



## Three Reasons to Load Test Your Production Web Site

Most Web sites are mission critical and, in many cases, are a major source of revenue for their company. Although many companies fully realize the value of load testing, often they are concerned about applying load tests to the production Web site. At Keynote we firmly believe that the benefits of load testing a production Web site greatly outweigh the disadvantages. This brief white paper describes how to plan and execute such a test in a way that minimizes impact on the site and its users.

## Why Load Tests Should Be Run on a Production Web Site

There are three, overwhelming, reasons why load tests should be run against a production site, rather than a *clone* of the site:

- Validity of the results
- Cost in terms of time and money
- The *fire-drill* value

### 1. Validity of the Results

Most Web sites are implemented using a wide variety of software, hardware, and services. The software typically includes operating systems, application servers, databases, etc., the hardware includes firewalls, routers, load balancers, servers, etc., and the services may include content distribution networks, ad servers, credit card verification systems, etc. These components come from different vendors, and each component has a unique performance and scalability profile. To make matters even more complicated, Rich Internet Applications using Web 2.0 technologies such as AJAX and Silverlight introduce the client-side processes, which may also be connected to partner applications. So, unless the system under test is *extremely* simple, creating a clone of the production system with the same performance and scalability characteristics as the original is very difficult. Since a difference in any of these components can dramatically change the scalability of the entire system, the results of any load test applied to the clone system cannot be applied to the production system with any degree of confidence.

### 2. Cost in Terms of Time and Money

Even if the production system can be cloned, the costs associated with recreating such a system with the accuracy required to make the load test results on the clone be applicable to the production system will be significant. Since in order to be valid, a complete Web site load test must include all components (hosting company, firewall, router, load balancer, etc.) you would have to purchase, install, and configure each and every component. Some people may be tempted to clone a subset of their system; instead of cloning the load balancer and four servers, for example, they decide to load test just one server and then multiply the scalability result by four to arrive at their scalability of their entire system. This would be a valid test for the scalability of one server, but not of the Web site as a whole. This load test would not identify any problems with the load balancer, for example, and would not confirm that the hosting company network could handle four times the load used in this smaller scale test. Over years of commercial testing we have seen numerous load balancer problems (mostly due to mis-configuration) and numerous surprises when the bandwidth requirements at a hosting company were stressed. The additional complexity of Web 2.0's asynchronous applications can create zero, one or many server requests. The addition of client side testing components to emulate user behavior is prohibitively expensive in a laboratory environment.

### 3. The Fire-Drill Value of a Production Site Load Test

One of the major reasons why some companies don't want to run a load test on their production site is that they are not looking forward to performing activities such as the database back-ups and restores that a production site load test requires. We believe that such *drills* are a positive side-benefit of load testing. We have all heard stories of Web sites crashing under heavy load and taking many hours, or even days, to bring their system back. Practicing system back-ups and recovery following a major load is an extremely important component of Web site preparedness. A load test on your production site will not only show you how well your system can handle a large load, but how well and how quickly your system and crew can recover from a site crash due to overload. If there are problems with post-crash recovery, the right time to discover and fix them is during a test, not while your site is experiencing real traffic peaks.

## Minimizing the Impact of Production Site Load Tests

Even though there are clear benefits to performing a load test on a production site, it's clear that such a test will have an impact on the site, its managers, and its users. To minimize this impact we recommend the following strategy:

- Conduct the load tests during the slowest hours of the slowest day of the week.
- Redirect users to a "Temporarily Out of Service" page.
- Increase load volumes gradually to avoid complete system crashes.

### 1. Test During Off-Hours of Off-Days

The simplest and most obvious way to minimize the impact of a load test on a production site is to conduct the test during the days and hours of the week when the traffic is lightest. Almost all Web sites have very clear volume patterns with peak and lows that can be easily identified by analyzing server log files. Such an analysis might reveal, for example, that the hours between 2AM and 5AM on Saturday nights represent less than 0.2% of your total weekly volume and a load test conducted at that time will not impact 99.8% of your weekly traffic. The 0.2% of sessions that might be affected is a small price to pay for the performance and scalability information and improvements that a good load test can give you.

### 2. Redirect Users to a "Temporarily Out Of Service" Page

Even if you schedule the load test during off hours, you might not want to any users to experience your Web site during a load test. One of the things we recommend is to redirect users to a Web page which explains that, in order to serve them better in the future, the Web site might be temporarily slow or unavailable because it's undergoing maintenance and testing. Then thank them for their interest and understanding and invite them to come back in a few hours.

### **3. Increase Load Volumes Gradually to Avoid Complete System Crashes.**

Another technique to minimize the impact of load testing on your regular users is to increase load volumes gradually and stop the test when the Web site's response time exceeds a threshold that you consider unacceptable. Then look at the load test results, identify the bottlenecks, fix them, and test again. In other words, instead of pushing the Web site to the breaking point in one single load test, gently push it to a load level when it begins to show performance deterioration. With several load test and performance improvements iterations, you will be able to gradually increase the scalability of your Web site while maintaining the Web site's availability and keeping the response time within acceptable parameters. This way the few users you will be affecting may have a slightly slower than usual experience, but will be able to conduct their business.

#### **The Bottom Line**

Even though conducting a load test on a production site might cause some disruption, the benefits of such a test greatly outweigh its disadvantages. Furthermore, if the test is planned and timed properly, the disruption to your team and your Web site visitors can be reduced to an almost negligible level.

On the other hand, an untested production site that crashes when subjected to real load by actual user will fail, by definition, during a period of peak traffic when its failure will cause significant disruption to the greatest number of users.

In our opinion, the choice is clear. Unless you can create a perfect clone of your production environment, you'll sleep much better at night if you do a thorough load test on the production environment itself.