

Devising a **Cloud Strategy** that Doesn't Cast a Shadow on App Performance



By Nick Cavallancia

Performance is the penalty that you pay in exchange for the agility, flexibility and cost savings of the cloud.

The move to the cloud comes with visions of always on applications, happy, productive users, and anytime, anywhere access. It touts itself as the answer to all of your on-premises application problems, providing a highly scalable and secure infrastructure on which to host your most critical applications. But that same move is actually burdened by performance issues like low response times or high latency, at a time in our industry when users have the highest level of expectation around application speed and availability. Performance is the penalty that you pay in exchange for the agility, flexibility and cost savings of the cloud. All this leads to poor end-user experience and can cause a resulting negative impact on the organization's ability to operate efficiently and effectively.

Whether you're moving to the cloud, or are already there and are trying to figure out a better way to do the cloud, the question becomes is there a way to utilize the cloud that ensures the highest level of performance possible?

You fall neatly into one of two camps when reading this paper – either you have an application on-premises and are looking for the best way to move to the cloud, or you are already in the cloud and are experiencing performance issues, causing you to look for a better way. Regardless of which camp you belong to, you're still in the same position of looking for the right way to have an application hosted in the cloud with the highest level of achievable performance possible.

In the quest to find out, we first need to look at what's casting a shadow on performance in the cloud today, and then cover a few possible strategies that give you the best chance for performance.

Cloud + Apps = Performance?

We've all heard the "promise of the cloud" – that mantra of everything's better/faster/cheaper/safer/etc. when you put an application in the cloud. Some of that mantra is well-founded, as there are definite benefits around cost savings, scalability, and accessibility when utilizing the cloud over hosting an application on-premises.

But should you simply assume that when you throw an application up into the cloud, it's going to inherently make good on all those promises?

App performance in the cloud is a bit more complex than that because you're making the switch from a network that you have (or can) optimize to one you have little or no control over at all. With cloud applications, bandwidth needs, traffic utilization, and impact on other applications all coming into play and being far more relevant than when on-premises, you shouldn't just take app performance for granted in the cloud – even though you're relying on the cloud to deliver the same level of performance you've enjoyed on premises.

Why the Cloud Must Perform

Part of the decision to move to the cloud involves meeting ever decreasing response time expectations of your applications by its users. Lack of app performance has a lot to do with the user's experience with the application, which, in turn, has an affect on the user's performance within the application. These two together somewhat equate to the impact on the user's productivity. Take a simple slowdown within a given application of only a second. As it continues to happen frequently, the user begins to think their machine is slow, causing a ripple effect. They eventually call the helpdesk and someone from IT has to look at their machine to see if anything's amiss. What started as a simple performance lag of a second has killed the productivity of two employees.

It seems that the simplest litmus test for application performance is the user experience.

This kind of potential impact on your organization means as you move applications to the cloud, IT needs to be laser focused on the performance optimization of those applications – whether SaaS-based, in a private cloud, or even apps that take a hybrid approach. The more comprehensive a cloud solution, the more control over performance is relinquished. Take SaaS-based applications that deliver turn-key software in the cloud. That's great, except now IT owns nothing! Not the application itself, nor the middleware, nor the underlying infrastructure. The more turn-key, the more acute the problem. This is where a loss of control is at it's absolute worst and

requires creative solutions to address performance issues. Without taking the optimization of the connection between user and app into consideration, organizations risk more costly migrations, longer time-frames to move to the cloud, and lowered user productivity.

So, as you come to understand that performance may not be all it's cracked up to be and yet is vitally important, you quickly realize it's time to get serious and figure out what's keeping your applications from reaching their performance potential. But before you can actually do that, you need to first understand how IT's grip on application performance has changed once you move to the cloud.

Despite how awesome the cloud can actually be, there are aspects of the cloud that will affect IT's ability to ensure a secure and improved user experience.

The Optimization Shadow Cast on IT

As you move applications to the cloud – even before you experience any kind of loss in performance, the “shadow” of the cloud's lack of optimization begins to loom over IT. In this case, the shadow is those parts of the cloud no marketing department ever wants to tell you about. Despite how awesome the cloud can actually be, there are aspects of the cloud that will affect IT's ability to ensure a secure and improved user experience.

