CHAPTER 4

Operating a VMware vCloud

IN THIS CHAPTER

4.1 Overview

This chapter offers practical, operations-focused guidelines to help you implement a VMware® vCloud®. Based on the *vCloud Operations Framework,* the guidelines have the near-term goal of supporting Infrastructure as a Service (IaaS) within a comprehensive, service-focused, operational framework. The long-term goal for IT operations is full implementation of IT as a Service (ITaaS), so this chapter also discusses many considerations with ITaaS in mind. These guidelines should be useful to both service providers and enterprises.

The following vCloud chapters are designed to be used together throughout the lifecycle of a VMware vCloud computing implementation. In combination with a service definition, these chapters provide a comprehensive view of VMware vCloud computing:

- Chapter 3, "Architecting a VMware vCloud," provides design guidance, design considerations, and design patterns for *constructing* a vCloud environment from its constituent components.
- Chapter 4, "Operating a VMware vCloud," includes design guidance and considerations for *operating and maintaining* a vCloud environment. It covers the people, process, and technology involved in running a vCloud environment.
- Chapter 5, "Consuming a VMware vCloud," covers considerations for *consumers* who choose to leverage vCloud computing resources.

Additionally, Chapter 6, "Implementation Examples," provides modular examples that show how to use VMware component software to implement a vCloud. Chapter 7, "Workflow Examples," and Chapter 8, "Software Tools," also provide useful information for IT operations.

NOTE

Detailed implementation procedures for installing a vCloud are available in the VMware vCloud product documentation (www.vmware.com/support/pubs/vcd_pubs.html).

4.1.1 Audience

This chapter is intended for IT personnel who are involved in the IT business, service, operations, and infrastructure governance, along with operational control, for one or more instances of vCloud delivering cloud services. The reader is assumed to be familiar with IT service management principles and VMware vSphere® and vCloud concepts.

4.1.2 Scope

This chapter focuses on operating a vCloud from the perspectives of organizational structure, service management, operations management, and infrastructure management.

4.2 Cloud Computing

Cloud computing leverages the efficient pooling of on-demand, self-managed virtual infrastructures, which are consumed as services. Figure 4.1 illustrates key cloud computing principles and service layers.

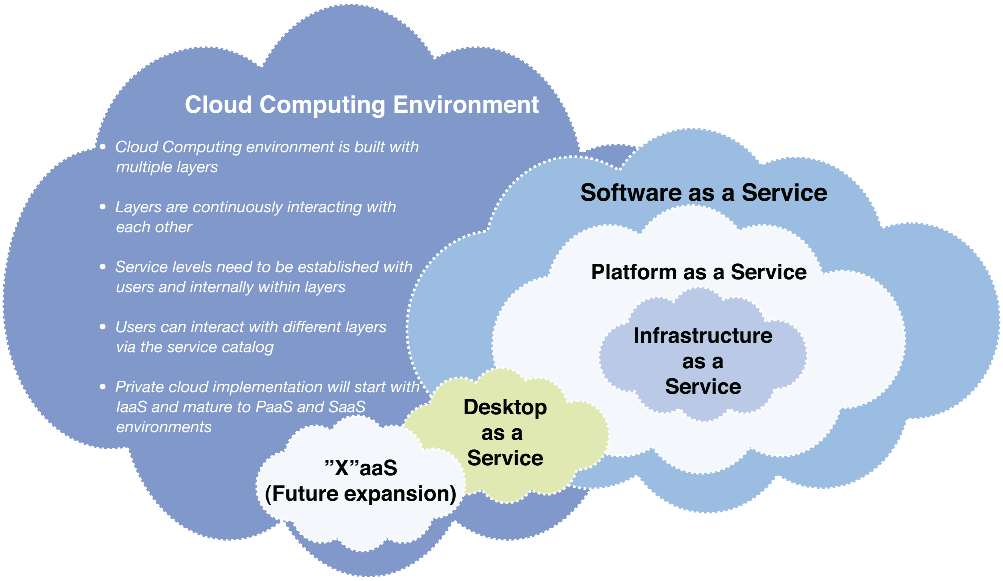

Cloud Computing Environment
• Cloud Computing environment is built with multiple layers
• Layers are continuously interacting with each other
• Service levels need to be established with users and internally within layers
• Users can interact with different layers via the service catalog
• Private cloud implementation will start with IaaS and mature to PaaS and SaaS environments
Software as a Service
Platform as a Service
Infrastructure as a Service
Desktop as a Service
"X"aaS (Future expansion)

FIGURE 4.1 Cloud computing layers

The National Institute of Standards and Technology (NIST) specifies three service layers in a cloud. VMware defines the following service layers:

- **Software as a Service (SaaS):** Business-focused services presented directly to users in a service catalog
- **Platform as a Service (PaaS):** Technology-focused services for application development and deployment presented directly to application developers through a service catalog
- **Infrastructure as a Service (IaaS):** Infrastructure containers for better agility, automation, delivery of components, and related purposes

Additional service layers are expected to become available as other services, such as Desktop as a Service, are developed.

Companies adopt cloud computing to improve quality of service, business agility, and operating cost efficiency.

- **Quality of service:** Standardized and automated service offerings, with associated availability levels and service management, help promote quality of service. Customers can provision a reliable vCloud service with predictable service levels to get the service they need, as they need it, within expected timeframes. To provide standardized, repeatable service, IT must introduce operational efficiencies and control the underlying infrastructure and applications.
- **Business agility:** A proactive, service-driven model helps IT provide and manage services, which are added to a service catalog to facilitate end-user self-service. IT retains control of the environment (for example, by protecting against oversubscription). Providing the reliable, dynamic services expected from a vCloud requires automation of complex, time-consuming, error-prone tasks. This model reduces IT lag, improving business agility and increasing speed to market.
- **Increased cost efficiency:** The key to increased cost efficiency is reduction of operational expenses. The current operational cost and burden of managing IT—approximately 70% operational expenses (OpEx) and 30% capital expenses (CapEx)—must change, especially as IT becomes more service driven. IT operational processes for vCloud computing must be enhanced by automation and the use of tools for management, compliance, and process governance. IT organizational structures must be optimized to support vCloud operations and management.

4.2.1 vCloud Operations Framework

An IT organization's adoption of IT as a Service (ITaaS) in a vCloud is evolutionary: The people, processes, and tools continue to evolve over time. The organizational structure and critical processes that support the adoption of ITaaS via vCloud computing are defined by the underlying VMware vCloud Operations Framework (VOF), as Figure 4.2 shows.

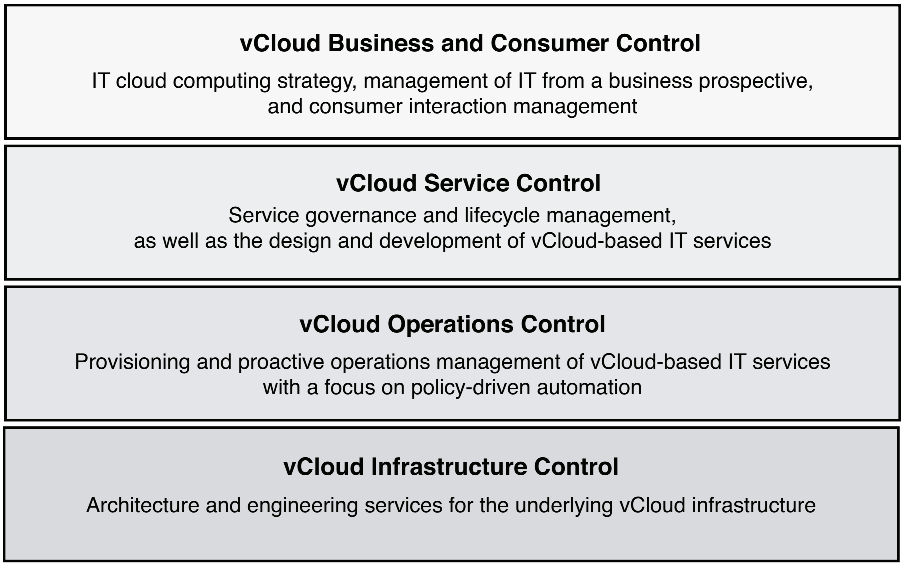

vCloud Business and Consumer Control
IT cloud computing strategy, management of IT from a business prospective, and consumer interaction management
vCloud Service Control
Service governance and lifecycle management, as well as the design and development of vCloud-based IT services
vCloud Operations Control
Provisioning and proactive operations management of vCloud-based IT services with a focus on policy-driven automation
vCloud Infrastructure Control
Architecture and engineering services for the underlying vCloud infrastructure

FIGURE 4.2 vCloud Operations Framework

The vCloud Operations Framework consists of the following layers:

- **vCloud Business and Consumer Control:** Addresses business-driven strategy in the context of consumer-driven requirements and demand for vCloud services, management of IT from a business perspective, and consumer interaction management.
- **vCloud Service Control:** Converts the consumer-driven requirements and demand, supported by business drivers, into vCloud service definitions. It also manages service development, creates service-level agreements (SLA), reports SLA compliance results back to the business and its consumers, and manages the lifecycle of services included in the service portfolio.
- **vCloud Operations Control:** Defines, deploys, and executes vCloud operations-related processes and supporting tools, and proactively manages the operations and delivery of vCloud services, with an emphasis on policy-driven automation.
- **vCloud Infrastructure Control:** Architects and deploys the underlying vCloud infrastructure on which the services are offered, provisioned, and run.

As Figure 4.3 shows, all these layers are required for support of IaaS, PaaS and SaaS services.

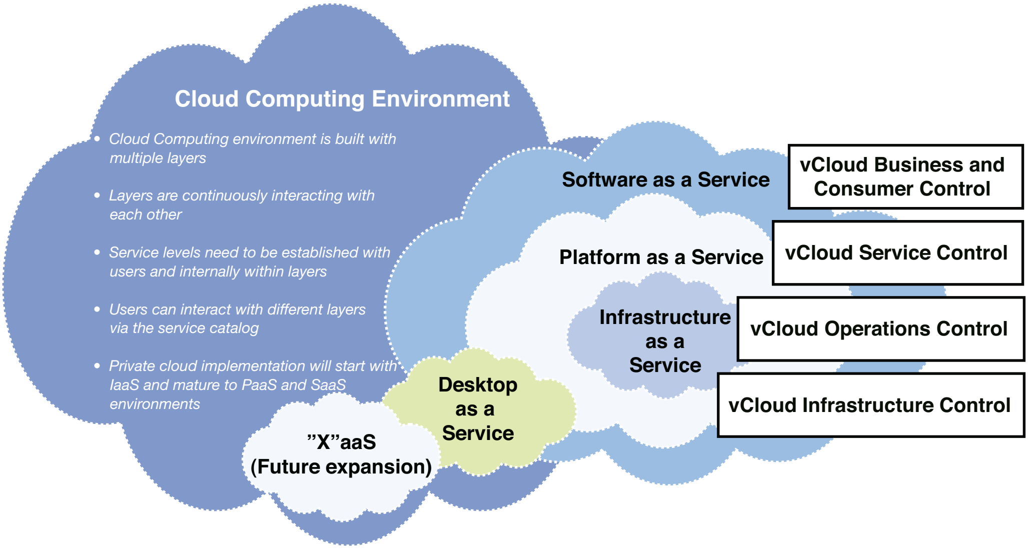

Cloud Computing Environment
Cloud Computing environment is built with multiple layers
Layers are continuously interacting with each other
Service levels need to be established with users and internally within layers
Users can interact with different layers via the service catalog
Private cloud implementation will start with IaaS and mature to PaaS and SaaS environments
Software as a Service
Platform as a Service
Infrastructure as a Service
Desktop as a Service
"X"aaS (Future expansion)
vCloud Business and Consumer Control
vCloud Service Control
vCloud Operations Control
vCloud Infrastructure Control

FIGURE 4.3 vCloud Operations Framework mapped to service layers

4.3 Process Maturity for vCloud Operations

Cloud computing is changing how resources are shared and consumed. Instead of relying on dedicated machines and workloads, vCloud relies on pooling and sharing of resources that are dynamic in nature. Within vCloud environments, new models are needed to effectively evaluate process maturity.

4.3.1 Traditional versus Maturity Models Specific to VMware

Traditional process maturity scales (ITIL, COBIT, CMM based) focus solely on optimizing processes in the physical world and are not capable of assessing the maturity of vCloud operations environments. Assessing process maturity in a vCloud environment requires a new scale.

Figure 4.4 represents the core differences between a traditional scale and a maturity scale based on VMware vCloud.

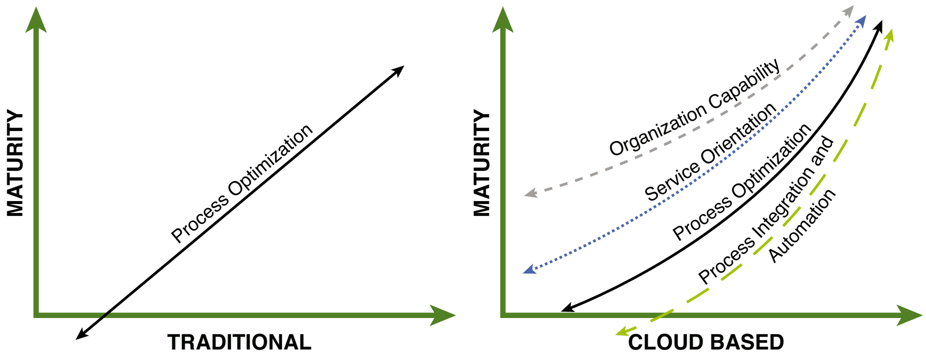


FIGURE 4.4 Core differences between traditional and vCloud maturity scales

In addition to process optimization, the scale based on vCloud focuses on process integration and automation, and on the organization's service orientation and capabilities, instead of on process optimization. The resulting maturity scale includes these elements:

- *Organization capability* is a measure of an organization's capability to use resource allocation, resource knowledge, and organizational setup to support vCloud operations.
- *Service orientation* is a measure of an organization's maturity and capability to align IT services with business user needs.
- *Process optimization* focuses on establishing and enforcing consistent, repeatable, and documented processes throughout an organization. On the maturity scale based on vCloud, process optimization is extended to the virtualization and vCloud computing stacks. In addition, process refinement is anticipated and planned for, to keep up with dynamic nature of the vCloud computing.
- *Process integration and automation* measures the evolution of traditional IT processes and their adoption for vCloud.

4.3.2 Process Maturity Scale Specific to VMware

Figure 4.5 shows the process maturity scale based on VMware vCloud. Organizations move from left to right on this scale over time, with the final goal of delivering IT as a Service (ITaaS) within a vCloud.

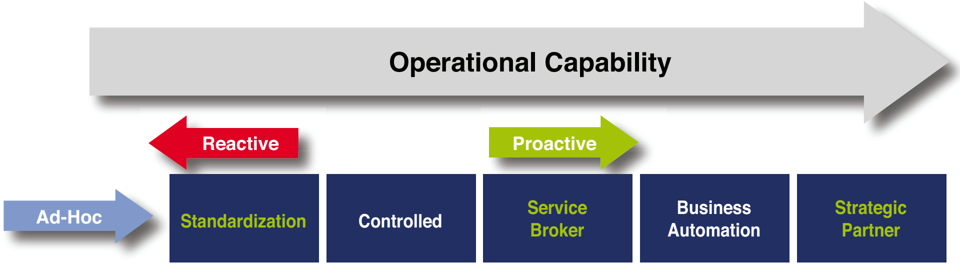


FIGURE 4.5 vCloud process maturity scale

Table 4.1 describes the process maturity scale states.

TABLE 4.1 Process Maturity Scale Legend

Maturity Level	Description
Standardization Level 1	Basic operational processes and tools are adapted for core virtualization, but not for vCloud computing. Processes objectives are defined, but activities are performed manually.
Defined/Controlled Level 2	Limited operational processes and tools are adapted for vCloud computing. Processes objectives are documented, and organization roles and responsibilities are defined. Limited, automated integration with existing IT processes (change, configuration, others) takes place.
Service Broker Level 3	Complete operational control is established over processes and tools, and a vCloud COE is in place. The organization is more service driven and offers services directly to business users through a service catalog. Operational processes are service focused and proactive. Service, design, and development procedures are clearly defined.
Business Automation Level 4	Automated process-management policies and operational controls are in place. Organization focus moves toward business agility—critical business services are offered through the vCloud with complete operational control. Detailed measurements and metrics are automatically collected and available for consumption. An expanded COE is established to support vCloud operations.
Strategic Partner Level 5	Operational control is automated and policy driven. Automated self-healing operations remediate errors and maintain quality of service. All processes are integrated, and the organization can consistently achieve ITaaS objectives and satisfy business demands.

4.3.3 Evolution of vCloud Operations

The following sections address how the people, processes, and tools that comprise the VMware vCloud Operations Framework evolve as an IT organization adopts ITaaS in a vCloud.

4.3.3.1 People

IT organizational evolution for vCloud computing begins with the VMware design guideline for virtualization operations: *Create a Center of Excellence (COE)*. The COE is central to successful operation of a virtualized environment, and it is essential for vCloud operations. For a detailed description of a COE, see Section 4.5.2.1, "vCloud Infrastructure Operations Center of Excellence."

Initially, the COE is the focal point for architecting, engineering, and administering the vCloud infrastructure. As vCloud computing takes on a more prominent role and purpose-built management tools mature, the organization adopts an increasingly service-driven approach. This paves the way for the next phase of the vCloud operations evolution, in which the focus shifts from the vCloud infrastructure to the services offered through it.

In the second phase, the COE includes responsibilities for current and new roles focused on the following:

- Construction of service offerings
- Service provisioning management
- Proactive operations management
- Integration and automation management

At this point, automation capabilities begin to drive increased operational efficiency, which leads to increased productivity and frees people up to work on other value-add initiatives. Meanwhile, improved management tool capabilities enable greater visibility into the infrastructure, applications, and user experience, all of which help to identify problems before they can lead to unacceptable performance or outages.

Eventually, purpose-built management tools and automation of business-aligned services progress to the point at which vCloud operations evolve from proactive service operation to predictive, policy-driven, end-to-end, vCloud-based service operation. Discrete operational and functional roles evolve into COE roles and skill sets that are focused on management tools and automation capabilities. Operational domain knowledge remains essential but now supplements deep management tools and automation expertise.

4.3.3.2 Process

As the IT organization continues to evolve, the maturity of vCloud management tools and automation drives vCloud process governance and implementation.

Initially, vCloud computing uses the same operational process approach as virtualization. This approach is effective while pilot studies are conducted and the vCloud is used for development and testing. At this stage, results depend on the maturity of the operational processes themselves, their integration with broader enterprise operational processes, and how successfully the combination has been adapted for virtualization. Maturity levels of the operational processes range from those characterized as reactive and immature—those still based on operating a physical infrastructure—to more mature processes adapted for the unique capabilities of virtualization.

As the IT organization evolves and becomes more service driven, operational processes must become more proactive and service focused. This requires implementing management tools that are purpose-built for vCloud and process automation. Traditional, discrete operational process and functional areas continue to exist, but management tools and automation begin to support some process consolidation and efficiency gains.

At this stage, the vCloud operational model cannot be CMDB driven. Instead, multiple federated configuration-management systems need to manage and interact with each other to support the dynamic nature of a vCloud. vCloud operational processes focus on delivering consumer-facing or infrastructure-related services; both are subject to the same service governance and lifecycle management, while blueprint and policy drive their design development.

The nature of vCloud computing forces proactive operations and management:

- Optimal performance and reliability of the vCloud requires enhanced performance management.
- Capacity management relies increasingly on forward-looking demand projections so that resources can be in place before users need them.
- IT financial management can no longer be project driven—it must become resource investment driven.

These factors position IT vCloud operations for the next phase, in which the company expands the vCloud environment and migrates business-critical services to it.

As vCloud management tools and underlying automation mature, the IT organization evolves from proactive to predictive operations and management. Operational process and functional areas are consolidated as management tools provide more intelligent, end-to-end operational capabilities. Configuration and compliance management become policy driven, with automated drift remediation and built-in auditability. Operations management can now be based on predictive analytics, with automated remediation reducing the number of incidents and/or systemic problems. Consumer-facing services that are deployed for subsequent on-demand self-service provisioning, automatically deployed infrastructure, and resource access supplemented by transparent bursting to an external vCloud provider can all use fully automated provisioning. The ultimate goal of *zero-touch* operations is now within reach.

4.3.3.3 Tools

Management tools must mature to support the evolution of vCloud operations. VMware envisions a dramatic change in vCloud operational processes over time based on the evolution of vCloud management tools, an increasing focus on policy-driven automation, and management tool maturity. For example:

- Proactive operations move to predictive operations that directly impact how event, incident, problem, availability, and performance management are realized.

- Configuration management moves from being CMDB based to adopting a more virtualized, on-demand approach in which configuration and relationship information is collected by multiple, federated configuration-management systems that independently manage and interact with each other, and provide data to management tools as required.
- Service offering development evolves from static and discrete vApp-based development to dynamic, blueprint-based and policy-based vApp construction.

These capabilities, along with increases in operational efficiency, process and functional area consolidation, and zero-touch operations, all depend on the evolution of vCloud management tools and a focus on policy-driven automation.

4.4 Changing Role of Information Technology Organizations

IT is undergoing tremendous change. Modern mobile devices such as tablets and smartphones are replacing traditional desktop and laptop platforms, and business users now expect on-demand accessibility of services on mobile platforms. Internal IT organizations are also seeing growing competition from external service providers. *Shadow IT* (IT solutions built and used inside organizations without official approval) continues to grow because some business needs are not serviced internally. The business expects IT to deliver services that do the job well for a fair price today, not six months from now. These trends make it necessary to rethink, reshape, and reimagine the function of IT and its relationship to business.

4.4.1 IT and Business Relationship

The relationship between IT and business must become more service driven, with IT in the role of the preferred supplier and business as the consumer. As the supplier, IT is responsible for providing services when they are needed. The following core IT disciplines apply to this relationship:

- Provisioning focuses on providing on-demand services while responding rapidly to changing business needs.
- IT economics focuses on increasing efficiency and reducing IT costs, optimizing CapEx and OpEx expenditures for the IT organization, and maintaining the expected quality of service.

4.4.2 Rethink IT

IT must move toward *IT as a Service* (ITaaS). IT organizations must become more service oriented, aligning IT services to business-consumable services that must be available on-demand and must be capable of scaling with business growth.

Becoming a service orientation is transformational for an IT organization. The first step in the transformation, server virtualization, has already been taken. Virtualization facilitates resource sharing, and IT organizations are investigating other initiatives to further enhance this capability:

- Implementing a comprehensive vCloud strategy
- Automating infrastructure management and operations
- Virtualizing business-critical applications
- Building new, modern applications for a post–personal computer era

Cloud computing is critical to the success of the ITaaS model. For VMware, it is a logical follow-up to virtualization. A VMware vCloud enables IT to realize cost-effective pooling and sharing of resources without increasing overall IT complexity and costs. vCloud models also allow for a consistent, repeatable architectural approach that reduces support costs.

In this new model, IT moves to a more proactive role, that of an effective business partner that seeks to meet business objectives. The IT supplier and business consumer come together to focus on better quality of delivered services, using negotiated service-level agreements and a process of continuous improvement. This enhances communication between IT and business, improving transparency, flexibility, and cost visibility.

4.5 Organizing for vCloud Operations

A transformative aspect of vCloud computing is its impact on the IT organization. By definition, vCloud computing provides on-demand service delivery and requires a service-driven IT organization. From an organizational perspective, delivering a service based on vCloud impacts all layers of the VMware vCloud Operations Framework: vCloud Business and Consumer Control, vCloud Service Control, vCloud Operations Control, and vCloud Infrastructure Control. It also directly impacts the relationships of these entities with other organizational teams within IT and with customers, who are the key IT shareholders.

4.5.1 Organizational Overview

vCloud Operations focuses on two organizing concepts, vCloud Tenant Operations and vCloud Infrastructure Operations, as well as their relationships to application development, the Network Operations Center (NOC), and customers. (See Figure 4.6.)

