UST Global Approach on Test Data Management using Synthetic Test Data Generators
Test Data Management (TDM) deals with creation of non-production data sets that mimic an organization's actual data. This necessitates that test requirements must be satisfied as documented. Organizations embrace new technologies such as cloud, database as a service (DBaaS), and big data to meet the growing demand for data insights in real-time, from new data sources in a variety of formats.

According to a Gartner report, "The average organization loses $14.2 million annually through poor Data Quality". Data professionals spend a large amount of time and effort to manage large and diverse data stores. With an increase in demand from users and restrictions on budget, there will be a surge in market-place requirements to automate data management tasks.

Test data should be constructed like real data but manipulated in such a way that the underlying real world items are not identifiable. For testing purposes, the provisioned data needs to be of optimum size, neither a huge volume of production data nor a small unrepresentative data sample. Different techniques can produce these data packs:

- **Production Cut**
  Identify and extract test data from production and load into the test environment

- **Data Upload using SQL Scripts**
  Create and maintain SQL Scripts that aids data upload to test environments

- **Manual Generation of Test Data**
  Create test data using manual methods, well defined processes and proprietary utilities

- **Automated Test Data Generation**
  Automated synthetic test data creation using tools or automation scripts

Overall, Test Data Management begins when a software project begins and terminates only when the project comes to an end. This whitepaper explains the effectiveness of synthetic test data generation in TDM and evaluates an appropriate tool to automate it.

As per the World Quality Report, "44% of organizations face challenges in maintaining the right test data set versions with different test versions". Synthetic data based on well-defined data models allow patterned and conditioned data to be accurately provisioned for virtually any test case scenario. This is the main reason that drives testing teams to go “Synthetic”. Test data can be modelled for future systems as well, thereby ensuring future test requirements are met. Synthetic test data also ensures data privacy; going forward, organizations will have to strictly adhere to regulatory criteria and deliver secured, quality products. Agile development environments can save time mocking up and synthesizing data instead of manually creating it. A test data generation tool automatically creates millions of rows of interrelated test data, based on the test criteria, in minutes. Also, with synthetic generation, test data give emphasis to both quality and quantity. Optimized intelligent data sets reduce testing costs drastically. The data synthesized will have all the characteristics of a live database but none of the sensitive content. This avoids data breach in testing and development.

For functional testing needs, synthetic data generators offer capabilities for the design, generation, provisioning and management of test data. It overcomes the challenges of test data instability and consumption in diverse industries.

In non-functional test scenarios, only huge amounts of realistic data can capture data-driven performance issues. Here, an automated approach is preferred to manual data generation as automation ensures that there is diversification in data while applications are performance tested at scale. It maintains the complexity of test data and produces highly accurate performance measurements.
Business Challenges

Testers face many challenges in obtaining the right set of data for testing due to many reasons.

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertain volume and format specifications</td>
<td>Establish parent-child relations and RI between tables</td>
</tr>
<tr>
<td>Inadequate knowledge to produce right set of data</td>
<td>Inadequate knowledge to produce right set of data</td>
</tr>
<tr>
<td>Partial test data coverage</td>
<td>Gather right test data from different teams</td>
</tr>
<tr>
<td>Absence of clear data requirements</td>
<td>Production data which are inapt to test business scenarios</td>
</tr>
<tr>
<td>Restricted access to data sources</td>
<td>Lack of data generation tool usage</td>
</tr>
<tr>
<td>Demand for quick data readiness</td>
<td>GenRocket - a powerful tool for Synthetic test data generation defines test data as it relates to requirements and test cases.</td>
</tr>
</tbody>
</table>

These challenges may lead to testing with an inadequate set of data which in turn causes defect slippage and major risk to production. Appropriate Test Data Management tools offer many effective solutions to overcome this. ‘GenRocket’ - a powerful tool for Synthetic test data generation defines test data as it relates to requirements and test cases.

Solution

An efficient synthetic data generation tool offers many benefits over manual test data creation techniques. GenRocket performs smart identification of business scenarios and the creation of matching data sets. It improves the efficiency of testing by standardizing the method of test data preparation. The tool ensures consistent data creation procedures and masking techniques across different teams. If data quality is a potential area of concern in a business, it can be identified and resolved at a faster pace. The structuring and version control capabilities of GenRocket maximize test data coverage and reusability. It can eliminate data errors and data corruption by generating data only when needed. Stale data need not be stored and the data storage cost can be minimized.

