



Service Definition for Private Cloud

TECHNICAL WHITE PAPER

Table of Contents

Introduction	3
Phase I	3
Phase II	3
Workload Categories	4
Transient	4
Highly Elastic	4
Infrastructure	4
Use Cases	4
Common Use Cases	4
Service Tiers	5
Roles & Rights	6
Metering Using Chargeback	6
Suggested vApp Catalog	7
Operating Systems	7
Infrastructure Apps	7
Other Considerations	7
Operational Components	7
Sizing Assumptions	8

Introduction

The service defined here will describe the creation of a fully virtualized, pooled compute platform for use as an Enterprise Private Cloud, or business-internal organizational cloud computing service, commonly referred to as Infrastructure as a Service (IaaS). VMware technologies including vSphere, vShield Manager, vCenter Chargeback, and vCloud Director will be used to establish this service.

The result of this service includes secure multi-tenancy for Lines of Business, shared virtualized assets, a self-service user portal, standard catalogs of pre-defined virtual machines & applications with usage metering. Going forward, this service will be enhanced with greater capacity and the preparation for linking with other cloud services either within or beyond the boundaries of the Enterprise.

Phase I

Goals

- Deliver a fully operational private cloud infrastructure
- Maintain IT control of access to the system and resources
- Provide differentiated tiers of scale to align with business needs
- Allow for metering of the service for internal cost distribution
- Establish a Catalog of common infrastructure & application building blocks
- Size to provide capacity for 400 VMs

Timeline

- Operational by Q4-2010

Phase II

Goals

- Enhance the cloud service for federation to public cloud resources
- Provide for workload redundancy and continuity options
- Size to provide capacity for 1,000 VMs

Timeline

- TBD

Workload Categories For Private Cloud

Private Cloud use cases generally fall into 3 different categories of workloads.

Transient

A transient application is one that is used infrequently, exists for a short time period or is used for a specific task/need and the discarded. This type of workload is appropriate for a Pay As You Go allocation model in VMware vCloud Director.

Highly Elastic

An elastic application is one that is dynamically growing and shrinking it's resource consumption as it runs. An example of this would be a retail application that sees dramatically increased demand during holiday shopping seasons, or a travel booking application that expands rapidly as the fall travel season approaches. This bursty type of workload is appropriate for the Allocation Pool model in VMware vCloud Director.

Infrastructure

An infrastructure application is one that tends to run all the time at a predictably steady state. Examples of these include a print server, directory server, an e-mail relay server or systems monitoring engine. This type of workload is appropriate for a Reservation Pool model in VMware vCloud Director.

Use Cases

Common Use Cases

The table below lists common use cases that fall into the general cloud Workload cases. These come from VMware customer engagements and have been found to be the most popular workloads.

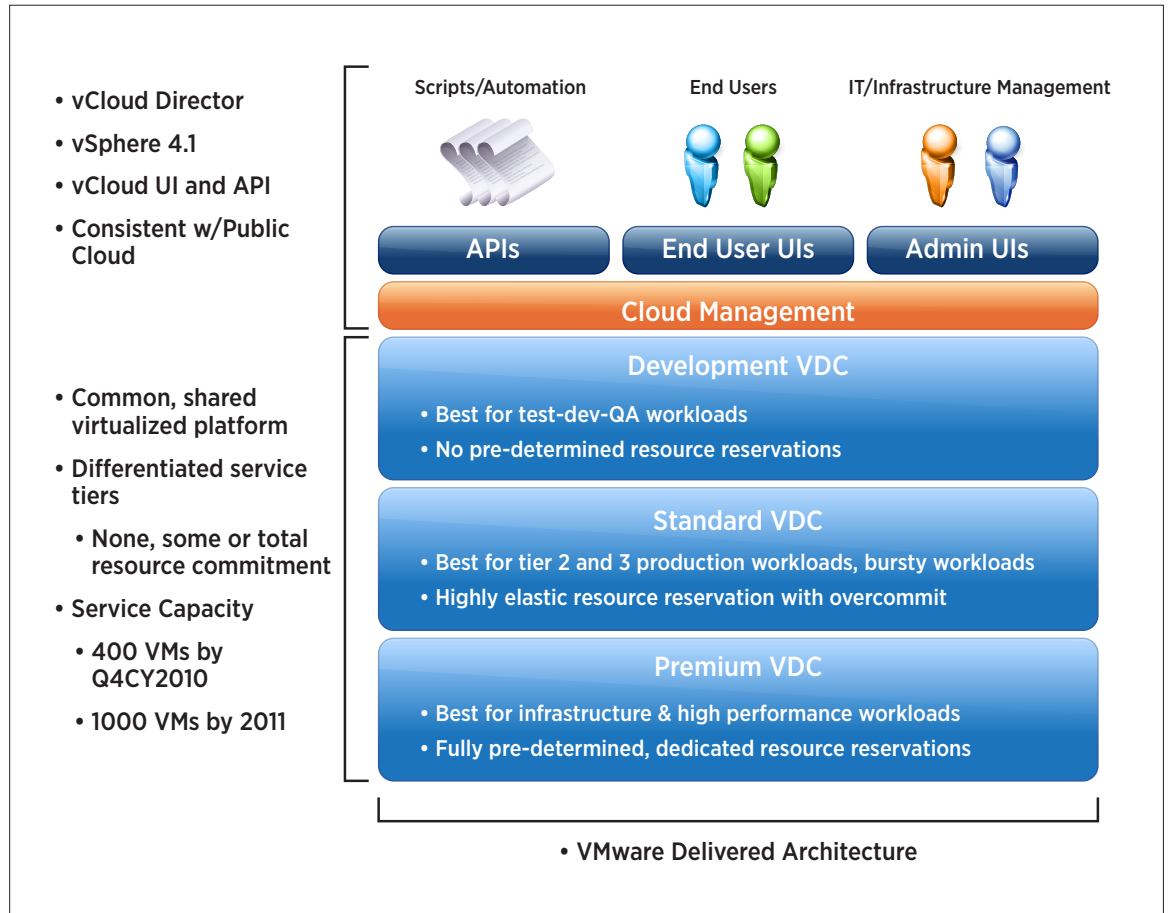
WORKLOAD EXAMPLES	
Software Testing	Software Development
Website Design	Custom Applications
File & Print Services	E-Mail
Sales & Demonstration	Contractor Access

