

# **VMware Virtual SAN: Deploy and Manage**

Lecture Manual  
Virtual SAN 6.2

VMware Confidential  
Internal Use Only



VMware® Education Services  
VMware, Inc.  
[www.vmware.com/education](http://www.vmware.com/education)

**VMware Virtual SAN:**

**Deploy and Manage**

Virtual SAN 6.2

Part Number EDU-EN-VSANDM62-LECT

Lecture Manual

Copyright © 2016 VMware, Inc. All rights reserved. This manual and its accompanying materials are protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The training material is provided “as is,” and all express or implied conditions, representations, and warranties, including any implied warranty of merchantability, fitness for a particular purpose or noninfringement, are disclaimed, even if VMware, Inc., has been advised of the possibility of such claims. This training material is designed to support an instructor-led training course and is intended to be used for reference purposes in conjunction with the instructor-led training course. The training material is not a standalone training tool. Use of the training material for self-study without class attendance is not recommended.

These materials and the computer programs to which it relates are the property of, and embody trade secrets and confidential information proprietary to, VMware, Inc., and may not be reproduced, copied, disclosed, transferred, adapted or modified without the express written approval of VMware, Inc.

Course development: Carla Gavalakis

Technical review: Javier Menendez, Roy Freeman, Jase McCarty, Jeff Hunter, John Nicholson

Technical editing: Shalini Pallat, Sheena Lakshmi

Production and publishing: Jen Myers, Philip Boyer

The courseware for VMware instructor-led training relies on materials developed by the VMware Technical Communications writers who produce the core technical documentation, available at <http://www.vmware.com/support/pubs>.

# CONTENTS

## MODULE 1

Course Introduction . . . . .	1
Importance . . . . .	2
Learner Objectives . . . . .	3
You Are Here . . . . .	4
Typographical Conventions . . . . .	5
References . . . . .	6
Software-Defined Enterprise . . . . .	7
Components of Software-Defined Data Center . . . . .	8
Components of the VMware Software-Defined Data Center . . . . .	9
Benefits of Software-Defined Data Center (1) . . . . .	10
Benefits of Software-Defined Data Center (2) . . . . .	11
Benefits of the Software-Defined Data Center (3) . . . . .	12
Software-Defined Storage . . . . .	13
Virtual SAN . . . . .	14
vSphere Resources . . . . .	15
VMware Education Overview . . . . .	16

## MODULE 2

Storage Fundamentals . . . . .	17
You Are Here . . . . .	18
Importance . . . . .	19
Learner Objectives . . . . .	20
Hard Disk Drives . . . . .	21
Solid-State Drives . . . . .	22
Storage Performance Factors . . . . .	24
Performance Factors: Disk Type . . . . .	25
Performance Factors: RAID Type (1) . . . . .	26
Performance Factors: RAID Type (2) . . . . .	27
RAID 0 and Pass-Through Mode . . . . .	28
Performance Factors: Read Caching and Write Buffering . . . . .	29
Performance Factors: Latency . . . . .	30
Performance Factors: Storage Path . . . . .	31
Proactive Reporting . . . . .	32
vSphere Storage Feature Comparison . . . . .	33
Storage Virtualization . . . . .	34
vSphere Datastore . . . . .	36
Traditional vSphere Storage Architectures . . . . .	37
Block Storage . . . . .	38
Storage Area Networks . . . . .	39
Fibre Channel SANs . . . . .	40
Fibre Channel over Ethernet . . . . .	41
iSCSI Storage Area Networks . . . . .	42
Direct-Attached Storage . . . . .	43
Logical Unit Number . . . . .	44

VMFS Datastore . . . . .	45
Network-Attached Storage . . . . .	46
NFS . . . . .	47
vSphere Virtual Volumes . . . . .	48
Flash Read Cache . . . . .	50
Virtual SAN . . . . .	51
Labs . . . . .	52
Lab 1: Licensing vSphere and Virtual SAN Components . . . . .	53
Lab 2: Basic Storage Commands . . . . .	54
Review of Learner Objectives . . . . .	55
Key Points . . . . .	56
 MODULE 3	
Introduction to Virtual SAN . . . . .	57
You Are Here . . . . .	58
Importance . . . . .	59
Learner Objectives . . . . .	60
Virtual SAN: Hybrid Architecture . . . . .	61
Virtual SAN: All-Flash Architecture . . . . .	62
Control and Virtual Data Planes . . . . .	63
The Control Plane . . . . .	64
The Virtual Data Plane . . . . .	65
Virtual SAN Requirements . . . . .	66
Virtual SAN Minimums and Maximums . . . . .	67
Virtual SAN and Object-Based Storage . . . . .	68
Objects (1) . . . . .	69
Objects (2) . . . . .	70
Components (1) . . . . .	71
Components (2) . . . . .	72
Mirroring . . . . .	73
Striping . . . . .	74
Mirroring Plus Striping . . . . .	75
Erasure Coding: RAID 5 . . . . .	76
Erasure Coding: RAID 6 . . . . .	77
About vSphere API for Storage Awareness . . . . .	78
Storage Providers . . . . .	79
Virtual Machine Storage Policies . . . . .	80
Multiple Storage Policies . . . . .	81
Replicas . . . . .	82
Witnesses . . . . .	83
Witness Example . . . . .	84
Virtual SAN Disk Formats . . . . .	85
Disk Groups . . . . .	86
Hybrid Disk Groups . . . . .	87

All-Flash Disk Groups . . . . .	88
Virtual SAN Datastore . . . . .	89
Space Efficiency . . . . .	90
Deduplication and Compression . . . . .	91
RAID 5/6 (Erasure Coding) . . . . .	92
Virtual Machine Swap File Efficiency (1) . . . . .	93
Virtual Machine Swap File Efficiency (2) . . . . .	94
Virtual SAN Health Service . . . . .	95
Additional Tools . . . . .	96
Review of Learner Objectives . . . . .	97
Key Points . . . . .	98
 MODULE 4	
Virtual SAN Configuration . . . . .	99
You Are Here . . . . .	100
Importance . . . . .	101
Module Lessons . . . . .	102
Configuring a Virtual SAN Network . . . . .	103
Learner Objectives . . . . .	104
Virtual SAN Network Traffic Flow . . . . .	105
Virtual SAN with vSphere Standard Switch . . . . .	106
Virtual SAN with vSphere Distributed Switch . . . . .	107
Multicast Requirement . . . . .	108
Layer 2 Multicast . . . . .	109
Changing the Multicast Address . . . . .	110
Virtual SAN Ports . . . . .	111
Configuring the Virtual SAN VMkernel Port . . . . .	112
Physical Network Adapters and NIC Teaming . . . . .	113
Lab 3: Configuring a Virtual SAN Network . . . . .	114
Review of Learner Objectives . . . . .	115
Configuring a Virtual SAN Cluster . . . . .	116
Learner Objectives . . . . .	117
Virtual SAN Cluster Requirements . . . . .	118
Enabling Virtual SAN on a vSphere Cluster . . . . .	119
Using Automatic Disk Claim Mode . . . . .	120
Using Manual Disk Claim Mode . . . . .	121
Enabling Deduplication and Compression . . . . .	122
Creating Disk Groups . . . . .	123
Disk Grouping by Disk Model or Size . . . . .	124
Disk Grouping by Hosts . . . . .	125
Creating All-Flash Disk Groups . . . . .	126
Viewing Disk Groups . . . . .	127
Labs . . . . .	128
Lab 4: Configuring a Virtual SAN Cluster . . . . .	129

Lab 5: Configuring Hybrid Disk Groups . . . . .	130
Lab 6: Configuring All-Flash Disk Groups . . . . .	131
Lab 7: Virtual SAN Storage Commands . . . . .	132
Review of Learner Objectives . . . . .	133
Key Points . . . . .	134
<b>MODULE 5</b>	
Virtual SAN Policies and Virtual Machines . . . . .	135
You Are Here . . . . .	136
Importance . . . . .	137
Module Lessons . . . . .	138
Storage Policy-Based Management . . . . .	139
Learner Objectives . . . . .	140
Review: Virtual SAN and Storage Policy-Based Management . . . . .	141
Storage Policies . . . . .	142
Rule Sets . . . . .	143
Virtual Machine Storage Policy Capabilities for Virtual SAN . . . . .	144
Number of Failures to Tolerate . . . . .	145
Number of Disk Stripes Per Object . . . . .	146
Failure Tolerance Method . . . . .	147
Mirroring Versus Erasure Coding . . . . .	148
Flash Read Cache Reservation . . . . .	149
Force Provisioning . . . . .	150
Object Space Reservation . . . . .	151
IOPS Limit for Object . . . . .	152
Disable Object Checksum . . . . .	153
Viewing Object Placement . . . . .	154
Assigning Storage Policies . . . . .	155
Storage Policies and Multiple VMDKs . . . . .	156
Labs . . . . .	157
Lab 8: Deploying Virtual Machines to Virtual SAN . . . . .	158
Lab 9: Creating Storage Policies . . . . .	159
Review of Learner Objectives . . . . .	160
vsanSparse Snapshots . . . . .	161
Learner Objectives . . . . .	162
About vsanSparse . . . . .	163
Review: VMware Snapshots . . . . .	164
vsanSparse: Always Sparse . . . . .	165
vsanSparse Memory Cache . . . . .	166
Snapshot Flow . . . . .	168
Uniform Snapshot Format . . . . .	169
Read Cache Considerations . . . . .	170
Snapshot Considerations . . . . .	171
Verifying Snapshot Format . . . . .	172

Review of Learner Objectives . . . . .	173
Key Points . . . . .	174
<b>MODULE 6</b>	
Managing and Operating Virtual SAN . . . . .	175
You Are Here . . . . .	176
Importance . . . . .	177
Learner Objectives . . . . .	178
Maintenance Mode Options . . . . .	179
Managing Hardware Storage Devices . . . . .	180
Resynchronizing Components . . . . .	181
Proactive Rebalance . . . . .	182
Understanding Failure Events (1) . . . . .	183
Understanding Failure Events (2) . . . . .	184
Failure Scenario: Restoring I/O Flow . . . . .	185
Failure Scenario: Rebuilding Components (1) . . . . .	186
Failure Scenario: Rebuilding Components (2) . . . . .	187
Handling Failure . . . . .	188
Cache Device Failure: Instant Mirror Copy . . . . .	189
Capacity Device Failure: Instant Mirror Copy . . . . .	190
Host Failure: 60-Minute Delay . . . . .	191
Storage Controller Failures . . . . .	192
Network Failure: 60-Minute Delay . . . . .	193
Replace and Upgrade Devices . . . . .	194
Virtual SAN Scalable Architecture . . . . .	195
vCenter Server Alarms . . . . .	196
Fault Domains . . . . .	197
Example of a Fault Domain . . . . .	198
Upgrade Overview . . . . .	199
Performing the On-Disk Format Upgrade . . . . .	200
Labs . . . . .	201
Lab 10: Using Maintenance Mode Options . . . . .	202
Lab 11: Scaling Out a Virtual SAN Cluster . . . . .	203
Lab 12: Working with Fault Domains . . . . .	204
Review of Learner Objectives . . . . .	205
Key Points . . . . .	206
<b>MODULE 7</b>	
Monitoring and Troubleshooting Virtual SAN . . . . .	207
You Are Here . . . . .	208
Importance . . . . .	209
Module Lessons . . . . .	210
Monitoring with vSphere Web Client . . . . .	211
Learner Objectives . . . . .	212
Monitoring with vSphere Web Client . . . . .	213

Example: Storage Policy Issues . . . . .	214
Example: Virtual SAN Health Issues . . . . .	215
About the Health Service and Performance Service . . . . .	216
Health Service Tests . . . . .	217
Example: Troubleshooting a Network Health Issue (1) . . . . .	218
Example: Troubleshooting a Network Health Issue (2) . . . . .	219
Example: Troubleshooting a Network Health Issue (3) . . . . .	220
Example: Troubleshooting a Network Health Issue (4) . . . . .	221
Enabling the Performance Service . . . . .	222
Performance Views . . . . .	223
Cluster Metrics: Virtual Machine Consumption . . . . .	224
Host Metrics: Disk Group and Disk . . . . .	225
Virtual Machine Metrics: Virtual Disk . . . . .	226
Viewing Virtual SAN Capacity Details . . . . .	227
Running Proactive Tests . . . . .	228
Updating the HCL Database . . . . .	229
Uploading Support Bundles . . . . .	230
Lab 13: Using the Health and Performance Services . . . . .	231
Review of Learner Objectives . . . . .	232
Monitoring with vRealize Operations Manager . . . . .	233
Learner Objectives . . . . .	234
Overview of vRealize Operations Manager . . . . .	235
Global View . . . . .	236
Example: Troubleshooting a Health Issue (1) . . . . .	237
Example: Troubleshooting a Health Issue (2) . . . . .	239
Virtual SAN Alerts . . . . .	240
Monitoring the Virtual SAN Cluster . . . . .	241
Monitoring Entity Usage . . . . .	242
Viewing Device Information . . . . .	243
Review of Learner Objectives . . . . .	244
Monitoring from the Command Line . . . . .	245
Learner Objectives . . . . .	246
About vSphere ESXi Shell . . . . .	247
esxcfg-info Command . . . . .	248
esxtop Command: d Option . . . . .	249
esxtop Command: u Option . . . . .	250
esxtop Command: v Option . . . . .	251
esxtop Command: m Option . . . . .	252
esxtop Command: c Option . . . . .	253
esxtop Command: n Option . . . . .	254
esxtop Command: x Option . . . . .	255
esxcli Command . . . . .	256
Ruby vSphere Console . . . . .	257

Log In to the Ruby vSphere Console . . . . .	258
Navigating the vSphere and Virtual SAN Infrastructure . . . . .	259
Using RVC to Get Help. . . . .	260
Using RVC to List Virtual SAN Commands . . . . .	261
Special Objects and Commands . . . . .	262
Viewing Virtual Machine Object Information . . . . .	263
Displaying Snapshot Delta Object Information . . . . .	264
Displaying Swap File Object Information. . . . .	265
Viewing Host-Specific Information . . . . .	266
Viewing Host Disk Information . . . . .	267
Viewing Host Resource Statistics . . . . .	268
Viewing Data Center Disk Statistics . . . . .	269
Activating Virtual SAN Observer . . . . .	270
Virtual SAN Observer Web Server . . . . .	271
VSAN Disks Tab . . . . .	272
VSAN Disks (Deep-Dive) Tab . . . . .	273
PCPU and Memory Tabs . . . . .	274
VMs Tab . . . . .	275
Lab 14: Using Ruby vSphere Console and ESXi Commands . . . . .	276
Review of Learner Objectives . . . . .	277
Key Points . . . . .	278
<b>MODULE 8</b>	
Stretched Clusters and Two-Node Clusters. . . . .	279
You Are Here . . . . .	280
Importance . . . . .	281
Module Lessons . . . . .	282
Stretched Clusters and Two-Node Clusters . . . . .	283
Learner Objectives . . . . .	284
About Virtual SAN Stretched Cluster . . . . .	285
Stretched Cluster Use Cases . . . . .	286
Stretched Cluster Architecture . . . . .	287
Single-Site Versus Stretched Clusters . . . . .	288
About the Witness Host . . . . .	289
Sizing the Witness Host . . . . .	290
Networking Requirements . . . . .	291
Network Latency and Bandwidth Recommendations . . . . .	292
About the Preferred Site . . . . .	293
Read Locality . . . . .	294
Deploy the Witness Host . . . . .	295
Create a Stretched Cluster . . . . .	297
Creating a Virtual SAN Cluster . . . . .	298
Configuring the Stretched Cluster . . . . .	299
Configuring DRS Affinity Groups and Rules . . . . .	300

Interoperability with vSphere HA . . . . .	301
Management and Maintenance . . . . .	302
About the Two-Node Cluster . . . . .	303
Two-Node Cluster Use Case . . . . .	304
Two-Node Cluster and Stretched Cluster . . . . .	305
Lab 15: Creating a Stretched Cluster . . . . .	306
Review of Learner Objectives . . . . .	307
Stretched Cluster Failure Scenarios . . . . .	308
Learner Objective . . . . .	309
Stretched Cluster Heartbeats . . . . .	310
Data Site Failure . . . . .	311
Failure of Single Host in Data Site . . . . .	312
Witness Host Failure or Loss of Network Connection . . . . .	313
Network Failure Between Data Sites . . . . .	314
Site Failure Where vCenter Server Is Hosted . . . . .	315
Disk Failure . . . . .	316
Lab 16: Configuring vSphere HA and DRS in the Stretched Cluster . . . . .	317
Review of Learner Objective . . . . .	318
Key Points . . . . .	319
<b>MODULE 9</b>	
Interoperability with vSphere Features . . . . .	321
You Are Here . . . . .	322
Importance . . . . .	323
Learner Objectives . . . . .	324
Virtual SAN and vSphere HA . . . . .	325
vSphere HA Networking Differences with Virtual SAN . . . . .	326
One Host Isolated: vSphere HA Restarts Virtual Machine . . . . .	327
Two Hosts Isolated: vSphere HA Restarts Virtual Machine . . . . .	328
Hosts Isolated: vSphere HA Fails to Restart Virtual Machine . . . . .	329
vSphere Data Protection . . . . .	330
vSphere Replication and Site Recovery Manager . . . . .	331
Stretched Cluster with vSphere Replication and SRM . . . . .	332
vSphere Fault Tolerance . . . . .	333
vRealize Automation . . . . .	334
OpenStack Framework . . . . .	335
View . . . . .	336
vSphere PowerCLI Cmdlets . . . . .	337
File Services with NexentaConnect . . . . .	338
Virtual SAN and Hytrust DataControl . . . . .	339
Virtual SAN and Oracle RAC . . . . .	340
Virtual SAN and Windows Server Failover Clustering . . . . .	341
Review of Learner Objectives . . . . .	342
Key Points . . . . .	343

## MODULE 10

Designing a Virtual SAN Deployment . . . . .	345
You Are Here . . . . .	346
Importance . . . . .	347
Learner Objectives . . . . .	348
Following the Compatibility Guide . . . . .	349
Cluster Design Considerations (1) . . . . .	350
Cluster Design Considerations (2) . . . . .	351
Using All-Flash Architectures . . . . .	352
Disk Group Design . . . . .	353
Cache Capacity Sizing . . . . .	354
Choosing Devices for Cache and Capacity Tiers . . . . .	355
Capacity Planning Considerations . . . . .	356
Establishing Baseline Capacity Requirements (1) . . . . .	357
Establishing Baseline Capacity Requirements (2) . . . . .	358
Virtual SAN TCO and Sizing Calculator . . . . .	359
VMware Infrastructure Planner . . . . .	360
Network . . . . .	361
Network I/O Control . . . . .	362
Virtual SAN Storage Policy Considerations . . . . .	363
Host Design Considerations: CPU and Memory . . . . .	364
Host Design Considerations: Storage Considerations . . . . .	365
Use Cases . . . . .	366
Use Case: Virtual Desktop Infrastructures . . . . .	367
Use Case: Management Cluster . . . . .	368
Use Case: Private Cloud for Testing and Development . . . . .	369
Review of Learner Objectives . . . . .	370
Key Points . . . . .	371

VMware Confidential  
Internal Use Only

VMware Confidential  
Internal Use Only

## MODULE 1

# Course Introduction

Slide 1-1

Module 1

*VMware Virtual SAN: Deploy and Manage*

VMware Confidential  
Internal Use Only

## Importance

Slide 1-2

VMware Virtual SAN™ is a policy-driven software-defined storage tier that is integrated with VMware vSphere® Hypervisor. Virtual SAN simplifies storage provisioning and management in the software-defined enterprise.

VMware Confidential  
Internal Use Only

# Learner Objectives

Slide 1-3

By the end of this course, you should be able to meet the following objectives:

- Describe the Virtual SAN architecture
- Identify Virtual SAN features and use cases
- Configure Virtual SAN networking components
- Configure a Virtual SAN cluster
- Deploy virtual machines on a Virtual SAN datastore
- Configure virtual machine storage policies
- Perform ongoing Virtual SAN management tasks
- Outline the tasks for upgrading to Virtual SAN 6.2
- Use the Virtual SAN health service to monitor health and performance
- Monitor Virtual SAN with VMware ESXi™ commands and Ruby vSphere Console
- Configure a stretched cluster and observe failover scenarios
- Describe Virtual SAN interoperability with VMware vSphere® and other products
- Plan and design a Virtual SAN cluster

VMware Confidential  
Internal Use Only

# You Are Here

Slide 1-4

- 1. Course Introduction**
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
- 10. Designing a Virtual SAN Deployment**

VMware Confidential  
Internal Use Only

# Typographical Conventions

Slide 1-5

The following typographical conventions are used in this course.

**Monospace**

Filenames, folder names, path names, command names:  
Navigate to the `VMS` folder.

**Monospace bold**

What the user types:  
Enter `ipconfig /release`.

**Boldface**

User interface controls:  
Click the **Configuration** tab.

