

BUILDING TEST DATA THAT DELIVERS ON THE PROMISE OF AGILE DEVELOPMENT

PeerPaper Report



BASED ON ACTUAL USER
EXPERIENCES & OPINIONS

ABSTRACT

The process of identifying, creating and delivering test data has long been a time- and skills-intensive drain on QA and test department resources. The emergence of agile software development methodologies and new, related architectures like web services and containers, has compounded the problem. Not only are the testing cycle times significantly compressed, the interdependencies between applications and data sources in testing have also become more complex. This paper explores how test data automation solutions can help testing teams speed up the test data creation process in order to keep up with new agile requirements. It draws on product reviews from experienced testing professionals published on IT Central Station.

CONTENTS

- Page 1.** Introduction
- Difficulties in Traditional Test Data Creation
- Page 2.** The Impact of Agile Development on Testing and Test Data Creation
- The Potential of Test Data Automation Solutions in Agile Software Testing
- Page 3.** Success Factors in a Test Data Automation Tool for Agile Development
- Finding the Right Data
- Speed of Data Generation
- Page 4.** Ability to Create Synthetic Data/Subsetting
- Data Masking and Compliance Issues
- Page 5.** Scalability
- Service Virtualization Simulation
- Page 6.** Demonstrable Return on Investment (ROI)
- Technical Support
- Page 7.** Conclusion

INTRODUCTION

Software testing requires test data, often quite a lot of it. Producing that data can be quite time consuming. Identifying, creating and managing test data has long been a time- and skills-intensive drain on QA and test department resources. In addition, the increasing prevalence of agile software development methodologies and new, related architectures like web services and containers, has made the issue even more urgent. Cycle times between tests get compressed. Interdependencies between applications and data sources also grow more complex. Automated test data generation tools offer a solution. This paper explores the potential of this technology. Using reviews from actual software testing professionals on IT Central Station, it looks at how Test Data Management (TDM) tools can improve the test data creation process.

Difficulties in Traditional Test Data Creation

The effectiveness of automated software testing depends on several factors. These include the quality of the test case parameters, scripts and the test data used to simulate the software's functionality. The better the test data, generally, the more accurate the test. For example, if a database's performance is affected by variability in the contents of certain fields, the test data should emulate that variability. Otherwise, the test may show performance results that are not accurate.

Test data may take the form of SQL test data, system test data, performance test data or xml, to name a few types. Methods of creating it include cloning

production data and manually "masking" it as well as creating new data from scratch. This is known as "synthetic" data generation. Manual test data creation, however, is viewed as a notorious drain on time and resources.

Most testers use actual live production data for testing. The production data is copied and masked prior to be used in a test. This approach offers certain benefits, such as realism in seeing how a piece of software will handle the data it is meant to use. Production data presents a host of challenges, however. The volume of it may be too large to handle in a testing environment. Testing may interfere with

regular workloads. Production data does not factor in outliers, negative test scenarios or other “what if” scenarios. In other words, sometimes production data is too good and doesn’t break the software. And, there are security risks inherent in the practice. The testing environment may not be secure enough for confidential information contained in the data even when masking has occurred, for example. Using production data in testing may even be illegal in certain places due to privacy laws.

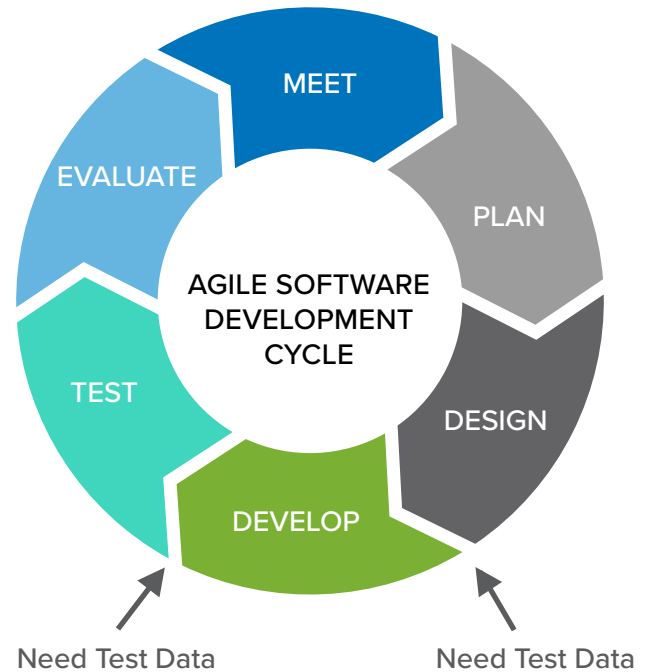
The Impact of Agile Development on Testing and Test Data Creation

