



VMware vCloud™ Service Definition for a Public Cloud

Version 1.6

TECHNICAL WHITE PAPER

© 2011 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. This product is covered by one or more patents listed at <http://www.vmware.com/download/patents.html>.

VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

VMware, Inc
3401 Hillview Ave
Palo Alto, CA 94304
www.vmware.com

Table of Contents

1. Introduction	4
1.1 Enterprise Hybrid Cloud	4
1.2 VMware vCloud Datacenter	5
1.2.1 Target Market and Use Cases	5
1.2.2 Challenges Solved	5
2. Service Definition for a Public Cloud	6
2.1 Service Overview	6
2.2 Service Offerings	6
2.2.1 Basic vDC	7
2.2.2 Committed vDC	7
2.2.3 Dedicated vDC	8
3. Compliance Definition	8
3.1 Compliance Controls	9
3.2 Compliance Visibility and Transparency	11
3.3 Compliant Architecture	12

1. Introduction

Cloud computing delivers convenient, on-demand access to shared pools of data, applications, and hardware. The cloud computing paradigm—made possible by sophisticated automation, provisioning, and virtualization technologies—differs dramatically from today’s IT model because it decouples data and software from the servers and storage systems running them and allows IT resources to be dynamically allocated and delivered as a service, either in component parts (where users subscribe to specific applications or simply lease computing power) or as an integrated whole.

Cloud computing is changing the way IT resources are utilized. Users want the ability to access infrastructure resources how and when they choose. IT teams are asked to accommodate this shift in the consumption model, but still deal with the security, compatibility, and compliance issues associated with delivering that convenience to application business owners and developers.

1.1 Enterprise Hybrid Cloud

A *private cloud* is the cloud infrastructure operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise. A public cloud is the cloud infrastructure made available to the general public or a large industry group and is owned by an organization selling cloud services.

A *hybrid cloud* is the cloud infrastructure composed of two or more clouds (private or public) that remain unique entities but are bound together by standardized technology that enables data and application portability (for example, cloud bursting for load-balancing between clouds).

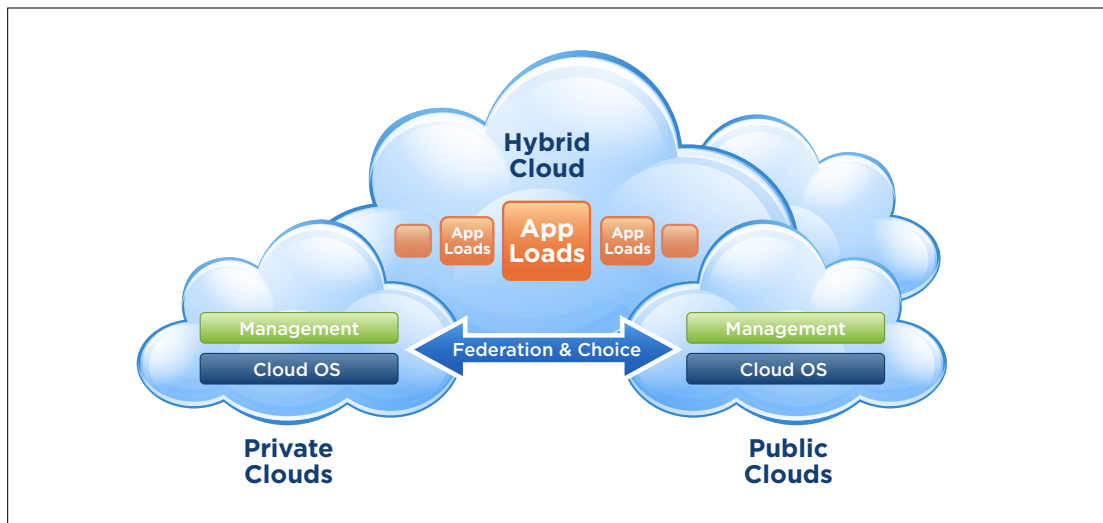


Figure 1. Hybrid Cloud

The common misperception is that cloud computing implies an “external” cloud, based on public cloud services, for example, Amazon. The fact is that cloud computing is how you approach IT; it is “a way of doing computing,” not a destination. Ultimately, most enterprises will benefit from adopting cloud computing within their own datacenters, building “private clouds,” and getting there in an evolutionary way through their existing virtualization journey.

Together with our leading service provider partners, we can also make hybrid clouds a reality through a common platform built around VMware vSphere and VMware vCloud Director, with common management and security models to give the enterprise the confidence they need, and in an environment that provides on-demand application portability.