*Italic*

Book titles and placeholder variables:

- *vSphere Virtual Machine Administration*
- *ESXi\_host\_name*

VMware Confidential  
Internal Use Only

# References

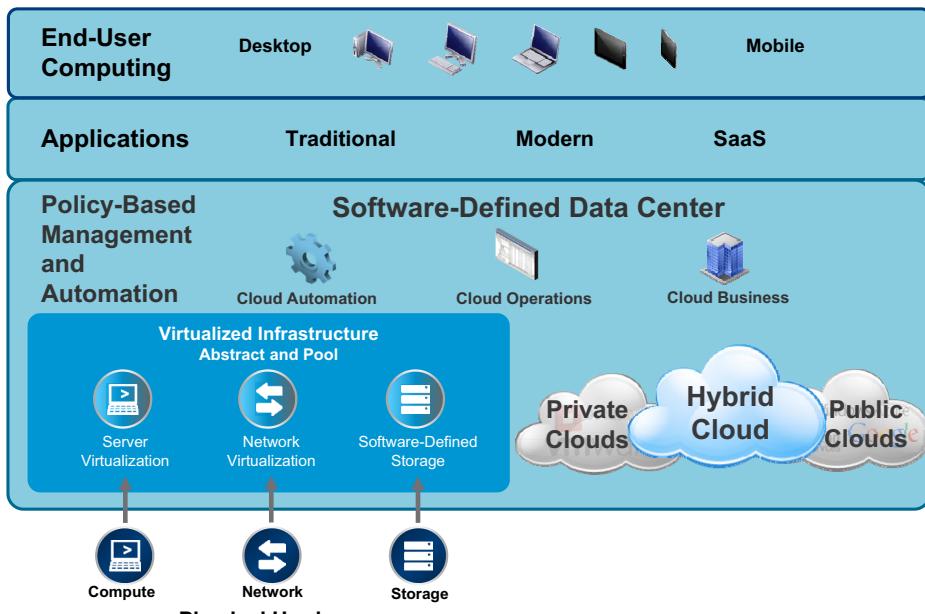
Slide 1-6

Title	Location
<i>vSphere Storage</i>	<a href="https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html">https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html</a>
<i>vSphere Installation and Setup</i>	<a href="https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html">https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html</a>
<i>VMware Virtual SAN 6.2 Design and Sizing Guide</i>	<a href="http://www.vmware.com/files/pdf/products/vsan/virtual-san-6.2-design-and-sizing-guide.pdf">http://www.vmware.com/files/pdf/products/vsan/virtual-san-6.2-design-and-sizing-guide.pdf</a>
<i>VMware Virtual SAN 6.2 Space Efficiency Technologies</i>	<a href="http://www.vmware.com/files/pdf/products/vsan/vmware-vsani-6.2-space-efficiency-technologies.pdf">http://www.vmware.com/files/pdf/products/vsan/vmware-vsani-6.2-space-efficiency-technologies.pdf</a>
<i>VMware Virtual SAN 6.2 Stretched Cluster &amp; 2 Node Guide</i>	<a href="http://www.vmware.com/files/pdf/products/vsan/VMware-Virtual-SAN-6.2-Stretched-Cluster-Guide.pdf">http://www.vmware.com/files/pdf/products/vsan/VMware-Virtual-SAN-6.2-Stretched-Cluster-Guide.pdf</a>

VMware Confidential  
Internal Use Only

# Software-Defined Enterprise

Slide 1-7



The software-defined enterprise uses intelligence and automation to shift the focus of data center management from process to policy through a common computing platform for all applications. This platform is based on a highly automated software-defined data center architecture. The architecture includes a common cloud management approach that provides transparent governance of infrastructure and service components, including services that are traditionally outside IT control.

The software-defined enterprise is implemented with a layer of management that automates tasks and activities in the compute, storage, and network components. The software-defined enterprise is incomplete without a management and automation layer that interfaces programmatically with the APIs exposed for the underlying components.

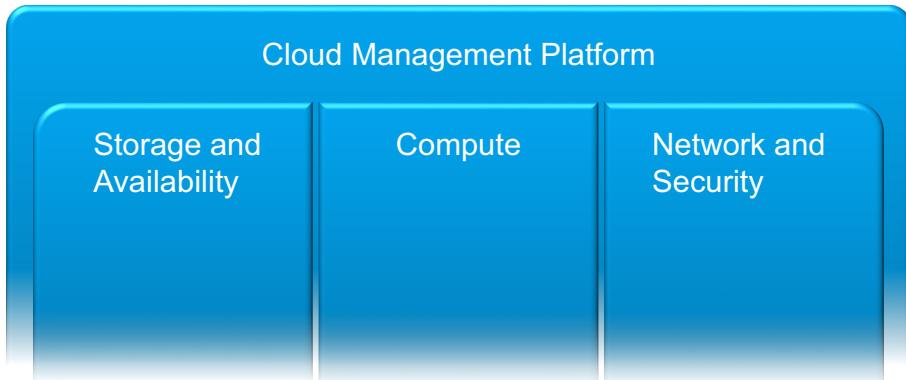
A software-defined environment requires policy-based automation and a common management platform across the entire infrastructure. The logical design of a software-defined data center includes compute, storage, virtualized network, and cloud management platform components. A software-defined enterprise abstracts compute, storage, and networking from the underlying physical hardware. Thus, IT can quickly and efficiently meet rapid business changes in an organization.

VMware offers management solutions to manage the business and financial dimensions of your cloud implementation. These offerings include cloud automation, cloud operations, and solutions.

# Components of Software-Defined Data Center

Slide 1-8

The conceptual design of the software-defined enterprise is built on a software-defined data center.

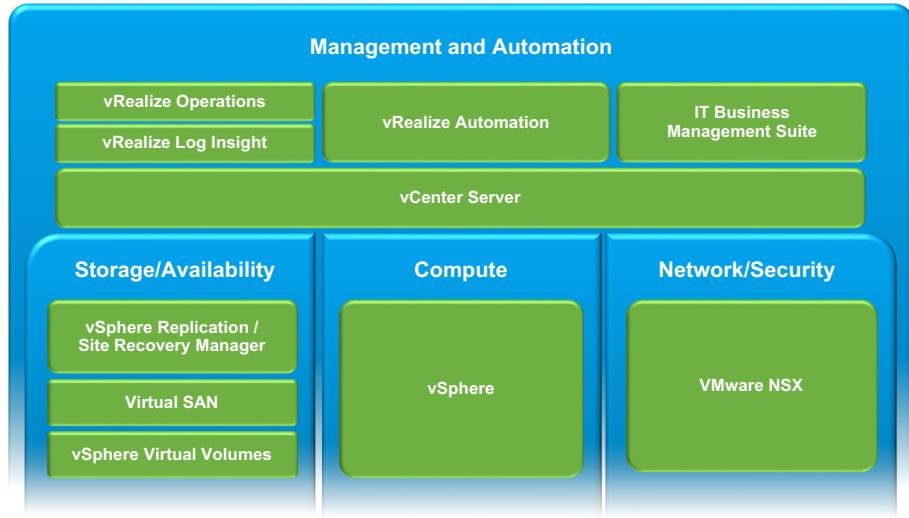


The software-defined data center includes software-defined storage, compute, and virtualized network components. A cloud management platform unifies the infrastructure and automates activities within and between the supporting pillars. By abstracting and bringing resources together with analytics-based operations management, the software-defined enterprise architecture results in greater efficiency and agility. The cloud management platform (CMP) component of the software-defined data center brings increased agility to a virtual infrastructure. The CMP offers self-service provisioning that is based on policy-driven automation to ensure that users are limited to the scope of their job role.

# Components of the VMware Software-Defined Data Center

Slide 1-9

VMware offers products that fit into each of the components of the software-defined data center.



VMware has a suite of products that you can use to create a software-defined data center. VMware products are available for all three pillars of the software-defined data center. VMware vSphere® provides processor and memory resources for the compute pillar. VMware Virtual SAN™, VMware vSphere® Replication™, and VMware Site Recovery Manager™ provide storage and availability services. VMware NSX® provides network virtualization for the software-defined data center.

Virtual SAN can be deployed and can coexist with VMware NSX in a vSphere infrastructure. Virtual SAN and VMware NSX are not dependent on each other to work. The Virtual SAN network stack can be configured over layer-2 and layer-3 network topologies.

VMware NSX does not support the configuration of VMkernel interface traffic over the VXLAN overlay. Examples of VMkernel interface traffic are VMware vSphere® vMotion® traffic, iSCSI traffic, management traffic, and Virtual SAN traffic.

At the highest level, VMware also provides additional tools for managing and automating a software-defined data center.

## Benefits of Software-Defined Data Center (1)

Slide 1-10

A software-defined data center provides the following major benefits when compared to traditional data centers:

- Agility
- Efficiency
- Choice
- Control
- Simplicity

VMware Confidential  
Internal Use Only

## Benefits of Software-Defined Data Center (2)

Slide 1-11

### Agility:

- Service provisioning across multiple platforms and multiple clouds
- Policy-driven automation
- Self-service portal

### Efficiency:

- Compute virtualization
- Virtualized network and security
- Software-defined storage
- Automated operations management
- Reduced hardware cost

The software-defined data center delivers efficiency and agility by virtualizing components of the physical infrastructure: compute, storage, and networking. Virtualized infrastructures can quickly adapt to change in business demands because they are not tied to silos of hardware. By abstracting and bringing these resources together, and by analytics-based operations management, the software-defined data center architecture results in reduced operational and capital expenditures.

The CMP component of the software-defined data center brings increased agility to a virtual infrastructure. The CMP offers self-service provisioning that is based on policy-driven automation to ensure that users are limited within the scope of their job role.

## Benefits of the Software-Defined Data Center (3)

Slide 1-12

### Choice:

- Removal of vendor-agnostic hardware dependencies
- Hybrid cloud extensibility
- Application deployment across multiple hardware stacks: physical or virtual
- Support for third-party hypervisors (Hyper-V, KVM) and clouds (OpenStack, Amazon AWS)

### Control:

- Automated business continuity or disaster recovery
- Virtualization-aware security and compliance across clouds
- Management across private and public clouds
- Operational analytics

### Simplicity:

- Simplified storage provisioning and management

The software-defined data center delivers increased control and choice. For example, by leveraging a policy-based governance framework, the software-defined data center provides the correct availability and security for every application or infrastructure service. Enabling the highest levels of application uptime, through automated business continuity and virtualization-aware security and compliance, helps to reduce downtime.

Through the self-service component of the software-defined data center, you can offer provisioning choices to users. Users can be entitled to provision virtual machines in private, public, or hybrid cloud infrastructures and on physical machines. This flexibility to span machine types provides the freedom to leverage multiple hardware or software stacks. You can use the software-defined enterprise solution to deploy any application on demand and on any platform.

# Software-Defined Storage

Slide 1-13



Software-defined storage delivers the following aspects of storage as software:

- Automated storage management model based on consumption through policies
- Extensible framework for ecosystems with hardware-agnostic data services
- Hypervisor-based storage abstraction on heterogeneous hardware

Software-defined storage abstracts, pools, and automates heterogeneous storage hardware resources. Providing a software-defined storage layer gives the same benefits of a virtualized compute layer: efficiency, agility, and flexibility.

Software-defined storage is an approach that delivers the following aspects of storage as software:

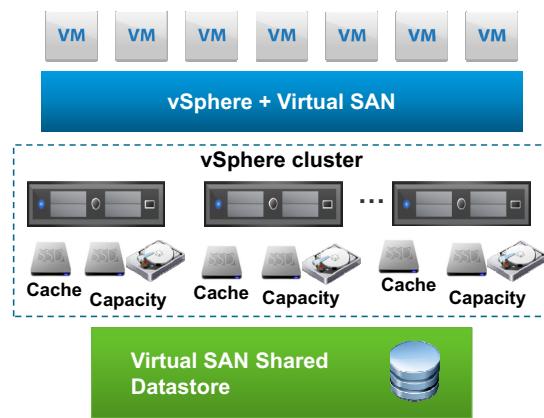
- Policy-based management that allows automated storage consumption through application-centric policies.
- Virtualized hardware-agnostic data services that allow detailed management of data services at the virtual machine level without the need of purpose-built hardware appliances.
- Abstraction of heterogeneous hardware resources that are presented to virtual machines as easily consumable pools of resources. These hardware resources might include direct-attached storage (DAS), or traditional SAN, or network storage (NAS).

# Virtual SAN

Slide 1-14

Virtual SAN is a distributed layer of software that runs natively as a part of the ESXi hypervisor.

Virtual SAN aggregates the local or direct-attached storage disks on the ESXi hosts in a vSphere cluster and creates a single storage pool that is shared across all hosts in the cluster.



VMware vSphere® High Availability, vSphere vMotion, and VMware vSphere® Distributed Resource Scheduler™ (DRS) require shared storage. Virtual SAN supports these features and eliminates the need for an external shared storage. Virtual SAN simplifies storage configuration and virtual machine provisioning activities.

Virtual SAN virtualizes local physical storage resources of ESXi hosts. Virtual SAN virtualizes these resources and presents them as a single pool of storage.

# vSphere Resources

Slide 1-15

For technical information on vSphere, use the following resources:

- VMware vSphere Documentation
  - <http://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html>
- Technical Papers
  - <https://www.vmware.com/vmtn/resources>
- Compatibility Guides
  - <http://www.vmware.com/guides.html>
- Knowledge Base
  - <http://kb.vmware.com>
- VMware Training and Certification
  - <http://www.vmware.com/training>

VMware Confidential  
Internal Use Only

# VMware Education Overview

Slide 1-16

Your instructor will introduce other Education Services offerings available to you:

- Learning Paths
- On-Demand Training
- VMware Learning Zone
- New Certification Framework

VMware Confidential  
Internal Use Only

## MODULE 2

# Storage Fundamentals

Slide 2-1

Module 2

VMware Confidential  
Internal Use Only

# You Are Here

Slide 2-2

1. Course Introduction
- 2. Storage Fundamentals**
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 2-3

vSphere supports different types of storage, which gives you the flexibility to set up your storage based on the cost, performance, and manageability requirements. You can choose from different types of disks, access methods, and connection technologies to implement the correct type of storage suitable for your environment.

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 2-4

By the end of this module, you should be able to meet the following objectives:

- Define common storage terminologies
- Identify characteristics of storage devices: magnetic and flash-based devices
- Identify and explain different types of storage architectures
- Identify SAN performance factors

VMware Confidential  
Internal Use Only

## Hard Disk Drives

Slide 2-5

Hard disk drives (HDDs) include single or multiple platters rotating at a specific speed to provide data access. The common HDD rotational speeds include the following:

- 15,000 RPM
- 10,000 RPM
- 7,200 RPM

In Virtual SAN, mechanical or magnetic drives are used in the capacity tier for hybrid configurations.



Hard disk drives (HDDs) are digital storage devices. HDDs use an electrical actuator to move a mechanical arm assembly with a magnetic head over rapidly rotating disks (platters) to read and write information.

## Solid-State Drives

Slide 2-6

Solid-state drives (SSDs) include a collection of memory semiconductors organized as a disk drive that uses either a SATA or PCIe interface.

SSD Category	General Lifespan of NAND Flash Devices
Single-level cell (SLC)	100,000 write cycles
Multi-level cell (MLC)	3,000 to 10,000 write cycles
Triple-level cell (TLC)	1,000 write cycles
Enterprise Multi-level cell (eMLC)	20,000 to 30,000 write cycles

Newer, high-density flash technology, such as PCIe and Non-Volatile Memory express (NVMe), have high performance and predictably low latencies:

- Virtual SAN 6.1 introduces support for Intel NVMe.

In Virtual SAN 6.x, flash devices can be used in both the cache tier and capacity tier, known as an all-flash configuration.

In addition to regular HDDs, VMware ESXi™ supports solid-state drives (SSDs). SSDs use semiconductors as their storage medium and have no moving parts.

On several storage arrays, the ESXi host can distinguish SSDs from traditional hard disks. To tag the SSD devices that are not detected, you can use Pluggable Storage Architecture Storage Array Type Plug-in claim rules.

In single-level cell (SLC), each cell can store a single bit (0 or 1) of information. In multilevel cells (MLC), NAND flash uses multiple levels per cell to allow more bits to be stored. An MLC can usually have four values (00, 01, 10, and 11). MLC is cheaper than SLC, but typically has a shorter life span. Because MLC uses the same number of transistors as SLC, a higher risk of errors exists. To reduce the risk of errors, another type of NAND flash called enterprise MLC (eMLC) has been designed. eMLC is a compromise between the cost and lifespan of SLC and MLC.

A triple-level cell (TLC) stores three bits in each cell. By storing more bits per cell, a TLC memory card achieves slower transfer speeds, higher error rates, and lower cell endurance than both SLC and MLC. The advantage of TLC memory is that the memory chip is physically smaller than SLC and MLC chips for a given memory capacity. TLC memory chip requires less power to operate than MLC memory and is cheaper to produce. TLC flash technology is used mostly in low-end memory devices where speed and reliability are not important.

Non-Volatile Memory express (NVMe) is a new specification developed specifically for SSDs. NVMe allows for greater parallelism to be utilized by both hardware and software, which results in improved performance.

VMware Confidential  
Internal Use Only

## Storage Performance Factors

Slide 2-7

Storage performance is based on several components and the configuration of a storage system. The factors include:

- Disk type
- RAID type
- Read cache and write buffer
- Latency
- Storage path

The performance of a storage system is typically measured in input/output operations per second (IOPS).

VMware Confidential  
Internal Use Only

## Performance Factors: Disk Type

Slide 2-8

### Magnetic disks:

- Slower speed
- Lower costs
- Larger capacity per device



### SSD:

- Increased speed
- Higher cost
- Lower capacity per device



The performance of the storage device depends on the disk type. A magnetic disk provides greater storage levels at slower speeds and at a much lower cost. SSD storage provides much faster speeds but with a significant increase in cost for the same raw capacity. A combination of these two solutions is often implemented to provide a compromise between speed and capacity.

## Performance Factors: RAID Type (1)

Slide 2-9

The common RAID types in use include the following:

- RAID 0: Striped
  - Advantages: Fastest performance. Data is striped across all disks.
  - Disadvantages: No redundancy. All data is lost if a single drive fails.
- RAID 1: Mirrored
  - Advantages: Good performance. An exact copy of the data is duplicated across all disks, so no data is lost if a disk fails.
  - Disadvantages: Usable capacity is reduced by 50 percent.
- RAID 10: Mirrored plus striped
  - Advantages: Best scalable performance and redundancy. Even if multiple disks are lost there is no data loss, as long as one mirror of each stripe remains.
  - Disadvantages: Expensive. Usable capacity is reduced by 50 percent.

When creating storage for use with servers, RAID type can also affect performance. RAID provides data protection and can increase the performance of storage systems.

In RAID 0, the data is striped across all disks in the RAID group. If one drive is lost, then the data is unrecoverable due to a lack of redundancy.

RAID 1 is also known as disk mirroring. The data on one disk is mirrored to a second disk. This RAID does not improve performance, but if either disk fails, a full copy of the data is available.

RAID 10 is a combination of RAID 0 and RAID 1. RAID 10 is faster than RAID 5 or 6 and carries redundancy of RAID 1. The main disadvantage of RAID 10 is that half of the storage capacity is lost and it is one of the most expensive RAID configurations.

## Performance Factors: RAID Type (2)

Slide 2-10

The common RAID types in use include the following:

- RAID 5: Striped plus parity
  - Advantages: Good performance and redundancy.
  - Disadvantages: Slower when writing to disk because parity is calculated and written along with the data.
- RAID 6: Striped plus double parity
  - Advantages: Increased data protection.
  - Disadvantages: Write performance is reduced because two parity instances are calculated and written along with the data.

In a RAID 5 configuration, the data is written across all disks. The number of disks in the RAID is directly proportional to the performance. In this configuration, there is no data loss if a single disk fails. The storage capacity is reduced by the capacity of one disk for the parity data.

RAID 6 is similar in function to RAID 5, except that an additional parity block is added. The disk group supports up to two concurrent disk failures as two parity blocks are written along with the data. These additional writes result in slower performance and the usable storage capacity is reduced by two disks in the RAID set.

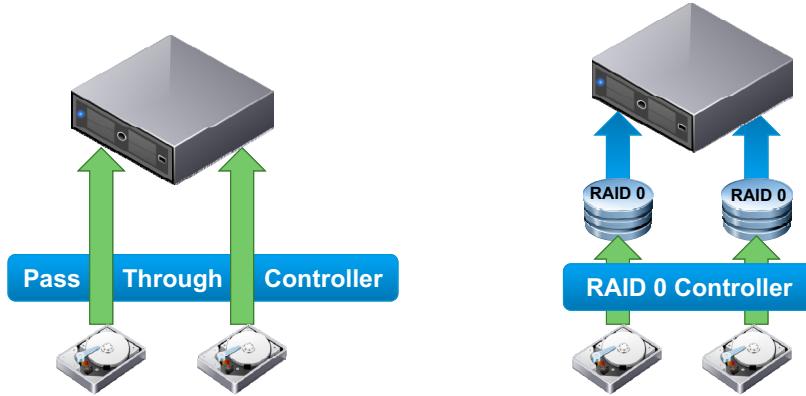
## RAID 0 and Pass-Through Mode

Slide 2-11

Some controllers require each magnetic disk to be configured as a RAID 0 volume so that the magnetic disks are visible to an ESXi host.

Controllers that support the pass-through mode present disks directly to a host.

Examining the storage details for a drive indicates whether it is configured with a RAID 0 volume.



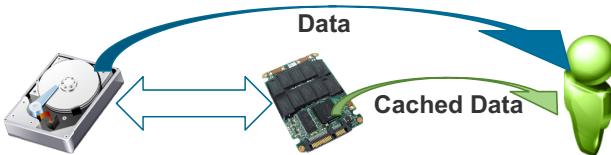
In the pass-through mode, a controller presents magnetic disks directly to the ESXi host. Otherwise, the controller requires each of the magnetic disks to be configured as a RAID 0 volume before the ESXi host can see the magnetic disks. When examining the properties of a disk using an `esxcli` command, the Model field lists LOGICAL VOLUME when the controller uses RAID 0 to present the drive.

# Performance Factors: Read Caching and Write Buffering

Slide 2-12

Storage systems use buffers to store data in memory or flash drives to improve performance:

- Read cache stores frequently accessed data in faster storage.



- Write buffers store writes to the storage system until a later time.



Read and write caches reduce the response time or latency for reading and writing to the storage system. A read cache stores frequently accessed data to faster storage locations, such as memory or SSDs, so that data does not have to be read from the slower magnetic disks. A write buffer stores incoming information in memory or to flash disks temporarily and acknowledges the write. The data is saved to the magnetic disks at a later time, which improves response times to applications that write the data.

