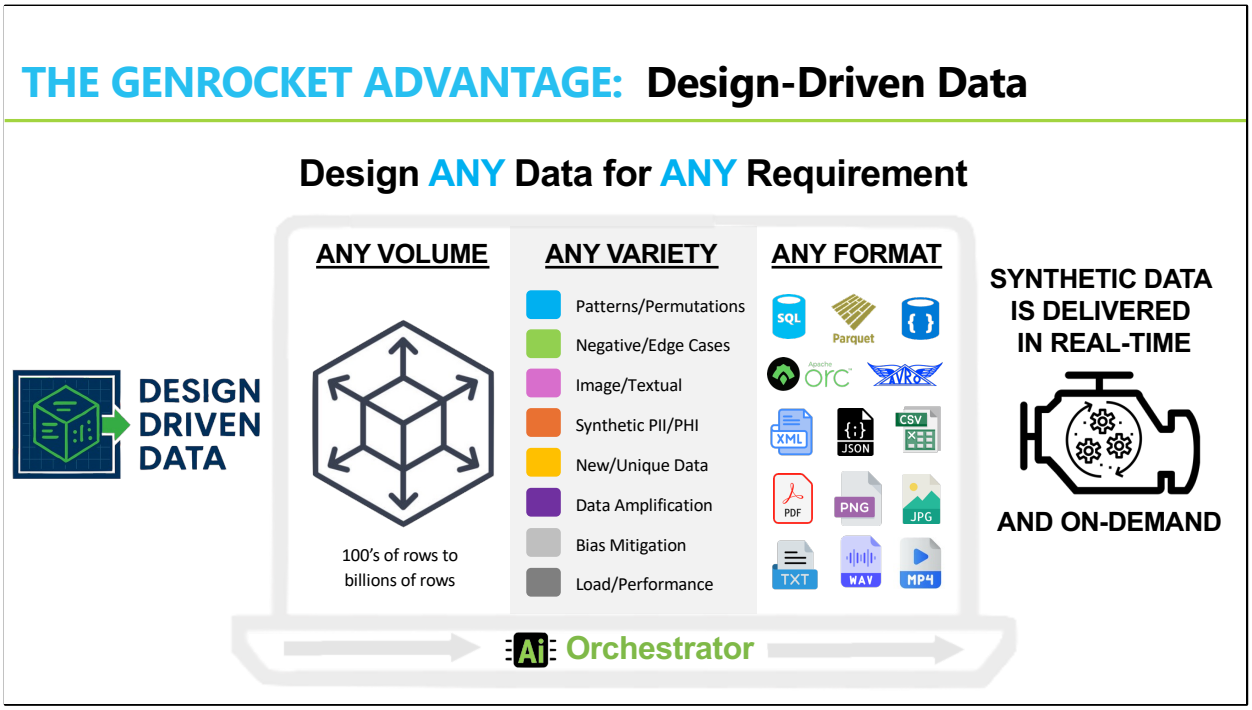


Our ***TDM Bridge*** preserves the production data provisioning methods that are commonplace in most organizations. However, it doesn't remove the limitations on data quality that are imposed by the use of production data for testing.

With production data, testers are limited to the data values that are present in the database. But what if they require specific patterns and permutations that are not present in the data to maximize coverage? And what if their test cases require edge case values for testing business rules and policies? And what if they need data for negative testing? Or new and unique data values for testing greenfield applications?

Unfortunately, testers must find and fill these data gaps themselves - often with manually created spreadsheets or custom scripts.

As as result, the traditional TDM approach usually limits test coverage and delays the testing process.



GenRocket has innovated a new and different approach to synthetic data generation. We call it **Design-Driven Data**. It allows our customers to literally design any volume, variety, or format of synthetic data they need – either for the purpose of testing software and training AI models.

When it comes to controlling data variety, we mean the ability to represent any pattern or permutation of data, any and all edge cases, and the combined use of positive and negative data values.

And for AI-enabled applications, ***Design-Driven Data*** offers the ability to mitigate bias and amplify key signals to improve the performance and accuracy of AI models.

