

Service Management for the Private Cloud

How to Apply the Key Principles

Version 1.0

Published: October 2011

For the latest information, please see [www.microsoft.com/mof](http://www.microsoft.com/mof)

Copyright © 2011 Microsoft Corporation. All rights reserved. Complying with the applicable copyright laws is your responsibility. By using or providing feedback on this documentation, you agree to the license agreement below.

If you are using this documentation solely for non-commercial purposes internally within YOUR company or organization, then this documentation is licensed to you under the Creative Commons Attribution-NonCommercial License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/2.5/ or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

This documentation is provided to you for informational purposes only, and is provided to you entirely "AS IS". Your use of the documentation cannot be understood as substituting for customized service and information that might be developed by Microsoft Corporation for a particular user based upon that user’s particular environment. To the extent permitted by law, MICROSOFT MAKES NO WARRANTY OF ANY KIND, DISCLAIMS ALL EXPRESS, IMPLIED AND STATUTORY WARRANTIES, AND ASSUMES NO LIABILITY TO YOU FOR ANY DAMAGES OF ANY TYPE IN CONNECTION WITH THESE MATERIALS OR ANY INTELLECTUAL PROPERTY IN THEM.

Microsoft may have patents, patent applications, trademarks, or other intellectual property rights covering subject matter within this documentation. Except as provided in a separate agreement from Microsoft, your use of this document does not give you any license to these patents, trademarks or other intellectual property.

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places and events depicted herein are fictitious.

Microsoft, Hyper-V, Windows Azure, and Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries and regions.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

You have no obligation to give Microsoft any suggestions, comments or other feedback ("Feedback") relating to the documentation. However, if you do provide any Feedback to Microsoft then you provide to Microsoft, without charge, the right to use, share and commercialize your Feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies and services to use or interface with any specific parts of a Microsoft software or service that includes the Feedback. You will not give Feedback that is subject to a license that requires Microsoft to license its software or documentation to third parties because we include your Feedback in them.

Contents

[Service Management and the Private Cloud 1](#_Toc307404392)

[Audience 1](#_Toc307404393)

[What Is a Cloud? 1](#_Toc307404394)

[Cloud Service Models 2](#_Toc307404395)

[What Is a Private Cloud? 2](#_Toc307404396)

[Key Principles that Drive New Thinking 3](#_Toc307404397)

[Applying IT Service Management to the Private Cloud 4](#_Toc307404398)

[Managing the Private Cloud 4](#_Toc307404399)

[Governance, Risk, and Compliance 5](#_Toc307404400)

[Change and Configuration Management 6](#_Toc307404401)

[Team 7](#_Toc307404402)

[Planning for the Private Cloud 7](#_Toc307404403)

[Key Planning Tasks for the Private Cloud 8](#_Toc307404404)

[Delivering to the Private Cloud 12](#_Toc307404405)

[Key Delivery Tasks for the Private Cloud 13](#_Toc307404406)

[Operating in the Private Cloud 15](#_Toc307404407)

[Key Operating Tasks in the Private Cloud 16](#_Toc307404408)

[What Does Microsoft Offer? 18](#_Toc307404409)

[Summary 19](#_Toc307404410)

[Version History 20](#_Toc307404411)

[Acknowledgments 21](#_Toc307404412)

[Feedback 21](#_Toc307404413)

# Service Management and the Private Cloud

The promise of public cloud computing is compelling: move to the cloud and you get all the benefits of information technology (IT) with fewer headaches. Get the computing resources you need for less money while someone else worries about how to provide them.

The promise of the private cloud is also compelling, but a little less clear. This is because the private cloud, in some cases, may be only a stop along the road to public cloud computing, and not the destination itself. Unless it is a hosted solution, private cloud computing might not offer the biggest advertised benefits of the public cloud: own less and do more. With an on-premises private cloud solution, you still have to own the capital expenditure part of the equation.

Even more importantly, getting to the private cloud is not simply deciding to go there. It requires discipline in the form of effective service management; however, there are some real benefits: elasticity, scalability, automation, and reduced time-to-market, which combine to make it a worthwhile destination.

This paper addresses how to apply IT service management principles addressed by the Microsoft Operations Framework (MOF) to get the most out of a private cloud environment to best realize those benefits.

## Audience

This guide is intended for IT managers, IT pros, and others interested in how to effectively operate and manage a private cloud environment.

## What Is a Cloud?

The National Institute of Standards and Technology (NIST) defines *cloud computing* as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (for example, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

The key to the NIST cloud model is that it promotes availability and features five essential characteristics:

* **On-demand self-service.** Consumers can provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction.
* **Broad network access.** Capabilities are available over the network through a variety of platforms, such as mobile phones, laptops, and PDAs.
* **Resource pooling.** Computing resources are pooled to serve multiple consumers with different physical and virtual resources assigned and reassigned according to consumer demand.
* **Rapid elasticity.** Capabilities can be rapidly provisioned, in some cases automatically, to quickly scale out, and rapidly released to quickly scale in.
* **Measured service.** Resource usage can be monitored, controlled, and reported, so the provider and consumer of the service understand how much is used.

There are three variations on cloud computing—public, hybrid, and private. A *public cloud* offers resources that are shared over the Internet and used as needed. Typical public cloud offerings are applications and services, available on pay-per-use models. A *hybrid cloud* typically refers to a blend of public and private clouds. A *private cloud* is a variation of cloud computing using resources that are dedicated to your organization.

