Enterprise-Grade Virtual Private Cloud from a State-of-the-Art Reference Implementation

Dramatically improve IT agility and reduce infrastructure costs with a Virtual Private Cloud optimized for automated workload management, elastic delivery of IT services, and demand-driven utility pricing.

TR174 Addendum C
Version 0.1
Notice

Copyright © TeleManagement Forum 2012. All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published, and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this section are included on all such copies and derivative works. However, this document itself may not be modified in any way, including by removing the copyright notice or references to TM FORUM, except as needed for the purpose of developing any document or deliverable produced by a TM FORUM Collaboration Project Team (in which case the rules applicable to copyrights, as set forth in the TM FORUM IPR Policy, must be followed) or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by TM FORUM or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and TM FORUM DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY OWNERSHIP RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

TM FORUM invites any TM FORUM Member or any other party that believes it has patent claims that would necessarily be infringed by implementations of this TM Forum Standards Final Deliverable, to notify the TM FORUM Team Administrator and provide an indication of its willingness to grant patent licenses to such patent claims in a manner consistent with the IPR Mode of the TM FORUM Collaboration Project Team that produced this deliverable.

The TM FORUM invites any party to contact the TM FORUM Team Administrator if it is aware of a claim of ownership of any patent claims that would necessarily be infringed by implementations of this TM FORUM Standards Final Deliverable by a patent holder that is not willing to provide a license to such patent claims in a manner consistent with the IPR Mode of the TM FORUM Collaboration Project Team that produced this TM FORUM Standards Final Deliverable. TM FORUM may include such claims on its website, but disclaims any obligation to do so.

TM FORUM takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this TM FORUM Standards Final Deliverable or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on TM FORUM's procedures with respect to rights in any document or deliverable produced by a TM FORUM Collaboration Project Team can be found on the TM FORUM website. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a
general license or permission for the use of such proprietary rights by implementers or users of this TM FORUM Standards Final Deliverable, can be obtained from the TM FORUM Team Administrator. TM FORUM makes no representation that any information or list of intellectual property rights will at any time be complete, or that any claims in such list are, in fact, Essential Claims.

Direct inquiries to the TM Forum office:

240 Headquarters Plaza,
East Tower – 10th Floor,
Morristown, NJ 07960 USA
Tel No.  +1 973 944 5100
Fax No.  +1 973 944 5110
TM Forum Web Page: www.tmforum.org
Table of Contents

Notice .................................................................................................................................................................. 2
Table of Contents ............................................................................................................................................. 4
List of Figures ................................................................................................................................................ 6
List of Tables ................................................................................................................................................ 6
Executive Summary ........................................................................................................................................ 7
1. Introduction .................................................................................................................................................. 8
   1.1. Document Structure .......................................................................................................................... 9
   1.2. Introduction ....................................................................................................................................... 9
   1.3. Issues and Appendices ....................................................................................................................... 10
   1.4. Terminology used within this document ......................................................................................... 10
2. Traditional IT Infrastructure Challenges ................................................................................................. 12
3. Optimal Infrastructure for “Everything-as-a-Service” ........................................................................... 13
4. Enterprise Adoption of Virtual Private Clouds .................................................................................... 14
5. Requirements for an Enterprise-Grade VPC ......................................................................................... 16
6. VPC Reference Architecture Overview ............................................................................................... 18
7. Workload Lifecycle Management ........................................................................................................ 19
   7.1. VPC Reference Implementation ..................................................................................................... 19
8. Automated Infrastructure as a Service .................................................................................................... 21
   8.1 VPC Reference Implementation ..................................................................................................... 21
   9.1 VPC Reference Implementation ..................................................................................................... 23
10. Optimized Physical Infrastructure ....................................................................................................... 25
    10.1 VPC Reference Implementation .................................................................................................. 25
11. Reference Implementation Topology .................................................................................................... 26
    11.1 Reference Implementation Topology .......................................................................................... 26
12. Reference Implementation Test Results ............................................................................................... 28
    12.1 Workload Types Used for Testing ............................................................................................... 28
    12.2 Capacity Units Per Blade .............................................................................................................. 28
    12.3 VM Density .................................................................................................................................. 29
    12.4 VM Costs ...................................................................................................................................... 29
    12.5 On-demand Provisioning ............................................................................................................ 30
    12.6 Auto Scaling ............................................................................................................................... 32
    12.7 Dynamic Security Configuration with Virtual Firewalls .......................................................... 33
13. Summary ................................................................................................................................................ 34
14. References ............................................................................................................................................... 35
    14.1 IPR Releases and Patent Disclosures ............................................................................................ 35
15. Administrative Appendix ....................................................................................................................... 36
List of Figures
Figure 1: VPC Reference Architecture Overview 18
Figure 2: Agility Platform 20
Figure 3: Agility Platform 22
Figure 4: Agility Platform 24
Figure 5: Reference Implementation Topology 27
Figure 6: On-Demand Provisioning 31
Figure 7: Provisioning Time 31
Figure 8: Auto Scaling 32
Figure 9: Auto Scaling 32

