



# Windows<sup>®</sup>IT Pro

## Monitoring and Managing Services in the Cloud

By Brian Desmond



A QUEST SOFTWARE<sup>®</sup> COMPANY

## → Contents

Monitoring and Managing Services in the Cloud .....	2
Monitoring Service Availability .....	3
Service Quality Monitoring.....	3
Dynamic End-User Devices.....	4
Service Provisioning .....	4
Auditing and Reporting.....	5
Assessing the Cloud.....	6



A QUEST SOFTWARE® COMPANY


# Monitoring and Managing Services in the Cloud

By Brian Desmond

The cloud presents many opportunities for organizations to expand service offerings, save money, and deliver increased service to their end users and customers. The rapid proliferation of cloud offerings has placed increasing pressure on IT administrators as decision makers learn about the value of the cloud and look to move the organization to it. There are many key points important to making an educated decision about moving to the cloud, but one important point is monitoring and managing the cloud provider as well as the organization's connectivity, usually with the Internet. The touch points for monitoring and managing a cloud provider will vary depending on whether you're using a cloud provider to host servers or to deliver a specific application; but, fundamentally, many of the touch points are the same.

While many cloud services come with availability guarantees in the form of Service Level Agreements (SLAs), it is less common to find accurate and readily accessible reporting from the service provider as to how they met their SLAs in a given time period. Likewise, while the cloud service may be available, if the organization's Internet connectivity is not sufficient to sustain the quality and degree of connectivity required, the perceived availability of the cloud service may suffer substantially.

On-premise services often take advantage of central authentication services (i.e., Active Directory) and are also able to connect directly to human resources databases and identity management systems in order to obtain user information and receive provisioning and deprovisioning requests for user access in an automated fashion. When services move to the cloud, this information needs to be transmitted securely and only when it is deemed necessary and of a low enough risk that it can leave the organization's network. More importantly, end-user access needs to be provisioned and deprovisioned from cloud services in order to ensure timely access, reduce risk, and limit the amount of effort required by IT administrators to manage the cloud service.



As servers and applications move off premise into the cloud, security, auditing, and reporting requirements must still be maintained. When a service is delivered locally from an on-premises datacenter, it is often much easier to report on user access, collect audit trails, and generally satisfy regulatory compliance requirements. With cloud services, the entire security ecosystem is no longer under the organization's control so proper planning and tooling must be in place to collect the necessary security data for reporting, as well as monitor for breaches and compliance challenges. As end users supply their own mobile devices with increasing frequency, it is even more important to understand where data is being accessed and how in order to manage the risk of data loss.

## Monitoring Service Availability

Any cloud service that is important to an organization should have a commercially backed SLA that includes penalties for the service provider in the event the SLA is breached. While some service providers will provide in-box reporting capabilities and dashboards, it is also prudent to validate these metrics, track availability against SLAs, and alert on outages in a timely basis so that resources such as the help desk can be prepared to manage these situations.

There are numerous components to the availability of any service whether it's delivered as a service cloud, hosted on servers managed by the organization in a public or private cloud hosting scenario, or in an on-premises environment. Every service is different, so it's critical to cater the monitoring to the specifics of the server or service, as well as to select a service monitoring product that has native support for that class of service (e.g., email, web site). At a high level, any service will require monitoring of basic metrics such as general reachability (via ICMP ping, for example), latency, and so forth. If you're monitoring servers hosted in the cloud, you'll need to monitor the same basic health metrics that apply on-premise, such as CPU utilization, memory pressure, and storage performance. In the case of public cloud services where billing is based on CPU utilization, it's even more important to track this data for accurate forecasting.

Aside from basic metrics, every service will require some customized monitoring. A simple web site or


web application might monitor the time it takes for the web server to return a response as well as ensure that the correct response is always returned. If an incorrect response (such as an HTTP 500 server error) is returned, then the monitoring tool should alert the service owner and other interested parties immediately.

On the other hand, an email service will have numerous monitoring touch points. These touch points will likely include functionality such as message routing. That is, how long it takes for an email to travel between on-premises SMTP servers and the cloud service and to ensure that mail is being accepted at all times and not bounced due to an error. Depending on the type of email service in use, other services might include IMAP/POP availability monitoring with a synthetic logon, or validating the availability and performance of the web mail interface.

