Ptak / Noel

Are Your Cloud Applications Performance-Ready?

Summary:

Cloud infrastructure scalability does not guarantee that applications will perform optimally in the cloud. Cloud applications still require performance testing and tuning to ensure that they perform optimally. This paper provides a high level overview of application performance testing considerations when moving applications to the cloud.

This paper is sponsored by Apica Systems

Application Performance and Moving to the Cloud

Cloud computing is on the radar of many companies – many are in the planning stages, some are in the process of deploying private clouds, and still others are already using clouds. The promise of cloud scalability, flexibility and agility are driving cloud adoption. However, as IT staff move their applications to the cloud, many of them are finding that cloud infrastructure scalability does not automatically eliminate application performance issues or the need to do performance testing of new cloud applications.

Early cloud adopters are also discovering that cloud application load and stress testing is not the same as testing applications in traditional environments. They are finding that applications cannot always scale up, even when running on scalable cloud infrastructures. This becomes clear when applications that performed adequately in traditional environments "break" when subjected to higher loads and dynamic scalability in the cloud.

This highlights an important point for IT staff that are moving to the cloud. Typically, cloud initiatives focus on building the cloud infrastructure, but IT staff must also focus on the readiness of applications that will run in the cloud. Cloud application performance depends on the scalability of the infrastructure, but it also depends on building scalability and performance into the design and architecture of the application itself.

Bridge engineers painstakingly build in and test strength and resiliency into their bridge design, to ensure that the bridge stands up to the traffic and weight loads in actual use. Similarly, application developers must build in and test their applications to ensure they are cloud-ready for performance.

> 'Apica provided us with automated testing that, when accompanied by their human touch and personal analysis of the data, allowed us to troubleshoot all of our bottlenecks. It was literally like being able to attach a hundred plus users to a debugger.'
>
> *Richard Bethell, Software Development Director*

This paper provides a high-level discussion on the approach for managers and application stakeholders to test cloud application performance. If you are new to cloud or not sure if you are approaching performance testing for your cloud applications correctly, this paper provides a starting point for evaluating your processes.

Why Load Test the Cloud?

Cloud computing's advantages of scalability and flexibility might lead one to ask, "Why load test in the cloud, when I can just scale up resources?" Although the question seems logical at first, the need for cloud managers to wisely manage the scaling up of resources for performance while balancing the costs for those additional resources is even more important in cloud environments. Load testing can help cloud managers plan resource consumption while optimizing the balance between cost and performance.

Cloud performance testing must "test the limits" by ensuring that cloud infrastructure elasticity is matched with applications that scale well. Figure 1 summarizes several reasons why load testing is important for cloud applications.

