Nutanix Complete Cluster Reference Architecture for Virtual Desktop Infrastructure

with

vmware®

VMware vSphere 5 and View 5

Table of Contents

1. Executive Summary

Introduction

This document provides a reference architecture for deploying virtual desktops using the Nutanix Complete Cluster for the virtualized infrastructure. The testing was completed using VMware vSphere 5 as the hypervisor and VMware View 5 as the Virtual Desktop Infrastructure (VDI) connection broker.

This reference architecture is intended for use by customer, field and reseller architects and engineers who will be configuring VDI environments. The goal of this reference architecture is to demonstrate the ease of deployment of a large-scale VDI environment using these products, with detailed performance and configuration information of the testing that was done in a lab environment.

Unlike many VDI reference architectures that are available today from other vendors, the Nutanix approach of modular scale-out enables customers to select any pod size and grow in more granular 50-75 virtual desktop increments. This removes the hurdle of a large up-front infrastructure purchase that a customer will need many months or years to grow into, ensuring a faster time to value for the VDI implementation.

The Nutanix Solution for Virtual Desktop Infrastructure (VDI)

The Nutanix Complete Cluster is a converged virtualization appliance for desktop and server virtualization that merges enterprise storage with x86 servers into a single high-performance tier. This innovative converged appliance approach delivers the only truly modular way to grow your datacenter, resulting in predictable performance and radically simple scalability with each unit of growth, much like a Google or Facebook datacenter. Each modular "block" includes compute, storage and networking along with the vSphere 5 hypervisor preloaded so customers can start provisioning virtual machines in less than 30 minutes.

A Nutanix Block ships as a rackable 2U unit containing four high-performance server nodes with local storage to run and store virtual machines. Fusion-io ioDrives combined with heat-optimized tiering of data deliver the high performance of SSDs at the cost of hard drives to tackle VDI performance issues like boot storms with ease. This design greatly reduces overall cost and complexity while increasing performance and scalability in a fully integrated VDI-in-a-box solution.

VMware View 5

VMware View enables the simplification of desktop and application management while providing an optimized user experience with high security. View allows IT to centrally manage desktops, applications, and data while increasing flexibility and customization at the end-point for the user. This enables higher availability and agility of desktop services unmatched by traditional PCs while reducing the total cost of desktop ownership by up to 50%.

2. Introduction to Project Colossus

Project Colossus is a fifty-node Nutanix cluster that is used to demonstrate the linear scalability of the Nutanix architecture from four nodes (one Block) to fifty nodes (12.5 Blocks or 3/5 of a rack). In previous benchmarking, Project Colossus delivered 375,000 random write IOPS and 224 Gbps of sequential write bandwidth.

In this reference architecture, we use the same configuration of 50 nodes to validate a large-scale VMware View 5 environment. In architecting this environment, we created a real world architecture that stressed both the server and storage capabilities of the Nutanix Complete Cluster architecture.

In this reference architecture, we will demonstrate the ease of scaling from one Nutanix block of 300 virtual desktops to ten blocks hosting 3,000 virtual desktops.

2.1. Problem Statement and Purpose

Sizing and architecting a virtualized infrastructure deployment to ensure adequate performance has been a consistent challenge since companies started virtualizing their servers over a decade ago.

Building a new virtualized infrastructure environment for a use case like VDI usually requires planning and budgeting for a system that will support current and future needs. Administrators would rather have excess capacity and performance versus not enough, which may result in oversizing.

Once the infrastructure has been created, growing it can be a challenge. The need for more compute resources requires the addition of multiple servers. As more storage is needed, more disks are added to the SAN. As this type of environment grows, bottlenecks are often created. Typically, the remedy is to upgrade the storage controllers or the infrastructure. In addition to being very expensive to build and maintain, this can create a piecemeal model that doesn't scale easily and incurs high costs.

Nutanix seeks to address these challenges with a scalable and predictable virtualization appliance. The Nutanix Complete Cluster contains the compute, storage, and networking in one package for turnkey virtualization. This allows the administrator to grow the virtual infrastructure in a much more efficient manner, adding a node at a time for a more granular scale-out of the virtualized environment.

Many vendors offer pre-packaged virtualized infrastructure stacks that integrate existing compute, storage and networking products, but only Nutanix has a truly converged, scale-out solution that merges these individual tiers into a single tier for high-performance virtualization. The smaller building block size enables customers to buy only what they need, instead of buying and re-architecting a large infrastructure stack that is both costly and may be more capacity than is actually needed.