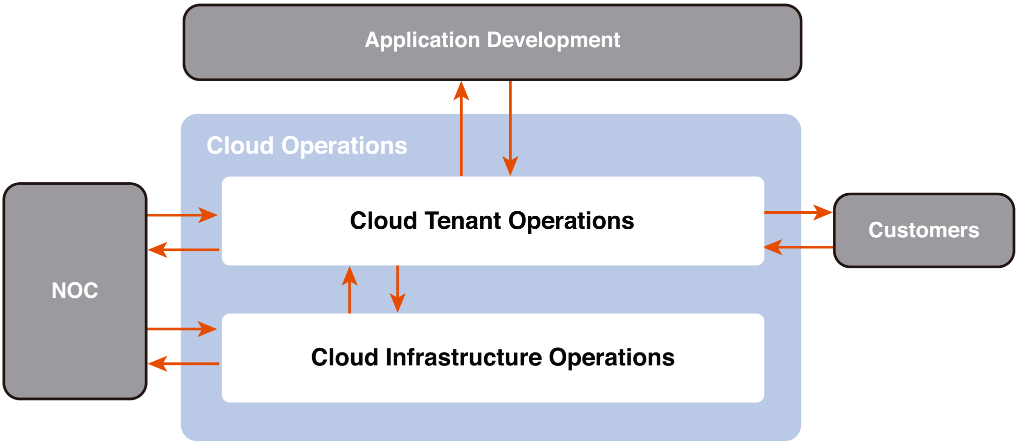


FIGURE 4.6 vCloud organizational overview

In non-vCloud environments, application development is responsible for designing, developing, integrating, and testing a company's custom applications and databases, and integrating and testing third-party applications. It fills the same role in a vCloud environment. The difference is that the vCloud environment promotes an agile approach to development, coupled with modern Platform as a Service–based tools and a tighter relationship to vCloud operations (specifically, vCloud Tenant Operations). Section 4.5.3, "vCloud Tenant Operations," discusses this relationship. Multiple application development teams can interact with vCloud Tenant Operations.

For vCloud computing, the design guideline for the Network Operations Center (NOC) is to become a center for proactive vCloud monitoring, event management, and remediation. From an organizational perspective, the requirement is to add vCloud-specific subject matter experts (SMEs) and begin migrating Tier 2 support responsibilities to the NOC. Instrumenting the NOC with purpose-built vCloud management tools is critical to achieving this. The NOC interacts with vCloud Tenant Operations and vCloud Infrastructure Operations for Tier 3 support as needed. Section 4.5.2, "vCloud Infrastructure Operations," and Section 4.5.3, "vCloud Tenant Operations," discuss this interaction.

vCloud Tenant Operations is responsible for managing end-customer organization relationships and governing, developing, releasing, provisioning, and operationally managing the services offered on the vCloud computing infrastructure. Organizationally, it represents the vCloud Service Control layer of the vCloud Operations Framework and the vCloud Operations Control layer as it relates to the offered services. Service offerings can include applications that an application development team provides.

vCloud Infrastructure Operations is responsible for architecting, engineering, deploying, and operationally managing the underlying logical and physical vCloud computing infrastructure.

4.5.2 vCloud Infrastructure Operations

vCloud infrastructure management encompasses the vCloud Operations Control and vCloud Infrastructure Control layers of the vCloud Operations Framework. It is responsible for architecting, engineering, deploying, and operating the underlying vCloud infrastructure. In VMware terms, the underlying vCloud infrastructure is defined as VMware vCloud Director®, its supporting components such as VMware vCloud Networking and Security™ and VMware vCenter Chargeback™, and the VMware vSphere and the physical infrastructure.

The vCloud Operations Control layer defines operating the vCloud infrastructure. This layer includes the functional operational areas that affect or are most affected by vCloud. They are divided into the following categories:

- Proactive operations management:
  - Change management
  - Configuration and compliance management
  - Capacity management
  - Performance management
  - Access and security management
  - Availability and continuity management
  - Monitoring, event, incident, and problem management
  - Analytics, trending, and metrics
- Integration and automation management

The vCloud Operations Control layer applies to the vCloud infrastructure and to vCloud service operations. For more information, see Section 4.5.3, "vCloud Tenant Operations."

vCloud Infrastructure Operations benefits considerably by reorganization. Traditional infrastructure operations consist of operational functional domains overlaying siloed infrastructure domains with little cross-domain interaction, unless interaction is required for a particular project or deployment. Infrastructure virtualization provides the most recent and compelling opportunity for the infrastructure management component of infrastructure operations to break from this traditional approach by creating a Center of Excellence (COE).

4.5.2.1 vCloud Infrastructure Operations Center of Excellence

The vCloud Infrastructure Operations Center of Excellence (COE) model is an extension of the VMware Center of Excellence model. Many organizations of various sizes have used the VMware Center of Excellence model to facilitate the adoption of VMware technology and to simplify the complexity of managing a VMware virtual infrastructure.

The vCloud Infrastructure Operations COE model defines cross-domain vCloud Infrastructure Operations management accountability and responsibility within team roles across an organization. These team roles enable an organization to consistently measure, account for, and improve vCloud infrastructure operations management.

The COE model further extends operations by including many of the responsibilities previously reserved for the traditional operations team. As vCloud-specific infrastructure operations tools advance, they (combined with automated remediation capabilities) reduce the need for dedicated operations roles. Roles evolve to have a deeper relationship to tools and associated operations. For example, instead of having an Availability Management role, availability management capabilities are built into the infrastructure architecture using a tool such as the VMware vCenter Operations Management Suite™ to proactively monitor availability. Automated remediation scripts help resolve anomalies before services are affected.

The vCloud Infrastructure Operations COE is a focused "virtual" team of vCloud infrastructure operations specialists and related functional groups that together form a vCloud Infrastructure Operations COE ecosystem (see Figure 4.7). The ecosystem serves as the focal point for all decisions and actions involving vCloud infrastructure operations.

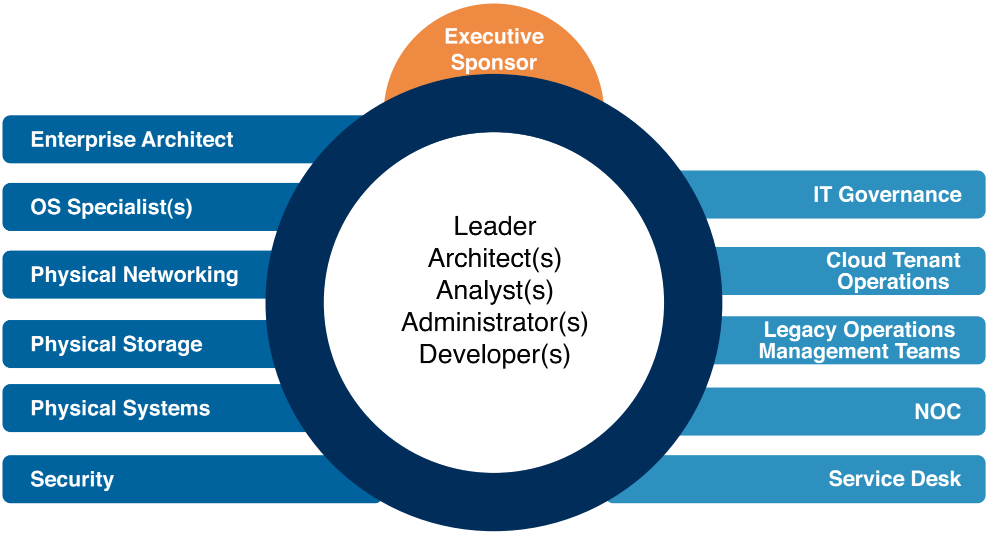


FIGURE 4.7 vCloud Infrastructure Operations Center of Excellence ecosystem

vCloud Infrastructure refers both to internally provided vCloud infrastructure and to infrastructure provided by an external vCloud provider. The following sections describe the primary roles for members of the vCloud Infrastructure Operations COE core team.

4.5.2.1.1 Executive Sponsor

- Provides clear messaging, leadership, and guidance to the entire IT organization and affected organizations about the vCloud Infrastructure Operations COE.
- Drives the cross-domain alignment required to establish a successful, functioning vCloud Infrastructure Operations COE extended team. This level of sponsorship is important for breaking down organizational barriers and mandating integrated

process design and implementation across the affected organizations. Cross-domain alignment and integrated process implementation are required to sustain a vCloud infrastructure at the level needed to support service offerings based on vCloud and associated service levels.

4.5.2.1.2 vCloud Infrastructure Operations COE Leader

Figure 4.7 refers to this function as Leader.

- Provides leadership and guidance to vCloud Infrastructure Operations COE members.
- Has a direct line of communication to the executive sponsor.
- Works with vCloud Tenant Operations regarding the planned vCloud-based service offering portfolio and any portfolio changes.
- Is responsible and accountable for making sure that the vCloud infrastructure can support service offerings based on vCloud and service levels.
- Actively promotes awareness of the impact of the vCloud infrastructure on service offerings and service level support and delivery.
- Facilitates integration of the vCloud infrastructure—for example, for change management—into existing traditional IT operations management processes, as needed.
- Coordinates and assists with planning cloud infrastructure initiatives.
- Provides guidance to Change Management for changes related to the vCloud infrastructure. Might authorize low-risk, low-impact changes to the vCloud infrastructure. Lobbies on behalf of the vCloud Infrastructure Operations COE for preapproved changes.
- Facilitates development and maintenance of vCloud infrastructure capacity forecasts.
- Manages the acquisition and installation of vCloud infrastructure components.
- Maintains management-level relationships with the vCloud Infrastructure Operations COE ecosystem teams.
- Is involved in managing vendor relationships for vCloud infrastructure components.
- Is involved in managing provider relationships with external vCloud providers.

4.5.2.1.3 vCloud Infrastructure Operations COE Architect

Figure 4.7 refers to this function as Architect(s).

- Is responsible for including operational considerations in vCloud infrastructure architecture and design
- Is responsible for developing and maintaining vCloud infrastructure architecture and design documents and blueprints

- Works closely with storage and network groups to architect and design vCloud infrastructure extensions
- Works with enterprise architects to make sure that the vCloud infrastructure architecture is aligned with company architectural standards and strategies
- Is responsible for architecting and designing the vCloud layer in support of the planned service offering portfolio based on vCloud and any portfolio changes
- Is responsible for working with the IT security team to make sure any architecture or design decisions address security and compliance
- Is responsible for architecting and designing solutions for vCloud infrastructure integration points with ecosystem team systems
- Provides subject matter expertise to support build, configuration, and validation processes
- Maintains awareness of VMware software patches and their impact on the environment
- Develops and maintains operational guidelines for the maintenance and support of the vCloud infrastructure
- Mentors and provides subject matter expertise to vCloud Infrastructure Operations COE core and ecosystem team members
- Assists with Tier 3 support to resolve issues related to vCloud infrastructure
- Develops software and hardware upgrade plans
- Develops and maintains the availability policy for the vCloud infrastructure, consistent with operating-level agreement (OLA) requirements

4.5.2.1.4 vCloud Infrastructure Operations COE Analyst

Figure 4.7 refers to this function as Analyst(s).

- Is responsible for developing and maintaining the vCloud infrastructure capacity forecast
- Is responsible for the day-to-day capacity and resource management of the vCloud infrastructure
- Works with the IT security team to make sure that the vCloud infrastructure aligns with IT security and compliance policies; assists in developing automated compliance policies
- Initiates requests for new vCloud infrastructure components
- Assists with Tier 3 support for issues related to vCloud infrastructure capacity and performance
- Assists with the change-management process as applied to the vCloud infrastructure
- Is responsible for maintaining the vCloud infrastructure asset-management data

- Is responsible for tracking and analyzing vCloud infrastructure performance, usage, and other operational analytics
- Is responsible for validating billing metering data collected for the service offerings based on vCloud

4.5.2.1.5 vCloud Infrastructure Operations COE Administrator

Figure 4.7 refers to this function as Administrator(s).

- Deploys and configures vCloud infrastructure components
- Executes the validation plan when deploying new infrastructure components
- Works with vCloud Infrastructure Operations COE ecosystem team members to configure vCloud infrastructure components
- Is responsible for auditing vCloud infrastructure component configuration consistency
- Develops and maintains vSphere and vCloud internal user access roles
- Creates, configures, and administers vCloud provider-related components, such as vCloud Networking and Security, vCenter Chargeback, and vCloud-specific operational management tools
- Works with the IT security team to implement vCloud-related security and compliance policies
- Determines maintenance windows for the vCloud infrastructure consistent with operating-level agreement requirements
- Provides Tier 3 support of the vCloud infrastructure
- Tests and installs vCloud infrastructure patches
- Verifies that the vCloud infrastructure is correctly instrumented for monitoring and logging purposes
- Is responsible for working with developers and other teams to implement any required vCloud integration with external systems
- Works with developers to implement workflows that impact the vCloud infrastructure

4.5.2.1.6 vCloud Infrastructure Operations COE Developers

Figure 4.7 refers to this function as Developer(s).

- Works with COE ecosystem teams to implement any required vCloud integration with other applications
- Develops, tests, and deploys vCloud-impacting automation workflow
- Evangelizes and mentors vCloud COE ecosystem teams about vCloud integration and automation

- Develops and maintains vCloud integration and automation workflow documentation
- Works with vCloud COE members and the ecosystem team to establish integration and automation monitoring
- Works with vCloud COE members and the ecosystem team to establish automated event remediation wherever possible and appropriate
- Provides Tier 3 vCloud integration and automation workflow support

Because these roles and responsibilities require unique skills, a different person should fill each role. With the exception of the vCloud Infrastructure Operations COE Leader, the number of people taking on each role depends on the scale and scope of the vCloud infrastructure.

4.5.2.2 Role of vCloud Infrastructure Operations COE in Standardization

In a traditional organization, multiple business units drive IT. The business unit (BU) controls IT funding, and each BU can enforce separate infrastructure policies and procedures. This approach leads to disjointed architectures and a lack of standardization. IT groups that support such an environment struggle to achieve agreed-upon operating levels, leading to end-user frustration, IT support inefficiencies, and possibly even financial liability.

The implementation of a vCloud changes this scenario. A vCloud is built as a shared resource that requires enforcement of consistent standards across the entire IT organization. To define and enforce these standards, all infrastructure policies and procedures associated with the vCloud should be driven by the vCloud Infrastructure Operations COE team instead of by BUs. This shift poses a significant challenge for organizations that try to move into a vCloud-appropriate infrastructure operating model. The vCloud Infrastructure Operations COE needs to negotiate with different business groups and rely on executive sponsorship and support during the transition to vCloud. More rigorous standards need to apply across the whole organization.

One recommended approach is to align vCloud Tenant Operations with the organization's phased development approach, adding a *vCloud-first policy* during the analysis and design phase for all new projects. Other recommendations include running vCloud Tenant Operations–driven assessments on applications that are being considered for migration to the vCloud. Assessments determine gaps and set expectations with business units on expected changes. The key to success is the capability to balance agility to meet business needs with stringent enforcement of defined standards within the vCloud.

4.5.2.2.1 Layers of Standardization

The vCloud is a shared resource running on infrastructure supported by the vCloud Infrastructure Operations Center of Excellence and core infrastructure teams. Whereas the vCloud Infrastructure Operations COE sets standards for the vCloud, core infrastructure teams might develop standards for the infrastructure that supports the vCloud. For example, the storage team might create standards for how new logical unit number (LUN) storage is presented for vCloud consumption. This layer of abstraction allows the storage team the flexibility to choose the most cost-effective SAN vendor and, if required, support a multivendor environment.

4.5.2.2.2 Measurement with Industry Benchmarks

vCloud technology is evolving at a rapid pace. After a vCloud is established within an organization, a continuous improvement cycle needs to be set up with annual reviews to make sure that the organization's vCloud is not lacking any current industry standards or benchmarks. The vCloud Infrastructure Operations COE is responsible for running this assessment and presenting the results, including recommendations for remediation, back to the leadership team.

4.5.3 vCloud Tenant Operations

vCloud Tenant Operations is central to governing, developing, and providing vCloud service offerings. It incorporates the Service Control layer of the vCloud Operations Framework and the Consumer Management component of the IT Business and Consumer Control layer. It also includes an Operations Control layer specifically applied to services. Figure 4.8 shows a high-level view of Tenant Operations and its ecosystem.

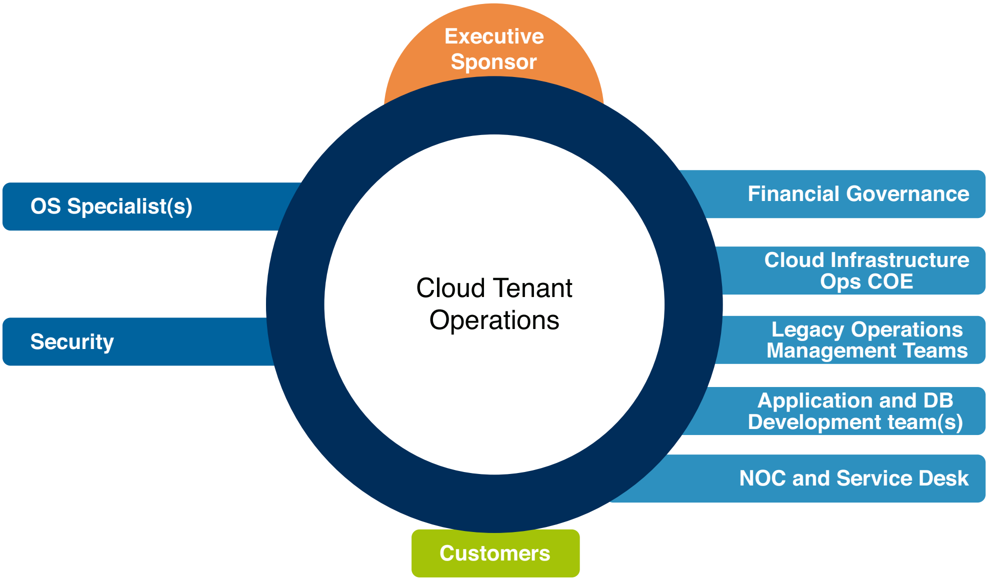


FIGURE 4.8 Tenant Operations

The following roles and responsibilities are involved in vCloud Tenant Operations:

- vCloud Service Leader:
  - Provides leadership and guidance to vCloud Tenant Operations members
  - Has a direct line of communication to the executive sponsors
  - Maintains a working relationship with the vCloud Infrastructure Operations leader

  - Actively promotes awareness of tenant operations team to end-user organizations
  - Maintains management-level relationships with the tenant operations ecosystem teams
  - Assigns vCloud Service Offering responsibilities to service owners

- Customer Manager:
  - Is responsible for establishing and maintaining a working relationship with end-user organizations.
  - Determines and collects business requirements for end-user organization service offerings. Works with the designated vCloud Service Owner to translate the business requirements into a vCloud Service Definition.
  - Works with end-user organizations to understand project service offering demands.
  - Is responsible for end-user organization issue escalation.

- vCloud Service Owner:
  - Is responsible for overall definition and delivery of the vCloud service offering.
  - Works with vCloud Consumer Management to collect end-user requirements and translate them into a vCloud service definition.
  - Works with IT Financial Management to determine a price for the vCloud service offering and determine whether multiple prices are appropriate if the service offering is provided in multiple service tiers.
  - Provides the required information to Service Catalog Management, to correctly set up the service catalog offering.
  - Develops service-level agreements (SLA) and operating-level agreements (OLA) for the vCloud service offerings for which they are responsible. Also negotiates updated SLAs and OLAs as the service offering is updated.
  - Leads development and enhancement efforts and works with vCloud Service Architects on the vCloud service offerings.
  - Is responsible for Tier 3 support and escalations for the vCloud service offerings.
  - Makes sure that the service levels are met through corresponding OLAs with vCloud Infrastructure Operations.
  - Regularly monitors and reports on service level attainment for the vCloud service offerings.

- vCloud Service Portfolio Manager:
  - Develops and maintains vCloud Service Portfolio policy, including criteria for acceptance and rejection.
  - Manages the portfolio of vCloud services and works with IT management to develop the vCloud service offering strategy that determines what services are included in the portfolio. Makes sure that the service offering strategy aligns with the IT strategy.
  - Proactively identifies potential vCloud service offerings based on demand information gathered from vCloud Consumer Managers or other sources, such as requests coming in through the service desk.
- vCloud Service Catalog Manager
  - Manages the vCloud service offering catalog and makes sure that all the information contained in the catalog is accurate and up-to-date
  - Maintains the consumer self-service catalog portal information
- vCloud Service Architect
  - Defines a vCloud service offering based on the requirements provided by the vCloud service owner after it is determined that the service offering is to be included in the vCloud Service Portfolio
  - Translates vCloud business requirements into technical requirements that can be used to architect a vCloud service offering
- vCloud Service Developer:
  - Works with the vCloud Service Architect to understand technical requirements for the vCloud service offering
  - Works with the application development team to incorporate custom or third-party applications into vCloud service offerings as needed
  - Develops new vCloud service offering components into blueprints, or constructs blueprints from existing vCloud service offering components for automatic provisioning
  - Develops and maintains vCloud service offering blueprint documentation
  - Works with the vCloud Service Analyst and application development to define service monitoring
  - Works with the vCloud Service Analyst and application development to establish automated event remediation wherever possible and appropriate
  - Works with the vCloud Service Analyst and application development to make sure security, operations, and chargeback metering capabilities are built into vCloud service offerings

  - Provides support for Tier 3 vCloud service offerings
  - Develops service-related and service integration workflows
  - Develops customizations for and maintains the online consumer self-service catalog capability

- vCloud Service QA:
  - Develops test plans, and tests and accepts services as ready for release to production, regardless of whether the services were developed in-house or by third parties, or are SaaS-based. Also performs post-release validation of services.
  - Develops test plans, and tests and accepts service-related and service integration workflows as ready for release to production. Also performs post-release validation.
  - Develops test plans, and tests and accepts online consumer self-service catalog capabilities as ready for release to production. Also performs post-release validation.
  - Is responsible for making sure that service desk personnel are trained to support the services that are put into production.

- vCloud Service Analyst:
  - Develops and maintains service capacity forecasts.
  - Is responsible for the day-to-day capacity and resource management of services.
  - Works with the IT security team to verify that services align with IT security and compliance policies. Assists in developing automated compliance policies.
  - Initiates requests for new or expanded service capacity.
  - Assists with Tier 3 support for issues related to tenant-deployed services.
  - Monitors and analyzes service performance, availability, usage, and other operational analytics.
  - Verifies that the NOC can support released services.
  - Works with the service QA to release services into production and coordinates any required change management. This responsibility decreases over time as the release process is automated and services consisting of previously released components are considered preapproved from a change management perspective.

- vCloud Service Administrator:
  - Administers tools vCloud Tenant Operations use to govern, develop, and operate services.

  - Administers customer vCloud environments.
  - Administers customer vApps and applications contained in vApps, if offered as a service. This is not usually applicable for development and test customers.

Either a single person or multiple people can satisfy these roles and responsibilities. The decision to employ one or multiple people depends on the number of vCloud service offerings. For a new vCloud environment, initial staffing should include the following roles and responsibilities:

- A single Consumer Manager.
- A single vCloud Service Portfolio Manager who is also responsible for vCloud Service Catalog management.
- One or more vCloud Service Owners, each responsible for conducting vCloud service development and working with other teams to make sure that the agreed-upon vCloud service levels for their vCloud service offering or suite of vCloud service offerings are maintained. The number of vCloud Service Owners depends on the number and complexity of the services to be offered, as well as the rate of service offering change.
- A single vCloud Service Architect.
- One or more vCloud Service Developers, depending on the number and complexity of the services to be offered and the rate of service offering change.

4.5.3.1 Relationship to Application Development

vCloud Tenant Operations interacts with application development teams from the following perspectives:

- Application development team as a customer
- Service development
- Production operations

The application development team is a customer of vCloud Tenant Operations. It uses a service that can provide virtual resources for deploying a development environment or a service in the form of PaaS. vCloud Tenant Operations monitors the environment for availability and is also involved in the release of the application as a service in the vCloud environment.