Commonly, virtual lab automation use cases are good fits for a Private Cloud, including integration & performance testing, dev-test-QA workflows for software authoring and general access for developers. Also, a Private Cloud is well suited for sales demonstrations, training staff and partners on software and easily accessing pre-built configurations of complex application suites. Finally, a Private Cloud can be used to establish expanded capacity outside of the realm of a traditional MSP (hosting provider) for virtual lab use or lab bursting based on project needs.

With the knowledge of the workload types and common uses cases you can begin to define the service tiers that you will offer inside of your enterprise.

Service Tiers

Tiering of service within the Enterprise can be an effective method of differentiating types, location or owners of infrastructure as reflected in a Private Cloud. The following service tiers map to the needs defined by the 3 main workload types mentioned above.



	DEVELOPMENT VDC	STANDARD VDC	PREMIUM VDC
Consumption Model	Pay-as-you-go	Allocation pool	Reservation pool
Use Case(s)	Dev-Test-QA	Tier 2/3, Seasonal or Bursty	Tier 2, Print/Directory/E-Mail
Workload Category	Transient	Highly Elastic	Infrastructure
Chargeable Unit	CPU, Memory, Storage	Resource pool	Resource pool
Catalog	Private	Private	Private

Roles & Rights

There are several roles defined in the access model of vCloud Director. There are administrative roles at both the system level and at the Organization or Private Cloud level. Within the Organization, there are levels of rights granted to pre-defined roles that have an important impact on how users interact with the cloud UI.

At a minimum, you'll want to define a system administrator, an organizational administrator for each tenant and a secondary, non-administrative user per tenant/Private Cloud.

ACCOUNT TYPE	NEEDS	PURPOSE
System Administrator	One (minimum)	Highest level admin; has superuser rights
Organization Administrator	One per Organization	Administrator in the Org over systems & users
Organization Author	One or more, as needed	Allows vApp & Catalog creation; no infrastructure mgmt..

Metering Using Chargeback

A concept highly desired by the Enterprise, but historically very difficult to impossible to implement, is the notion of metering the use of the virtual infrastructure for the purposes of inter-departmental billing. Cloud computing enables this capability and thus should be considered as part of a standard service installation.

How this metering data is used falls to the architects of the infrastructure, but at a minimum the Resource Pools that back each Provider vDC should be monitored with VMware vCenter Chargeback and the resulting reports provided to the line of business on a quarterly basis (for the purposes of 'showback').

The following table gives some examples only of workload VM sizing & costing.

VIRTUAL MACHINE TYPE	SIZING	STORAGE	COST MODEL	
Large	4 vCPU x 8GB RAM	200G	Provision: \$400	Operate: \$200/mo.
Medium	2 vCPU x 2GB RAM	60G	Provision: \$300	Operate: \$100/mo.
Small	1 vCPU x 1GB RAM	30G	Provision: \$200	Operate: \$50/mo.

Suggested vApp Catalog

The following is a list of suggested vApp templates that would help promote immediate use of the Private Cloud.

Operating Systems

- Microsoft Windows Server 2003 R2 Enterprise Edition
- Microsoft Windows Server 2008 R2 Enterprise Edition
- RHEL 5.x
- Centos 5.x
- Novel SUSE Linux Enterprise Server 11
- Ubuntu Server 10.04

Infrastructure Apps

- Databases
 - Microsoft SQL Server 2000/2005/2008
 - Oracle 11g
 - MYSQL 5.x
- Web/App Servers
 - Microsoft IIS
 - Spring tcServer
 - Apache Tomcat
 - IBM Websphere Application Server 7
- Simple n-Tier Apps
 - 2-Tier app with a web front-end and database backend
 - 3-Tier app with web, processing and database
 - Enhanced 3-Tier with added monitoring
- Load balancer

Other Considerations

Operational Components

When creating a private cloud there are other operational components that you will need if you do not already have them in place in order to maintain a good experience for the consumers of the cloud. Some of these operational components are:

- Ticketing
 - Helpdesk support (online or phone)
 - Optional: online ticket viewing
- Monitoring
 - Monitoring of all infrastructure components
- Metering
 - Quarterly showback reporting to the LoB
 - 12 months of metering history

Sizing Assumptions

The following are assumptions that can be used for service capacity planning.

- Capacity Assumption
 - 400 VMs by Q4-2010
 - 1,000 VMs by Q4-2011
 - 75% small / 20% medium / 5% large
- Distribution Assumption
 - 45% small (1 GB, 1 vCPU)
 - 35% medium VMs (2 GB, 2 vCPU)
 - 15% large VMs (4 GB, 4 vCPU)
- Storage Assumption
 - Average 60 GB per VM

The following table gives some examples of workload VM sizing & utilization.

VIRTUAL MACHINE TYPE	SIZING	CPU UTILIZATION	MEMORY UTILIZATION
Large	4 vCPU x 8GB RAM	>50% average	High (upwards of 90%)
Medium	2 vCPU x 2GB RAM	20-50% average	Moderate (50% - 75%)
Small	1 vCPU x 1GB RAM	10-15% average	Low (10% - 50%)

