Technical Report

NetApp StorageGRID Point of Deployment (POD) Architecture

David Bickle and Albert Chen, NetApp
March 2011 | TR-3885

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The NetApp® StorageGRID® point of deployment (POD) is a deployment architecture combining NetApp storage systems with VMware® vSphere® hypervisors running StorageGRID software. The POD facilitates the rapid deployment, configuration, and support of an integrated solution for StorageGRID in an enterprise environment.

This document provides a high-level summary of the StorageGRID POD, including design, architecture, and integration. This report explains StorageGRID POD concepts to facilitate an understanding of StorageGRID among NetApp field sales, customers, partners, and resellers.

Users of this guide should already be familiar with NetApp storage solutions, NetApp StorageGRID, and VMware vSphere.

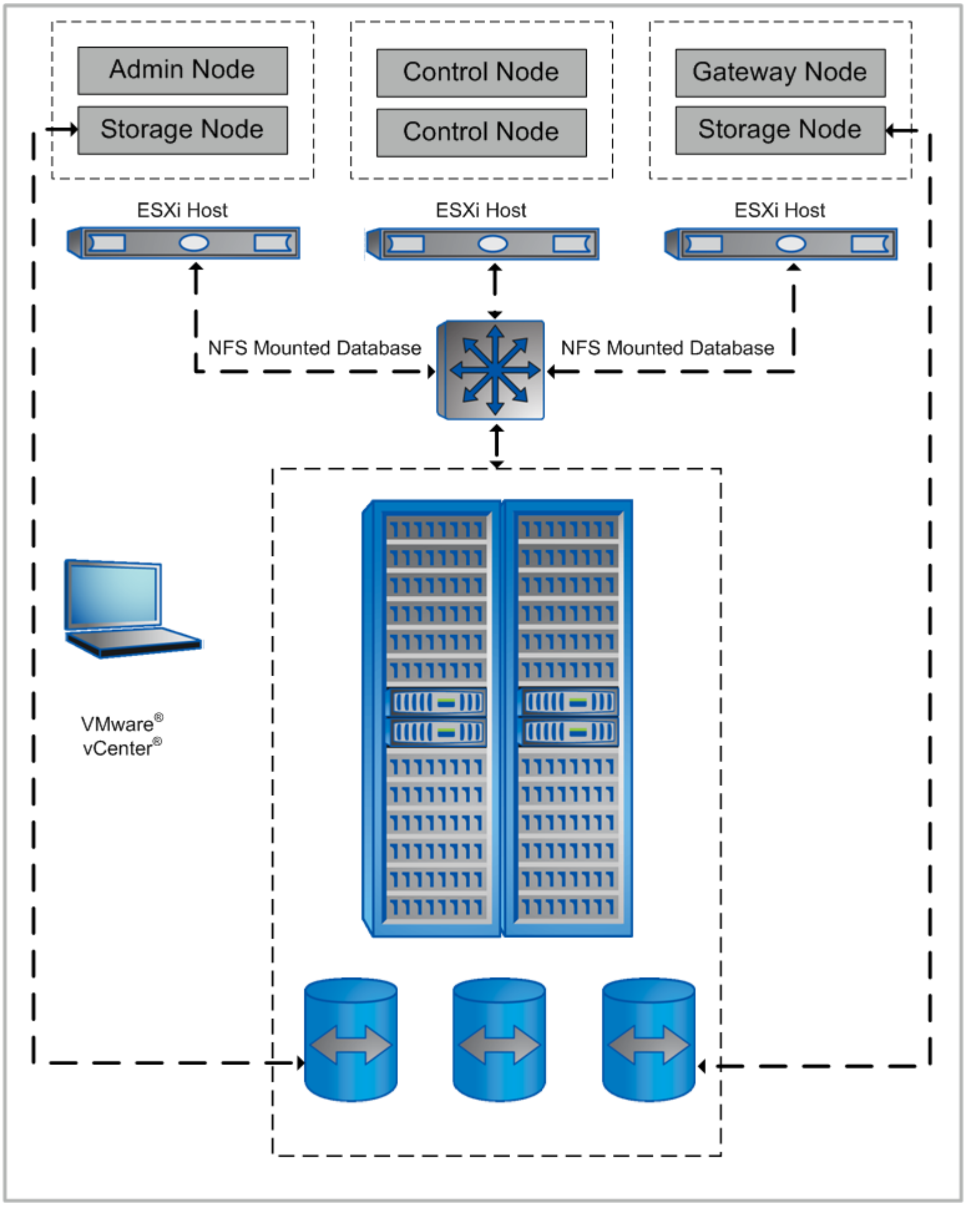
2 POINT OF DEPLOYMENT ARCHITECTURE

A POD, which is a combination of software and hardware used to deploy a StorageGRID solution, is composed of:

- NetApp StorageGRID software hosted on VMware vSphere
- NetApp Data ONTAP® software
- NetApp storage system

A deployment of the StorageGRID solution includes a minimum of two PODs.

Figure 1) Single instance of a POD.



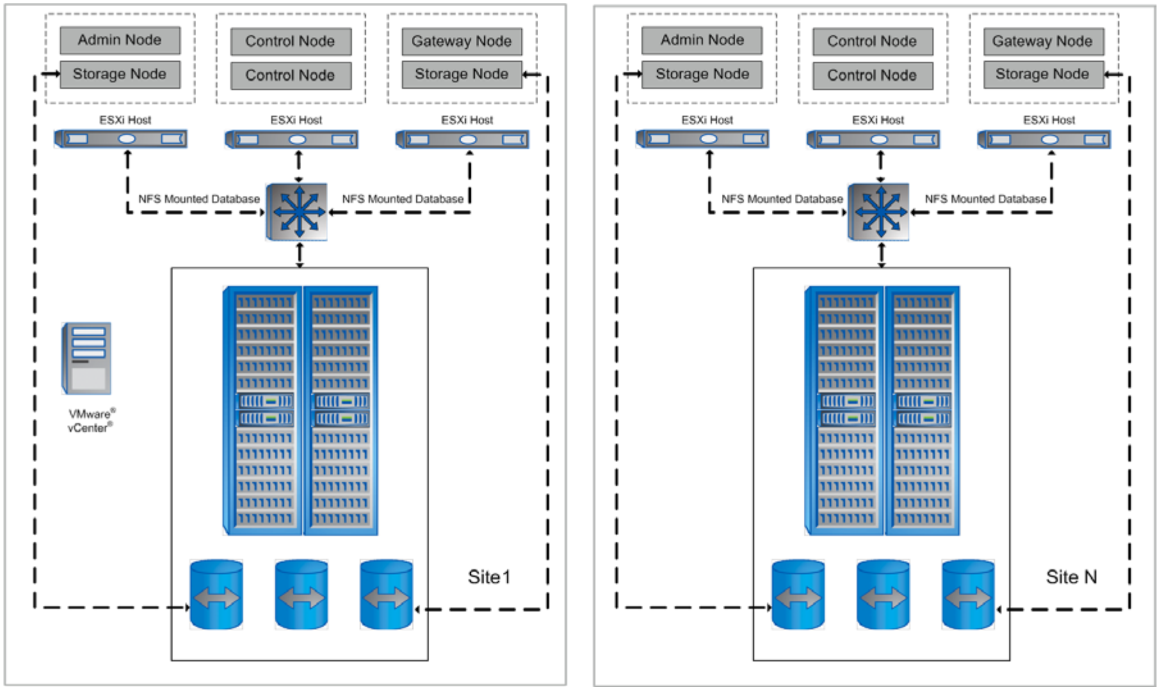
As shown in Figure 1, the StorageGRID portion of a POD typically consists of several servers with an ESXi hypervisor running on each server. Each ESXi hypervisor can host several virtual machines with each virtual machine hosting one grid node.

