SYNTHETIC TEST DATA AUTOMATION

Enterprise Features and Their Value

MODELING Your Application or Database

In order to generate test data that accurately reflects the structure of your application or database, it must be easy to model and maintain each test data project as changes to the data model occur throughout the lifecycle of the application.

Test Data Modeling

Referential Integrity	Support	Why this is Important
Maintain referential integrity of parent/child/sibling relationships across the data domains within an application database or across mulitple databases used by multiple applications.	YES	Maintaining referential integrity is essential for test data within an application or database and across applications or databases. GenRocket holds the only U.S. patent for generating test data with referential integrity.
Ensure the consistency and integrity of synthetic data attributes across applications, data sources and targets. For example, a customer name must always match the same customer ID across multiple transactions simulted by real-time synthetic data generation.	YES	Data consistency is essential for regression testing where the exact same data needs to be generated on every test run. Data consistency is also essential for test data as data flows through API's and applications in complex testing.

Data Model Setup	Support	Why this is Important
The number of methods to build a test data project based on a database or data model	YES	Maintaining referential integrity is essential for test data within an application or database and across applications or databases. GenRocket holds the only U.S. patent for generating test data with referential integrity.
Ability to read a SQL database as a data source	YES	Customers want to quickly and accurately create their data model as a test data project for SQL databases. XTS (Extract Table Schema) reads the meta data of any SQL database and uses that to quickly set up a test data project for that application or database. A wizard is provided to quickly define the data relationships.
Ability to read a NoSQL database as data source	YES	Customers want to quickly and accurately create their data model as a test data project for NoSQL databases.
Ability to read JSON data structures as data source	YES	Customers want to quickly and accurately create their data model as a test data project using JSON data structures.
Ability to build models of data on the fly without direct reference to a database as a data source	YES	Scratchpad is a convenient tool that quickly sets up a test data Domain with associated Attributes. This tool is convenient for Agile teams that want to quickly model and design new Domains.
Ability to visually display the dependencies between the different data structures	YES	Most testers struggle to understand the relationships between different tables or objects in their data model which limits their ability to test. GenRocket's graphic visualization allows testers to see all the Domain relationships (parents, children and siblings).

Life Cycle Management	Support	Why this is Important
Ability for the data model, relationships and data generators to automatically adapt when a change occurs (A new attribute is added to the Customer Domain)	YES	Data models change in applications over time, some change frequently. These natural "life cycle" changes should be handled automatically with limited effort on the part of architects or testers. G-Refactor automatically updates all impacted Domains, Attributes and Scenarios when a change is made to the data model in GenRocket.
Ability to use machine learning algorithms to detect data source changes in order to apply the changes to the data model relationships and generators	NO	Data models change in all applications over time, some change frequently. Automatic data model change and Admin notification ensures that test data keeps pace with these "life cycle" changes. G-Delta is in development. Release scheduled for March 2021.
Ability to manage multiple versions of data models and sources	YES	Test data projects should be easily versioned so that test data can be provided for many different versions of the software application / database. GenRocket versioning is fully automated and can be done with the click of a button.

DESIGNING Your Test Data

In order to generate test data that accurately reflects your test case requirements you need an easy to use interface that allows any form of test data (positive, negative, patterns, permutations, boundary values, dynamic, load, etc.) to be designed for a given data model.

Test Data Design

Graphical User Interface	Support	Why this is Important
User-friendly, Self Service Web interface with intuitive and intelligent navigation	YES	Testers do not want to "request and wait" for their test data. Test data cases should be able to be created and modified by testers themselves. GenRocket offers extensive self service capabilities for testers to create, copy and share Test Data Cases and apply different Test Data Rules against those Test Data Cases

Ability to preview data in the user interface prior to data generation	YES	Seeing a preview sample of the data in the user interface greatly speeds up the process of creating and validating test data. GenRocket offers a real time data preview.
Built in help throughout the user interface	YES	Users prefer to have context sensitive help to simplify use of the the tool. GenRocket has integrated help throughout the user interface
Built in help throughout the user interface	YES	Users prefer to have context sensitive help to simplify use of the the tool. GenRocket has integrated help throughout the user interface
Built in intelligent automations to speed up common user tasks	YES	Tester productivity is important. Automating repetitive tasks (copying, modifying, versioning, etc.) is essential. GenRocket has many automations built into the platform as well as tools to accelerate repetitive tasks.
Data Generators	**Support**	**Why this is Important**
Number of data generators currently available	650	A useful synthetic test data platform will offer all of the data generators needed to generate the required data in the exact data structure needed. GenRocket offers over 650 data generators and a "linked Generator" feature where Generators can be combined to give additional flexibility about what data is generated.
Data generators have parameters that allow user defined control over the data that is generated (random, patterned, sequential, null, %, etc.)	YES	Random test data is not ideal for many test case requirements. Synthetic data should be able to be patterned, null, split by percent and percent within percent as needed. GenRocket offers full control over data patterns and not just random data.
Ability to configure Generators so that data is generated in exact length and types	YES	To accept data, databases often require data of an exact length and type. GenRocket allows exact control over length and type of data.