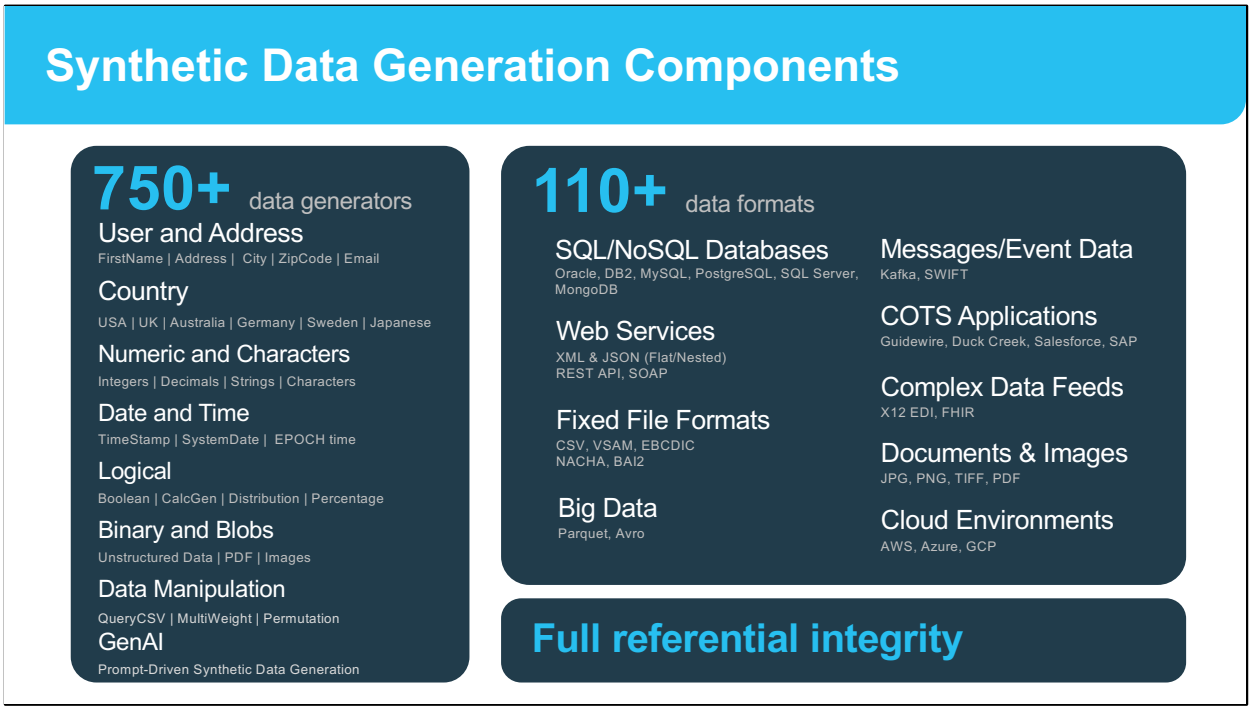
Designed Synthetic Data Delivers Full Test Coverage

Pattern	Realistic	Sequential	Random	Negative	Null
firstName1	Ms. Tereasa F. Saldana	001-01-0001	749-40-0182	749-40-0182	749-40-0182
firstName2	Mr. Everette Q. Groom II	001-01-0002	797-59-7445	797-59-7445	null
firstName3	Mr. Jules U. Hackney Jr.	001-01-0003	135-93-8060	135-93-8060	135-93-8060
firstName4	Mrs. Kristina J. Brick	001-01-0004	214-82-8447	214@82&8447	null
firstName5	Mr. Francisco M. Grimes II	001-01-0005	170-60-5224	170-60-5224	null
firstName6	Dr. Iona D. Starrett	001-01-0006	302-76-0978	302-76-0978	null
firstName7	Ms. Patricia O. Ingraham III	001-01-0007	266-20-5659	266-20-5659	266-20-5659
firstName8	Ms. Tracee M. Farah	001-01-0008	005-57-7667	005#57%7667	005-57-7667
firstName9	Mr. Alva I. Ziegler Jr.	001-01-0009	490-48-8084	490-48-8084	null
firstName10	Dr. Mike T. Youngblood II	001-01-0010	471-29-7519	471-29-7519	null

Here is an example of synthetic data that has been designed to meet various test case objectives. As you look at the data samples in each column, you can see the use of patterns, realistic values, sequential data, randomized data and negative or null data.