With use of the VMware Reference Architecture Workload Simulator (RAWC), we set up an environment that scaled in one block increments to demonstrate how a customer's initial deployment of 300 VMs could grow to 3000 VMs in a plug-and-play fashion with zero redesign and linear performance scaling.

3. Nutanix Complete Cluster Setup

The Nutanix Complete Cluster is comprised of Nutanix Blocks that are rackable 2U units that contain four high-performance server nodes. Each of the ten Blocks in this reference architecture is configured as follows:

4 x86 Server Nodes Per 2U Nutanix Complete Block	
Per x86 Server Node	**Per Nutanix Complete Block**
Dual 6-core Intel Xeon X5650 CPUs	8 CPUs/48 cores
96GB RAM	384GB RAM
320GB Fusion-io PCIe SSD 300GB SATA SSD 5TB HDD (5x1TB; 7200 rpm SATA)	1.3TB Fusion-io PCIe SSD 1.2TB SATA SSD 20TB HDD (15x1TB; 7200 rpm SATA)
Power for Block: Dual (Active-Active) Configuration 1400W:180–240V, 50–60Hz, 7.0–9.5 Amp	

Figure 1 - Nutanix Hardware

4. Networking Setup

The Nutanix Complete Cluster uses 10 Gigabit Networking and is configured per VMware recommended best practices. Multiple VMkernel ports were configured to isolate the different types of network traffic.

Figure 2 below illustrates the networking setup that can be used for multiple pods of Nutanix Blocks to scale from 3,000 desktops to 12,000 desktops. Note that pods can be of any desired size.
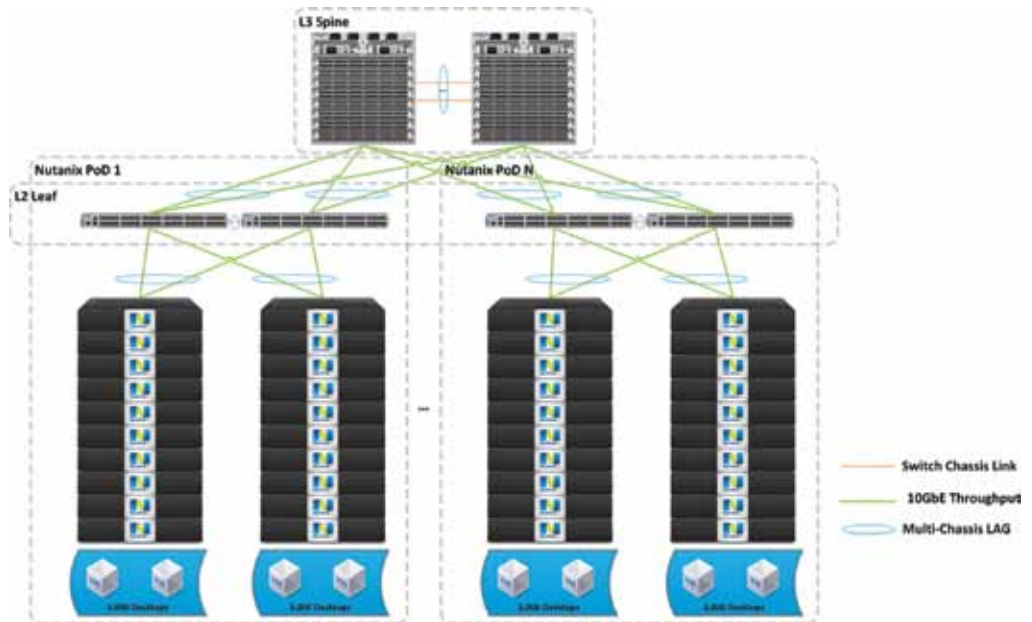

Figure 2 - Network Diagram

5. Storage Setup

The Nutanix Complete Cluster is unique in its grow-as-you-go scalability, enabled by its scale-out architecture. Because the Nutanix virtualization appliance uses internal storage, the need for an external storage network or SAN is removed from the architecture. This removes the bottlenecks that are created when all the storage traffic needs to route from the ESX hosts to the SAN.