For service development, vCloud Tenant Operations interacts with an application development team if a custom application is needed to provide the service. Application development is seen as a partner (as well as a customer) in the service development process. vCloud Tenant Operations works with application development to make sure the application is properly instrumented for meaningful monitoring, security, and metering (for showback or chargeback). In addition, the teams work closely together to release the service into production.

The final perspective is production operations. In this case, vCloud Tenant Operations interacts with an application development team, if needed, in a Tier 3 support capacity.

4.5.4 Evolution of Organizational Structure for vCloud

The traditional IT organizational structure must evolve to support a model based on vCloud.

4.5.4.1 Traditional Organization Structure

Figure 4.9 shows an example of a traditional organization structure.

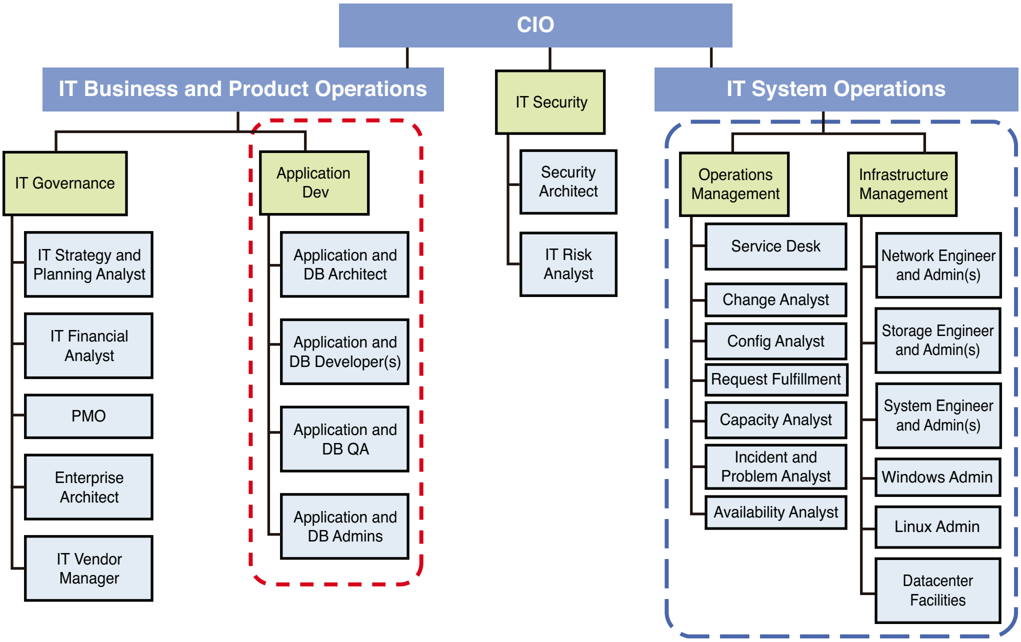


FIGURE 4.9 Traditional organization structure

This structure represents a traditional organization with two core groups: application development and infrastructure. The application development team focuses on application creation, and the infrastructure team focuses on management of hardware resources and daily operation of components. This model applies for most VMware customers, but there is limited focus on the cloud services. VMware associates this model with the reactive state in the vCloud capability model.

The traditional organization has limited focus on cloud management. Responsibilities for supporting the vCloud are typically handled by the roles that manage the physical and virtual world. This organizational structure has limitations and needs to evolve to fully realize the benefits of a cloud.

4.5.4.2 Organization Structure Focused on vCloud

Figure 4.10 shows an example of how the organization structure might evolve to support and effectively manage a vCloud.

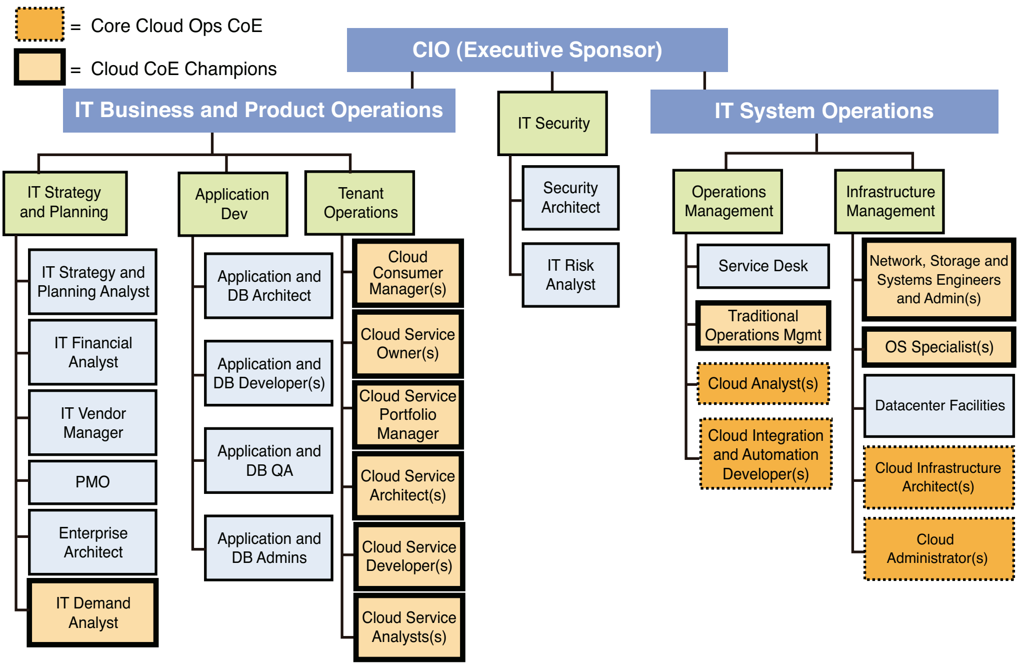

= Core Cloud Ops CoE
= Cloud CoE Champions
CIO (Executive Sponsor)
IT Business and Product Operations
IT Security
IT System Operations
IT Strategy and Planning
Application Dev
Tenant Operations
Security Architect
IT Risk Analyst
Operations Management
Infrastructure Management
IT Strategy and Planning Analyst
IT Financial Analyst
IT Vendor Manager
PMO
Enterprise Architect
IT Demand Analyst
Application and DB Architect
Application and DB Developer(s)
Application and DB QA
Application and DB Admins
Cloud Consumer Manager(s)
Cloud Service Owner(s)
Cloud Service Portfolio Manager
Cloud Service Architect(s)
Cloud Service Developer(s)
Cloud Service Analysts(s)
Service Desk
Traditional Operations Mgmt
Cloud Analyst(s)
Cloud Integration and Automation Developer(s)
Network, Storage and Systems Engineers and Admin(s)
OS Specialist(s)
Datacenter Facilities
Cloud Infrastructure Architect(s)
Cloud Administrator(s)

FIGURE 4.10 Organization structure focused on vCloud

The organizational structure based on vCloud represents a modern organization focused on vCloud with three core groups: application development, Tenant Operations, and Infrastructure Operations. The following describes how the model based on vCloud is different from the traditional model.

- IT System Operations:
  - The organizational model based on vCloud includes creating a focused group, the vCloud Infrastructure Operations Center of Excellence (COE), to support and manage vCloud infrastructure and operational components.
  - Under Infrastructure Management, a focused vCloud Infrastructure Operations COE champion role is added within the core infrastructure and operating system groups. The infrastructure-focused vCloud champions are responsible for pooling core physical infrastructure components to support the vCloud platform. The operating system group aligned vCloud champions work with the focused vCloud Infrastructure Operations COE to manage and maintain operations system standards for autodeployment packages within the vCloud.

  - vCloud COE champions also act as the liaison between their respective groups and the vCloud Infrastructure Operations COE team. The goal is enhanced communication and alignment to support vCloud agility.
  - Under Operations Management, the traditional operations management process teams allocate vCloud COE Infrastructure Operations champions' roles focused on operational governance and process automation. The standard ITSM processes are still valid, but they need to evolve and become proactive to support the dynamic nature of the vCloud. The vCloud Infrastructure Operations COE champions in the operational space work closely with the vCloud architect and analyst to support this goal.
  - The service desk needs to closely work with the vCloud Infrastructure Operations COE team. This interaction is critical to successfully operationalize the vCloud. The service desk needs to act on proactive alerts before incidents occur. This alignment requires a dedicated service desk representative to take on vCloud Infrastructure Operations COE champion roles.

- IT Business and Product Operations:
  - The creation of a service-focused vCloud Tenant Operations group is a significant shift from the traditional organization model.
  - The vCloud Tenant Operations group is essential to achieving higher maturity in the cloud because it focuses on supporting services instead of applications. Services are at the core of the vCloud concept. vCloud Tenant Operations moves the overall organization to a service mindset; the primary IT objectives are to manage and maintain services offered in the vCloud.

4.6 vCloud Business and Consumer Control

The vCloud Business and Consumer Control layer deals with an organization's overall IT vCloud computing strategy, management of IT from a business prospective, and consumer interaction management.

4.6.1 Introduction to IT Business Management

IT Business Management (ITBM) is part of the top two service layers of the vCloud Operations Framework: Business and Consumer Control and Service Control (see Figure 4.3). ITBM addresses the business-driven strategy, as well as consumer-driven requirements and demand for vCloud services to be offered. It offers IT executives the visibility and control required to run IT as a business. In addition, it simplifies and automates the strategic business aspects of IT service delivery by optimizing the customer-specific cost elements and service-level requirements that directly influence IT service value.

4.6.1.1 ITBM Process Components

The ITBM layer is divided into the following major subcomponents:

- **IT Governance:** Focuses on financial transparency, with the capability to collect, model, and report costing data aligned to IT services. This subject area includes the following:
    - IT Financial Management
    - Demand Management and Budget Planning
    - IT Risk Management
    - IT Vendor Management
    - Accounting and Billing
- **Consumer Management:** Aligns IT services with customer requirements and makes sure that the customer catalog satisfies business requirements. It is responsible for managing customer expectations and providing customers with control and governance across the IT service portfolio. This subject area includes the following:
    - Consumer Service Catalog Management
    - Consumer Management and Reporting
- **Service Governance and Lifecycle Management:** Provides the methodology and control over the proposal, acceptance or rejection decision, definition, and end-to-end disposition of services and service offerings, along with governance and control over the quality of services available. This subject area includes the following:
    - Service Portfolio Management
    - Service Level Management
- **Service Design and Development:** Provides methodology and structure for the creation of new IT services. It enhances cost efficiency by reviewing service costs, chargeback, and metering as part of the development cycle. It also adds agility and speed when creating services by adding appropriate blueprints, and it allows for bundling and tiering of IT infrastructure resources. This subject area includes the following:
    - Infrastructure Architecture and Engineering Services
    - Service Chargeback and Metering

4.6.1.2 VMware Product Alignment

VMware addresses ITBM with the following products to help customers:

- **vCenter Chargeback:** End-to-end cost reporting solution for virtual environments that leverages integration with vSphere and vCloud Director.
- **VMware IT Business Management Suite:** Set of SaaS business applications that automate key processes for IT business management. Through its proactive planning, billing, and cost optimization capabilities, ITBM provides the visibility and

predictability that enable stakeholders to improve value and align spending with business goals. It also automates the core financial processes needed to easily plan, charge, and optimize the cost and value of IT. The ITBM suite includes these components:

- **IT Costing:** Maps the connections between IT services and their underlying cost drivers using an intuitive graphical approach that enables total cost of ownership (TCO) and unit cost tracking
- **IT Demand Management and Budget Planning:** Facilitates accurate, fact-based IT budgeting, planning, and forecasting
- **IT Showback and Chargeback:** Gives business units visibility into IT costs and alternatives, including fully itemized billing and chargeback
- **IT Cost Optimization:** Automatically identifies potential areas for ongoing cost reduction, such as candidates for virtualization and consolidation, storage tiering, SLA reduction, end of life, deferral of upgrades, and support reduction
- **Vendor Manager:** Provides a control and optimization mechanism for vendor agreements that proactively governs contractual commitments
- **SLA Manager:** Sets, tracks, and reports on SLAs, key performance indicators (KPIs), and key value indicators (KVIs) for services, vendors, and customers, and performs root cause and business impact analysis at all levels

4.6.1.2.1 Relationship Between Chargeback and ITBM Suite

vCenter Chargeback collects virtualization and vCloud cost data by integrating with vSphere and vCloud Director. It then provides cost data to the ITBM suite for inclusion in cost models.

Both products are connected by the vCenter Chargeback Connector, which scans vCenter Chargeback for a specific hierarchy and creates a report schedule to generate cost reports for this hierarchy on a daily basis. The connector also retrieves both generated and archived reports and provides the cost data for each virtual machine in the hierarchy to the IT Business Management Suite.

Based on the cost data collected by the connector, the IT Business Management Suite populates detailed analysis reports in its cost model and CIO dashboard. This integration provides visibility to CIOs across all IT assets and enables them to easily identify cost reduction opportunities by comparing virtualization, vCloud, and physical costs.

4.6.1.2.2 Cost Models

The ITBM Suite provides out-of-the-box (OOTB) cost models. A *cost model* is a multitiered set of allocation rules that map the financial relationships from the general ledger up to the business units within the organization. The relationships reveal which entities drive the cost of other entities.

The *cost browser* provides a simple way for users (typically the IT finance administrator) to create and modify a cost model that defines the cost relationships in their business structure. Cost models can be modified periodically by adding or deleting elements and changing dependencies to reflect the current contributory relationships between cost object types.

The OOTB cost model does not necessarily reflect all financial aspects of a fully mature IT organization. Instead, it provides immediate value to typical IT organizations and introduces design guidance for object types and common allocation rules that allocate cost end to end from the general ledger to the business units. If needed, the model can be enhanced to reflect any organization cost structure and data sources.

4.6.1.2.3 Integration with vCloud

Using the ITBM Suite, the customer gains unprecedented visibility and transparency across all IT components (physical, virtual, and vCloud). The ITBM Suite enables automatic tracking and processing of IT cost and service data across the organization. ITBM dashboards provide a 360-degree view of what IT services cost to deliver and the service levels that are provided. This visibility enables IT to run like a business and helps IT executives make fact-based decisions.

4.7 vCloud Service Control

vCloud Service Control deals with service governance and lifecycle management, and the design and development of vCloud-based IT services.

4.7.1 vCloud Service Governance and Lifecycle Management

The purpose of vCloud Service Governance and Lifecycle Management is to implement a standard methodology and control over the proposal, acceptance or rejection decision, definition, and end-to-end disposition of services and service offerings. It also provides governance and control over the quality of available services. Elements include Service Portfolio Management, Catalog Management, and Service Level Management.

4.7.1.1 Service Portfolio and Catalog Management

The purpose of the *service portfolio* is to accept or reject service proposals and maintain the overall catalog of services, whether rejected, under development, deployed, or retired. A primary responsibility of Service Portfolio Management is to verify that the services accepted for development and deployment align with the strategic and business requirements of the organization and its customers. This includes continuous review to allow adjustment of services due to new requirements and retirement of existing services due to lack of demand or replacement with a newer service.

The purpose of a *service catalog* is to maintain the active set of services. Active services are those under development or currently offered to customers for use in the vCloud environment. In this context, the service catalog is part of the overall service portfolio (as opposed to the consumer service catalog, from which customers deploy service offerings). The tool that supports the service portfolio and contains the service catalog provides a mechanism for automatically populating the consumer self-service catalog from the service offerings defined in the service catalog as part of the service offering release process. Regular reviews of the service catalog should be performed and adjustments made in line with feature changes in future releases of vCloud Director, vSphere, or other supporting products.

4.7.1.1.1 vCloud Service Catalog Components

The service catalog for the vCloud that vCloud Director supports offers service components to the end customer. At a minimum, the service catalog must define the following:

- **Organization container:** The *container* for the customer's IaaS, with attributes that hold basic, default service configuration information. Typically, only one organization container is purchased per customer.
- **Organization virtual datacenters:** The boundaries for running the virtual machines within the IaaS service, configured with sizing information based on the customers' requirements, with an appropriate SLA assigned to them. A minimum of one organization virtual datacenter is required for a customer to offer a service. Additional organization virtual datacenters can be requested, if required.

In addition to these core vCloud components, an organization can establish a standard set of offerings within the vCloud service catalog to provide customers with vApps (standardized groupings of preconfigured virtual machines) and media (installable software packages).

After being accepted into the service portfolio, the service and constituent service offerings should be defined with at least the following components:

- Service description
- Service requirements
- Service-level agreements
- Support terms and conditions
- Service lifecycle considerations
- Projected demand information for capacity planning
- Pricing and chargeback requirements
- Compliance requirements (regulatory and otherwise)
- Security requirements
- Monitoring and other operational requirements

Including pricing and chargeback, compliance, security, and operational requirements in the service definition is critical because these are core considerations during the service design and development process.

4.7.1.1.2 Service Types

Service types include business user services and technology services.

- **Business user services:** Defined as Software as a Service (SaaS) offerings, these services are generally directly consumed by users and are available as part of the organization's enterprise service catalog.

- **Technology services:** Defined as Infrastructure as a Service (IaaS) or Platform as a Service (PaaS), these technology services are not consumed directly by users, but they enable infrastructure automation that enhances an IT organization's capability to provide business user services.

4.7.1.1.3 Service Interrelationships

For optimal vCloud business user services, all types of technology services must be seamlessly integrated, usually with a workflow engine named the *orchestration layer*. Invoking a business user service can automatically trigger one or more technology services. The rules governing these workflows need to be preconfigured and preapproved for control. They are also needed to provide an agreed-to level of service to the business user. This agreed-to level of service is known as a *service-level agreement* (SLA).

4.7.1.1.4 vCloud Service Catalog Evolution

To improve the vCloud service catalog process and help realize vCloud benefits, as many service offerings as possible should be made available to users through automated provisioning.

In the virtualization world, the initial process for procurement of virtual machines generally follows the model that is applied to physical infrastructure. Although it is effective, it is not the most efficient mechanism for providing services, and vCloud benefits cannot be fully realized unless the process is changed. Figure 4.11 gives a logical representation of the evolution of the vCloud service catalog from this current state to the desired end state.

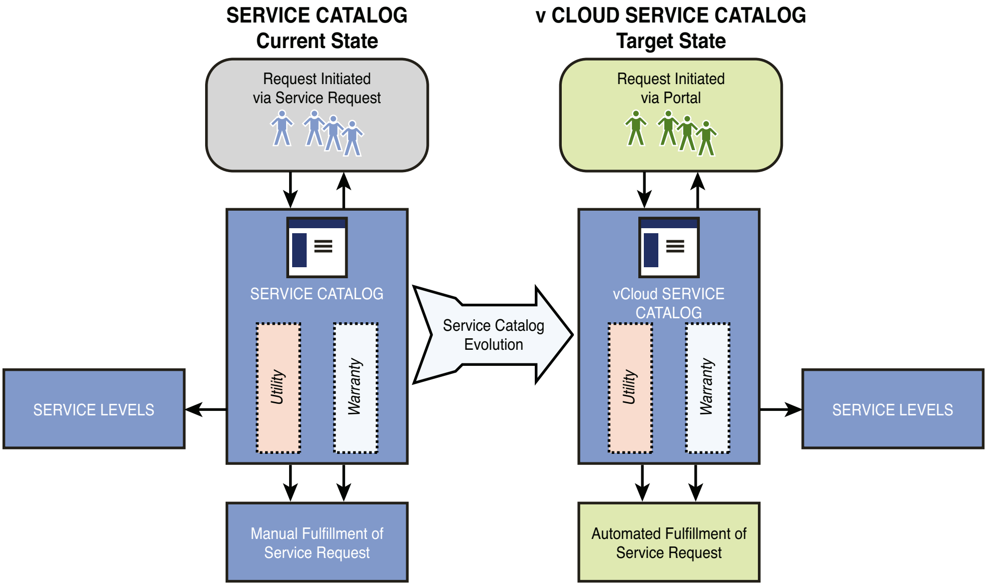

SERVICE CATALOG
Current State
Request Initiated via Service Request
SERVICE CATALOG
Utility
Warranty
SERVICE LEVELS
Manual Fulfillment of Service Request
Service Catalog Evolution
v CLOUD SERVICE CATALOG
Target State
Request Initiated via Portal
vCloud SERVICE CATALOG
Utility
Warranty
SERVICE LEVELS
Automated Fulfillment of Service Request

FIGURE 4.11 Service catalog evolution

In the service catalog current state, when a new service is requested, a service request is submitted to select and provision an offering from the service catalog. In addition to the utility (vApp or organization virtual datacenter) to be provided to the customer, the request includes the required service level provided by the virtual datacenter in which the vApp is to be provisioned, as well as any built-in availability features within the vApp itself. After the service is ordered, the end customer must wait for staff to fulfill the service request for the virtual machines to provide the service to be provisioned.

To satisfy the self-service, on-demand attribute of vCloud computing, the customer should be able to connect to a portal, select the required service offering, and have it automatically provisioned. This removes the need for manual selection from the service catalog and also removes delay in the provisioning processes. Figure 4.11 shows this process as the vCloud service catalog target state.

vCloud Director provides the capability to manage these requests from the service catalog. For vApps, an organization administrator can determine who within the organization has rights to request and provision vApps and thus provide end-to-end self-service. With vCloud Director, the user can select and provision the vApp and also specify the organization's virtual datacenter in which it is to be deployed. Because organization virtual datacenters are associated with provider virtual datacenters, the user is essentially selecting the required level of service.

To transition to the target state vCloud service catalog:

1. Continue with the service request process until the vCloud service catalog is available on the portal.
2. Enable IT staff to perform vCloud service catalog requests with automated provisioning on behalf of the user, including required approvals.
3. Add the capability for users to access the vCloud service catalog and request services that result in automated provisioning of the corresponding vApps, including required approvals.

4.7.1.1.5 Standardization of vCloud Offerings into the Service Catalog

Standardization of service offerings is essential to achieving a scalable, cost-efficient vCloud environment. Typically, compute resource-based service offerings (CPU, memory, and storage) provide a baseline for vCloud consumption and should be standardized as much as possible, regardless of whether they apply to organization virtual datacenters or vApps (and their associated virtual machines).

Compute resources for organization virtual datacenters available in the service catalog should be standardized into various sizes. The required compute resource configurations vary depending on the selected vCloud Director allocation model (allocation pool, pay as you go, or reservation pool) because attributes such as CPU speed and CPU or memory guarantee vary. Combining these two components means that the service catalog can offer differently sized organization virtual datacenters for each type of allocation model.

Similarly, to create a vApp catalog item (public or organization), standardization should be used as possible. From a compute resource point of view, standard-sized virtual machines should be created to use in a *pick list* of machines for vApp creation. These standardized virtual machines can vary in resource size for CPU, memory, and storage (for example, Standard, Standard Plus, Advanced, Premium, and Premium Plus). Because a vApp consists of one or more individual virtual machines, the appropriately sized virtual machines can be selected from the pick list during the vApp catalog creation process.

In addition to the basic compute offerings of the virtual machines within the vApps, it is necessary to develop the service catalog to include vApp software configurations. These can be basic groupings of compute resources and can be expanded over time to offer more advanced services. Table 4.2 shows sample vApp offerings.

TABLE 4.2 Sample vApp Offerings

vApp	Configuration
2-Tier Standard Compute	1x Standard RHEL Web virtual machine 1x Standard Windows Server 2008 Application virtual machine
3-Tier Standard Compute, Advanced Database	1x Standard RHEL Web virtual machine 1x Standard RHEL application virtual machine 1x Advanced MySQL Database virtual machine
3-Server Standard Plus Compute (not necessarily tiered)	3x Standard Plus Windows Server 2008 Application virtual machine

4.7.1.1.6 Establish Service Levels for vCloud Services in the Service Catalog

To provide an appropriate level of service for the vCloud customers' requirements, services should be further differentiated by their corresponding service levels. Service levels can be defined with availability and recoverability attributes such as Recovery Time Objective (RTO), Recovery Point Objective (RPO), and incident response times. The attributes can be applied to the different components within the service catalog.

It is possible to design for different service levels for the virtual machines contained in a vApp. For example, a vApp could contain multiple web servers to provide resilience in the event of server failure, and thus a lower RTO for the service.

Virtual datacenters provide abstracted physical and virtual resources. Different service levels can be defined by using (or not using) the underlying hardware technology (such as server capabilities, storage array technologies, storage protocols, and replication) and virtualization technology (HA, DRS, VMware vSphere vMotion®, and others).

