Seeing Through the Fog: Managing Application Performance in the Cloud

By Peter Sevcik and Rebecca Wetzel
February 2011

As your business applications recede from desktops into data centers and from there into the great beyond that is the public cloud, your view into the performance of important applications becomes increasingly shrouded in mist. The network interface becomes the best point at which you can measure performance, thus you should instrument your network to measure performance there. If you have good network visibility, it will enable you to inventory all your applications and baseline their performance before moving them into the cloud, and to measure performance continuously thereafter to ensure it meets your business needs.

This report describes the cloud and its inherent dangers for those inadequately prepared to move applications into it. We describe the essential role network-based performance measurement plays in maintaining service visibility when applications are drawn into the cloud, and we discuss what to look for in a measurement tool. Finally, we assess Riverbed's Cascade visibility solution.

This report was sponsored by Riverbed Technology

riverbed
Think fast

NetForecast Report 5106

What is "the Cloud"?

The term "cloud" is the latest name for an old idea—using remote rather than local computing power and storage to run applications that users access via the corporate WAN or the public Internet. From a user's perspective computing resources are "elsewhere"—in many cases outside an enterprise's control.

Applications consist of three components: *production*, *transport*, and *consumption*. Production encompasses computing and storage. Consumption is the work performed on PCs, laptops, and/or mobile devices that deliver applications to users, and transport is the networking technology that binds production to consumption.

Think of production as the "heavy lifting" needed to run an application. As Figure 1 shows, when computers were scarce and expensive, production happened via time sharing, i.e., renting time on computers owned by third parties located elsewhere. Production then moved to mainframes housed in data centers, and subsequently to remote distributed computing. In time, much production migrated onto users' desktops.

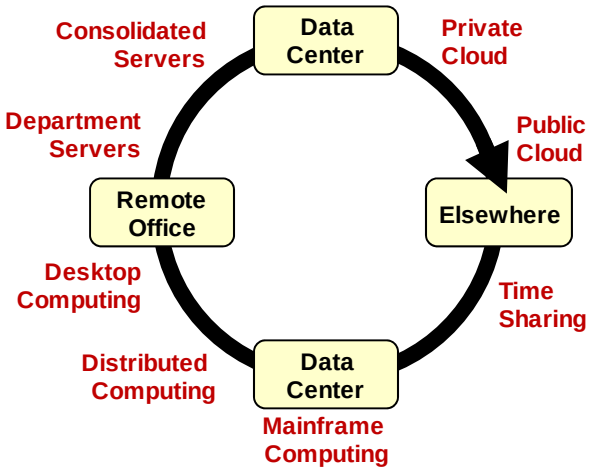

Figure 1 – Computing Production Circles Back to the Cloud

Then things began to come full circle as computing production migrated from the desktop to local servers, and those servers were then consolidated into data centers or moved into a private cloud. Today's trend is for heavy lifting to be relegated once again to third parties. This time the production resources live in public cloud service locations beyond the purview of the enterprise.

Performance Management Challenges in the Cloud

For decades network distances shrank as applications moved closer to users—often running on the user's desktop. With applications on the desktop, measuring application performance is moot because production and consumption happen in the same machine. But when applications run on a server in a remote office, measuring performance becomes helpful because production and consumption are separated and are bound by a network. In this environment, server performance statistics reflect the user experience closely enough that no special instrumentation is necessary. When production moves into the cloud, however, measurement not only becomes paramount, it presents a number of challenges.

If you are consolidating applications into data centers, the server measurements which served well in remote office-based server environments no longer provide an accurate view of performance because they cannot reflect the effect network distance exerts on the user experience. As distance increases (see Figure 2), latency increases, thus degrading the user experience. This problem has spawned an entire WAN optimization industry—but even with these technologies in place to improve application performance, it is still important to measure the user experience so you can further improve it.