### Cloud Service Models

The NIST cloud definition also includes these three service models:

* **Cloud software as a service (SaaS).** Also called on-demand software. SaaS allows consumers to use software that is hosted centrally—typically on the Internet. Consumers do not have to manage any of the underlying infrastructure. Microsoft® Office 365 is an example of SaaS.
* **Cloud platform as a service (PaaS).** PaaS is a way to rent hardware, operating systems, storage, and network capacity over the Internet. The consumer is able to rent virtualized servers and associated services for running existing applications or developing and testing new ones. Microsoft’s Windows Azure™ offering is an example of PaaS. A private cloud provisioned inside a public cloud is another example of PaaS.
* **Cloud infrastructure as a service (IaaS).** Also known as on-demand data centers. IaaS provides compute power, memory, and storage, typically priced per hour and based on resource consumption. You pay only for what you use, and the service provides all the capacity you need, but you are responsible for monitoring, managing, and patching your on-demand infrastructure.

Figure 1 illustrates the differences between IaaS, PaaS, and SaaS relative to what the customer manages versus what others manage.

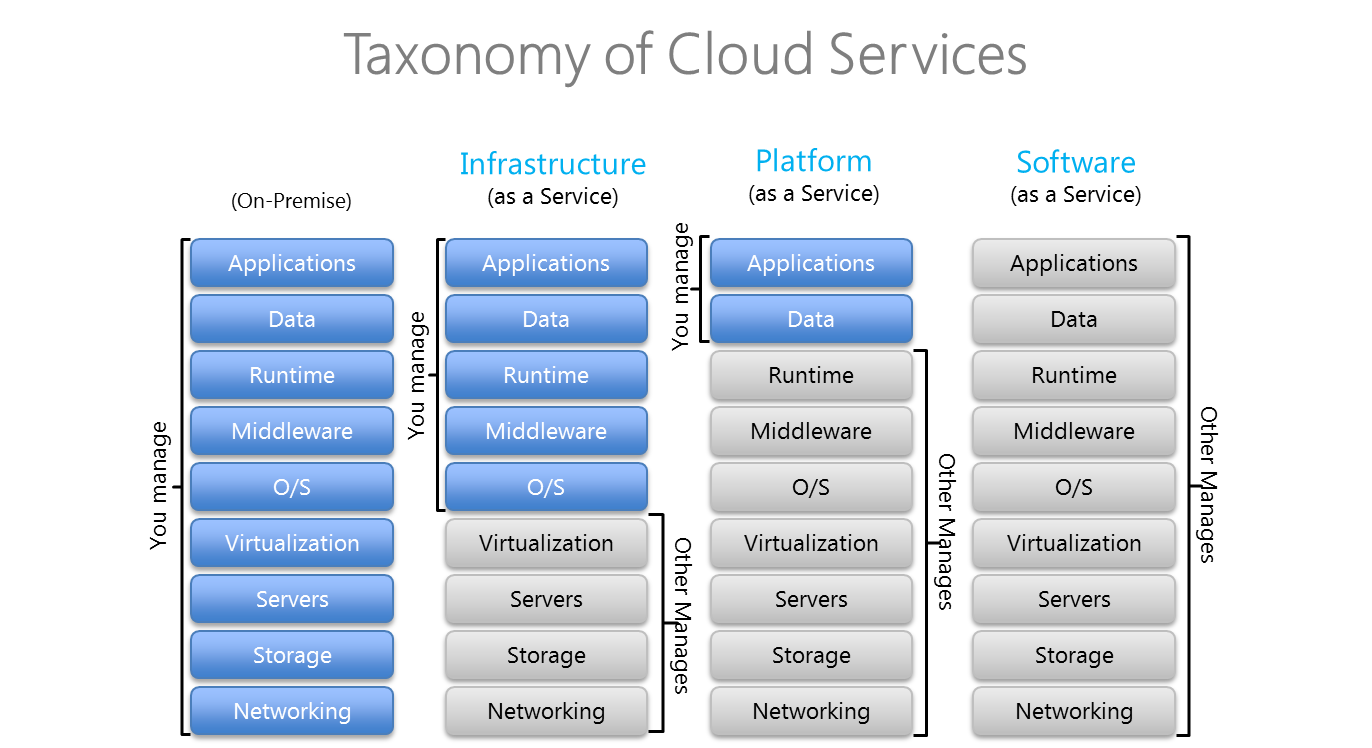


Figure 1. Cloud services taxonomy

## What Is a Private Cloud?

A *private cloud* is a variation of cloud computing using resources that are dedicated to your organization, whether they exist on-premises or off-premises. With a private cloud, you get many of the benefits of public cloud computing—including self-service, scalability, and elasticity—with the additional control and customization available from dedicated resources.

## Key Principles that Drive New Thinking

In addition to the NIST essential characteristics, several key principles drive new thinking around the private cloud. These are highlighted throughout the paper and should be part of the conversation for any organization venturing to a private cloud.

* **Create a perception of infinite capacity.** As far as the consumer is concerned, there is no apparent limit to the amount of service they can use; however, this needs to be balanced with the business desire to encourage more cost-effective use of IT resources. That can be done by clearly tying consumption costs to levels of service, which sends the message to the consumer that you have to pay for what you get and so you should not ask for more than you need.
* **Create a perception of continuous availability.** The consumer does not notice any interruption to service, even if failures occur within the cloud environment.
* **Provide predictability.** The private cloud should remove as much variation from the environment as possible to increase predictability.
* **Offer a service provider’s approach to delivering infrastructure.** IT organizations should adopt a service provider model; the provider delivers infrastructure on demand.
* **Develop a resiliency-over-redundancy mindset.** The provider’s focus should be on maintaining service availability through the resiliency of the service application architecture, rather than redundancy. Resiliency focuses on quickly repairing services so the user does not notice a service is unavailable. (See the “Incident and Problem Management” section for further discussion.) In a real sense, resiliency is also the tolerance for error, of being able to sustain a service’s performance in the case of an infrastructure error (such as disk failure).
* **Minimize human involvement.** Automation is essential to achieving resiliency, error-proofing operations, and containing costs.
* **Optimize resource usage.** Providers should optimize resource use to get the maximum use with the least excess capacity.
* **Encourage desired resource-consumption behavior.** Use cost of services to discourage over-use of resources or use of the wrong resources, and to encourage use of the preferred resources.