Agile development methodologies further compound TDM challenges. With rapid cycles like the two-week sprint, agile development routinely reconfigures the testing requirements. Test data created for the first iteration of code may become obsolete as the latest sprint adds more functionality to the code base for testing. When the process involves Continuous Delivery (CD) or Continuous Integration of new code, the application being tested is both live in production and needing to be retested frequently. The old data might not result in an effective test so testers have to create a new batch of test data—back to the already onerous grindstone of data generation and management.

The Potential of Test Data Automation Solutions in Agile Software Testing

Responding to the challenges inherent in test data creation, a number of vendors have produced software that automates software test data generation. These tools are able to mass-produce test data based on testing requirements. For instance, if a test needs 10,000 database records

containing name, address and phone numbers of fictitious customers, the TDM solution can generate this quite quickly using the synthetic approach. Synthetic TDM tooling is often sophisticated, with the ability to simulate variability in data that might appear in production.



The advantages of TDM in agile environments are quite clear and compelling. Using TDM tools, a testing team can generate suitable test data in the same tempo as the agile, CD or CI cycles. For instance, if a sprint results in additional fields being added to a database, the TDM tool can generate synthetic data with those new fields. As a [Senior Test Data Management Specialist](#) at a transportation company put it, “TDM has tons of great solutions involved in one package. For me personally, what I find to be most valuable is its ability to do synthetic data creation. I love that because it has a lot of flexibility and you do not have to worry about one specific database or how you are going to manage all the data points. What you can do is instead of taking everything from production and wondering what you are going to get from there, you can just create it all from the get-go yourself. That is a beautiful thing to be able to do.”

Success Factors in a Test Data Automation Tool for Agile Development

Of course, nothing is quite that simple, as real-life test automation users on IT Central Station reveal. Test automation users on IT Central Station share their insights into what makes the best test data automation solution. In many cases, identifying the right TDM solution involves thinking across the entire IT department as well as the broader organization.

A [Sr Test Manager](#) at a transportation company explained, “We want to work with somebody that’s going to help us solve the problem that we’re trying to solve. Do it in a way that works for our enterprise. We don’t want to do something that only works for one team or one solution, something that’s scalable. We have a lot of different technologies in our organization, so we need something that will work with all of them.” In the spirit of this insight, IT Central Station users highlight the following TDM success factors:

FINDING THE RIGHT DATA

Testers spend lot of time trying to find the right data. They search across databases or other systems looking for specific data that can be used in their test. They enlist the help of database administrators and other test data engineers. Without the right tools, testers can spend hours upon hours searching for the right data. Some testers resort to manually building their own test data if none can be found on existing systems. This is where TDM solutions can help speed the process.

An [IT Manager](#) at an energy company stated “The benefits are that TDM allows us to find the right test data for the test that we need, and then it also allows us to get the required data inputs into our API test, so that we can do a full test.” Testing is a complex

undertaking and without the right tools companies will be hard pressed to achieve their continuous testing goals.

SPEED OF DATA GENERATION

Speed counts, especially when the testing team needs large volumes of data for automated testing. A [Practice Manager](#) (Testing Services) at a financial services firm with over 1,000 employees said, “In terms of the speed to market, because we don’t manually produce data anymore, we use intelligent profiling techniques, test data matching, we massively reduce the time we spend finding data, and we also can produce data on the fly, which turns around test data cycles. In terms of cost, because we’re doing it a lot quicker, it’s a lot cheaper.”