1.2 VMware vCloud Datacenter

vCloud Datacenter is a VMware program designed to raise the bar and define a whole new class of enterprise-class cloud computing infrastructure services. In a market currently denominated by commodity, low performance, unsecured public cloud offerings, vCloud Datacenter defines a new enterprise-class cloud computing segment. vCloud Datacenter is the hybrid cloud solution for enabling enterprises to extend their private cloud to the public cloud with flexibility, scalability, security, and operational efficiency. Through a common platform built around VMware vSphere and vCloud Director, with common management and security models, in an environment that provides on-demand application portability, enterprise customers and leading global service providers are delivering cloud compatible, connected and integrated hybrid clouds.

vCloud Datacenter services are cobranded by the service provider and VMware, and initially offered by a small number of service providers worldwide. This offers VMware's enterprise customers a choice of 100% compatible services that are based on a VMware cloud architecture and certified by VMware.

1.2.1 Target Market and Use Cases

VMware vCloud Datacenter services are designed to serve the corporate and departmental IT teams inside the medium to large enterprises (1,000 to 50,000+ employees), and the government/federal sectors. The services allow these IT teams to augment their private cloud with public cloud capacity, in order to support pre-production workloads such as test and development workloads, and production workloads such as web applications, marketing/brochure sites, and messaging/collaboration applications.

The services will support both new and existing workloads, as well as the following use cases:

- Pre-Production environments
 - Development/test/stage
- Production environments
 - Web applications
 - Marketing/brochure sites
 - Multi-tiered web applications
 - eCommerce websites
 - Corporate portals and intranet sites
 - Messaging and collaboration applications
 - SharePoint
 - Content/document management
 - Internal wikis/blogs

Note: While the target markets and use cases listed above can be served by a vCloud Datacenter service, this list is not all inclusive of markets and use cases that a service provider can or should target.

1.2.2 Challenges Solved

There are challenges for enterprises to adopt the public cloud today.

- Trust/security
 - Alignment with existing process and tools
 - Regulatory and standards compliance
 - Secure connectivity
 - Data location
- Management
 - Consistent identity and access management
 - Single pane of glass resource management
 - Compatible platforms and API

2. Service Definition for a Public Cloud

The following service definition is largely derived from the VMware vCloud Datacenter program and is being provided as an example for a service provider to consider. More detailed specifications are provided to a service provider that joins the program.

2.1 Service Overview

The VMware-powered public cloud delivers three classes of on-demand, self-service virtual datacenters (vDCs):

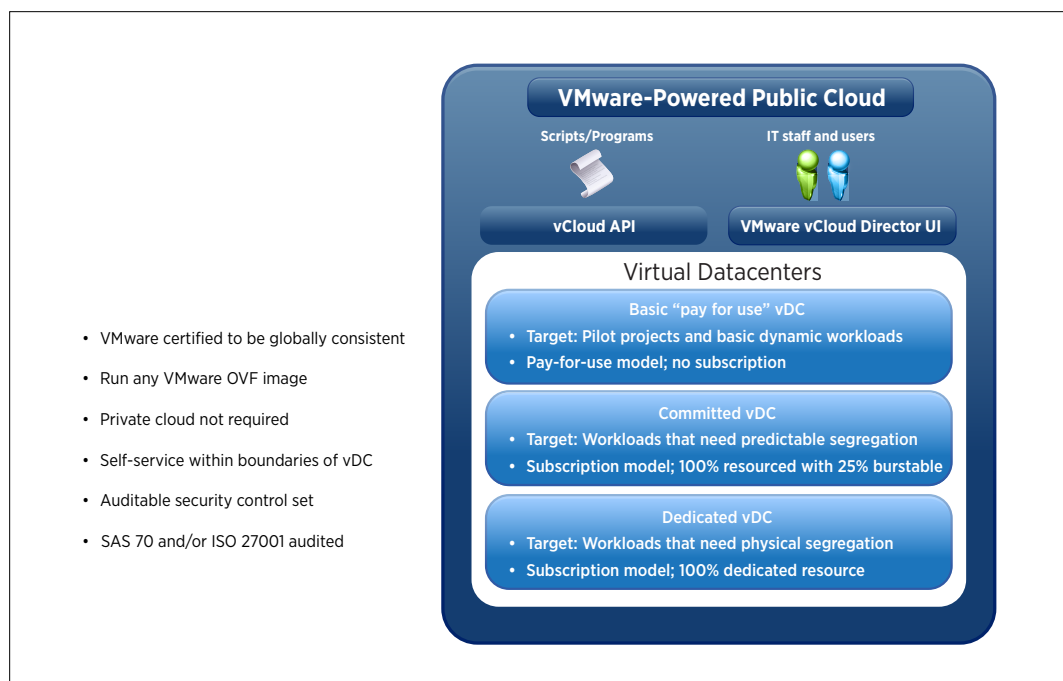


Figure 2. VMware-Powered Public Cloud Service Classes

The service is designed to make it as easy as possible for enterprises to move their workloads to the VMware-powered public cloud. Any existing VMware virtual machine (VM) or virtual application (vApp) can be run with little or no modification in the public cloud, and compatibility with existing enterprise VMware deployments is a key design objective. There is no requirement for an enterprise to deploy a private cloud—any VMware virtualized infrastructure is compatible.

