



BIG DATA TESTING

BY REGG STRUYK

Europe, Middle-East, Africa: Polarion Software GmbH
Kesselstraße 19 — 70327 Stuttgart, GERMANY
Tel +49 711 489 9969 - 0
Fax +49 711 489 9969 - 20
www.polarion.com - info@polarion.com

Americas & Asia-Pacific: Polarion Software, Inc.
1001 Marina Village Parkway, Suite 403, Alameda, CA 94501, USA
Tel +1 877 572 4005
Fax +1 510 814 9983
www.polarion.com - info@polarion.com

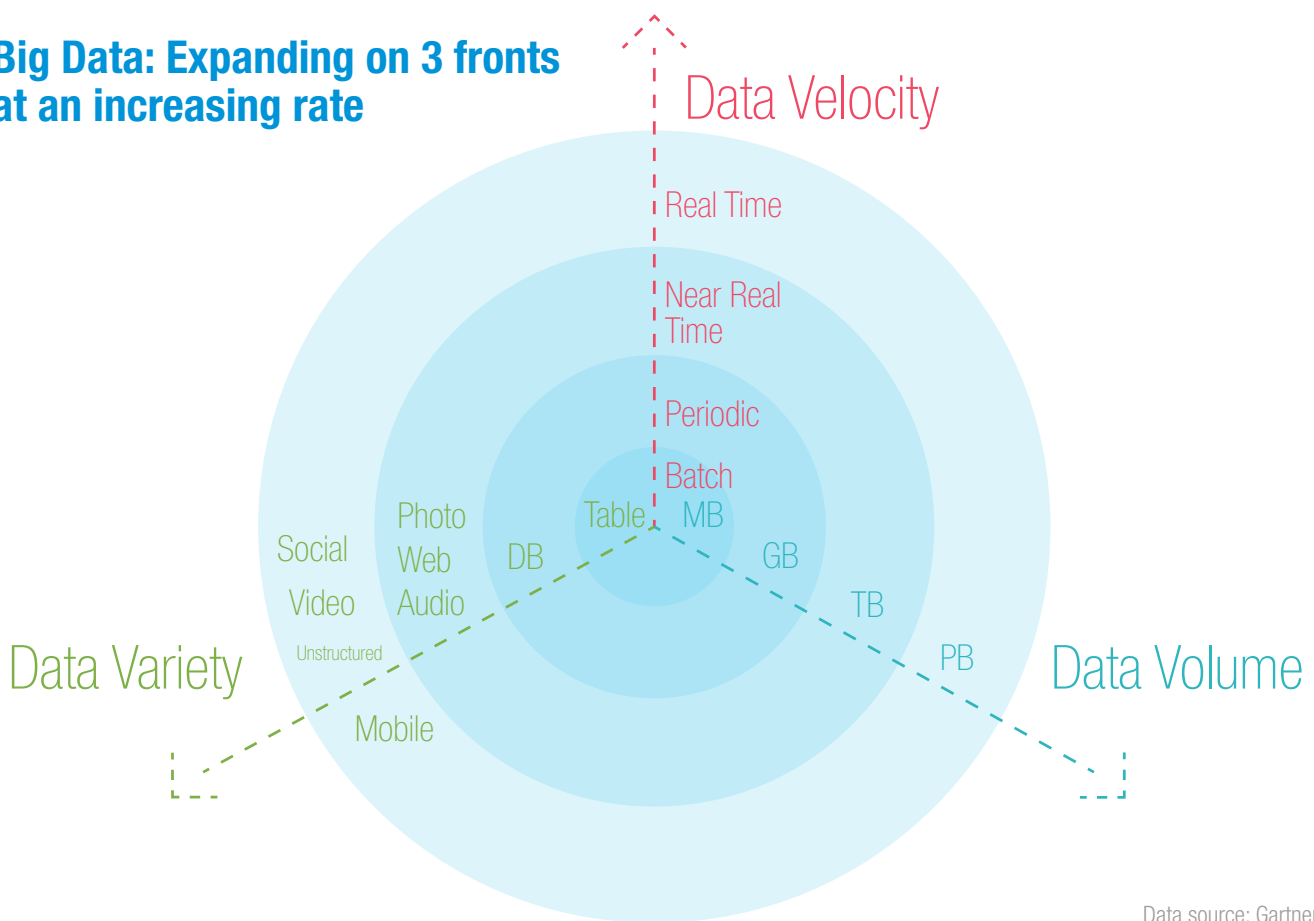
Big Data

Big data is a big topic these days, one that has made its way up to the executive level. Most organizations may not yet fully understand what big data is, exactly, but they know he or she needs a plan for managing it.