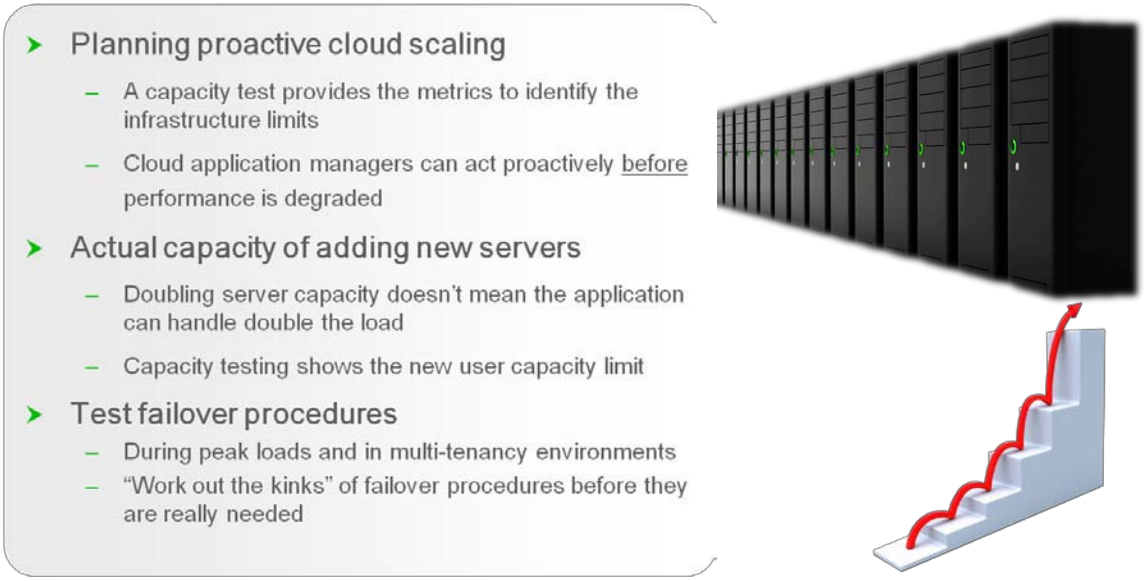

Figure 1: Why Load Test Cloud?

In addition, consideration for performance issues related to third party content and trends can be accommodated through cloud performance testing. For example:

- Optimizing performance despite the increasing complexity of web-based applications and inclusion of third party content
- The impact of social media
- Search engines using URL response time to assemble the first page of search results. This is especially important as 62% of searchers don't look beyond the first page.

A Multi-Faceted Cloud Performance Testing Approach

The unique nature of cloud environments and applications running in the cloud compel a different approach to testing cloud application performance. Cloud application performance testing must use a multi-faceted approach that goes beyond just measuring response time. This multi-faceted approach is described as a series of steps, outlined in the sections below.

However, keep in mind that performance testing should be a regular iterative process that is integrated into the application development lifecycle – and built into your quality processes. This is particularly important because of today's accelerated release schedules, which frequently result in inadequate performance testing as developers rush to meet aggressive business deadlines. This "corner cutting" is alarming when considering that inadequate performance directly impacts revenue and the business.

Understanding Cloud Application Characteristics

The first step is to understand the performance characteristics of your cloud application. Understanding how your application will perform under various load levels provides the ability to plan effectively and proactively manage your cloud application's performance. ***Load Profiles*** show how your application behaves at different load levels for selected use cases, with a variety of metrics revealing different aspects of performance. For example:

> Average response time per web page – shows how response times increase with increasing loads. This enables IT staff to determine when action (tuning or adding more resources) is required to prevent response time from reaching unacceptable levels.
>
> Transaction rate – shows the effect of increasing loads on transaction throughput. Useful in identifying when the maximum server/resource capacity is reached.

The benefit of understanding how your application performs along the load curve is that it enables IT staff to know when increased resources are required to prevent unacceptable performance degradation and poor end-user experiences.

The second step is to understand the ***transaction characteristics.*** This involves learning the behavior of the transaction and gives insight into tuning, resource planning and diagnosing performance problems. For example, understanding what resources are most sensitive to your transaction's performance (server CPU, database, etc.), knowing the composition of the transaction response time and where time is spent.

The third step is to understand the ***geographic profiles of load sources***, which means understanding geographic-related issues that affect the application's end-user experiences in dispersed global locations. For example, local network connectivity issues, local ISP response times, etc.

Performance goals should be defined simultaneously with the steps above and provide guidance for the performance testing and resulting actions. This includes performance goals that define performance (response time, throughput, number of concurrent users, etc.) measurements for desired performance, acceptable performance (expressed as a range), and unacceptable performance. Performance goal definitions could be more specific (by geography, transaction, etc.) or general, depending on what makes most sense for the application.

Stress Testing Cloud Applications

One of the ways to determine your cloud service's performance-readiness is to stress test it. Cloud services can involve complex interactions of various components, such as load balancers, databases, web front-end servers, application servers, and cache servers. Stress testing a cloud service can reveal a lot by watching the performance indicators as the load levels increase. When heavier transaction loads (stress) are placed on the cloud service, "weak links and components" in the cloud service either break or begin showing slowdowns. Identifying these performance vulnerabilities enables IT staff to remediate these weak links before they affect real customer transactions.

Configuration errors and inefficient programming techniques are typical issues that affect cloud application performance, and stress testing exposes them.