Unfortunately, adding WAN optimization solutions to mitigate the performance hampering effects of distance makes it harder to accurately measure performance. Compound that with the increasingly complex, multi-tiered application architectures common in today's data centers, and the performance management challenges generated by moving applications into the cloud rise exponentially.

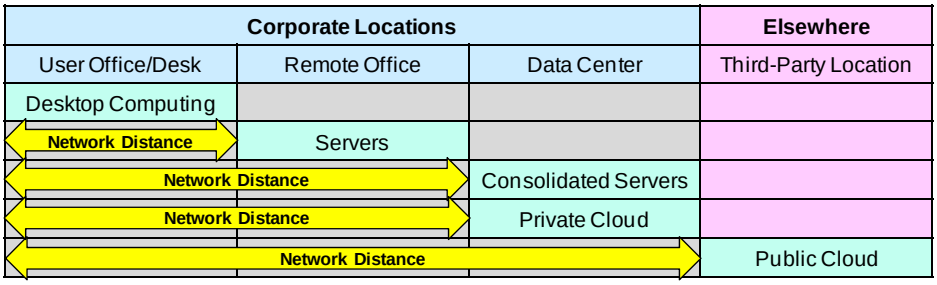

Figure 2 – Moving Applications to the Cloud Increases Network Distance

Cloud computing presents yet more performance management challenges. Cloud services add layers to production, yet they provide little or no instrumentation to shed light on performance. When computing is performed in a cloud owned by third parties, the interface becomes the only point at which performance can be managed, and that interface point is a network connection.

Don't Fly Blindly into the Cloud

It is unstoppable that more and more applications will migrate into the cloud—and server and data center-focused performance management tools simply don't address the performance management challenges the cloud environment presents. If you fail to address these challenges, you risk flying blindly into the cloud and very possibly into

unseen obstacles because you lack the visibility to ensure that applications deliver required performance before, during, and after migrating into the cloud.

Because the best place to gain the visibility needed to manage performance in a cloud environment is from the network, ***network-based application performance measurement tools constitute the best means to maintain service visibility***. To keep from striking unseen performance obstacles, it is important to have WAN-based measurement capabilities before you undertake data center consolidation and move applications into a private cloud, and it is vital before you move applications into the public cloud.

With the right network-based performance measurement tools and processes in place at the outset, you will be prepared to enter the cloud. Otherwise the cloud migration, especially to the public cloud will become a performance management crap shoot.

To prepare for the migration, you must be able to:

- Identify all of the applications running on the network,
- Map dependencies among the elements involved in delivering your applications,
- Measure traffic volume along each application delivery path, and
- Measure response times along each application delivery path.

Discovery: Identify Your Networked Applications

Before you move applications into the cloud, you need to know the good, the bad, and the ugly facts about what is running over your network. Without this information you risk promulgating bad application behavior, or inadvertently turning off or "breaking" applications, which could cripple your business.

Knowledge is the bedrock upon which good performance is built, but it is short supply within most enterprises. After gaining visibility into network traffic, we often hear network managers say things like: "We were shocked to learn that what was actually happening on our network was far different from what should have been happening." For example, to its surprise a financial services firm we interviewed learned that some data streams were being resent up to five times. This inefficiency would have been impossible to detect without a performance monitoring tool.

To stay out of trouble, we strongly recommend you perform application discovery before heading into the cloud.

Map Your Application Flows and Measure Traffic Volumes

Before, during, and after you move applications into the cloud, you need to know what is happening where. To understand this, you need to map application flows within the network. Without mapping you can't identify dependencies among the elements involved in delivering applications, and you risk degrading performance and/or disrupting application delivery.

Figure 3 shows just how complex these dependencies can be. All of the interdependent parts must function properly for the application user to have a good experience. Each mapped flow consumes resources that must be handled by support services, and each flow that will be supported by a public service represents traffic to and from the cloud that must be properly provisioned. When you map your flows you may find some that don't belong on your network, such as flows created by users downloading movies for recreational use.