Combined, vApps and the capabilities of the virtual datacenters on which they can be deployed offer the capability to create a powerful and extensive vCloud service catalog.

4.7.1.2 Service Level Management

Service Level Management defines the service-level agreement (SLA) associated with a vCloud service offering or a tier of service, negotiates corresponding operating-level agreements with the service provider to support the SLAs, and regularly monitors service levels and reports on results.

4.7.1.2.1 Definition of Service-Level Agreement

A *service-level agreement* (SLA) is a predetermined agreement between the service consumer and the service provider that measures the quality and performance of the available services. SLAs can be of many types, from those that measure pure service availability to those that measure response time for service components and process workflows as experienced by users.

Services run at every layer of the vCloud stack, so service consumers might be business users or internal IT groups who access the vCloud primarily for technology and infrastructure services. SLAs for base technology services that business users do not consume directly but are needed to make sure that downstream operations and infrastructure components support the business users' SLAs are referred to as *operational-level agreements* (OLAs).

4.7.1.2.2 vCloud Layers and SLAs

A typical vCloud computing environment consists of multiple layers (IaaS, PaaS, SaaS, and possibly others). The customer chooses how to implement the vCloud stack based on business requirements. Options include creating a private vCloud, using a public vCloud provider, or creating a hybrid vCloud model with both private and public vCloud resources. The enabler for this flexibility is an organization's capability to guarantee availability and performance at every vCloud layer. Signing SLAs externally with service providers or, for a private vCloud, creating SLAs with internal user organizations and supporting OLAs with the IT organization, achieves this.

4.7.1.2.3 Example

Figure 4.12 shows an example use case for an organization with an IaaS layer hosted by a public vCloud provider and the PaaS and SaaS layers maintained internally.

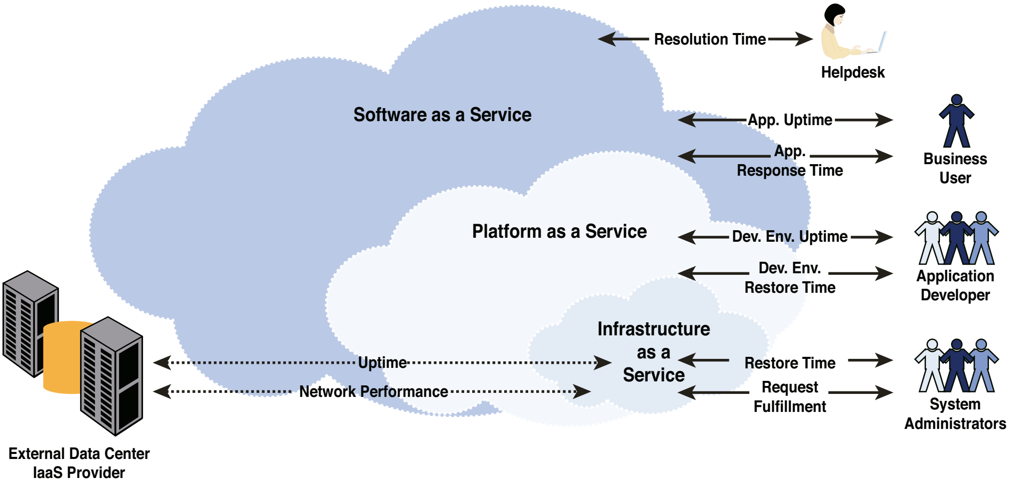


FIGURE 4.12 Example organization with public vCloud IaaS and private vCloud PaaS/SaaS layers

The SLAs shown are for illustration purposes only and are a subset of the total number of SLAs created within an organization in such a case.

The example includes the following SLAs:

- IaaS layer:
  - Uptime/availability SLA signed with the external vCloud service provider
  - Network performance SLA signed with the external service provider
  - Request fulfillment SLA—measure of response time for provisioning and access configuration requests
  - Restore time SLA
- PaaS layer:
  - Uptime/availability SLA for development environment
  - Uptime/availability SLA for critical development environment components
  - Restore time SLA for development environment
- SaaS layer:
  - Uptime/availability SLA specific to an application
  - Application response time SLA—measure of how the application is performing for the business users
  - Time to resolution SLA—time to recover an application in case of a failure

Given this example, the following are some key conclusions:

- SLAs, OLAs, and KPIs are relevant at all levels within a vCloud stack. These agreements are required to provide efficiency and accountability at every layer, for both external providers and internal IT groups.
- These SLAs, OLAs, and KPIs need to be managed within every layer to help isolate systemic problems and eliminate delays.
- SLAs can be between external vendors or providers of vCloud services, or between internal IT groups. An organization can choose whether to implement a private, public, or hybrid vCloud. At every layer, SLAs give organizations flexibility by guaranteeing availability and quality of service.
- Interrelationships exist between SLAs set up at different vCloud layers. A change in quality of service or breach of an SLA at a lower vCloud layer can impact multiple SLAs in a higher vCloud layer. In the example, if a breach of a performance SLA results in the external vCloud provider's incapability to support OS performance needs, the breach has a ripple effect at the SaaS layer, decreasing application performance and response time for business users.
- SLAs need to be continuously managed and evaluated to maintain quality of service in a vCloud. Business needs are continuously evolving, resulting in changing vCloud business requirements. SLAs must be continuously updated to reflect current business requirements.

Consider the impact of adding another 1,000 users to a particular application so that the application becomes mission critical. SLAs supporting the application might need to be updated to provide increased uptime and availability. This might lead to increased demands at the IaaS layer, so SLAs with the external IaaS provider might also have to be expanded.

4.7.1.2.4 vCloud SLA Considerations

vCloud SLA considerations include the following:

- **Uptime/availability SLA:**
  - To what timeframe does the SLA pertain? Timeframes are generally divided into tiers depending on business criticality (9 to 5, 24 by 7).
  - Are maintenance windows (for configuration changes, capacity changes, and OS and application patch management) included or excluded from availability SLAs?
  - Do multi–virtual machine vApps need to be treated as a single entity from an SLA perspective?

- **End user response time SLA:** This is generally focused on overall user experience, measuring response time from local and major remote sites to get a representative view. Measurement is implemented with remote simulators and by running automated robotic scripts.

- **Recovery (system, data) SLA:** What recovery time objectives and recovery point objectives need to be met?
  - Are backups required?
  - Is high availability required?
  - Is fault tolerance required within the management cluster?
  - Is automated disaster recovery failover required within certain time parameters?
- **Privacy SLA (data security, access and control, compliance):**
  - Do data privacy requirements (encryption, others) exist?
  - Are there regulatory requirements?
  - Are specific roles and permission groups required?
- **Provisioning SLA:** Are there provisioning time requirements?
- **SLA penalties:**
  - How are SLA penalties applied?
  - Are they applied as service credits?
  - What legal liabilities apply, and how are they covered?
  - Is there a termination for cause clause in the SLA?
  - What defines an outage, and who bears the burden of claim?
  - What is the track record for delivering on SLAs? These SLA considerations should be applied to external service providers.

4.7.1.3 Roles and Responsibilities

The following are primary roles associated with Service Governance and Lifecycle Management:

- Service Portfolio Manager
- Service Catalog Manager
- Service Owner
- Service Level Manager

4.7.1.3.1 Service Portfolio Manager

The Service Portfolio Manager role is the gatekeeper for accepting proposed services and constituent service offerings into the overall portfolio of vCloud services. Responsibilities include the following:

- Developing service/service offerings analysis and acceptance criteria
- Reviewing and accepting or rejecting service proposals
- Continuously reviewing the overall portfolio of services for applicability and demand
- Providing initial service/service offering demand information for vCloud capacity planning
- Authorizing a service owner to define and develop, or retire, a service or service offering

4.7.1.3.2 Service Catalog Manager

The Service Catalog Manager role manages the "active" service catalog component of the overall service portfolio. The active service catalog contains the definitions for those service/service offerings currently either under development or available to consumers for deployment. Responsibilities include the following:

- Maintaining information about services and service offerings contained in the active service catalog
- Verifying that service and service offering information is accurate and complete, and providing it to consumers through the consumer self-service portal

4.7.1.3.3 Service Owner

The Service Owner role has end-to-end responsibility for defining, developing, maintaining, and decommissioning a specific service or set of services and their component service offerings. Responsibilities include the following:

- Translating business requirements into a service definition
- Defining service/service offering composition details, pricing, service levels, support terms and conditions, operational considerations, and any service-specific compliance requirements
- Working with the Service Architect to translate the service definition into service design and development technical details
- Managing development, deployment, update, and retirement of the services and service offerings
- Tracking demand and service requests for service updating and retirement

4.7.1.3.4 Service Level Manager

The Service Level Manager role establishes and maintains SLAs, and reports on service level attainment. Responsibilities include the following:

- Developing service-level agreements for customers
- Tracking and reporting on service level attainment
- Developing operating-level agreements with the service provider in support of SLAs

4.7.1.3.5 Staffing Considerations

As with most vCloud operations-related roles, staffing depends on scale. Initially, a single person can fill the Service Portfolio and Service Catalog Manager roles. As the number of services and service offerings and the activities involving them increase, the Service Portfolio Manager and Service Catalog Manager roles might each require a person. The same is true for the Service Owner and Service Level Manager roles.

4.7.2 vCloud Service Design and Development Management

The purpose of vCloud Service Design and Development Management is to implement a standard methodology with governance and control across all service development groups within an organization. This area focuses on design and architecture consistency, cost transparency, and service metering based on consumption. It includes subareas for service development, service showback, and metering management.

4.7.2.1 Service Development Management

The process for Service Development Management enforces a structured approach to maintain quality and consistency during service development. The following sections describe the main process components.

4.7.2.1.1 Service Requirements

The Service Requirements process manages the interaction between the service development teams and the business user during development of new services. A clear communication channel is set up, and a standard service definition document template is created and used to capture business requirements. Continuous review takes place, and signoffs occur after every significant service development phase. This function requires analysis time for alignment with the Service Portfolio process. Business scenario and use cases are developed, and a cost-benefit analysis is completed before service development begins.

Focus areas include the following:

- Involvement of all necessary stakeholders
- Documentation of business drivers and requirements that can be translated into appropriate service definitions and SLAs
- Clear definition of operational requirements, along with alignment with appropriate service tiers
- Definition of business scenarios and use cases
- Definition of the business users and roles that interact with services development so that user-centric services are created

- Workflow representation of the service to understand the components of the service, interactions, and sequence of interrelated actions
- Cost-benefit analysis (internal versus external)

4.7.2.1.2 Service Requirements—Initiation Workflow

Figure 4.13 shows a sample service initiation workflow for creating a new service and the roles that are involved in the process.

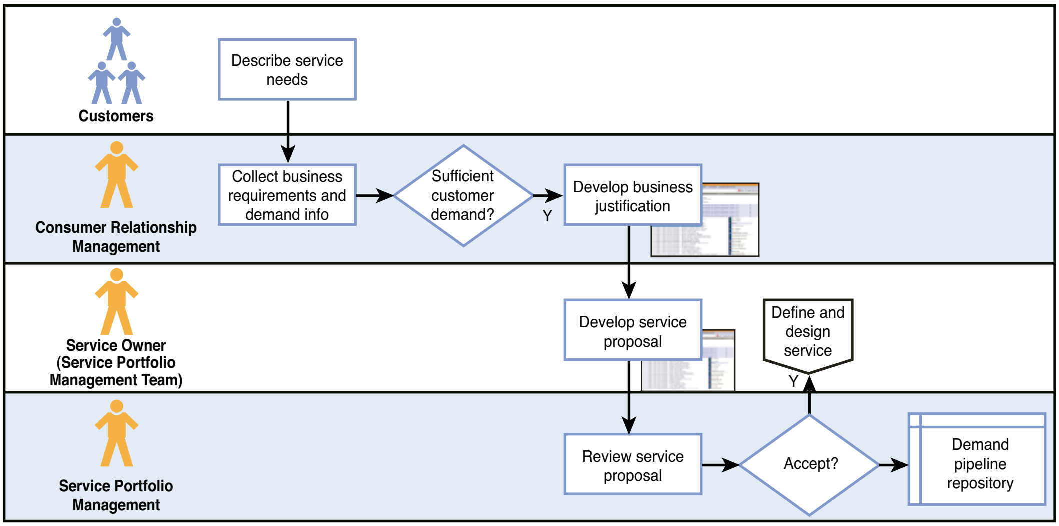

Customers
Describe service needs
Consumer Relationship Management
Collect business requirements and demand info
Sufficient customer demand?
Y
Develop business justification
Service Owner (Service Portfolio Management Team)
Develop service proposal
Define and design service
Y
Service Portfolio Management
Review service proposal
Accept?
Demand pipeline repository

FIGURE 4.13 Service requirements workflow

See Section 4.5, “Organizing for vCloud Operations,” for more information about the organization and roles.

4.7.2.1.3 Service Design

The Service Design process focuses on creating consistent architecture and design for new services. This function is responsible for creating architecture blueprints and service templates for rapid service creation.

Key focus areas:

- High-level design representation of the service, to understand its components, interactions, and sequence of interrelated actions and expected SLAs
- Integration and alignment with the service portfolio and catalog process areas
- Integration and alignment with the service showback and metering process
- Business user signoff on service design

4.7.2.1.4 Service Design—Workflow

Figure 4.14 shows a sample service design workflow and the roles that are involved in the process.

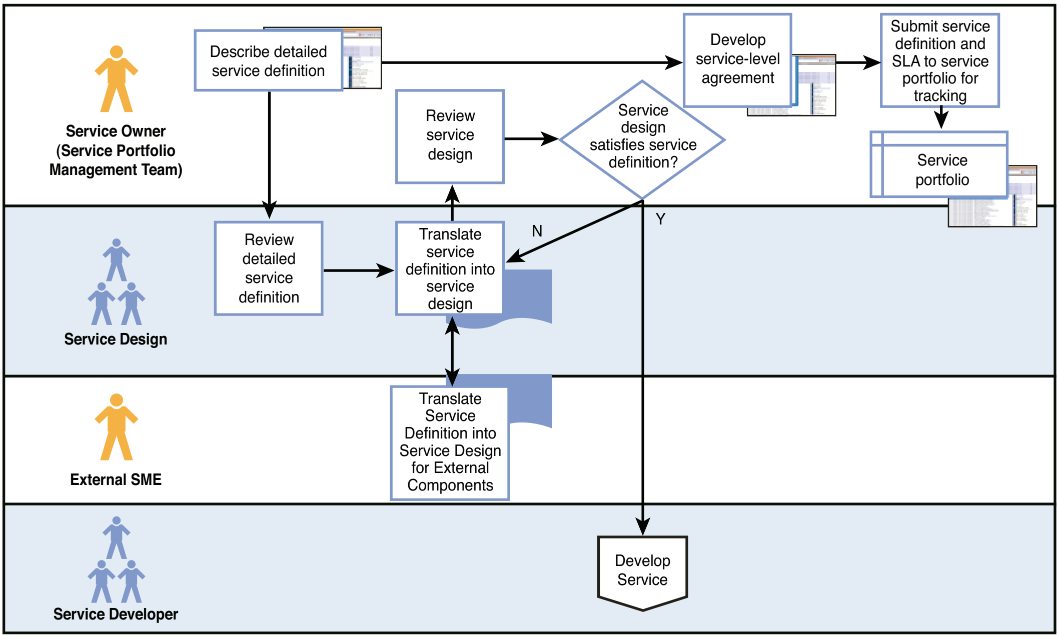

Service Owner (Service Portfolio Management Team)
Describe detailed service definition
Develop service-level agreement
Submit service definition and SLA to service portfolio for tracking
Review service design
Service design satisfies service definition?
Service portfolio
N
Y
Service Design
Review detailed service definition
Translate service definition into service design
External SME
Translate Service Definition into Service Design for External Components
Service Developer
Develop Service

FIGURE 4.14 Service design workflow

Refer to Section 4.5, “Organizing for vCloud Operations,” for more information about the organization and roles.

4.7.2.1.5 Service Development

The Service Development process focuses on developing the services and aligning service development methodologies for an organization. This function requires speed and agility to respond quickly to changing business needs. This function also manages and controls the overall service development environment, platforms, and tools used in the overall service development process.

Key focus areas:

- Agility and rapid response
- Definition of service development methodology (in general, Agile development is recommended)
- Integration and alignment with other operational process areas:
  - Performance SLAs (application response time, bandwidth including burst, time to respond, time to resolution)

  - Availability SLAs (uptime, backup, restore, data retention)
  - Continuity SLAs (RPO, RTO)
  - Scalability
  - Manageability (user account management, supportability)
  - Security (application/data access, management/control access, user accounts, authentication/authorization)
  - Compliance (regulatory compliance, logging, auditing, data retention, reporting)
- Alignment with the service showback and metering management process for service costing, pricing, metering, and billing
- Development of service controls:
  - Continual service reporting, service quality analysis, and trending
  - Automated remediation scripts and integration workflows

4.7.2.1.6 Service Development—Workflow

Figure 4.15 shows a sample service development workflow and the roles that are involved in the process.

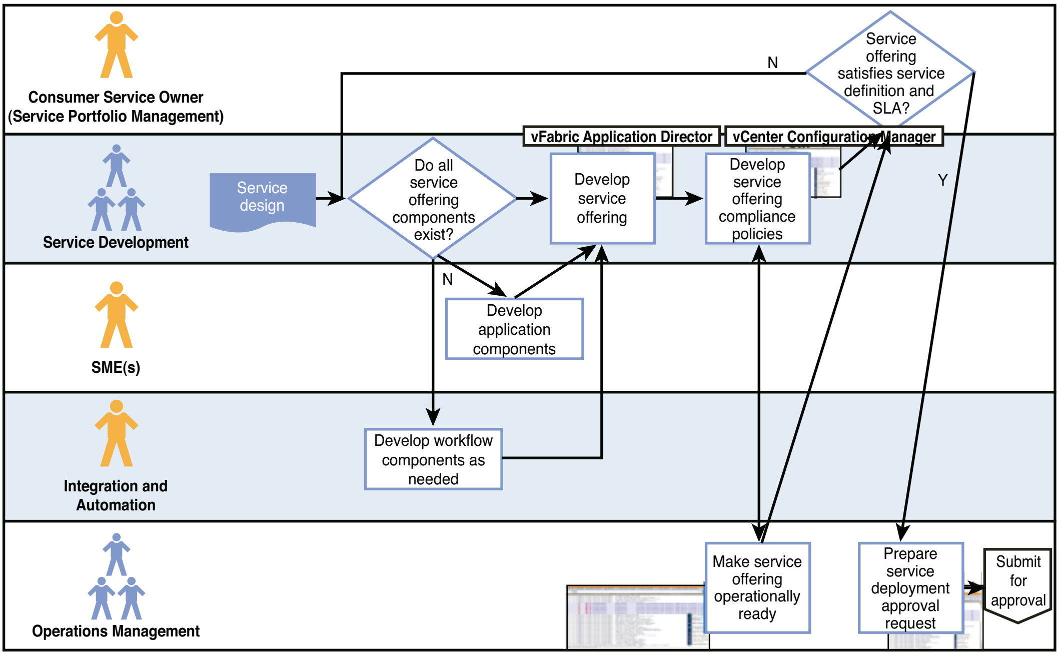

Consumer Service Owner (Service Portfolio Management)
Service Development
SME(s)
Integration and Automation
Operations Management
Service design
Do all service offering components exist?
N
Develop application components
Develop workflow components as needed
vFabric Application Director
Develop service offering
vCenter Configuration Manager
Develop service offering compliance policies
Service offering satisfies service definition and SLA?
N
Y
Make service offering operationally ready
Prepare service deployment approval request
Submit for approval

FIGURE 4.15 Service development workflow

See Section 4.5, “Organizing for vCloud Operations,” for more information about the organization and roles.

4.7.2.1.7 Common Service Development Characteristics

The following are common service development characteristics of a vCloud service:

- **On-demand self-service:** A vCloud service needs to be designed and developed to allow for on-demand provisioning via a service catalog that uses automated workflows with minimal human interaction from the service’s provider side.
- **Service mobility:** Services should be designed to be accessible from multiple end-user computing mobility platforms, such as tablets, phones, and other thin or thick end-user platforms.
- **Resource pooling:** Services should designed to use pooled computing resources, not bound to physical infrastructure. There should be a sense of location independence—the service consumer generally has no control or knowledge over the exact location of the provided resources.
- **Rapid elasticity:** Services should be capable of using the vCloud elastic and bursting feature to support high-utilization timeframes. Services should be designed to automatically scale and release computing resources.
- **Measured service:** Services must be designed with capability to leverage metering capability based on service consumption. This is critical to make the service a viable business investment for the service provider. This feature is at the heart of the service showback and metering process.

4.7.2.2 Service Showback and Metering

The service showback and metering process provides a mechanism for calculating service costs for end users. The short-term goal is to raise awareness of costs based on service consumption usage. For an organization in an initial maturity state, no formal accounting procedures and billing are involved, but as the maturity within the organization increases, the service showback and metering process integrates with the IT business control layer and supports automated IT chargeback.

Key focus areas:

- **Early alignment during the service design and development process:** Showback enables service subscribers to see costs associated with service usage. Showing the cost of consumption is the first step toward moving an organization to IT chargeback, in which consumers pay for services they consume.
- **Showing and calculating true service costs:** Service costing is complex in a vCloud because services are designed to run on pooled resources and have inherent elasticity features. The key to success is understanding and aligning to vCloud cost models, and being able to break down individual service component costs and understand their interrelationships.

4.8 vCloud Operations Control

vCloud Operations Control deals with provisioning and proactive operations management of IT services based on vCloud, with a focus on policy-driven automation.

4.8.1 Provisioning Management

In IT, generally, and in vCloud, *provisioning* typically refers to one of the following:

- Provisioning virtual machines or vCloud components (vApps, business applications, services) as a result of a consumer request
- Provisioning the underlying infrastructure that supports virtualization or vCloud platforms

Provisioning resulting from a consumer request is evolving in vCloud computing. A primary goal when implementing a vCloud computing environment is to lower ongoing OpEx costs. Provisioning that results from a consumer's request is an activity from which significant OpEx savings can be realized. Savings are realized by the following:

- Providing a self-service portal through which a consumer can make requests from a service catalog
- Automating the resulting provisioning process to satisfy the consumer's request

4.8.1.1 Consumer Self-Service Portal

From a consumer perspective, vCloud computing is driving the following new expectations:

- Self-service
- Flexibility and granularity of choices
- Instant gratification
- On-demand services

A private or public vCloud provider can meet these expectations by providing an online consumer self-service capability. For a private vCloud, this capability is deployed internally. When using a public vCloud, consumers might have access to the public vCloud self-service, online service catalog. By providing access, the provider expects to benefit by being able to deliver vCloud services quickly and inexpensively, while still maintaining control over the process. The consumer's expectations are met by automating the provisioning process.

Initially, the online, self-service capability must provide an easy way for the consumer to provision resource-based services in the form of vApps. Ultimately, this must be extended to offer self-service access to any and all IT services, whether as a wholly contained development environment, Software as a Service–based applications, or applications for mobile devices. Providing this addresses consumer expectations regarding flexibility and granularity of choices.

The online, self-service portal should provide the capability to do the following:

- Get secure access
- View available services, costs, and service tiers
- Request vApps and other services based on organizational maturity
- Obtain any required approvals through automated workflows
- Track request status
- View items successfully provisioned
- Perform tasks such as start, stop, and add capacity (at least this minimal set)
- Receive notifications
- Decommission items
- View basic consumption reports
- View the health of provisioned items

4.8.1.2 Provisioning Process Automation

Automating the provisioning process to satisfy consumer requests is a key element in meeting custom expectations and enables the provider to realize OpEx savings. Initially, process automation applies to vApp provisioning, but provisioning of other IT services can also be automated.

Automated vApp provisioning consists of the following:

- An automated vApp provisioning process that handles the entire lifecycle of a vApp
- Automated interaction between the vApp provisioning process and other required processes and associated systems

Figure 4.16 illustrates an example vApp provisioning process that can be fully automated.