He added to this by noting, “Within our organization we have many, many platforms, many, many different technologies. One of the interesting challenges we always have is in terms of, especially when we’re doing performance testing, can we get the kind of the volumes of data in sufficient times, and we use things like data explosion quite often and it does what it needs to do and it does it very quickly.”

A [Practice Leader](#) for DevOps at an IT consultancy described his TDM tool by saying, “[It] is enormously helpful to us. We assist our customers by speeding up the application development process using real-time test data and synthetic test data, which mimics the real test data.” The [Sr Test Manager](#) at the transportation company said, “It’s really allowed us to focus on the craft of testing, and not focus on the creating data, spending time setting up the scenario that we need, or trying to find that sort of thing. It’s allowed us to be more focused on our efforts, it’s allowed us to be faster.” A [Senior Specialist](#) at an automotive publishing company added, “It’s just always a much more dynamic world that we can react a lot faster to, and attribute most all of that to [our TDM tool].”

Some testers want the ability to have complete instances of databases made available to each of their testers. Data Virtualization is the term used to describe this process. Some TDM solutions provide virtual test data cloning. One [user in the transportation industry](#) said, “database cloning, the ability for them to have full control of their environment is the most important aspect of testing sometimes. Now, you do not have to worry about what is on your left, what is on your right, and who you are going to be hurting by trying to do the best testing that you can. Instead you have just you are own set that you can work with. You can spin it up and burn it down when you are done.”

ABILITY TO CREATE SYNTHETIC DATA/ SUBSETTING

Synthetic data is generated to resemble the production data the software will eventually use. A TDM tool must have extensive synthetic data creation capabilities. An [AVP Quality Assurance](#) at a large financial firm commented on this, saying, “I think from my synthetic generation, a lot of times generating synthetic data can be cumbersome. TDM, with some of the rules aspect of it, you can generate it and have your rules in place that you know your data’s going to be very consistent.” Load size matters in this respect. As a [Quality Assurance](#) at a logistics company with over 1,000 employees said, “With synthetic data generation, we can test applications with three or four times the production load.”

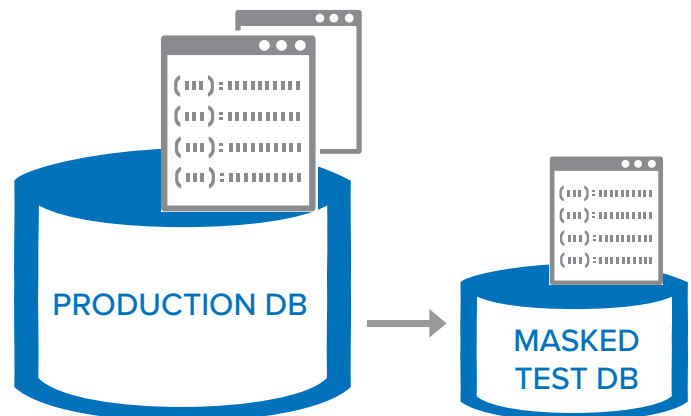
A [COE Consultant Test](#) at a financial services firm with over 1,000 employees valued his TDM tool’s synthetic data generation capabilities while a [Senior Test Data Management Specialist](#) at an engineering company with 1,001-5,000 employees was pleased with his tool’s synthetic data creation featuring flexible built in data functions. An [IT Manager](#) at a financial services put synthetic data into a broader context, saying, “It doesn’t only cover our data

masking needs, but also our data subsetting and synthetic data generation needs. Finally, it gave us the idea of making test data as a service within our organization.”

Testing also often requires working on data subsets. As a result, subsetting features are highly valued in TDM tools. The [IT Manager](#) at the financial services firm explained that he valued his TDM tool’s “Data masking, subsetting, and synthetic data generation.”

DATA MASKING AND COMPLIANCE ISSUES

For testers who do automated testing with production data, it’s necessary in most cases to mask any sensitive information that might be contained in the records. For example, a production database with social security numbers needs to be masked so the testing process won’t violate any regulatory compliance rules.