The accepted definition for Big Data talks about exploiting “data sets whose size is beyond the ability of commonly used tools to process it within acceptable time”. Big Data is one of the most discussed topics in recent times, the volume of data we’re handling is growing exponentially, with the popularity of Social media and the ever growing internet of things, the mass of data produced by smart electric grids, intelligent traffic systems, etc. According to IBM, “Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone”

The big data construct theory described by Gartner’s has 3Vs – data Volume, Velocity, and Variety), are all growing dramatically and we’re being asked to process data ever more quickly so we can respond to events as they happen, and that data is coming from an ever wider array of channels, sensors and formats.

Big Data: Expanding on 3 fronts at an increasing rate



Data source: Gartner

Data Volume

The new norm is that more sources of data are added on a continuous basis. For companies, in the past, all data was generated internally by employees. Currently, the data is generated by employees, partners and customers. For a specific group of companies, the data is also generated by machines. For example, Hundreds of millions of smart phones send a variety of information to the network infrastructure; Multiple sensor readings from factories, pipelines, etc. This data did not exist five years ago and the result is that more sources of data with a larger size of data combine to increase the volume of data that has to be analyzed and tested. This is a major issue for those looking to put that data to use instead of letting it just disappear.

Peta byte data sets are common these days and Exa byte is not far away.

Data Velocity

Big data can be described by its velocity or speed. Or you may prefer to think of it as the frequency of data generation or frequency of data delivery. For example, think of the stream of data coming off of any kind of sensor, say thermometers sensing temperature, microphones listening for movement in a secure area, or video cameras scanning for a specific face in a crowd. This isn't new; many firms have been collecting click stream data off of Web sites for years, using streaming data to make purchase recommendations to Web visitors. With sensor and Web data flying at you relentlessly in real time, data volumes get big in a hurry. Even more challenging, the analytics that go with streaming data have to make sense of the data and possibly take action—all in real time.

Data Variety

Ranging from excel tables to databases, data structure has evolved to unstructured and to add hundreds of different formats. We now have to deal with pure text, photo, audio, video, web, GPS data, sensor data, relational data bases, documents, SMS, pdf, flash etc. This is compounded by the fact that testers have limited to no control over the input data formats. Structure can no longer be imposed like in the past in order to keep control over the analysis. As new applications are introduced new data formats come to life and have to be dealt with during the software testing process.

The key discussion for this paper are the main challenges of testing big data and what is the threshold for managing large quantities of test data with existing tools and resources such as MS Excel.

The Excel Dilemma

Microsoft Excel is one of the flagship applications of the Microsoft Office suite and by far the most popular tool used in software testing.

According to Naysawn Naderi, a program manager at Microsoft, who played an integral role in the R&D process of defining and developing Microsoft Test Manager 2010, Microsoft Excel, a tool developed in the '90s, continues to be the most popular testing tool in the industry for manual testing. While some manual testers just jot down their test cases on a piece of paper, many prefer Excel, since it can be reused, is digital, and can be shared. However, it provides no customized or optimized testing experience. *"It amazes me that testers live with it, as it is roughly equivalent to having a developer code using Microsoft Word."*

While many still use MS Excel to manage tests, it clearly leaves much to be desired. This includes: taking up a full window while testing, no options for collaborating effectively with the development team, no easy way of sharing what was tested, no ability of sharing steps between test cases, and no linkage to a defect tracking or requirements system.

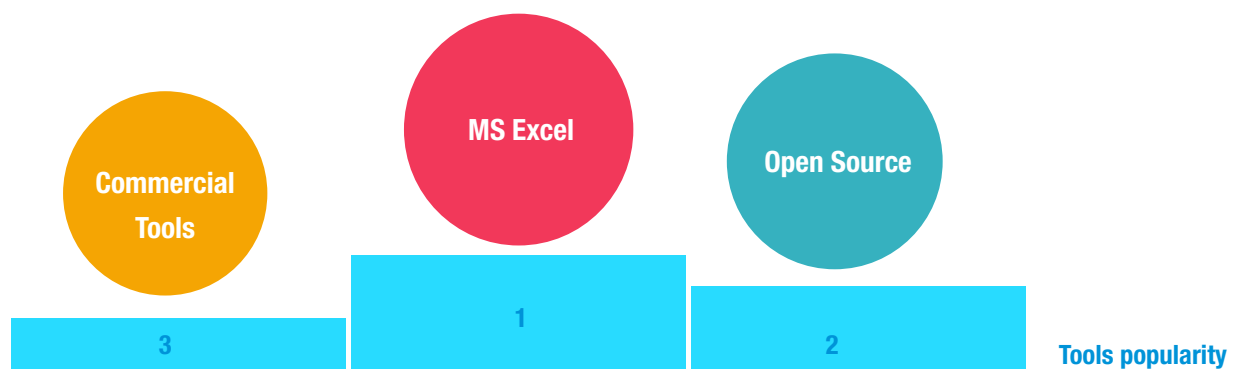
Typically, testers use MS Excel spreadsheets (or, less typically, Word tables) to record testing steps, expected results, and pass/fail state in the required timeframes. They compile the results of testing either manually or with scripted programs that import the Excel spreadsheets or Word tables, process the test results, and then produce somewhat usable reports. However, anyone attempting to use Excel spreadsheets or Word tables to manage manual test scripts will run into these problems:

- The need to constantly scroll horizontally and vertically makes these methods inefficient and difficult to use
- The typically wide formats of Excel spreadsheets or Word tables and the need to reproduce headings on each page make it difficult to print reports
- It's difficult to organize scripts by grouping testing steps.
- There's no standard way to identify expected results within test scripts.
- There's no standard way to report test results.
- It's hard or impossible to reuse test script lines or groups of lines in different test scripts.

- not easily sharable across multiple locations
- not scalable with row and cell limits
- limited data validation
- minimal reporting capability

As you can see there are several shortcomings that make MS Excel a sub-standard test management software tool, however the characteristics of big data make Excel virtually unusable. Excel has been used on big data for years, it's true. But not directly on big data. With a row limit of around 1 million, it certainly can't be used as a direct window into continuous volumes of data. Alternatively, test analysts use it for front end views or other subsets of the original data set, which has been heavily cleansed and normalized.

A tool that solves these problems would enable test departments to take full advantage of their enlightened decision to incorporate manual test scripts into the application development process.



Part 1 – A Solution: What's After Excel? Big Data and the Future of Spreadsheets

First, software engineering teams need to adopt new technologies and to learn — or hire — new skill sets to work with data at this scale. New roles such as data scientists and data technologists are arising to meet these challenges. Training and cross-discipline collaboration can also help to raise everyone's data IQ.

A software solution should solve the needs of manual testers who have to manage large amounts of data, which typically are short of time, work as a part of a team, collaborate extensively with developers, and are passionate about application quality. To begin, testers need a tool that just allows them to test manually. Specifically, they need a tool which allows them to perform exploratory test, author manual test cases, run them while focusing on the application under test, fast forward through the uninteresting parts, file defects, share the results of their testing with their team, and validate that bugs are fixed without having to worry about managing large volumes of test data and the various formats provided.

Second, since manual testers and developers work collaboratively in the development process, each requires an integrated toolset that provides visibility into each other's activities and efficiently allows them to work together.

Third, manual testers need a tool that will provide them with a workflow of going from a manual test to an automated test to a test which is run as a part of a build process with minimal to no interruption.

Part 2: Why use a commercial Test Management solution?

A great test management application will reduce the cost of managing big data by providing structure in the form of test cases. Many organizations have their test scripts scattered across multiple documents and formats. This quickly becomes cumbersome when trying to track test execution, and reporting metrics around test passes turns into a painful exercise of tallying up results from multiple different sources.

Reduce cost of managing test suites by mitigating the pain of managing test cases

A quality test management system should make it easy to import test cases stored in Excel and create structured test cases. Once those cases are imported it should be easy to group those tests into lists, execute and track the results of test list runs thereby eliminating the consolidating stats from multiple Excel sheets!

Reduce time for executing your test

Manual test scripts require a lot of time for setup and configuration before the true testing begins. In many test environments it is common for test scripts to be written, modified and run on an ad hoc basis, where no thought is given to formal approval and release for the test scripts that need to be run. It's not uncommon to spend 10 – 20 minutes doing setup work for a test that takes only a few moments to complete. Advanced testers will often cobble together utilities and small scripts to perform these setup actions. In addition scripting can be more complex and time consuming when dealing with larger and more frequent data structure changes. Ideally, a test management solution should simplify and reduce the time and complexity of creating and managing test scripts.

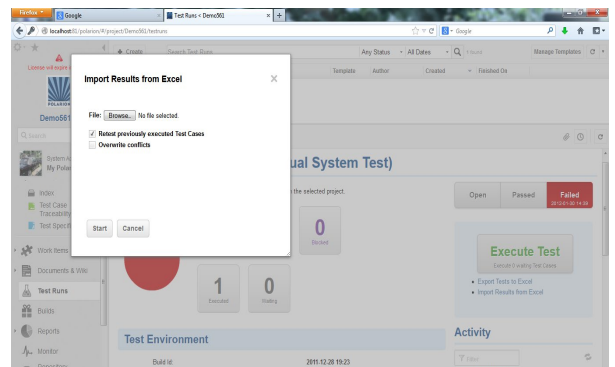
Polarion QA – The Collaborative Test Management Solution for BIG DATA

Manage and aggregate all of your testing activities from one central quality management platform. Most QA departments already use multiple testing tools, have been working with defined processes, and usually do not want to change their processes because of the introduction of a new tool.

