A critical customer facing system

[https://www.testingforschools.com] is the flagship digital assessment system operated by the education services provider GL Assessment (GL) [https://www.gl-assessment.co.uk] .

GL launched Testwise as service in 2005 and it remains at the heart of their digital strategy. The service supports many schools in the UK and around the world, enabling them them to assess and determine their pupil's educational needs. In the UK, September is an ultra-busy time for Testwise, when schools assess pupils at the start of the academic year.

The importance of quality of service

It is critical to GL's reputation that Testwise is available and error-free, especially during key periods, like September. Any downtime or failure in these periods is extremely disruptive to the schools using the system. Disruption that leads to dissatisfied customers, bad publicity and reduced uptake of the digital option.

Preparing for growth in 2010

Forecasts for 2010 showed 50-100% growth in the use of Testwise, as schools responded to strategies requiring digitisation. This anticipated growth followed a year when GL had already delivered a record number of digital tests. The GL leadership team recognised that, to meet this demand and to allow for growth in future years, Testwise had to move to a more sophisticated and scalable platform.

After a wobble in 2009

The imperative to make a move was not only driven by these growth forecasts. During September 2009 Testwise had wobbled under load. During the most important and sensitive time of the year the load had proved a bit too much and there had been disruption. A repeat in 2010 had to be avoided come what may.

An updated and re-engineered platform

Whilst the core software technologies of Testwise would remain the same, the changes to be made in time for September 2010 were quite extensive. There were physical network infrastructure and host changes, a major Oracle database version uplift and extensive reconfiguration of the database and JBOSS servers. The GL team realised that this was not a risk-free move and would require rigorous testing, a judgement that proved to be well founded.

Selecting the right test partner

Rigorous load and failure testing was recognised as being fundamental to the success of the exercise. Not a cosmetic, tick in the box, exercise, rather, testing that would push the system and the skills of the delivery team to the limits.

In search of this, GL contacted mainstream tool suppliers and specialist testing consultancies to solicit proposals. At the end of this process, in January 2010, GL's Service Delivery Manager, development team and Operations Director all agreed on a decision to use SQC [https://www.sqc.co.uk] to undertake the work. In their own words…

"Out of all the potential partners they were quite simply the most focused. Their understanding of our needs was far beyond that of any of the other partners and they went into a great deal of detail explaining how they would approach the testing. It was clear from the outset that they would bring a lot of much needed experience to the process"

Paul Webster, GL's Operations Director 2010

"We required more than just a company that could provide a load testing application, one that was able to go that extra mile by providing advice and expertise on how best to proceed through the exercise. SQC were able to work closely with the in-house staff, effectively becoming a highly efficient virtual team, allowing for diagnosis of problems encountered and for them to be solved by drawing on the diverse knowledge pools both teams provided.

The other aspect which made them stand out from the other companies that were tendering for the work was the attention to detail SQC gave to the initial quotation. They spent a lot of time and effort going through the details of the task they would have to undertake and produced a thorough outline of their proposal. This allowed us to have confidence that they would be able to provide everything we required to make the project a success."

Stuart Cairns, Lead Developer Testwise

"I was set the challenge of subjecting our online testing software to Load Testing, to ensure if could withstand forecasted customer usage with additional headroom for the foreseeable future. Given the already aggressive workload on our in-house development and technical teams, I needed to outsource this challenge to a third party. A rough tender was put together and was issued to three different potential suppliers. All three of which responded with their proposals and associated commercials. Out of the three it was immediately obvious SQC were the preferred third party to do the work."

Sean Dotchin, Testwise Service Delivery Manager

GL have identified the deciding factors in this choice as being:

- The quality of both the technical and delivery proposals
- Value for money
- Confidence in the capabilities of the delivery team

A disciplined, engineering approach

SQC knew that a successful outcome for GL was not going to come from simply scripting some user journeys and running basic tests. Success would require a whole range of enabling activities covering knowledge development, work load assessment, non-functional requirements crystallisation, risk analysis, data analysis, load design, monitoring design and technical investigation. This project was a collaborative non-functional engineering engagement.

Knowledge is key

Knowledge and understanding were fundamental, without this everything else would have been ineffective. The SQC delivery team developed a solid understanding of Testwise and of what it would experience. They learnt about the technical environment, the hardware, the network infrastructure and the software architecture. They investigated and understood the behaviour of the user base. Understood how schools used the system and what this would mean in September.