# Applying IT Service Management to the Private Cloud

Most of the principles of IT service management (ITSM) are relevant in the context of the private cloud with some differences in how they apply.

If you want agility, as one example and a key component of cloud technology, users should have the ability to rapidly and inexpensively reprovision technological infrastructure resources. If you use a process-heavy approach to change management, this will be difficult, but agility is more likely if you adopt a standard changes approach to provisioning.

Creating a private cloud with automated virtual machine provisioning means IT can define standard profiles that can be automatically provisioned (small, medium, and large). These profiles can be ordered from a user portal and be implemented as standard or preapproved changes. This eliminates a complex process with several potential human points of failure and replaces it with an automated process with very little human intervention.

Other ITSM examples:

* Service catalogs play a big role in a cloud environment because of the importance of letting users know what is available, at what costs, and at what service levels.
* Service level management is more important than ever because of the private cloud’s emphasis on self-service, and the interdependency of its components.
* Problem management is important because of its emphasis on root cause analysis and proactive avoidance of incidents.

Two of the better known ITSM frameworks are the Information Technology Infrastructure Library (ITIL) and the Microsoft Operations Framework (MOF). Both offer a structured approach to effectively managing IT services. This paper uses the structure of MOF, which is Microsoft’s service management framework, to explain the role of ITSM in the private cloud.

MOF’s guidance comes in the form of this IT service management lifecycle:

* Manage. Provide operating principles and best practices to ensure that IT delivers expected business value at an acceptable level of risk.
* Plan. Plan and optimize an IT service strategy that supports business goals and objectives.
* Deliver. Ensure that IT services are developed effectively, are deployed successfully, and are ready for operations.
* Operate. Ensure that IT services are operated, maintained, and supported in a way that meets business needs and expectations.

More information about MOF can be found at <http://technet.microsoft.com/en-us/solutionaccelerators/dd320379.aspx>.

## Managing the Private Cloud

There are three service management functions (SMFs) representing activities that occur through the entire IT service management lifecycle. These SMFs are in the MOF Manage Layer:

* Governance, Risk, and Compliance (GRC)
* Change and Configuration Management
* Team

**Governance, Risk, and Compliance** focuses on these activities or outcomes:

* Define the regulations and standards to which IT must abide.
* Create policy to reflect regulations and standards.
* Identify and prioritize risks.
* Establish controls to mitigate risks.
* Monitor controls and report.
* Determine laws and regulations to which IT must comply.
* Evaluate and maintain compliance.
* Provide reporting.

**Change and Configuration Management** focuses on these activities or outcomes:

* Baseline the IT cloud.
* Identify and classify the change request.
* Approve and/or deny the request, and communicate the approved change.
* Implement and validate the change.
* Update the baseline to reflect the change.

**Team** focuses on these activities or outcomes:

* Identify who is responsible for each task, activity, or area.
* Ensure that every task, activity, or business area has an owner.
* Confirm that adequate skills exist for each task, or provide them.

### Governance, Risk, and Compliance

*Governance, Risk, and Compliance* clarifies who has the authority to make decisions, who is accountable for them, and how the outcome of decisions will be measured. In addition, GRC identifies risks to success and how to manage those risks to avoid negative outcomes. It also ensures that regulations, policies, and procedures that senior management decides on are followed.

In a private cloud, mandated compliance to government regulations should be considered when planning IT services, deploying or delivering those IT services, and in the daily support and operations of those IT services. One example is the United States government’s Health Insurance Portability and Accountability Act (HIPAA) guidance that mandates the protection of patient data. This protection does not stop at any stage in an IT service lifecycle, but must be considered in each and every IT activity that might come in contact with patient data.

Organizations often choose the private cloud option because of GRC concerns. Public cloud benefits are offset by security and compliance concerns about storing or managing data outside of the normal boundaries of an IT organization.

The private cloud has most of the characteristics of a public cloud (on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service), but within the safer and better understood policy and process boundaries of an existing IT organization.

But IT organizations are still subject to the rules and regulations of governments, industries, and their own business organizations. Some characteristics of the cloud may present challenges that will need to be mitigated—for example, resource pooling may not be allowed between different business units because of legal constraints about data co-mingling on devices. Risks and compliance issues need to be identified and managed across all layers of the private cloud.

### Change and Configuration Management

*Change management* is about enabling healthy and necessary change, while minimizing any disruption to the production environment. Change management is usually thought of in terms of changing IT systems, but changes to IT strategy or to major IT initiatives can be just as disruptive to IT service delivery so they should also be managed in a controlled and predictable manner.

In a private cloud, where the perception of continuous availability is important, driving predictability and minimizing human involvement are core principles for achieving stable services. Driving predictability means defining and deploying processes and systems that will provision, manage, and support the new virtualized environment effectively. Minimizing human involvement means automating as many of those processes as possible and identifying and automating standard changes that are unique to the virtualization environment.

Many virtualization technologies and their management systems allow for dynamically performing operational tasks such as automatically detecting and responding to failure conditions in the environment. They often allow for quick migrations to other virtualized systems; however, all of these actions are changes that come with risk. Each change type must be categorized based on risk and processed through an appropriate approval process, the same as in a traditional data center.