Ability for data to be dynamically generated based on variables that change during the execution of complex workflows.	YES	In complex testing, where data is flowing through multiple systems, the data needs to be dynamically generated based on an evaluation of the data in the current state. So if a customer has a balance of above $50,000 in a savings account "gold" customer type is generated. GenRocket offers a rich set of dynamic data generation functionality.
Ability to generate all valid permutations & combinations of data	YES	In regulated industries, where compliance and full testing is required, partial testing isn't enough. Full testing comes from testing all permutations and combinations. Permutation testing with production data is not possible and attempting to do this in spreadsheets is almost impossible. GenRocket automates data generation for all valid permutations and combinations.
Data generators can generate identical data every time data is generated	YES	In automated testing, data for each test often needs to be identical for every run of that test to ensure the validity of the test. GenRocket's "seed" feature allows test data that is generated by a generator to be identical every time the data is generated.
Data Generators can insert data into an existing database that already contains data. The added synthetic data needs to accurately add to the data in the database	YES	When generating data to be added to existing data in a database, there needs to be a method to ensure data generation starts from the last data value in the database. GenRocket evaluation Generators can determine the last ID number or Account number (for example) and accurately add incremental synthetic data to the production data.
Data Generators can remember where they stopped generating data and start again at a later time while preserving state - this is called stateful data generation.	YES	Data Generators can remember where they stopped generating data and start again at a later time while preserving state - this is called stateful data generation.

Ability to combine (link) multiple generators to increase the versatility of the data that is generated	YES	With synthetic data generation there is an almost infite variety of data that is needed. To allows testers to create the exact data they need on the fly, it's helpful to allow testers to combine or link data generators together to generate more specific data. This is a powerful GenRocket feature and dramatically increases the variability of test data.
Generators must be intelligent and be able to maintain contextual relationships. For example a city must match a valid state and postal code.	YES	To be useful and valid, synthetically generated data values need to relate to other data values. For example a City, State and Postal Code should all match. GenRocket Generators have the ability to "talk" to each other to ensure that generated data is always valid.
Ability to generate a dataset based on percentages as well as percentages within percentages.identical data every time data is generated	YES	Testing usually requires control over the test data to ensure that the weighting (percentage) of data generated is realistic or intentionally unrealistic (negative, edge). Percentages and percentages within percentages should be able to be generated. GenRocket has extensive data weighting (percentage) generation functionality.
Ability to store or add re-usable assets into a Generator (e.g. a list of car types: sedan, convertible, hatchback)	YES	Not all data needs to be synthetically generated. When there is a known list of values, such as car types, it's useful to store the required values and use them. GenRocket offers a "List Generator" for this purpose as well as 17 different Query Generators that can query data such as a list of real postal codes from a file (spreadsheet) or from a database during the data generation process.
Ability to generate transactional or historical data for past or future values	YES	Some test cases need to test transactions that have not yet happened (future) or need to be simulated for the past. GenRocket easily generates transactional data in the future or past.

Test Data Cases, Stories & Epics	Support	Why this is Important
Ability to create and organize many different cases or scenarios that can be used / re-used with different test cases	Yes	Testing at the enterprise level involves the creation of hundreds to thousands of test cases. Each test case needs data. The ability to quickly create test data for each test case is essential. GenRocket has powerful capabilities to create and organize test data generation cases (Test Data Cases) in a structured, searchable layout.
Ability to organize many test data cases / scenarios into larger stories or epics that can be used / re-used with different test cases or workflows	Yes	End-to-End or workflow testing requires test data that is referentially accurate and aligns with the data requirements of the stage in the workflow. GenRocket Test Data Cases and Rules can be packaged into Test Data Stories and Test Data Epics to ease the process of creating and deliving test data for complex workflow testing.

Test Data Rules	Support	Why this is Important
Ability to create simple or complex test data rules with no programming language skills required	Yes	Most software applications have business logicthat must be tested with valid data and invalid (negative) data. The data to test business rules need to exactly match the test case requirements. In GenRocket, simple or complex rules can be created, copied and applied against Test Data Cases, without programming skills, to control the data that is generated.
Data generation rules can be used and reused in multiple scenarios or test data cases	Yes	For versatility of the data generation platform, rules should be able to be created and re-used to control data generation for many different scenarios. In GenRocket, Test Data Rules can be applied to one or as many Test Data Cases as required.