Forming the technical picture was relatively easy, the delivery team were highly collaborative and shared their knowledge and experience. Forming a model of what was to happen in September was more challenging. As with many systems, the collection of data on what users are doing was variable. Piecing together clues from historic user behaviour, from the exploration of the activity records held in the database and log files and from interviews and brainstorming sessions with operational and support staff allowed SQC to form and agree a plausible model of what to expect in September 2010.

GL were also concerned about failure containment in the new architecture. It was designed with redundant network switches, redundant firewalls, redundant proxies, redundant JBOSS services and redundant database servers supported on redundant high availability processing and SAN. The interesting thing about such arrangements is that what happens when something does fail can be a surprise. GL's engineers were aware of this and wanted to test failure scenarios under load. SQC worked with the engineers to explore failure scenarios and identify which ones needed testing.

Together with the data

Load testing eats data and, in this case, vast amounts of it. Repeated high load runs and endurance runs of many hours meant the rate of consumption would be high. To avoid continual database rollbacks, it was necessary to create and load multiple large data sets.

These data sets were complex relating pupils to schools to registers and registers to product subscriptions and the availability of specific test to be taken. Not only were they complex, they need to be representative. There were concerns around performance being sensitive to the characteristics of the data and so the data used had to be representative of real-world data.

Synthesising the data

To meet this need SQC developed a bespoke data synthesiser application. Loaded the profiles of a set of runs the synthesiser incrementally 'grew' the data needed to support those runs. It simulated the runs thousands of times using the data it had already 'grown' when it could, 'growing' an additional data item when it could not.

At each level, pupil, class and school it chose whether to reuse or add a new entity. If it added a pupil then it decided whether to put that pupil in an existing class or add a new one. If it added a class then it made similar decisions around the school. All these decisions were governed by rules that aimed to keep the nature of the generated data aligned with the real-world data. The synthesiser even tracked how long a pupil would be occupied on a visit, allowing their reuse later in the run if they had completed a previous visit.

Getting the most out of the available tools

GL already owned a commercial load testing tool. They had purchased this five years earlier, during a phase of heavy product development, but it was no longer in use. GL had forgotten about it, one of SQC's engineers had not.

The challenge was that GL had 1000 virtual user licenses, on a tool that was no longer supported by the tool supplier, but the load forecast for 2010 created a requirement to support 10,000 simultaneous visitor sessions. The simple, but costly, answer would have been to purchase a new tool with more virtual users. The smart answer was to apply SQC's model of using virtual users as multiplexing workers to simulate multiple, simultaneous, browser sessions on each virtual user. This approach reduced GL's costs by over £50k.

Realistic and repeatable patterns of load

SQC aimed to make the load applied both realistic and repeatable. This ruled out the traditional technique used in load testing, randomised 'think times' during visits. This technique generates a different pattern of load on each run, so it is not repeatable. Furthermore, there are questions about whether it generates 'realistic' load or is just a convenient way of creating the illusion of a realistic load.

The answer was to implement an off-line load synthesiser. A program that generated the timed sequences of actions to occur in advance. The outputs were fed into the data generator to grow the test data they needed. Each load test run executed one of these sequences. Test runs that used the same sequence were, effectively, identical tests, under which Testwise behaved identically, suffering at the same points during the test. Offload load synthesis delivered the two objectives of realism and repeatability.