Traffic volumes should be rational. By understanding your traffic volumes and flows, you can determine if there are resource uses that should not be part of a functional tier.

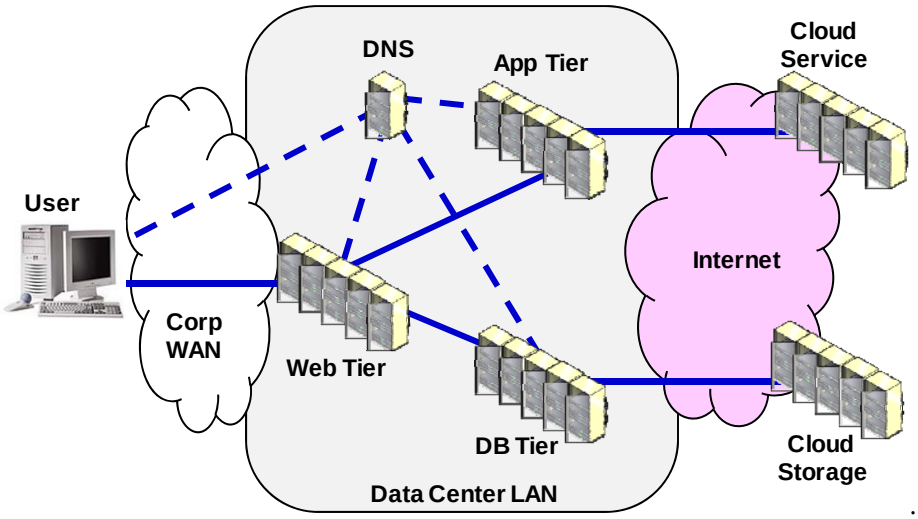

Figure 3 – Example of Application Flow Mapping

Measure Response Times

Before you move applications into the cloud you need to flag areas of potential performance weakness, so you can implement measures to prevent poor performance. To do this you need to benchmark the "normal" response at each of your dependency points. Without this information it is hard to determine where actual or potential performance problems lurk.

After moving applications into the public cloud you should perform another baseline to determine what the "***new*** normal" is, and perform ongoing performance measurement to help find and fix problems as well as continuously improve performance.

If you are using public cloud services, you should monitor response time as far as you can see. You should ask your service provider to give you reports that will complete the performance picture.

This not a one-shot effort—a performance history allows you to identify what is acceptable and to take action when behavior degrades.

Hypothetical Example

To illustrate the importance of mapping and measuring flows, let's explore the hypothetical example of a company staff portal. Our hypothetical company runs an internal web portal that serves as a central information hub for all employees and contractors.

Many of the portal's applications are owned and operated by the company and housed in two data centers under corporate control, while other applications are provided by third parties as cloud-based services. Human resources, for example is responsible for providing information such as a company holiday calendar, personnel policies, and security procedures. The finance department is responsible for operating and managing an online timesheet application. These web-based services are hosted on servers in the two data centers and provide acceptable performance to the users in remote offices.

However, some applications are supplied by third parties as cloud-based Software as a Service (SaaS) applications. These include applications from Teleo for internal job postings, WorkforceLogic for contractor management, Staples for office supply ordering, and Iron Mountain for desktop backup services.

Users simply tab to these services from within the web portal. They don't know that the browser tab dealing with contract management, for example, is running on an additional service tier in the cloud. But users expect the same application response time as from the internal services. The goal is to make in-house and outsourced applications perform the same. A performance management tool must be able to easily track and compare response times accordingly.

If outside services fail or don't perform correctly, then a major section of the staff portal will not support the business.

