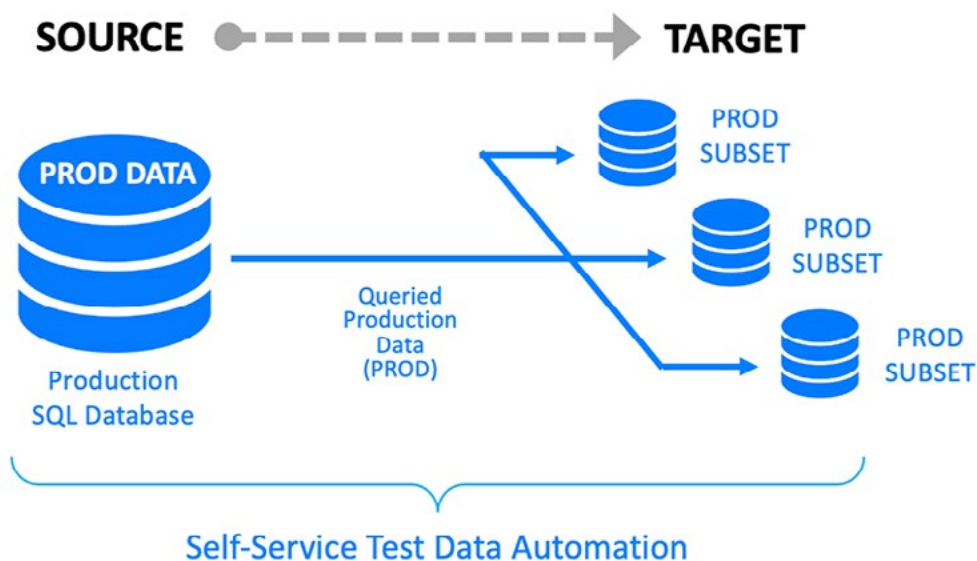


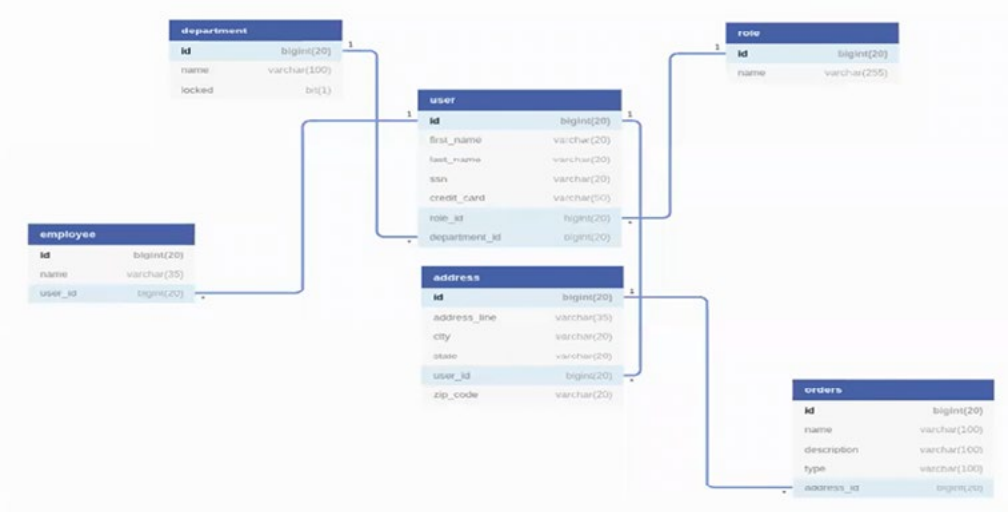
# INTELLIGENT DATA SUBSETTING

March 2021

Depending on the application and its data environment, there may be a requirement for migrating a subset of production data into a dev or test environment. GenRocket addresses this requirement with a solution called *G-Subset*, a module included in the premium, enterprise and unlimited editions of its *Test Data Automation* platform.



*G-Subset* allows testers to extract a portion of a source production database, move it to a target dataset using the identical data model and format, and provision the data for automated testing during continuous integration. A data subset may represent all data tables in the source database or a specific family of related data tables. *G-Subset* provides testers with the ability to specify the tables to be retrieved and the rules for migrating data based on the values in table columns. And *G-Subset* ensures that referential integrity between parent, child and sibling relationships is always maintained during the subsetting process.