- **Loss of Visibility** – On-premises affords you having applications running on a server down the hall, and users just around the corner. You know if an application is up or down almost immediately, whether there are power issues in the building, etc. The network path from user to application is simple, with only a limited number of pieces of network hardware involved. But with the cloud, you no longer know where the application resides (and what, potentially may be impacting it), where the user resides, and in some cases even what's in between them! And when problems arise, you have little ability to understand where the problem resides within the underlying infrastructure supporting the app delivery (from user to application), making it difficult to even participate in the troubleshooting or optimization process.
- **Loss of Control** – You knew you'd give up some control the moment you thought of putting an application into the cloud, but it was always a question of how much. When an application isn't performing adequately, you lack the ability to change the path applications take over WAN and Internet links, make QoS policy

changes, or address security concerns. In fact, because you lack visibility, it's likely you won't even know what the proper action should be.

- **Lack of Maintaining an SLA** – Of course, every cloud provider will have a service level agreement, but should they fail to meet it, the repercussions to the provider are limited and rarely compensate for the loss of business. Ideally, just like your on-premises applications, your organization's management team is looking for you to provide and be able to uphold some kind of SLA.

It's having that visibility and control into network and server performance that translates into an ability to begin to optimize application performance. And, in general, the cloud natively isn't always helpful to provide the levels necessary. It may be that due to the significant cost savings, you're aware of and are willing to make some concessions around who owns and controls app performance. But even so, it's important to understand what you should be optimizing so you can begin to determine a strategy to minimize performance hits and maximize your investment in the cloud.

Eliminating the App Performance Shadow

Sticking with the analogy of the shadow, to improve application performance, you simply need to shed light on as many parts of the application experience as possible, and have an ability to optimize them.

So, what application performance factors in the cloud should be optimized?

There are a number of variables that come into play when a user is connecting to a cloud-based application and consuming it:

- **The Path** – Think of the app in the cloud as one end of the path, and the user as the other end. If your app is on a public cloud and accessed from a user remotely, the path is a complete unknown to you. Even if users are located within your office, you only are aware of the path up until you hit your Internet connection. You need an ability to dictate what path is taken, as well as an ability to identify where in the path a problem exists, should one arise.

- **Bandwidth Reservation** – Setting up a reservation is easy when you own the networking hardware. But since you no longer have control over the path (nor the hardware involved in that path) to the cloud, the ability to measure and control the bandwidth consumed (and do something about it to avoid congestion) will be critical to ensure performance.
- **Quality of Service** – Tied in with bandwidth reservations, similarly you should have an ability to place a network priority to a given cloud application, other than within your own organization's network.
- **Latency** – In some ways, latency becomes the sum total of the previous three variables. With nothing more than the application traffic traveling in an unknown path with no QoS, your users can experience latency. Now add on top of that potential problems on the cloud infrastructure hosting the application, or an issue with the application itself. Latency only multiplies the impact of those issues, making the user feel the application is less and less responsive. So by gaining control over optimizing the path, the bandwidth and the QoS, you will logically have better control over any latency experienced.

Admittedly, without an ability to optimize an application once in the cloud, it does sound all gloom and doom because you no longer are truly in control. But the reality is if you simply put an application in the cloud, with no consideration of how to maximize application performance, you will be stuck with no visibility, no control, and no ability to respond to issues.

So, what's the right cloud strategy to ensure application performance?

Crafting a Cloud Strategy w/out the Shadow

It was once said the best cloud approach is the one the users know nothing about. Makes sense – if you can put an application in the cloud, and the users simply believe it's on-premises, you can put this paper down, congratulate yourself, and know you've achieved what some can only dream of.