List of Tables
Table 1: Capacity Units Per Blade 29
Table 2: Dynamic Security Configuration with Virtual Firewalls 33
**Executive Summary**

Virtual Private Clouds (VPCs) help address critical business needs for a more agile and less costly IT operating model with greater vendor contestability. This paper can assist organizations in jump starting their VPC initiatives by detailing the capabilities and requirements of a state-of-the-art VPC, providing a reference architecture, and quantifying the benefits and metrics from an enterprise-grade VPC reference implementation that was tested in a lab environment. These results showed that VPCs can deliver high levels of automation, time savings, and can operate VMs at costs up to 75% less than comparable public cloud offerings.

The intent of this paper in the context of TMForum was to illustrate an initial field test of the Infrastructure as a Service (IaaS) requirements in Technical Report 174. As a follow-up, The Enterprise Cloud Leadership Council will use the “lessons learned” from this initial exercise to develop additional proofs of concepts for other IaaS requirements. The goal is to further refine the context created in this implementation in new Catalyst projects.

The document was contributed to TM Forum on July 6, 2012. It was originally developed by ServiceMesh, Cisco, Savvis, Oracle, and Commonwealth Bank of Australia.

This document is primarily intended for IT leaders responsible for cloud strategy development, cloud implementations, enterprise data centers, and IT operations. It is also intended for IT practitioners involved in the design, architecture, and operation of cloud-based technologies ranging from high density computing, storage, security, governance, and lifecycle management.
1. **Introduction**

Enterprises today are struggling with unprecedented levels of change, complexity, and pricing pressure in the marketplace. This has put tremendous pressure on IT organizations to adapt to rapidly changing business needs while simultaneously lowering capital spending and reining in operating costs.

One of the primary hindrances to achieving these goals is an inflexible and siloed IT infrastructure. Traditional IT infrastructure often consists of tightly bound proprietary hardware and software solutions that result in high maintenance and support costs, slow and costly application deployments, low computing resource utilization rates, and high vendor switching costs. This high fixed-cost infrastructure often consumes the majority of the IT budget, leaving few resources left for IT to innovate.

There is a pressing need for IT infrastructure that adapts more rapidly and costs less to operate, and organizations are adopting Agile IT operating models that leverage cloud computing to accomplish this. These Agile IT operating models leverage a federation of internal and external cloud providers, and infrastructure-, platform-, and software-as-a-service that are readily selected and changed to provide the best fit and value. The resulting benefits can be significant:

- Faster time-to-market through on-demand provisioning and a compressed software development lifecycle.
- Greater agility to start, stop, and scale projects with a variable cost infrastructure requiring less investment.
- Cost savings through “pay-as-you-go”, demand-driven consumption.
- Lower costs by mitigating vendor lock-in and enabling price arbitrage between providers.
- Increased transparency and more direct revenue-to-cost operating models.
- Improved standardization and lower risk operating environments.

The technology and tools are available to implement Agile IT operating models today, and a critical component of those initiatives is the Virtual Private Cloud (VPC). A VPC is a combination of one or more internal and external clouds that are securely combined into common resource pools that enterprises can use to deploy platforms and software while achieving an optimal balance of scale, cost, and performance. This document describes the reference architecture and reference implementation of a state-of-the-art, enterprise-grade Virtual Private Cloud that can deliver on the business objectives of a more agile and less costly IT operating model.
1.1. Document Structure

1.1 Introduction

Executive Summary: Summarize the main points from the document and highlight the problem statement being addressed, the main results, the conclusions drawn and the next steps as appropriate.

Section 1 Introduction: Provides overview of this the document and outlines its structure and defines essential terms used in the document.

Section 2 Traditional IT Infrastructure Challenges: Traditional IT infrastructures exhibit a few common characteristics that limit their ability to adapt rapidly to change and introduce high costs and risks into IT initiatives including application upgrades and migrations.

Section 3 Optimal Infrastructure for “Everything-as-a-Service”: Every large enterprise over time is going to transition their IT operating model towards a combination of internal and external cloud providers, and infrastructure, platform, and software as a service.

