Realtime publishers

Deploying and Managing Private Clouds

The Essentials Series

Tips and Best Practices for Managing a Private Cloud

sponsored by

Dan Sullivan

Copyright Statement

Tips and Best Practices for Managing a Private Cloud

Private clouds are a relatively new model for delivering computing and storage services, but this model builds on a long history of IT infrastructure management. Private clouds are a delivery model that builds on well-established IT practices, such as virtualization, network management, systems administration, and operations management. These practices have developed over years of repeated use and refinement in a wide variety of application areas. We draw from these practices here and highlight three areas that are especially applicable to private cloud management:

- Establishing policies and procedures
- Standardizing hardware and application stacks
- Formalizing discovery and monitoring procedures

Together, these help to establish a sustainable management framework that promotes the efficient use of cloud resources without creating unnecessary management burdens for IT staff.

Establishing Policies and Procedures

The first set of tips and best practices is not about some arcane technology that enables cloud computing but is instead about management practices. In many ways, the best hardware in the optimal configuration will only continue to perform well for so long before changes in demands, hardware failures, and software revisions start to adversely impact operations. Private clouds require a minimal set of operating policies and procedures that are implemented by automated systems and support staff to ensure the private cloud continues to deliver computing, storage, and networking services.

Some of the most important policies and procedures entail:

- Cost allocation and reporting
- Image management
- Security and patch management
- Monitoring
- Backup and disaster recovery

Cost Allocation and Reporting

Cloud computing allows you to efficiently allocate computing and storage resources on demand on an as-needed basis. When the finance department has a large number of end-of-quarter reports to generate, they can allocate multiple virtual servers in the private cloud for as long as needed to complete the reports. A data warehouse project with a large amount of legacy data can use the cloud for the initial data extraction, transformation, and load process to rapidly add legacy data to a new data warehouse. When advertising campaigns are more successful than anticipated and a there is a surge in orders, the online business can scale up by adding application servers and Web servers to accommodate the demand. Unless all of these services are provided without charge to end users, you must have a mechanism in place to track usage.

A cost recovery system can use data from the self-service management system to track which users are allocating virtual servers, how long they run, and which applications are run on the virtual servers. The latter is important to recover the cost of software licenses. Similar data is required on the amount of data storage used over time as well as the amount of network bandwidth used while running applications in the private cloud.

Policies are needed so that IT providers can plan to recover their costs and possibly build capital to finance additional infrastructure purchase. Users need these policies so that they can plan how to efficiently use the cloud. An advantage of cost recovery policies is that they can be used to distribute jobs across time. For example, if the cost of an hour of CPU time is the same at all times of the day, users have no incentive to run their jobs at any particular time of the day. If, however, the cost of a CPU hour was 50% less during non-business hours, users with batch reporting jobs might move their jobs to off-hours leaving more resources to time-critical applications.

Image Management

Part of a private cloud's service offerings is a service catalog. This set of virtual machine images is available for use in the cloud. Policies and procedures should be in place that define what types of images will be available in the service catalog as well as rules governing the use of privately created and managed images in the cloud.

Policies should define a process for adding new images to the service catalog and reviewing, and possibly removing, images. The goal is to maintain the set of images that are needed by users while staying in compliance with software licenses and reducing security risks to vulnerabilities that may exist in the operating systems (OSs) and applications within these images. This begins to get into the realm of security within the private cloud.

Security and Patch Management

The need for preserving the confidentiality of information, the integrity of data, and the availability of resources are the key drivers behind IT security. A private cloud should build on existing security policies, especially with regard to:

- User authentication and authorizations
- Software allowed to run on IT-managed hardware
- Vulnerability scanning
- Operations monitoring
- Patch management

A private cloud by definition is restricted to a specific set of potential users. Policies and procedures should be in place to ensure that only qualified users are allowed to access cloud resources, that authorizations to use software and hardware are aligned with a user's roles and responsibilities, and that those authorizations and privileges can be easily modified as needed.

Policies can also be used to balance the need of IT administrators to control which applications run in the cloud with the specialized needs of some users. For example:

- If a department hires a team of consultants to design a custom database application, what kind of review process is required to add it to run in the cloud?
- Can users run any application that uses a standard database management system, such as Microsoft SQL Server?
- What if it uses a database management system not supported by IT?

Planning for how to make decisions such as this are best done while planning for the private cloud; this helps to reduce the need for ad hoc decision making with regards to policies and procedures.