Finding the Right Visibility Toolset

The best solution combines all of the above capabilities into a small number of tools that are well integrated, so you do not have to manually cobble together results from many tools

The tools must provide dependency mapping for:

- User connections to the data center
- Cross-tier connections
- Connections from data center to cloud services

. . . . and must be able to measure these parameters:

- ✓ Server utilization
- ✓ Bandwidth utilization
- ✓ Transactions processed
- ✓ Server errors
- ✓ Availability
- ✓ Server query response time
- ✓ Packet ping time to user locations
- ✓ Traffic flow data (by application)
- ✓ TCP transaction response time
- ✓ End-user page response time

The key to choosing the right tool set is to place your performance tool choices into a "big picture" context (or framework) that helps you define what you need from your tools—and map those needs to the current crop of tools.

Here's the underlying concept of the NetForecast Performance Management Tool Framework. As Figure 4 illustrates, there are three application performance management functions:

1. Gathering performance data,
2. Interpreting the data, and
3. Using the results to support the organization.

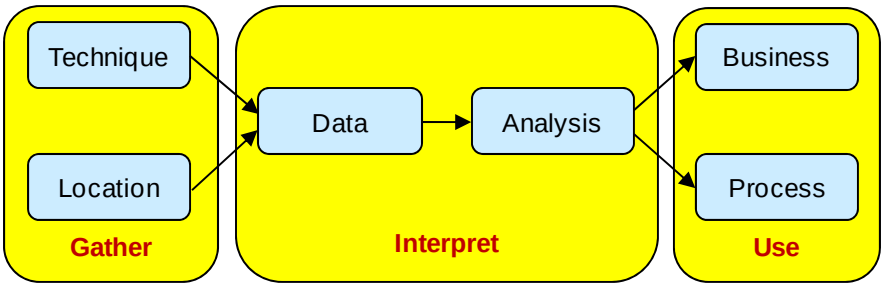

Figure 4 – Performance Management Tools Framework

Gathering the Data: There are two key aspects to gathering application performance data—technique and location. Technique determines how to measure performance, and location determines where to measure it.

In the case of a public cloud, you will need to receive at least some data from your cloud provider. The cloud provider should provide you with raw technical data that you can feed into your tools and management processes. If the cloud service provider does not offer technical data, then you should ask them to provide comprehensive reports to you that include the parameters listed at the beginning of this section. This approach augments, but does not entirely replace your own monitoring.

Interpreting the Data: Data interpretation involves first determining what to measure (i.e., what data to gather), and then to analyze the data to make sense of it. Your tool should be able to interpret data about and impacts of typical technologies that reside in today's data center such as WAN optimized and non-optimized traffic and both traditional and virtualized servers.

Using the Data: When all is said and done, the results of the analysis must support IT processes—and ultimately the business as a whole. Information must be based on analysis of performance by time of day, location, and business function. It must provide trend analysis over long periods, integration and correlation of data from multiple sources (e.g. other tools and outside data feeds)—and it must be stored in a long-term data repository for data mining.

Some tools are good on the front end of this functional flow, but require lots of human gray matter to make sense of the data, put it to use, and keep a record of it. Although there is no substitute for a smart geek, there is much to gain from incorporating geek-smarts into automated tools that analyze data and apply it to improve performance.

The back-end functional flow (analysis, process, business reports) is where a good tool can make your life easier. A tool that automatically tracks and reports on performance can do what you don't have time for, and free you up for more meaningful work. Such a tool can automatically generate daily reports detailing how well you are managing performance.

The Riverbed Cascade Solution

The Riverbed Cascade solution (see Figure 5) is a passive network performance monitoring system that sees and reports on user interactions with application delivery infrastructure such as application servers, web servers, and load balancers. It collects network flow data, identifies applications and users, and applies behavioral analysis to help you interpret that data, and use it to support your business. The Cascade solution learns what "normal" is and sends alerts when changes indicate actual or potential problems. It monitors all communications to and from the data center, including cloud services. For more information about Cascade visit http://www.riverbed.com/us/products/cascade/.