The [IT Manager](#) at the financial services firm noted, “We started the TDM journey due to our data masking needs.” The [Senior Specialist](#) at the automotive publisher praised his TDM tool by saying, “The data masking is a powerful aspect of the tool ...A lot of people, when they first started looking at the tool, started immediately jumping in and looking at the data masking, the data subsetting that it can do, and it works fantastically to help with the compliance issues for masking their data. That’s a very powerful aspect of the tool.”

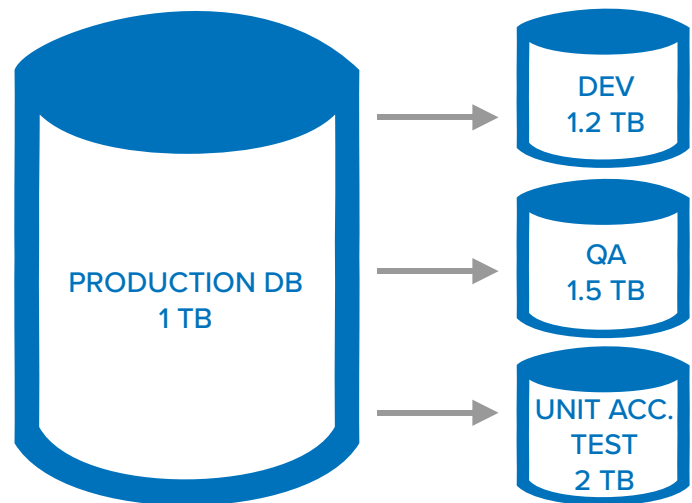
The [COE Consultant Test](#) said, “Using data masking techniques we can comply the rule of non-disclosure of personally identifiable information.” The [Practice Leader](#) of DevOps added, “The most valuable features for us are masking, data profiling, and creating data subsets. More specifically, we are able to assist our clients with data privacy and the regulatory recommendations that come from the government. We help them to comply with PI, IP, HI and PCI regulations.”

Recent data breaches have highlighted the importance of safeguarding all types of data. Production data and data masked for testing must be protected equally. The ability for a TDM solution to profile and identify PII data prior to using it for testing gives organizations a head start on demonstrating compliance for any regulation.

SCALABILITY

IT Central Station users point out that testing sometimes means testing data loads larger than what is even available in the production database. TDM scalability is essential. A [Quality Assurance](#) at a logistics company with more than 1,000 employees said, “It [our TDM tool] has really changed the culture in the company because nobody could ever imagine generating millions of records. Even production systems have just a couple of million records. When you want to test your applications with three or four times the production load, you can never actually achieve it because there is no other way besides synthetic data generation. You can’t have that volume of data in your DBs. Even if you subset your entire production, you would get just one X of it. To get three or four X of it, you have to go to either data cloning or to synthetic data generation.

The [Senior Specialist](#) at the automotive publisher praised his TDM tool by saying, “It has scaled



tremendously. Especially, again, I don’t want to harp back too much on it, but when you start looking at data generation, your options are endless in the way you want to incorporate that into your environment.”

A [Client Partner](#) at a mid-sized financial services firm noted, “I think one of the main features of TDM is how you can scale from a small organization; how you can use it in a very big organization. In our company, that is everything. Only because of that very feature, scalability, we are considering TDM.” The [Senior Test Data Management Specialist](#) at the engineering company said, “The many databases supported and data delivery formats available provide a seemingly endless supply of options to meet the ever growing demand of our testing teams.”

SERVICE VIRTUALIZATION SIMULATION

The emergence of APIs and virtual services has created a demand for testing tools that fit these criteria. The [AVP Quality Assurance](#) at the financial firm explained, “We intend on taking that, partnering it with the SV solution and being able to generate the data for the service virtualization aspect.” He added, “I also push the service virtualization record to respond to the request of the loan, hitting the credit bureau, returning a certain credit score, which then gets us within that target zone for that loan we’re looking for, to trigger a rule.”

The [Senior Specialist](#) at the automotive publisher said, “It’s not something that I would often give, but I do give this a perfect rating. We’ve been able to solve any of the data issues that we were having initially when we first brought it in, and it’s expanded everything that we can do as we looked into the future right now of where we want to go with this. That includes its tie-ins for service virtualization; that includes the way that we can build out our environments in a way that we’d never considered before. It’s just always a much more dynamic world that we can react a lot faster to, and attribute most all of that to Test Data Manager.”