This slide is just an example to make the point that data can be designed for both simple and complex use cases. GenRocket's data generators are intelligent, meaning they can be configured and combined to perform complex operations. They can, for example, produce data that is dynamic and stateful for testing program logic, business rules and complex multi-stage workflows.

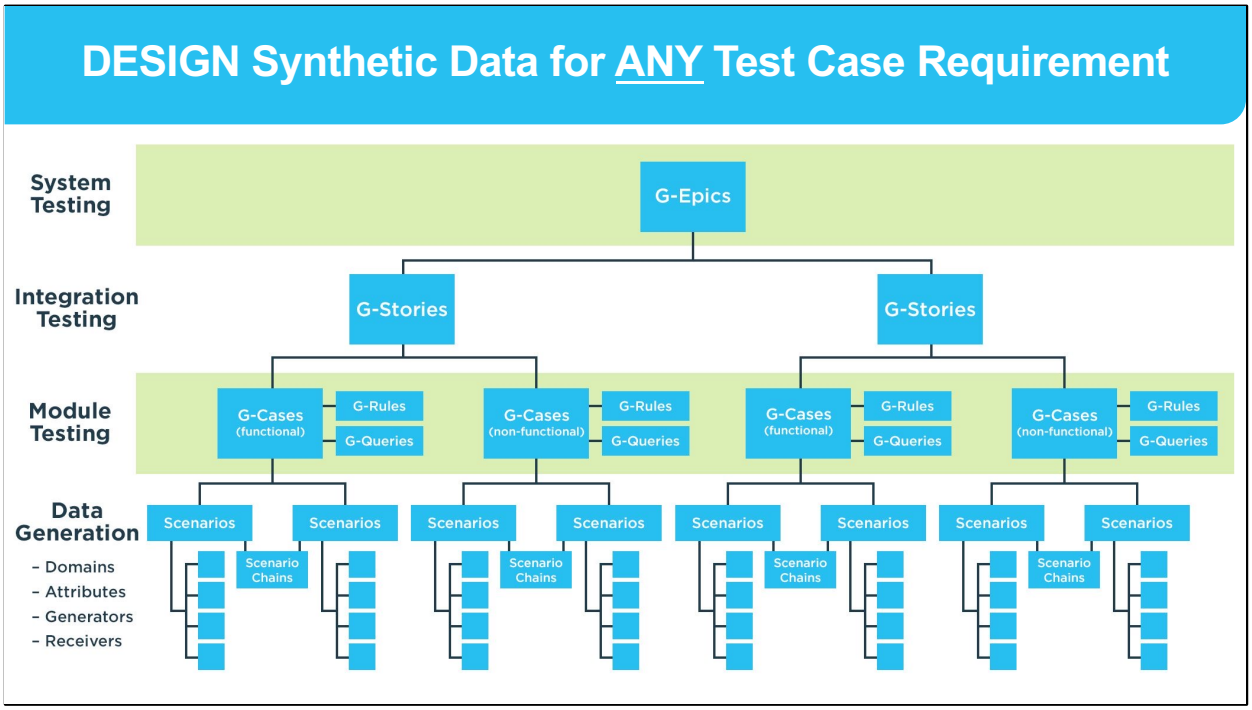
With GenRocket, synthetic data can be formatted to simulate virtually any structured, structured, or semi-structured data environment.



GenRocket leads the industry with the most comprehensive library of intelligent data generators with more than 750 generators in our platform.

GenRocket also offers the greatest diversity of output data formats of any synthetic data platform with more than 110 formats available. As a core design principle, GenRocket always ensures full referential integrity for all generated synthetic data.