Aside from basic alerting and monitoring, it's important that the monitoring tool be able to provide metrics in an easy-to-read format that can be compared with the service provider's SLAs, as well as the organization's internal service level commitments. It's also useful to periodically compare these metrics with any reporting the cloud provider might offer about its service, either at the application or server level. Service owners should review this data regularly to ensure that service levels are being met. For public cloud scenarios, the ability to independently approximate resource utilization metrics tied to billing will inevitably be of great use, also.

## Service Quality Monitoring

Some applications and services may be especially susceptible to service degradations, if the quality of the organization's Internet connection is poor. An example of such a service is a cloud-based Voice over Internet Protocol (VoIP) provider. Voice traffic is susceptible to both latency and jitter. Latency is the amount of time it takes for a network packet to travel from the source to the destination; jitter is the variability in latency. Videoconferencing is another example of an application that is susceptible to degradation from latency and jitter. To ensure that applications such as VoIP perform as expected, it is important to monitor these metrics and to configure network equipment to perform Quality



of Service (QoS) traffic management for sensitive network traffic.

When a service or application is moved to the cloud, the network traffic travels over the organization's Internet connection. It's important to plan for this because many applications require significant bandwidth either individually or at scale. If proper planning does not occur, then it is very possible that the deployment of the cloud service might overload the organization's Internet connection. Some public and private cloud providers allow organizations to create Virtual Private Network (VPN) tunnels over the Internet so that traffic to the cloud is encrypted and the path is controlled. VPN tunnels come with increased processing and management overhead, so the need to monitor connection quality is all the more important.

One common example of an individual application that could cause the Internet connection to be overloaded is video streaming. Take, for example, a scenario where an organization moves the streaming of an internal event from an on-premises system to the cloud. Each employee accessing the video stream would require Internet bandwidth. On the other hand, some applications might not seem to be large consumers of Internet bandwidth at first glance. Email is one example that might come to mind where individual users are not likely to cause a substantial amount of Internet bandwidth utilization, but a large number of users in one office collectively accessing email or another application could generate a measurable amount of Internet bandwidth utilization.

To make an informed decision about Internet bandwidth utilization by a potential new cloud service or by connections to servers located in the cloud, it's critical to have historical data showing average and peak utilization of the organization's Internet connection(s). Based on this data, an informed decision can be made as to whether or not capacity must be added to accommodate the application. In addition to evaluating this data while considering the use of the cloud, it's also critical to review trending data regularly (e.g., quarterly) to determine whether or not the organization's bandwidth utilization has grown organically to the point where additional capacity will be required.

## Dynamic End-User Devices

Over the past several years, organizations have faced an increasing trend of mobile devices that are often supplied by the end user with an expectation that the device will have access to information. This data might be available through a web-based interface or through an application written specifically for the device (e.g., an iPhone or iPad). The risk introduced by end users supplying their own devices can be significant, in addition to increased support costs.


When company information is accessed via a device that isn't managed by the organization's IT department, the ability to control the lifecycle of the data is greatly reduced. Mobile devices are inherently susceptible to data loss risk given they are easy to lose. These risks can be managed to a degree with basic security constraints and training for end users.

In the case of end user training, it is important to communicate the varying degrees of impact that information might have. When users understand this, and data is labeled, they can make informed decisions about where to access documents and data, where to store copies of them, and so forth. For example, company policy may prohibit accessing High Business Impact (HBI) information on a mobile device.

In addition to the human side of the problem, there are technical solutions to managing these devices. Simple approaches such as requiring encryption of the device and a security PIN to unlock are a good first step. Some organizations may decide to deploy a management agent or tool on each device in order to maintain a greater degree of central control for both policy and application management purposes.

## Service Provisioning

In order for end users to take advantage of any service or application, whether it is hosted on-premise or in the cloud, the user typically requires some degree of access to be provisioned in the application. In an on-premises application, this provisioning is often simply linking a user's existing Active Directory account to an identity in the application. For cloud-based applications, this might require a separate username and password



to be provisioned for the user in the application, or complex identity federation to be configured.

There are numerous ways to solve the user provisioning problem, but fundamentally they boil down to three approaches. The first approach is to provision separate user accounts in each cloud service and rely on the user to maintain separate usernames and passwords for each service. For organizations using a very limited number of cloud service providers or cloud-hosted servers, this may be a manageable approach, but support costs will certainly rise as the number of credentials a user has to maintain increases.