## Performance Factors: Latency

Slide 2-13

The following factors affect latency:

- Data locality
- Disk speed
- Congestion on the storage path
- High CPU usage on the storage system
- Read cache/write buffer capacity
- Distance between the server and the storage system

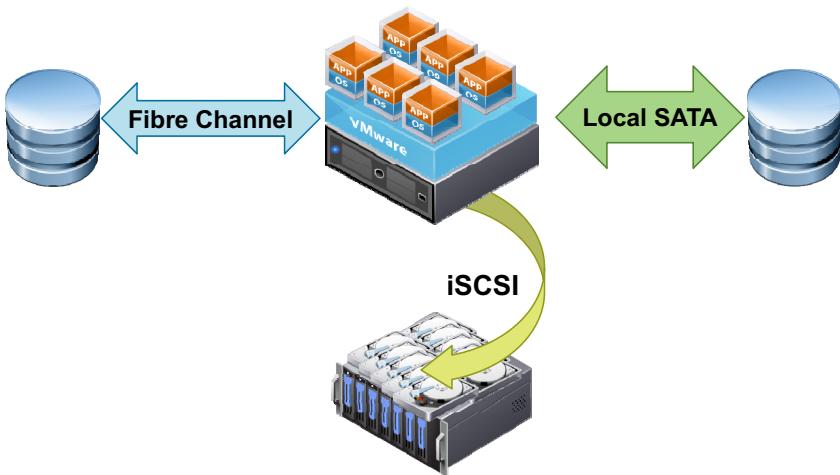
Latency is the delay between when a system requests a read or a write to a storage system and when the storage system completes the request. Several factors can affect the latency between the request and response to that request.

VMware Confidential  
Internal Use Only

## Performance Factors: Storage Path

Slide 2-14

The route that data travels between the server and the storage system is the storage path.



Based on the protocol that is used, the equipment, the size, and the number of paths can affect performance. Raw throughput speed in the storage path might not provide an accurate measurement of performance. For example, although an iSCSI solution has a capacity of 10 Gbps, the performance is reduced due to overhead for Ethernet.

# Proactive Reporting

Slide 2-15

ESXi supports disk drives that are enabled with Self-Monitoring, Analysis and Reporting Technology (SMART):

- Not all drives support the SMART feature.
- This feature is specifically useful for SSD.
- You can use the `esxcli` command to display the information for a drive.
  - `esxcli storage core device smart get -d device`

```
[root@esxi02:/dev/disks] esxcli storage core device smart get -d
t10.ATA_WDC_WD10EZEX2D00BN5A0_WD2DW
CC3F6LDJTSJ
Parameter          Value  Threshold  Worst
-----
Health Status      OK     N/A        N/A
Media Wearout Indicator  N/A     N/A        N/A
Write Error Count  N/A     N/A        N/A
Read Error Count   200    51         200
Power-on Hours    100    0          100
Power Cycle Count 100    0          100
Reallocated Sector Count 200    140        200
Raw Read Error Rate 200    51         200
Drive Temperature  121    0          107
Driver Rated Max Temperature N/A    N/A        N/A
Write Sectors TOT Count 200    0          200
Read Sectors TOT Count 100    0          253
Initial Bad Block Count N/A    N/A        N/A
[root@esxi02:/dev/disks]
```

If a drive supports Self-Monitoring, Analysis and Reporting Technology (SMART), useful information is available from the disk drive, especially solid-state disks. This information helps proactive reporting, for example, this information can help determine the prevalence of read or write errors, or drive overheating.

When RAID 0 is used on the controller to present disk devices to the ESXi host, the controller frequently blocks the SMART and disk type information. So controllers that support pass-through mode must be used where possible.

# vSphere Storage Feature Comparison

Slide 2-16

Storage Protocol	Supports Boot from SAN	Supports vSphere vMotion	Supports vSphere HA	Supports DRS	Supports Raw Device Mapping
Fibre Channel	•	•	•	•	•
FCoE	•	•	•	•	•
iSCSI	•	•	•	•	•
NFS		•	•	•	
DAS		•			•
vSphere Virtual Volumes		•	•	•	
Virtual SAN		•	•	•	

The slide shows a feature comparison between different storage solutions that are supported by vSphere.

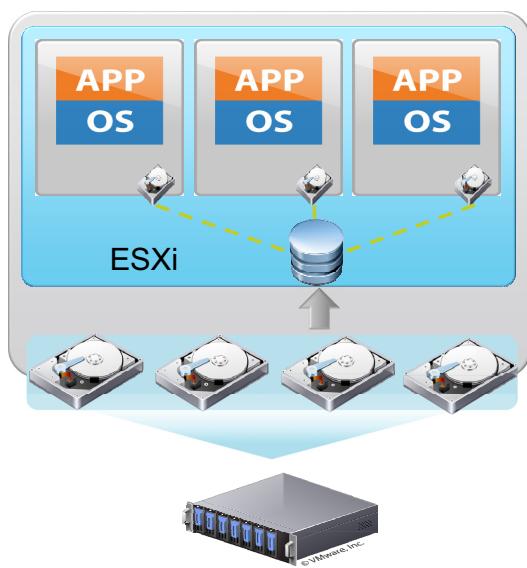
# Storage Virtualization

Slide 2-17

ESXi provides host-level storage virtualization, which logically abstracts the physical storage layer from virtual machines.

Virtual machines are encapsulated in sets of files.

Virtual machines use virtual disks to store their operating system, program files, and other data that is associated with their activities.



A virtual disk is a file or a set of files that can be copied, moved, archived, and backed up as easily as other files. You can configure virtual machines with multiple virtual disks.

To access virtual disks, a virtual machine uses virtual SCSI controllers. These virtual controllers include BusLogic Parallel, LSI Logic Parallel, LSI Logic SAS, and VMware Paravirtual.

Each virtual disk resides on a datastore that is deployed on physical storage. To the virtual machine, each virtual disk appears as a SCSI drive connected to a SCSI controller. The guest operating system and applications running on the virtual machine can determine whether storage adapters or network adapters are accessing the actual physical storage on the host.

In addition to virtual disks, vSphere offers a mechanism called raw device mapping (RDM). RDM is useful when a guest operating system in a virtual machine requires direct access to a storage device.

ESXi supports Fibre Channel, iSCSI, Fibre Channel over Ethernet (FCoE), and NFS protocols. Regardless of the type of storage device that your host uses, the virtual disk always appears to the virtual machine as a mounted SCSI device. The virtual disk hides a physical storage layer from the virtual machine's operating system. You can run operating systems that are not certified for specific storage equipment, such as SAN, in the virtual machine.

The graphic shows how storage access is different for a physical machine and a virtual machine. The operating system of a physical server directly accesses local or network-based storage devices. Storage devices are statically and individually mapped and configured per host. The operating system of a virtual machine interacts with installed hardware through the hypervisor. The hypervisor provides storage resources dynamically to virtual machines as needed to support the operation of the virtual machines. Using the hypervisor, virtual machines can operate with a degree of independence from the underlying physical hardware. For example, virtual disks can be moved from one type of storage system to another without affecting the functioning of the virtual machine.

VMware Confidential  
Internal Use Only

# vSphere Datastore

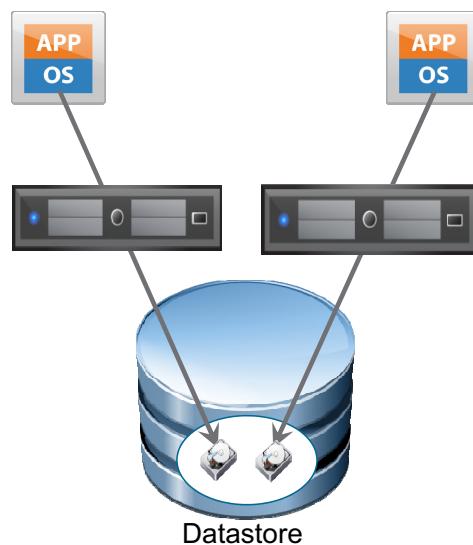
Slide 2-18

A datastore is a logical container that can use disk space on one physical device or span several physical devices.

Datastores provide the functional storage capacity to store virtual machines and other files.

vSphere provides native support for the following datastores:

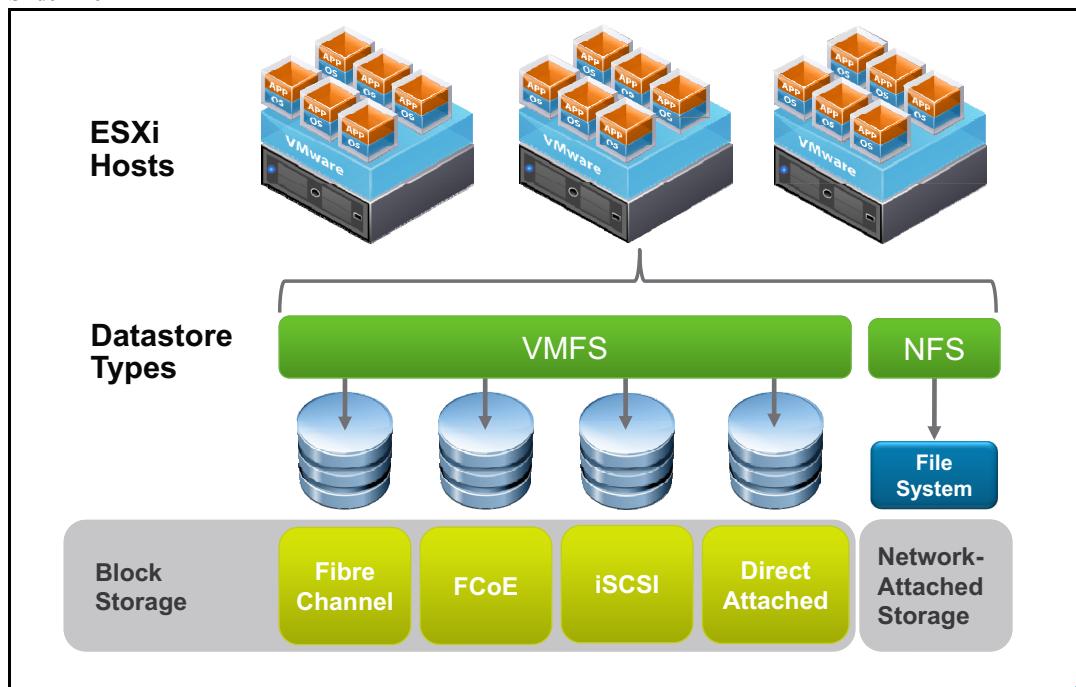
- VMware vSphere® VMFS
- NFS
- Virtual SAN
- VMware vSphere® Virtual Volumes™



When a virtual machine communicates with its virtual disk stored on a datastore, the virtual machine issues SCSI commands to its virtual disk. The hypervisor then intercepts these commands and issues its own commands (SCSI or NFS) to the appropriate datastore. Datastores can exist on various types of physical storage. Thus, these commands are encapsulated into other forms, depending on the protocol that the ESXi host uses to connect to a storage device.

# Traditional vSphere Storage Architectures

Slide 2-19



The slide shows examples of the different storage architectures that vSphere supports.

# Block Storage

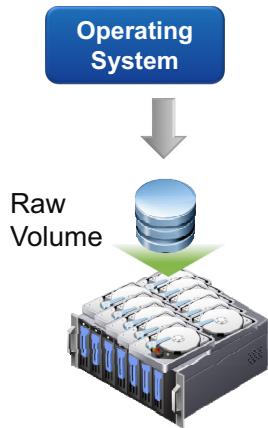
Slide 2-20

A server has block-level access to a storage device in the following scenarios:

- Local disk
- Fibre Channel
- iSCSI
- Fibre Channel over Ethernet (FCoE)

Block storage requires one of the following File systems:

- NTFS
- EXT4
- VMFS



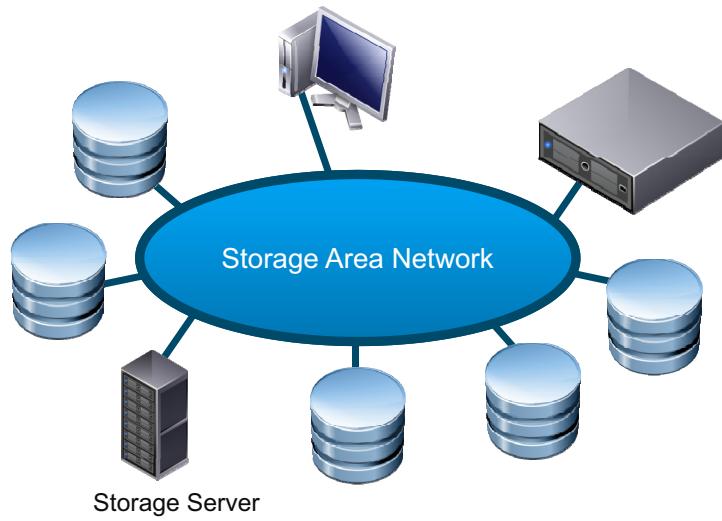
Block storage devices typically abstract data from the underlying physical hardware and move the data in a sequence of bits called a block. Block-based storage is a sequence of bytes with a fixed length (a block size). In this type of structure, data is stored in blocks. Block data is normally read in multiple blocks at a time. Block storage is normally abstracted by a file system. The physical or logical volumes accessed through block I/O might be devices internal to a server, or directly attached through SCSI or Fibre Channel. Distant block-based devices can be accessed through a SAN that uses protocols such as iSCSI and FCoE.

## Storage Area Networks

Slide 2-21

The most common types of SAN protocols include the following:

- Fibre Channel
- iSCSI
- FCoE



A SAN is a dedicated network architecture that provides access to block-based storage devices.

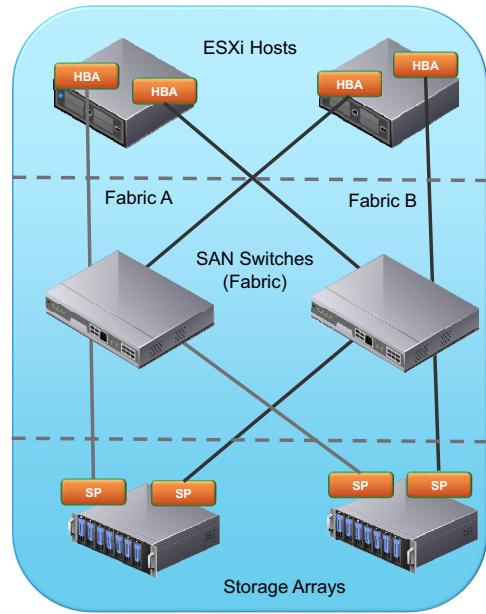
## Fibre Channel SANs

Slide 2-22

Fibre Channel SANs are a network storage solution that provide block-based storage.

vSphere provides native support for Fibre Channel protocols and speeds:

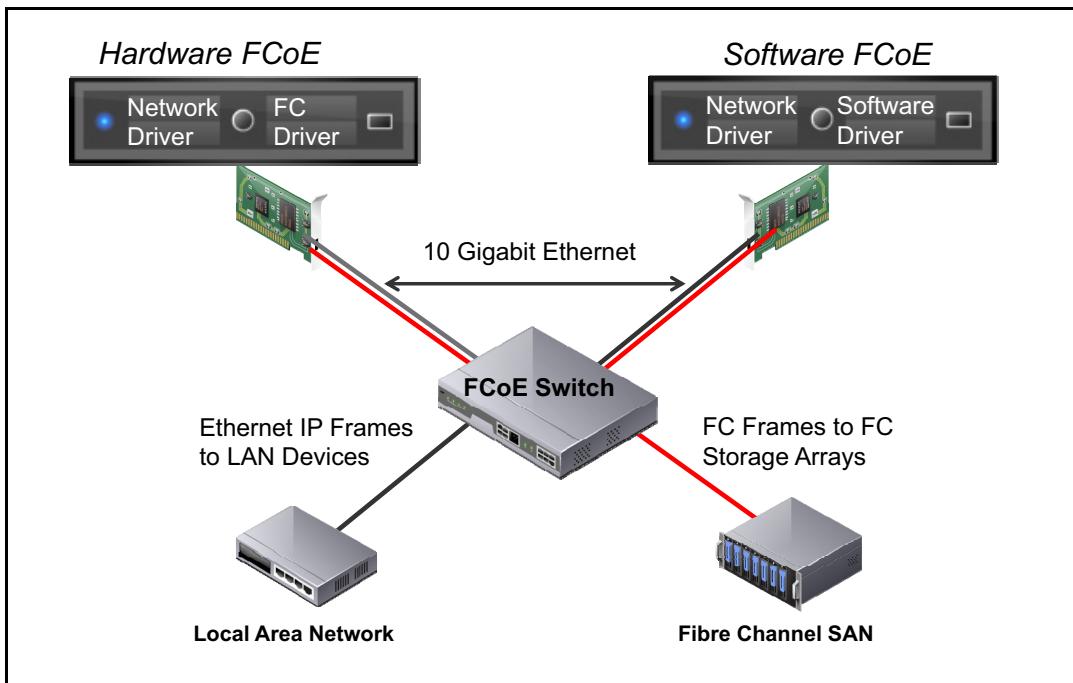
- Fibre Channel speeds from 2 Gbps to 16 Gbps
- FCoE



Fibre Channel is a high-speed network technology that is primarily used to connect storage components over a storage area network. Fibre Channel solutions require dedicated network storage devices that are not typically accessed through other LAN devices. Fibre Channel is commonly used for VMware vSphere® VMFS datastores and boot logical unit numbers (LUNs) for ESXi.

# Fibre Channel over Ethernet

Slide 2-23



FCoE adapters are used by ESXi hosts to access Fibre Channel storage over Ethernet. The Fibre Channel traffic is encapsulated into Ethernet frames. These FCoE frames are converged with networking traffic. By enabling the same Ethernet link to carry both Fibre Channel and Ethernet traffic, FCoE increases the use of the physical infrastructure. FCoE also reduces the total number of network ports.

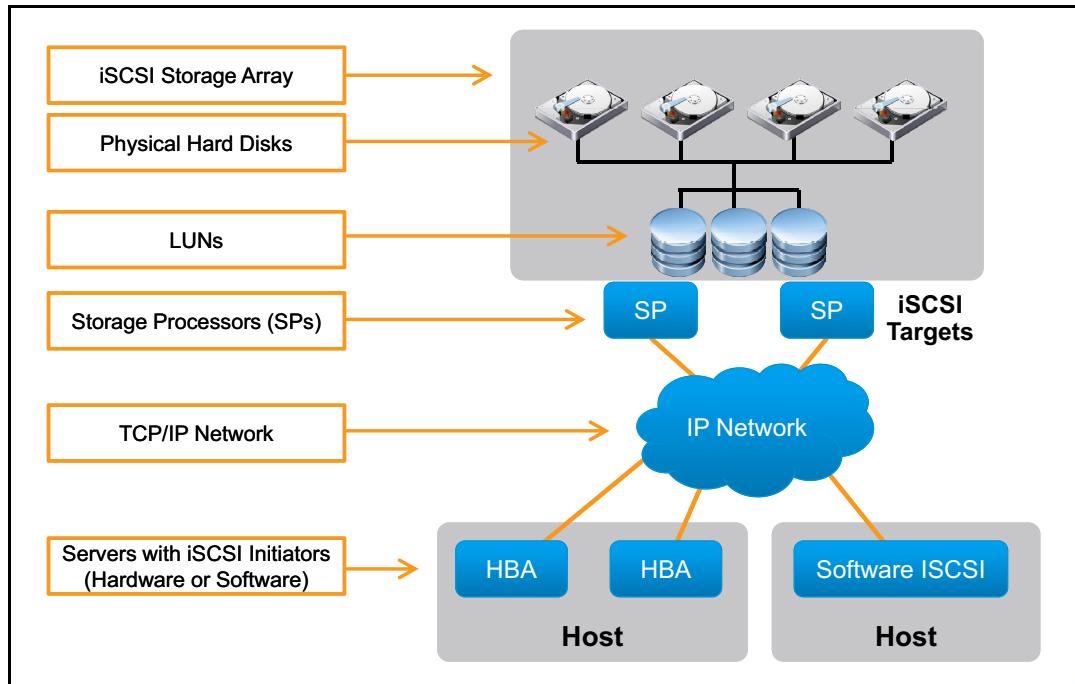
vSphere supports hardware and software FCoE adapters. Software FCoE adapters are network adapters with support for FCoE functionality. These adapters can be used with several NICs that support partial FCoE offload. Hardware FCoE adapters are called converged network adapters (CNAs). These adapters contain network and Fibre Channel functionalities on the same card. When this adapter is installed, your host detects that this adapter can use both Fibre Channel and network components. In vSphere Web Client, the networking component appears as a standard network adapter (vmnic) and the Fibre Channel component appears as an FCoE adapter (vmhba).

The main difference between a CNA and a NIC with FCoE support is that a CNA provides a high level of performance. The level of performance is not limited by the overhead of higher-layer TCP/IP protocols.

For more information about the NICs that support software FCoE, see VMware Compatibility Guide at <http://www.vmware.com/resources/compatibility>.

# iSCSI Storage Area Networks

Slide 2-24



iSCSI is an IP-based storage protocol that is used to connect storage devices over an IP network. iSCSI SANs use Ethernet connections between computers, hosts, and high-performance storage systems. iSCSI is commonly used for VMFS datastores and SAN-based boot devices for ESXi. iSCSI does not require special-purpose cables and can also be run over long distances.

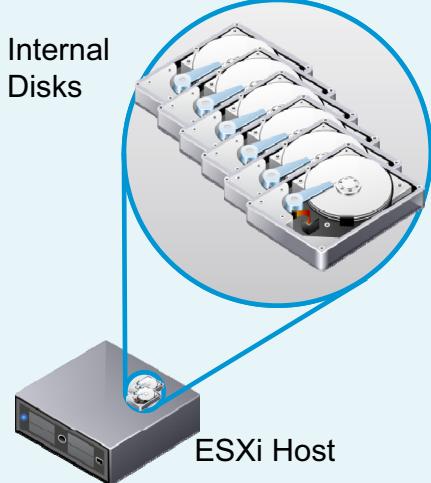
ESXi provides native support for iSCSI through the following:

- Independent iSCSI hardware initiators
- Dependent iSCSI hardware initiators
- iSCSI software initiators

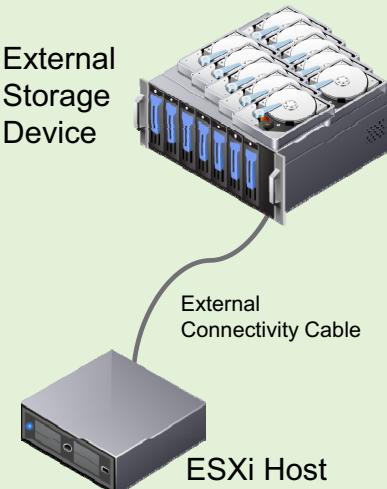
# Direct-Attached Storage

Slide 2-25

Internal direct-attached storage (DAS) devices are managed using ESXi.