Polarion® QA™ provides you with all the benefits of test management software with minimal interruption to your existing environment while reducing cost and improving the quality of your software. We make it easy to create test cases and to link them back to your requirements without disrupting your current workflow .

MS Excel Round-trip

Recognizing the fact that most organizations continue to use MS Excel in their testing environment, Polarion offers the most robust integration that allows testers to easily import existing MS Excel based test data to quickly bring those tests into Polarion QA where you'll be able to edit, track, and execute them with ease with none of the limitations from Excel. Bringing your manual tests in to Polarion also makes it a snap to manage your manual tests in version control.



Continue to use MS Excel when required

As well you can leverage your existing environment using Document “RoundTrip”™ - Changes made outside of Polarion can be imported back seamlessly while keeping your original formatting. All your tests can still be executed via MS Excel while taken advantage of Polarion QA’s advanced features such as detailed traceability from requirements to defects, forensic level auditing, and customizable reporting.

Maximize your true testing time with workflow automation

Polarion QA’s workflow engine and template creator lets you quickly accomplish the same tasks within one consolidated environment. You have an existing test script that creates the setup steps for your testing environment? Bring it in to your manual test straight away! Need to write some code to handle database initialization or system configuration? You can do that too!

Increase Velocity while Reducing Risk

The challenges of big data management result from the expansion of all three properties, rather than just the volume alone – however the sheer amount of data to be managed becomes a major issue. Ironically, as the quantity of test data increases, the expected rate of testing such data increases as well. With increased velocity, Polarion QA can mitigate risk while accelerating the quantity of data consumed during the testing process. For example - when a tester creates a bug, not only are we able to make it extremely rich (complete with video information, a readable log of the actions that took place in the application under test, links to the test case that caused it, and audit logs), but also we are able to present it to developers in an actionable way, where they can immediately jump to the line of code that failed. We feel that if developers are presented with bugs that are irrefutable to reject, they will fix them, and the state of application quality will improve. Further, when the developers fix the bug, they may change some of the lines of code, which will cause some tests that previously passed now to now fail. To check against this, Polarion QA tracks the lines of code your tests cover and recommends that testers rerun the test cases that cover the code that has changed from build to build.

Conclusion

Big testing gets its bigness from three distinct things:

1. A willingness to experiment with big, bold ideas — to try something that may fail. Risk is mitigated by testing such ideas on a small scale, not by avoiding daring new concepts.

2. The pursuit of experimentation is championed by management — it’s a “big deal,” an integral part of the culture. High failure rates of experiments, as long as they’re run well, are not treated as failures of the individuals running them. On the contrary, aggressive testing is recognized and rewarded.

3. A large number of people throughout the organization are empowered to run experiments in their work. This is what gives big testing its scale, allowing a large number of experiments to be run across different facets of the company.

Conventional thinking and software tools clearly do not meet the requirements of being able to properly manage and test big data. As well, for the better part of twenty years, Microsoft Excel has been the defacto test application on the planet. So how do you replace it? The simple answer is that you don’t. Many have tried. All have failed. The best strategy in competing with Office generally, as evidenced by the success of Google Docs, is to reframe the problem and solution. To compete where Office does not. In the case of testing there are several areas that true test management software will exceed the limitations of MS Excel.

Big data is good. It’s an amazing source of new hypotheses for marketing. But to truly unlock the value from big data, companies must embrace Big Data Testing.

About Polarion Software



The creators of the world's fastest enterprise scale web-based ALM solution. The success of Polarion Software is best described by the hundreds of Global 1000 companies and over 1 Million users who rely daily on Polarion's Requirements Management, Quality Assurance and Application Lifecycle Management solutions in their business processes. Polarion is a thriving international company with offices across Europe and North America, and a wide ecosystem of partners world-wide. For more information, visit www.polarion.com.



Handy Links

- [Customer Testimonials](#)
- [Polarion Events & Webinars](#)
- [Contact Polarion Software](#)

 <http://blog.polarion.com/archives/category/testing-test-management-qa>

 <http://www.linkedin.com/company/polarion-software>

 <http://twitter.com/polarionnews>

 <http://www.youtube.com/polarionsoftware>

 <https://www.facebook.com/pages/Polarion-Software/1391000377783078>

 <https://plus.google.com/u/0/110152892433674780669/>

Europe, Middle-East, Africa: Polarion Software GmbH

Kesselstraße 19 — 70327 Stuttgart, GERMANY

Tel +49 711 489 9969 - 0

Fax +49 711 489 9969 - 20

www.polarion.com - info@polarion.com



Americas & Asia-Pacific: Polarion Software, Inc.

1001 Marina Village Parkway, Suite 403, Alameda, CA 94501, USA

Tel +1 877 572 4005

Fax +1 510 814 9983

www.polarion.com - info@polarion.com