2.2 Service Offerings

The VMware-powered public cloud will have three different service offerings. A single customer may have one or more of the three offerings:

- **Basic vDC:** unreserved “pay per use” class. Designed to quick start pilot projects, or for workloads such as software testing that doesn’t need reservations and high performance.
- **Committed vDC:** provides reserved compute resources (subscription model) with the ability to burst above committed levels if additional capacity is available. Offers predictable performance by reserving resources for workloads within a multi-tenant infrastructure, while also allowing on-demand self-service.
- **Dedicated vDC:** provides dedicated compute resources (using specific hardware dedicated to a given customer), sometimes known as “virtual private cloud.” Offers predictable performance by reserving dedicated resources, which is useful for situations where security or compliance requirements require physical separation.

	BASIC vDC	COMMITTED vDC	DEDICATED vDC
Consumption Model	Pay-as-you-go	Allocation pool	Reservation pool
Chargeable Unit	CPU, Memory, Storage	Resource pool	Resource pool
Targeted Use Cases (but not limited to)	Pre-production (test/dev/stage)	Tier 2/3 production	Production, high performance, security, compliance requirements
Metering	Hourly	Monthly	Monthly

Table 1. Service Offering Differences

We suggest an initial VM capacity of 1,500 VMs, split across the three service offerings. Initially you should have this capacity set so that 50% of the total number of virtual machines will be run in the reservation pool model and 50% will be run in the Pay-As-You-Go model. Furthermore, you should split the reservation pool into small, medium, and large pools with a respective split of 75%, 20%, and 5%. For more detailed information, see the Sizing the Cloud section in the *Architecting a vCloud* document.

2.2.1 Basic vDC

The Basic vDC offering is an instance-based pay-as-you-go resource consumption model. Each virtual machine provisioned in this vDC is charged separately and separate billing records are produced for each virtual machine. Customers using the Basic vDC service will be charged for each hour or part hour of consumption. For example, if a customer uses a machine for 5 minutes they will be charged for one hour of usage. If a customer changes the virtual machine size after 5 minutes, this starts a new hour.

Instance-based model refers to the bundling of vCPU and memory together into a single virtual machine instance and price. Pricing will be charged by hour of consumption.

Memory will be offered in the following unit options: 512 MB, 1 GB, 2 GB, 4 GB, 8 GB, and 16 GB.

vCPU will be offered in the following unit options: 1, 2, 4, and 8.

Pricing will be charged by hour of consumption along the 24 combinations of memory and vCPU unit options.

Actual prices are not specified here because pricing is specific to a service provider's infrastructure, capabilities, location, and market conditions.

2.2.2 Committed vDC

The Committed vDC service offering uses the allocation-pool consumption model. (See the vCloud Director documentation for more details on using an allocation-pool consumption model.) A user is allocated a vDC that contains a certain amount of CPU (GHz), memory (GB), and storage (GB). The allocation pool model is defined using two parameters, the reservation percentage and the total allocation (also called limit). The reservation percentage is how much resource will be guaranteed or committed for the customer. The total allocation or limit is the maximum amount of resource the customer can consume.

For the Committed vDC, the reservation percentage is always 75% of the total allocation/limit. This means the customer can burst up to an additional 25% of resources they originally requested. So, for example, if a customer buys a virtual datacenter with 10 GHz of CPU resources, a virtual datacenter is created for the customer and 10 GHz is allocated for the customer. This is the maximum amount of CPU the customer can ever consume. Of this 10 GHz, the service provider should reserve 75%, which is 7.5 GHz. This is the amount of CPU that's guaranteed for the customer. The 25%, or 2.5 GHz, will be available to the customer if the underlying cluster has available resources. This model allows service providers to charge a price that's generally higher than just 7.5 GHz given the additional burstable capacity, but gives customers the benefit of potentially paying less than if the resources were fully guaranteed.

In the Committed vDC model, we have the following virtual datacenter sizes defined.