Section 4 Enterprise Adoption of Virtual Private Clouds: VPCs are poised to become a dominant implementation model of cloud computing for most large enterprises.

Section 5 Requirements for the Enterprise-Grade VPC: VPCs exhibit some unique requirements to satisfy enterprise-grade customers that are typically not cost effective or practical to implement for external public clouds and a general purpose customer base.

Section 6 VPC Reference Architecture Overview: This VPC reference architecture addresses the previously described enterprise-grade VPC requirements in a modular, flexible design with minimal complexity.

Section 7 Workload Lifecycle Management: An enterprise-grade platform is needed to ensure consistent workload governance, security and lifecycle management across all IaaS, PaaS, and SaaS offerings and their cloud deployment environments.

Section 8 Automated Infrastructure as a Service: Configuring the underlying infrastructure of enterprise applications can be highly complex, with countless interdependencies among supporting network and security services, along with the complexity of the application topology.

Section 9 Zoned Security Models: Workload security requirements can come from a broad range of sources both internal (such as keeping competitive information private) and external (such as government privacy regulations for customer data).
Section 10  **Optimized Physical Infrastructure**: Key objectives for the physical infrastructure are to support rapid implementation, reduced interoperability challenges, lower operating and support costs, and simplicity of design.

Section 11  **Reference Implementation Topology**: A high level topology of the VPC reference implementation has been provided. This reference implementation was designed to meet the aforementioned enterprise-grade requirements including high availability.

Section 12  **Reference Implementation Test Results**: The VPC testing methodology and results are summarized.

**1.2 Issues and Appendices**

Appendix A  **References**

Administrative Appendix provides document revision history, acknowledgements for work completed and information about the TM Forum.

**1.2. Terminology used within this document**

This section identifies the important terms, abbreviations and acronyms introduced by the project and are necessary for the understanding of this document. More detailed definitions and definitions for other terms are shown in Appendix A.

This document makes use of the following terms:

**Stateless Computing**
Stateless computing refers to separating and abstracting each application architecture tier, including the underlying hardware and operating system, so that workloads can be readily scaled and moved from one infrastructure platform to another.

**Infrastructure-as-a-Service (IaaS)**
Underlying compute, storage, network, and security services to enable higher level platforms and applications

**Platform-as-a-Service (PaaS)**
Platforms that enable teams to produce software such as development platforms, testing environments, etc.

**Software-as-a-Service (SaaS)**
End user software services often delivered on-demand via subscription or pay-as-you-go models
Internal Private Clouds
Hosted internal to the enterprise and secured from the public through internal networks and security infrastructure.

External Public Clouds
Hosted from data centers not owned or leased by the enterprise, and typically consisting of standardized offerings available to the general public.

External Private Clouds
Hosted from data centers not owned or leased by the enterprise. Consists of customized offerings tailored to a client or group of clients that are accessed securely over the Internet or via a secure dedicated network connection.

Virtual Private Clouds
One or more Internal Private Clouds and External Private Clouds that together form a composite solution. VPCs allow enterprises to deploy workloads across multiple resource pools to achieve the optimal balance of scale, cost, and performance.
2. Traditional IT Infrastructure Challenges

Traditional IT infrastructures exhibit a few common characteristics that limit their ability to adapt rapidly to change and introduce high costs and risks into IT initiatives including application upgrades and migrations. One common problem is siloed infrastructure that is designed and deployed for the needs of one application or project with limited or no thought towards repurposing the infrastructure afterwards. Enterprise data centers reflect this diversity today, with dozens of versions of operating systems, hundreds of makes and models of equipment, and thousands of unique configurations. Repurposing this infrastructure can be extremely time and resource intensive, with many systems not decommissioned at all, remaining well beyond their useful life.

Another key challenge relates to the lack of portability of software platforms and applications (collectively called “workloads different underlying infrastructure”) across different underlying infrastructure. Applications and platforms are bound to their underlying infrastructure are called “stateful.” These tightly bound application architectures limit agility, and require a lengthy reinstall effort with many labor intensive steps to achieve portability. This can lock you into specific vendors and their upgrade cycles, limit your ability to rapidly scale up or down, and miss opportunities to utilize lower cost vendors and providers.