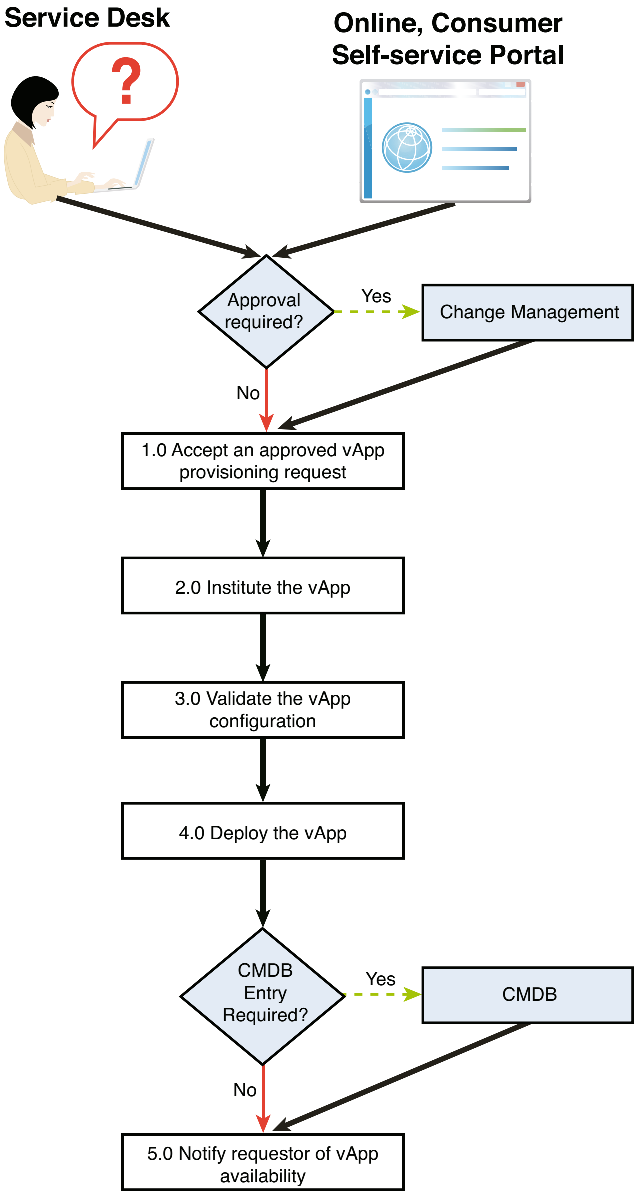

Service Desk
?
Online, Consumer Self-service Portal
Approval required?
Yes
Change Management
No
1.0 Accept an approved vApp provisioning request
2.0 Institute the vApp
3.0 Validate the vApp configuration
4.0 Deploy the vApp
CMDB Entry Required?
Yes
CMDB
No
5.0 Notify requestor of vApp availability

FIGURE 4.16 Provisioning workflow

This vApp provisioning process can be fully automated using VMware vCenter Orchestrator™. The vCenter Orchestrator plug-in directly supports automating the following tasks:

- Instantiating the vApp
- Validating the vApp configuration
- Deploying the vApp

Additional vCenter Orchestrator standard protocol plug-ins (email, SOAP, HTTP REST) and VMware partner application plug-ins provide the mechanisms for automating integration with Change Management and Configuration Management, and with other third-party applications and systems as needed. For information regarding Orchestrator plug-ins, see www.vmware.com/products/datacenter-virtualization/vcenter-orchestrator/plugins.html.

OpEx savings are realized by automating what were previously a set of manually executed steps typically driven by work queues. After automating, a process that previously took days or weeks might take only minutes or hours. In addition to OpEx savings, consumer satisfaction increases as their expectation of instant gratification is met.

4.8.1.3 Provisioning Process Analyst

After provisioning processes are automated, a provisioning process analyst role is needed. Provisioning process analyst responsibilities include the following:

- Working with service development to understand the provisioning implications of new services to be offered
- Providing Integration and Automation Management with requirements for workflow implementation, modification, maintenance, and integration with systems in other process areas
- Working with Service Level Management to understand operating-level requirements for vApp provisioning
- Working with monitoring to properly instrument the provisioning process, create thresholds, and implement monitoring
- Working with Release Management on coordination and validation:
    - Provisioning of workflow releases
    - Updates to existing service catalog entries that affect the provisioning process
    - New service offerings added to the service catalog

Staffing levels for the provisioning process analyst role depends on the following factors:

- Number of distinct service offerings
- Rate of service releases
- Provisioning operating agreements tied to service-level agreements

Whether a part-time or full-time provisioning process analyst or multiple analysts are needed depends on the use and stability of the automated provisioning process and any related operating-level agreements tied to service-level agreements. In most cases, a part-time or, at most, a single full-time analyst is sufficient. Additional OpEx savings can be realized through IT role consolidation. After the provisioning process is automated, the provisioning process analyst responsible for provisioning management should maintain the automation workflows and integration points with other applications. With the appropriate skills and training, this role can be shared with other roles that have integration and automation activity responsibilities.

Cloud computing drives customer expectations toward increased agility and choice and requires less time to deliver. By providing an online consumer self-service portal backed by an automated service provisioning processes, providers can realize OpEx savings and satisfy consumers' expectations.

4.8.2 Capacity Management

Capacity Management focuses on providing vCloud capacity to meet both existing and future needs in support of vCloud service offerings.

For the vCloud provider, the goal of capacity planning is to provide sufficient capacity within the vCloud infrastructure to meet the current and future needs of the services offered to customers. Sufficient reserve capacity must be maintained in the vCloud infrastructure to prevent virtual machines and vApps from contending for resources under normal circumstances, thus breaching agreed service levels.

The vCloud provider components must manage the following:

- Management cluster that contains all the components used to create and manage the vCloud
- Resource clusters that provide resources to the vCloud consumers

The sizing of the management cluster is generally predictable, but consideration needs to be given to the number of vCloud Director cells and the size of the vCloud Director, vCenter, and Chargeback databases. Additionally, if VMware vCenter Operations products are used, the storage required for vApps needs to be considered because it can be substantial in large environments. Initial sizing guidelines for the management cluster are provided in Chapter 3.

Usage can be unpredictable for vCloud consumer resources such as vApps and organization virtual datacenters. To size consumer resources, estimate the initial capacity required and use vCloud Capacity Management techniques, which predict future usage needs based upon past usage trends.

Capacity planning is required to make sure that the vCloud resources supplied to the tenant are used appropriately, are available when required, and expand or contract depending on current and future demand.

4.8.2.1 Capacity Management Process Definition and Components

Historically, capacity management is usually performed when the system is implemented and covers the capacity requirements for the entire lifetime of the system. This creates significant waste during the system's early life because the excess capacity is not required until later, if at all. Potential exists for significant waste during the system's entire lifecycle because of many other factors, including overestimation of usage or early retirement from technology evolution.

Even with virtualization, ensuring that sufficient capacity is readily available is always a concern. Virtualized environments manage capacity by reducing the contention for resources, usually by reducing the ratio of virtual machines to hosts. This approach wastes resources as low ratios are adopted.

To make a vCloud implementation successful, resource waste must be avoided. The capacity-management process must become proactive, with adjustments to capacity configuration as conditions change. It is not sufficient to "set it and forget it." Focusing on proactive capacity management makes it possible to increase the density of virtual machines on hosts. This enables the provider to realize the cost benefits of implementing a vCloud without compromising the services that run on it.

Figure 4.17 illustrates a high-level, proactive capacity-management process. This process is applicable to both vCloud providers and tenants.

Although the proactive capacity-management process appears the same as the traditional capacity-management process, the dynamic nature of the vCloud requires that the proactive process be more agile and rely less on manual intervention. Manual capacity management might be appropriate in a physical infrastructure or during early virtualization adoption stages, but only tooling and automation can provide the proactive capacity-management required for the vCloud.

Long-term capacity issues should be identified early so that the vCloud service is not impacted. With appropriate tooling, early warning is possible. Historical capacity usage behavior can be identified and combined with known future demand to provide a vCloud capacity forecast.

Short-term capacity breaches also need to be identified early so that remediation can be put in place to keep from compromising vCloud SLAs.

With this short-term and long-term knowledge, automation can help make the required resources available in the appropriate environment. For short-term breaches, automation can help identify underused resources in one environment and temporarily transfer them to an environment that is under-resourced. For long-term capacity issues, the automated provisioning process for new resources is predictable and well defined. This makes it possible to provision new resources such as hosts, clusters, or organization virtual datacenter capacity as required, without service breaches.

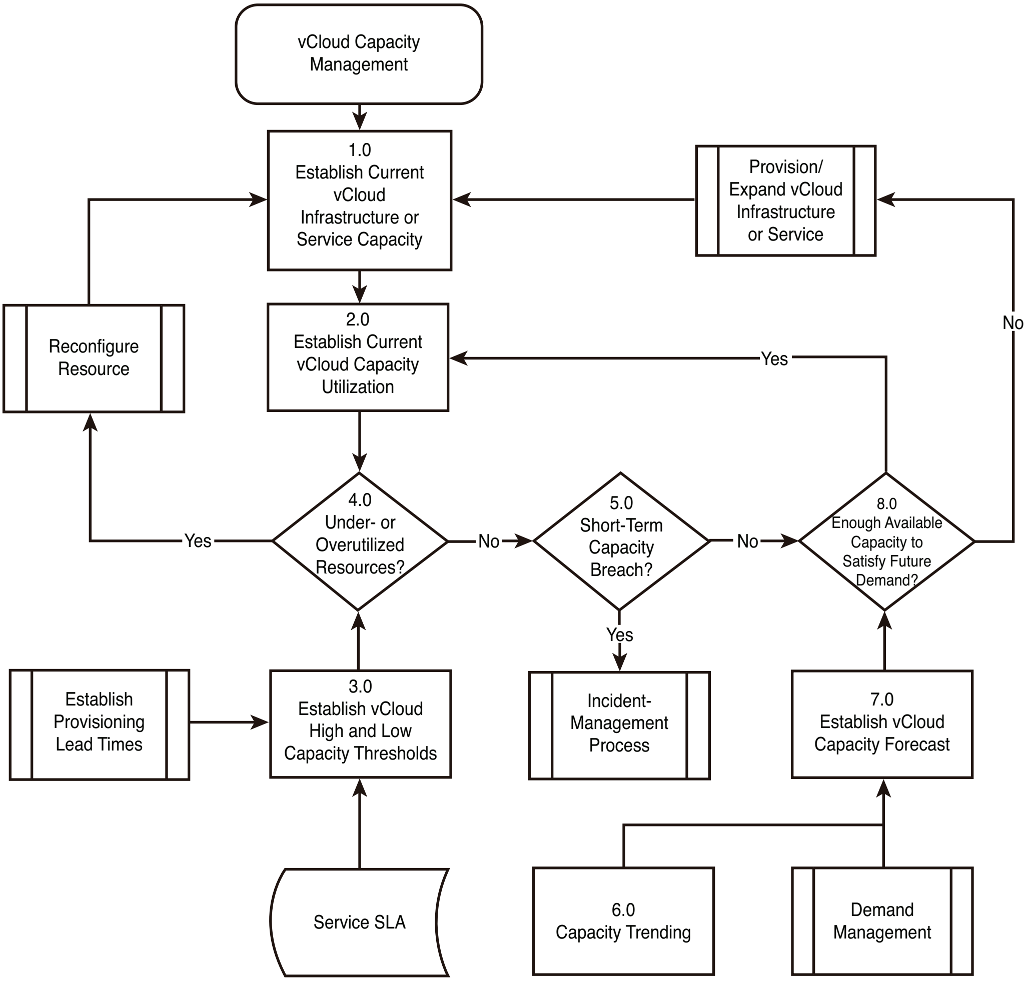

vCloud Capacity Management
1.0 Establish Current vCloud Infrastructure or Service Capacity
Provision/ Expand vCloud Infrastructure or Service
Reconfigure Resource
2.0 Establish Current vCloud Capacity Utilization
Yes
No
4.0 Under- or Overutilized Resources?
Yes
No
5.0 Short-Term Capacity Breach?
No
8.0 Enough Available Capacity to Satisfy Future Demand?
Yes
Establish Provisioning Lead Times
3.0 Establish vCloud High and Low Capacity Thresholds
Incident-Management Process
7.0 Establish vCloud Capacity Forecast
Service SLA
6.0 Capacity Trending
Demand Management

FIGURE 4.17 High-level proactive capacity management process

From a high level, capacity management involves the following:

1. Determining current capacity reserves
2. Forecasting new requirements
3. Planning for additional capacity

Continuous improvement activities are critical to extracting the most value from the vCloud infrastructure. Results can be achieved through simple, periodic planning activities supported by regular capacity augmentation and operational day-to-day activities.

4.8.2.2 Process Evolution for vCloud Capacity-Management Operations

To provide robust capacity management, automate and remove the need for manual intervention wherever possible. Capacity management takes time and effort to evolve, so work on maturing processes in stages instead of trying to do everything in a single step.

Initially, the challenge is to document and maintain capacity-management processes, policies, and methods. Any tools used to assist with vCloud capacity must be carefully selected and suitable for the purpose. All capacity-management roles and responsibilities should be clearly defined.

Over time, vCloud organizations mature and become more vCloud service focused. Tool automation is introduced so that incorrectly sized vCloud components can be easily identified and adjusted with minimal manual interaction. Evaluate automation possibilities to identify other capacity scenarios that can be made more efficient. Specific vCloud KPI metrics should be identified and reported to key stakeholders. Short- and long-term capacity plans should become ingrained within the organization.

To fully integrate capacity management into the vCloud service offering, implement automated capacity-management remediation to stabilize the environment and make sufficient capacity available for services. The COE should be responsible for end-to-end vCloud Capacity Management using highly optimized capacity-management tools and processes.

4.8.2.3 Process Automation and Tool Alignment and Integration

Capacity Management cannot depend on manual processes and activities in a vCloud. Given its ever-changing nature, effective management of vCloud capacity requires an up-to-date view of usage and available capacity of services and infrastructure. Manual processes and most capacity tools cannot provide real-time capacity data.

vCloud providers must provide the capacity for vCloud consumers required to meet the agreed-to SLAs. For the provider to realize ROI, some level of resource sharing is required. Intelligence must be built into Capacity Management tools so that the dynamic usage of the vCloud environment is better understood and any recurring usage behavior is clear. There must be a view of the vCloud customer's environment and virtual datacenters to understand the capacity provisioned, the demand for the resources, and any recurring resource usage behavior.

To provide agile capacity management, it is important that no other process impact the delivery of additional capacity. For example, the change-management process must be closely aligned with the provisioning process so that additional capacity can be rapidly put in place. Capacity provisioning can be at an infrastructure layer (hosts, storage, and vSphere) and at a service layer (new virtual datacenters, additional capacity to existing virtual datacenters). If the change-management process involves lengthy change tickets and CAB attendance, some of the benefits of the vCloud are lost. A lengthy change-management process can delay the introduction of additional capacity into the vCloud, which, in turn, could negatively impact the vCloud consumer's services and associated SLAs.

You must understand the impact of each vCloud Director allocation model on the underlying vSphere infrastructure before effective Capacity Management can be implemented.

Otherwise, it is not be possible to understand how the organization virtual datacenters and virtual machines can use the available capacity.

Each vCloud Director organization virtual datacenter has an underlying vSphere resource pool that supports it and provides the resources to all the virtual machines in the deployed vApps. The configuration of an organization virtual datacenter has a direct relationship to the configuration of the vSphere resource pool and the virtual machines in it. For example, percent guarantees in vCloud Director translate to reservations in the underlying vSphere components. The relationship between vSphere reservations and vCloud guarantees varies depending on the selected vCloud Director allocation model.

Using VMware vCenter Operations Manager™—part of the VMware vCenter Operations Management Suite—makes it possible to understand the complexity in vCloud implementations because the vSphere adapter provides specific vSphere metrics and an analysis of their impact on the environment. The *Risk badge* (see Figure 4.18) in the vCenter Operations vSphere UI provides vSphere Capacity Management functions.

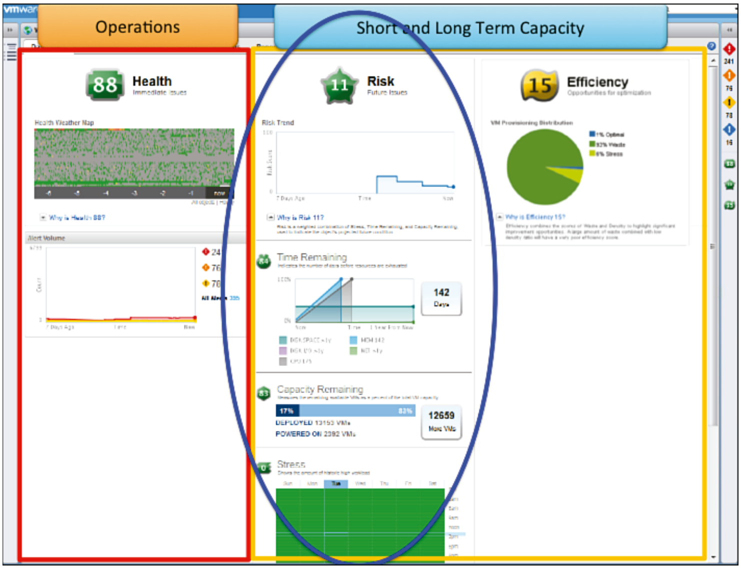


FIGURE 4.18 vCenter Operations Capacity Management in the vSphere UI

In vCenter operations, the analytics functionality analyzes the current and past usage patterns of resources in a vCloud environment, and what-if scenarios help establish future capacity requirements.

The VMware design guideline states that a provider virtual datacenter should be supported by an entire vSphere cluster. Then you can view the capacity information for the provider virtual datacenter in the vCenter Operations vSphere UI by selecting the underlying vSphere cluster.

Because vCenter operations can connect to multiple vCenter instances using the same adapter, you can manage the capacity of multiple resource provider virtual datacenters and the management cluster in a single implementation—provided that the implementation remains within the sizing guidelines for vCenter operations.

4.8.2.4 Roles and Responsibilities for Capacity Management

The vCloud Center of Excellence (COE) model supports capacity management of the vCloud services and the supporting infrastructure. Depending on the size and vCloud maturity of the vCloud organization, the primary capacity management responsibility lies with either the COE analyst or, for smaller organizations, a dedicated Capacity Management individual or team. The primary responsibility is to maintain an accurate and up-to-date capacity-management plan and forecast. Achieve this by granting access to the capacity data and metrics by using appropriate capacity health-monitoring tools such as vCenter Operations.

Automation is essential for capacity management of the vCloud, and the COE analyst, COE developer, and capacity management champion are responsible for making sure that the capacity-management tools and processes for this automation work effectively. Validate effectiveness by auditing the data used in the capacity forecasts and the tools used for the capacity plan. The ultimate goal is to automate as much as possible with minimal administrative interaction.

For more information about vCenter operations, see the latest VMware vCenter Operations Management Suite documentation (www.vmware.com/products/datacenter-virtualization/vcenter-operations-management/technical-resources.html).

4.8.3 Performance Management

This section focuses on addressing vCloud performance issues in support of vCloud service offerings. For a vCloud provider, the goal of performance management is to avoid or quickly resolve performance issues in the vCloud infrastructure and meet the performance requirements for the services offered to consumers. Monitoring is required for the VMware vCloud infrastructure to prevent agreed-upon services levels from being breeched.

Although performance management is performed in the context of normal event, incident, and problem management, it is specifically called out in a vCloud environment because of its importance in persuading potential vCloud users that concerns about additional layers of virtualization and the vCloud have been addressed and that their SLAs will continue to be met. In a traditional physical environment, servers are typically oversized to such a degree on dedicated hardware that performance issues are unlikely to occur. In a shared vCloud environment, users must feel confident that the provided services will meet their needs.

4.8.3.1 Performance Management Process Definition and Components

The high-level event, incident, and problem processes for performance management in Figure 4.19 apply for both vCloud providers and tenants. These processes look the same as any traditional Performance Management process. However, the dynamic nature of the vCloud and the drive to reduce OpEx means that the process has to be more agile and must rely less on manual intervention. Manual performance management might be appropriate for physical infrastructure world and early virtualization adoption stages, but only tooling and automation can provide the level of performance management required for the vCloud.

At a high level, the objectives of event, incident, and problem processes for performance management are to automate as much as possible and maximize the number of tasks that can be performed by Level 1 operators (instead of Level 2 administrators or Level 3 subject matter experts [SME]). The following are possible ways to handle events, incidents, or problems, listed in order of preference:

1. **Automated workflows:** These workflows are totally automatic and can be initiated by predefined events or support personnel.
2. **Interactive workflows:** These workflows require human interaction and can be initiated by predefined events or support personnel.
3. **Level 1 support:** Operators monitor systems for events. They are expected to follow runbook procedures for reacting to events, which might include executing predefined workflows.
4. **Level 2 support:** Administrators with basic technology expertise handle most routine tasks and execute predefined workflows.
5. **Level 3 support:** SMEs for the various technologies handle the most difficult issues and are also responsible for defining the workflows and runbook entries that allow Level 1 operators and Level 2 administrators to handle more events and incidents. Section 4.8.4, "Event, Incident, and Problem Management," describes this in more detail.

4.8.3.1.1 Event Management Process for Performance Management

As Figure 4.19 shows, performance events are generated in multiple ways:

- **vCenter Operations Manger Early Warning Smart Alerts:** These alerts arise when multiple metrics show a change in behavior. Level 2 administrators typically review them to determine whether an incident has occurred.
- **vCenter Operations Manager Key Performance Indicator (KPI) Smart Alerts:** These alerts result from anomalous behavior from predefined KPIs or Super Metrics. Because these alerts are more specific, they are more readily automated with workflows.
- The Service Desk receives a call from a user to report a performance issue.
- The Level 1 operator receives an alert from the monitoring system regarding a performance issue.

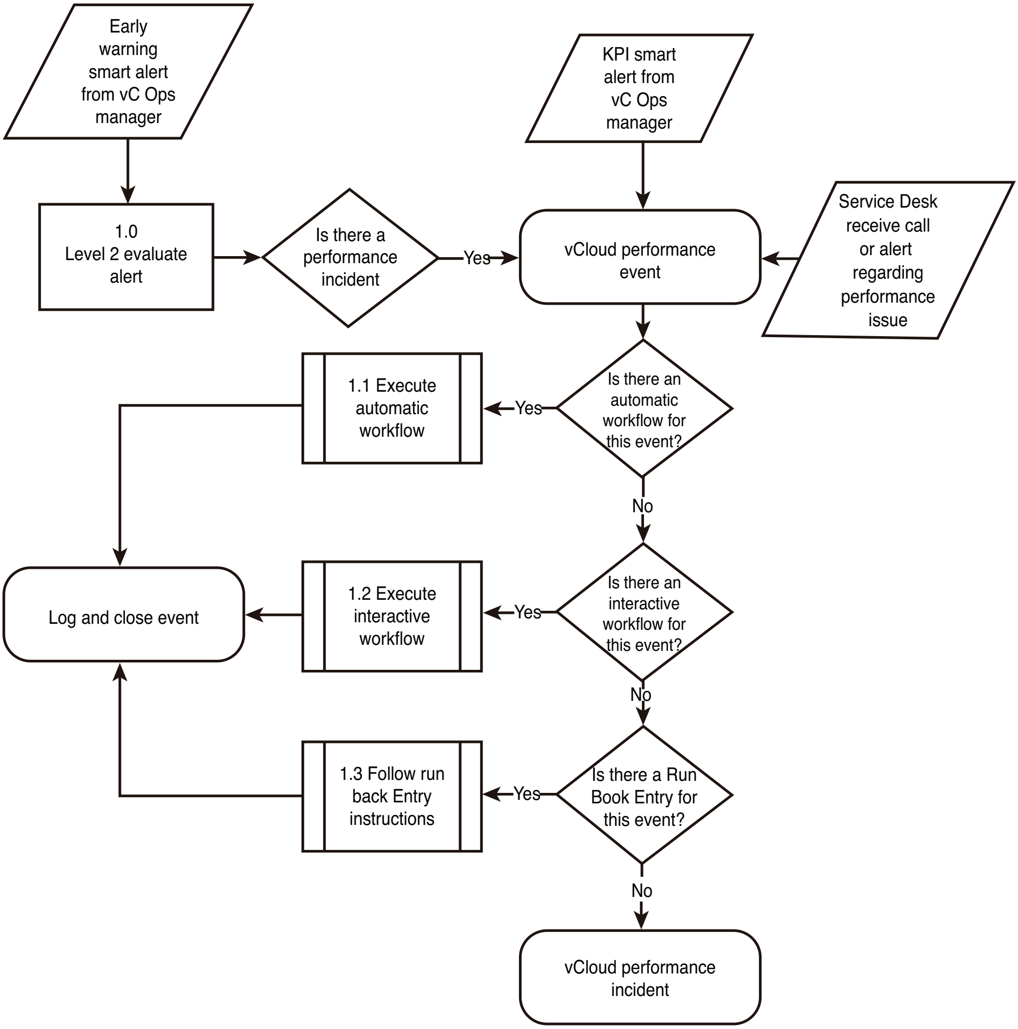

Early warning smart alert from vC Ops manager
KPI smart alert from vC Ops manager
1.0 Level 2 evaluate alert
Is there a performance incident
Yes
vCloud performance event
Service Desk receive call or alert regarding performance issue
1.1 Execute automatic workflow
Yes
Is there an automatic workflow for this event?
No
Log and close event
1.2 Execute interactive workflow
Yes
Is there an interactive workflow for this event?
No
1.3 Follow run back Entry instructions
Yes
Is there a Run Book Entry for this event?
No
vCloud performance incident

FIGURE 4.19 High-level event-management process for performance

If a performance event is identified as a known issue, it might trigger a predefined action such as an automated workflow, interactive workflow, or runbook procedure. If the event does not have a definition, it becomes an incident that a Level 2 administrator or Level 3 SME must resolve.