	SMALL	MEDIUM	LARGE	X-LARGE	CUSTOM
CPU Reserved	7.5 GHz	18.75 GHz	37.5 GHz	75 GHz	N/A
CPU Limit	10 GHz	25 GHz	50 GHz	100 GHz	N/A
Memory Reserved	15 GB	37.5 GB	75 GB	150 GB	N/A
Memory Limit	20 GB	50 GB	100 GB	200 GB	N/A
Storage	400 GB	1 TB	3 TB	6 TB	N/A
Virtual Machine Limit (Can be changed)	20	50	100	200	N/A
Approx Virtual Machines (Not limit)	10-20	25-50	50-100	100+	N/A

Table 3. Committed vDC Model Size Definitions

In most virtualized environments, memory is the gating factor in terms of resource utilization. Based on the distribution ratio below, we calculate that on average a virtual machine consumes 2.15 GB of memory. For a 20 GB virtual datacenter, that's approximately 10 virtual machines. If a customer runs more small virtual machines, that count can increase. Thus, setting the virtual machine limit to be around 20 will give customers some flexibility to run more small virtual machines but still manage the capacity effectively.

2.2.3 Dedicated vDC

The Dedicated vDC service offering uses the reservation pool-based model. A customer works with a service provider to provision a cluster of servers that's dedicated to this customer. The hardware (network, storage, servers) is not shared with other customers. The customer gets full control over the reservation and limit of this set of resources. This service offering will likely be a fixed price monthly subscription.

3. Compliance Definition

Security and compliance continues to be one of the biggest barriers for enterprise customers to adopt the public cloud. Most regulations and mandates in the industry, including SOX, PCI, HIPAA, COBIT, and ISO, all have two areas of requirements: transparency/visibility and control. Transparency is an absolute must, as the cloud consumers need to know who's accessed what data, when and where, and potentially why based on some documented evidence. PCI requirement 10 is a good example of the need for visibility and transparency. Control is also necessary component of compliance for the cloud consumers. For example, the cloud consumers need to be able to control who can access, configure and modify the cloud environment, what firewall ports are open, when to apply patches and where the data resides. Cloud consumers, and especially the enterprise customers, believe in "you can outsource responsibility, but you can't outsource accountability." At the end of the day, the cloud consumers are still accountable for being compliance.

The VMware-powered public cloud is designed to tackle this very problem, and, along with service provider partners, will do so in three areas:

- Ensure compliance through ISO27001 certification or SAS70 Type II audit, based on a standard set of controls.
- Provide compliance logging and reports to the customers so they have full visibility into their public cloud environments.
- Architect the service so customers can have control the access to their cloud environments.

3.1 Compliance Controls

In order to ensure the enterprise customers feel secure and safe in the public cloud, and that they have the necessary information and visibility into the service to meet their own internal audit requirements, the VMware-powered public cloud services must have one of the following completed within three months of launching the service as GA:

- ISO27001 certified, which proves that security management processes are in place, and have a relevant subset of the ISO27002 controls in place as specified by VMware's Compliance Architecture and Control Matrix; or
- SAS 70 Type II audits based on the same relevant set of controls

VMware will supply the standard set of compliance controls (see table below), and the service provider is responsible for the actual ISO or SAS70 audits using third party auditors. The compliance controls will be published to the provider's customers so they understand that the vCloud Datacenter-certified public cloud service is not only compliant, but that the customers have full visibility into what controls the services were audited against.

ISO 27002	CONTROL TITLE
A.06.2.2	Addressing security when dealing with customers
A.07.1.1	Inventory of assets
A.07.1.2	Ownership of assets
A.07.2.1	Classification guidelines
A.07.2.2	Information labeling and handling
A.08.1.1	Roles and responsibilities
A.08.3.3	Removal of access rights
A.09.2.4	Equipment maintenance
A.09.2.6	Secure disposal or re-use of equipment
A.10.01.1	Documented operating procedures
A.10.01.2	Change management
A.10.01.3	Segregation of duties
A.10.01.4	Separation of dev, test and ops
A.10.03.1	Capacity management
A.10.03.2	System acceptance
A.10.04.1	Controls against malicious code
A.10.04.2	Controls against mobile code
A.10.05.1	Information back-up
A.10.06.1	Network controls
A.10.06.2	Security of network services
A.10.07.1	Management of removable media
A.10.07.2	Disposal of media
A.10.07.3	Information handling procedures
A.10.07.4	Security of system documentation
A.10.08.4	Electronic messaging