Virtualization technologies are becoming more prevalent in the enterprise today, and although virtualization can help close the agility and cost savings gap, it is not sufficient in itself to address the underlying challenges. For example, virtualization can consolidate workloads onto a server, but these workloads are still bound to the underlying hypervisor or operating system, limiting their portability. Virtualization also doesn’t address automated workload management across multiple deployment environments. One common result is “VM sprawl”, where virtual machines cannot be provisioned and decommissioned fast enough to keep up with business needs, creating an overabundance of unmanaged VMs and a point of diminishing returns for infrastructure utilization and agility.
3. Optimal Infrastructure for “Everything-as-a-Service”

We believe that every large enterprise over time is going to transition their IT operating model towards a combination of internal and external cloud providers, and infrastructure, platform, and software as a service. A key enabler is an underlying IT infrastructure optimized for the new demands of automated workload management and utility-based pricing and consumption. This can be accomplished by sharing pools of modular components of compute and storage, combined with management software that orchestrates and dynamically configures workloads through fully automated, policy-based government frameworks.

These pools of common computing resource are designed to be delivered as on-demand Infrastructure-as-a-Service (IaaS), which can support higher level service offerings and diverse workloads including development platforms and end-user applications. The underlying infrastructure is critical in enabling the platforms and applications above it to benefit from core capabilities such as on-demand provisioning, automated policy enforcement, dynamically configured network and security services, elastic capacity scaling, and more. Together, the combination infrastructure, platforms, and software services can provide a highly efficient, cost effective and agile “everything-as-a-service” operating model that supports a broad range of enterprise IT needs.
4. Enterprise Adoption of Virtual Private Clouds

VPCs are poised to become a dominant implementation model of cloud computing for most large enterprises. Virtual Private Clouds allow enterprises to deploy platforms and software to internal or external cloud resource pools, or to straddle multiple resource pools, to achieve the optimal balance of scale, cost, and performance. There are multiple scenarios where an optimized and standardized VPC can be implemented by an enterprise and their external private cloud providers:

- An enterprise implements their VPC reference architecture internally within their datacenter.
- An enterprise specifies a VPC reference architecture for an external cloud provider to implement as a condition of doing business with them.
- An enterprise works with an external cloud provider to implement their VPC reference architecture through some combination of cost/risk sharing:
  - Commitments to purchase some level VPC capacity over a period of time.
  - Capital expenditure cost sharing.
  - Other arrangements.
- An enterprise implements their VPC reference architecture in an external cloud provider’s datacenter as an initial “skunk works” project, which is used as a lever to implement internal change and later roll out.

To help jump start the adoption of VPCs, ServiceMesh has included information on the requirements, reference architecture, and reference implementation of an optimized, state-of-the-art VPC. This information can assist organizations in several ways:

- Define enterprise-grade requirements and a corresponding enterprise-grade VPC reference architecture to support RFI/RFP creation and vendor selection efforts.
- Describe state-of-the-art VPC capabilities that may not have been previously considered to better assess VPC value and relevance in addressing business needs.
- Share VPC reference implementation metrics and results.
- Describe the schedule and timeline to procure, deploy, configure, and test the VPC reference implementation in this document.

Regarding the last point, many enterprises are understandably concerned with the perceived complexity and vendor induced confusion associated with cloud-based technologies such as
a VPC. However, this marketplace noise does not reflect a lack of ability to implement a fully functional enterprise-grade VPC in a matter of weeks. In our example, a small but experienced team ranging from 2 to 4 team members on a rotating basis was able to procure, deploy and test the VPC reference implementation in 7 weeks. This included three weeks for vendor selection and procurement, three weeks for system setup and configuration, and one week for testing. Of course, direct translation of this timeline into internal enterprise settings cannot be assumed. However, it does demonstrate that an initial VPC implementation scenario using experienced resources in a “green field” setting such as an external service provider's data center can be achieved in this approximate timeframe.
5. Requirements for an Enterprise-Grade VPC

VPCs exhibit some unique requirements to satisfy enterprise-grade customers that are typically not cost effective or practical to implement for external public clouds and a general purpose customer base. We define “enterprise-grade” as the ability to deploy business critical services that match or exceed the level of security, quality, reliability, and availability found in internal corporate data centers. These requirements include:

- High levels of automation, including automated provisioning and dynamic configuration of diverse workloads and multi-tier application topologies including web servers, application servers, database servers, security components (e.g. firewalls) and infrastructure components (e.g. DHCP).
- Unified workload management interface across all internal and external cloud providers.
- A vendor agnostic solution that is not locked into a hardware vendor, software vendor, or service provider.
- Portability to move running workloads from one physical host to another in an automated manner.
- The ability to create and enforce automated workload management policies for stateless computing architectures including:
  - Instance management (e.g. creating and deleting VMs, starting and stopping VMs, etc.)
  - Storage configuration and management (e.g. creating and deleting storage volumes, attaching volumes to VMs, initiating backups/snapshots, etc.)
  - Network configuration and management (e.g. VPN, NAT, VLANs, IP address overlays, etc.)
  - Security configuration and management (e.g. creating and deleting security zones, dynamically configuring firewalls in accordance with security zones, etc.)
  - User account management (e.g. authentication and authorization)
  - Service level management (e.g. instance scalability limitations, automated recovery, etc.)
  - Capacity management (e.g. auto-scaling, cloud-bursting, etc.)
- Encryption services for data in transit and for data at rest.
- Secure management of encryption keys (including private key certificates or pre-shared keys).
- Flexible reporting including support for billing and accounting processes and the capture of historical data for predictive usage models and analytical efforts.

  Integrated monitoring and metering of workloads, including monitoring of bandwidth, performance, latency, storage, etc., which can be provided through a portal or an API.
6. VPC Reference Architecture Overview

This VPC reference architecture addresses the previously described enterprise-grade VPC requirements in a modular, flexible design with minimal complexity. There are four major components to the VPC reference architecture that will be described in the sections below. Immediately following each section, we will briefly describe how that component was implemented in the VPC reference implementation. For readers seeking more detailed information, please contact ServiceMesh.

![Figure 1: VPC Reference Architecture Overview](image-url)
7. Workload Lifecycle Management

An enterprise-grade platform is needed to ensure consistent workload governance, security and lifecycle management across all IaaS, PaaS, and SaaS offerings and their cloud deployment environments. Key capabilities include automated workload management and self-service portals that enable users to bring workloads online and offline at the click of a button and allow resource consumption on-demand. However, lifecycle management requires far more than just starting and stopping workloads in the cloud. It also extends across an organization’s software development lifecycle, operations, security, and corporate governance. You need to intelligently govern, manage and secure your workloads as they progress across different roles and stages in the lifecycle to drive more significant cost savings and agility improvements.

The workload lifecycle management platform provides a single, consolidated interface to manage, govern, and monitor workloads deployed across internal and external clouds accessed by the VPC. This includes:

- Automation of the plan, build, share and run lifecycle for stateless workloads.
- Create and enforce policy-driven governance across infrastructure, platforms, apps and users.
- Provide consolidated workload monitoring and metering.
- Enable comprehensive security at the network, instance, data, and access levels including easy-to-use and secure key management.
- Provide federated identity management across deployment environments inside and outside the organization.
- Extend and enhance existing IT systems with integration into asset management, authentication/authorization, audit/governance, performance monitoring, chargeback/billing systems, and more.

7.1. PC Reference Implementation

The Agility Platform was used to provide a unified workload management platform for internal and external cloud providers, and supports the lifecycle management of stateless workloads. Agility Platform enabled the portability to move stateless workloads across deployment environments in a fully automated manner, and was used to create and enforce policies for user access, resource access, workload sizing, workload placement, security services, and infrastructure services.
Figure 2: Agility Platform
8. Automated Infrastructure as a Service

Configuring the underlying infrastructure of enterprise applications can be highly complex, with countless interdependencies among supporting network and security services, along with the complexity of the application topology. These multi-tier application topologies typically consist of multiple web servers, application servers, database servers, and other dependent services, each requiring a specific startup sequence and registration of services to ensure successful deployment of the system.

Automated IaaS tools orchestrate and automate the configuration of complex workloads and are responsible for:

- Setup and configuration of hypervisors and hardware
- Integration with enterprise services such as DHCP servers, DNS, proxy servers and others
- Creating local repositories for configuration data and settings
- Configuring workload-specific topologies including load balancers, app servers, etc.

These tools drastically reduce configuration time through automation, eliminate manual errors including mis-keyed scripts/commands, and ensure greater standardization and compliance.

8.1 VPC Reference Implementation

The Agility Platform, xCat, and VMware’s ESX were used for Automated IaaS functions:

- The Agility Platform automated the management and policy enforcement for workloads, scripts, and topology deployment. The Agility Platform’s virtual DHCP services were also used to control IP assignments and ensure consistent IPs for virtual servers if they failed and restarted.
- xCat, an open source initiative, provided orchestration to facilitate distributed computing management and the deployment and administration of clusters
- VMware ESX and VMware ESXi were used as bare-metal hypervisors that install directly on top of the physical server and partition it into multiple virtual machines that can run simultaneously, sharing the physical resources of the underlying server. Each virtual machine represents a complete system, with processors, memory, networking, storage and BIOS, and can run an unmodified operating system and applications.
Figure 3: Agility Platform