A Change Advisory Board (CAB) will need to evaluate each change type and determine if a given change can be categorized as a standard change. Standard changes are changes that have been preapproved by the CAB and can be fully automated because no further approval is necessary. For more information on CABs, see <http://technet.microsoft.com/en-us/library/cc543211.aspx>.

Standard change candidates often include patching, virtual machine creation, starting and stopping virtual machines, virtual machine live migration, and scaling out workload for just-in-time capacity as well as fault conditions.

More information about standard changes can be found in the *Using Standard Changes to Improve Provisioning* guide, which can be downloaded from the Microsoft Operations Framework website at <http://technet.microsoft.com/en-us/solutionaccelerators/dd320379.aspx>.

*Configuration management* provides information needed by many IT processes, beyond just change. If you have a clear understanding of the state of a service’s environment, you can avoid such issues as planning duplication, troubleshooting complications, and failed releases.

Configuration management is about maintaining a known, baselined production environment state. Other key characteristics include providing relationships among components, component information versioning, protecting data from unwanted access (while assuring appropriate access), and ensuring that all data is subject to change management.

A predictable and stable private cloud requires an environment that is as standardized as possible. That will result in lower support and management costs, increased automation opportunity, and more predictability.

Although initial decisions regarding standardization should come from planning, continual assessment will need to be done as lifecycles rotate. Along the way, normal support issues will require decisions about replacing existing components with non-standardized components. Non-standardized workloads add complexity and support and management costs, which can interfere with targeted private cloud benefits.

A key to avoiding those issues is a configuration management database (CMDB). Identifying the necessary components and services that need to be controlled is a first step toward creating a CMDB.

A service map of the private cloud ecosystem will help with that identification process. A service map is a graphical display of a service that illustrates the various components on which successful delivery of that service relies. By visually presenting a service-centered view of a private cloud, a service map helps interested parties better understand what it will take to deliver the service. Additionally, the service map will help identify root causes if the services provided by the private cloud should fail.

A service map can help with many ITSM areas, such as enabling service level management by identifying where operating level agreements (agreements between IT groups) are needed, and knowing what to target for availability and service continuity.

### Team

*Team* management ensures that the right work gets done by making sure that someone is accountable for getting it done.

Effective team management helps those who plan, deliver, and operate private cloud services to:

* Understand the business and operational needs for the cloud service and make sure those needs are met.
* Effectively and efficiently deploy the cloud service with as little disruption to the business as the service levels specify.
* Operate a cloud service on which the business trusts and relies.

Team management principles also help ensure that the right people with the right skills are doing the right things toward effective operation of a cloud service.

## Planning for the Private Cloud

Planning provides an opportunity to ensure your private cloud services are reliable, compliant, and cost-effective, and that they continuously adapt to the ever-changing needs of the business. Focus your planning efforts on:

* Increasing automation to minimize human involvement to reduce costs and errors.
* Providing self-service.
* Offering flexible capacity to optimize resource usage.
* Offering the perception of continuous availability.
* Driving for standardization and predictability.
* Providing a service-oriented approach.
* Offering resiliency and reliability.
* Providing consumption-based pricing.

There are four SMFs in the MOF Plan Phase. They are:

* Business/IT Alignment
* Reliability
* Policy
* Financial Management

**Business/IT Alignment** focuses on these activities and outcomes:

* Agree on an IT service strategy.
* Identify and map services that support that strategy.
* Identify demand for services and decide how to manage requests.
* Create and publish an IT services portfolio.
* Manage services so they meet business needs.

**Reliability** focuses on these activities and outcomes:

* Plan for reliability by defining service requirements and analyzing how to meet them.
* Develop and implement such plans as availability, capacity, data security, disaster recovery, and monitoring.
* Monitor service reliability, analyze and report on reliability trends, and review reliability.

**Policy** focuses on these activities and outcomes:

* Decide where policies are needed.
* Create policies.
* Validate that they are the right policies.
* Publish policies.
* Enforce policies.
* Review and keep policies up to date.

**Financial Management** focuses on these activities and outcomes:

* Decide what IT services are needed and budget for them.
* Manage IT finances.
* Track finances and report the actual costs.

### Key Planning Tasks for the Private Cloud

The key areas for private cloud planning are:

* Business/IT alignment
* Service level management
* Demand management
* Financial management
* Service catalog management
* Reliability, which includes these components:
* Capacity management
* Availability management
* IT service continuity management
* Security management

#### Business/IT Alignment

Getting clear alignment between what the business needs and what IT can provide is crucial to succeeding with the private cloud. It helps define two important things: key business requirements and service requirements.

The key business drivers for a private cloud are:

* Cost. If you want cost reduction and/or containment, then focus on clearly identifying and marketing what the private cloud can do for cost management. Because one of the highest costs in operating a data center is power (the other is human resources), focus on the cloud’s ability to virtualize servers. Capitalize on its ability to reduce the cost and complexity of provisioning and managing hardware redundancy layers. Also, focus on how the automation of tasks will present opportunities to reduce costs both in manpower to deploy and reduction in human error, and therefore support cycles in the IT environment.
* Reliability. Getting reliability may require that you plan how to incorporate dividing a service across multiple data centers for disaster recovery purposes. Also, it is critical to tightly define workloads and ensure their distribution across multiple virtual machines.
* Agility. Getting agile might require the flexibility of multiple-sized virtual machine templates and capacity buffers—such as excess capacity built into the system reflected in additional physical hosts, excess bandwidth, storage, and so on—to accommodate business changes. There are multiple definitions of agility, so it is important to understand business expectations and align with the capabilities of a private cloud.