Test Data Queries	Support	Why this is Important
Ability to query and blend real production data values (e.g. account#, member ID#) with synthetically generated data for data validity	Yes	While synthetic test data meets most testing requirements, there are times when "enumerated" data, real production data values, (e.g., Member ID#, Account #, ICD-10 healthcare codes) that is necessary for ensuring data validity in a system. GenRocket offers 17 query generators that can query specific enumerated data values. Test Data Queries make it easy to query specific data values and blend those data values into a Test Data Case with synthetically generated data.

Test Data Workflows	Support	Why this is Important
Ability to deliver data to test complex workflows across heterogeneous environments with the ability to keep track of the mapping name key value pairs between components	Yes	There are multiple ways to accomplish this with different components and features including scenarios, scenario chains, test data cases and mapping tables.
Ability to design workflows for complex data generation	Yes	For the most complex test data workflows the GenRocket API is recommended. A new module, G-Workflow, will become available to make modeling complex workflows easier in Q2 2021

DEPLOYING Your Test Data

In today's Agile world, test data needs to be deployed on demand as part of a CI/CD pipeline. Test data needs to be available in a wide variety of formats, directly inserted into databases, and integrate seamlessly with any test automation tool or framework

Deployment and Integration with Test Environments		
Test Data Deployment & Formats	Support	Why this is Important
Number of data Receivers (formats) currently available	73	For a test data platform to be useful it needs to being able to quickly and easily deliver test data in any desired data format such as XML, JSON or to any SQL or NoSQL database. GenRocket accomplishes test data deployment through a "Receiver" component. Over 70 Receivers are currently offered and new Receivers are added in 2 to 3 weeks for our customers. Here is the current list of supported data formats: https://genrocket.freshdesk.com/support/solutions/articles/19000114411-genrocket-supported-output-formats
Ability to directly insert data into SQL databases	Yes	Test data solutions should be able to quickly insert data into SQL databases. GenRocket can connect to just about any SQL database over JDBC and also offers "bulk load" Receivers for high speed data insertion into databases, when millions to billions of rows of data are required.
Ability to directly insert data into NoSQL databases	Yes	Test data solutions should be able to quickly insert data into NoSQL databases. GenRocket can connect to just about any NoSQL database over JDBC.
Ability to deliver files to AWS S3 and Snowflake	Yes	Test data solutions should be able to quickly deliver data to the cloud such as an AWS S3 bucket or Snowflake. GenRocket works with most popular cloud databases and applications.
Ability to deliver files in real time over REST API's and SOAP	Yes	Data in most complex applications is not static; it moves over API's so test data also needs thea ability to be sent over a REST API or SOAP. GenRocket supports both SOAP and REST API's.

Ability to simulate complex data feeds with nesting and calculations	Yes	With most industries transforming to cloud solutions, data is often delivered between cloud applications in the form of a data feed. Test data solutions should be able to generate and format data to exactly simulate data feeds. GenRocket's architecture is perfectly suited to simulate complex data feeds with any level of nesting and complexity.
Ability to simulate SWIFT and KAFKA messages	Yes	KAFKA messages are used by 80% of the Fortune 100. The ability to simulate and test KAFKA and SWIFT messages is important for an enterprise class test data solution. GenRocket's architecture is perfectly suited to simulate message formats like KAFKA and SWIFT.
Ability to generate files in flat and nested XML and JSON	Yes	A standard requirement for any test data solution is to be able to simulate flat and nested XML and JSON data formats; these formats are fully supported by GenRocket
Ability to generate X12 EDI transaction set data	Yes	In the North American healthcare market, X12 EDI (Electronic Data Interchange) is the industry standard for how electronic data is sent such as an X12 EDI 837 I claim. GenRocket is the only test data company to partner with X12, the industry standards body, and fully supports all HIPAA mandated EDI formats such as EDI 837, EDI 835 and EDI 834.
Data Subsetting & Masking	**Support**	**Why this is Important**
Ability to select and copy a subset of data from a production database copy to a destination database	Yes	Data subsetting is the ability to connect to a production database copy and instead of copying the entire database, querying and copying specific data values from that database. Data subsetting is available through GenRocket's G-Subset feature.