In addition, Apica performance experts report that customers' cloud application performance issues are identified repeatedly in these areas:

> Database backend - problems can be caused by inefficient database access (programming) or database configuration issues.
>
> Load balancing – problems can be caused by configuration issues.
>
> Web front-end programs that sit behind load balancers – problems can be caused by inefficient programming techniques, or configuration issues.

Stress testing is also useful to see what happens when the application's capacity is exceeded. This enables IT staffs to put contingency plans in place, so their plans can be executed if the application ever approaches its capacity.

Capacity Testing for Cloud Applications

At first glance, "capacity testing for cloud applications" may sound like an oxymoron, given cloud's scalability. But there are several reasons why capacity testing is a sound practice for cloud applications. The first is for ***planning proactive cloud scaling***. A capacity test provides the metrics to identify when the existing cloud infrastructure will reach its limits, enabling cloud application managers to proactively spin up additional servers before performance is degraded. They can easily compensate for the lag time to spin up servers by putting in their scale-up order just before it is needed. The result is that scale-up actions are based on more than just CPU threshold metrics (which many cloud providers use as triggers for adding servers.)

Cloud managers must also understand the impact of ***adding new servers***. Doubling server capacity does not necessarily mean that the cloud application can handle double the number of users. Capacity testing gives analytical data to predict the increased number of concurrent users that can be serviced when an additional server is deployed.

In addition, capacity testing can be used to ***test failover procedures*** during peak loads and in multi-tenancy environments, enabling staff to "work out the kinks" of their failover procedures before they really need it.

Performance Testing to Select the Best Cloud Infrastructure Option

Cloud performance testing offers managers the opportunity to test various cloud options to determine which infrastructure or cloud provider option delivers the best performance for the price. For example, performance testing helps with comparing the alternatives, such as, whether it is better to have a few big servers or a larger number of medium-sized servers. In addition, performance testing can also provide performance comparisons between cloud providers, exposing specific provider nuances that affect performance. For example, with Amazon's cloud service, medium-sized servers are set up as multi-tenant servers but its large servers are dedicated servers, which will likely

affect performance. Performance testing both infrastructure options provides the necessary information to make an informed decision to pick the right alternative.

Apica's Application Performance Approach

Apica's approach to cloud application performance focuses on solving customer performance problems by delivering a combination of performance testing data and "application performance and optimization strike team expertise." Apica's strike teams use their best practices, methodologies and know-how to help their customers interpret and use performance information to find performance problems, tune and optimize cloud applications.

Although customers have the option to "do-it-themselves", one of Apica's differentiators is the expertise they provide as they work alongside their customers to diagnose and solve performance issues.

> "The Apica team worked closely with our engineering team…They not only provided incredible loads and quick results but they also provided great insight on various configuration settings. During the test cycles they became an embedded part of our team because of their commitment to seeing our results improve. If you want someone who is committed to your success, then I highly recommend you go with Apica."
>
> ***Roberto Monge***
> ***Lead Architect, Transgaming***

The Final Word

Cloud application performance testing is an essential part of the cloud application lifecycle. As teams rush new applications to the cloud, performance testing is not the place to cut corners because there is so much at stake. Important decisions must be made when using public clouds – managers must choose between infrastructure choices that balance price and performance, as well as selecting the right cloud provider partner.

However, it can be confusing because there are many vendors offering cloud application performance testing solutions, touting the amount of performance data that they provide. Although having the right performance data is necessary, what is most important with application performance management is not the amount of data that you produce, but understanding how to best use that information to find and fix performance problems, or fine-tune application performance for new applications before they are placed into production.

A cloud manager's perspective is best expressed as, "I don't just want to know that I have a problem and know where to look, more importantly I want to fix the problem quickly." Cloud performance testing is a fundamental way to minimize the risk of unpredictable and costly behavior in the live environment.

Publication Date: May 10, 2012

About Ptak, Noel & Associates LLC

We help IT organizations become "solution initiators" in using IT management technology to business problems. We do that by translating vendor strategy & deliverables into a business context that is communicable and actionable by the IT manager, and by helping our clients understand how other IT organizations are effectively implementing solutions with their business counterparts. Our customers recognize the meaningful breadth and objectively of our research in IT management technology and process.
www.ptaknoel.com