The service requirements for a private cloud are:

* Self-service customer portal with controlled access.
* Different-sized, predetermined virtual machine templates to accommodate the different business unit needs.
* Cost model determination and use of chargeback methods.
* Determination of future cloud strategies and whether the virtual machines should be designed to one day migrate to a public cloud.
* Workload compatibility with the private cloud services (what applications are targets for the cloud—which cannot and/or should not be considered?).

Real success with getting and keeping the business and IT aligned on private cloud planning requires program-based continuity and sustainment. That program-based approach should include agreed-on checkpoints between IT and the business to make decisions and encourage alignment.

#### Service Level Management

*Service level management* (SLM) is traditionally concerned with establishing and codifying expectations between the customers and IT. That need does not go away, but there are some changes in how it gets applied in a private cloud environment.

One key change is in service level agreements (SLAs), which document expectations between customers and IT. SLAs in a private cloud environment need to address such things as:

* On-demand services with self-service attributes. What new interfaces and processes will be defined?
* Automation innovations. For example, moving from a six-month server delivery time to hourly virtual machine server delivery.
* Targets for virtual machine cost, quality, and agility by service class, and the metrics for measuring their successful achievement.
* Determining how an SLA is affected by the resource pooling required to serve multiple business units. Are there compliance or security requirements that present obstacles or boundaries?
* Chargeback models based on consumption, and possible penalties for SLA breaches.

Operating level agreements (OLAs) are expectations, or ground rules, between IT groups in support of SLAs. Underpinning contracts (UC) are similar contractual agreements between IT and third-party vendors. Both will likely change in a private cloud environment. For example, a number of different IT groups and vendors may be involved in provisioning and deploying a virtual server. Each of these groups will bring an area of specialty that enables a private cloud service, such as storage, infrastructure, or Internet service provider (ISP). Each will be a possible point-of-failure in the chain to provision a virtual server, and therefore a threat to meeting SLAs for timely server delivery. OLAs and UCs will need to specifically define expectations and timeframes to support SLA boundaries.

#### Demand Management

*Demand management* involves understanding and influencing the customer demand for services. The IT service provider will need to scale capacity (up or down) to meet these demands. The private cloud creates the perception of infinite capacity and continuous availability, but is still subject to real boundaries. To deliver those perceptions requires providing a resilient, predictable environment, and managing capacity in a predictable way. You can use such factors as cost, quality, and agility to influence consumer demand for private cloud services. Demand management is critical to understanding and balancing expectations and actions needed to ensure a successful private cloud ecosystem.

Demand management activities should focus on determining the rhythm of the business. For example, you need to account for seasonal or cyclical capacity requirements, such as a holiday shopping season for an online retailer. What are the business growth expectations for the next quarter, year, and three years? What is the likelihood that there will be an accelerated growth requirement (caused by new markets, mergers, acquisitions, and so on) in that timeframe?

#### Financial Management

Because *financial management* focuses on managing the service budgeting, accounting, and charging requirements, it raises several private cloud-based decisions and considerations:

* Does the business want to benchmark the cost of a private cloud versus a public cloud or a traditional data center? If the business wants to compare the cost of building and delivering a private cloud capability internally with that of the market, it will need to clearly measure the costs for building and operating the service.
* Does the business want to drive consumer behavior? With cost transparency—metering and reporting—consumers can better understand the cost of the services they are consuming. Consumption-based pricing will drive customers to consume only what they really need. And leveraging price-based service classes will allow customers to choose more or less expensive classes of service based on feature differences—for example, more redundancy, perhaps across data centers. This may drive service owners to build or buy applications that do not require hardware redundancy and qualify for the least expensive class of service, such as using stateless applications when they are available.

#### Service Catalog Management

*Service catalog management* is about defining and maintaining a catalog of services offered to consumers. As a single view into all operational services, it is where you should list the private cloud and all related services. It is an expectation-setting tool for both the business and IT. By clearly defining expectations, you can better plan and implement necessary tasks and responsibilities to ensure SLAs can be met.

Private cloud services in the catalog could include attributes such as:

* Availability targets
* Service contacts
* Ordering information
* Cost and chargeback
* Performance targets
* Quality targets
* Backup/restore activities
* Service continuity plans and triggers in the event of a disaster
* Administration (who has rights to carry out certain actions)
* Operational level agreements (lead time to complete actions by certain teams)

If you define different classes of service, these attributes will vary between the classes.

#### Reliability

A *reliable* service or system is dependable, requires minimal downtime for maintenance, performs without interruption, and allows users to quickly access the resources they need. These characteristics are not only true for business-as-usual conditions, they must also apply during times of business change and growth and during unexpected events.

Reliability ensures that service capacity, service availability, service continuity, data integrity, and confidentiality are aligned to the business needs in a cost-effective manner. The following are requirements for reliability.

##### Capacity Management

*Capacity management* defines what is needed to create the perception of infinite capacity in a private cloud. It is strongly driven by the output of demand management.

You need to manage capacity to meet peak demands while controlling the cost of under-utilization. Developing and executing a mature capacity plan will ensure that the perception of infinite capacity is successfully maintained.

You also need to understand reserve capacity requirements, current usage, projected growth (from demand management), bursting requirements, and the length of the procurement process for additional core capacity and factor those into the decision about how much buffer capacity you need to maintain.

##### Availability Management

*Availability management* defines what is needed to achieve the perception of continuous availability. One major shift needed in a private cloud environment is around availability management. With the ability to use virtualization technologies, multiple logical servers can exist on a single physical server. In addition, current technology allows for workloads to be quickly transferred between virtual servers.

Traditional data center environments typically dictated that high availability needed expensive hardware redundancy. But the new virtualized world, with the ability to seamlessly move workloads from one virtual server to the next, dictates a rethinking of availability management to a concept of resiliency.