Ability to mask sensitive data	Yes	Any production data values that contain PII or PHI need to be anonymized (i.e., masked). GenRocket's data masking approach is called "Synthetic Data Replacement" and is an integrated solution with the data subsetting (G-Subset) solution where a subset of data is copied and sensitive data is replaced in real time with 100% secure synthetic data.
Ability to auto-detect sensitive (PHI, PII) data for masking	No	Some test data solutions have algorithms that help identify sensitive data values. GenRocket does not offer an automated solution to identify sensitive data values as synthetic data, by definition, is 100% secure.
Test Environment Integration	**Support**	**Why this is Important**
In-House test automation framework integration	Yes	Some organizations have developed their own in-house test automation frameworks and need to have test data cleanly integrated with that framework. GenRocket's Runtime, GMUS (GenRocket Multi User Server), REST and Socket engines offer many different ways to integrate data with any testing framework. GenRocket offers a wide range of flexible components to handle any of these requirements. Data can be delivered into the CI/CD pipeline on demand, in real time.
CI/CD integrations (AWS, Azure, etc.)	Yes	Most organizations are moving towards Agile and CI/CD pipelines need to have test data quickly and easily integrated. GenRocket's Runtime, GMUS (GenRocket Multi User Server), REST and Socket engines offer many different ways to integrate data with AWS, Azure or any other CI/CD pipeline.
Parallel data generation in various test environments	Yes	In some use cases, data not only needs to be delivered in one format in one test environment but also, at the same time, in another format or database in another test environment. GenRocket Scenarios can incorporate multiple Receivers in the same Scenario so can easily accomplish this requirement.

Ability to generate data very fast and in high volume (rows per seconds) in a test environment	Yes	Fast data generation is defined in a number of ways. 1) How quickly can data be delivered to a simple unit or integration test 2) How many rows of data per second can be generated 3) how long does it take to generate millions to billions of rows of data. GenRocket can generate data for a simple unit or integration test in about 100 milliseconds, can generate on average about 10,000 rows per second, and when running multiple instances in parallel through the Partion Engine, millions to billions of rows of data in minutes.
Ability for multiple users to request and run multiple scenarios simulataneously and deliver test data on demand to the automation framework	Yes	When test data solutions are being used in larger organizations it is likely that multiple testers will have test data requirements at the same time; the solution needs to be able to scale and handle large volumes of simultaneous requests, possibly from hundreds of testers. For GenRocket, this requirement is handled by the GenRocket Multi User Server (GMUS).
Ability to read and process an API response in a test environment	Yes	Testing more complex use cases with dynamic data may require the test data system to read and process an API response. GenRocket has an extensive, well-documented API that can read and process API commands.
Ability to automatically maintain & import test data scenario and test data cases	Yes	Enterprise accounts could easily have thousands of test cases with matching test data scenarios and test data cases. It's important for the test data system to have an automated test data repository so that testers are always using the latest test data versions. GenRocket's G-Repository Server and Client automate the maintenance and importing of test data scenarios, cases, rules, queries, stories, epics and configuration files.

Dynamic Data Generation via an API	Support	Why this is Important
Real time dynamic data generation via an API	Yes	Computers can make decisions thousands of times faster than a person. Generating data dynamically, in real time via an API creates the possibily of an "electronic tester" making automated decisions and generating test data on the fly. GenRocket has a powerful, well-documented API that can be used to create an electronic tester.

MANAGING Your Test Data Environment

High performing test data platforms provide the features needed to manage, monitor and secure the test data environment

Management, Reporting & Security		
Management	Support	Why this is Important
Access controls and permissions across teams and projects	Yes	Enterprise customers want to ensure that they have control over the levels of test data system access their users have; permission controls are important when there are multiple teams working in the system. GenRocket offers extensive security control through its Team Permissions feature.
Reporting	Support	Why this is Important
Ability to view user activity and data generation activity within the test data automation platform	Yes	Managers and users of test data systems want visibility into how the system in being used. A comprehensive reporting / analytics dashboard is a necessity. GenRocket's G-Analytics dashboard provides detailed views of users activities and test data Scenario and Test Data Case runs in the platform.

Security	Support	Why this is Important
Secure architecture that does not expose any customer data in the cloud	Yes	The ideal design for a test data platform is cloud collaboration where test data scenarios can be created, modified, shared and organized with all data generation on premise for full security and performance. GenRocket's hybrid cloud /on premise architecture perfectly meets these requirements as *no customer data is stored in the cloud* and all data is generated on premise. A security assessment report by a Wall Street security expert is available upon request.
Supports Single Sign On (SSO)	Yes	Larger enterprises use SSO to reduce the number of attack surfaces because users only log in once each day and only use one set of credentials. Reducing logins to one set of credentials improves enterprise security. GenRocket supports multiple SSO vendors and methods.