External DAS devices are not managed through ESXi.



Direct-attached storage (DAS) is the most basic storage solution. DAS devices are located in a host computer or directly connected without a storage network.

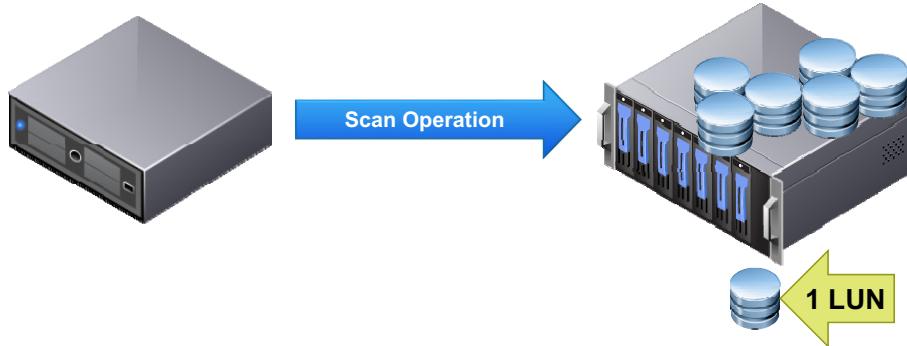
DAS solutions are low-cost storage solutions when compared to other storage area network-based solutions. DAS solutions do not require dedicated network equipment to create the storage connection.

DAS is used as a VMFS datastore and as ESXi boot devices.

## Logical Unit Number

Slide 2-26

From an ESXi host perspective, a LUN is a single raw storage block device or disk.



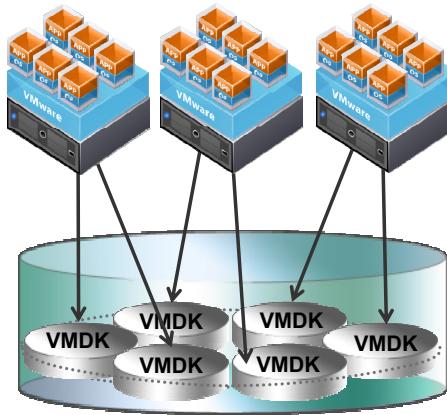
A LUN is an allocation of block storage presented to a system. LUNs are based on a unique identification given by a block device resource to a system after scanning a SAN array.

## VMFS Datastore

Slide 2-27

VMFS can be used on a wide variety of block storage devices:

- Fibre Channel SANs
- iSCSI SANs
- Local storage



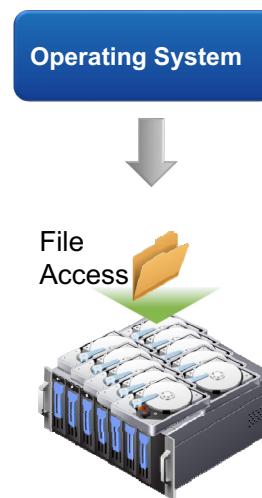
VMFS is a clustered file system that gives read and write access to multiple hosts for the same block storage devices simultaneously. vSphere currently supports VMFS 5.

# Network-Attached Storage

Slide 2-28

The storage capacity is accessed by using IP-based protocols like:

- NFS
- Apple Filing Protocol
- File Transport Protocol
- SMB



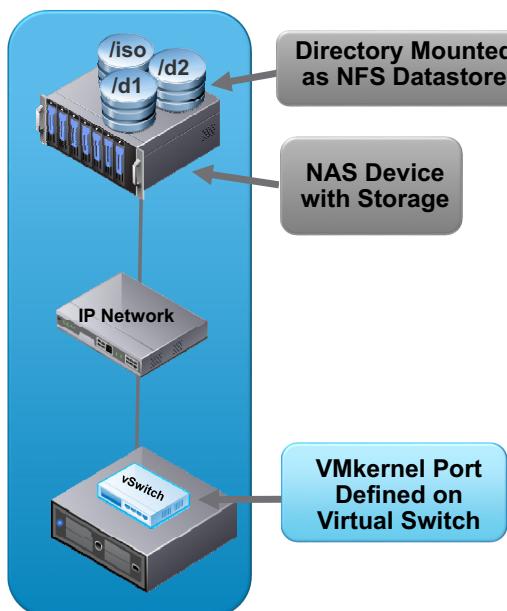
File level storage is commonly accessed through NAS devices. NAS devices typically control and manage access and the on-disk file structures. File level storage devices are easy to manage and are easily configured.

# NFS

Slide 2-29

NFS is the only NAS protocol supported by vSphere. vSphere supports the following:

- NFS versions 3 and 4.1 over TCP/IP
- Simultaneous host access to NFS volumes



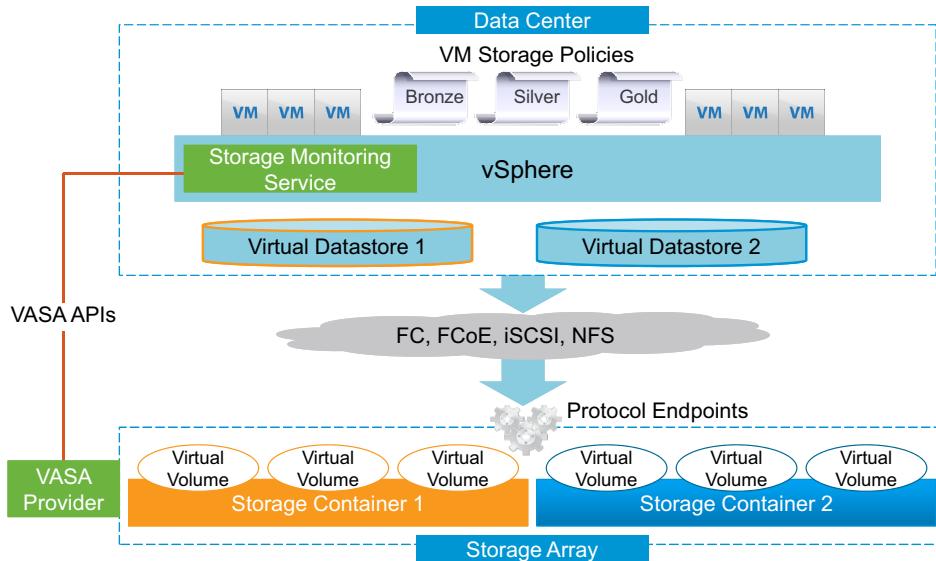
NFS is an IP-based file sharing protocol that is used by NAS systems to allow multiple remote systems to connect to a shared file system.

NFS uses file-level data access. The target NAS device controls the storage device. You cannot initialize or format a NAS target from the remote server. The NAS server is responsible for the file system where the data is stored. vSphere supports both NFS 3 and NFS 4.1.

# vSphere Virtual Volumes

Slide 2-30

vSphere Virtual Volumes defines a virtual disk container (virtual volume) that is independent of the underlying storage representation.



VMware vSphere® Virtual Volumes™ is an object that is exported by a compliant storage array and typically corresponds one-to-one with a virtual machine disk and other virtual machine-related files. A virtual volume is created and manipulated out-of-band, not in the data path, by a VASA provider. A VASA provider, or a storage provider, is developed through VMware vSphere® API for Storage Awareness™. The storage provider enables communication between the vSphere stack and the storage system.

ESXi hosts access vSphere Virtual Volumes through an intermediate point in the data path, called the protocol endpoint. Protocol endpoints establish a data path on demand from virtual machines to their respective vSphere Virtual Volumes and serve as a gateway for direct in-band I/O between ESXi hosts and the storage system.

vSphere Virtual Volumes resides in storage containers that logically represent a pool of physical disks on the storage array. In the vSphere stack, storage containers are presented as virtual datastores. A single storage container can export multiple storage capability sets. As a result, when you create a virtual machine with different performance and availability requirements, you can use different storage policies to place vSphere Virtual Volumes in the same storage container. Thus, the diverse storage needs of a virtual machine are met.

Environments that have dynamic resource requirements can benefit from vSphere Virtual Volumes and Storage Policy-Based Management. For example, during a seasonal workload increase, an automated process might update a given storage policy to deliver higher levels of performance.

Multitiered applications can benefit from using vSphere Virtual Volumes. Each virtual machine of the multitiered application can have an appropriate policy attached to it at the time of deployment, and is instantiated as a virtual volume with the required storage characteristics, even when deployed to the same container.

VMware Confidential  
Internal Use Only

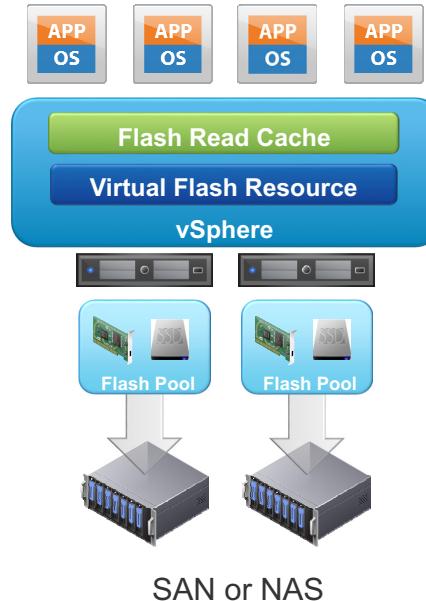
# Flash Read Cache

Slide 2-31

VMware vSphere® Flash Read Cache™ is a flash-based storage solution that is integrated with vSphere:

- Uses host-local flash devices to accelerate read operations from external storage arrays
- Provides read caching that is configurable per virtual machine disk

SSDs that are in use by Flash Read Cache cannot be used for Virtual SAN.



VMware vSphere® Flash Read Cache™ enables you to accelerate virtual machine performance by using host-resident flash devices as a cache.

You can reserve Flash Read Cache for individual virtual disks. Flash Read Cache is created only when a virtual machine is powered on. Flash Read Cache is discarded when a virtual machine is suspended or powered off. When you migrate a virtual machine, you can choose to migrate the cache.

Flash Read Cache supports read caching (or write-through), but does not support write caching (or write-back). Data reads are satisfied from the cache, if present. Data writes are dispatched to the backing storage, such as a SAN or NAS. All data that is read from or written to the backing storage is unconditionally stored in the cache.

The Flash Read Cache infrastructure includes the virtual flash resource and the flash resource management. All host flash devices can be pooled together as a virtual flash resource. Virtual machines and virtual machine disks can reserve virtual flash resources. The configuration is done in VMware vSphere® Web Client.

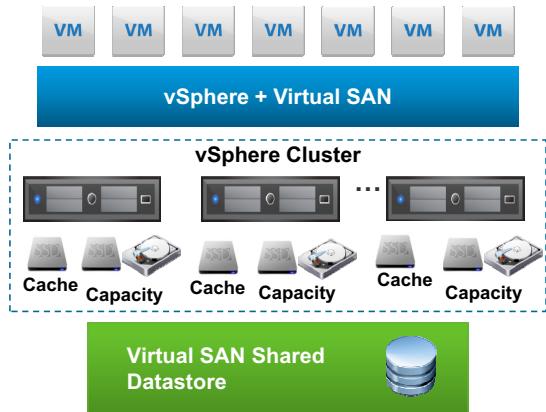
Flash Read Cache can be allocated from the same flash resource at the same time. Host swap cache reservations are immediately allocated. Flash Read Cache reservations are allocated only when the virtual machine powers on. If enough flash resource capacity is not available, the virtual machine fails to power on.

# Virtual SAN

Slide 2-32

Virtual SAN has the following features:

- Software-defined storage built in to vSphere
- Aggregates locally attached storage from each ESXi host in a cluster.
- Virtual machine-centric data operations and policy-driven management.
- Resilient design based on a distributed RAID/RAIN architecture:
  - No single points of failure
  - Fully integrated with vSphere



Virtual SAN creates a cluster of ESXi hosts. The ESXi hosts communicate with one another over a dedicated Virtual SAN network. Most of the hosts participating in the cluster have local storage devices. If a host has a local storage device, then the host must also have a local SSD to participate in the Virtual SAN cluster.

The local hard disks and the local SSD combine to form a disk group. The disk group uses the SSD for read caching and write buffering. The hard disks are used for persistent storage. Only one SSD can be used per disk group, but you can have multiple disk groups per host.

The disk groups of all the ESXi hosts in the Virtual SAN cluster are combined to create a Virtual SAN datastore. The datastore uses the storage components of each host in the cluster. Only one datastore can be created for each Virtual SAN cluster.

# Labs

Slide 2-33

Lab 1: Licensing vSphere and Virtual SAN Components

Lab 2: Basic Storage Commands

VMware Confidential  
Internal Use Only

# Lab 1: Licensing vSphere and Virtual SAN Components

Slide 2-34

Access vSphere Web Client and license vSphere and Virtual SAN components

1. Prepare the User Interface
2. Verify That the vSphere Components have Valid Licenses
3. Add New vSphere Licenses if Licenses have Expired
4. Verify That ESXi Hosts are Connected to vCenter Server
5. Add a Virtual SAN License

VMware Confidential  
Internal Use Only

## Lab 2: Basic Storage Commands

Slide 2-35

Become familiar with basic storage commands

1. (Optional) Prepare the Environment
2. Execute Basic Storage Commands

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 2-36

You should be able to meet the following objectives:

- Define common storage terminologies
- Identify characteristics of storage devices: magnetic and flash-based devices
- Identify and explain different types of storage architectures
- Identify SAN performance factors

VMware Confidential  
Internal Use Only

## Key Points

Slide 2-37

- vSphere supports a wide array of storage technologies.
- File and block storage can be used to store virtual machines.
- Shared storage, in the form of a centralized SAN or a distributed architecture, is integral to vSphere features like vSphere vMotion, vSphere HA, and DRS.

Questions?

VMware Confidential  
Internal Use Only

## MODULE 3

# Introduction to Virtual SAN

Slide 3-1

Module 3

VMware Confidential  
Internal Use Only

# You Are Here

Slide 3-2

1. Course Introduction
2. Storage Fundamentals
- 3. Introduction to Virtual SAN**
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 3-3

Understanding the logical architecture and relationships between components of Virtual SAN provides the foundation that is needed to build the software-defined data center.

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 3-4

By the end of this module, you should be able to meet the following objectives:

- Describe the Virtual SAN architecture and components
- Describe the differences between the Virtual SAN hybrid and all-flash architectures
- Describe the space efficiency features of Virtual SAN

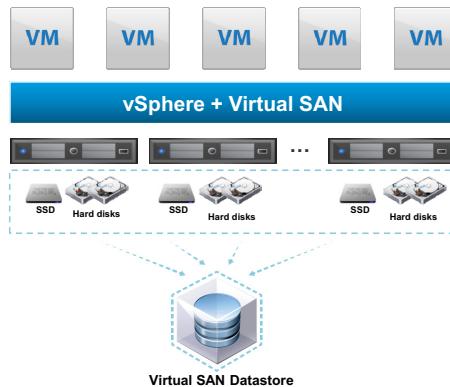
VMware Confidential  
Internal Use Only

## Virtual SAN: Hybrid Architecture

Slide 3-5

Introduced in vSphere 5.5 U1, the hybrid architecture includes a radically simple hypervisor-converged storage software:

- Pools local HDDs into a shared distributed datastore
- Ensures high performance through flash acceleration
- Highly resilient: Zero data loss in the event of some hardware failures



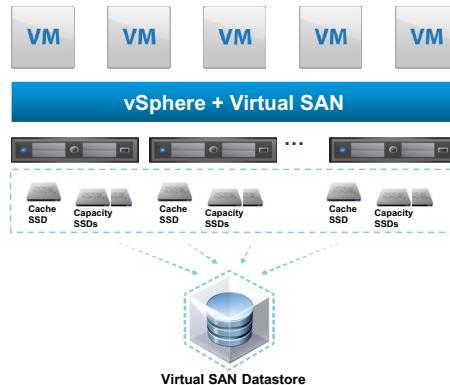
Virtual SAN aggregates the locally attached disks of hosts that are members of a vSphere cluster to create a distributed shared storage solution. Virtual SAN enables the rapid provisioning of storage in VMware vCenter Server®. Virtual SAN is the first policy-driven storage product that is designed for the vSphere environments. Virtual SAN simplifies and streamlines storage provisioning and management. Using VM-level storage policies, Virtual SAN automatically and dynamically matches with underlying storage resources.

## Virtual SAN: All-Flash Architecture

Slide 3-6

Introduced in Virtual SAN 6.0, the all-flash architecture provides better performance than the hybrid architecture:

- Uses flash devices instead of magnetic devices in the capacity tier
- Supports 10 Gb networks only

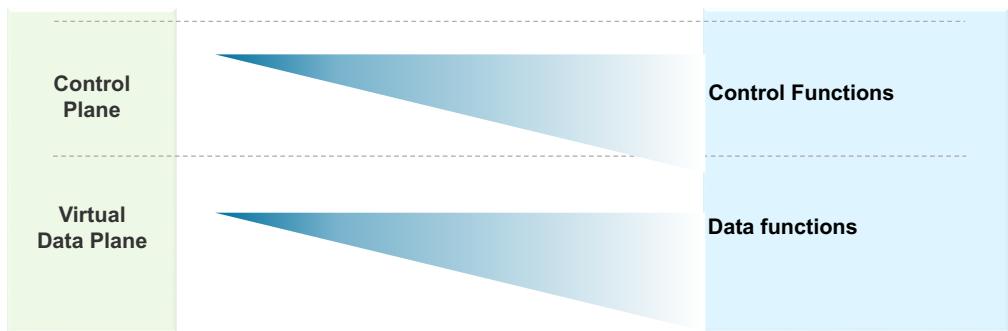


When compared with a hybrid configuration, an all-flash Virtual SAN configuration brings improved, highly predictable and uniform performance regardless of the workload.

## Control and Virtual Data Planes

Slide 3-7

The Virtual SAN architecture includes the control plane and the virtual data plane.



VMware Confidential  
Internal Use Only

# The Control Plane

Slide 3-8

The control plane acts as the bridge between applications and storage infrastructure that are responsible for controlling and monitoring storage operations.

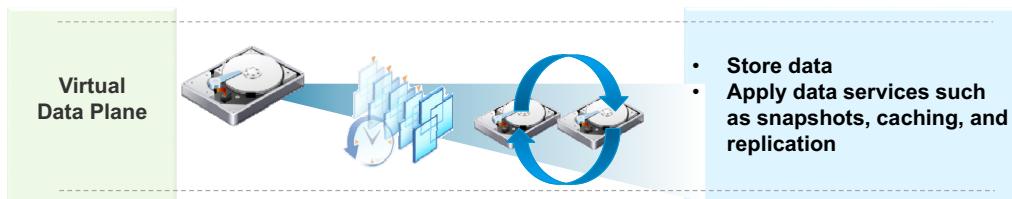


Using Virtual SAN, the storage classes of service become logical entities that are controlled entirely by policy. Defining and making adjustments to these policies automates the provisioning process at scale, while controlling individual service levels over individual virtual machines at any time.

# The Virtual Data Plane

Slide 3-9

The virtual data plane is responsible for both storing data and applying data services.



In the physical model, the data plane operates on constructs like logical unit numbers (LUNs) or storage volumes that are typically allocations of storage services. These storage services are independently defined from applications. In Virtual SAN, the data plane is virtualized by abstracting physical hardware resources and aggregating these resources into logical pools of capacity by using virtual datastores.

To simplify the delivery of storage service levels for individual applications, the virtual data plane makes the virtual disk the fundamental unit of management around which all storage operations are controlled.

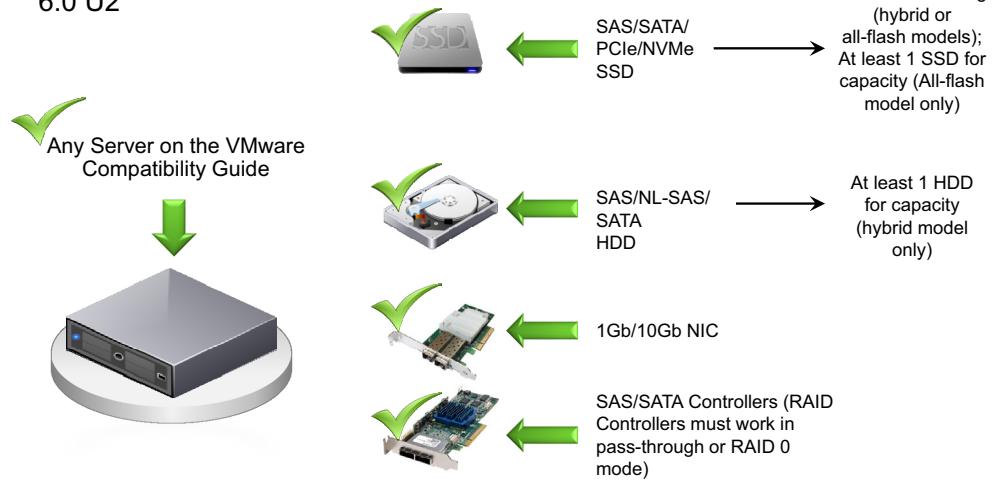
For each virtual machine that is deployed, the data services that are offered can be applied individually. Thus, each application can have unique storage service level and capabilities. Per-application storage policies ensure simpler yet individualized management of applications without mapping applications to broad infrastructure concepts like a physical datastore.