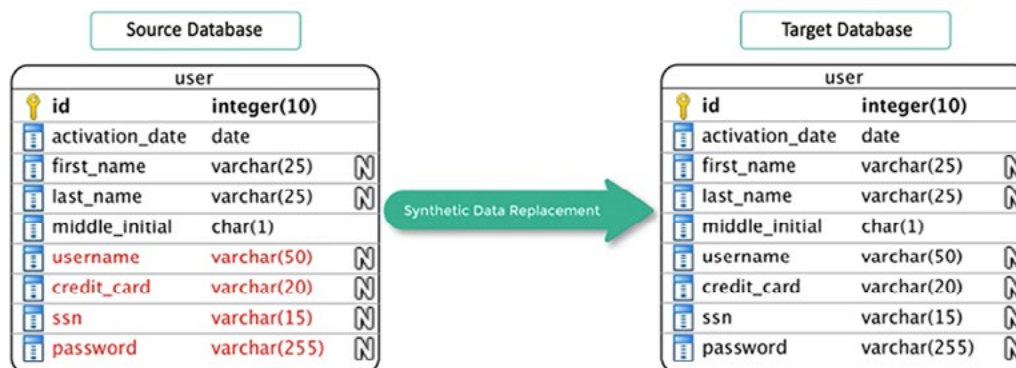


**Intelligent Data Subsetting** is a powerful concept that allows testers to reduce the size of a massive database down to a manageable data subset, making it faster and easier for testers to find the data needed for their test cases. For example, a database containing millions of transactions for customers in all 50 United States can be subsetting to contain transactions from just one or two states. The size of the data subset will be a small percentage of the source database while providing a consistent representation with the same referentially intact data tables.

## Synthetic Data Replacement

*G-Subset* also allows testers to take sensitive data elements in the production data subset and replace them with 100% secure synthetic data. This real-time synthetic data replacement (SDR) process addresses the need for masking sensitive production data to ensure data privacy as it accelerates test data provisioning.

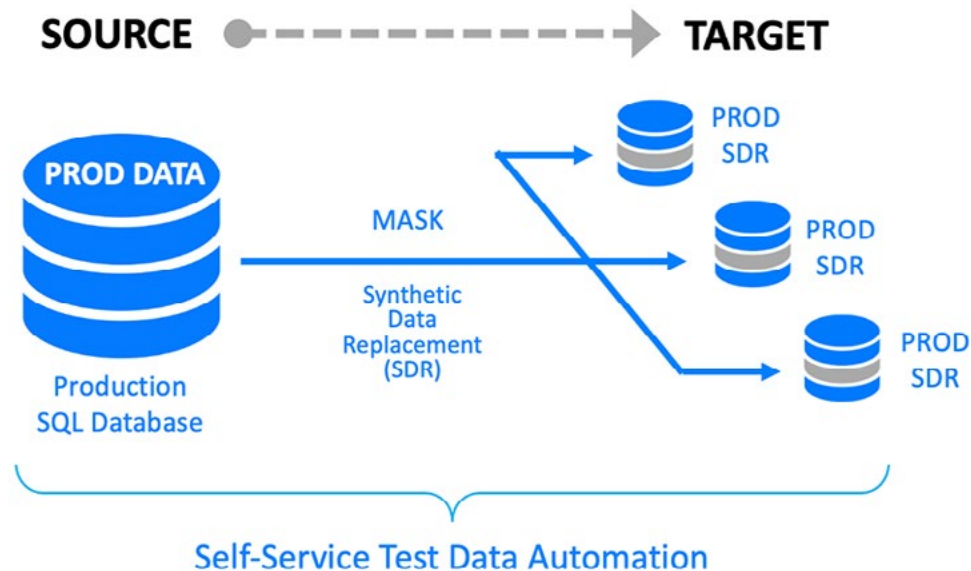
### Subset & Mask Using Synthetic Data Replacement



The following steps occur during the data subsetting process:

- Non-sensitive data is obtained from the production database using queries
- Synthetic data is generated and merged with the non-sensitive production data for each specified table and column
- Merged data is inserted into the testing database using the same data model and database format

With G-Subset, sensitive data (PII, PHI) is never touched, stored or copied.