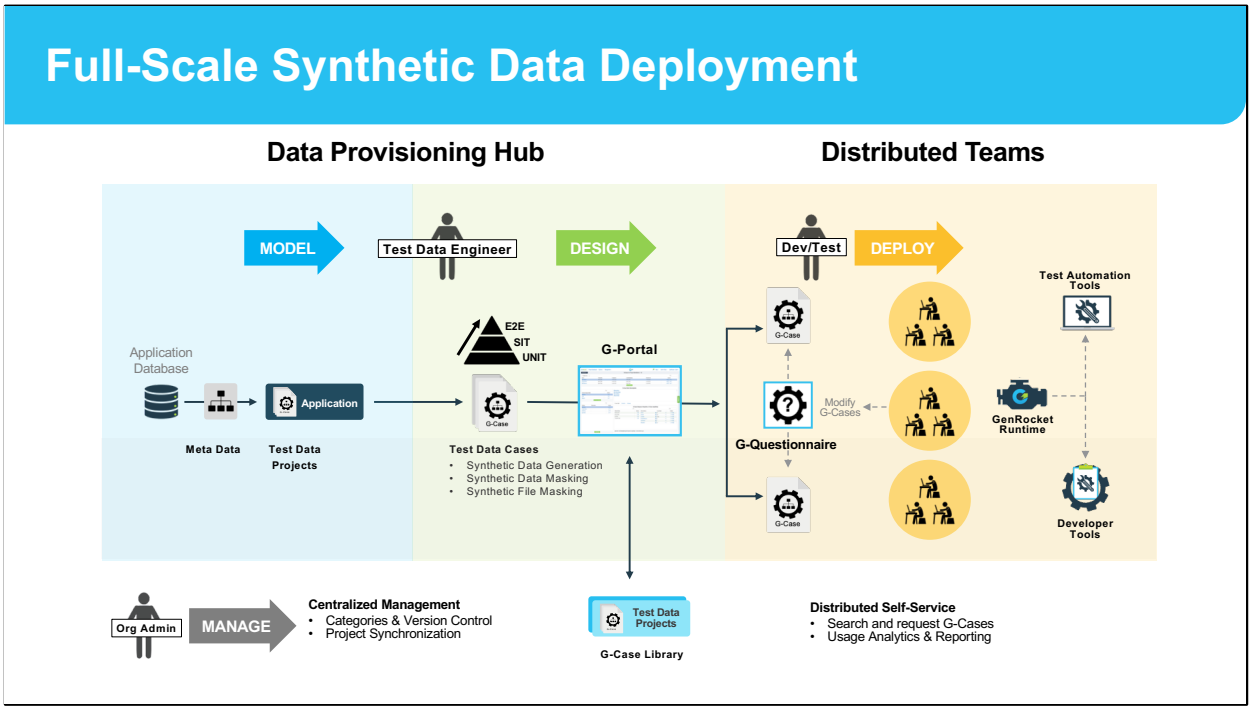
If there is a requirement for a data generator or output format that is not currently supported by our platform, we can quickly create new ones to meet the test and training data needs of our customers.



GenRocket allows you to design synthetic data to meet the needs of any test case from unit testing to integration testing to end-to-end system testing.

It's component based architecture enables a modular design approach where simple test data cases can be aggregated into more complex stories and epics to maximize reusability. And test data cases can be version controlled and repurposed for multiple test categories such as functional, performance, and regression testing.

All Test Data Projects are stored and organized in a repository that can be distributed across the organization. When a change is made to one repository, it is synchronized across all other repositories. This organized approach to ***Design-Driven Data*** enables true enterprise scalability.



We have found that the most effective deployment model is to establish a centralized team of experts to serve as a data provisioning hub for distributed teams of developers and testers.

Using this model a small team of test data engineers can design test data cases and deliver them to distributed teams quickly and efficiently. Users request the data they need through ***G-Portal*** and the centralized team returns executable test data cases ready for deployment into their test environment.

Additionally a component called ***G-Questionnaire*** allows users to make quick modifications to their test data cases using a simple wizard like interface. This deployment model can be centrally managed by an organization administrator who has centralized control over access to the system and the utilization of it's resources.