The StorageGRID portion of a POD includes all grid nodes and grid services necessary for the successful deployment of a StorageGRID system. When deploying a StorageGRID solution, each instance of a POD is bound by all current StorageGRID design constraints, including gateway node, control node, and storage node counts.

Note: Any POD within a site can display unique characteristics based on customer requirements, including data profiles, high availability, and performance.

As shown in Figure 2, in a multisite deployment (data center plus disaster recovery) of the StorageGRID solution, each site includes at least one POD. Grid nodes are interconnected using a pair of network switches providing a fully redundant configuration for failover. All of the storage for the virtual machine hosts is mounted from the NetApp storage system.

Figure 2) Multisite (Site 1 + Site N) deployment.



PODs are connected across sites using the customer network, where data replication, failover, migration, and recovery can occur across a network infrastructure.

POD designs vary depending on customer requirements for application, topology, data profiles, replication, information lifecycle management (ILM), and performance. Currently, there are two POD types:

- FAS3170HA-based configuration with 150TB and greater of attached storage and 10GbE or 1GbE interconnect
- FAS2040HA-based configuration with 20TB to 120TB of attached storage and 1GbE interconnect

Two types of replication groups exist within both POD types: the gateway replication group and the content management service (CMS) replication group. The gateway replication group has an initial capacity of 200 million objects. This capacity can be extended in increments of 200 million objects. The content management service (CMS) replication group has an initial capacity of 200 million objects. This capacity can be extended, in increments of 200 million objects, up to 2 billion objects.

2.1 POINT OF DEPLOYMENT EXTENSION

The demands of the StorageGRID system increase over time. Therefore, the grid must be expanded by extending the POD configurations. This extension consists of adding new servers and grid nodes to the StorageGRID system. The following sections describe the grid extension for POD configurations.

The extension of a POD includes the addition of physical servers hosting VMware ESX or ESXi and the addition of storage system capacity. The storage system needs to increase both SAS and SATA disks to support the additional grid nodes.

POD extensions are unique to each POD deployment option (FAS3170HA or FAS2040HA configuration). Each deployment option has a number of extension alternatives depending on data profiles. Control nodes can be added to an existing POD to increase metadata processing and capacity. Gateway nodes can also be added to PODs for additional application throughput. Archive nodes provide a middleware layer enabling PODs to connect to external archiving solutions.

2.2 STORAGEGRID SYSTEM

Within the POD architecture, the NetApp StorageGRID system leverages virtualization technology to make efficient use of the StorageGRID system's server and storage infrastructure. With the StorageGRID system:

- All StorageGRID services run within a virtualization layer: Grid nodes, grid services, and the underlying operating system run on a virtual machine hosted by a hypervisor.
- Each grid node makes use of its own virtual machine resources. Grid nodes cannot share a virtual machine resource.
- All grid nodes are configured to mount storage volumes residing on storage systems. System and database volumes are mounted as VMware virtual disks, backed by NFS-exported SAS storage on the storage systems. All storage volumes are mounted directly through NFS on storage nodes and reside on SATA storage.

2.3 STORAGEGRID OPERATIONAL LAYERS

The proper layout and configuration of the various operational layers that make up the StorageGRID system are central to the deployment of a POD. The following sections summarize the system, network, virtualization, and StorageGRID software layers.

SYSTEM LAYER

The system layer consists of all hardware resources required to support the virtualization layer. These hardware resources include server resources (CPU and memory) and NetApp storage systems (disks, aggregates, and volumes).

NETWORK LAYER

The network layer is responsible for interconnecting system resources to provide sufficient reliability, failover, and performance. This layer includes network design considerations such as VLAN partitioning, VLAN tagging, and failover using link aggregation. The StorageGRID system relies on the proper configuration of the network and virtualization layers to support transparent system failover and performance.