The Testing Process

Key aspects of the process that was used to load test Testwise are illustrated below.
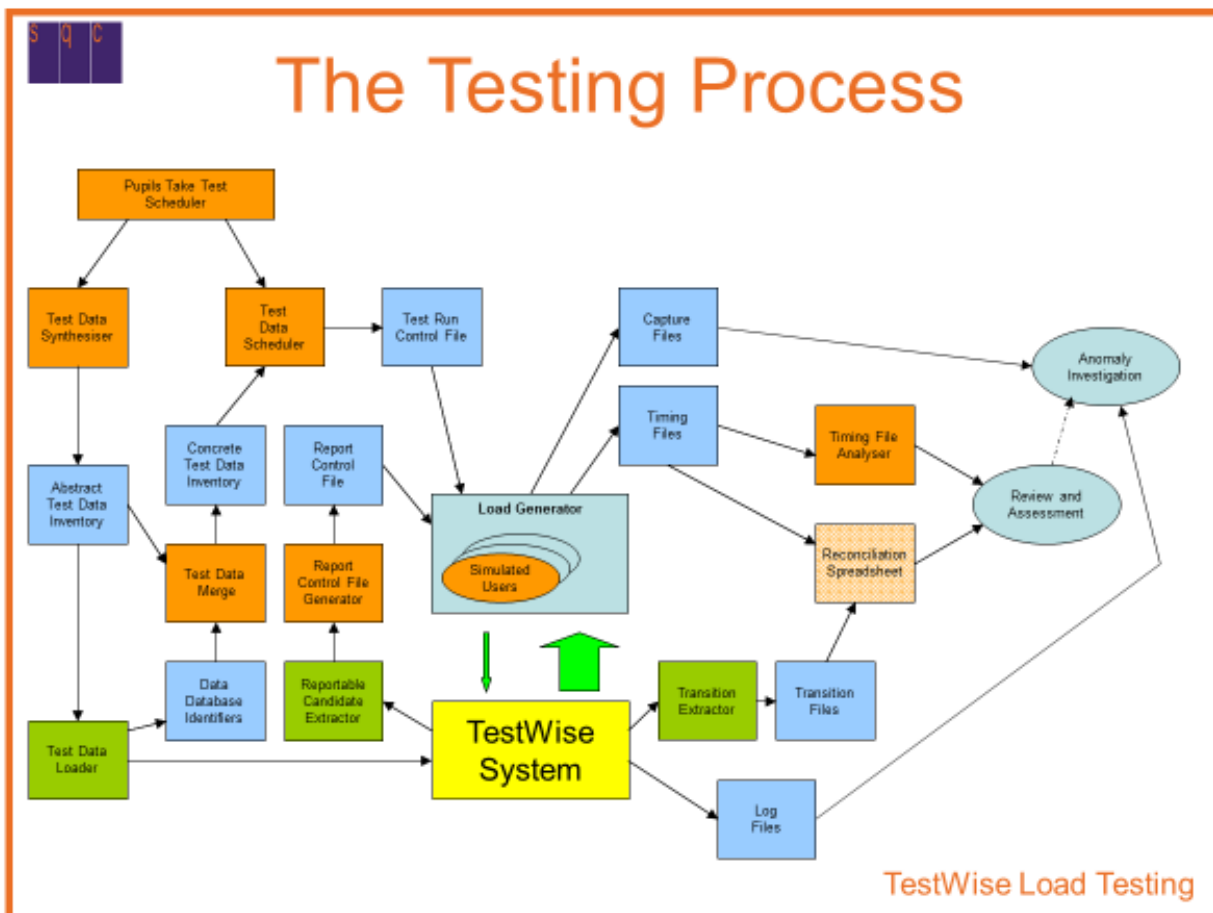
A lengthy load testing campaign

Testing started in April 2010 and sign off was achieved at the end of July 2010, just in time to allow migration of customers ahead of the September busy period. Extensive testing was performed in a full-scale load test environment before final validation was done on the new production platform. It was a long arduous journey delivered with a high degree of collaboration between the two teams.

Round One, a bit of a massacre

Early testing saw issues emerge across a wide range of system components. This cycle saw the demise of GL's initial choice of load balancer appliance. This commercial appliance exhibited unpredictable behaviour under heavy load and so GL switched to a different one.

Six weeks of test and investigation surfaced a variety of issues that the GL engineering team needed to address. They responded with mix of tuning changes, some code changes and the introduction of additional infrastructure. By the end of Round One, Testwise had changed dramatically and was meeting the target response times under load. The implementation had become more efficient, using less resources, and more scalable.

Round Two, building up stamina

Being able to run, without incident, at target load for one or two hours cannot be used as a basis for trusting a system to provide a critical service under the same load. Trust must be based on seeing the system survive and remain consistent under such loads for many times the duration that it could experience such loads in production.

Confidence comes in two ways. Firstly, repeated tests, applying extended periods of continuous load, establish confidence that the system is not near the edge and surviving by chance. Secondly, the cumulative runtime and number of transactions processed across multiple runs provide detection of random failure events and of failures caused by the accumulation of processing records, outputs and processing by-product's.

Testwise was subjected to hundreds of hours of testing within which it performed millions of transactions. The engineering team tackled process stalls and go-slows, application server crashes, unstable load balancer responses and database outages. At the start, runs of three of four hours nearly always resulted in failures. By the end, Testwise was experiencing peak load for thirty-six hours, close to twenty times the expected duration of busy periods, without blip.