Nutanix Distributed File System (NDFS) is at the heart of Nutanix' clustering technology. Inspired by the Google File System, NDFS delivers a unified pool of storage from all nodes across the cluster, leveraging techniques including striping, replication, auto-tiering, error detection, fail-over and automatic recovery. This pool can then be sliced and diced to be presented as shared-storage resources to VMs for seamless support of features like vMotion, HA and DRS, along with advanced data management features. Additional nodes can be added in a plug and-play manner in this high-performance scale-out architecture to build a cluster that will easily grow as your needs do.

NDFS uses virtual storage controllers, one per node, which control the management of local storage of each node. In Figure 3, we illustrate how these storage controller VMs communicate with the other ESX hosts (nodes) in the cluster create a clustered storage system. The Fusion-IO ioDrives create the top tier of primary storage to provide high performance for most recently accessed data, and when the data becomes cold, the heat-optimized tiering feature automatically migrates that data to the lower tier of SATA HDDs.
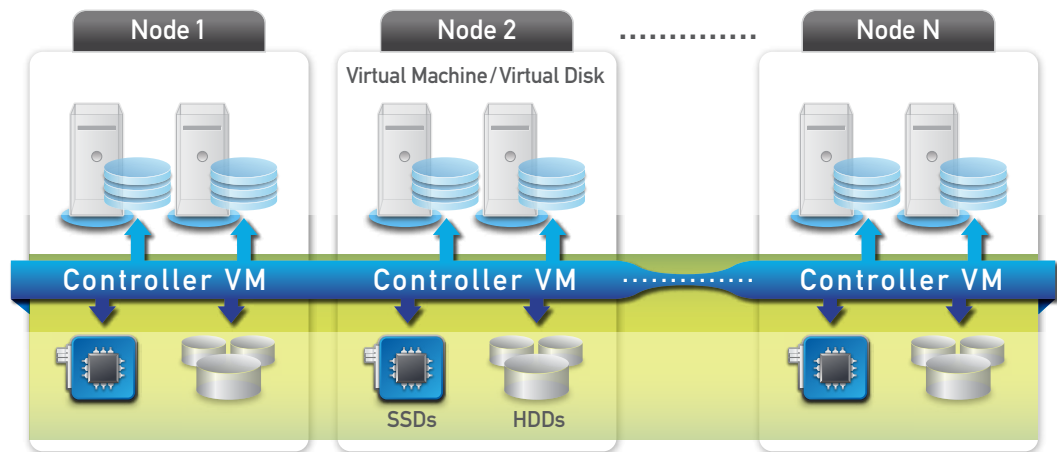

Figure 3 – Nutanix Distributed Scale-Out Architecture

@nutanix
www.nutanix.com

Before the storage configuration details are discussed, it is good to know how the Nutanix Complete Cluster's storage differentiates from traditional storage arrays.

Shared Nothing Distributed Architecture

The Nutanix Complete Cluster is a pure distributed system. This means that the compute gets its storage locally, and does not need to traverse the network. All the problems around network, interconnect and core switch bottlenecks are avoided completely.

Distributed Metadata

A big problem affecting scalability in a traditional SAN is metadata access. Most storage arrays have 1-4 controller heads and all metadata access needs to go through these heads. This causes contention issues and performance drops as more servers try to access the same storage array.

In the Nutanix Cluster, metadata is maintained in a truly distributed and decentralized fashion. Each of the nodes in the Nutanix Cluster maintains a part of the global metadata, which means that there is no single bottleneck for metadata access.

Distributed Metadata Cache

Most storage arrays do a very poor job of maintaining metadata caches. In storage arrays that do have a metadata cache, the caches live on the limited number of controller heads. Access to the cache is limited by the bottlenecks around the network, interconnect and switch as discussed above. In the Nutanix Cluster, metadata is cached on each of the controller VMs.

Most metadata access is served up by cache lookups. Each controller VM maintains its own cache. This means that however large the Nutanix Cluster grows as you continue to add nodes, the performance cost of metadata access stays the same.

Lock-free Concurrency Model

The standard approach to ensuring correctness for metadata access is to use locking. Unfortunately, in a distributed system, locking can become very complicated. Excessive locking can better ensure correctness, but creates significant performance drag.

The Nutanix Cluster implements an innovative lock free concurrency model for metadata access. This model ensures the correctness of metadata but at the same time ensures high performance.