Workload security requirements can come from a broad range of sources both internal (such as keeping competitive information private) and external (such as government privacy regulations for customer data). Ensuring a secure computing environment while maximizing business agility and productivity is a difficult balance to maintain. Zoned security models help achieve this balance by ensuring that adequate security controls are in place while allowing IT practitioners and users as much latitude as possible within those controls. Zoned security models include a combination of capabilities:

- **Access security** including federated identity management and granular role-based access control to instances and asset repositories.
- **Network security** including encrypted overlay networks across internal and external clouds, and firewall integration with support for multicast, static IP management, point-to-point routing, and more.
- **Instance Security** including images with embedded host-based intrusion detection systems and virus scanning.
- **Data Security** including images that utilize configurable encrypted block storage as well as SDKs for non-block storage requirements.

9.1 VPC Reference Implementation

The Agility Platform and Altor Network’s virtual firewall product were used to implement a zoned security model:

- The Agility Platform managed encryption keys, logs, data encryption, instance security, and provided federated identity management. Agility Platform creates and enforces security zone policies that encompass VLANs, firewall rules, port groups, access lists, persistent storage, instance placement, and more.

- Altor Network’s hypervisor–based, virtual firewall product was used to inspect all packets to and from virtual machines, and allowed administrators to enforce virtual firewall policies for individual VMs, logical groups of VMs, or all VMs.
Figure 4: Agility Platform
10. Optimized Physical Infrastructure

Key objectives for the physical infrastructure are to support rapid implementation, reduced interoperability challenges, lower operating and support costs, and simplicity of design. The physical infrastructure uses a modular design that can accommodate incremental scalability and a wide range of power, storage, and space configurations. The infrastructure must accommodate a wide range of workloads, achieve high VM density, incorporate off-the-shelf components, and incorporate open standards:

- X86 for lower cost commodity hardware
- Ethernet to reduce interoperability challenges
- NFS to avoid the higher capital expense and operating costs associated with SAN solutions

10.1 VPC Reference Implementation

The reference implementation consisted of two computing racks to evaluate two different blade computing options, one storage rack, and a 10 Gbps Ethernet backbone:

- HP Compute Rack #1 consisting of HP BL460 G6 blade servers with 48 blades for a total of 576 cores (12 cores per blade). General memory configuration for 47 Blades is 96GB (or 4608 GB total), and the memory configuration for one blade for high end testing is 192 GB.
- Cisco Compute Rack #2 consisting of a Cisco B230 blade server with 16 cores, and a Cisco B200 blade server with 8 cores. General memory configuration is 96 GB, with the high end memory configuration at 255 GB.
- Oracle Storage Rack consists of an 7420 storage cluster with 100 TB of storage and 500 GB of memory.
- Networking consists of a 10 Gbps Ethernet backbone to ensure adequate performance of NFS storage and includes two Cisco 5548 10 Gbps Ethernet switches, each with 480 Gbps of network bandwidth. The reference implementation made heavy use of VLANs to isolate guest and management networks.
11. Reference Implementation Topology

A high level topology of the VPC reference implementation has been provided. This reference implementation was designed to meet the aforementioned enterprise-grade requirements including high availability.

There are multiple redundancies across each layer of the solution, including the ability to tolerate the failure of a storage node, network switch, and various other components. It’s designed so that a blade chassis can be taken offline for maintenance, and the guests can be seamlessly migrated to other chassis in the rack and back again.

A failure of a blade may result in an outage for the guests running on it, and the system can be configured to restart those guests on another blade as required. For solutions needing fault tolerance inclusive of the failure of a blade, this can also be configured. It is possible to deploy a similar or smaller VPC implementation and replicate to it for disaster recovery purposes. For systems requiring zero data loss, an active-active application topology can be deployed on top of the reference implementation.
Figure 5: Reference Implementation Topology
12. Reference Implementation Test Results

The VPC testing methodology and results are summarized below. Additional findings and results are available by contacting ServiceMesh.

12.1 Workload Types Used for Testing

- Virtual desktop workload (representing typical Office productivity apps on a Windows OS)
- Generic server workload (representing a mid-range business app running a LAMP stack)

12.2 Capacity Units Per Blade

Each ESX host has a different set of capacity units that can be utilized depending on the type of blade and workload. The two hardware platforms referenced in this document had capacity units defined as the reservation of 1GHz compute power and 2.5GB of memory. This is essentially equivalent to a guest accessing 1GHz of compute power using a modern core with 2.5GB of memory. These capacity units listed do not include the effects of hyper-threading or turbo boost.