ISO 27002	CONTROL TITLE
A.10.08.5	Business information systems
A.10.10.1	Audit logging
A.10.10.2	Monitoring system use
A.10.10.3	Protection of log information
A.10.10.4	Administrator and operator logs
A.10.10.5	Fault logging
A.10.10.6	Clock synchronization
A.11.1.1	Access control policy
A.11.2.1	User registration
A.11.2.2	Privilege management
A.11.2.3	User password management
A.11.2.4	Review of user access rights
A.11.3.1	Password use
A.11.4.2	User authentication for external connections
A.11.4.3	Equipment identification in networks
A.11.4.4	Remote diagnostic and config. port protection
A.11.4.5	Segregation in networks
A.11.4.6	Network connection control
A.11.4.7	Network routing control
A.11.5.1	Secure log-on procedures
A.11.5.2	User identification and authentication
A.11.5.3	Password management system
A.11.5.4	Use of system utilities
A.11.5.5	Session time-out
A.11.5.6	Limitation of connection time
A.11.6.1	Information access restriction
A.11.6.2	Sensitive system isolation
A.12.1.1	Security requirements analysis and specification
A.12.2.1	Input data validation
A.12.2.2	Control of internal processing
A.12.2.3	Message integrity
A.12.2.4	Output data validation
A.12.3.2	Key management
A.12.4.1	Control of operational software
A.12.4.2	Protection of system test data

ISO 27002	CONTROL TITLE
A.12.5.1	Change control procedures
A.12.5.2	Technical review of apps after op. system changes
A.12.5.3	Restrictions on changes to software packages
A.12.5.4	Information leakage
A.12.6.1	Control of technical vulnerabilities
A.13.1.1	Reporting information security events
A.13.1.2	Reporting security weaknesses
A.13.2.3	Collection of evidence
A.14.1.2	Business continuity and risk assessment
A.14.1.3	Developing and implementing continuity plans
A.14.1.5	Testing, maintaining and reassessing plans
A.15.2.2	Technical compliance checking

Table 3. Compliance Controls

3.2 Compliance Visibility and Transparency

Log management is built into many of the compliance frameworks such as ISO, HIPAA, PCI and COBIT. It is required to meet the requirements of these audit standards. Enterprise customers not only need visibility into their private clouds, but they also demand that the service providers provide them visibility into their public cloud environments. For example, enterprise customers are looking for all the necessary logs and reports around user activities, access control, firewall connections and others.

To meet the requirements of being compliant with the controls listed above, service providers must provide cloud consumers visibility and transparency into the vCloud Datacenter-certified service. To accomplish this, service providers must be able to collect and maintain logs for all components of the service and be able to provide relevant logs back to the cloud consumers. The service provider may choose to keep the logs for the underlying infrastructure private to the service provider. In this case, the service provider must be willing to provide these to a customer in the case of an audit. In general, vCloud Datacenter-certified services should have logs covering the following components of a customer’s environment and make them available to their customers in a proactive fashion:

- vCloud Director
- vShield Edge

The VMware-powered public cloud is based on a set of products that have been battle tested in many secure environments, and the products such as VMware vCloud Director and vShield generates a set of logs that gives customers visibility into all the user activities and firewall connections. (VMware will provide vCloud Datacenter program service providers the necessary blueprints and best practices so that service providers can capture this set of logs and provide customers the ability to download them.) The logs must be available to the customers for a minimum of 6 months.

In addition to the logs, service providers should provide basic compliance reports to the enterprise customers so they understand all the activities inside their cloud environment. VMware will provide a set of best practices in this area to ensure the vCloud Datacenter-certified service meets the enterprise customer requirements. The service provider will be responsible for ensuring the successful logging of their vCloud Datacenter-certified service as well as their customer’s environments as defined above. This capability must be implemented and validated before the service is launched as GA..

3.3 Compliant Architecture

All vCloud Datacenter-certified services offer unparalleled security. vCloud Datacenter services are built on vSphere, the most secure virtualization platform with ELA4+ (vSphere 4 in progress) and FISMA certifications, and VMware vCloud Director, a cloud delivery platform offering secure multi-tenancy and organization isolation. With the vCloud Datacenter services, enterprises can exercise the defense-in-depth security best practice as the platform offers both per-organization firewalls and per-vApp firewalls; and all organizations are isolated with their own layer 2 networks. Access and authentication can be performed against the enterprise's own LDAP/AD directory, which means that the enterprise can manage its own user base and provide role-based access according to its own policies. All vCloud Datacenter infrastructures and datacenters have been audited against a standard set of compliance controls for SAS 70 Type II or received ISO27000 certifications. In addition, all vCloud Datacenter services will provide customers relevant audit logs and compliance reports for their cloud environments to ensure enterprises can meet their own internal audit requirements.