The new resiliency comes from the fact that, if a service has a failure, restoration to another location is so timely that no one notices.

You should consider availability across the whole stack: infrastructure, platform, application, and data in the private cloud. You can reduce costs by designing an application to expect and handle failure. Stateless applications are extremely powerful in the private cloud world.

Resiliency, measured in mean-time-to-restore-service (MTRS), minimizes the need for hardware redundancy, measured in mean-time-between-failures (MTBF). Even if the number of outages increases, the duration of each outage is very low, maintaining a high availability experience for the user.

##### IT Service Continuity Management

*IT service continuity management* defines how risk will be managed in a disaster scenario and ensures that minimum, agreed-to service levels are maintained. IT service continuity plans, and the business continuity plan, should be driven by business needs. You should regularly test these plans and incorporate them into the business continuity plans.

With the strong reliance on virtualization in the private cloud, the most important aspect of continuity is ensuring that virtual machines are replicated and can be restarted in a recovery environment. Key enablers in the private cloud will be hardware redundancy and clustering (failover technologies), as well as the ability to monitor for warnings that suggest hardware may be about to fail—for example, servers, uninterruptible power supply, switches, and storage arrays.

A fault domain is a set of hardware components—computers, switches, and more—that share a single point of failure. Because hardware failure can trigger a disaster recovery plan—imagine if 4 of 6 servers in a fault domain failed—it would be necessary to define this as a trigger for the IT service continuity management (ITSCM) plan. Disasters are not strictly defined by the cause (flood, terrorist, and so on), but by the amount of risk that is needed to trigger activation of the plan.

##### Security Management

*Security management* ensures that data integrity, confidentiality, and availability are risk-managed and accommodated for all IT services. Private cloud services fall under the same umbrella of security need as any IT service.

Certain dynamics in a private cloud environment may need particular attention:

* Security implications of possible multi-tenanting of customer data.
* Policy or compliance issues with resource pooling.
* New access mechanisms used that introduce risk that needs to be mitigated (for example, web browser or mobile device access).
* Any portability decisions for future use of public clouds that might introduce security modifications to current designs or plans.

The levels of concern will differ based on private cloud services offered, but all have to be addressed and mitigated.

## Delivering to the Private Cloud

Delivering a private cloud solution very likely represents the biggest changes in how service management principles are applied. Although the overall goals of ensuring that IT services are envisioned, planned, built, stabilized, and deployed in line with business requirements and the customer’s specifications remain the same, the ways in which those goals are met are significantly different.

There are two key reasons for that:

* Any service has to be designed with resilience in mind.
* Cloud services have to be deployed with continuous availability.

The Deliver Phase of MOF has five SMFs. They are:

* Envision
* Project Planning
* Build
* Stabilize
* Deploy

**Envision** focuses on these activities and outcomes:

* Organize the project team.
* Create a high-level description of what is to be built as well as how and when it will get built.
* Get team and business stakeholder approval for high-level plans.

**Project Planning** focuses on these activities and outcomes:

* Decide on the tools that will be used to build.
* Document in detail what will be built.
* Create a detailed project plan.
* Create a detailed project schedule.
* Get team and business stakeholder approval for those plans.

**Build** focuses on these activities and outcomes:

* Prepare for development.
* Develop, document, and test the solution.
* Get ready to release the solution.
* Get agreement that the solution is ready to release.

**Stabilize** focuses on these activities and outcomes:

* Determine the solution is stable enough to be released.
* Pilot-test the solution.
* Get team and customer stakeholders to agree the solution is ready for release.

**Deploy** focuses on these activities and outcomes:

* Deploy core IT service solution components.
* Deploy sites.
* Stabilize deployment.
* Review the Deployment Complete Milestone.

### Key Delivery Tasks for the Private Cloud

Traditional ITSM delivery is about ensuring that a service is conceived, project-planned, built, tested, and deployed in line with business requirements and the customer’s specifications, including its operability and manageability.

Delivery for the private cloud shares those goals, but focuses specifically on:

* Release and deployment management
* Testing

#### Release and Deployment Management

Although change management is based on an approval mechanism, *release and deployment management* determines how those changes will be implemented.

In a private cloud environment, two distinct types of release and deployment management activities need to be managed (there are others, as with any IT service, but these are unique tasks that are driven by the virtualization setting):

* How to deploy new virtual machines.
* How to deploy new workloads inside the virtual environments (the tenant applications that sit on top of the servers).

Deploying virtual machines involves the successful coordination of different IT groups and possibly vendors. The different services and groups engaged to deploy a virtual machine may include:

* Directory services
* Database or storage services
* Networking services
* Server and/or virtualization teams
* Security management
* Management and monitoring services
* Web/portal services
* Provisioning services
* Backup/restore services
* Disaster recovery planning
* Change and configuration management
* Capacity and availability management
* Patching or software update teams

Traditionally these groups are divided into departments. Normal operational and support tasks, shifting responsibilities, and varying and/or rotating team members can often interfere with the virtual machine deployment process. Yet customer expectations will be for predictable and timely virtual machine deployments.

That makes it critical to look to service level management for its guidance on creating appropriate operating level agreements (agreements between IT groups) and underpinning contracts (contracts between IT and vendors). A service-oriented perspective and culture are required for these agreements to enable a successful private cloud.

In addition, IT organizations may want to create a center of excellence (CoE) team with designated members from each needed team to form a project-based focus team. Once processes and executions are appropriate and predictable, the function should evolve into a free-standing program that delivers a consistent, repeatable service to customers.