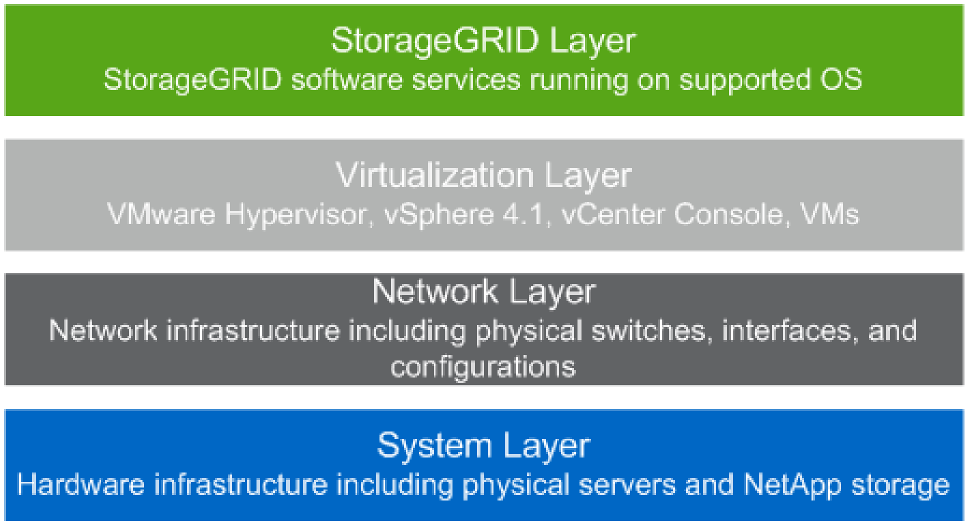
VIRTUALIZATION LAYER

The virtualization layer allows consolidated and efficient system resources. This layer also helps to reduce the complexity of the overall design of StorageGRID. In addition, the virtualization layer aggregates StorageGRID services across network resources for efficient throughput and reliability. The virtualization layer must be configured to make the best use of the underlying physical resources for computational, storage, and networking tasks.

STORAGEGRID SOFTWARE LAYER

The StorageGRID software layer consists of the software required to run the object storage, metadata management, and ILM capabilities of the StorageGRID system. This layer includes all StorageGRID software services and components, including the SUSE Linux® Enterprise 10 64-bit operating system.

Figure 3) StorageGRID operational layers.



2.4 NETWORK LAYER

A supporting network is required to interconnect the system layer. The StorageGRID system requires separate network interconnects for system management, customer, grid, and storage access. The following sections summarize these networks and their design constraints.

MANAGEMENT NETWORK

All VMware ESXi servers require remote access to configure, manage, monitor, and audit VMware services. Because management traffic is separated from the rest of the grid, separating this network traffic protects both the security and performance of the grid. Access to this network should be restricted to StorageGRID administrators and network operators.

PRIVATE NETWORK

The private network is responsible for forwarding traffic among grid nodes. This traffic is internal to the StorageGRID system and includes data transfers and messaging traffic. Maintaining a separate network for grid traffic protects system performance and security, because intragrid node traffic remains within the StorageGRID system.

CUSTOMER NETWORK

Some grid nodes have separate interfaces to facilitate customer network access to grid management interfaces, file shares, datastores, and APIs. Grid nodes serving as gateways govern access to the grid while separating external network traffic from the private network.