# Virtual SAN Requirements

Slide 3-10

Virtual SAN supports the following versions of vSphere:

- vSphere 5.5 U1 and 5.5 U2
- vSphere 6.0, 6.0 U1 and 6.0 U2



Virtual SAN requires several hardware components that hosts do not normally have:

- One Serial Attached SCSI (SAS) or SATA solid-state disk (SSD) or PCIe flash device and one or more magnetic drives for each hybrid disk group.
- One SAS or SATA SSD or PCIe flash device and one or more flash disks with flash capacity enabled for all-flash disk groups.
- NVMe is a specification that allows for greater parallelism to be utilized by both hardware and software, resulting in improved performance.
- Dedicated 1 Gbps network (10 Gbps recommended) for hybrid disk groups.
- Dedicated 10 Gbps network for all-flash disk groups.
- The Virtual SAN network must be configured for IPv4 or IPv6 and support multicast.

In addition, each host should have 32 GB or more of memory to accommodate a maximum number of 5 disk groups and a maximum number of 7 capacity devices per disk group.

# Virtual SAN Minimums and Maximums

Slide 3-11

Virtual SAN 6.2 supports a wide array of values.

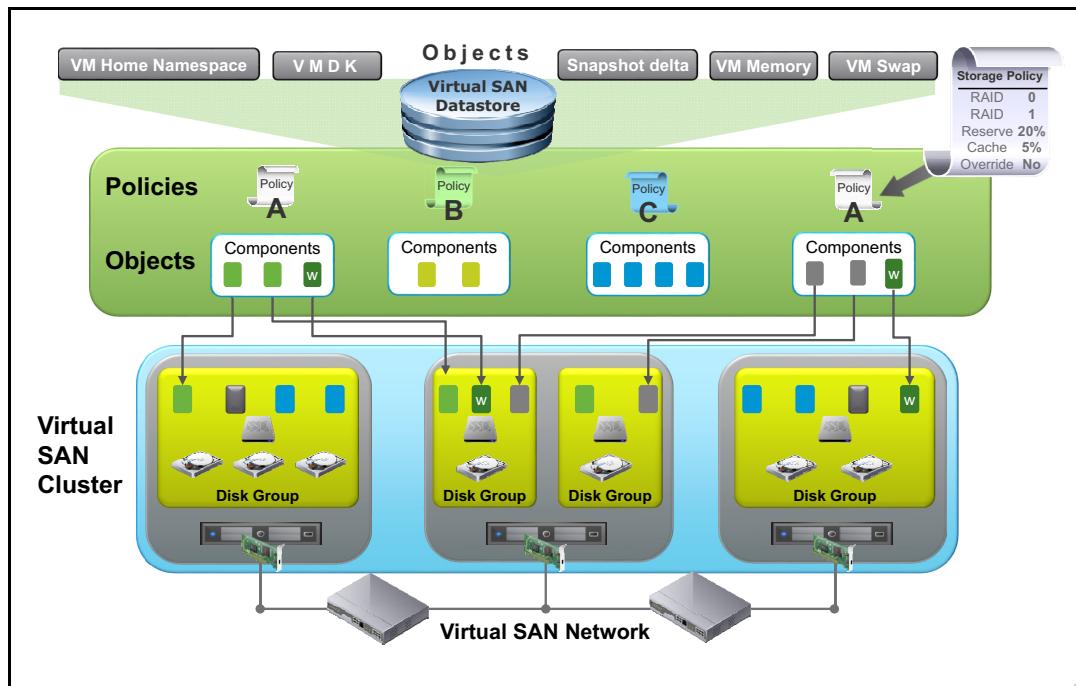
Feature	Minimum	Maximum
ESXi hosts	3	64 per Virtual SAN cluster
Virtual machines	None	6,400 per Virtual SAN cluster
Virtual machines protected by vSphere HA	None	6,400
Capacity disks (flash or magnetic)	1 per host	35 per host
Virtual SAN cache flash disks	1 per host	5 per host
Disk groups	1 per host	5 per host

Virtual SAN supports the following configurations:

- A Virtual SAN cluster supports either 64 all-flash hosts or 64 hosts that are running hybrid mode.
- Virtual SAN supports up to 200 virtual machines per host.
- vSphere HA can protect all virtual machines that are supported by the cluster.
- Up to 35 magnetic or flash disks are supported to provide storage capacity for each host.
- Up to 5 flash devices are supported for each host to use as cache disks for Virtual SAN disk groups.
- Up to 5 disk groups are supported on a host participating in a Virtual SAN cluster.

# Virtual SAN and Object-Based Storage

Slide 3-12



Virtual SAN stores and manages data as flexible data containers called objects. The object store file system manages data as objects. Each object includes its own data, part of the metadata, and a unique ID. By using this unique ID, the object can be globally addressed by more than the filename and path. The use of objects allows a detailed level of configuration on the object level, for example, RAID type or disk usage at a level higher than the physical disk blocks.

In a block-level file system, blocks are arranged in a RAID set or logical disk first and the file system is created on top of the RAID set. The file system includes the metadata or file allocation table that defines filenames, paths, and data location. In this environment, the file blocks are placed on the disk according to the file system structure and the data protection is based on the logical disk or RAID set.

# Objects (1)

Slide 3-13

Virtual machines that are created in a Virtual SAN datastore include the following objects:

- VM Home Namespace
- VMDK files
- VM swap
- Snapshot deltas
- VM memory

Virtual SAN Object	Traditional VM Files
VM Home Namespace	.nvram, .vmsd, .vmx, vmx-*.*.vsdp, .log, .hlog
VMDK	-flat.vmdk
VM swap	.vsdp
Snapshot delta	-delta-0000#.vmdk
VM memory	.vmem

An object is a VMDK file, a snapshot, or the virtual machine home folder (namespace). Objects are logical volumes with data and metadata that are distributed and accessed across the entire cluster.

When you provision a virtual machine on a Virtual SAN datastore, a set of objects are created. These objects are of the following types:

- VM Home Namespace: Stores the virtual machine metadata (configuration files).
- VMDK: Virtual machine disk.
- VM swap: Virtual machine swap file, which is created when the virtual machine is powered on.
- Snapshot deltas: Created when snapshots of the virtual machine are taken.
- VM memory: Virtual machine's memory state when a virtual machine is suspended, or when a snapshot is taken of a virtual machine and its memory state is preserved.

## Objects (2)

Slide 3-14

By default, the VM Home Namespace object is found in  
/vmfs/volumes/vsanDatastore.

The flat VMDK files or delta VMDK files do not exist in the  
vsanDatastore directory because these files are not plain files.

This virtual machine does not  
have the flat VMDK or the  
snapshot delta file.

```
# cd /vmfs/volumes/vsanDatastore/carla01
#
# ls
carla01-000001.vmdk      carla01.nvram
carla01-03920e70.vswp    carla01.vmdk
carla01-03920e70.vswp.lck carla01.vmsd
carla01-Snapshot1.vmem   carla01.vmx
carla01-Snapshot1.vmsn   carla01.vmx.lck
#
#
```

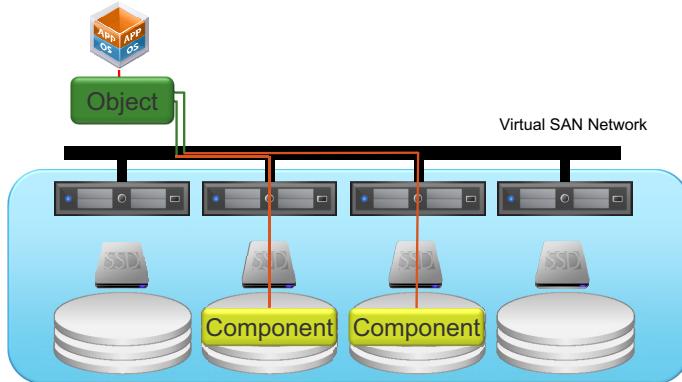
All virtual machine files (excluding VMDK files, virtual machine swap files, snapshot delta files, and the memory state file) reside in the VM Home Namespace. The VM Home Namespace includes the log files (.log and .hlog), the .nvram file, the .vmsd file, and the .vmx file.

## Components (1)

Slide 3-15

Each storage object is deployed in Virtual SAN as a RAID tree and each leaf of the RAID tree is a component:

- For example, a storage policy with a stripe width of two creates a VMDK striped across two magnetic disks.
- The VMDK is the object and stripes are components of that object.
- Components have a maximum size of 255 GB.



Components are transparently assigned caching and buffering capacity from cache devices, with their data at rest on the capacity disks. The number of components that an object is split into depends on the rules of the storage policy applied to that object.

Objects have the following characteristics:

- A VMDK can be up to 62 TB.
- Objects are limited to 255 GB in size.
- A 62 TB object equates to approximately 250 x 255 GB components.
- Each host has a maximum of 9,000 components, giving a cluster a maximum of 576,000 components.

The 62 TB VMDK limitation is the same as VMFS and NFS. Thus, virtual machines with 62 TB VMDK files can be cloned and migrated by using vSphere vMotion between Virtual SAN and other datastores.

## Components (2)

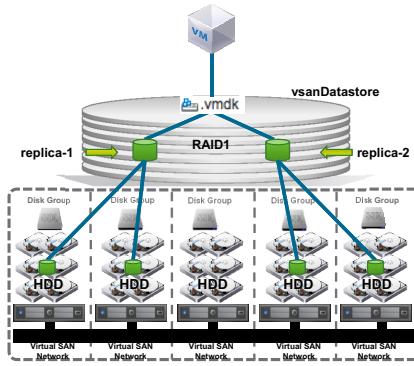
Slide 3-16

Virtual SAN uses a distributed RAID architecture to distribute data across the cluster.

Components are distributed by using the following techniques:

- Striping (RAID 0)
- Mirroring (RAID 1)
- Mirroring plus striping (RAID 10)
- Erasure coding (RAID 5/6)

The number of component replicas and copies that are created is based on the object policy definition.



Virtual SAN components are chunks of objects that are distributed across multiple hosts in a cluster. Virtual SAN can tolerate simultaneous failures and meet performance requirements. Components are created as a result of the storage policies applied to a virtual machine.

# Mirroring

Slide 3-17

Mirroring creates multiple replica copies of an object to increase the levels of availability.

The number of replicas created per object is based on the configured virtual machine storage policies.

Virtual SAN supports two-way, three-way, and four-way mirroring.

The screenshot shows the VMware vSphere Web Client interface. The left sidebar shows a tree view of the vCenter inventory, including a 'Javier Cluster' under 'SA-Datacenter'. The main content area has tabs for 'Getting Started', 'Summary', 'Monitor' (which is selected), 'Manage', and 'Related Objects'. In the 'Monitor' tab, there are sections for 'Issues', 'Performance', 'Policies', 'Tasks', 'Events', and 'Utilization'. A search bar labeled 'Filter' is at the top right. Below the tabs is a table with columns 'Name', 'VM Storage Policy', 'Compliance Status', and 'Last Checked'. It lists 'VM home' and 'Hard disk 1' as compliant. Under 'Physical Disk Placement', it shows a 'RAID 1' setup for 'javier01 - Hard disk 1' across three hosts: 'sb-esxi-03.vcl...', 'sb-esxi-04.vcl...', and 'sa-esxi-01.vcl...'. All components are listed as 'Active'.

Type	Component State	Host	Fault Domain	Cache Disk Name
RAID 1				
Component	Active	sb-esxi-03.vcl...		Local VMw...
Component	Active	sb-esxi-04.vcl...		Local VMw...
Witness	Active	sa-esxi-01.vcl...		Local VMw...

Virtual SAN uses synchronous mirroring across hosts to meet the availability and reliability policies of objects. A single-mirrored data set is commonly called RAID1. Virtual SAN supports multiple mirrors of an object. Mirroring is controlled by the `NumberofFailuresToTolerate` parameter of virtual machine storage policies.

# Striping

Slide 3-18

Striping splits the data of a given object into multiple stripes, also known as segments. Different data stripes are accessed simultaneously, increasing performance.

Virtual SAN allows for 12-way striping maximum.

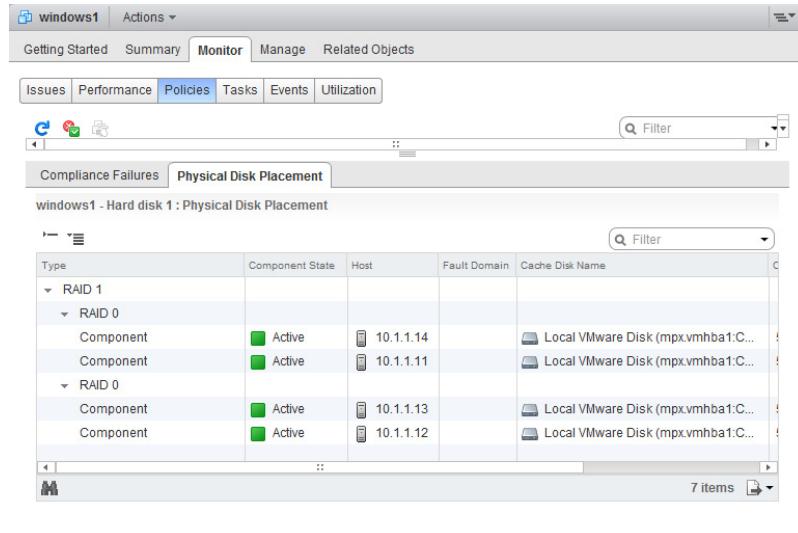
The screenshot shows the VMware Virtual SAN interface. At the top, there's a navigation bar with tabs for 'Getting Started', 'Summary', 'Monitor' (which is selected), 'Manage', and 'Related Objects'. Below the navigation bar, there are tabs for 'Issues', 'Performance', 'Policies' (selected), 'Tasks', 'Events', and 'Utilization'. A search bar labeled 'Filter' is located at the top right. The main area displays two tables. The first table, titled 'Compliance Failures', lists 'VM home' and 'Hard disk 1' under 'Name', both with 'Stripe=3 FTT=0' under 'VM Storage Policy', 'Compliant' under 'Compliance Status', and '4/5/2016 4:39 PM' under 'Last Checked'. The second table, titled 'Physical Disk Placement' for 'Hard disk 1', shows 'RAID 0' components across three hosts ('sb-esxi-04.vcl...', 'sa-esxi-02.vcl...', 'sa-esxi-01.vcl...') with 'Local VMw' cache disk names. Both tables have a '2 items' count at the bottom right.

Striping increases performance by allowing a request for data to be concurrently served by multiple Virtual SAN disk groups.

# Mirroring Plus Striping

Slide 3-19

Mirroring plus striping can be used together to provide availability and performance benefits.



Virtual SAN uses the mirroring and striping technologies together to provide redundant access to data and performance improvements. Additional drive space is consumed because of this feature.

# Erasure Coding: RAID 5

Slide 3-20

Erasure coding provides the same levels of redundancy as mirroring, but with a reduced capacity requirement.

With RAID 5, a minimum of four hosts are required.

The screenshot shows the VMware vSphere interface. On the left, the inventory tree displays a Datacenter named 'SA-Datacenter' containing a Cluster named 'Javier Cluster' with hosts 'sa-esxi-01.vclass.local', 'sa-esxi-02.vclass.local', 'sb-esxi-03.vclass.local', and 'sb-esxi-04.vclass.local'. A selected host 'javier03' is shown with its IP address '192.168.1.10'. Below it is another host 'sc-witness-01.vclass.local'. In the center, a table titled 'Physical Disk Placement' lists 'Hard disk 1' with 'Raid5' storage policy, marked as 'Compliant' with a last check date of '3/30/2016 5:58 AM'. The table has columns: Name, VM Storage Policy, Compliance Status, and Last Checked. At the bottom, a detailed view for 'javier03 - Hard disk 1: Physical Disk Placement' shows four components of type 'RAID 5' placed across four hosts ('sb-esxi-03.vcl...', 'sa-esxi-02.vcl...', 'sa-esxi-01.vcl...', 'sb-esxi-04.vcl...') with 'Active' status and assigned to 'Local VMwrs'.

Name	VM Storage Policy	Compliance Status	Last Checked
VM home	Raid5	✓ Compliant	3/30/2016 5:58 AM
Hard disk 1	Raid5	✓ Compliant	3/30/2016 5:58 AM

Type	Component State	Host	Fault Domain	Cache Disk Name
RAID 5				
Component	Active	sb-esxi-03.vcl...		Local VMwrs
Component	Active	sa-esxi-02.vcl...		Local VMwrs
Component	Active	sa-esxi-01.vcl...		Local VMwrs
Component	Active	sb-esxi-04.vcl...		Local VMwrs

In general, erasure coding is a method of taking data, breaking it into multiple pieces, and spreading it across multiple devices, while adding parity data. The method of spreading the data across multiple devices while adding parity data allows the data to be recreated if one or more of the data pieces is corrupted or lost.

Although several methods of erasure coding exist, Virtual SAN 6.2 supports a RAID 5 and RAID 6 type data placement and parity pattern as a new method of surviving failures and providing space efficiency versus mirroring.

With RAID 5, the data is placed in a 3+ 1 pattern across hosts. If a single host fails, data is still available.

# Erasure Coding: RAID 6

Slide 3-21

With RAID 6, the number of failures to tolerate is two.

A minimum of six hosts are required.

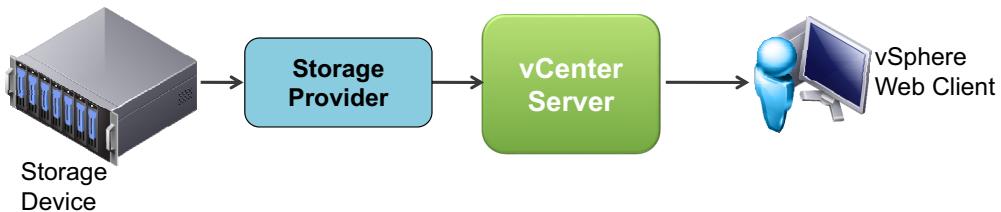
Type	Component State	Host	Fault
RAID 6			
Component	Active	sb-esxi-03.vclass.local	
Component	Active	sb-esxi-04.vclass.local	
Component	Active	sa-esxi-02.vclass.local	
Component	Active	sa-esxi-01.vclass.local	
Component	Active	sb-esxi-02.vclass.local	
Component	Active	sa-esxi-03.vclass.local	

With RAID 6, the data is placed in a  $4 + 2$  pattern across hosts. If a host fails, data is still available and protected from an additional failure.

## About vSphere API for Storage Awareness

Slide 3-22

Using VMware vSphere® API for Storage Awareness™, storage vendors can develop a software component for storage arrays that can integrate with VMware vCenter Server® through server-side plug-ins or storage providers.



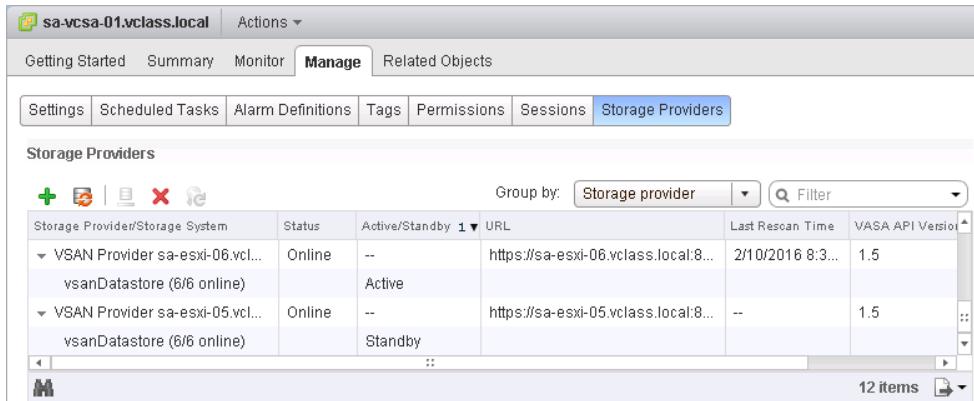
Using vSphere API for Storage Awareness, storage arrays integrate with vCenter Server for management functionality through server-side plug-ins called storage providers. Storage providers might exist on either the storage array service processor or might be standalone. For Virtual SAN, the storage provider is available on the ESXi host.

# Storage Providers

Slide 3-23

Virtual SAN automatically registers and configures a storage provider for each host in a Virtual SAN cluster as part of enabling Virtual SAN.

Virtual SAN uses vSphere API for Storage Awareness version 1.5.



The screenshot shows the vSphere Web Client interface with the following details:

- Header:** sa-vcsa-01.vclass.local, Actions ▾
- Top Navigation:** Getting Started, Summary, Monitor, **Manage** (selected), Related Objects
- Sub-navigation:** Settings, Scheduled Tasks, Alarm Definitions, Tags, Permissions, Sessions, **Storage Providers** (selected)
- Section Header:** Storage Providers
- Toolbar:** +, -, X, Filter
- Table Headers:** Storage Provider/Storage System, Status, Active/Standby, URL, Last Rescan Time, VASA API Version
- Data:** Two rows of storage providers:
  - VSAN Provider sa-esxi-06.vcl... (Online, Active/Standby: 1, URL: https://sa-esxi-06.vclass.local:8..., Last Rescan Time: 2/10/2016 8:3..., VASA API Version: 1.5)
  - VSAN Provider sa-esxi-05.vcl... (Online, Active/Standby: 1, URL: https://sa-esxi-05.vclass.local:8..., Last Rescan Time: --, VASA API Version: 1.5)
- Bottom:** 12 items, a refresh icon, and a download icon.

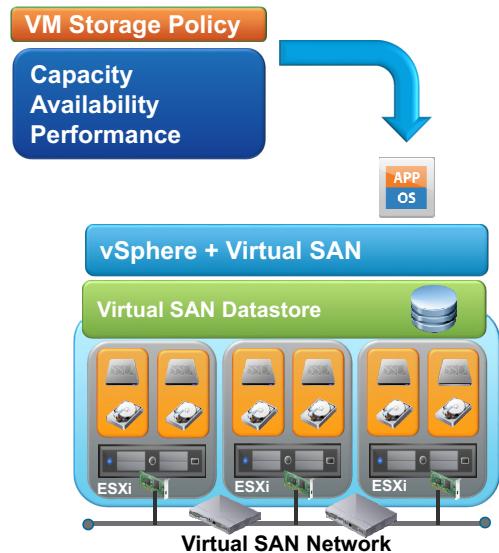
Storage providers gather information from storage arrays about available storage topology, capabilities, and state. The information includes both external storage providers, such as the storage providers that are used by vSphere Virtual Volumes, and internal storage providers, such as Virtual SAN. Unlike vSphere Virtual Volumes, the storage provider does not need to be manually registered for Virtual SAN.