The same approach should be used for deploying workloads into the virtualized environment—only the tenants (or customers requesting the virtual machines) need to be included in the service loop. The other consideration with tenants is the determination of how much of the virtual machine environment should be their responsibility, and how much IT will assume. This will reflect some of the topics in the list above (who will patch the applications, perform backups, do monitoring, and so on). This is also covered in the Monitoring and Management section of the “Operating in the Private Cloud” section below.

#### Testing

The primary driver for comprehensive *testing* is protecting the production environment from the unplanned consequences of changes and releases. The private cloud environment relies heavily on virtualization technologies. These technologies have historically proven a valuable tool for replicating environments and enabling the testing of new services and systems. The use of hardware in multiple virtual environments significantly reduces the costly identical hardware requirement typically needed for production-mirroring test environments.

In addition, with the dynamics of resiliency (the ability to failover more quickly than business effects are noticed), it is possible to push testing scenarios closer to, or actually into, production. With carefully defined and managed rollback plans and execution triggers, organizations can consider streamlining testing with virtualization. Given the prevalence of virtualization in a private cloud solution, the essential building blocks of a robust testing environment are readily available.

Although testing of the original virtual machine provisioning fabric should be a requirement, you should also define and enable testing environments for tenant applications, separate from production environments.

## Operating in the Private Cloud

Operating a private cloud involves the daily management and support of all the components and services needed to deliver private cloud services. This means managing proactively and effectively, monitoring proactively and continuously, and restoring services to health when problems occur.

In traditional ITSM the focus is on:

* Managing services proactively and effectively.
* Monitoring the health of services proactively and continuously.
* Restoring services to health when things go wrong.

In a private cloud, those certainly remain important goals, but they have to be looked at in the context of increased automation for operations, and the implications of outsourcing. Monitoring and responding will likely be automated, so the emphasis will have shifted to the design of controls and responses. Customer service and problem management will play key roles.

In addition, compliance issues will be critical, because a prime motivation for moving to the private cloud instead of to the public cloud is related to compliance requirements and lack of experience/trust in the public cloud’s ability to manage compliance requirements.

The four SMFs in the Operate Phase of MOF are:

* Operations
* Service Monitoring and Control
* Customer Service
* Problem Management

**Operations** focuses on these activities and outcomes:

* Define, write, and maintain daily, weekly, monthly, and ad-hoc tasks.
* Execute daily, weekly, monthly, and ad-hoc tasks.
* Report on tasks.

**Service Monitoring and Control** focuses on these activities and outcomes:

* Define what needs to be monitored.
* Define what *healthy* means.
* Define notification triggers needed.
* Define who to notify.
* Define historical reporting needs.
* Add monitoring tasks to Operations SMF.

**Customer Service** focuses on these activities and outcomes:

* Receive and record requests from users or systems.
* Classify the request (information, issue, or request).
* Determine if the request is supported.
* Prioritize the request (impact and urgency).
* Resolve request directly or with escalation assistance.
* Record metrics/reports.

**Problem Management** focuses on these activities and outcomes:

* Document, classify, and prioritize the problem.
* Research the problem.
* Apply fix or workaround.
* Update processes if necessary to prevent future recurrence.

### Key Operating Tasks in the Private Cloud

More complex dynamics emerge from the increased automation provided by virtualization management technologies, and the management of both the underlying technology fabric that provides the virtual machines, as well as the platforms that sit on that layer.

Key operations tasks are:

* Incident and problem management
* Monitoring and management
* Tooling and automation

#### Incident and Problem Management

The goal of *incident management* is to resolve events that are impacting or could impact services as quickly as possible. The goal of *problem management* is to identify and resolve the root cause of incidents, as well as identify and prevent, or minimize the impact of, incidents that may occur—in other words, proactive problem management.

The private cloud environment requires another perspective shift to fully utilize its promised benefits. Traditionally, if a server fails, response is quick, and restoring the individual server immediately is the primary concern. But features in a private cloud can now be used for more options. Those features include:

* Resiliency. This was discussed in the “Availability Management” section of this paper. Virtualization technologies can be used to move failed workloads quickly between virtual servers.
* Resource pools. A collection of multiple physical servers, all running virtual machines.
* Fault domain. Physical servers (particularly virtualization hosts) in the resource pool organized into distinct groups that ensure that points of failure are not shared with any other group. This allows both clustering servers and spreading distributed applications across fault domains. It also allows for resource decay (see next bulleted item).
* Resource decay. Tolerating failure of a server (or of many) in any fault domain because of availability of a buffer of capacity in a fault domain.

The concept of resource decay is what changes the way a hardware failure can be handled. Rather than treat a failed server as an incident that requires immediate resolution, you can treat the failed server as a part of a maintenance or lifecycle schedule, or as an actionable failure only when a fault domain, or larger resource pool, reaches a certain predetermined threshold of decay.

In other words, if a server fails you no longer need to treat the failure as an incident that must be fixed immediately. Rather, it may be more efficient and cost effective to treat the failure as part of the decay of the resource pool.

Cost savings come from predictable procurement planning and reduced cost of replacement. Emergency replacement is generally more expensive in terms of resource cost and support agreements.

You need to balance cost, efficiency, and risk in determining the server buffer and decay threshold you are willing to accept.

Incident and problem management will also be assisted by the automation of responses to monitoring events—covered in the following “Monitoring and Management” section.

In addition, as with any new IT service, you need to establish specific incident and problem escalation matrices to ensure that issues are addressed in a timely and effective manner. Problem management will also need the appropriate skillsets to support troubleshooting the private cloud environment, which will likely include support to vendors to accommodate any new technologies—for example, virtualization.