Benefits of using GenRocket

GenRocket is an on-demand, real-time synthetic data generator that addresses all key challenges encountered in test data generation. The tool is accessed through a web browser and generates test data in real time with parent-child relationships. It generates realistic, random and conditioned test data at high speed for a variety of test applications including integration, functional and load tests. This transforms the time-intensive process of test data generation for the Agile/DevOps practitioner by integrating test data generation with continuous testing to ensure that the software is tested completely. GenRocket’s smart data modelling and generation techniques provide software developers and testers with the comprehensive test data they need to fully test their software. Once data models are defined, users can download the test data sets from anywhere through the web interface. Hence different teams in a project can use the same data and avoid any duplication of efforts.

How the tool works

The GenRocket user interface makes it easy to model, update and generate data. The test data is generated in different formats – XML, JSON, SQL, and CSV. The tool uses a component based design to produce test data. The five major components are: Domains, Attributes, Generators, Receivers and Scenarios. Initially, a project is defined and Domains (or tables) are built from DDL, CSV or existing preset values. Other components such as Attributes define the characteristics of a Domain and Generators generate specific types of data. Receivers receive the data generated from a Generator and
transforms the data into usable format (XML, SQL, JSON, REST etc.). Ultimately, Scenarios are built to provide instructions to select one or more Domains to generate data. Users define Global Variables within Scenarios and these are shared globally with all other Domains in a specific project. In GenRocket, users establish referential relationships between Domains. This enables the creation of Scenarios that contain related Domains used to generate data with referential integrity.

The Scenarios modelled can be downloaded locally from anywhere and then run to generate test data. This innovative and powerful tool facilitates automated testing with real-time and model-driven test data.

**Solution Components**

Some of the key differentiators that the tool offers are:

- 100+ Generators
- 20+ Receivers
- Create Referential Integrity between Domains
- Implement Simple or Complex Business Logic
- Create Scenarios on the Web and Run Scenarios Locally
- Fast to Generate, Change and Update Test Data
- Build Quality Data On-demand and in High Volume

**Conclusion**

Today’s digital business environment demands reduced time to market with shorter software development cycles. This stresses the testing process and creates a need for continuous testing backed by on-demand, real-time test data. Identifying a reliable automated solution to build and manage test data has become extremely important.

GenRocket allows software engineering teams to easily share models of required data or import existing data models and schemas. They can create their own Scenarios and decide how they want test data to be generated. Most important, the solution is user-friendly, economical and fills the need for dynamic testing environments. The speed and quality of testing is dramatically accelerated by its intelligent techniques for generating realistic, patterned, negative and conditioned test data sets.

**References:**

1) Blogs and online references for Test Data Management and Generation
   - Back to basics: Fundamentals of test data management
   - [https://www.sticky minds.com/article/synthesize-your-test-data](https://www.sticky minds.com/article/synthesize-your-test-data)
2) [www.genrocket.com](http://www.genrocket.com) - Official Website of GenRocket
ABOUT UST GLOBAL®

UST Global® is a fast-growing company that provides advanced computing and digital services to large private and public enterprises around the world. Driven by a larger purpose of Transforming Lives, and the philosophy of “fewer Clients, more Attention”, we bring in an entrepreneurial spirit that seeks the fastest path to value in today’s digital economy. Our innovative technology solutions and pioneering social programs make us stand apart. Our clients include Fortune 500 companies in Banking and Financial Services, Healthcare, Insurance, Retail, High Technology, Manufacturing, Shipping, and Telecom. We believe in building long-lasting, strategic business relationships through agile and client-centric global engagement models, combining expert on-site and local resources with cost, scale, and quality advantages of offshore operations.

For more information, please visit www.ust-global.com

ABOUT THE AUTHOR

Sreeja Raju is a Senior Test Analyst in the Digital Assurance practice and has more than 10 years of testing industry experience. She specializes in test design, SOA testing and test automation. She is an ISTQB Advanced Level Certified professional and has deep Domain expertise in Biometrics, Retail and CRM.