Round Three, taking the blows

Attention then turned to the resilience of the system. GL prides itself on not causing disruption to schools and their pupils. Testwise is designed for high availability and to guarantee completion of all 'inflight' visits, so that no one loses their work. Once a pupil has commenced a test they should be able to complete it and their answers will not get lost, not matter what happens to the Testwise system.

However, things will break at inconvenient times with unforeseen consequences. This occurs within the infrastructure and off the shelf components that support an application and within the software of the application itself. It is unavoidable, particularly under heavy load. Therefore, GL's answer was not to pretend this would not happen, rather it was to act to ensure that, when it did happen, the end users would not know it had happened.

GL and SQC worked together to identify a catalogue of 'interesting' potential failures. These covered a wide range of domains including network anomalies, server outages, process crashes and database node failures. The teams then embarked on a campaign, testing how the rest of the system responded when failures were injected whilst the system was under load. Tests checked the impact on active users and assessed the way the overall system recovered its 'poise' immediately after the event and on restoration of a failed element.

When failures were injected and the outcomes fell short of the high standards expected of Testwise the engineering team acted to refine or replace aspects of how the system worked. The result of this? Almost complete isolation of inflight end users from any anomalies in the operation of Testwise. From individual worker process failures to wholesale failure of the active database or SAN, nothing disrupted the experience of the pupils taking their, to them, very important and stressful assessments.

The final step, making sure production worked

The final hurdle GL faced was ensuring the production system was operationally identical to the load test system. With many new components and a need to deploy and configure them in exactly the same way as they were setup in the test environment there was an ample opportunity for small mistakes that could undermine all the good work done by the teams.

To counter this threat, it was decided to test the production platform, moving the load test from the lab into the outside world. SQC established a second load injection capability using virtual infrastructure hosted on the Amazon EC2 platform. This solution provided load generators that could be activated and deactivated on demand and high bandwidth connectivity out into the internet. Though a common solution now this was not the norm at the time.

The first thing the production testing broke was nothing to do with Testwise, rather it was a backbone internet router, located in London, on the path between the Amazon data centre in Ireland and the Testwise data centre in the North of England. Latent problems within the router's configuration were revealed by the intensity of the traffic created by the load test. Once this was diagnosed the way was cleared for final corrections to Testwise and the system was declared ready the weekend before the, September, start of the new school year in the UK.

The overall picture

Around two hundred load tests were executed in the load test environment, with active load durations ranging from two to six hours. Endurance campaigns applied six, six-hour, tests back to back, keeping the system busy for thirty-six hours. Overall the system was exposed to more than six hundred hours of operation under full load. This involved over 2,500,000 simulations of a pupil downloading and completing a test.

The journey was a tough one, there were challenges for both the SQC and the GL teams. Long hours and seven-day weeks. Investigations of strange phenomena and creative problem solving. Testing started in March and it took until July before the end of the road came into sight. The full extent of the work undertaken is indicated, visually, below.
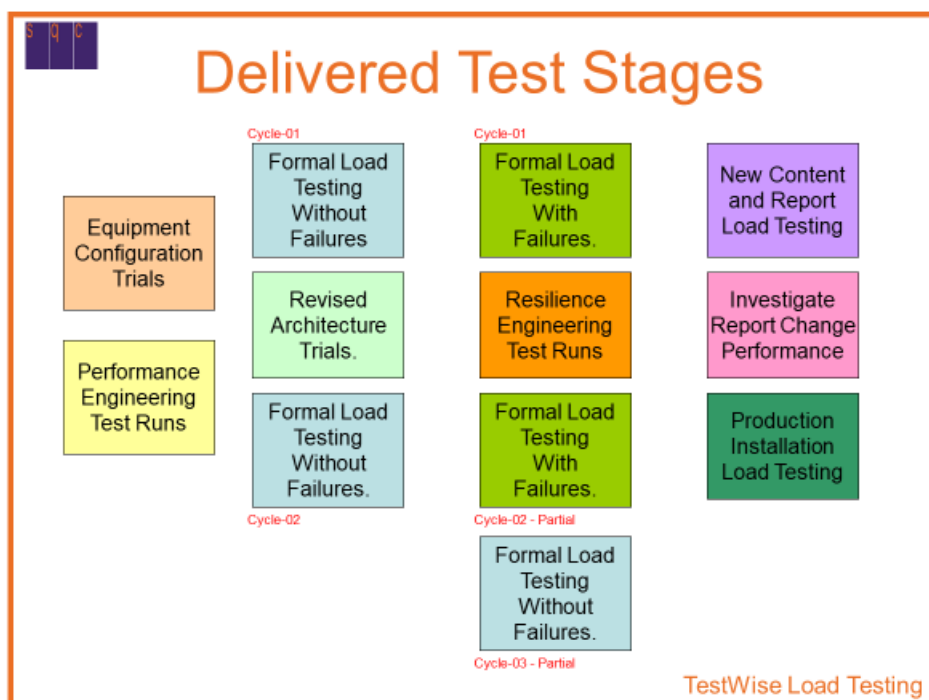
Successful in many respects