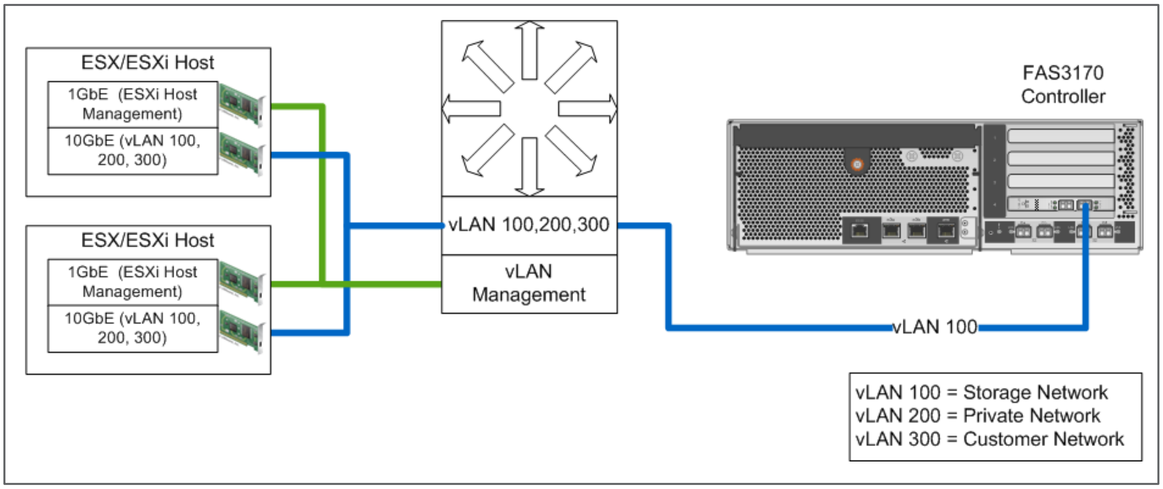
STORAGE NETWORK

Data storage and retrieval have an impact on the overall system performance. To improve performance and manageability, access to external storage is provided on a dedicated network. Maintaining a separate storage network optimizes the performance of the entire system.

All StorageGRID networks must be interconnected through a series of network switches providing redundant network paths to NetApp storage systems. Fundamental to this network design are resiliency and performance, which are enabled through bonded network interfaces for failover and bandwidth optimization. Given this design principle, each element of the network infrastructure must have sufficient network interfaces to support the virtualization layer. All servers must have at least four physical interfaces, five in the case of a high-availability (HA) gateway node. An HA gateway node requires an extra heartbeat interface to facilitate failover. These interfaces are cross-connected across a switching infrastructure that supports VLAN, tagging, and link aggregation.

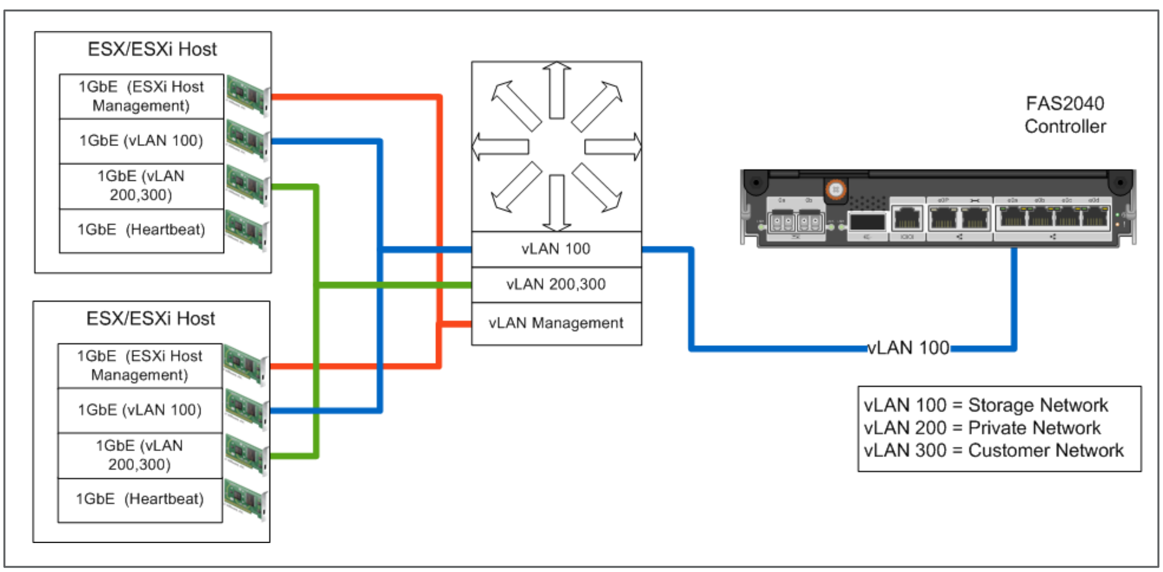
As illustrated in Figure 4, each server maintains two separate connections to different switches for each network. For the customer, private, and management networks, one interface is connected to Switch A while the second interface is connected to Switch B. NIC teaming provided by the virtualization layer provides load balancing and flow control based on NetApp best practices. This configuration is consistent with the storage network that has the same characteristics.