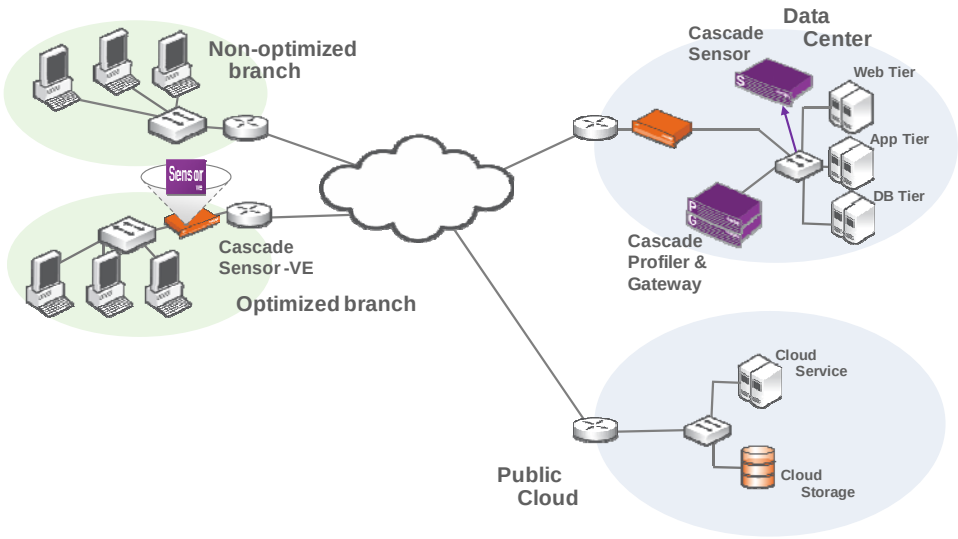

Figure 5 – Example of a Riverbed Cascade Deployment

NetForecast Assessment of Riverbed Cascade

Figure 6 maps the Riverbed Cascade capabilities to the NetForecast performance management tools framework (see page 5). The purpose of this matrix is to help you assess at a glance whether the Riverbed Cascade solution meets your needs.

Gather		
Technique	Touch: Passive/Active	Passive
	View: Internal/External	Internal
Location	Remote LAN	✓
	Private WAN points	Optimized/Non-opt, Cloud/Non-cloud
	Data center LAN	✓
Interpret		
Data	Bandwidth utilization	Optimized/Non-opt, Cloud/Non-cloud
	Network time to user locations (RTT)	Optimized/Non-opt, Cloud/Non-cloud
	Traffic flow data (by application, location, user)	Optimized/Non-opt, Cloud/Non-cloud
	TCP transaction response time (by application, location, user)	Optimized/Non-opt, Cloud/Non-cloud
Analysis	Functional performance tracking	✓
	Trend analysis over long periods of time	✓
	Correlation analysis across measurement data types	✓
	Analytics automate converting data into information	✓
Use		
Process	Incident: resolving performance degradation	✓
	Capacity: proper service resources	✓

Figure 6 – Riverbed Cascade Performance Management Capabilities

Conclusions

What you can't see CAN hurt, or at least hobble performance. Therefore, when you move your applications into private and/or public clouds, you need to measure performance in the network. This visibility will enable you to inventory all your applications and baseline their performance before moving them into the cloud, and it will enable you to measure performance continuously thereafter to ensure it meets your business needs. In NetForecast's assessment, Riverbed Cascade is indeed a good solution to provide the visibility needed to plan for migration into a cloud environment and to manage performance once your applications are at home in the cloud.

About the Authors

Peter Sevcik is President of NetForecast and is a leading authority on Internet traffic, performance, and technology. Peter has contributed to the design of more than 100 networks, including the Internet, and holds the patent on application response-time prediction. He can be reached at peter@netforecast.com.

Rebecca Wetzel is an Associate of NetForecast and a veteran of the data networking industry with unparalleled inside knowledge of the Internet service and product markets. She works with network product vendors and service providers to develop and implement product strategies. She can be reached at rebecca@netforecast.com.

NetForecast helps enterprises and vendors understand and improve the performance of networked applications.

Additional information is available at: www.netforecast.com