Distributed MapReduce for Data/Metadata Consistency

For a large scale deployment, consistency checking for data and metadata becomes a challenge. The Nutanix Complete Cluster implements a fully distributed MapReduce algorithm to ensure data and metadata consistency. The distributed nature of the MapReduce implementation ensures that there is no single bottleneck in the system. MapReduce has been shown by Google to scale to 1000's of nodes, and is a key ingredient in the incremental scalability of the Nutanix Cluster.

Distributed Extent Cache

Caching is not a new concept to storage arrays. The challenge however, is that caches are located on the limited number of storage controllers in a storage array. Not only does the limited of storage controllers cause contention, but the fact that these caches live in the storage array means that access to cached data needs to traverse the core switch. This brings latencies around the network, interconnect and switch that we discussed above. The Extent Cache caches the data served up by the controller VM from local storage, and similar to the Metadata Cache, leverages the Fusion-io ioDrives for high performance. The compute tier can easily access the cache locally without having to hop across the network and the core switch. This approach helps enable incremental scale-out while maintaining high performance.

6. VMware View Setup

6.1. VMware View 5 Configuration

The VMware View configuration settings used during all performance tests are shown in Table 1 below.

View 5 Configuration	Details
Pool Definition	
Desktop Pool Type	Automated Pool
User Assignment	Persistent Desktops: Dedicated with automatic assignment
vCenter Server Desktop VM provisioning type	View Composer Linked Clones
Pool Settings	
Remote Desktop Power Policy	Take no power action
Default Display Protocol	PCoIP
View Composer Disks	
Persistent Disk	All changes (writes to OS, temporary system and application and user data) are saved to persistent disk
Provisioning Settings	
Provisioning	Provision all desktops up-front
Datastores	OS Disks on separate datastores from replica disks. Storage over-commit set for conservative (1) 100GB VMFS 5 Datastore for each replica (8) 2TB VMFS 5 Datastores for Linked clones

Table 1 - View 5 Pool Configuration

6.2. Guest VM Setup - Windows 7 Virtual Machine

The operating system chosen for the testing was Windows 7 Enterprise x86.

Several desktop configurations were considered based on the virtual desktop user types. There are three basic types of Desktop Users: Task Workers, Knowledge Users and Power Users. The most common type of virtual desktop users are Knowledge Users. Our Knowledge User desktop configuration assumes approximately 9-12 IOPS average workload and 20 IOPS Peak during Boot and Logon.

The Windows 7 virtual machines for Knowledge Users were configured as follows:
- VMware Hardware Version 8
- Windows 7 SP1 x86 Enterprise Edition
- 1GB Ram
- 2 vCPU
- 40GB HDD (Thin Provisioned)
- 1 vNIC (VMXNET 3)

The virtual machine also had the following software installed.
- Microsoft Office 2007 with
  - Outlook
  - Word
  - PowerPoint
  - Excel
- Adobe Acrobat Reader 9.4

The Windows 7 Desktop Gold Image was optimized using the VMware View Optimization Guide for Windows 7.

7. Testing

The VMware Reference Architecture Workload Simulator (RAWC) 1.2 Tool was used to create a real world workload for performance testing of the virtual desktops.

The VMs were provisioned using VMware View 5 View Composer's Linked Clone technology.

7.1.   Tools

RAWC generated all workloads in the environment, as indicated by the screenshot in Figure 6. The selected applications were chosen in collaboration with VMware to generate an application mixture that matched a Knowledge User workload of 9-12 IOPS/desktop. The following information is also significant:

- • SendMail on a Linux VM was used for IMAP e-mail access.
- • The virtual desktops were set to run Outlook in cached mode.

All logins with workload scenarios were conducted in the same manner. Users logged in over PCoIP through RAWC. On average, 5 logins were initiated every 3 seconds, and all logins were completed within 28 minutes. To achieve this, 60 RAWC launchers were used. The launchers were each configured to log into one virtual desktop every 15 seconds. The launchers were logged in themselves every 5 seconds serially. After 3 minutes, all launchers were setting up new PCoIP sessions concurrently every 15 seconds.