4.8.3.1.2 Incident Management Process for Performance Management

As Figure 4.20 shows, performance incidents are resolved in different ways, depending on how they are generated.

- **Lack of tenant capacity:** When a tenant's capacity is fully used, events can be triggered depending on how the tenant's lease is defined. If the tenant purchased a *bursting* capability, additional resources can be added at a premium cost if they are in excess of their base usage. If bursting has not been purchased or is not available, the tenant should be notified that their capacity is fully used.
- **Lack of provider capacity:** This should never happen if the design guidance for proactive capacity management is established and effective. If capacity is fully used, the service provider must either add more capacity or move capacity around to address the issue. This condition should be reported to Capacity Management and can result in SLA breaches for tenants.
- **Hardware or software failure:** Performance issues can be the result of software or hardware error such as host failures, configuration errors, bad software updates, or other repairable issues. If insufficient redundancy is built into the overall vCloud, these types of errors can also result in SLA breaches for tenants.

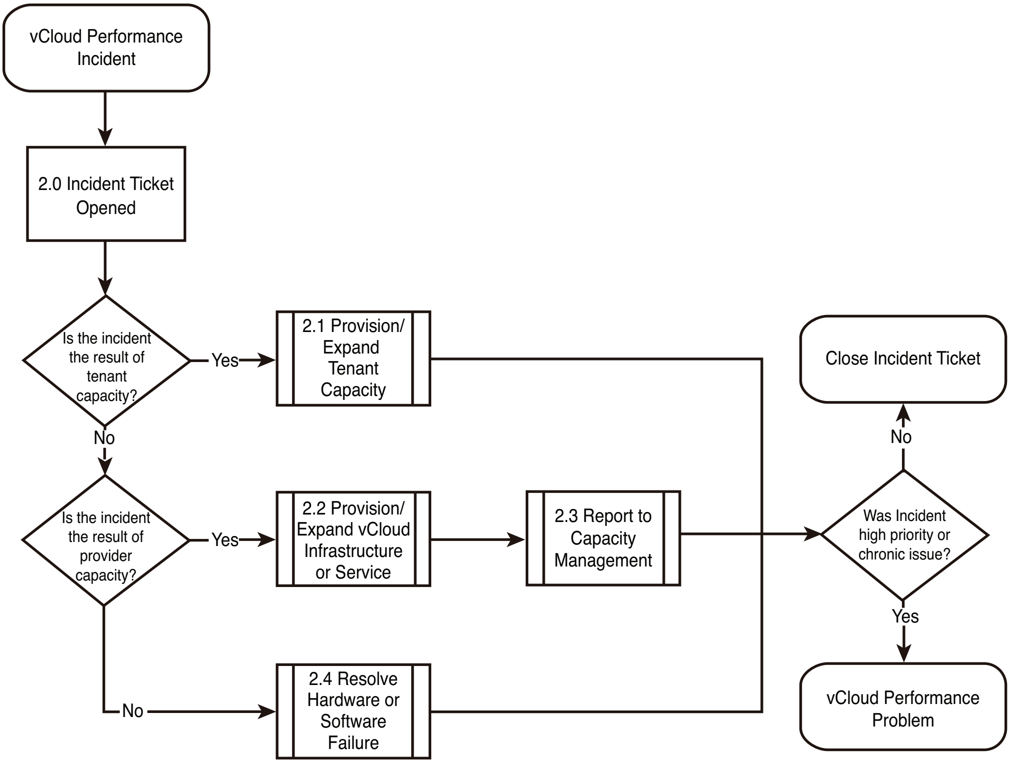

vCloud Performance Incident
2.0 Incident Ticket Opened
Is the incident the result of tenant capacity?
Yes
2.1 Provision/ Expand Tenant Capacity
No
Is the incident the result of provider capacity?
Yes
2.2 Provision/ Expand vCloud Infrastructure or Service
2.3 Report to Capacity Management
No
2.4 Resolve Hardware or Software Failure
Close Incident Ticket
No
Was Incident high priority or chronic issue?
Yes
vCloud Performance Problem

FIGURE 4.20 High-level incident management process for performance

If the incident is high priority or a chronic issue, turn it over to Problem Management for further analysis.

4.8.3.1.3 Problem Management Process for Performance Management

As Figure 4.21 shows, the primary goal of Problem Management is to identify the root cause of a problem. After the root cause is identified, develop and implement an action plan to avoid the problem in the future.

- The preferred method is to fix the root cause so that the problem never occurs again.
- If the problem cannot be eliminated, workflows and runbook entries must be defined so that the problem can be quickly resolved if it occurs again. KPIs and Super Metrics can be defined to help identify an issue before it becomes a problem.

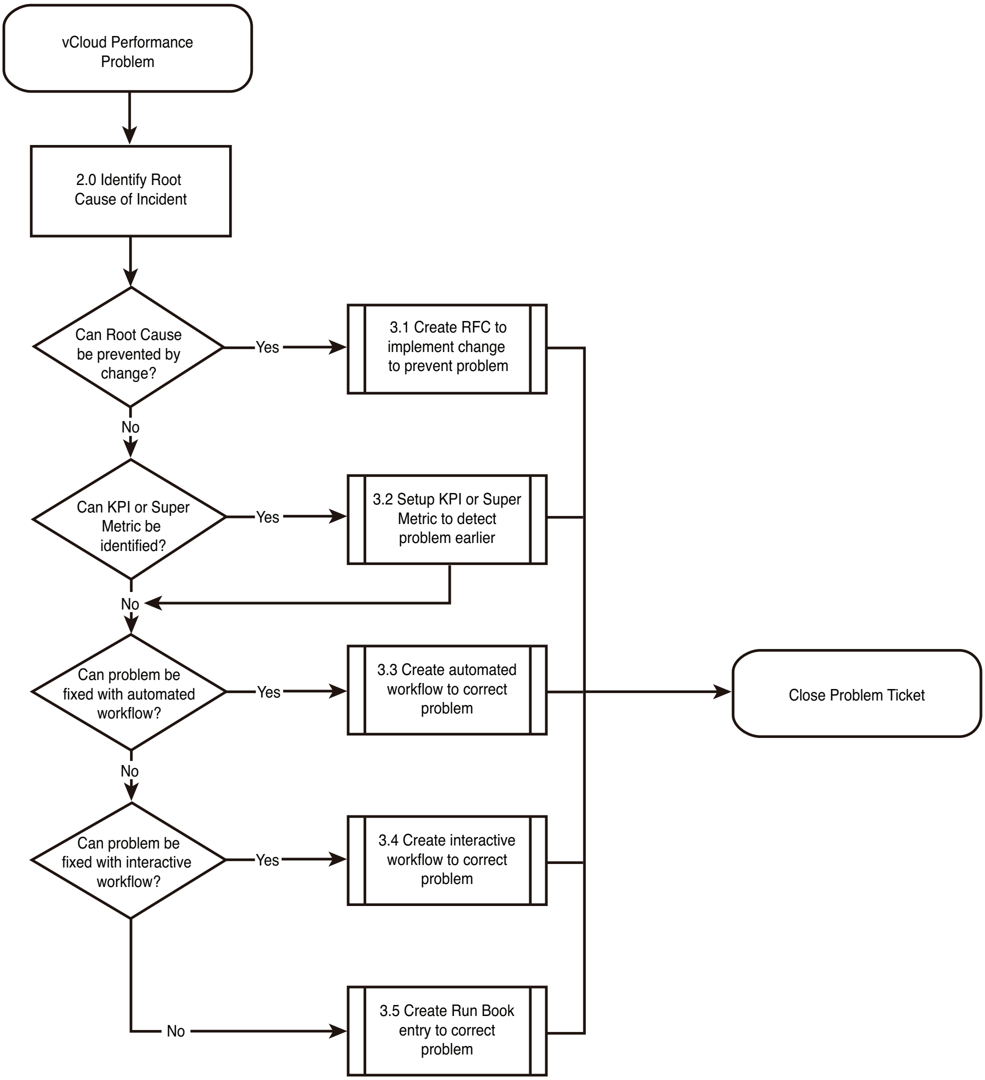

vCloud Performance Problem
2.0 Identify Root Cause of Incident
Can Root Cause be prevented by change?
Yes
3.1 Create RFC to implement change to prevent problem
No
Can KPI or Super Metric be identified?
Yes
3.2 Setup KPI or Super Metric to detect problem earlier
No
Can problem be fixed with automated workflow?
Yes
3.3 Create automated workflow to correct problem
Close Problem Ticket
No
Can problem be fixed with interactive workflow?
Yes
3.4 Create interactive workflow to correct problem
No
3.5 Create Run Book entry to correct problem

FIGURE 4.21 High-level problem-management process for performance

Process Evolution for Cloud Operations

To provide a robust performance-management process, automate and remove the need for manual intervention wherever possible. Evolving the performance-management process takes time and effort. Work on maturing processes in stages instead of trying to do everything in a single step.

Initially, the challenge is to document and maintain the performance-management processes, policies, and methods. Any tools used to assist with vCloud Performance Management must be carefully selected and suitable for the purpose. All performance-management roles and responsibilities should be clearly defined.

Over time, vCloud organizations mature and become more vCloud service focused. Tool automation is introduced so that performance issues can be easily identified and rectified with minimal manual interaction. Evaluate automation possibilities to identify other performance scenarios that can be made more efficient. Better metrics and event coverage are necessary for all aspects of an application, including the capability to collect performance metrics for the following:

- Components (appliances, operating systems, devices)
- Middleware (databases, web servers, Java, messaging)
- Applications
- Virtualization
- vCloud
- Services, including active and/or passive end-user experience monitoring

Specific vCloud KPI metrics should also be identified and reported to key stakeholders.

To fully integrate performance management into the vCloud service offering, implement automated performance remediation to stabilize the environment and provide satisfactory performance for services. The COE is responsible for end-to-end vCloud Performance Management using highly optimized performance-management tools and processes.

4.8.3.2 Process Automation and Tool Alignment/Integration

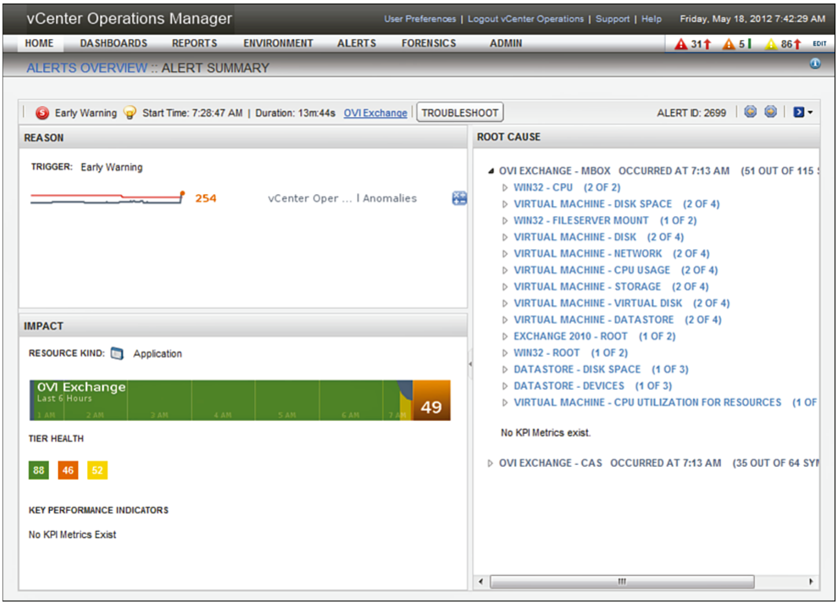
Performance management cannot depend on manual processes and activities in a vCloud. Given its dynamic nature, effective management of vCloud performance requires tooling and instrumentation to be in place. Manual processes and traditional performance tools that focus primarily on up or down status cannot provide the required level of performance data.

For effective performance management, you must understand the impact of *metric coverage*. Having instrumentation at all levels of the application stack enables much better insight into the overall performance of an application. This is particularly true with *end-user experience monitoring*, which provides information to administrators about the consumer experience. Traditionally, administrators have relied on component-level monitoring to approximate a service's availability or performance. This approach provides only partial results and rarely identifies actual performance problems.

To solve this problem, an analytics tool is needed to analyze more than just the up or down status of traditional monitoring tools. An analytics tool enables an administrator to see the relative performance of a system based on dynamically generated baselines. By using VMware vCenter Operations Manager (part of the VMware vCenter Operation Management Suite), this level of detail in vCloud implementations is understood and can be instrumental in revealing more complex performance-management issues.

4.8.3.3.1 Event Management

A key feature of vCenter Operations Manager is the capability to establish dynamic baselines on millions of metrics within an organization's environment. These baselines also take into account time of day, day of week, and other cyclical patterns to understand normal behavior. The baselines are then used to determine early warning smart alerts if too many metrics start behaving abnormally at the same time. If KPIs or Super Metrics have been defined to capture known problem areas, KPI smart alerts that have associated automated or interactive workflows can be triggered. Figure 4.22 shows the custom user interface used for vCenter Operations Manager event management.



Friday, May 18, 2012 7:42:29 AM

Early Warning | Start Time: 7:28:47 AM | Duration: 13m:44s OVI Exchange

ALERT ID: 2699

TRIGGER: Early Warning

254 vCenter Oper ... l Anomalies

RESOURCE KIND: Application

No KPI Metrics Exist

OVI EXCHANGE - MBOX OCCURRED AT 7:13 AM (51 OUT OF 115

- WIN32 - CPU (2 OF 2)
- VIRTUAL MACHINE - DISK SPACE (2 OF 4)
- WIN32 - FILESERVER MOUNT (1 OF 2)
- VIRTUAL MACHINE - DISK (2 OF 4)
- VIRTUAL MACHINE - NETWORK (2 OF 4)
- VIRTUAL MACHINE - CPU USAGE (2 OF 4)
- VIRTUAL MACHINE - STORAGE (2 OF 4)
- VIRTUAL MACHINE - VIRTUAL DISK (2 OF 4)
- VIRTUAL MACHINE - DATASTORE (2 OF 4)
- EXCHANGE 2010 - ROOT (1 OF 2)
- WIN32 - ROOT (1 OF 2)
- DATASTORE - DISK SPACE (1 OF 3)
- DATASTORE - DEVICES (1 OF 3)
- VIRTUAL MACHINE - CPU UTILIZATION FOR RESOURCES (1 OF

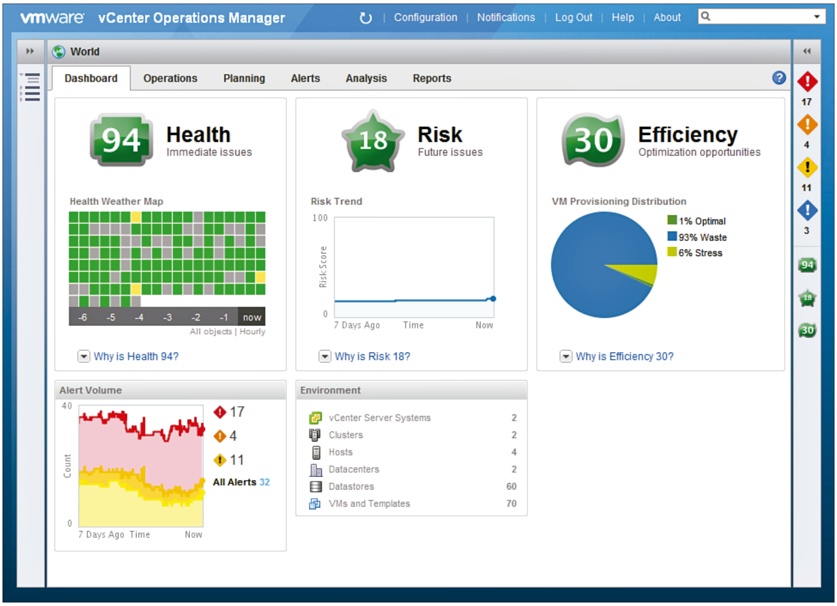
No KPI Metrics exist.

OVI EXCHANGE - CAS OCCURRED AT 7:13 AM (35 OUT OF 64 SYI

FIGURE 4.22 vCenter Operations Manager event management within the custom UI

4.8.3.3.2 Incident Management

When a performance incident is identified, an administrator can use vCenter Operations Manager to locate the responsible underlying system. The Health badge can provide insight into performance-management incidents (see Figure 4.23).



Environment

vCenter Server Systems	2
Clusters	2
Hosts	4
Datacenters	2
Datastores	60
VMs and Templates	70

FIGURE 4.23 vCenter Operations Manager performance management in the vSphere UI

The vCenter Operations Manager analytics capability analyzes the current and past usage patterns of resources in a vCloud environment and provides users with both a high-level and detailed view of the health of their environment.

4.8.3.3.3 Problem Management

After an incident is resolved, an administrator can use vCenter Operations Manager to identify the responsible system and the root cause of the issue. Examining the underlying system that was responsible for a performance issue can expose the relationship to other tiers within an application, any smart alerts that are associated with it, and the performance history of affected components. This process can help identify the root cause of the issue.

4.8.3.4 Roles and Responsibilities for Performance Management

The vCloud Center of Excellence (COE) model supports performance management for vCloud services and the supporting infrastructure. Depending on the size and maturity of the vCloud organization, the primary performance management responsibility lies with either COE analyst or administrator for smaller organizations, or with a dedicated performance management individual or team within the COE. The primary responsibility is to address performance issues and quickly mitigate them when they arise. This is achieved by granting access to the performance data and metrics by means of appropriate performance health–monitoring tools such as vCenter Operations Manager.

Automation and instrumentation are essential for vCloud Performance Management, and the COE analyst, COE developer, and performance management champion must be responsible for making sure that the performance-management tools and processes for this automation are working effectively. Validation should be conducted by auditing the data used in the performance forecasts and the tools used for the performance plan. The goal is to automate this process as much as possible, with minimal administrative interaction.

For information on vCenter Operations, see the latest VMware vCenter Operations Management Suite documentation (www.vmware.com/products/datacenter-virtualization/vcenter-operations-management/technical-resources.html).

4.8.4 Event, Incident, and Problem Management

Traditionally, *Event, Incident, and Problem Management* focused on monitoring the services offered from the vCloud and on minimizing impact from unplanned events. Restoring service as rapidly as possible and preventing repeat events from affecting services were also core functions. Today there is an increased emphasis on reducing vCloud OpEx cost and increasing reliability. This can be achieved by increasing automation, allowing operators to handle more routine tasks, and proactively detecting and eliminating incidents before they impact end users.

Event Management focuses on how to categorize and handle outputs from monitoring and analytics tools. Based on predefined rules, inputs to event management are called *events*. They can be associated with a variety of possible actions, ranging from suppression, to triggering an automatic workflow, to triggering an incident to be created in the case of a performance incident or an actual outage.

Incident Management focuses on how to handle performance incidents or outages. Such occurrences are referred to as *incidents*. The primary focus of incident management is to manage the incident until it is resolved. Recurring incidents or incidents that are high priority can be referred to Problem Management for further investigation.

Problem Management focuses on identifying root causes for recurring and high-priority incidents. After a root cause has been identified, a plan of action is generated that, ideally, repairs the underlying problem. If the problem cannot be fixed, additional monitoring and event management handling might be implemented to minimize or eliminate future occurrences of the problem.

One of the main benefits of implementing a vCloud environment is to lower ongoing OpEx costs. A key to realizing this goal is vCloud Event, Incident, and Problem Management process automation that consists of the following:

- Automating responses to events when possible
- Creating highly automated workflows to other events where some operator input is required as part of decision support
- Creating runbook entries, workflows, and automations so that operators (instead of administrators or subject matter experts) can handle many more events
- Automating interaction between the vCloud Event, Incident, and Problem Management process and other required processes and associated systems
- Identifying, instrumenting, and developing key performance indicators (KPIs) that can develop workflows and automations

4.8.4.1 Event, Incident, and Problem Management Process Definition and Component

The following must be in place for successful vCloud event, incident, and problem management:

- Monitoring of the vCloud environment
- An event-management system, such as a Manager of Manager (MoM), for applying rules to events that can launch workflows or route events to the appropriate support teams
- A ticketing system and methodology so that various support teams are allocated tickets in an efficient manner
- Defined incident priorities and severities
- Well-understood roles and responsibilities
- The capability to view KPI status

Figure 4.24 shows the overall Event, Incident, and Problem Management process and the interrelationship among the components. All three subject areas are shown together because they are intrinsically linked. Event Management feeds into Incident Management, which feeds into Problem Management. Problem Management then feeds back into Event Management to complete the cycle. Because IT is ever evolving and changing, Event, Incident, and Problem Management must be continually updated to keep pace.

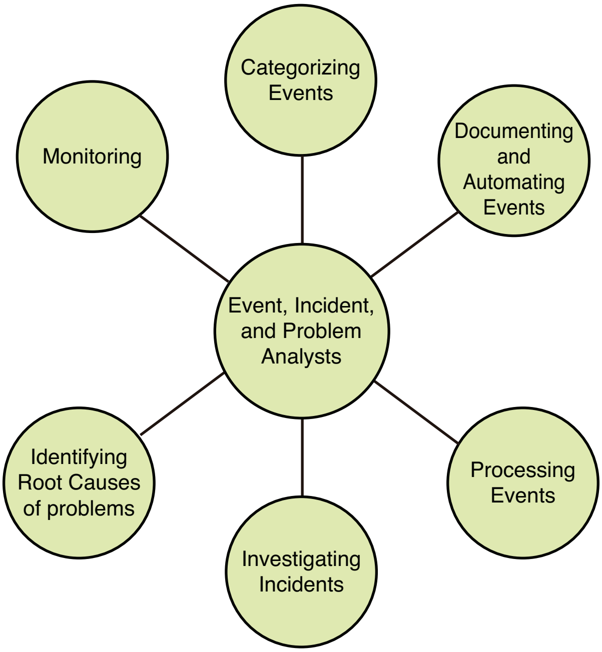

Categorizing Events
Monitoring
Documenting and Automating Events
Event, Incident, and Problem Analysts
Identifying Root Causes of problems
Processing Events
Investigating Incidents

FIGURE 4.24 High-level Event, Incident, and Problem Management processes

One of the first steps in Event Management is to monitor components and services. Events can then be fed into an Event Management system, such as a MoM, and metrics can be fed into an analytics engine, such as vCenter Operations Manager, for processing.