In the case of applications running on servers hosted in a public or private cloud, it may be possible to establish a VPN tunnel such that the cloud servers are able to integrate directly with the organization's Active Directory. If this is the case, users may simply be granted access to the server or application on the basis of a pre-existing Active Directory account. This approach is not always possible depending on the architecture of the cloud offering.

The third approach is to leverage identity federation. Federation enables users to access cloud applications using their on-premises credentials (i.e., from Active Directory) using standard protocols via HTTP. With federation, when users attempt to access the cloud application, they are redirected to a federation web server in their organization. The federation server authenticates the users, and if they are successfully authenticated, they are redirected back to the cloud service. The application in the cloud receives information ("claims") about the users during this process and can then allow the users access to the application as appropriate.

In all three approaches, a process must be in place to provision and maintain user information in the cloud service. This process often takes the form of an identity management tool that is capable of synchronizing data to various systems from an authoritative source such as a Human Resources database. In some cases (i.e., without federation), it may be possible to synchronize passwords from an on-premises directory to the cloud service. However, this approach introduces risk when considering that user passwords that

control access to sensitive information are in the hands of a third party.


In addition to provisioning users to the cloud service, it is important to consider how user access will be deprovisioned. Deprovisioning is especially important because many cloud services are priced on a per user basis. Thus, if unnecessary user accounts exist in the service, costs will be higher. Deprovisioning may take two forms depending on the cloud service in question. In some cases deprovisioning may simply consist of removing a user's access to the service when they are terminated. On the other hand, it may be necessary to manage permissions and roles inside the cloud service as the user changes jobs and responsibilities in the organization.

## Auditing and Reporting

As data moves into the cloud, it will become even more important to be able to audit and report on who has access to that data, as well as any modifications to it. Compliance requirements continue to grow and the burden of meeting them increases proportionally. As part of the process of evaluating a cloud service or application, decision makers must evaluate the impacts of the service offering on compliance and regulatory requirements. This evaluation may take the form of access to logging information from servers or security devices (e.g., virtual firewalls) in the cloud, or audit trail information from a cloud-based application or service. One example is that some organizations are moving extremely sensitive information (i.e., Personally Identifiable Information [PII]) to cloud-based Human Resources and Payroll solutions.

Once the risk assessment has been performed, it will be important to plan how the data necessary for proper reporting and auditing will be collected and preserved. Some cloud services may only offer the ability to retrieve this data on-demand inside of the application, while other more complex public/private cloud solutions may offer an interface for programmatically retrieving auditing information.

Once the information has been collected, it will need to be stored and made accessible for reporting. Some reports should be reviewed regularly, while others may only need to be retrieved as needed. An example of a report that should be



reviewed regularly might include information about which users have elevated (e.g., administrative) access to a system. On the other hand, audit trail information for documents accessed might only be required annually to satisfy an audit. It is very important to many organizations to have the necessary tools and capabilities to track this type of data, as well as report on it.

## Assessing the Cloud

The promise of the cloud is winning the minds of many IT decision makers, but to make a successful move to the cloud for servers and virtual machines or cloud-based services applications, a number of variables and risks need to be evaluated and then managed over time. These include core network infrastructure components such as bandwidth and latency, as well as security questions around mobile devices, user access management and authentication, and security auditing and reporting.

Network management is critical to ensuring that access to the cloud is delivered to the degree of speed and quality that end users are accustomed to when accessing on-premises applications. When servers and applications move to the cloud, there is frequently no longer a capability to maintain

direct access to a central authentication service such as Active Directory. Consequently, it becomes important to examine how authentication will be performed, as well as how user access will be managed.

Finally, security requirements that have traditionally applied to on-premises applications are still applicable and even more critical than before with cloud-based servers and applications. The ability to access auditing data and create reports for monitoring, compliance, and audit purposes is crucial. As end users begin accessing data from mobile devices that are often not supplied by the company, understanding where data is accessed from as well as controlling how and where it is stored is extremely important.

**Brian Desmond** is senior consultant with Moran Technologies. He has been a Microsoft MVP for Directory Services since 2003 and is the author of *Active Directory, 4th Edition* from O'Reilly. Brian is an Active Directory and Exchange focused consultant leading and delivering on projects primarily for large enterprise (40K – 500K seat) customers. His website can be found at [www.briandesmond.com](http://www.briandesmond.com).