The launching of a test run simulates a morning logon boot storm and the test workload would simulate a user workload. The testing used the default login rates as defined in the RAWC Administration Guide (available to VMware partners). The order in which the applications were run was random on each virtual desktop.
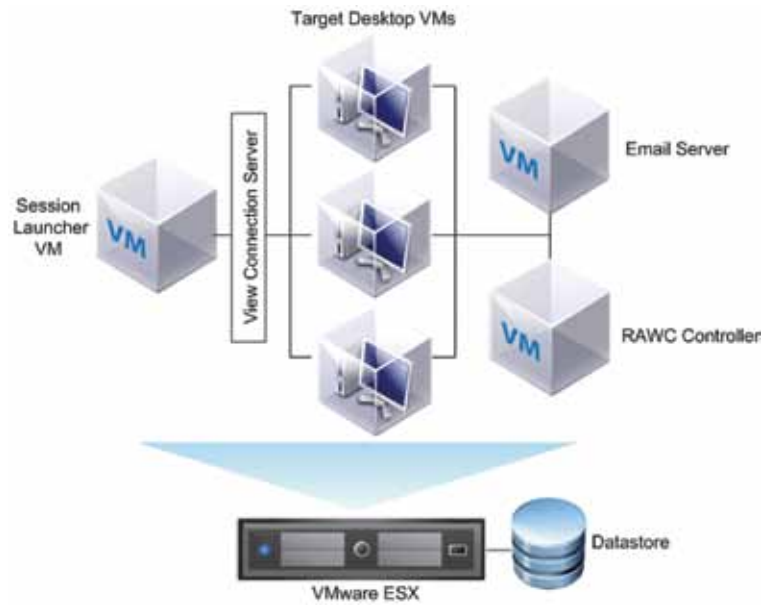
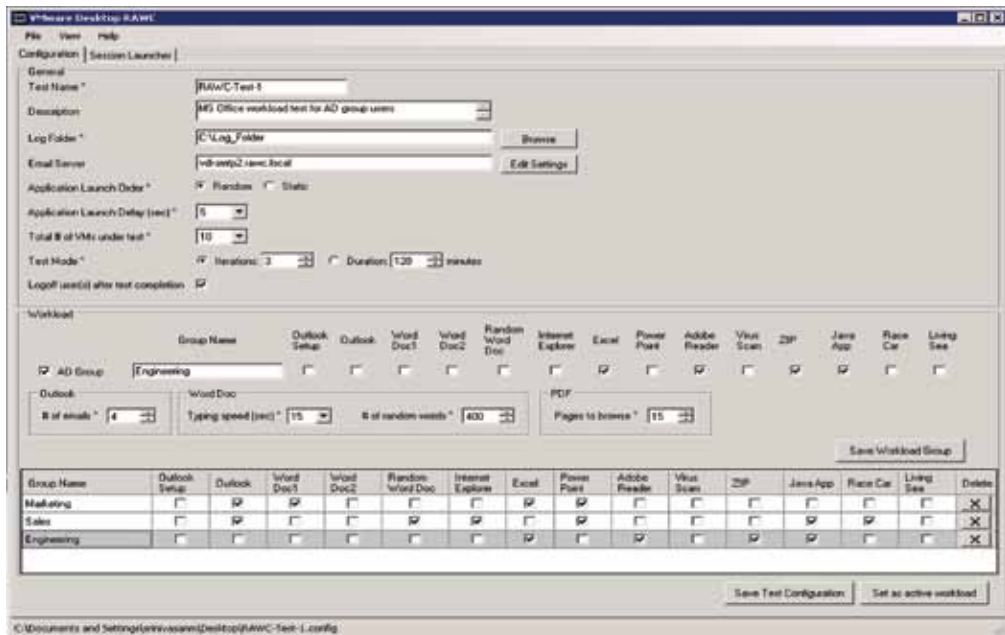*Figure 4 - RAWC Test Architecture (Courtesy of VMware)*

Figure 5 - RAWC Screen Showing Workload Configuration

TESTING

@nutanix
www.nutanix.com

7.2. Procedure

Shared storage was provisioned for the VMs to enable use of VMware advanced features such as HA, DRS and vMotion. One Replica Datastore was created per VMware cluster and one Linked Clone datastore was created per ESX host (Nutanix Node).

VMware View Composer has a limit of 8 ESX hosts per View Linked Clone Pool. We created multiple VMware clusters with 2 Nutanix Blocks (4 nodes/block) each.

The process of provisioning the datastores to the ESX servers is simple and can be done through the Nutanix Management Interface.