# Virtual Machine Storage Policies

Slide 3-24

Storage policies define how the files that are included in a virtual machine are stored:

- Storage policies are based on storage capabilities.
- Storage policies are defined for a virtual machine at deployment time.
- Storage policies can be applied at a later date.
- Storage policies can be changed at any time.
- Storage policies cannot be deleted if they are in use.



Virtual machine storage policies are a set of rules that an administrator configures for virtual machines. Each of these storage policies reflects a set of capabilities that meet the availability, performance, and storage of the application or service-level agreement for that virtual machine.

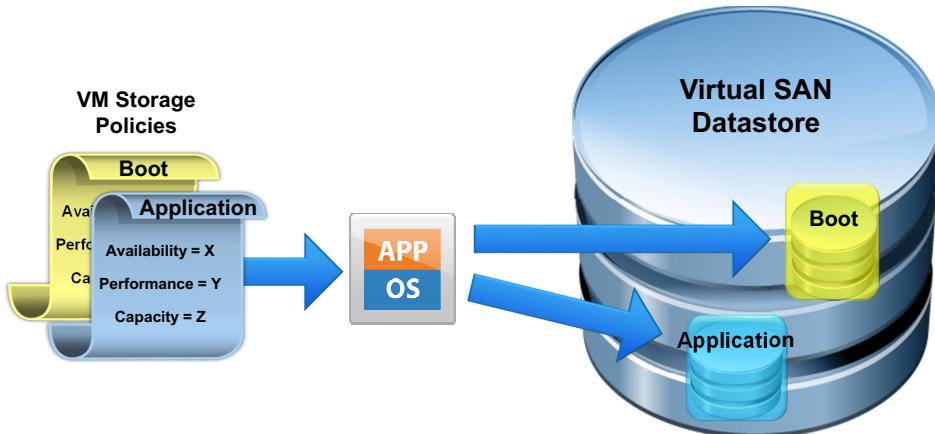
Storage policies should be created before the deployment of the virtual machines that require these storage policies. Storage policies can be applied and updated after deployment. A vSphere administrator who is responsible for the deployment of virtual machines can choose policies that are created based on the storage capabilities and for those virtual machines. Based on the policy that is selected for the object virtual machine, these capabilities are pushed back to the Virtual SAN datastore. The object is created across ESXi hosts and disk groups to satisfy these policies.

## Multiple Storage Policies

Slide 3-25

You can assign multiple Virtual SAN storage policies to a single virtual machine:

- For example, if a virtual machine has two virtual disks, then each disk can be assigned a different policy.



Virtual SAN datastores support the use of multiple policies simultaneously in the same datastore. Virtual SAN enforces different performance, availability, and capability levels for each virtual machine or each virtual disk based on the policies that are applied.

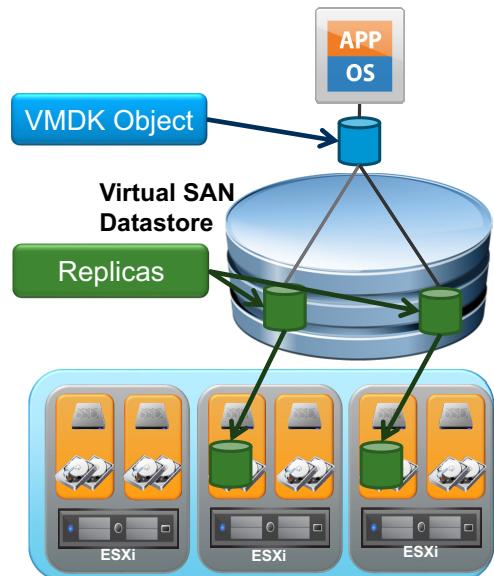
For example, a standard user workstation virtual machine has a basic policy. This virtual machine is hosted on a single disk in a single disk group. A server that provides business-critical services has a policy that requires several hardware failures to be tolerated and is also set up to be striped across multiple disks for performance reasons. Both these virtual machines can exist in the same Virtual SAN datastore with disparate storage policies.

# Replicas

Slide 3-26

Replicas are component copies of storage objects:

- Replicas are created when a virtual machine storage policy defines an availability capability.
- The availability configuration dictates the number of replicas.
- Replicas are used to provide redundancy.



If a VMDK object is mirrored, each mirror is a component of that object. Each component mirror is called a replica. Virtual machines deployed to Virtual SAN have an availability policy setting. This setting ensures that at least one additional copy of the virtual machine data is available. This setting affects both persistent data and write cache contents. If a host fails, Virtual SAN maintains a copy of the in-cache data so that no corruption happens. Virtual machines reuse the replicated copy of the cache and the replicated disk data.

# Witnesses

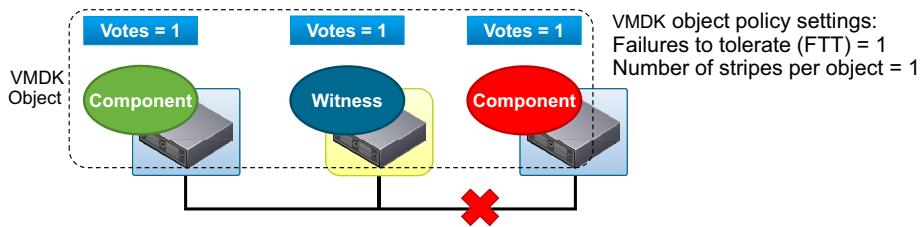
Slide 3-27

A witness is a component of an object that can serve as a tiebreaker when availability decisions are made in a Virtual SAN cluster:

- On disk, the witness component is 2 MB in size.
- Witnesses are created automatically, when needed.

Virtual SAN supports a quorum-based system:

- If a quorum exists, then the object is accessible.
- Each component has one or more votes.
- Quorum is achieved when more than 50 percent of the votes is available.
- Virtual machines can have zero or more witnesses.



The witness component contains only metadata and does not contain any actual application data.

Virtual SAN supports a quorum-based system where each component might have more than one vote to decide the availability of virtual machines. To be more precise, 50 percent of the votes that make up a virtual machine's storage object must be accessible at all times. When fewer than 50 percent of the votes is accessible to all hosts, the object is not available to the Virtual SAN datastore.

The default storage policy states that any object can sustain at least one component failure. This illustration represents RAID 1 with two replicas on two separate magnetic disks in two different hosts. A witness can be created and placed on a third host. If a component or the access to a component is lost, the witness is used. The component that can communicate with the witness is declared to have integrity and is used for all data read/write operations until the broken component is repaired.

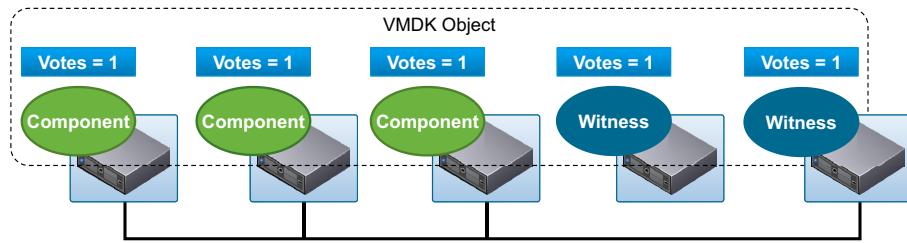
## Witness Example

Slide 3-28

Virtual SAN manages and controls the witness layout:

- Witness management is transparent to the end user.

This example shows a three-way mirror with two witnesses.



In this example, each component is given one vote. Two witnesses are used to guarantee an adequate quorum if a component failure occurs. The witness count is dependent on how the components and data are placed on the nodes in the cluster.

# Virtual SAN Disk Formats

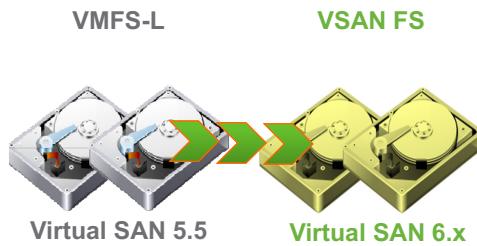
Slide 3-29

Virtual SAN 6.x supports both Virtual SAN disk formats:

- VMFS-L from Virtual SAN 5.5
- VSAN FS for Virtual SAN 6.x

The VSAN FS on-disk format improves on VMFS-L:

- Higher performance characteristics
- Efficient and scalable high performance snapshots and clones



Virtual SAN 5.5 snapshots are based on vmfsSparse. Performance degrades over time as the number of snapshots and their use increases.

Virtual SAN 6.0 introduces a new delta disk type called vsanSparse. The vsanSparse-based snapshots and clones deliver performance comparable to native SAN. Performance does not degrade based on the number of snapshots used, their size, or time of use.

You can examine the disk format in the vSphere Web Client. The Disk Format Version column indicates the current format:

- 1 indicates a VMFS-L format
- 2 indicates a VSAN FS format
- 3 indicates a VSAN FS format with deduplication and compression, and checksum features

# Disk Groups

Slide 3-30

Disk groups are Virtual SAN management constructs that include one cache device and one or more capacity devices.

A disk group requires the following:

- One flash device for caching
- One to seven capacity devices for storage
- Maximum of five disk groups per host



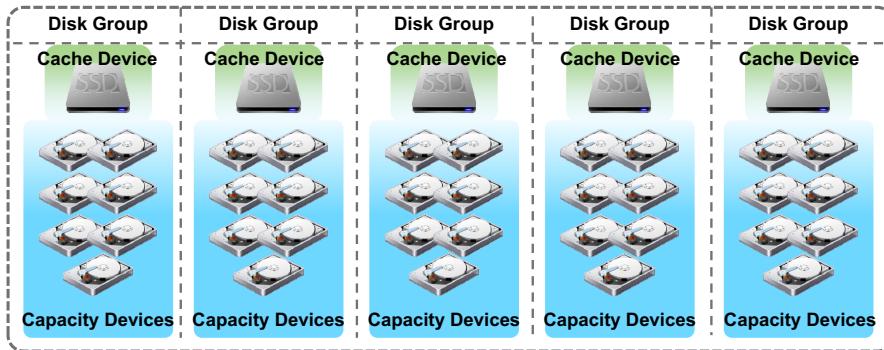
Virtual SAN uses the concept of disk groups to pool together cache devices and capacity devices as single management constructs. A disk group is a pool of one cache device and one to seven capacity devices. A host can support a maximum of five disk groups.

## Hybrid Disk Groups

Slide 3-31

The Virtual SAN hybrid disk group configurations include at least one cache device (SSD or PCI-e) and one capacity device (magnetic HDD):

- Cache devices are used for performance (read cache and write buffer):
  - 70 percent of the available cache is for frequently read disk blocks.
  - 30 percent is for write buffering.
- Capacity devices are used for storage capacity.



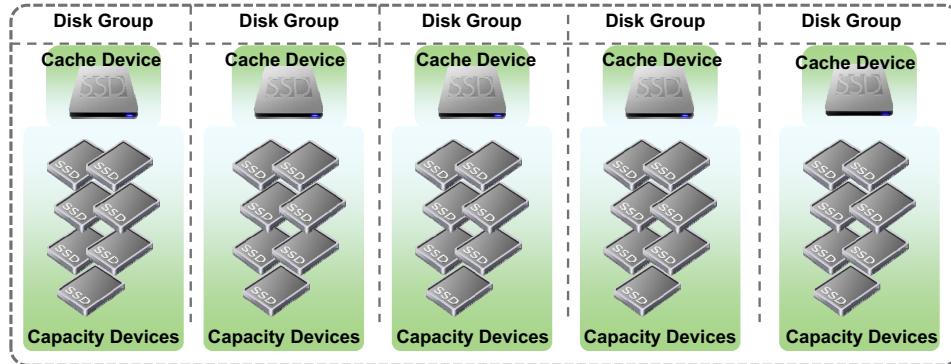
A hybrid disk group uses a single flash device to provide high-performance read caching and write buffering while using traditional magnetic drives to store data. Using a single flash device creates a distributed shared datastore that abstracts the storage hardware and provides a software-defined storage tier for virtual machines.

# All-Flash Disk Groups

Slide 3-32

The Virtual SAN all-flash disk group configurations include at least one cache flash device and at least one capacity flash device:

- Flash devices are used in a two-tier format for caching and capacity:
  - 100 percent of the available cache is for write buffering.
- The administrator decides which flash devices to use for the capacity tier.

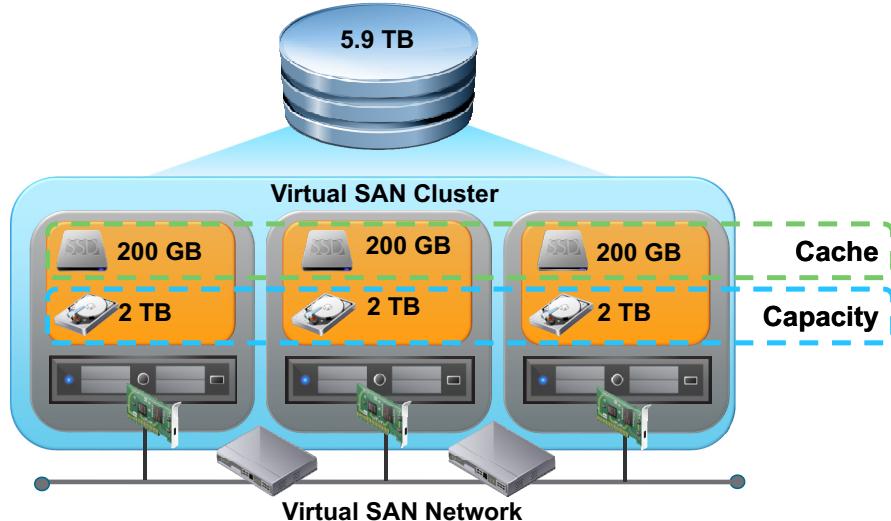


An all-flash disk group consists of one flash-based device that is used as a write cache while one to seven additional flash-based devices provide capacity. This new implementation delivers high performance in the range of 100,000 or more I/O operations per second for each host. The Virtual SAN all-flash architecture allows tiering of flash-based devices. A performance write-intensive, high-endurance caching tier for the writes and a read-intensive, durable cost-effective flash-based device tier for capacity reduces the overall cost of an all-flash architecture.

## Virtual SAN Datastore

Slide 3-33

A Virtual SAN cluster always has only one Virtual SAN datastore regardless of the number of disks and hosts in the cluster.



The disk groups of all ESXi hosts in a Virtual SAN cluster are combined to create a Virtual SAN datastore. The size of a Virtual SAN datastore is governed by the number and size of capacity disks on each ESXi host and the number of ESXi hosts in the cluster. The Virtual SAN datastore is used to store virtual machine files.

## Space Efficiency

Slide 3-34

Virtual SAN 6.2 introduces the following space efficiency features:

- Deduplication and compression
- RAID 5/6 erasure coding
- Swap object efficiency

These features help to minimize storage capacity consumption while ensuring performance and availability.

VMware Confidential  
Internal Use Only

# Deduplication and Compression

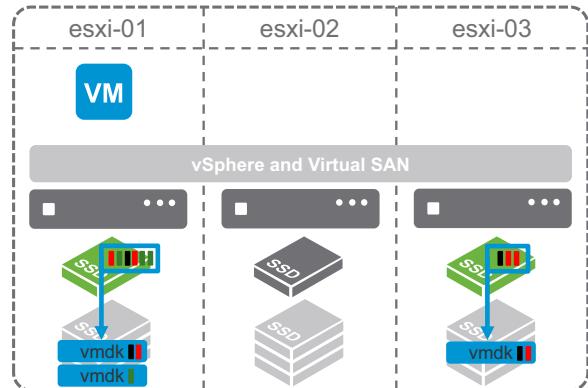
Slide 3-35

Deduplication and compression are space-saving features that can reduce the amount of storage consumption by as much as seven times:

- Actual reduction numbers vary:
  - Spreadsheet files compress well.
  - Video files do not compress well.
- Enabled at the cluster level.
- Occurs when data is destaged from the cache tier to the capacity tier.

This feature is disabled by default.

This feature is available for all-flash configurations only.



Deduplication removes redundant data blocks, whereas compression removes additional redundant data within each data block. These techniques work together to reduce the amount of space required to store the data. Virtual SAN applies deduplication followed by compression as it moves data from the cache tier to the capacity tier.

Deduplication occurs when data is destaged nearline, that is, from the cache tier to the capacity tier. The deduplication algorithm utilizes a 4K-fixed block size and is performed within each disk group. Redundant copies of a block within the same disk group are reduced to one copy, but redundant blocks across multiple disk groups are not deduplicated. Deduplication at the disk group level by using a 4K-block size provides a good balance of efficiency and performance.

The compression algorithm is applied after deduplication has occurred, before the data is written to the capacity tier. Considering the additional compute resource and allocation map overhead of compression, Virtual SAN only stores compressed data if a unique 4K block can be reduced to 2K or less. Otherwise, the block is written uncompressed.

## RAID 5/6 (Erasure Coding)

Slide 3-36

Virtual SAN 6.2 introduces RAID 5/6 (Erasure Coding):

- An alternative failure tolerance method to RAID 1 (Mirroring).
- Ensures the same levels of availability as RAID 1, while consuming less storage capacity.
- Requires additional overhead:
  - Impact to latency and IOPS are negligible due to the inherent performance of flash devices.

This feature is available for all-flash configurations only.

While mirroring techniques excel in workloads where performance is the most important factor, they are expensive in terms of capacity required. RAID 5/6 (Erasure Coding) data layout can be configured to ensure the same levels of availability, while consuming less capacity than RAID 1 (Mirroring).

Use of erasure coding reduces capacity consumption by as much as 50 percent versus mirroring at the same fault tolerance level. This method of fault tolerance requires additional write overhead in comparison to mirroring as a result of data placement and parity. This additional overhead is common among any storage platform. Because erasure coding is only supported in all-flash Virtual SAN configurations, effects to latency and IOPS are negligible in most use cases due to the inherent performance of flash devices.

# Virtual Machine Swap File Efficiency (1)

Slide 3-37

Prior to Virtual SAN 6.2, virtual machine swap files were always thick provisioned.

The storage capacity used for virtual machine swap files with Virtual SAN can be significant:

- Virtual machine swap file size equals virtual machine memory size minus memory reservation size.
- When placed on a Virtual SAN datastore, RAID 1 (mirrored) policy is always applied, regardless of the virtual machine's storage policy.

$$\begin{array}{rcl} \text{VM memory size} & - & \text{Memory reservation size} \\ 4 \text{ GB} & - & 1 \text{ GB} \end{array} = \begin{array}{l} \text{Swap file size} \\ = \\ 3 \text{ GB} \end{array}$$

With mirrored policy applied, total swap file size is 6 GB.

In larger deployments with thousands of virtual machines, this additional capacity is substantial.

Virtual swap files are created when virtual machines are powered on. If the physical host memory is exhausted, the virtual swap file is used in place of physical memory for a virtual machine. Virtual swap files are sized according to the allocated memory minus reserved memory.

A virtual machine with 4 GB of RAM allocated, with a 1 GB memory reservation, creates a 3 GB virtual swap file. On Virtual SAN, that swap file is created with a mirrored policy, resulting in 6 GB of space consumed. 100 virtual machines with the same configuration consume 600 GB of capacity. In larger deployments of hundreds or thousands of virtual machines, this additional capacity can be substantial and require significant capacity.

## Virtual Machine Swap File Efficiency (2)

Slide 3-38

With Virtual SAN 6.2, you can save disk space used by virtual machine swap files by modifying the SwapThickProvisionDisabled advanced host setting.

Use the `esxcfg-advcfg` command with the `-s` option to modify this setting.

- No reboot is necessary for the change to take effect.

```
[root@sa-esxi-01:~] esxcfg-advcfg -g /VSAN/SwapThickProvisionDisabled  
Value of SwapThickProvisionDisabled is 0  
[root@sa-esxi-01:~] esxcfg-advcfg -s 1 /VSAN/SwapThickProvisionDisabled  
Value of SwapThickProvisionDisabled is 1  
[root@sa-esxi-01:~] █
```

When this setting is enabled, virtual swap files are created as sparse objects (thin-provisioned files).

Use this setting only if virtual machines do not overcommit memory.

In addition to the use of deduplication and compression, and erasure coding, using the Virtual SAN SwapThickProvisionedDisabled advanced host setting can be advantageous as it can provide additional space savings.

Enabling this setting creates virtual swap files as a sparse object on the Virtual SAN datastore. Sparse virtual swap files consume capacity on Virtual SAN as they are accessed. As a result, significantly less space might be consumed on the Virtual SAN datastore, provided the virtual machines do not experience memory overcommitment, requiring the use of the virtual swap file.

# Virtual SAN Health Service

Slide 3-39

The Virtual SAN health service is designed to deliver troubleshooting and health reports to vSphere administrators about Virtual SAN subsystems and their dependencies.

The screenshot shows the vSphere Web Client interface for the VSAN module. The top navigation bar includes 'VSAN' (selected), 'Actions', 'Getting Started', 'Summary', 'Monitor' (selected), 'Manage', and 'Related Objects'. Below the navigation is a breadcrumb menu with 'File Compliance', 'Performance', 'Utilization', 'Tasks', 'Events', 'Resource Reservation', 'vSphere DRS', 'Virtual SAN' (selected), and navigation arrows. On the left, a sidebar lists 'Physical Disks', 'Virtual Disks', 'Resyncing Components', 'Health' (selected), 'Capacity', and 'Proactive Tests'. The main content area displays 'Virtual SAN Health (Last checked: Today at 12:54 PM)' with a 'Retest' button. A table lists test results: Failed (Cluster), Warning (Hardware compatibility, Performance service), and Passed (Network, Physical disk, Data, Limits). At the bottom, there is a message 'Select a test to view its details' and a '7 items' indicator with a download icon.