#### Monitoring and Management

Monitoring and managing the private cloud environment is complicated because of the separation of the fabric (the underlying physical infrastructure) and virtual machines that sit on top.

For the fabric, IT organizations have to define what hardware is going to be monitored (for example, servers, switches, and storage), as well as what software can be used to monitor that hardware. They need to decide what hardware warnings/failures should trigger migrating virtual machine workloads to another server, and how much automation is triggered by these warnings/failures.

By dividing physical servers into upgrade domains, IT organizations can accommodate upgrades and patching of the fabric without disrupting service delivery. The concept simply demands that servers are grouped across the fault domains (grouping one server from each fault domain into Upgrade Domain #1, and then grouping another server from each fault domain into Upgrade Domain #2). All servers in an upgrade domain undergo maintenance simultaneously; this mitigates the risk in any fault domain (and the entire resource pool). Each upgrade domain is targeted in turn. This allows workloads to be migrated away from the upgrade domain during maintenance and migrated back after completion.

There is also the question of monitoring and managing the virtual machine workloads, and how much autonomy is given to the tenants who provisioned the virtual machines. In other words, will IT monitor and manage the workloads, or will the tenants monitor and manage them? The former will add management complexity and overhead, and necessitate gathering monitoring and management requirements—for example, when to patch. The latter would essentially outsource monitoring and managing the individual workloads to the tenants (or to whom they choose to outsource the work).

You will need to do ongoing management and further automation and tuning of the management tools for monitoring. Automation is key to the reduction of costs and to increasing the value of the service to the business.

Also, if IT will monitor and manage the virtual machines, should it use a different systems management infrastructure, or capitalize on the infrastructure used to manage the fabric?

#### Tooling and Automation

Automation is crucial to effectively operating a private cloud. Without automation deeply embedded across all layers of the infrastructure, dynamic processes will grind to a halt as soon as user intervention or other manual processing is required.

# What Does Microsoft Offer?

Using the infrastructure as a service model, the Microsoft solution for the private cloud, built on Windows Server® 2008 R2 SP1 Hyper-V® and Microsoft System Center, is a key part of Microsoft’s approach to cloud computing, enabling you to build out a dedicated cloud environment to transform the way you deliver IT services to the business.

There are three options:

* Build your own private cloud with help from the Microsoft Private Cloud Deployment Guides and Microsoft Private Cloud partners. More information is available at [www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-deployment.aspx](http://www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-deployment.aspx).
* Get a prevalidated private cloud configuration from Microsoft Private Cloud Fast Track OEM partners. Microsoft Private Cloud Fast Track partners have worked with Microsoft to combine hardware and software offerings based on a reference architecture for building private clouds. More information is available at [www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-fast-track.aspx](http://www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-fast-track.aspx).
* Find a service provider in the Microsoft Private Cloud Service Provider Program who can host a dedicated private cloud for you. More information is available at [www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-service-providers.aspx](http://www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-service-providers.aspx).

Microsoft Services has designed, built, and implemented a Private Cloud/IaaS solution using Windows Server, Hyper-V, and System Center. Microsoft Services also used the NIST private cloud definition, but added several more requirements:

* Resiliency over redundancy
* Homogenization and standardization
* Resource pooling
* Virtualization
* Fabric management
* Elasticity
* Partitioning of shared resources
* Cost transparency

A team within Microsoft gathered and defined these principles. The team profiled the Global Foundation Services (GFS) organization that runs Microsoft’s mega-datacenters; MSIT, which runs the internal Microsoft infrastructure and applications; and several large customers who agreed to be part of the research. With the stated definitions and requirements accepted, the team moved on to the architecture design phase. Here, Services further defined the requirements and created an architecture model to achieve them. For more information, see:

* [www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-fast-track.aspx](http://www.microsoft.com/en-us/server-cloud/private-cloud/hyperv-cloud-fast-track.aspx)
* <http://technet.microsoft.com/en-us/cloud/private-cloud>

# Summary

The private cloud represents an important first step on the way to realizing the benefits of cloud computing. With a private cloud, you get many of the benefits of public cloud computing—including self-service, scalability, and elasticity—with the additional control and customization available from dedicated resources.

Realizing those benefits requires managing the private cloud well, which means applying the principles of IT service management.

# Version History

|  |  |  |
| --- | --- | --- |
| **Version** | **Description** | **Date** |
| 1.0 | First release. | October 2011 |

# Acknowledgments

The Microsoft Operations Framework team acknowledges and thanks the people who produced *Service Management for the Private Cloud*. The following people were either directly responsible for or made a substantial contribution to the writing, development, and testing of this paper.

**Lead Writers**

* Jerry Dyer *–* Microsoft
* Shawn LaBelle *–* Microsoft

**Reviewers**

* Sean Christensen *–* Microsoft
* Daniel Gracia
* Karl Grunwald *–* Microsoft
* John Joyner *–* ClearPointe
* Mike Kaczmarek *–* Microsoft
* Michael Kirst-Neshva *– ANK Business Services UG*
* Don Lemmex *–* Microsoft
* Pål McCarthy – *Nexans Norway*
* Mike McIver *– MacroTrenz*
* Betsy Norton-Middaugh *–* Microsoft
* Joseph Sgandurra *–* Microsoft
* Wallace Simpson *–* Microsoft
* Edward Walton *– B2B Technologies LLC*
* Kathleen Wilson *–* Microsoft

**Editors**

* Jude Chosnyk *– GrandMasters*
* Laurie Dunham *– Xtreme Consulting Group*

## Feedback

Please direct questions and comments about this guide to [mofpm@microsoft.com](mailto:mofpm@microsoft.com?subject=MOF%20-%20Service%20Management%20for%20the%20Private%20Cloud).