1. Create Shared VMFS LUNS.

    a. Create (4) 2TB VMFS DataStores per Block

    b. Create (1) 2TB VMFS 5 Datastore for Replica Disks per VMware Cluster (2 Blocks)
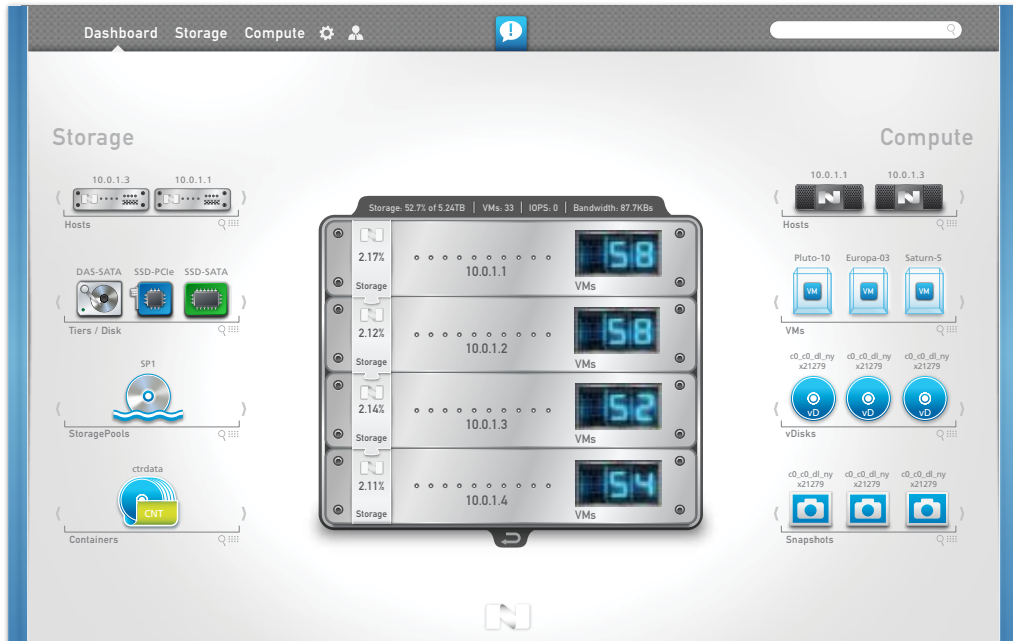
    c. Format VMFS 5

Figure 6. Nutanix Management Console: Converging Compute and Storage

Figure 7 - Nutanix UI Wizard for Provisioning Datastores

TESTING

2. Create 300 Linked Clones per Block using VMware View 5 Connection Server and VMware Composer 2.7:

    a. Select the Gold VM and set destination of replica to Replica Datastore

    b. Create 4 Dedicated Pools Per Block

    c. Create 150 VMware Linked Clones from each Gold replica to the Linked Clone Datastore

    3. Run RAWC and gather performance data

    4. Add additional Nutanix Block (Four Nutanix Nodes)

    5. Run next RAWC workload test.

After all tests were run, the RAWC tool generated the results data, which were compiled and charted in the next section.

Based on whether the user profile is that of a task, knowledge or power user, the desktop ratios per Nutanix block may be greater or less than 300 VMs per block. Power users' desktops may require as much as twice the CPU and storage of the knowledge worker profile that was used during this reference architecture testing, so advanced assessment and planning will need to be conducted for every VDI environment to ensure adequate performance. Keep in mind that each Nutanix block consistently delivers 25,000 random IOPS, so this process of mapping user requirements to the virtualized infrastructure will be much simpler than with traditional approaches that require balancing of server, SAN and networking components.

8. Test Results

The Nutanix Complete Cluster performed in a predictable scale-out fashion as additional Nutanix Blocks of 300 VMs were added. The test results below demonstrate the linear performance in consistent application response times from 300 to 3000 desktop VMs. Single nodes, instead of blocks of four nodes, could have been used as well.