Complex software can harbor vulnerabilities that can be exploited for malicious purposes. Vulnerability scanning is an established practice of checking deployed applications and OSs for known risks. This type of practice should continue with private clouds. Both public images in the service catalog and privately managed images should be checked according to a policy-defined schedule using vulnerability scanning tools that meet functional requirements defined in those policies.

Policies should also define the type of operational data to collect and the frequency with which it should be collected. The goal of this policy is to ensure IT administrators have the information they need to optimally manage the private cloud on a day-to-day basis. This policy also provides baseline data and trend information that managers can use for planning for the long-term growth of the private cloud.

Another policy should govern the patch management process and the related process of rebuilding images. After an image is built and stored in the service catalog, there may be OS upgrades and patches to applications that should be applied. A policy should describe conditions under which a patch is considered critical and should be applied immediately; it should also define routine patch schedules for non-critical updates. As with other policies, it is important to have this policy in place when deploying a private cloud to reduce the need for ad hoc policy making.

Backup and Disaster Recovery

A private cloud may be used for production operations, so it is important to have a backup and disaster recovery policy in place. The backup policy should define what data is backed up, how long backups are kept, as well as costs associated with those services. Similarly, in the event of a catastrophic failure of a private cloud, a failover plan should be in place. This plan may include using multiple data centers to host a private cloud or running jobs in a more conventionally organized cluster environment with manual management of jobs. The details of how to implement backup and disaster recovery will vary by your needs and resources, but it is essential for business continuity planning to have some policy in place.

Standardizing Hardware and Application Stacks

Another set of best practices focuses on standardizing hardware and application stacks. It is not that a variety of hardware or software is necessarily a bad thing, but it often requires additional time to manage. Consider a simple scenario: Suppose you build a cloud with servers from three vendors with different network and storage controllers. In order to minimize downtime, you maintain spare parts; however, you have multiple configurations, so you must maintain a larger set of spares than if you had a single standard configuration. The additional overhead does not stop with hardware. It is not hard to imagine that one configuration of Linux might work optimally given one hardware configuration but sub-optimally in another configuration and, as a result, you start to maintain two or more configurations.

The management objective with regards to standardizing is to have the minimal number of distinct hardware and software configurations that meet all user requirements. Fortunately, it is fairly easy to standardize hardware, especially if you are purchasing new servers and storage arrays. Even if you are working with legacy hardware, you can incrementally move to standard configurations as older hardware is retired or repurposed.

Businesses with a wide range of application needs will find that they must maintain a fairly broad service catalog of images. This is not necessarily a problem if you can at least standardize on some of the key components in the application stack:

- OS
- Application servers
- Network services
- Transaction processing servers

For example, a business may have one or two versions of Windows Server and two or three versions of Linux OSs for different tasks. Building on these, the IT department can offer .Net Framework applications on the Windows servers while providing Java applications on the Linux servers. Applications that require directory services may be able to run a standard LDAP server. Similarly, the private cloud may offer a preconfigured transaction processing server that is generalized enough to meet most user requirements.

Standardization does not require that you fit everyone's needs into a single set of application images. There will be exceptions and those should be accommodated. The purpose of standardization is to reduce management overhead, not constrain business requirements.

Formalize Discovery and Monitoring Procedures

Knowing what you are managing and understanding how it is used is essential to efficiently delivering cloud services. Businesses that deploy private clouds will likely have some resources dedicated to the cloud and others used outside the cloud. An ongoing objective will be to allocate servers, storage, and network services optimally between the private cloud and other uses. If servers are underutilized outside the cloud while at the same time job queues are growing in the cloud because there is not sufficient CPU capacity, then you should consider reallocating resources. To collect this kind of data, you need to have discovery and monitoring procedures in place.

Discovery and monitoring software can meet at least three management needs. Automated discovery helps to maintain an accurate inventory of resources. This is especially important when hardware is frequently moved and repurposed; manual recordkeeping can easily fall behind. A second objective is to use the network and server monitoring to collect data on utilization and availability. Cloud administrators can use this data to identify bottlenecks, potential hardware failures, and other areas that need their attention. Both discovery and monitoring data is useful for establishing operational baselines and planning for growth. This data can help justify the need for new hardware as well as changes to policies; for example, if job queues are filled during the day and relatively empty at night, a change in pricing policy could be used to spread demand more evenly throughout the day.

Summary

As private clouds evolve, so too will their management. Fortunately, you can leverage many IT best practices, particularly with respect to establishing policies and procedures, standardizing hardware and applications, and formalizing asset discovery and monitoring procedures.