The Virtual SAN health service checks a range of different Virtual SAN statistics and provides insight into the root cause of Virtual SAN issues. When troubleshooting Virtual SAN, begin with the Virtual SAN health service. After an issue is detected, the health service highlights the problem and directs administrators to the appropriate VMware knowledge base article to begin solving problems.

## Additional Tools

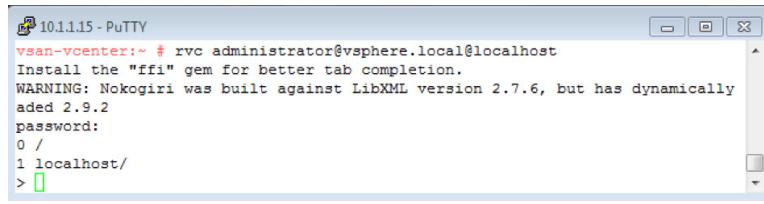
Slide 3-40

Use esxcli to run common system administration commands in VMware vSphere® ESXi™ Shell or VMware vSphere® Management Assistant.

```
[root@vsan-node4:~] esxcli storage
Usage: esxcli storage {cmd} [cmd options]

Available Namespaces:
  core          VMware core storage commands.
  nfs           Operations to create, manage, and remove Network Attached Storage filesystems.
  nfs41         Operations to create, manage, and remove NFS v4.1 filesystems.
  nmp           VMware Native Multipath Plugin (NMP). This is the VMware default implementation of the Pluggable Storage Architecture.
  san           IO device management operations to the SAN devices on the system.
  vflash        virtual flash Management Operations on the system.
  vmfs          VMFS operations.
  vvol          Operations pertaining to Virtual Volumes
  filesystem   Operations pertaining to filesystems, also known as datastores, on the ESX host.
  iofilter      IOFilter related commands.
```

Ruby vSphere Console (RVC) is a command-line console user interface for vCenter Server and VMware vCenter® Server Appliance™.



The esxcli command and Ruby vSphere Console (RVC) are valuable tools for examining details of Virtual SAN and storage that are not available in the client interface.

Use the esxcli command to run common system administration commands against ESXi systems from a machine with network access to those systems. Most vSphere command-line interface (CLI) commands also run on a vCenter Server system and target an ESXi system that is managed by the vCenter Server system. vSphere CLI includes the esxcli command set, vicfg- commands, and other commands.

Ruby vSphere Console is an interactive command-line console user interface for vSphere and vCenter Server. Ruby vSphere Console is bundled with both VMware vCenter® Server Appliance™ and the Windows version of vCenter Server.

## Review of Learner Objectives

Slide 3-41

You should be able to meet the following objectives:

- Describe the Virtual SAN architecture and components
- Describe the differences between the Virtual SAN hybrid and all-flash architectures
- Describe the space efficiency features of Virtual SAN

VMware Confidential  
Internal Use Only

## Key Points

Slide 3-42

- The data and control planes provide the necessary functions for Virtual SAN.
- Disk groups are Virtual SAN management constructs that include one cache device and one to seven capacity devices.
- Storage objects include multiple components such as replicas, stripes, and witnesses.
- Components are distributed across multiple disks and multiple ESXi hosts.
- Virtual machine storage policies are a critical part of the Virtual SAN implementation.
- All-flash configurations can take advantage of space efficiency features such as deduplication and compression and RAID 5/6 (Erasure Coding).

Questions?

VMware Confidential  
Internal Use Only

## MODULE 4

# Virtual SAN Configuration

Slide 4-1

Module 4

VMware Confidential  
Internal Use Only

# You Are Here

Slide 4-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
- 4. Virtual SAN Configuration**
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 4-3

Virtual SAN clusters have a number of specific requirements that must be implemented correctly.

Failure to properly configure the Virtual SAN cluster can affect Virtual SAN performance and availability.

VMware Confidential  
Internal Use Only

## Module Lessons

Slide 4-4

Lesson 1: Configuring a Virtual SAN Network

Lesson 2: Configuring a Virtual SAN Cluster

VMware Confidential  
Internal Use Only

# Configuring a Virtual SAN Network

Slide 4-5

## Lesson 1: Configuring a Virtual SAN Network

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 4-6

By the end of this lesson, you should be able to meet the following objectives:

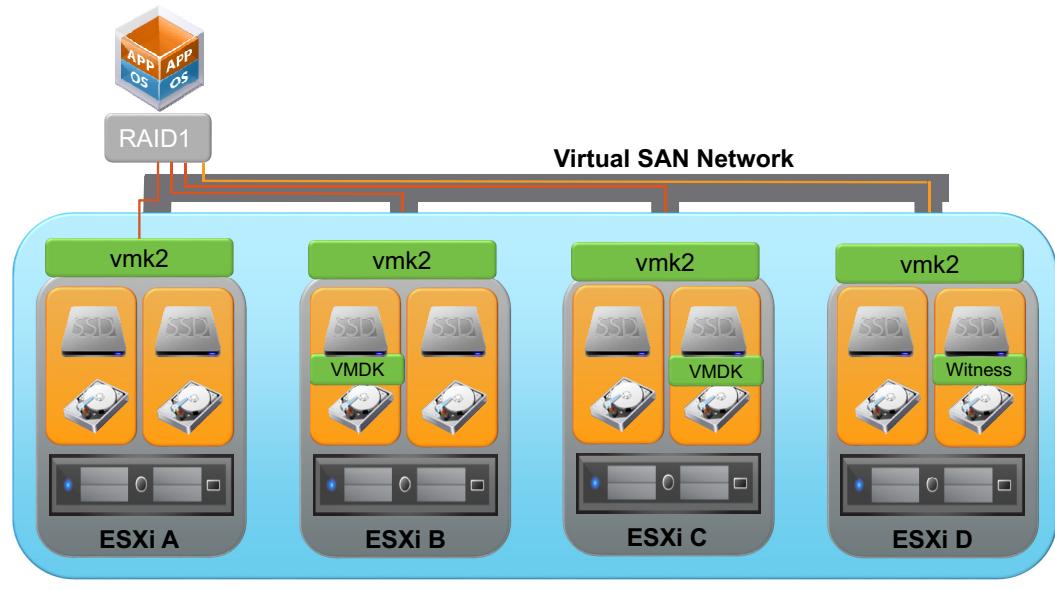
- Identify physical network configuration requirements
- Configure Virtual SAN networking

VMware Confidential  
Internal Use Only

# Virtual SAN Network Traffic Flow

Slide 4-7

The storage I/O traffic always goes over the Virtual SAN network.



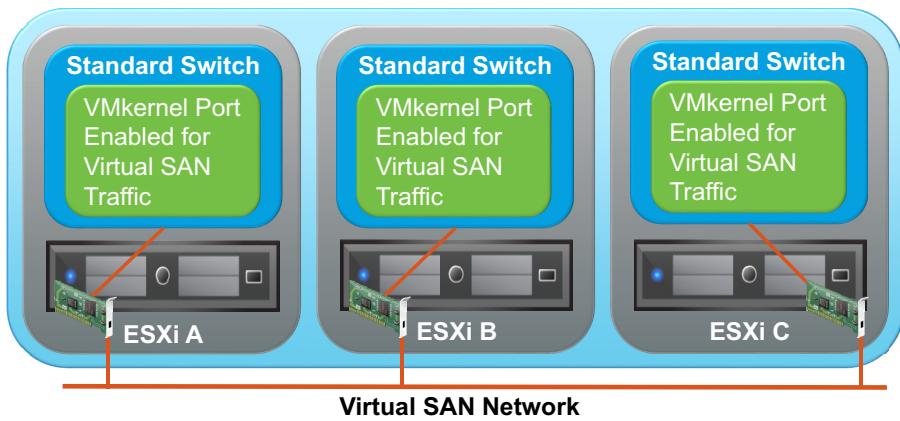
A virtual machine's compute and storage are always located on different ESXi hosts in the cluster. If a virtual machine is configured to tolerate one or more failures, objects and components can be spread across three or more nodes. All disk reads and writes for this virtual machine must traverse the Virtual SAN network. This unicast traffic forms the bulk of the Virtual SAN network traffic.

## Virtual SAN with vSphere Standard Switch

Slide 4-8

The Virtual SAN cluster hosts must have a Virtual SAN traffic-enabled VMkernel port that is connected to a virtual switch.

This network can be created with either standard switches or distributed switches. VMware recommends using distributed switches.

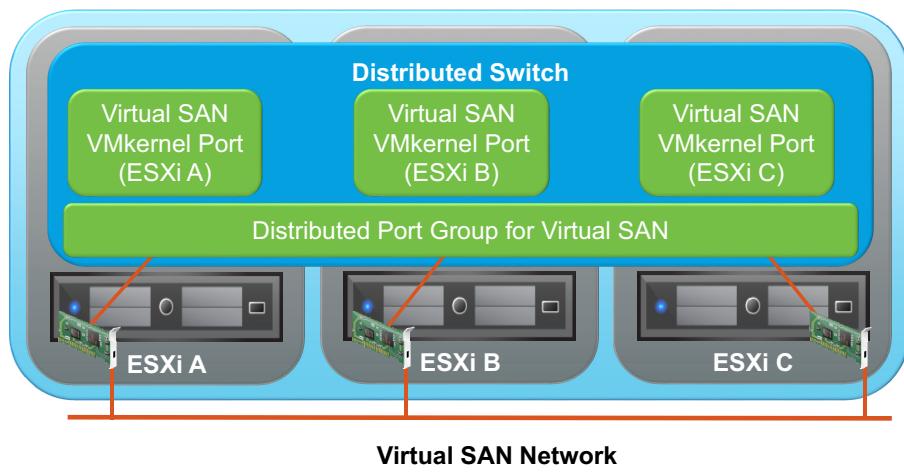


Virtual SAN supports both vSphere standard switches and vSphere Distributed Switch with either 1 Gb or 10 Gb Ethernet network uplinks. The virtual switch must be uplinked to a single Layer 2 Ethernet broadcast domain that provides connectivity to all hosts in the cluster. For the best security and performance, isolate the Virtual SAN network traffic to its own Layer 2 network segment.

# Virtual SAN with vSphere Distributed Switch

Slide 4-9

vSphere Distributed Switch simplifies data center administration by abstracting and storing the configuration of virtual networks in the vCenter Server system.



VMware recommends the use of vSphere Distributed Switch with 10 Gb network uplinks. Virtual SAN works on slower 1 Gb networks. However, the volume of traffic that is generated by Virtual SAN from replication and synchronization can saturate a 1 Gb network uplink. When using vSphere Distributed Switch, enable VMware vSphere® Network I/O Control to allocate share values to different types of traffic and mitigate the effect of contention.

## Multicast Requirement

Slide 4-10

Multicast must be enabled to support the heartbeat and exchange of metadata between the hosts in the Virtual SAN cluster.

Virtual SAN requires multicast to work:

- Layer 2 multicast must be enabled on physical switches.
- Multicast over Layer 3 is supported.

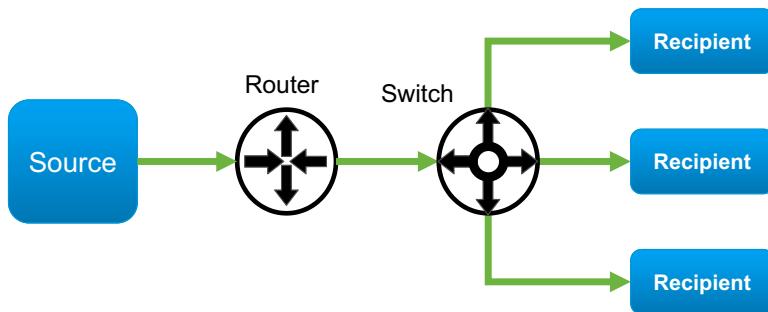
Multicast must be enabled on the physical switches and routers that handle Virtual SAN traffic for Layer 2, and optionally for Layer 3. You must configure an IGMP snooping querier on the physical switches to restrict delivery of multicast messages to the switch ports that are connected to the Virtual SAN network adapters. If multiple Virtual SAN clusters are configured on the same network, you must change the multicast address for the new cluster before you deploy an additional Virtual SAN cluster in production. Thus, the member hosts do not receive unrelated multicast messages from another cluster.

## Layer 2 Multicast

Slide 4-11

Virtual SAN requires that multicast (IGMP snooping) be enabled on the Layer 2 physical network segment that is used for Virtual SAN intracluster communication:

- If hosts participating in Virtual SAN span across multiple switches or across L3 boundaries, the network must be configured to enable multicast connectivity.
- Change the default multicast addresses if your network environment requires or if you are running multiple Virtual SAN clusters on the same L2 network.



Layer 2 multicast traffic can be limited to specific port groups by using IGMP snooping. Using the IGMP snooping feature, a network switch can listen to the IGMP conversation and maintain a map. This map provides details about the links that are needed by the different IP multicast streams. Multicasts can be filtered from the links that do not need them. Thus, the switch can control which ports receive specific multicast traffic. Virtual SAN does not require Layer 3 multicast for its network communication.

# Changing the Multicast Address

Slide 4-12

The Virtual SAN multicast address must be changed if multiple distinct Virtual SAN clusters are on the same Layer 2 network.

For details about changing the multicast address, see VMware knowledge base article 2075451 at <http://kb.vmware.com/kb/2075451>.

```
[root@sa-esxi-01:~] esxcli vsan network list
Interface
  VmNic Name: vmk1
  IP Protocol: IP
  Interface UVID: 14e4f656-a465-ad20-750b-00505601e068
  Agent Group Multicast Address: 224.2.3.4
  Agent Group IPv6 Multicast Address: ff19::2:3:4
  Agent Group Multicast Port: 23451
  Master Group Multicast Address: 224.1.2.3
  Master Group IPv6 Multicast Address: ff19::1:2:3
  Master Group Multicast Port: 12345
  Host Unicast Channel Bound Port: 12321
  Multicast TTL: 5
[root@sa-esxi-01:~]
```

If multiple Virtual SAN clusters exist on the same Layer 2 network, you must change the multicast address for each host in each cluster. Assign a unique multicast address so that each host receives traffic only for their specific cluster.

## To change the multicast address for an active Virtual SAN cluster

1. Disable Virtual SAN on the cluster.
2. Open an SSH connection to the target host in the Virtual SAN cluster.
3. Identify the VMkernel adapters for Virtual SAN.
4. Run the `esxcli vsan network list` console command and record the VMkernel adapter identifiers.
5. Run the `esxcli vsan network set` console command on each VMkernel adapter for Virtual SAN.
6. Repeat steps 2 to 5 for each host.
7. Re-enable Virtual SAN on the cluster.

## Virtual SAN Ports

Slide 4-13

Firewalls and other network filters must not block ports used for Virtual SAN traffic.

Name	Ports	Protocol	Traffic
Virtual SAN Clustering Service	12345, 23451	UDP	Multicast
Virtual SAN Transport	2233	TCP	Unicast
VASA Provider	8080	TCP	Unicast
VSAN Observer	8010	TCP	Unicast

Virtual SAN sends messages on certain ports on each host in the cluster. Verify that the host firewalls allow traffic on these ports.

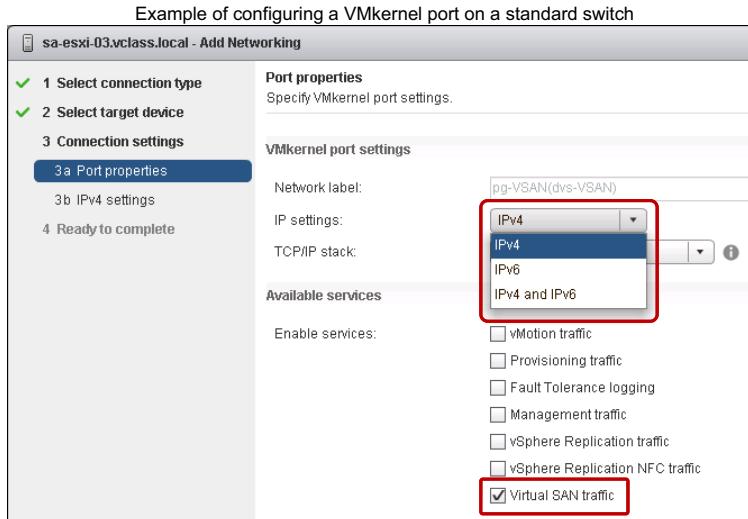
When fault domains are implemented, and the ESXi hosts are located in different data centers, the following ports must be opened between the data centers:

- Port 2233 for Reliable Datagram Transport (RDT)
- Ports 12345 and 23451 for Cluster Monitoring, Membership, and Directory Service (CMMDS)

# Configuring the Virtual SAN VMkernel Port

Slide 4-14

Each host that participates in the Virtual SAN cluster must have a VMkernel port with the Virtual SAN traffic service enabled.



Virtual SAN can support IPv4-only, IPv6-only, and IPv4/IPv6-combined networks.

When creating a VMkernel port for Virtual SAN, you must enable the Virtual SAN traffic on the Port properties page.

IPv6 is supported to address the needs of organizations that are moving to IPv6, such as service providers and government organizations.

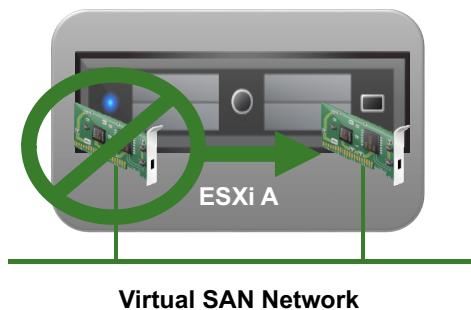
# Physical Network Adapters and NIC Teaming

Slide 4-15

NIC teaming provides Virtual SAN availability and redundancy.

Virtual SAN supports configuring a physical network as a failover adapter:

- The additional adapter is not used for load balancing.
- 10 Gb Ethernet adapters are recommended for best performance.
- You cannot have two VMkernel ports on the same subnet.



Network adapter teaming is supported in Virtual SAN as an availability and redundancy measure. Virtual SAN does not team physical network adapters for bandwidth aggregation. You can use Link Aggregation Control Protocol, or LACP, to provide network redundancy for the Virtual SAN network traffic. Thus, multiple VMkernel ports can be bound together in a NIC team by using a Route based on IP hash policy. However, a single misconfigured uplink on one host can prevent the Virtual SAN cluster from being formed. So administrators must document each uplink that is used for the Virtual SAN traffic.

## Lab 3: Configuring a Virtual SAN Network

Slide 4-16

### Configure the vSphere network for Virtual SAN

1. (Optional) Prepare the Environment
2. Create a Distributed Switch and Virtual SAN Distributed Port Group
3. Add Hosts to the Distributed Switch and Assign an Uplink to the Virtual SAN Port Group
4. Add a VMkernel Adapter for Virtual SAN Traffic
5. Add a vSphere vMotion Port Group
6. Assign an Uplink to the vSphere vMotion Port Group
7. Add a VMkernel Adapter for vSphere vMotion Traffic
8. Use esxcli to Check Virtual SAN Networking

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 4-17

You should be able to meet the following objectives:

- Identify physical network configuration requirements
- Configure Virtual SAN networking

VMware Confidential  
Internal Use Only

## Configuring a Virtual SAN Cluster

Slide 4-18

### Lesson 2: Configuring a Virtual SAN Cluster

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 4-19

By the end of this lesson, you should be able to meet the following objectives:

- Configure a Virtual SAN cluster
- Test and validate the Virtual SAN configuration and functionality

VMware Confidential  
Internal Use Only

## Virtual SAN Cluster Requirements

Slide 4-20

Virtual SAN can be enabled and made functional during cluster creation, provided the requirements are met:

- A minimum of three ESXi hosts (nodes) with the required cache and capacity disks:
  - A two-node configuration is supported, but requires a witness host.
- All hosts must have hardware validated with the VMware Compatibility Guide for Virtual SAN.
- All hosts must be connected to a VMkernel port with Virtual SAN traffic enabled.

When creating a Virtual SAN cluster, at least three ESXi hosts must be available to join the cluster. Each of these hosts must have a minimum of one compatible cache and one capacity drive to create a disk group. Virtual SAN networking must be configured for all hosts.

VMware Confidential  
Internal Use Only

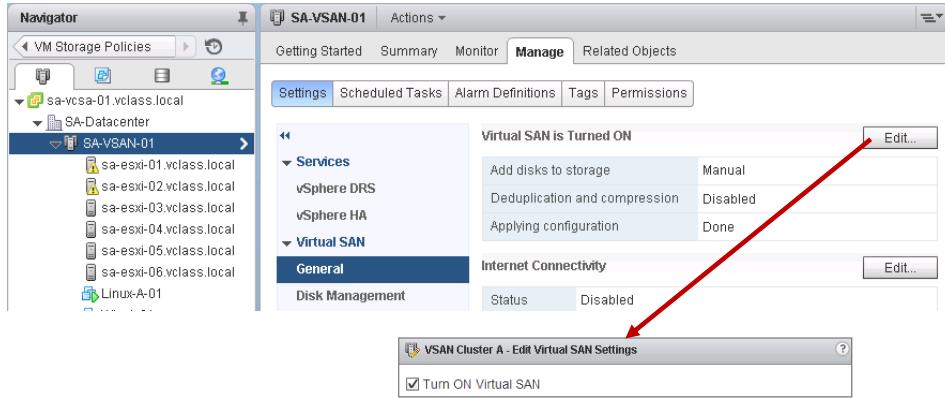
# Enabling Virtual SAN on a vSphere Cluster

Slide 4-21

Virtual SAN can be enabled on a new or an existing cluster:

- All requirements for a Virtual SAN cluster must be met.
- VMware vSphere® High Availability must be disabled.