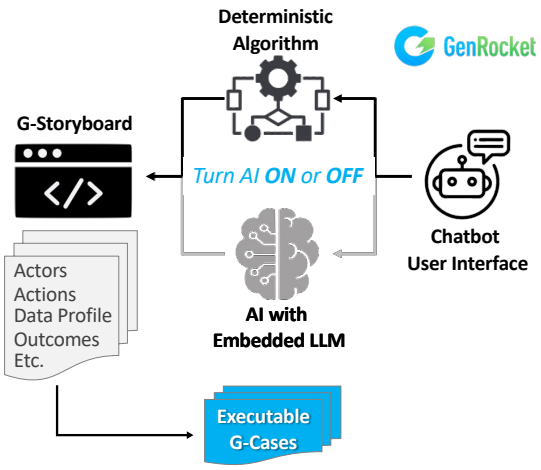
Quality Evolution

Leveraging AI
for Speed, Simplicity & Accuracy

The next phase of GenRocket's ***Quality Evolution Platform*** development will be to focus on the most effective way to leverage AI technology for increasing the speed and simplicity of its operations.

We will also be supporting expanded use cases for generating synthetic training data that maximizes the accuracy of AI models.

AI Orchestrator: Design Driven Synthetic Data – *Simplified*



- Synthetic data design will be radically simplified with a chatbot interface
- Initially a deterministic algorithm will guide the user through the process
- This will accommodate organizations that restrict the use of AI internally
- G-Storyboard will translate natural language requirements into G-Cases
- Embedded AI can replace this process with fully conversational approach
- The two methods can be configured to meet customer security requirements

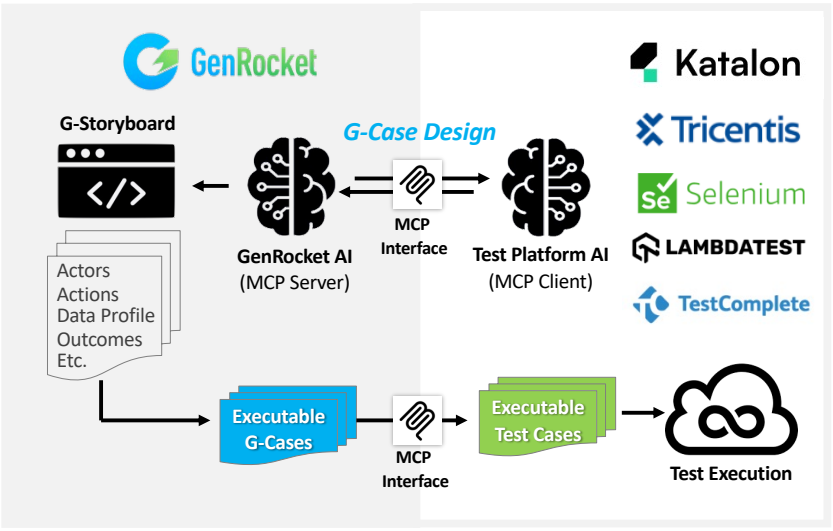
GenRocket will simplify the way synthetic data is designed through a new component called the ***AI orchestrator***. Using a chatbot interface, users will be guided through the data design process using a deterministic approach that is based on best practices. This approach avoids the use of AI technology for organizations that want to restrict its use. A new component known as ***G-Storyboard*** will translate user requirements into executable ***G-Cases***.

Additionally, an embedded large language model (LLM) will allow the chatbot interface to leverage the broader flexibility of AI to build out their G-Cases. An embedded LLM will provide locked-down security and will translates user requirements into ***G-Cases*** using ***G-Storyboard***.

Customers will have the option of choosing between the non-AI (deterministic) approach or the full AI-driven approach to design

their synthetic data.

