Realtime publishers

Deploying and Managing Private Clouds

The Essentials Series

Managing for the Long Term: Keys to Securing, Troubleshooting and Monitoring a Private Cloud

Dan Sullivan

Copyright Statement

Managing for the Long Term: Keys to Securing, Troubleshooting, and Monitoring a Private Cloud

Private clouds are dynamic systems with constantly changing application loads, hardware components, and groups of users. Managing for the long term requires that you adopt tools and procedures that allow you to secure cloud resources, troubleshoot components, and rapidly perform root cause analysis, as well as monitor key areas to ensure continued availability and meet service level agreements (SLAs). In this final article in the series, we will consider three topics that will be a constant concern for cloud managers and administrators:

- Securing a private cloud
- Troubleshooting a private cloud infrastructure
- Monitoring critical components of a private cloud

With the right tools and procedures in place, these tasks can be accomplished efficiently and effectively even as the size of the private cloud grows and usage increases.

Securing a Private Cloud

Many security practices common in IT work well in cloud environments. There will be, of course, cloud-specific adaptations and practices, but for the most part, the principles are the same within and outside of a cloud. Three topics of particular interest in private cloud security are:

- Identity management
- Image management
- Network security

These areas touch on three distinct aspects of private cloud computing: who is all allowed to use the cloud, what is allowed to run in the cloud, and how the cloud infrastructure is protected.

Identity Management

Managing user identities is a fundamental process. Within a private cloud, identity management combines a number of security functions: authentication, authorization, and some amount of auditing and logging. They are all based on the concept of a user as an agent who is allowed to perform operations in the cloud.

Authentication is the process of verifying who a user claims to be. Simply because a user identifies herself as a systems administrator is insufficient reason to allow this person to execute root-level operations. Authentication is often based on knowledge of a password or possession of a token. Directory services in a cloud can be used to track user identity information and store authentication data, such as encrypted passwords. It can also be used to store information about authentication services that perform verification operations before granting access to the cloud.

Once a user has been authenticated, there may be restrictions on what that person can do. For example, most users will be allowed to select an image from the service catalog and run it on a virtual instance within the cloud. Some, but not all, users may be able to install additional software on an instance and save the new version to the service catalog. Still other users may have privileges to alter billing and accounting records to correct for errors, such as forgetting to shut down a server when a job was complete and being charged for the additional time. These different levels of privileges are associated with varying authorizations.

In addition to supporting security, identity management systems are useful for accounting purposes. Auditing records and operational logs with identity information can be used to determine who performed what operations in the cloud. This is useful for both forensic operations as well as cost accounting.

Image Management

Besides knowing who is running operations in a private cloud, we need to be able to control what kinds of applications and operating systems (OSs) are run in the cloud. As a general starting point, we will want to restrict applications to those that meet a minimal set of criteria for running in the cloud, such as:

- Running software from an approved set of applications
- Running in service to business operations
- Not running any type of malicious software
- Not violating privacy, confidentiality or other governance requirements
- Not violating software license agreements
- Not undermining auditing and accounting data collection procedures

To meet these requirements, the service catalog must implement access controls to limit who can add, remove, and modify images in the catalog. The images in the service catalog should be periodically scanned for malware and vulnerabilities and patched as needed.

Network Security

Applications running in the cloud should operate within as secure a network environment as possible while still allowing for necessary business services. The perimeter of the cloud should be controlled to mitigate the risk of external threats. For example, only virtual private network (VPN) users may be granted access to cloud resources when a request comes from an external network source. Network traffic should be monitored to watch for suspicious activity, such as large file downloads outside the cloud outside of normal business hours. Similarly, frequent failed attempts to authenticate to a cloud service or methodical probes of ports can indicate unauthorized attempts to access cloud services. Monitoring network and server activity helps with troubleshooting as well as with security.

Troubleshooting Private Cloud Infrastructure

In order to keep a private cloud functioning, we need to be able to quickly identify problems and correct them. There are two key functions we need from our network and infrastructure management software: support for problem detection and root cause analysis.

Monitoring software should be in place to alert systems administrators when a problem condition exists. Rules can be established to define thresholds for problem events. For example, if a certain number of attempts to ping a server fail in a given time period, an administrator may be alerted. Similarly, if the number of write errors to a disk exceeds some threshold, an alert is sent to inform a manager of a potential hardware problem. In many cases, though, the cause of a problem may not be obvious and an alert may be more of an indication of a symptom than of an underlying problem.

Root cause analysis is the process of identifying the underlying cause of a problem. For example, if an application generates an error because it cannot update a database record, there may be multiple causes:

- A hardware problem with the storage array that is preventing data blocks to be written from cache back to the disk
- A network problem between the database server and the storage array, which is preventing the storage device from acknowledging that the data block has been written to disk
- An application error that fails to complete a logical transaction even though server, storage, and networking services are all working correctly

When troubleshooting problems in a cloud environment, it is helpful to have tools that can quickly isolate particular aspects of a problem, such as determining whether there is connectivity between a server and a storage device, if a storage device can correctly read and write from a disk, and whether a server can successfully write blocks of data to storage.

Tools that support root cause analysis are important for maintaining adequate levels of performance and availability. The longer a server, storage, or network problem persists, the more users it can adversely affect. Also, the longer it takes to diagnose a problem, the greater the cost of diagnosis. Quickly isolating the cause of a failure in a private cloud helps to improve availability and to keep maintenance costs under control.

Key Areas to Monitor

Long-term private cloud management is best built on a solid foundation of operational data. Businesses have a wide range of use cases for running jobs in a private cloud. Different departments may have different peak demand periods, and workloads across departments will vary from day to day. The more data you have about these usage patterns, the better able you are to manage growth.

Four areas of particular interest for private cloud monitoring are:

- Server utilization
- Network bandwidth utilization
- Availability
- Image use

Server utilization is a measure of how much available CPU time is actually used for productive work. Improving this one metric is a common business driver for adopting a private cloud. Too often, we purchase single servers for single applications and find ourselves with excess capacity. Ongoing monitoring can help identify periods when some servers can be shut down to save on power without adversely affecting performance. It can also provide information on common patterns, such as times of the week, month, or quarter where demand is unusually high or low. With detailed information about server utilization and network bandwidth utilization, private cloud managers can better assess their ability to support new business services that would put additional demands on the cloud.

You should also monitor image use. Doing so can help you to understand patterns of application use and support compliance with software licensing. In some cases, image management can help identify situations in which excess software licenses have been purchased and can be scaled back in the future.

Summary

Long-term management of a private cloud depends on several factors, such as securing private cloud infrastructure, troubleshooting operation problems, and monitoring assets and usage patterns. By starting with the right tools, you can secure and monitor a private cloud efficiently and effectively. Care should be taken when selecting tools to maximize their use; ideally a tool that supports troubleshooting will also have adequate logging features to support monitoring efforts as well.