There are two cluster configurations depending on the workload type. For server workloads, the typical constraining factor is CPU, but could also be memory or IO. For desktop workloads, the constraining factor is typically memory. Windows 7 and Windows Vista desktop were given 2GB memory reservations, while Windows XP was given 1GB. Capacity units can vary with different organizational needs, and many may find that 1.2GB of memory is enough for Window 7 virtual desktops.

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Server Capacity Units</th>
<th>Virtual Desktop Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco B200Blade(96GB)</td>
<td>22</td>
<td>60 Windows 7/Vista</td>
</tr>
<tr>
<td>Cisco B230Blade(255GB)</td>
<td>34</td>
<td>125 Windows 7/Vista</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250 Windows XP</td>
</tr>
<tr>
<td>HP Blade(96GB)</td>
<td>32</td>
<td>60 Windows 7/Vista</td>
</tr>
<tr>
<td></td>
<td></td>
<td>180 Windows XP</td>
</tr>
</tbody>
</table>
Enterprise-Grade Virtual Cloud from a State-of-the-Art Reference Implementation

| HP Blade (192GB) | 32 | 95 Windows 7/Vista  
|                 |    | 204 Windows XP (CPU constraints) |

Table 1: Capacity Units Per Blade

Capacity units were validated by analyzing the underlying CPU limits to guarantee performance. Each blade actually has additional capacity allocated in reserve to handle virtual firewall, hypervisor, and other functions. These capacity units assume no oversubscription of resources. It is possible to get a much higher numbers of guests if oversubscription is acceptable.

12.3 VM Density

VM density was extrapolated from blade capacity units and rack configuration, and was validated in the lab.

- Up to 1600 server workloads per compute rack
- Up to 600 desktop workloads per compute rack

12.4 VM Costs

A primary value proposition for VPCs is the ability to drive price arbitrage across a federation of internal and external cloud service providers. An accurate “apples-to-apples” comparison of internal and external cloud costs requires careful analysis, as a number of organization specific variable scan materially affect price.

**Key Finding:** The VPC reference implementation can cost up to 75% less than equivalent external cloud offerings depending upon the specific use case.

The team covered the VPC to equivalent external cloud providers under a variety of scenarios. Based on published cloud prices, the analysis showed that the VPC can cost up to 75% less than external cloud offerings depending upon the use case. A detailed breakdown of this analysis is beyond the scope of this document, however several high level factors are provided below for consideration.

- Network bandwidth costs must be carefully evaluated when using external providers. For internal clouds, the most applications will use internal network
bandwidth already allocated to the data center which will simply be redirected to the internal cloud using the same or less bandwidth.

- External cloud providers charge separately for storage. Network bandwidth costs for managing storage must be considered. There are also regulatory constraints, network latency impacts, and internal risk aversion to moving corporate data offsite when evaluating external storage options.
- Many external cloud providers offer a limited selection of high availability options. A clear understanding of high availability needs, including the percentage of workloads requiring high availability, is important to evaluate and compare external providers.
- Reduction in administrative headcount by migrating workloads to external providers requires analysis that is beyond the scope of this document. Cost savings must be balanced by the increased costs to maintain service provider relationships and support for new operational processes and roles.
- There is a wide range of SLA limitations associated with external cloud providers which need to be evaluated closely. In some cases, an internal cloud may end up being the de facto solution simply due to the lack of external cloud providers that can meet your SLA requirements.

### 12.5 On-demand Provisioning

The VPC uses ServiceMesh’s Agility™ Platform for automated provisioning of workloads in addition to other automated workload management functions. This is combined with the Agility Platform’s portal for self-service access to workloads and its policy enforcement engine to configure and secure each instance that is deployed. **As a result, the VPC can deploy server workloads from initiation of the workload request to running business functionality in approximately 4 minutes.** This delivers a quantum improvement in provisioning time relative to other approaches:
A broad range of provisioning scenarios was tested, including the simultaneously provisioning of multiple server instances to simulate peak demand for various sized production applications. The VPC demonstrated the ability to scale predictably and reliably at these volumes.

- 1 server instance = 4 minutes
- 25 simultaneous server instances = 13 min
- 50 simultaneous server instances = 22 min
12.6 Auto Scaling

In this test case, a pool of two server instances and a load balancer were load tested using a commercial load generating tool. Instance performance was monitored with the Agility Platform, and when CPU consumption for any instance went above or below an operating band of 30 to 85% capacity for a period of time, Agility Platform automatically provisioned additional instances or decommissioned existing ones as appropriate.