September 2010 and subsequent years

In 2010, Testwise experience its highest ever demand, the busiest week in September saw 80,000 tests taken compared with the previous year's record of 62,000. However, unlike 2009, the quality of service delivered to customers was exceptional. Perhaps the most notable outcome was the lack of stress experienced by the operations and support people during September, reported as...

"the situation was relaxed enough for time to be spent on other matters not related to keeping the system running, whereas the previous year the technical operations team had little time to spare".

Paul Webster, GL's Operations Director 2010

Did SQC's testing make a difference?

How much did load testing contribute to the success of the Testwise technology uplift? The opinion of GL's team is that it was fundamental to the success they achieved in September 2010...

"during the official load testing, our software and infrastructure needed a reasonable amount of configuration in order to meet the demands of the load testing exercise. Without doing these configuration changes, and without SQC helping us identify them, our software would not have been able to meet the demands of forecasted customer usage we originally set out to achieve."

Sean Dotchin, Testwise Service Delivery Manager

"I don't even want to think about what would have happened if we hadn't carried out such thorough testing."

Paul Webster, GL's Operations Director 2010

The work done on the Testwise system has given GL room for growth; the configuration was tested to well beyond the demand forecasted and the architecture ensures headroom can be maintained by scaling the system as demand grows. The exercise transformed Testwise into an industrialised, dependable, highly scalable solution. The platform that was tested in 2010 supported GL's digital ambitions and growth until 2017 when it was replaced by a cloud solution. Over that time the platform saw a fourfold increase in peak load. Each year it was load tested and each year, following appropriate horizontal scaling, it was able to sustain its quality of service. The platform never failed, nor displayed performance issues, under load.

GL's assessment of the service provided by SQC?

The story of the Testwise refresh, from the start of the testing in April 2010 to its first operational exposure in September 2010 was not one GL foresaw. The variety and depth of issues encountered was a surprise. Overcoming this required great flexibility, from both the GL team and from SQC.

SQC's collaborative approach to the work, doing what ever was necessary, and treating commercial concerns as something to be dealt with once the job was done, provided the flexibility and responsiveness needed to sustain a high velocity and, ultimately, complete just in time. GL recognised the partnership approach that lies at the heart of the way SQC approaches work...

"Throughout the load testing, SQC provided a service which exceeded any expectation we had when we initially started the exercise. Everything from the SQC Team individuals, the documentation, the support, the advice, the approach, far exceeded what any other potential load testing partner could have achieved, and we would ultimately have proceeded to have serious technical issues with our software had we not undertaken the exercise."	**Sean Dotchin, Testwise Service Delivery Manager**
"Would I recommend SQC to other customers? Without a shadow of a doubt, probably the most recommended and trusted partner I have the pleasure to work with in over twenty years in the industry."	**Paul Webster, GL's Operations Director 2010**
"The level of service was second to none. They provided a very professional service throughout, working very closely with the technical team. I would recommend SQC to anyone."	**Paul Webster, GL's Operations Director 2010**

An ongoing partnership

Eight years on, in 2018, SQC are still supporting GL, providing technical testing and load testing capabilities to support the ambitious Testwise road-map. Services provided have included:

- Nine years of a continuous, business as usual, load testing services.
- An annual pre-September validation of the Testwise service that now tests at six times the load testing in September 2010.
- Replacement of the original tool with a custom load testing solution able to support the higher loads now being tested.
- Full non-functional testing of the replacement of the JBOSS based Testwise system with an Azure hosted cloud solution that went live in 2017.

Need to know more?

To find out more about technical testing services, including load and performance testing, provided by SQC please visit the technical assurance section of our website [https://www.sqc.co.uk/perform/assurance/technical-assurance] or email or enquiry mailbox [enquiry@sqc.co.uk] and we will get back to you.