PRODUCTION DATA vs. SYNTHETIC DATA

6 Essential Criteria for Comparing the Value of Test Data

Software development cycles are getting faster and code is becoming increasingly modular. This puts new demands on software testing to operate at the speed of development for all levels of testing. Otherwise it becomes a bottleneck, or worse, fails to fin critical software defects prior to deployment. This has created a big push for testing automation and the implementation of CI/CD pipelines to better align software development and testing processes.

HERE'S THE PROBLEM

The current approach to provisioning test data is not able to keep up with this pace. The traditional approach for provisioning test data is out of date. Subsetting production data takes too long, limits data variations, and exposes sensitive customer data – in spite of best intentions to mask and obfuscate. Managing production test data has become overly complicated and costly. Centralized test data management platforms are difficult to learn and operate.

The current approach to provisioning test data is simply not able to keep up with the pace of development.

For many organizations, reducing cost and complexity means testers get pruned production data and modify it by hand, or create a few rows of data manually and reuse them. This severely limits the level of testing that can be done.

IT'S TIME TO THINK DIFFERENTLY ABOUT TEST DATA

What if synthetic test data could do everything real test data can do, only better and faster?

This was GenRocket's vision 7 years ago when it embarked on a mission to build a platform for generating any kind of synthetic test data on- demand. We determined that synthetic test data not only had to be every bit as good as real production data, it had to be even better.

We created a new standard for synthetic data we call Real-Time Synthetic Test Data.

We developed a list of essential criteria for test data to measure our standard for synthetic test data alongside of real production test data.

6 ESSENTIAL TEST DATA CRITERIA

SPEED
1000% Faster Provisioning

COST
10% of the Cost

QUALITY
Controlled, Accurate, Complete

SECURITY
Assured Data Privacy

SIMPLICITY
Easy to Learn and Use

VERSATILITY
Flexible Architecture

HERE ARE THE 6 ESSENTIAL CRITERIA FOR TEST DATA

CRITERIA #1 IS SPEED

GenRocket's Real-Time Synthetic Test Data is generated in a few seconds or minutes not hours or days. Our benchmarks show test data can be generated 1000% faster than provisioning production data through traditional methods.

CRITERIA #2 IS COST

Using production data requires a costly TDM platform – either purchased or home grown. But here's the thing, if you're generating Real-Time Synthetic Test Data on demand, you don't need a TDM to manage your test data, just generate what you need when you want it!

CRITERIA #3 IS QUALITY

Production data is fully representational data, but If you want to find software defects in all expected and unexpected scenarios, your test case will require all data variations to test the edge cases. Real-Time Synthetic Test Data can be generated for every conceivable data scenario.

CRITERIA #4 IS SECURITY

Privacy breaches are unacceptable and full compliance with security protocols is a must. But if you are using production data, there will always be a risk of exposing sensitive user data. Synthetic test data is not real data so it eliminates risk and is 100% secure.

CRITERIA #5 IS SIMPLICITY

Testing software is complicated, but creating test data doesn't have to be. Real-Time Synthetic Test Data Generation replaces a complex, centralized TDM process with an easy, decentralized self-service model that any tester or developer can learn in a few days.
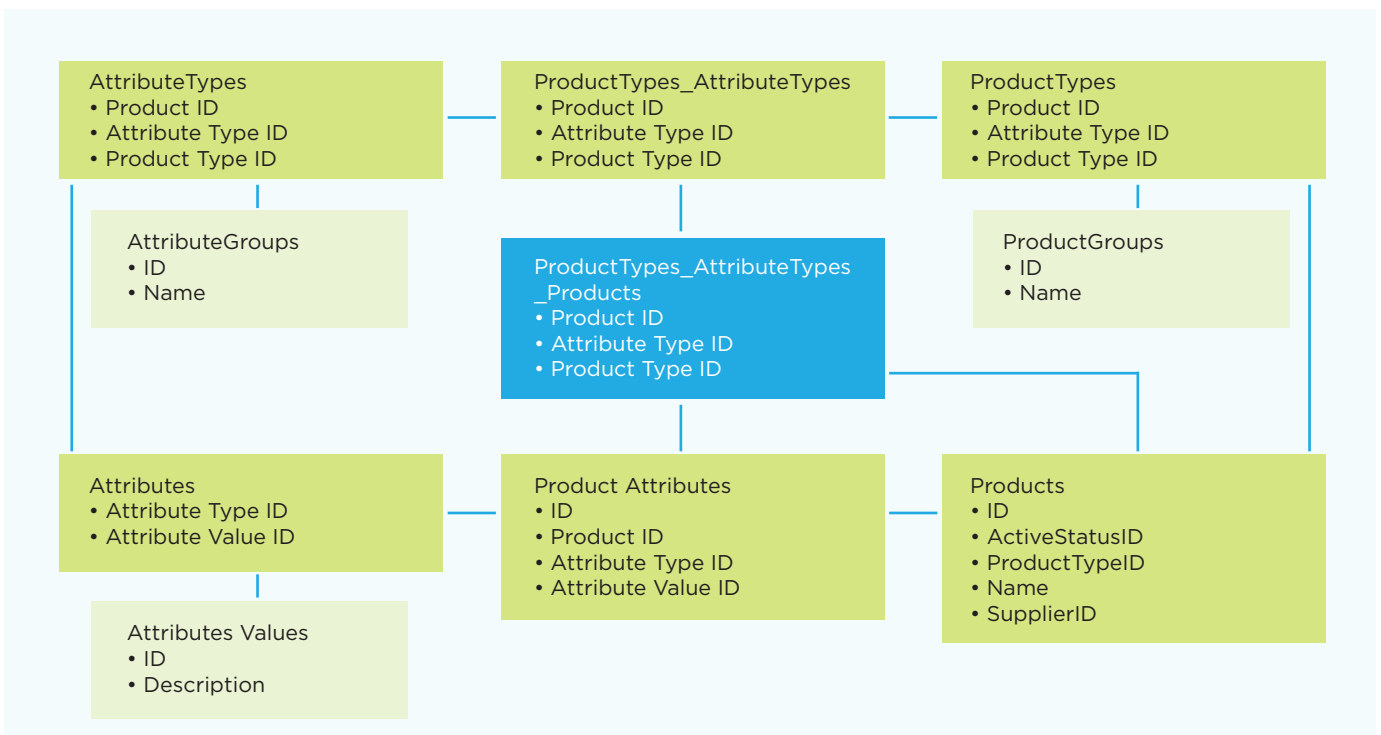
CRITERIA #6 IS VERSATILITY

Every organization has different test data requirements. GenRocket was designed with this in mind and can deliver test data in just about any format to testing tools, databases or frameworks.

Pattern	Realistic	Sequential	Random	Edge Case	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

Data Variations for test data must cover every data scenario in order to discover defects for both expected and unexpected outcomes.

AttributeTypes
• Product ID
• Attribute Type ID
• Product Type ID

ProductTypes_AttributeTypes
• Product ID
• Attribute Type ID
• Product Type ID

ProductTypes
• Product ID
• Attribute Type ID
• Product Type ID

AttributeGroups
• ID
• Name

ProductTypes_AttributeTypes_Products
• Product ID
• Attribute Type ID
• Product Type ID

ProductGroups
• ID
• Name

Attributes
• Attribute Type ID
• Attribute Value ID

Product Attributes
• ID
• Product ID
• Attribute Type ID
• Attribute Value ID

Products
• ID
• ActiveStatusID
• ProductTypeID
• Name
• SupplierID

Attributes Values
• ID
• Description

Referential Integrity among test data tables must be ensured for all parent-child relationships.