Figure 8: Auto Scaling
The system scaled smoothly both up and down as external load was applied. For abrupt increases in load, the system was able to auto-scale and bring all instances back within their operating band within approximately 60 to 90 seconds.

Figure 9: Auto Scaling
12.7 Dynamic Security Configuration with Virtual Firewalls

One of the benefits to using a virtual firewall is the ability to rapidly configure workloads across different hosts without compromising security. This requires security zone policies to be associated with each workload so that security settings can be dynamically configured as the workload is deployed across environments. This approach eliminates the manual administrative effort to configure IP addresses, port rules, and other settings using physical firewalls.

Virtual firewalls also offer the additional protection from guest-on-guest attacks in the same port group. This is significant as physical firewalls are ineffective inside virtual networks. With virtual firewalls, it is possible to create security boundaries at the guest level without having to resort to VLAN proliferation and sprawl. A brief comparison of the two security approaches is provided below:

<table>
<thead>
<tr>
<th>Instance Security Type</th>
<th>Configuration Time</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Firewall</td>
<td>Fully automated. 4 minutes per server instance during provisioning.</td>
<td>Security policies are enforced during instance provisioning. These policies include firewall, anti-virus and host intrusion detection configuration.</td>
</tr>
<tr>
<td>Physical Firewall</td>
<td>Typical 1-2 weeks for manual approval of new instance security settings for a new environment.</td>
<td>Manual approvals processes vary significantly between organizations, but typically cannot be accomplished in less than a week for new deployment environments.</td>
</tr>
</tbody>
</table>

Table 2: Dynamic Security Configuration with Virtual Firewalls
13. Summary

Virtual Private Clouds can enable a more agile and less costly IT operating model for enterprise customers. Our objective is to help accelerate enterprise adoption of VPCs with a flexible, vendor agnostic solution that meets enterprise-grade requirements. To support this objective, this document provides a high level overview of enterprise-grade VPC requirements, reference architecture, and reference implementation results which were tested and validated in a lab environment. These results showed that VPCs can be rapidly implemented, deliver high levels of automation and time savings, and operate VMs at costs up to 75% less than comparable public cloud offering. This paper establishes a bar for a current state-of-the-art VPC, and represents a foundation for continued improvements in the future.
14. References

14.1 IPR Releases and Patent Disclosures

This document may involve a claim of patent rights by one or more of the contributors to this document, pursuant to the Agreement on Intellectual Rights between the TM Forum and its members. Interested parties should contact the TM Forum office to obtain notice of current patent rights claims subject to this document.
15. Administrative Appendix

15.1 Document History

15.1.1. Version History

This section records the changes between this and the previous document version as it is edited by the team concerned. Note: this is an incremental number which does not have to match the release number and used for change control purposes only.

<table>
<thead>
<tr>
<th>Version Number</th>
<th>Date Modified</th>
<th>Modified by:</th>
<th>Description of changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>11/9/2012</td>
<td>Mary Amalfitano</td>
<td>Format document</td>
</tr>
</tbody>
</table>

15.1.2 Release History

This section records the changes between this and the previous Official document release. The release number is the 'Marketing' number which this version of the document is first being assigned to.

<table>
<thead>
<tr>
<th>Release Number</th>
<th>Date Modified</th>
<th>Modified by:</th>
<th>Description of changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;&lt;=Release Number &gt;&gt;</td>
<td>DD/MMM/YY</td>
<td>&lt;&lt;=name&gt;&gt;</td>
<td>Description e.g. first issue of document</td>
</tr>
</tbody>
</table>

15.2 Company Contact Details

<table>
<thead>
<tr>
<th>Company</th>
<th>Team Member Representative</th>
</tr>
</thead>
</table>
| ServiceMesh      | Anthony Skipper  
Vice President of Infrastructure and Security  
anthony.skipper@servicemesh.com |
| ServiceMesh      | Mark Thiele  
Vice President of Data Center  
Technologies  
mark.thiele@servicemesh.com |
| ServiceMesh      | Derick Townsend  
Senior Director of Marketing  
derick.townsend@servicemesh.com |
15.3 Acknowledgments

This document was prepared by the members of the TM Forum ECLC team:

Anthony Skipper, Vice President of Infrastructure and Security, ServiceMesh, Inc.
Mark Thiele, Vice President of Data Center Technologies, ServiceMesh, Inc.
Derick Townsend, Sr. Director of Marketing, ServiceMesh, Inc.

The authors can be contacted at knowledgecenter@servicemesh.com.