But getting there requires the right cloud deployment model to begin with. There are a few generally accepted models that you can consider, keeping in mind that you may actually want to use (or are already using) a hybrid mix of these:

- **Private Cloud** – With dedicated infrastructure and connectivity to only your organization, the private cloud provides organizations with a secondary data center to both host your applications, as well as act as a recovery site.
- **Central Cloud Access** – Using the public cloud (such as AWS or Azure) infrastructure to host your applications, you can centralize the connectivity to the application by routing all app traffic first through your organizations network, increasing security and providing some ability to monitor and control traffic.
- **Direct-to-Cloud** – Still using the public cloud, this model allows users to connect directly to the applications without needing to first connect to the organization's network.
- **SaaS** – This is the most familiar model, with applications like Office 365, Salesforce, Microsoft Dynamics, and others providing the application and the infrastructure, allowing users to connect directly, with no involvement of your organization's IT necessary. What does make this a viable choice is the fact that there is still a team dedicated to creating an optimal user experience.

Maximizing app performance depends a lot on your ability to attain and maintain visibility, and control – something the cloud doesn't normally offer. **Table 1** below shows a somewhat subjective representation of the visibility into, and control over, the application performance and user experience with each model.

Table 1: *Comparing Cloud Deployment Models*

Model	Visibility	Control
Private Cloud	Better	Better
Central Cloud Access	Good	Good
Direct-to-Cloud	Worse	Worse
SaaS	Worse	Worse

Keep in mind that no two cloud or SaaS providers are completely alike, so the values placed on each model are more a generality. And even with a rating of “better”, it doesn’t necessarily mean you’ll have complete visibility and/or control; it’s a stance relative to what you can get from models that are the “worse.”

So, which model provides the greatest app performance?

You might look at the table and simply say “Private Cloud.” But, it’s not that simple. The value you get from a particular model, such as SaaS, may outweigh the tradeoffs in visibility and control. There also may be applications constraints that come into play. For example, an internal application may need to be accessed from devices that don’t have any kind of VPN functionality, limiting your ability to use the Central Cloud Access model. Or a SaaS app provider may not have many options other than direct access to their cloud. You need to consider the benefits to the user experience each model brings, and then balance them against the level (or lack thereof) of visibility and control.

The value you get from a particular model, such as SaaS, may outweigh the tradeoffs in visibility and control.

The right model is actually any one of these... and most likely, a hybrid cloud mix as long as you’ve taken steps to address performance – accelerate app performance, limit bandwidth use (which saves costs), and achieve visibility and control.

The challenge is accomplishing all those steps.

In lieu of a third-party optimization and acceleration solution, you will need to work with your cloud or application provider to put as much visibility and control over the application performance as is possible in your hands. It’s difficult to near impossible in some cases (e.g. – Salesforce isn’t going to hand you the keys to their application metrics, right?). But, remember there are some metrics you can use in all cases to monitor application performance. By watching and measuring page load times, transaction times, and server up times, you have an ability to guess what the user experience is like, and have leading indicators around whether performance is less than optimal.

Going from Shadow to Strategy

The move to the cloud alone doesn’t guarantee application performance. And at a time when users are expecting sub-second

response times, you need to be sure your cloud strategy delivers the highest levels of performance possible. But with the lack of visibility and control, IT is often left reacting to issues well after they have impacted users.

Even so, IT is charged with crafting a strategy that provides the highest levels of performance and optimizing that performance throughout the life of an application, so choosing one or more of the cloud deployment models available is a solid first step, considering the benefits and tradeoffs of each, with the goal being how to maximize visibility and control over the variables that impact the user experience.

Identifying ways to monitor and optimize performance – whether using available performance metrics or by using a third-party WAN optimization and acceleration solution – will be a crucial first step to gaining visibility and, to varying degrees, control, eliminating the cloud's shadow on your application performance. ■

With nearly 20 years of enterprise IT experience, Nick Cavalancia is an accomplished consultant, speaker, trainer, writer, and columnist and has achieved certifications including MCSE, MCT, MCNE and MCNI. He has authored, co-authored and contributed to over a dozen books on Windows, Active Directory, Exchange and other Microsoft technologies. He has spoken at conferences such as the Microsoft Exchange Conference, TechEd, Exchange Connections, and on countless webinars and at tradeshow around the world.