DEMONSTRABLE RETURN ON INVESTMENT (ROI)

Saving time on test data automation translates into financial savings, or at least it should. Several IT Central Station members explained the ROI impacts of their TDM tools. The [Sr Test Manager](#) at the transportation company commented, “The return on investment for us has been great. In the last year, we captured some metrics around our ROI, and we’ve saved over eleven thousand hours in manual time of trying to create test data. In that same time, that eleven thousand hours, has translated to about eleven million dollars in cost savings, just from that. Additional benefits around being able to shorten our test cycle time, allow us to go deeper and further into our testing and freeing up people to actually do the work that they’re paid to do, and not being doing things like creating data so that they can actually do the work.

The [Senior Specialist](#) at the automotive publisher said, “When I look at the return on investment, there are not only huge financial gains on it. In fact, when I recently ran the numbers, we had about \$1.1 million in savings on just the financials from 2016 alone. What it came down to is, when we started creating our data using Test Data Manager, we reduced our hours used by about 11,800 in 2016. That’s real time. That’s a significant, tangible benefit to the company.”

TECHNICAL SUPPORT

The complexities of TDM make technical support an important aspect of the vendor selection process. The [COE Consultant Test](#) gave his TDM vendor a score of 9 out of 10, noting, “All our queries and functional defects were resolved within very little time. [The vendor] technical support people are proactive and we get the fixes in very little time.” The [AVP Quality Assurance](#) at the financial firm was also pleased with his TDM vendor technical support.

In the best case scenario, there is a true relationship between the customer and TDM vendor. The [Senior Specialist](#) at the automotive publisher put it this way: “Everyone on [the vendor’s] side has been there to support on any kind of small bug or enhancement that might come up along the way.” The [Practice Manager](#) (Testing Services) at the financial services firm commented, “We work in an organization where we use many tools from many different suppliers. I think that the kind of a relationship that my organization has with CA is kind of a much richer one in terms of, you know, it’s not just a tool support.”

CONCLUSION

IT Central Station members who work in software testing are embracing TDM. The technology saves time and facilitates complex testing in a fast-moving agile world. TDM options abound, however, so reviewers point out what matters to them. The best test data results come solutions that are scalable and effect a strong, demonstrable ROI. Organizations are looking for complete test data management solutions, not just a single aspect of managing test data. Users want extensive and flexible synthetic data creation as well as pure speed—the better to quickly generate data volumes that simulate heavy load levels. They want subsetting and masking and the ability to find existing data. The work is sufficiently sophisticated, though, that customers value more than just features. They want strong technical support. They want a vendor they can treat as a trusted advisor.

ABOUT IT CENTRAL STATION

User reviews, candid discussions, and more for enterprise technology professionals.

The Internet has completely changed the way we make buying decisions. We now use ratings and review sites to see what other real users think before we buy electronics, book a hotel, visit a doctor or choose a restaurant. But in the world of enterprise technology, most of the information online and in your inbox comes from vendors but what you really want is objective information from other users. IT Central Station provides technology professionals with a community platform to share information about enterprise solutions.

IT Central Station is committed to offering user-contributed information that is valuable, objective and relevant. We validate all reviewers with a triple authentication process, and protect your privacy by providing an environment where you can post anonymously and freely express your views. As a result, the community becomes a valuable resource, ensuring you get access to the right information and connect to the right people, whenever you need it.

www.itcentralstation.com

IT Central Station does not endorse or recommend any products or services. The views and opinions of reviewers quoted in this document, IT Central Station websites, and IT Central Station materials do not reflect the opinions of IT Central Station.

ABOUT CA TECHNOLOGIES

CA Test Data Manager provides a solution to one of the most time-consuming and resource-intensive problems in Continuous Delivery—the identifying, maintaining and provisioning of test data needed to rigorously test evolving applications within a 2 week sprint. CA Test Data Manager uniquely combines elements of data identification, data subsetting, data masking, on demand synthetic and virtual data creation, to help testing teams achieve continuous testing.

To learn more, go to: ca.com