AI Integration: Automated Synthetic Data Delivery

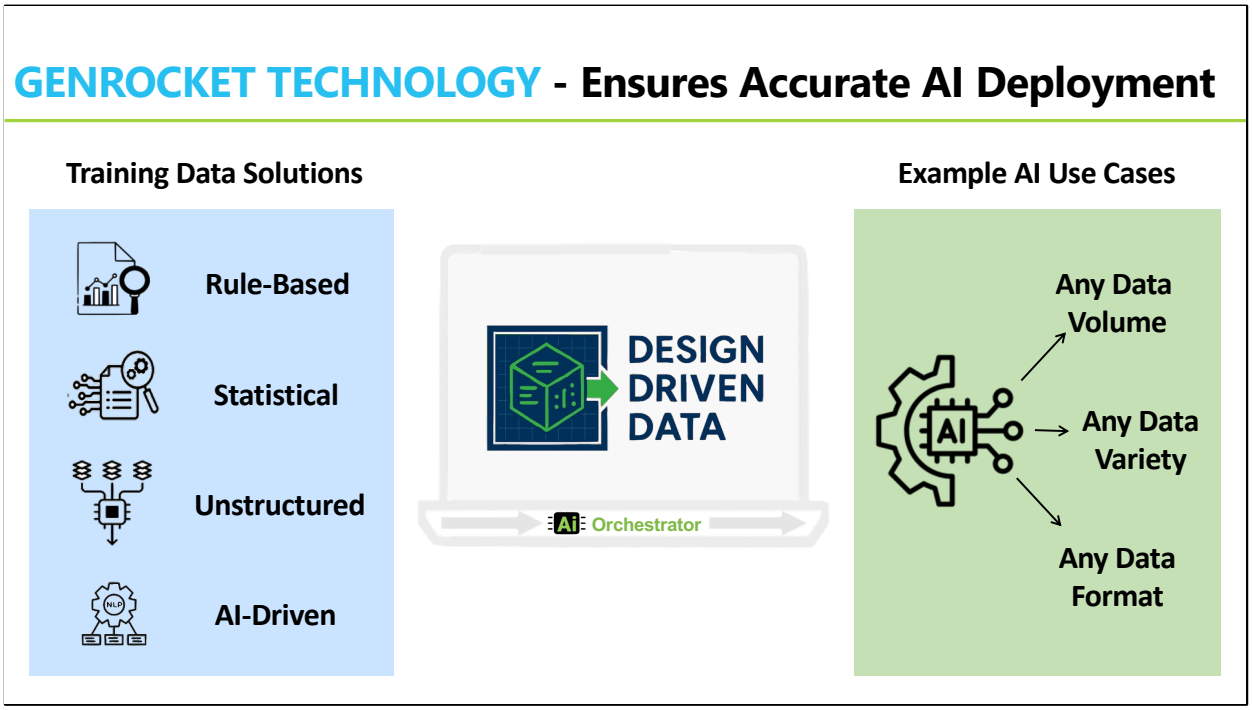


- Test Cases are designed with AI by test platforms
- MCP interface connects to the GenRocket platform
- Synthetic Data is designed using GenRocket AI
- Executable G-Cases are returned via MCP
- Test platform generates real-time synthetic data
- Integrated test & test data cases run in CI/CD pipeline

The same architectural approach will allow GenRocket to leverage its ***AI Orchestrator*** as as a method for integrating with third-party test automation tools and frameworks.

This will allow synthetic data to be requested by external test automation platforms at the same time they are using AI to build their test automation scripts. The Model Context Protocol (MCP) interface standard will be used to provide connectivity between platforms.

And once again, ***G-Storyboard*** will be the internal GenRocket component that will translate these synthetic data requirements into executable ***G-Cases*** that can be delivered to, and executed by, third-party test automation platforms.



GenRocket's AI technology will also support a wide variety of training data solutions that use the power of ***Design-Driven Data*** to maximize the performance and accuracy of AI models.

Here are four categories of training data solutions we support.

Rule based training data – for improving the performance a fraud detection systems.

Statistical training data- that leverages metadata to generate the statistical data profile required for synthetic training data

Unstructured training data- to combine forms, images and even audio and video with structured tubular synthetic data

And AI driven training data – to integrate the delivery of tabular and textual synthetic data with test automation systems