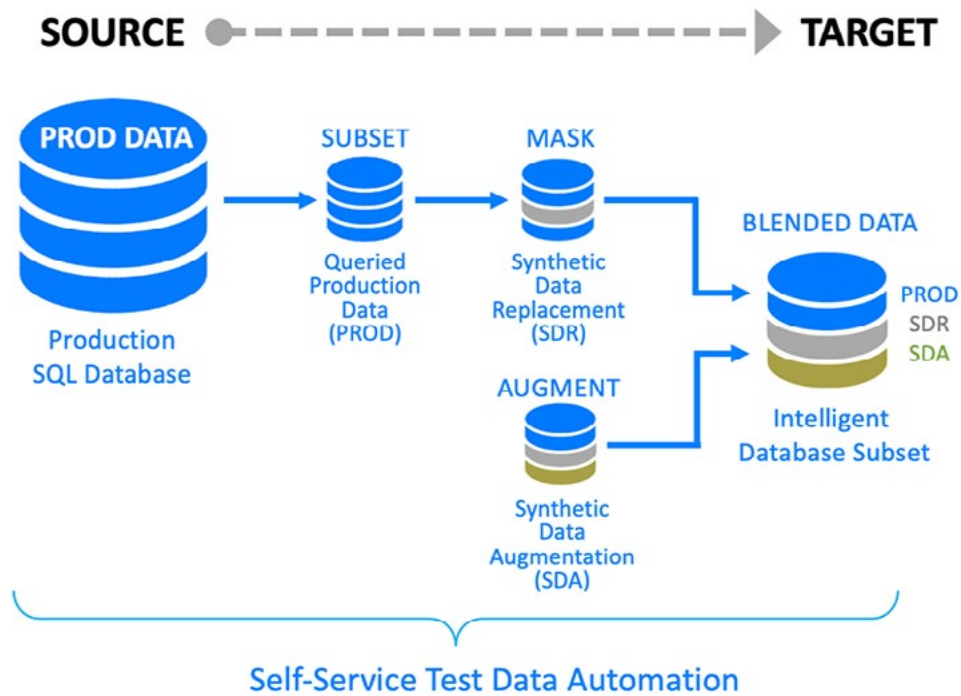
## Synthetic Data Augmentation

New applications, or existing applications with new features, may not have production data readily available for testing. In this case, G-Subset provides testers with the ability to augment production data with synthetic data to provide a complete data subset. Simply import the data model for the required synthetic data, specify the range of values and the volume of data required, and generate controlled synthetic data in combination with the subsetting process.

With the addition of synthetic test data, test coverage can be maximized in ways not possible with only production data values. This allows testing for both positive and negative conditions, edge cases, or any combination or permutation of data values. Here are some examples:

- Data combinations for account numbers, verification codes & expiration dates
- Business logic that controls reward point values based on cumulative spending
- Edge cases for validating credit approval based on transaction amount & credit limit
- Scalable data volume to simulate a peak transaction load for performance testing





Synthetic data augmentation is a practical way of introducing synthetic data into a QA environment. Testers accustomed to working exclusively with production data can gradually introduce synthetic data into their testing regimen to ensure data privacy as they increase test coverage. The result is the best of both worlds: The familiarity of working with existing and realistic production data combined with the flexibility of generating synthetic data to provision a more complete test data subset.

Synthetic data augmentation extends the power of traditional data subsetting and masking. With *G-Subset*, testers now have total control over the security, accuracy, variety and volume of the test data needed for any type of testing. And with GenRocket's *Intelligent Data Subsetting* solution, data refresh is as simple as dumping the database and re-running *G-Subset*. A fresh copy of any data subset that is needed is provisioned *on demand*.



## The Benefits of Intelligent Data Subsetting

GenRocket's *Intelligent Data Subsetting* solution brings the power of test data automation to the production data subsetting process. A brief summary of its benefits include:

**Model-based:** All queried production data and blended synthetic data is model-based to ensure the accuracy of the data structure and referential integrity of data relationships.

**Secure:** Sensitive production data is replaced with 100% synthetic data to provide total data privacy and allow control over the variations of data needed for testing.

**Intelligent:** Production data can be easily blended with synthetic data to fully meet test objectives.

**Self Service:** GenRocket provides an accessible self service platform with appropriate security controls for testers to create their own data subsets whenever they need them.

**Scalable:** GenRocket offers enterprise-class management of all test data designs and configuration files with scalable performance to generate millions to billions of rows of test data at the speed required by Agile and DevOps environments.

### SUPPORTING THE LEADING DATABASE TECHNOLOGIES

ORACLE

Microsoft  
SQL Server

IBM DB2

[IN DEVELOPMENT]

PostgreSQL

MySQL

MAKE A REQUEST  
FOR YOUR  
DATABASE SUPPORT

REQUEST A DEMO