Figure 4) Physical network layout for 10GbE configuration.



Note: The switch in this diagram denotes a generic 10GbE switch provided by the customer.

Figure 5) Physical network layout for 1GbE configuration.



Note: The switch in this diagram denotes a generic 1GbE switch provided by the customer.

The network switches that make up the fabric interconnecting StorageGRID servers with NetApp storage systems have common configurations. This is meant to simplify the physical network infrastructure so that it is fully transparent to the virtualization layer. To properly segment network traffic, the switch must manage the separate broadcast domains (VLANs) for each network.

The network configuration applied in VMware at the virtualization layer is mapped to the configuration of switches at the network layer. In the context of the network layer, it is important to understand that the networks that are configured at the virtualization layer converge on the same physical interfaces on the server. This has implications on the network configuration, because the network must be able to distinguish traffic transiting through that port as belonging to a particular StorageGRID network. Therefore, the traffic leaving the server must be identified to the switch as belonging to a particular network, and the switch must be able to forward Ethernet frames based on that identifier. This capability is made possible through VLAN tagging.

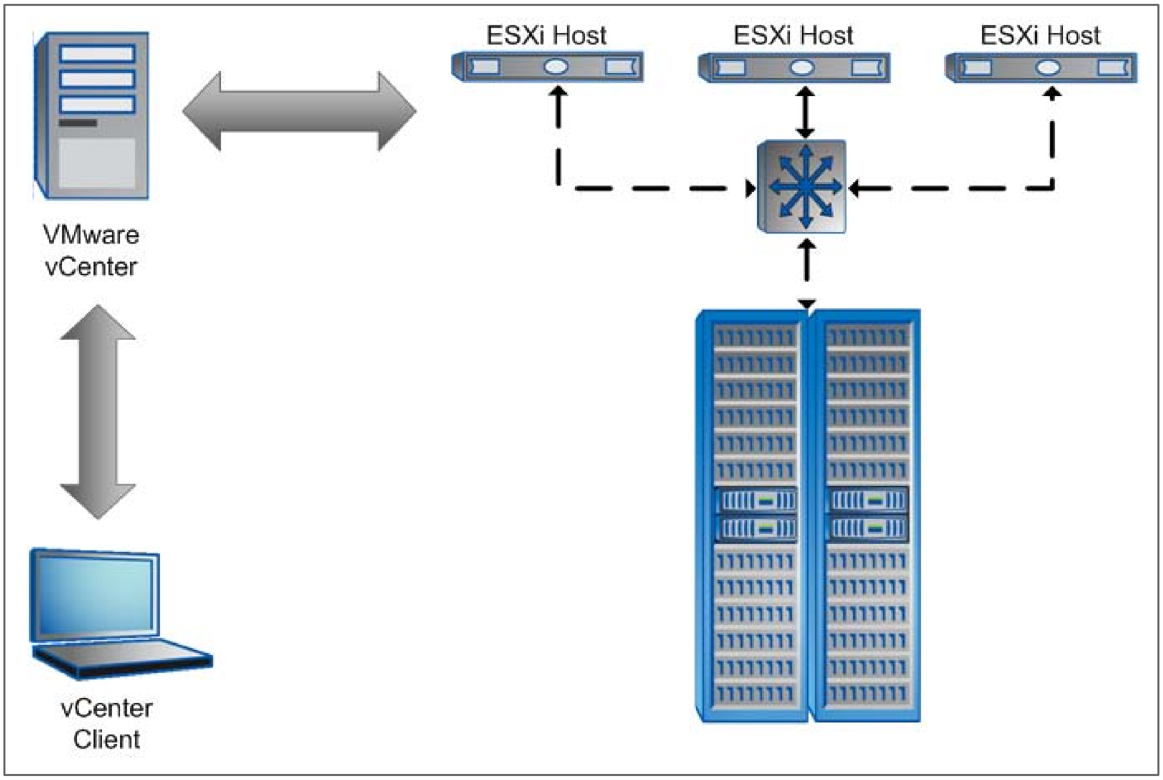
2.5 SYSTEM LAYER

The system layer consists of hardware systems that support the virtualization layer. These hardware systems include servers running VMware hypervisors and NetApp storage systems that provide the storage for the StorageGRID system.

VMWARE SERVERS

The StorageGRID solution requires a number of servers to support the virtualization layer. Servers are required to host the VMware infrastructure including the ESXi server and VMware vCenter.

Figure 6) VMware infrastructure.



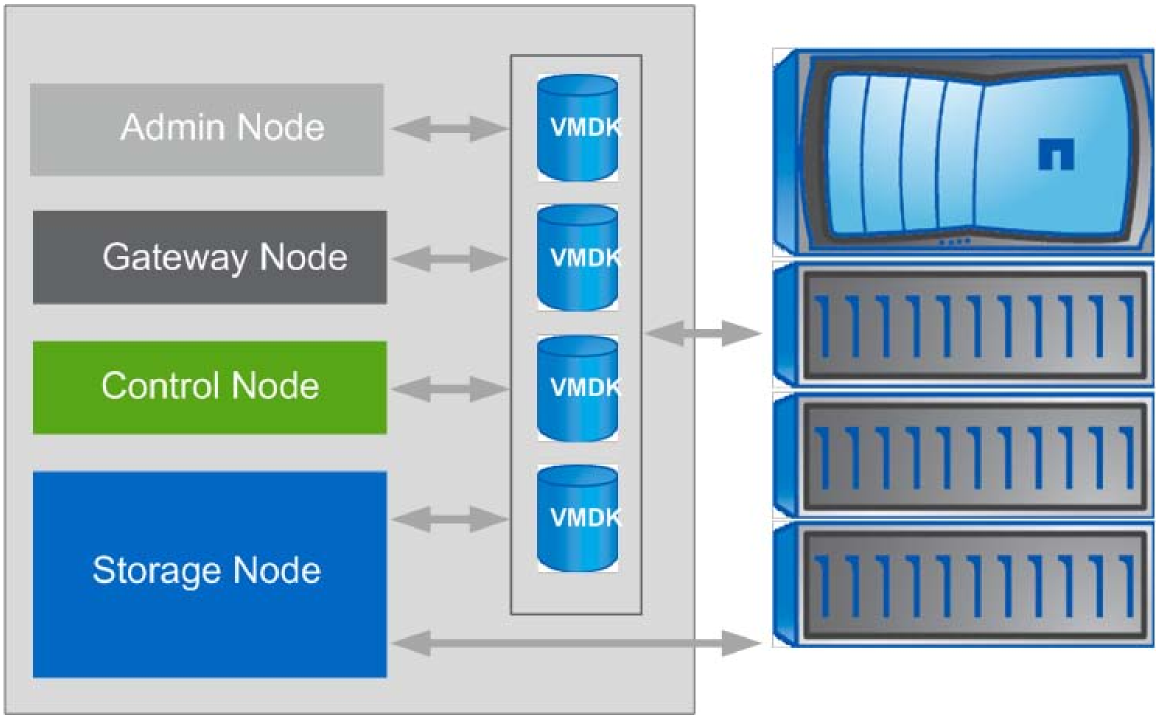
A VMware ESXi server acts as a host for a virtual machine. To enable the centralized management of ESXi servers, a VMware vCenter server is required. The VMware vCenter server manages system resources and virtual machines running on remote VMware ESXi servers. To administer ESXi servers and their hosted virtual machines, the vSphere client must connect to the VMware vCenter server management interface.