A key component of Event Management is event categorization. After an event is categorized, rules and documentation such as runbooks and workflows can be developed to handle the event the next time it occurs. This proactive approach leads to fewer new incidents and reduces the duration and severity of the outages and performance incidents that do occur.

Core process areas of Incident Management include managing support tickets by determining priority and impact, handling customer communications, facilitating technical and management communication (including phone bridges), and closing out tickets.

When an incident is recurring or high priority, it is sent to Problem Management to identify the root cause. After a root cause is identified, a solution is developed either to fix the problem or to establish monitoring or event handling to eliminate the problem or reduce the severity the next time the problem occurs.

4.8.4.2 Process Evolution for vCloud Operations

To provide a robust event, incident, and problem management process, automate and remove the need for manual intervention wherever possible. Evolving the process takes time and effort—work on maturing processes in stages instead of trying to do everything in a single step.

Initially, the challenge is to document and maintain the performance-management processes, policies, and methods. Any tools used to assist with vCloud Event, Incident, and Problem Management must be carefully selected and suitable for the purpose. All Event, Incident, and Problem Management roles and responsibilities should be clearly defined.

Over time, vCloud organizations mature and become more vCloud service focused. As a result, automated responses and analytics are necessary to help vCloud providers provide the required levels of service. As the analytics engine better understands the vCloud environment, rapid identification of events that could become incidents enables fixes to be put in place before services are affected. Initially, the fixes are manual, but with maturing processes in place, tool automation can be introduced so that future incidents can be easily identified and rectified with minimal manual interaction. Automation possibilities must be evaluated to identify other event, incident, and problem scenarios that can be made more efficient. Specific cloud KPI metrics should be identified and reported to key stakeholders.

4.8.4.3 Process Automation and Tool Alignment/Integration

The vCloud Event, Incident, and Problem Management processes depend on tooling. If the appropriate tools are not in place, it is difficult to manage and operate the environment while sustaining the required service levels. Traditionally, event, incident, and problem management has relied heavily on tooling; in a vCloud, the scope of the required tools increases. This is the result of additional vCloud requirements, such as a greater need for early warning for impending incidents and a higher level of automation. For early warnings, increased functionality of the tools (for example, smart alerts, dynamic thresholds, and intelligent analytics) helps fulfill this requirement. For a higher level of automation, additional tools, such as vCenter Orchestrator, are required.

To realize the vCloud benefits of reliability and lower OpEx costs, it is not sufficient to merely interpret events to highlight incidents and problems. It is also necessary to establish how incidents can be more efficiently identified, how remediation can be put in place quickly, and how to identify the root cause to prevent the problem from happening again.

Because the vCloud resources and services supplied to vCloud customers are based on underlying vSphere resources, it is possible to use tools that manage and monitor at the vSphere level.

As Figure 4.25 shows, vCenter Operations Manager can provide an up-to-date understanding of the health of the vSphere environment as it relates to the vCloud provider virtual datacenters.

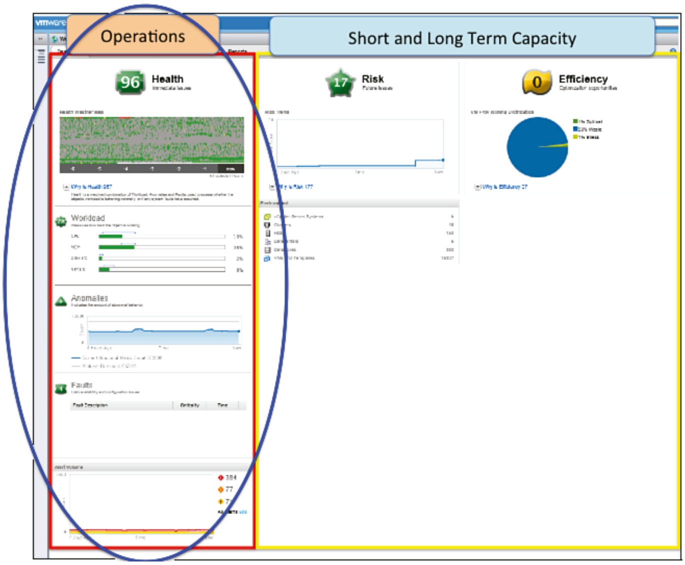


FIGURE 4.25 vCenter Operations Manager Event and Incident Management

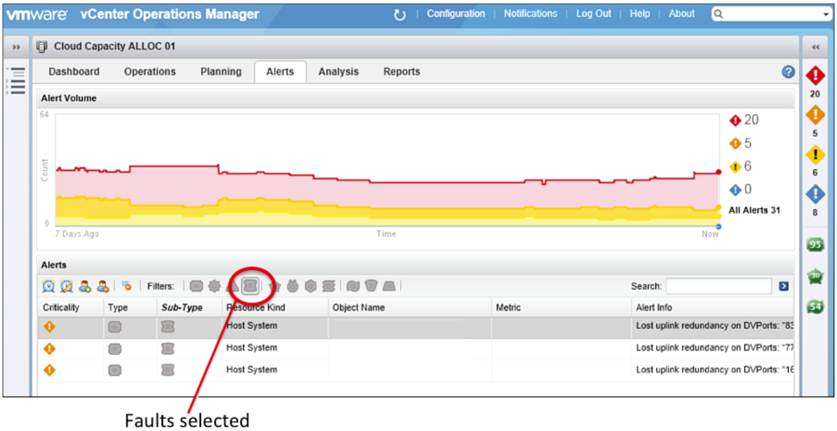
The Health badge shows a score that indicates the overall health of the selected object. The object can be a vCenter instance, vSphere datacenter, cluster, host, or datastore. The monitoring mechanism provides proactive analysis of the performance of the environment and determines when the health of the object reaches a level that indicates an incident might be about to occur. To enforce effective management, the vCloud NOC can be provided with a dashboard that shows key metrics that indicate the health of the environment.

The score shown for the Health badge is calculated from the following sub-badges:

- **Workload:** Provides a view of how hard the selected object is working
- **Anomalies:** Provides an understanding of metrics that are outside their expected range
- **Faults:** Provides detail on any infrastructure events that might impact the selected objects availability

For faults, active vCenter events or alerts are used. These can include host hardware events, virtual machine FT and HA issues, vCenter health issues, cluster HA issues, and so on. The vCenter alerts are supplied through the vSphere adapter into vCenter Operations Manager and can identify root cause. Additionally, vCenter Operations Manager generates alerts if a sub-badge score hits a predefined value.

The events or alerts appear as faults, as Figure 4.26 shows.

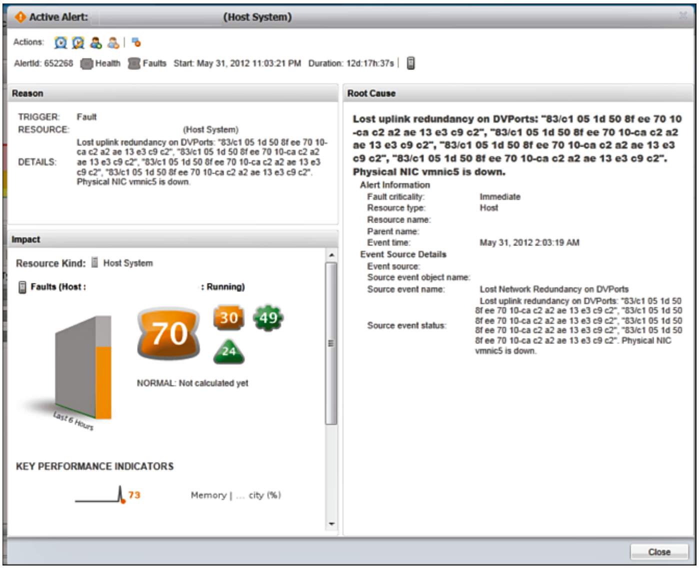


Criticality	Type	Sub-Type	Resource Kind	Object Name	Metric	Alert Info
◆			Host System			Lost uplink redundancy on DVPorts: "83
◆			Host System			Lost uplink redundancy on DVPorts: "77
◆			Host System			Lost uplink redundancy on DVPorts: "16

Faults selected

FIGURE 4.26 vCenter Operations Manager faults

Any fault can be selected to gain further information. In Figure 4.27, the event is associated with a host and indicates that an uplink has been lost.



AlertId: 652268 Start: May 31, 2012 11:03:21 PM Duration: 12d:17h:37s

TRIGGER: Fault
RESOURCE: (Host System)
DETAILS: Lost uplink redundancy on DVPorts: "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2". Physical NIC vmnic5 is down.

Resource Kind: Host System
Faults (Host : : Running)
NORMAL: Not calculated yet

Lost uplink redundancy on DVPorts: "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2". Physical NIC vmnic5 is down.

Alert Information
Fault criticality: Immediate
Resource type: Host
Resource name:
Parent name:
Event time: May 31, 2012 2:03:19 AM
Event Source Details
Event source:
Source event object name:
Source event name: Lost Network Redundancy on DVPorts
Source event status: Lost uplink redundancy on DVPorts: "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2", "83/c1 05 1d 50 8f ee 70 10-ca c2 a2 ae 13 e3 c9 c2". Physical NIC vmnic5 is down.

FIGURE 4.27 vCenter Operations alert

In addition to using vCenter Operations Manager for vSphere metrics and events, VMware vFabric™ Hyperic® can provide operating system and application metrics. Providing these metrics to vCenter Operations Manager further enhances the incident management toolset.

4.8.4.4 Roles and Responsibilities for Event, Incident, and Problem Management

The vCloud Center of Excellence (COE) model supports the Event, Incident, and Problem Management of the vCloud services and the supporting infrastructure. Depending on the size and vCloud maturity of the vCloud organization, the model for managing events, incidents, and problems is based on several levels for larger organizations. Each of these levels has an escalation path to the next level, until SMEs are required to help resolve incidents or problems.

1. Initial responsibility for any incident lies with the Level 1 Service Desk or operations center, such as a NOC. There, the intention is to resolve as many incidents as possible. KPIs are used for measurement.

2. Typically, a NOC with a general level of vCloud knowledge and skill provides Level 2 support.

3. Level 3 support comes from the vCloud COE subject matter experts (SMEs), as well as other technology specialists that provide resources and knowledge of the vCloud environment, such as network, storage, and security. Refer to Section 4.5.2.1, “vCloud Infrastructure Operations Center of Excellence,” for more information about the COE.

The COE analyst works with the Event Management analyst so that event routing rules, runbook entries, and workflows that define event handling are well defined and accurate. They implement additional monitoring for events that indicate an incident has occurred and define event routing rules, runbook entries, and/or workflows to handle known events. They need to understand the monitoring implications of new Event Management rules, automations, and workflows. They also need to provide requirements for automation and workflow implementation, modification, maintenance, and integration with other systems. In addition, they work to categorize events for promotion to a workflow or support queue.

Specific to Incident Management, the COE analyst and administrator work with the Incident Management analyst so that event routing rules, runbook entries, and workflows defining event handling are well defined and accurate. They identify recurring or high-priority incidents that Problem Management needs to examine for root cause analysis. They also work with the infrastructure and application teams to categorize, manage, and resolve incidents, and they work with the service desk to communicate status of incidents.

The COE analyst works with the Problem Management analyst to identify recurring or high-priority problems that need root cause analysis and assist in identifying the root cause. They implement monitoring, event routing rules, runbook entries, and/or workflows to handle problem events. They also develop a plan to address the root cause of a problem, which might include a permanent solution or might require a workaround that is coordinated with Event Management.

For information about vCenter Operations Manager, refer to the latest VMware vCenter™ Operations Management Suite documentation (www.vmware.com/products/datacenter-virtualization/vcenter-operations-management/technical-resources.html).

For information about VMware vFabric Hyperic, see the latest product documentation (http://support.hyperic.com/display/DOC/HQ+Documentation).

4.8.5 Configuration and Compliance Management

vCloud differs from traditional virtualization in its increasing reliance on automation, increased scale, and dynamic workload management. It is the equivalent of moving from a handcrafted workshop to a fully automated assembly line with the benefits of speed, reliability, and volume. To realize this goal, all the components that constitute the vCloud must be interchangeable and secure. This can be achieved through *Configuration and Compliance Management*.

Configuration Management focuses on defining and maintaining information and relationships about a vCloud and its components and services. This might involve a Configuration Management Database (CMDB) to store data centrally or a Configuration Management System (CMS) to federate data across multiple repositories. Another aspect of configuration is to maintain a record of the single source of truth for each piece of data and coordinate the exchange of data with external systems.

In contrast with Configuration Management, *Compliance Management* focuses more on maintaining corporate vCloud provider or tenant standards for systems that might include compliance standards such as PCI, SOX, or HIPPA. In addition to security settings and firmware, software, and patch levels, Compliance Management is concerned with change management, user access, and network security.

Together, Configuration and Compliance Management validate that configuration settings, firmware, software, and patch versions all follow predetermined standards and policies set by the controlling organization, which can be the vCloud provider, the tenant, or the subtenants.

A major goal of implementing a vCloud is to lower ongoing OpEx costs. To realize this goal, promote and maintain standardization of as many components as possible while maintaining a high level of security and compliance. The following practices are necessary to realize maximum OpEx savings:

- Automated provisioning of interchangeable components that meet vCloud provider or tenant standards and compliance policies
- Ongoing validation that standards and compliance policies are maintained over time
- Ongoing validation that the underlying vCloud infrastructure meets standards and compliance policies (*trusted cloud*)
- Ongoing reporting of noncompliant systems
- Ongoing remediation of noncompliant systems
- Tracking and propagating relationships between components to enhance impact analysis and troubleshooting of the vCloud
- Work with existing CMDB, CMS, or other vCloud provider or tenant data sources to understand where the sources of truth are for exchanging data with the rest of the organization

4.8.5.1 Configuration and Compliance Management Process Definition and Components

For effective configuration and compliance management, the following must be in place:

- Configuration and compliance tools to capture the current state of the vCloud environment
- Automation and workflow tools to detect, report, and remediate noncompliant systems
- A CMDB, CMS, or other corporate data schemas to identify where the single sources of truth exist within a vCloud provider or tenant organization
- Defined vCloud provider or tenant standards and compliance policies
- Defined vCloud provider or tenant Change Management policies for compliance remediation
- Defined vCloud provider or tenant access policies for user access and level of rights
- Defined vCloud provider or tenant network security policies

- Well-understood roles and responsibilities
- Capability to capture, record, and view KPI statistics

Figure 4.28 shows a high-level view of the Configuration and Compliance Management process.

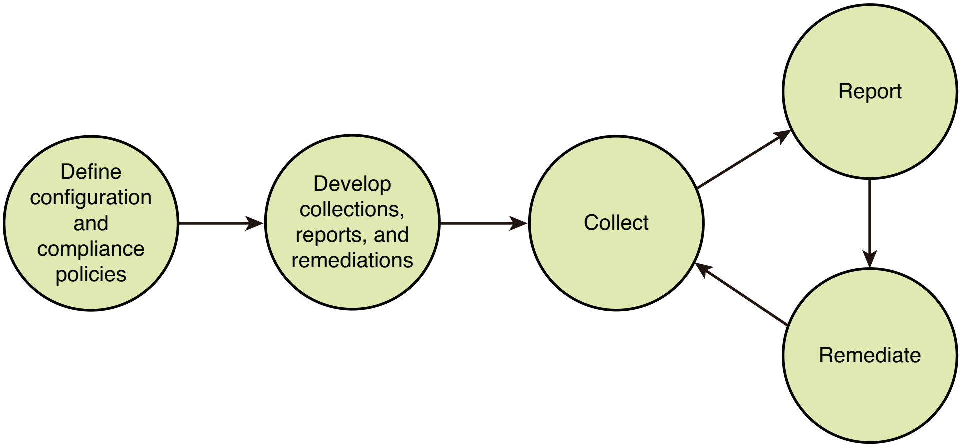

Define configuration and compliance policies
Develop collections, reports, and remediations
Collect
Report
Remediate

FIGURE 4.28 High-level Configuration and Compliance Management process

The process involves the following steps:

1. Define the standards and compliance policies. This is an ongoing process that must be updated as new components are developed and compliance policies evolve. Goals must be established for level of compliance and time to remediate.
2. Develop content for the following areas:
    - Collections to validate compliance
    - Reports to show levels of compliance
    - Automations and runbook entries to remediate noncompliance
3. As part of a regular cycle, gather information about the following:
    - Configuration settings for standardization and hardening
    - Firmware, software, and patch levels
    - Status and completeness of change records, especially for systems subject to compliance regulations
    - User access records such as rights allowed, logins, failed logins, commands used, and others
    - Network access records such as firewall rules, denied access, and so on

4. Evaluate the results and generate reports that show the level of compliance for each area.
5. Remediate if noncompliance is detected. Depending on the type of noncompliance and any impacted service levels, different levels of urgency might apply.

4.8.5.2 Process Evolution for vCloud Operations

To provide a robust Configuration and Compliance Management process, automate and remove the need for manual intervention wherever possible. People, process, and tools must be in place to support the overall process. Evolving the Configuration and Compliance Management process takes time and effort—work on maturing processes in stages instead of taking on the challenge as a whole in a single step.

Initially, the challenge is to define, document, maintain the following.

- **People:** All roles, responsibilities, and necessary skill sets
- **Processes:** Interactions with other processes, as well as other personnel
- **Tools:** Functionality required

Over time, vCloud organizations mature and become more vCloud service focused. As a result, automated collections, reports, and remediation are necessary to help vCloud providers meet the required levels of standardization and compliance. These efforts are initially manual, but as processes mature, tool automation can be introduced and expanded so that future standards and compliance policies can be implemented with minimal manual interaction. Automation possibilities must be evaluated to identify other configuration and compliance scenarios that can be made more efficient.

Configuration and Compliance Management processes should also include collection and reporting of specific vCloud KPIs to key stakeholders showing the overall state of the environment. Examples might include percent of noncompliant configuration items or services, time to remediate noncompliant systems, or percent of services made compliant through automated remediation.

4.8.5.3 Process Automation and Tool Alignment/Integration

The configuration and compliance processes for vCloud depend on tooling. The appropriate tools must be in place to effectively manage and operate the environment while sustaining the required service levels. Traditionally, Configuration and Compliance Management has been mostly manual, with few tools used. In a vCloud, additional tools are required due to additional requirements, such as a greater need for standardization and compliance, and a higher level of automation.

The following products are available to assist with process automation:

- **vCloud Director:** As the core of the vCloud, this is the single source of truth for all the vCloud components. vCloud Director manages all the vCloud relationships, including provider virtual datacenters, organization virtual datacenters, and vCloud networks and storage.

- **vSphere:** While vCloud Director provides a level of abstraction from the vSphere virtualization layer, vSphere provides the single source of truth for configuration and relationship information about the virtualization components that support the vCloud, such as hosts, virtual switches, and datastores. vSphere configuration information is usually not referred to directly for configuration and compliance management, but is used in other tools.
- **VMware vCenter Configuration Manager™:** Collects and validates configuration, software, and patch information for the vCloud infrastructure and the vCloud service components. It also remediates configuration settings and software and patch levels.
- **VMware vCenter Infrastructure Navigator™:** Collects and stores relationships between the virtual machines that make up and interact with an application or service.
- **VMware vCloud Networking and Security Manager™, VMware vCloud Networking and Security App™, and VMware vCloud Networking and Security Edge™:** Manages vCloud network policies, configurations, and settings.
- **vCenter Orchestrator:** Collects information, generates reports, and remediates issues through automated workflows. vCenter Orchestrator is the preferred method for interfacing with systems outside the VMware ecosystem.

For more information about these tools, see the latest documentation at www.vmware.com/products.

This suite of products is required to varying degrees, depending on whether configuration and compliance is from a provider or tenant perspective.

Tenants have visibility of all components in their domain but might not have visibility into components that make up a service that has been provided to them. For example, a public vCloud tenant will probably not have a view into the vSphere virtual infrastructure within the provider's environment. For this example, the scope of configuration and compliance management is limited to the virtual datacenter instance.

A vCloud provider will probably not have any view inside the components that it has provided to a tenant. This also applies to tenants who provide services to subtenants. For example, a Value Added Reseller (VAR) who buys an organizational virtual datacenter from a vCloud provider would not have visibility into the virtual machines that it resells to its customers.

A provider offers a vCloud service with infrastructure that might meet a certain level of compliance (for example, PCI or SOX), which would be reflected in the service level offered to its tenants. It is the provider's responsibility to make sure that this service level is adhered to and that all the components remain compliant (possibly including services consumed from other providers). It is each tenant's responsibility to make sure that the infrastructure and services built on top also adhere to the same compliance level.

4.8.5.4 Roles and Responsibilities for Configuration and Compliance Management

The COE model supports Configuration and Compliance Management for vCloud services and the supporting infrastructure. Refer to Section 4.5.2.1, "vCloud Infrastructure Operations Center of Excellence," for more information about the COE.

Depending on the size and maturity of a vCloud provider or tenant organization, the staffing levels for the roles described in this section can range from a single individual in a smaller organization who performs multiple roles, up to a team that performs a single role in a large organization.

In the vCloud environment, the COE analyst role (for vCloud providers) and the vCloud Service Analyst role (for vCloud tenants) are responsible for overseeing the running of the following core Configuration and Compliance Management processes: Responsibilities include the following:

- Defining vCloud configuration and compliance standards
- Developing collections to validate compliance
- Developing reports showing compliance levels
- Developing remediation
- Collecting and reporting Configuration and Compliance Management KPIs
- Coordinating integration with CMDB, CMS, or other data sources
- Overseeing collections
- Producing reports of compliance
- Coordinating remediation efforts
- Assisting in developing automated compliance policies

When required, the enterprise Configuration and Compliance Management analyst role works with the COE analyst or service analyst on the following tasks:

- Reviewing configuration and compliance standards and policies
- Assisting with the development of the collections, reports, and remediation

4.8.6 Orchestration Management

Orchestration Management is responsible for gathering and understanding service orchestration workflow requirements; managing their development, testing, and release; and interacting with the COE to integrate infrastructure-related automation workflows.

4.8.6.1 Orchestration Management Definition

Orchestration Management is the process responsible for governance and control over orchestration workflows and the resulting automation within the vCloud. The goal of Orchestration Management is to understand the impact of orchestration workflows on an organization's vCloud, on those who approve or benefit from the orchestration, and on the interrelations between orchestration and traditional IT service management processes.

4.8.6.2 Value of Orchestration Management in a vCloud

Orchestration capabilities contribute greatly to making a vCloud dynamic and to vCloud agility, elasticity, and self-healing properties.

Along with the benefits, elasticity raises some risks. A successful vCloud implementation must focus on delivering consistent quality of services. Orchestration Management adds the layer of control required to achieve consistency in a vCloud. Control also includes the capability to protect and secure the vCloud. Unwarranted actions in a vCloud cannot be tolerated, so orchestration workflows and actions must be tightly controlled.

The following sections provide information about how to control orchestration in a vCloud. Orchestration is a relatively new feature, and as organizations mature in their management of vCloud environments, the role of Orchestration Management becomes more important.

4.8.6.2.1 Orchestration Workflow Creation Control in a vCloud

Before implementing orchestration workflows in a vCloud environment, answer the following questions:

- Who approved the orchestration workflow?
- Why is it needed?
- What impact does the orchestration workflow have on the vCloud environment?
- Who needs to be informed when the workflow is executed?

Answer these questions for all orchestration workflows that are built into the vCloud. VMware recommends that the following teams be involved during development of orchestration workflows:

- The Orchestration Management team, which focuses on business requirements gathering and business unit negotiations
- The COE team, which focuses on technical development of workflows to facilitate the implementation of consistent standards across all orchestration workflows in the organization

Development of orchestration workflows is complex. Orchestration engages with multiple internal and external systems in a vCloud environment, so a complete development lifecycle must be followed with dedicated support from the application and business teams.

Appropriate testing should be completed at every stage of development, including unit, system, and integration testing, before moving orchestration workflows into production. As part of development testing, operational testing that includes performance and scalability scenarios for end-to-end automation processes must also be completed. In many cases, orchestration workflows themselves might be able to withstand new loads, but external or downstream systems might experience a performance impact. A clear roll-back procedure must be established for exceptions to protect against impacting production functions.