Note: The times listed below in Figure 8 were collected from the VMware RAWC benchmarking tool, and represent response times that yield a user experience roughly equivalent to that of a local desktop experience.
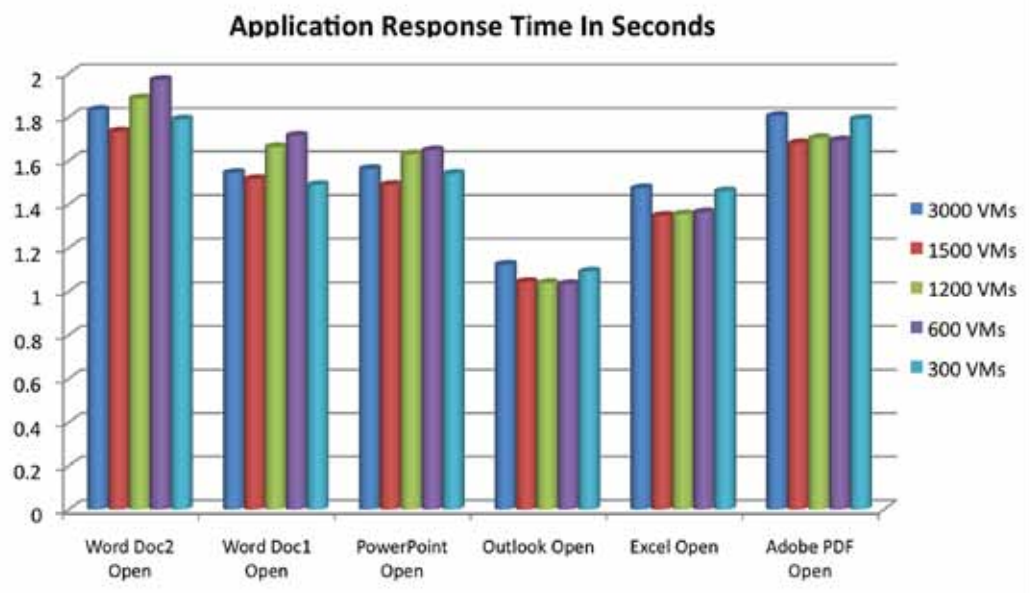
Application Response Time In Seconds

Figure 8 - RAWC Performance Chart of 300 to 3000 Desktop VMs on Nutanix Cluster

9. Conclusion

In this reference architecture, we have demonstrated that the Nutanix Complete Cluster delivers a modular scale-out solution for VDI infrastructure. The performance results prove the linear performance as the cluster grew with almost no difference in application response times of 300 VMs versus 3000 VMs. The ability to grow-as-you-go has tremendous benefits over the traditional virtualization infrastructure architectures in addressing two key problems of VDI deployments: high upfront cost and performance.

By scaling from 300 to 3000 virtual desktops, this validation shows the ability to start with a small VDI environment and expand easily. Whether it is a VDI proof of concept or a large-scale enterprise deployment, administrators are able to grow this environment easily and efficiently. Traditional server and SAN approaches require sizing for longer term capacity, which often results wasted or limited capacity. Expanding a Nutanix Cluster is as easy as adding an additional block for a more granular growth model that can deliver ROI in a much shorter time span.

Along with the grow-as-you-go scalability ease, another key benefit is the predictable linearity of performance that allows for easier performance design of VDI deployments. This validation shows there is guaranteed predictable performance going from 300 desktops to 3000 desktops, with an application response time difference of less than .3 seconds between the slowest and fastest times on all application tests. Nutanix delivers fast random IO as well as high sequential bandwidth, providing desktop users with a great experience in steady state and in the face of boot storms.

Another advantage demonstrated with this reference architecture is the ease of management for a VDI installation. The intuitive Nutanix user interface makes it easy to manage entire virtual environments without separate consoles for server and storage. This converged approach enables administrators to easily see the environment's capacity, utilization and overall health.

Ease of manageability combined with the high performance and 40-60% cost savings that the Nutanix solution delivers for virtualized infrastructure offers customers seeking to deploy VDI a new approach that removes the complexity that has held many deployments back. Nutanix combined with VMware View provides a radically simple approach to VDI that enables customers to realize the ROI on Windows 7 migration, better security and simpler desktop management much more rapidly than with traditional server and SAN products.

10. References

10.1. Table of Figures

10.2. Table of Tables

11. Authors and Acknowledgements

This reference architecture was designed and tested by Scott Werntz, Technical Marketing Engineer, and Steve Poitras, Solutions Engineer, at Nutanix.

Nutanix would like to thank the VMware View Technical Marketing team, including Mason Uyeda, Mac Binesh and Fred Schimscheimer, for their valuable guidance and ongoing collaboration during this project.