2.6 VIRTUALIZATION LAYER

After the system and network layers are set up, configure the VMware hosts and virtual machines to support the StorageGRID system. This configuration involves the creation of virtual machines on the VM host and includes the allocation of virtual machine resources for CPU, memory, storage, and networking.

The virtualization layer is central to the installation of a StorageGRID system. This layer is responsible for interfacing the system and network layers with the StorageGRID software. Figure 7 illustrates the configuration of a VM host connected to a NetApp storage system.

Figure 7) VM host connected to a NetApp storage system.



Each VMware host contains a number of virtual machines, each of which runs a grid node. Within the virtual machine, each grid node is configured to use a number of virtual devices. In Figure 7, each virtual grid node is mapped to a virtual disk in the host datastore. In addition to the standard mounts, a storage node mounts additional volumes from the external storage device to serve as its object datastore.

3 CONCLUSION

This document introduces the StorageGRID POD architecture for facilitating the rapid deployment and support of NetApp StorageGRID. This review of the StorageGRID POD operational layers gives NetApp field sales, partners, and customers a better understanding of how to deploy StorageGRID in their enterprise environment. For more information on how the NetApp StorageGRID POD can be used in your storage infrastructure, consult the StorageGRID overlay sales team.

4 RESOURCES

4.1 TERMINOLOGY AND DEFINITIONS

Table 1) Terms and definitions.

Term or Acronym	Definition
DC	Data center
DR	Disaster recovery
ESX/ESXi	VMware virtual machine servers running on bare metal hardware
POD	A combination of software and hardware used to deploy a StorageGRID solution; a minimum of two PODs are used
Virtual machine (VM)	Middleware layers that abstract operating systems from underlying system hardware by translating machine code into byte code for use by a hypervisor
VM hosts	Servers running virtualization software (hypervisor)
vSwitch	A virtual network switching interconnectivity among VM hosts running on a virtual machine
vCPU	A virtual CPU running in a virtual machine representing a single logical hardware processor; one physical core equals two vCPUs when hyperthreaded
VLAN	A virtual LAN representing a single broadcast domain or subnet for the communication of Ethernet frames
VLAN tag	An identifier within an Ethernet frame embedded by a host or switching device identifying the frame as belonging to a particular VLAN
Jumbo frames	Ethernet frames that are larger than their default size of 1,500 bytes, intended to improve performance; jumbo frames are typically 9,000 bytes in size
Virtual disk	A logical abstraction of physical disk volumes running in a virtual machine
10GbE	10-Gigabit Ethernet
1GbE	1-Gigabit Ethernet

4.2 ONLINE RESOURCES

- NetApp StorageGRID Web site
  www.netapp.com/us/products/storage-software/storagegrid/
- Introduction to vSphere:
  www.vmware.com/pdf/vsphere4/r40/vsp_40_intro_vs.pdf
- Performance optimization for VMware:
  www.vmware.com/pdf/Perf_Best_Practices_vSphere4.0.pdf
- ESXi and vCenter server installation information:
  www.vmware.com/pdf/vsphere4/r41/vsp_41_esxi_i_vc_setup_guide.pdf
- VMware vSphere CLI information:
  www.vmware.com/support/developer/vcli/

www.netapp.com

TR-3885-0311