4.8.6.2.2 Orchestration Workflow Execution Control in a vCloud

A vCloud is a dynamic environment where continuous changes are made to improve the quality of the services that run on it. Orchestration plays a key part in this agility, allowing automated actions to be performed as required by vCloud. Orchestration Management focuses on vCloud impacts and maintains flexibility in the environment. VMware recommends control for the execution of orchestration workflows developed for vCloud, with error handling built into the workflows. If workflow execution issues arise, notifications need to be sent to the operations team, with appropriate escalations and tiering for alerts.

4.8.6.2.3 Orchestration Management in Relation to Change Management

As orchestration matures, complex manual tasks are automated. Prior to implementation, workflows that will lead to changes in business services that directly impact users must be analyzed in detail. The Change Advisory Board (CAB) needs to preapprove actions on production applications. Additional controls might also be set to allow for notification back to the CAB upon execution of critical business that impacts orchestration workflows. This must be done in accordance with an organization's change control policies. Business impact should be the main driver for discussion between the orchestration team and CAB. The CAB should allow more flexibility to simple orchestration actions that impact vCloud internal background operations (for example, capacity-related actions) but that do not directly impact a business application or service and might not need require approval for them.

4.8.6.2.4 Orchestration Management in Relation to Configuration Management

Orchestration can be used to provision new vApps in a vCloud. Orchestration Management needs to integrate with and provide status on new or updated configuration items to the Configuration Management System (CMS) to provide consistency. Also, the CMS can trigger autoscaling actions for vApps executed by an orchestration workflow, to provide quality of service.

Another aspect of the relationship between orchestration and configuration management is understanding the physical layer that supports the vCloud environment. In mature implementations, orchestration can interact with the Configuration Management layer to identify gaps in the physical layer and remediate as needed to maintain environment stability (for example, adding new storage capacity).

4.8.6.2.5 Orchestration Management in Relation to Security

Services based on vCloud focus on business users, enabling them to request new services directly via the service catalog. Orchestration is critical to such automation and should have an API to communicate with external systems. Orchestration adds flexibility in a vCloud. With flexibility comes a requirement to add controls so that there are no security risks or exposure for the organization. Because the orchestration workflows have access rights to multiple systems, the orchestration workflow code needs to be protected. Encryption controls such as *Set Digital Rights* management need to be enabled while moving workflow code packages within servers. Access to the orchestration servers must be limited. VMware recommends that the COE exclusively control and manage access on these servers.

4.8.6.2.6 Orchestration Management in Relation to Audit and Compliance

Orchestration workflows allow vCloud to be more dynamic. Automated actions enhance key vCloud functions such as provisioning and self-service. Although enhanced automation is highly beneficial, it poses a challenge to organizations that are bound by tight audit, regulatory, and compliance rules. VMware recommends that orchestration engines running the orchestration workflows be centralized within an organization, with centralized error handling and logging for all workflows. Reporting features that checkpoint all workflow actions must be enabled for audit compliance. Centralized orchestration engines also enhance an organization's problem management and root-cause analysis capabilities.

Some recommended orchestration management principles currently cannot be fully automated and require manual configuration actions based on individual client needs. VMware continues to improve existing libraries, and as vCloud implementations mature, more packaged orchestrations with control and governance features should be available for clients to download.

4.8.7 Availability Management

Availability Management focuses on cost-effectively meeting or exceeding the agreed-upon service-level requirements for the level of availability provided for all vCloud service offerings. Managing availability in a vCloud environment depends on VMware vCloud Director component availability and on the resilience of the underlying infrastructure. vCloud Director works transparently with VMware vCenter Server to provision and deploy virtual machines on hosts. Architecting redundancy and protecting the infrastructure components is imperative. VMware vSphere High Availability (HA) can protect provisioned virtual machines; backup tools in the guest operating system or vStorage API can also protect them.

4.8.7.1 Uptime SLAs

VMware vCloud components support a 99.9% uptime SLA out of the box. This might be sufficient for noncritical applications or applications that are inherently highly available. For vCloud, uptime SLAs typically require the following verification:

- End customer workloads are running.
- End customer workloads are accessible (via the vCloud portal, the API, and remote access protocols).

In some cases, a provider (external service provider or internal IT) might want to increase the vCloud uptime SLA. VMware can control the resiliency of only its vCloud platform components and can provide recommendations to mitigate single points of failure (SPOF) in the underlying infrastructure. A provider can eliminate SPOF by providing redundancy. For example:

- Redundant power sourced from multiple feeds, with multiple whips to racks, and sufficient backup battery and generator capacity
- Redundant network components
- Redundant storage components
    - The storage design needs to be able to handle the I/O load. Customer workloads might not be accessible under high disk latency, file locks, and so on.
    - The storage design should be tied to business continuity and disaster recovery plans, possibly including array-level backups.
- Redundant server components (multiple independent power supplies, network interface cards (NICs), and, if appropriate, host bus adaptors (HBAs)
- Sufficient compute resources for a minimum of n+1 redundancy within a vSphere high availability cluster, including sufficient capacity for timely recovery
- Redundant databases and management

Appropriate Change, Incident, Problem and Capacity Management processes must also be well defined and enforced to make sure that poor operational processes do not result in unnecessary downtime. In addition to a redundant infrastructure, everyone responsible for operating and maintaining the environment and the supporting infrastructure must be adequately trained and skilled.

For more detailed information on increasing vCloud component resiliency, refer to Appendix A, "Availability Considerations."

4.8.8 Continuity Management

Continuity Management for vCloud focuses on making sure that the service offerings based on vCloud and the infrastructure upon which they are hosted can be resumed within an agreed-upon timeframe if service is disrupted—regardless of whether the outage is at the vApp level or whether it impacts an entire vCloud environment instance. In this context, VMware defines two components to Continuity Management: Disaster Recovery (strategic), and vApp Backup and Restore (tactical).

4.8.8.1 Disaster Recovery

Disaster Recovery (DR) focuses on recovering systems and infrastructure after an incident that interrupts normal operations. A disaster can be defined as partial or complete unavailability of resources and services, including software, the virtualization layer, the vCloud layer, and the workloads running in the resource groups. Different approaches and technologies are supported, but at least two areas require disaster recovery: the management cluster and consumer resources. Different approaches and technologies are supported.

4.8.8.1.1 Management Cluster Disaster Recovery

Good practices at the infrastructure level lead to easier disaster recovery of the management cluster. This includes technologies such as HA and DRS for reactive and proactive protection at the primary site. VMware vCenter Heartbeat™ can also protect vCenter Server at the primary site. For multisite protection of virtual machines, VMware vCenter Site Recovery Manager™ (SRM) is a VMware solution that works well because the management virtual machines are not part of a vCloud instance of any type (they run the vCloud instances). For a detailed description of using SRM to provide disaster recovery solution for the management cluster, see www.vmware.com/files/pdf/techpaper/vcloud-director-infrastructure-resiliency.pdf.

Disaster Recovery operational considerations for the vCloud management cluster are the same as for a virtualized environment. A vCloud infrastructure risk assessment must be undertaken to determine the threat risk exposure and the corresponding mitigation activities. The actions necessary for executing the mitigation activities, including those for the management cluster, should be captured in a vCloud infrastructure continuity plan. After the vCloud infrastructure disaster recovery planning and technical implementation are complete, awareness building, disaster recovery training, disaster recovery testing, and review/adjustment should be considered part of ongoing vCloud operations.

VMware vCenter Site Recovery Manager 5 can perform a disaster recovery workflow test of the cloud management cluster. This can be useful to verify that the steps taken to move the cloud management stack from the protected site to the recovery site complete without fail. But the SRM test feature is only validation of the workflow, not functional testing of connectivity (because of the fencing feature that protects the production vCloud management cluster).

4.8.8.1.2 vCloud Consumer Resources Disaster Recovery

The vCloud consumer resources (workloads or vApps) can be failed over to an alternate site, but VMware vCenter Site Recovery Manager (SRM) cannot be used. Although SRM is vCenter Server aware, it is not vCloud Director aware. Without collaboration between vCloud Director and SRM, the underlying mechanisms that synchronize virtual machines cannot keep vCloud consumer resources in sync.

A solution for vCloud consumer workload disaster recovery is to use storage replication. Storage replication can replicate LUNs that contain vCloud consumer workloads from the protected site to the recovery site. Because the LUN/datastores containing vCloud consumer workloads cannot currently be managed by SRM, manual steps might be required during failover. Depending on the type of storage used, these steps can potentially be automated by leveraging storage system API calls.

Operationally, recovery point objectives support must be determined for consumer workloads and included in any consumer service-level agreements (SLAs). Along with the distance between the protected and recovery sites, this helps determine the type of storage replication to use for consumer workloads: synchronous or asynchronous.

For more information about vCloud management cluster disaster recovery, see www.vmware.com/files/pdf/techpaper/vcloud-director-infrastructure-resiliency.pdf.

4.8.8.2 Backup and Restore of vApps

Some manual backup and restore procedures are required for the vApps that are deployed into the vCloud. Traditional backup tools do not capture the required metadata associated with a vApp, such as owner, network, and organization. This results in recovery and restoration issues. Without this data, recovery must include manual steps and requires configuration attributes to be manually reentered.

Within a vCloud environment, a vApp can be a single virtual machine or a group of virtual machines, treated as one object. Backup of vApps on isolated networks must be supported. Identifying inventories of individual organizations becomes challenging based on current methods that enumerate the backup items using vSphere. vSphere uses universally unique identifiers (UUIs) to differentiate objects, whereas vCloud Director uses object identifiers.

For backing up and restoring vApps, VMware recommends the use of VMware vSphere® Storage APIs—Data Protection backup technology. This technology has no agents on guest operating systems, is centralized for improved manageability, and has a reduced dependency on backup windows.

Guest-based backup solutions might not work in a vCloud because not all virtual machines are accessible through the network. Also, virtual machines might have identical IP addresses. Therefore, backups of vCloud vApps require a virtual machine-level approach.

Use the full name and computer name fields to specify realistic names that help describe the virtual machines when deploying virtual machines (as part of a vApp). If this is not done, the generic information in these fields can make it difficult to specify individual virtual machines. vApps and virtual machines that vCloud Director provisions have a large GUID template_name. Multiple virtual machines might appear to be similar, making it difficult for a user or administrator to identify and ask for a specific virtual machine to be restored.

4.8.8.2.1 VMware Solutions

VMware Data Recovery is a solution based on vStorage APIs for Data Protection. Other storage APIs for data protection–based backup technologies are available from third-party backup vendors. Currently, because of the issue of universally unique identifiers (UUIs) versus object identifiers, Data Recovery cannot be used with VMware vCloud Director.

For backup of vCloud workloads, VMware recommends that clients validate the level of support provided by the vendor to make sure client requirements are supported. Table 4.3 provides a checklist of vCloud vApp requirements to ask vendors about.

TABLE 4.3 vCloud vApp Requirements Checklist

vApp Requirement	Detail
vStorage API Data Protection integration	▶ vStorage API Data Protection that provides change-block tracking capability to reduce backup windows. ▶ Integration to enable backup of isolated virtual machines and vApps. ▶ Integration with vStorage API Data Protection to provide LAN-free and server-free backups to support better consolidation rations for vCloud and the underlying vSphere infrastructure. ▶ Use of the virtual machine universally unique identifier (UUI) versus the virtual machine name, to support multitenancy and avoid potential name space conflicts.
vCloud Director integration	▶ Interface support for vCloud provider administrator teams. In the future, some vendors might provide consumer (organization administrator and users) access. ▶ Include vCloud metadata for the vApps. This includes temporary and permanent metadata per virtual machine or vApp. This is required to make sure that recovery of the virtual machine or vApp has all the data required to support resource requirements and SLAs.
vApp requirements	▶ Provide vApp granularity for backups. Support the backup of multitiered vApps (for example, a Microsoft Exchange vApp that has multiple virtual machines included. The backup selection of the Exchange vApp would pick up all the underlying virtual machines that are part of the main vApp). This capability is not available today, but vendors are working to develop it.

4.8.8.2.2 Challenges

Challenges associated with backing up and restoring a vCloud include the following:

- vApp naming that poses conflict issues between tenants
- vApp metadata required for recovery
- Multiobject vApp backup (protection groups for multitiered vApps)
- Manual recovery steps in the vCloud
- Support for backup of vApps on isolated networks or with no network connectivity
- Enumeration of vApps by organization for use by the organization administrator
- Enumeration of vApps by organization and provider for use by the organization provider
- User-initiated backup/recovery
- Support of provider (provider administrator) and consumer (organization administrator and user)

For more detailed information about vCloud Business Continuity, see Appendix J, "Business Continuity."

4.8.9 Access and Security Management

Access and Security Management is essential for a vCloud architecture.

4.8.9.1 Workload Isolation

Additional security controls and network functionality can be added to a vCloud platform for greater versatility in hosting enterprise applications.

Using VMware vCloud Networking and Security technology to isolate Layer 2 traffic and persistent network policies, a vApp can have a number of private, vApp-only networks that never leak outside their environment. It is possible to clone this environment indefinitely, never changing an IP address or configuration file.

When a vApp is built, firewall rules created in VMware vCloud Networking and Security Edge (Edge) can permit or restrict access from external vSphere objects or physical networks to TCP and UDP ports of the application. See Figure 4.29.

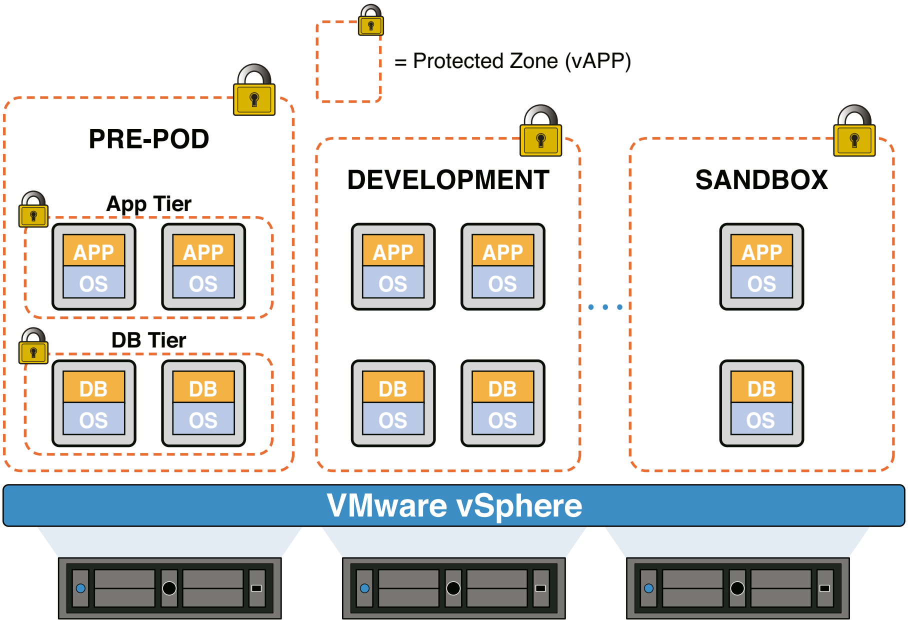


FIGURE 4.29 Workload isolation

Although the vApp is the recommended way to create the virtual infrastructure for multitiered applications, administrators can define security rules based on any of the following vSphere objects: datacenter, cluster, resource pool, vApp, port group, or VLANs. A rule that is created for a container applies to all resources in that container.

4.8.9.2 Access Management

Within a public or private vCloud environment, directory services must be configured for vCloud Director to enable user access to vCloud resources.

A mechanism for authorization and authentication is available within vCloud Director. Directory services based on Lightweight Directory Access Protocol (LDAP) and network authentication protocols such as Active Directory, OpenLDAP, or Kerberos v5 can be configured with vCloud Director. See the *VMware vCloud Director Administrator's Guide* (at www.vmware.com/support/pubs/vcd_pubs.html) for additional information about integrating these services with vCloud Director.

User authorization is controlled through *role-based access control* (RBAC) within vCloud Director. Careful consideration must be given to roles and responsibilities for managing vCloud Director, whether as a provider or as a tenant. The *VMware vCloud Director Administrator's Guide* contains details about permissions, roles, and settings that can be modified to fit the requirements for access control within the organization.

From a provider perspective, the system administrator role should be restricted to only individuals within the provider organization's vCloud operations team who need that level of access. Other individuals within the provider organization who require only vCloud Director organization access should use other roles. If possible, an LDAP group for the provider administrators should be created and imported into vCloud Director with the system administrator role applied to it. All users who require this level of access can then be managed through the LDAP system. The built-in admin account should not be used for vCloud administration, and the credentials must be stored securely.

From a tenant perspective, there are predefined roles. The organization administrator is the highest level of privilege, and it should be limited to individuals within the tenant organization's vCloud operations team that truly require that level of access. This can be achieved with LDAP groups by importing them so that vCloud Director roles can be applied to them. A variety of roles exist with vCloud Director for organizations, and if required, additional roles can be created with alternate privileges. A policy of *least privilege* (grant only privileges required to perform the role) should be applied to all individuals who require access to the vCloud organization, with continued use of LDAP groups to assist with managing this policy.

4.8.9.3 Log Management

Providing log data to customers is an important capability for providers offering vCloud services. The primary advantages include the following:

- **Regulatory compliance:** Aggregate log data for security review and analysis through applicable controls. Archive historical data and retrieve based on an audit window containing relevant data. Logs showing specific events such as a user authentication with a timestamp are examples of satisfactory evidence for auditors

- **Tenant requirements:** Tenants (customers or clients) should have access to logs that pertain to the use of their particular compute resources. Tenant log requirements are similar to those for a provider, but the capability to offer the data that corresponds to the specific tenant is an important capability in a vCloud environment.

- **Event correlation:** Log data can be forwarded to Security Information and Event Management (SIEM) tools for analytic analysis and correlation with unique behavioral signatures. This enables the possibility of early and possibly real-time detection of an attack, misconfiguration, and secondary capacity utilization reporting.
- **Operational monitoring:** For the automation of health and status reporting, logs can provide data that can be checked when required for state changes to applications, operating systems, and virtual machine hosts.
- **Simple troubleshooting:** Many applications and operating systems provide the capability to enable more verbose logging detail during runtime. When troubleshooting unexpected behavior, this additional detail can provide the information needed when attempting to remediate most problems.

4.8.9.3.1 Logging and Architecture Considerations

- **Redundancy:** The leading logging platform is syslog. Syslog is a UDP-based protocol, so the delivery of all log data is not guaranteed. To facilitate the integrity of log delivery over networks, try the following:
  - Design physical redundancy on logging equipment (redundant network interfaces, others).
  - Specify multiple syslog targets.
  - If only one remote syslog target is possible, configure local logging as well as one remote target.
  - Host the log targets on DRS-enabled hosts so that vCenter can manage availability of the syslog virtual machine and service.

- **Scalability:** When compared with customer-generated events, vCloud infrastructure components generate considerably less log data. However, customer components such as the vCloud Networking and Security Edge firewall generate a very high volume of logging. Logs from performance data such as IOPS, network throughput, and CPU utilization are critical, so the design guideline is to define standalone disk partitions for log collection and archiving on a collection server. Additionally, if possible, this data should be part of the vCloud monitoring solution using vCenter Operations Manager.
- **Reporting:**
  - Logs need to be available to customers in raw format from both vCloud Director and vCloud Networking and Security Edge that pertain specifically to their organization and networks.
  - Within vCloud Director, customer-specific activity is specified as an identifier for the customer's organization.
  - vCloud Networking and Security Edge applies descriptive and unique names to organization-specific traffic that SIEM products use to correlate log messages.

4.8.9.3.2 Logging as a Service

When enabling a formalized service for log collection and processing, a provider should consider offering the following types of log services to a customer:

- **Provider log management of customer logs for systems within the vCloud organization:** The customer sends logs to a provider for analysis and report generation of customer-specific events.
    - Pros:
        - Logs can be sent over a private VLAN within the provider's environment.
        - Cost savings for customer of licensing SIEM tools.
    - Cons:
        - Difficult to customize analysis and correlation to other customer-specific events.
        - Dedicated resources are required even with low utilization.
        - Billing does not follow the IaaS model because resource consumption is primarily for storage and analysis.
- **Provider forwarding logs to customer for management:** Logs from provider resources, such as network equipment, host server, and firewall appliances, are sent to the customer system for collection and analysis.
    - Pros:
        - vCloud resources are scalable and rely on distributed analysis within the customer environment.
        - The customer uses a tool of choice for analysis and reporting.
    - Cons:
        - The customer creates a duplicate copy of the infrastructure log, for audit purposes.
        - The log transmission requires network resources.
        - Due to multitenancy within the vCloud, a potentially complex implementation is required as a result of the need for a built-in intelligence engine in the log-forwarding mechanism.

4.9 vCloud Infrastructure Control

vCloud Infrastructure Control deals with architecture and engineering services for the underlying vCloud infrastructure. This layer includes infrastructure architecture services, infrastructure engineering services, and infrastructure deployment services. Operationally, the key for control and governance in these areas is to establish, document, and implement a standardized architecture vision and create consistent design principles and enterprise-wide blueprints for vCloud. Additional guidance on design principles and standards is provided in Chapters 3 and 6. The following are some key topic areas that provide operational guidance.

- Chapter 3, *Architecting a VMware vCloud*:
    - Section 3.2, "vCloud Architecture"
    - Section 3.3 "vCloud Management Architecture"
    - Section 3.6 "vCloud Metering"
    - Section 3.7 "Orchestration and Extension"
- Chapter 6, *Implementation Examples*:
    - Section 6.8, "vCloud Management and Monitoring Examples"

4.9.1 Monitoring

Monitoring the components of a vCloud Director implementation is essential to the health of a vCloud environment and is necessary to maintain capacity and meet service-level agreements. This section provides recommendations regarding what systems and associated objects to monitor, and introduces readily available tools that can be used to extract health-related metrics. This chapter does not go into details on specific limits or thresholds because they are available in the product documentation. This chapter does not attempt to provide specifics for setting up a monitoring solution because various service providers and enterprises might have very different monitoring solutions in place to be integrated.

4.9.1.1 Management Cluster

Design guidelines for monitoring the management cluster components are the same as the guidelines for monitoring vSphere components. A centralized monitoring tool such as VMware vFabric Hyperic HQ Enterprise can be used to monitor the core objects (Oracle Server, SQL Server, Active Directory Server, DNS Server, Red Hat Enterprise Linux Server, and Windows Server) that are needed to run a vCloud environment. A customer can use SNMP and SMASH to monitor the hosts on which the vCloud Director cells are installed and running, but the vCloud Director application itself cannot be monitored by SNMP or SMASH. However, SNMP can be integrated from vCenter. Alternatively, cells can be monitored through integration with a third-party monitoring platform via JMX Beans. Beyond JMX Beans monitoring, the vCloud and vSphere APIs provide component, resource, and activity metrics that can be used for health and capacity management.

4.9.1.2 Cloud Consumer Resources and Workloads

Design guidelines for monitoring the vCloud consumer resources and workloads are the same as for monitoring vSphere. However, there are additional vCloud-specific considerations for VMware vCloud Networking and Security Edge and vCloud consumer workloads.

4.9.1.2.1 vCloud Networking and Security Edge

VMware vCloud Networking and Security Edge appliances are self-contained environments that are stateless in nature. There is a "health check" API call that can be made to an edge appliance to determine whether it is functioning correctly. If the API returns negative, initiate a reboot of the edge device. At the time of reboot, configuration information is updated from the VMware vCloud Networking and Security Manager, and the edge device continues to function properly.

4.9.1.2.2 Cloud Consumer Workloads

Monitoring workloads provisioned by vCloud consumers might be desirable. vCloud Director does not provide any built-in monitoring of workloads for availability or performance. Several third-party solutions are available to monitor vSphere resources and workloads running on vSphere. However, these solutions might not work all the time when vCloud Director is in use. Isolated networking in vApps might prevent monitoring tools from acquiring the performance or availability information of a vApp. Furthermore, vApps might be provisioned and deprovisioned or power-cycled at any time by a vCloud consumer, and these actions might create false positives in the monitoring environment. Until solutions in the market are fully integrated with vCloud Director, providing detailed monitoring for vCloud consumer workloads might be difficult.