Example of enabling Virtual SAN on an existing cluster

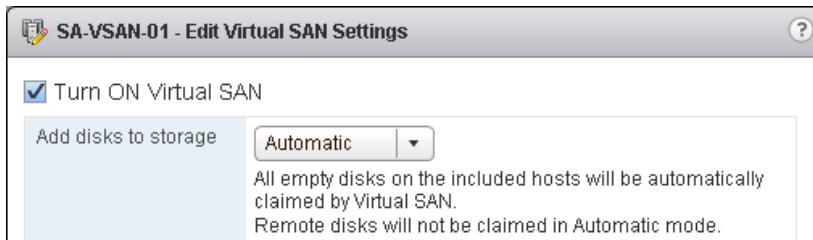


When configuring an existing cluster for Virtual SAN, the cluster must meet the requirements for a new Virtual SAN cluster. Additionally, if vSphere HA is enabled on an existing cluster, vSphere HA must be disabled before Virtual SAN can be enabled. After Virtual SAN is enabled on the cluster, vSphere HA can be enabled.

# Using Automatic Disk Claim Mode

Slide 4-22

With the Automatic disk claim mode, the storage cluster is created and all available local disks are used to create as many disk groups as each host supports.



Automatic disk claim mode is the default mode when creating a Virtual SAN cluster through vSphere Web Client.

If **Automatic** disk claim mode is selected when creating a Virtual SAN cluster, Virtual SAN discovers local empty disks on each host. Virtual SAN builds disk groups on each host in the cluster. Because each disk group can only contain a single cache solid-state drive (SSD), multiple disk groups might exist. Each of these disk groups might contain one cache SSD and one or more capacity disks. After the disk groups are created, a datastore is created. The size of the datastore reflects the capacity of all the capacity disks across all the hosts in the cluster.

If a host is added to the Virtual SAN cluster and Automatic disk claim mode is selected, the empty local storage on this host is claimed by Virtual SAN. Disk groups are created and the Virtual SAN datastore expands accordingly. No additional administrative tasks are required other than moving the new ESXi host into the Virtual SAN cluster.

## Using Manual Disk Claim Mode

Slide 4-23

When the Manual disk claim mode is selected, the Virtual SAN datastore is created, but its initial size is 0 bytes. Disk groups need to be created on each participating host to specify which drives to use for cache and capacity.



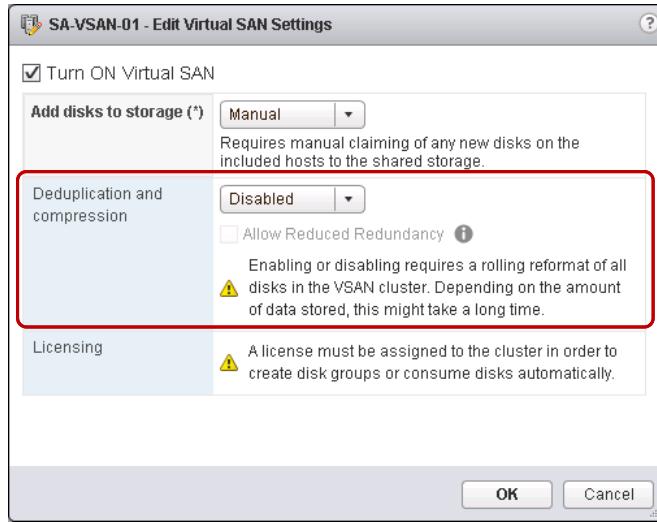
When **Manual** disk claim mode is selected, the Virtual SAN cluster is created. Virtual SAN creates a single distributed datastore but does not automatically create any disk groups. The resulting datastore's size is 0 bytes. After the cluster is created, the administrator must manually create disk groups per host. As disk groups are created, the size of the datastore expands according to the amount of capacity that is added.

# Enabling Deduplication and Compression

Slide 4-24

Deduplication and compression are enabled as a single feature.

The feature is disabled by default.



The deduplication and compression feature is a cluster-wide setting. A rolling reformat of every disk group on every host in the Virtual SAN cluster is required, which can take a considerable amount of time. However, this process does not incur virtual machine downtime. Deduplication or compression cannot be enabled individually.

If you enable deduplication and compression, you can select the **Allow Reduced Redundancy** check box. This option allows Virtual SAN to reduce the protection level of the virtual machines, if necessary, while enabling deduplication and compression. You can use this option only if your setup is at the limit of the protection level configured by the storage policy (for example, using the failures to tolerate setting or the failure tolerance method setting). For example, you might allow reduced redundancy if a virtual machine is configured to use RAID 5/6 failure tolerance method, and the cluster has four hosts.

# Creating Disk Groups

Slide 4-25

Disks are assigned to disk groups, either for caching purposes or capacity.

One individual disk can only be used in one disk group.

The screenshot shows a software interface titled "VSAN - Claim Disks for Virtual SAN Use". It displays a list of disks grouped by model and size. A red box highlights the "Claim For" column, which contains three options: "Do not claim" (unchecked), "Cache tier" (checked), and "Capacity tier" (unchecked). The table includes columns for Disk Model/Serial Number, Claim For, Drive Type, Total Capacity, Disk Distribution/Host, and Transport Type. The "Disk model/size" filter is set to "Disk model/size".

Disk Model/Serial Number	Claim For	Drive Type	Total Capacity	Disk Distribution/Host	Transport Type
▶ VMware Virtual disk , 6.00 GB disks	<input type="checkbox"/> Do not claim	HDD	48.00 GB	2 disks on 4 hosts	Parallel SCSI
▶ F VMware Virtual disk , 4.00 GB disks	<input checked="" type="checkbox"/> Cache tier	Flash	16.00 GB	1 disk on 4 hosts	Parallel SCSI
▶ F VMware Virtual disk , 6.50 GB disks	<input type="checkbox"/> Capacity tier	Flash	52.00 GB	2 disks on 4 hosts	Parallel SCSI
▶ F VMware Virtual disk , 4.50 GB disks	<input type="checkbox"/> Capacity tier	Flash	18.00 GB	1 disk on 4 hosts	Parallel SCSI

You can select which disks should be claimed for the cache tier and which disks should be claimed for the capacity tier in the Virtual SAN cluster. A recommended selection has been made for you based on the available devices in your environment. You can modify the selection if necessary.

The number of capacity disks must be greater than or equal to the number of cache disks claimed per host.

# Disk Grouping by Disk Model or Size

Slide 4-26

Disks can be grouped by drive type and drive capacity.

The screenshot shows the 'VSAN - Claim Disks for Virtual SAN Use' interface. At the top, it says: 'Select which disks should be claimed for cache and which for capacity in the VSAN cluster. The disks below are grouped by model and size or by selection has been made based on the available devices in your environment. The number of capacity disks must be greater than or equal to the number of cache disks claimed per host.' Below this is a table with columns: Disk Model/Serial Number, Claim For, Drive Type, Total Capacity, Disk Distribution/Host, and Transport Type. A dropdown menu 'Group by:' is set to 'Disk model/size'. The table data is as follows:

Disk Model/Serial Number	Claim For	Drive Type	Total Capacity	Disk Distribution/Host	Transport Type
▶ VMware Virtual disk , 6.00 GB disks	Do not claim	HDD	48.00 GB	2 disks on 4 hosts	Parallel SCSI
▶ F VMware Virtual disk , 4.00 GB disks	Cache tier	Flash	16.00 GB	1 disk on 4 hosts	Parallel SCSI
▶ F VMware Virtual disk , 6.50 GB disks	Capacity tier	Flash	52.00 GB	2 disks on 4 hosts	Parallel SCSI
▶ F VMware Virtual disk , 4.50 GB disks	Capacity tier	Flash	18.00 GB	1 disk on 4 hosts	Parallel SCSI
F Local VMware Disk (mpx.vmhba1:C0:T4:L0)	Capacity	Flash	4.50 GB	sa-esxi-01.vcl...	Parallel SCSI
F Local VMware Disk (mpx.vmhba1:C0:T4:L0)	Cache tier	Flash	4.50 GB	sa-esxi-02.vcl...	Parallel SCSI
F Local VMware Disk (mpx.vmhba1:C0:T4:L0)	Capacity tier	Flash	4.50 GB	sb-esxi-03.vcl...	Parallel SCSI
F Local VMware Disk (mpx.vmhba1:C0:T4:L0)	Do not claim	Flash	4.50 GB	sb-esxi-04.vcl...	Parallel SCSI

You can claim disks for a disk group in bulk, based on disk model and size. Or, you can claim disks individually. Claiming the disks that are members of the groups individually is better when you want to control the capacity disks that are paired with specific cache disks.

# Disk Grouping by Hosts

Slide 4-27

Disks can also be grouped by hosts.

The screenshot shows a VMware interface titled "VSAN - Claim Disks for Virtual SAN Use". It displays a list of disks grouped by host. The top section contains a message: "Select which disks should be claimed for cache and which for capacity in the VSAN cluster. The disks below are grouped by model and size or k selection has been made based on the available devices in your environment. The number of capacity disks must be greater than or equal to the number of cache disks claimed per host." Below this is a toolbar with icons for refresh, search, and other functions. A dropdown menu labeled "Group by:" is set to "Host". The main area is a table with columns: Name, Claim For, Drive Type, Total Capacity, and Transport Type. The "Name" column lists disk names grouped by host: "sa-esxi-01.vclass.local", "sa-esxi-02.vclass.local", "sb-esxi-03.vclass.local", and "sb-esxi-04.vclass.local". The "Claim For" column shows "Custom" for all. The "Drive Type" column includes "HDD" and "Flash". The "Total Capacity" column shows values like 6.00 GB, 4.00 GB, etc. The "Transport Type" column shows "Parallel SCSI" for all. A red box highlights the host group "sa-esxi-01.vclass.local" and the "Group by:" dropdown.

Name	Claim For	Drive Type	Total Capacity	Transport Type
sa-esxi-01.vclass.local	Custom			
sa-esxi-02.vclass.local	Custom			
sb-esxi-03.vclass.local	Custom			
sb-esxi-04.vclass.local	Custom			
Local VMware Disk (mpx:vhba1:C0:T2:L0)	Do not claim	HDD	6.00 GB	Parallel SCSI
Local VMware Disk (mpx:vhba1:C0:T1:L0)	Cache tier	Flash	4.00 GB	Parallel SCSI
Local VMware Disk (mpx:vhba1:C0:T6:L0)	Capacity tier	Flash	6.50 GB	Parallel SCSI
Local VMware Disk (mpx:vhba1:C0:T5:L0)	Capacity tier	Flash	6.50 GB	Parallel SCSI
Local VMware Disk (mpx:vhba1:C0:T4:L0)	Capacity tier	Flash	4.50 GB	Parallel SCSI
Local VMware Disk (mpx:vhba1:C0:T3:L0)	Do not claim	HDD	6.00 GB	Parallel SCSI

You can claim disks for a disk group based on host instead of disk model and size. You can claim disks in bulk or individually on a host-by-host basis.

# Creating All-Flash Disk Groups

Slide 4-28

All-Flash disk groups use an SSD cache and one or more SSDs as capacity devices.

The `vdq` command can be used to verify whether an SSD is marked as a cache device or capacity device.

```
[root@sa-esxi-01:~] vdq -qH
DiskResults:
  DiskResult[0]:
    Name: mpx.vmhba1:C0:T2:L0
    VSANUUID: 52372898-b16e-de4e-0b87-7584078eaa75
    State: In-use for VSAN
    Reason: None
    IsSSD?: 1
    IsCapacityFlash?: 1
    IsPDL?: 0
      Yes, SSD is a capacity device.

  DiskResult[1]:
    Name: mpx.vmhba1:C0:T1:L0
    VSANUUID: 528e7f13-18c1-b195-d23c-3dd1eaf93c6d
    State: In-use for VSAN
    Reason: None
    IsSSD?: 1
    IsCapacityFlash?: 0
    IsPDL?: 0
      No, SSD is not a capacity device.
```

Hybrid disk groups contain one flash-based disk drive and one or more magnetic disks. Virtual SAN 6.x can mark other flash-based disks as capacity disks to enable higher performance than a hybrid group.

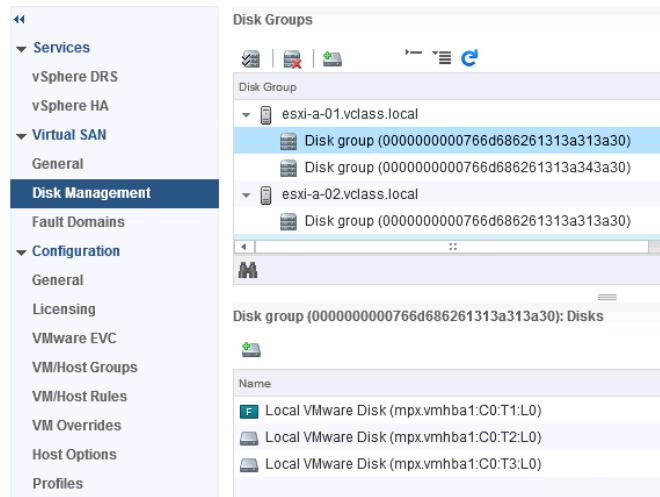
You can use either vSphere Web Client or the `vdq` command to verify whether an SSD is marked as a cache device or capacity device.

# Viewing Disk Groups

Slide 4-29

In VMware vSphere® Web Client, you can view which disks are assigned to which disk group.

Cluster's Manage > Settings tab



From the disk group view, an administrator can see which disks are members of which disk groups. This view helps in identifying disks that are malfunctioning and causing disk group problems.

## Labs

Slide 4-30

- Lab 4: Configuring a Virtual SAN Cluster
- Lab 5: Configuring Hybrid Disk Groups
- Lab 6: Configuring All-Flash Disk Groups
- Lab 7: Virtual SAN Storage Commands

VMware Confidential  
Internal Use Only

## Lab 4: Configuring a Virtual SAN Cluster

Slide 4-31

### Configure a Virtual SAN cluster

1. (Optional) Prepare the Environment
2. Create a Virtual SAN Cluster to Use Manual Disk Claim Mode
3. Move Three ESXi Hosts into the Virtual SAN Cluster
4. Assign the Virtual SAN Enterprise License to the Cluster

VMware Confidential  
Internal Use Only

## Lab 5: Configuring Hybrid Disk Groups

Slide 4-32

Configure hybrid disk groups for a Virtual SAN cluster

1. (Optional) Prepare the Environment
2. Mark the 6.00 GB Disks of the Hosts as Magnetic Disks
3. Create Disk Groups for All Hosts in the Cluster
4. Identify and Rename a Virtual SAN Datastore

VMware Confidential  
Internal Use Only

## Lab 6: Configuring All-Flash Disk Groups

Slide 4-33

Create an all-flash disk group configuration

1. (Optional) Prepare the Environment
2. Verify that the 4.50 GB and 6.50 GB Disks of the Hosts are Flash Disks
3. Create an All-Flash Disk Group for Each Host in the Cluster
4. Delete the Hybrid Disk Groups
5. Mark the 6.00 GB Disks of the Hosts as Flash Disks
6. Create a Second All-Flash Disk Group for Each Host in the Cluster

VMware Confidential  
Internal Use Only

## Lab 7: Virtual SAN Storage Commands

Slide 4-34

Use ESXCLI to examine the Virtual SAN disks

1. (Optional) Prepare the Environment
2. Examine the Virtual SAN Disks

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 4-35

You should be able to meet the following objectives:

- Configure a Virtual SAN cluster
- Test and validate the Virtual SAN configuration and functionality

VMware Confidential  
Internal Use Only

## Key Points

Slide 4-36

- Correctly configured host networking is critical to a functional Virtual SAN deployment.
- The different disk claim modes can result in completely different disk group configurations.

Questions?

VMware Confidential  
Internal Use Only

## MODULE 5

# Virtual SAN Policies and Virtual Machines

Slide 5-1

Module 5

VMware Confidential  
Internal Use Only

# You Are Here

Slide 5-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
- 5. Virtual SAN Policies and Virtual Machines**
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 5-3

Virtual SAN uses virtual machine storage policies to ensure that the performance and availability requirements of virtual machines are met.

Virtual SAN supports vsanSparse snapshots. Snapshots are an important mechanism for providing restore points for virtual machines.

VMware Confidential  
Internal Use Only

## Module Lessons

Slide 5-4

Lesson 1: Storage Policy-Based Management

Lesson 2: vsanSparse Snapshots

VMware Confidential  
Internal Use Only

# Storage Policy-Based Management

Slide 5-5

## Lesson 1: Storage Policy-Based Management

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 5-6

By the end of this lesson, you should be able to meet the following objectives:

- Explain how storage policies work with Virtual SAN
- Define and create a virtual machine storage policy
- Apply and modify virtual machine storage policies
- Change virtual machine storage policies dynamically
- Identify virtual machine storage policy compliance status

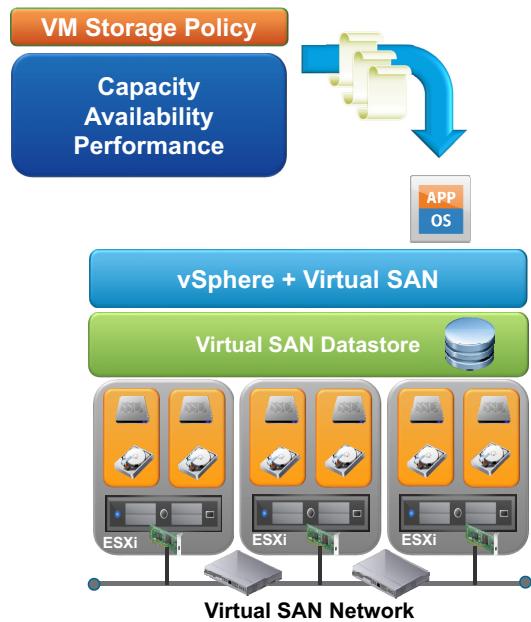
VMware Confidential  
Internal Use Only

# Review: Virtual SAN and Storage Policy-Based Management

Slide 5-7

A storage policy defines a set of capability requirements for virtual machines:

- Policies are based on Virtual SAN capabilities
- Storage policies can be changed at any time
- Policies are monitored for compliance



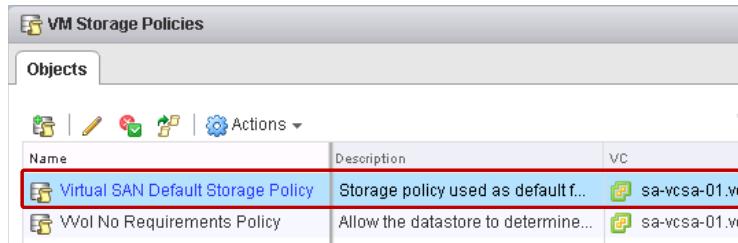
Virtual SAN monitors and reports on the policy compliance during the lifecycle of the virtual machine. If a policy becomes noncompliant, Virtual SAN takes remedial actions. Virtual SAN reconfigures the data of the affected virtual machines and optimizes the use of resources across the cluster. The reconfiguration processes occur with minimal effect on the regular workload.

## Storage Policies

Slide 5-8

Multiple virtual machine storage policies can be created for use by a single Virtual SAN cluster:

- Virtual SAN has a default virtual machine storage policy.
- Custom storage policies can be created using the Virtual SAN capabilities.
- The default Virtual SAN storage policy is used, unless a specific storage policy is selected.
- You use vSphere Web Client to view, create, and modify policies.



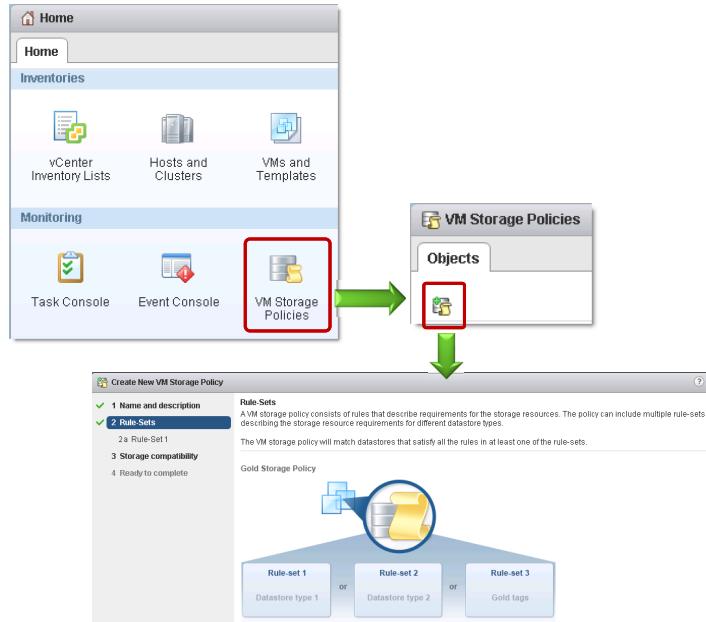
Storage policies define virtual machine storage requirements, such as performance and availability, in the form of a policy. Virtual SAN ensures that the virtual machines deployed to Virtual SAN datastores are assigned at least one virtual machine storage policy. If a storage policy is not explicitly assigned to the virtual machine that is provisioned, a default storage policy is applied to the virtual machine from the datastore. If a custom policy has not been applied to the Virtual SAN datastore, then Virtual SAN Default Storage Policy is used.

# Rule Sets

Slide 5-9

Storage policies include rule sets that define requirements for virtual machine storage:

- Rule sets include one or more rules.
- Rules define specific capabilities to be provided by storage resources.



Rule sets are the guidelines that define the storage requirements for a storage policy.

# Virtual Machine Storage Policy Capabilities for Virtual SAN

Slide 5-10

Storage policies can be created from one or more Virtual SAN rules.

Storage Capability	Use Case	Values	Default
Number of failures to tolerate	Redundancy	0 to 3	1
Number of disk stripes per object	Performance	1 to 12	1
Failure tolerance method	Performance or capacity	RAID 1 or RAID 5/6	RAID 1
Flash Read Cache reservation (%)	Performance	0 to 100%	0
Force provisioning	Override policy	Yes or no	No
Object space reservation (%)	Capacity planning	0 to 100%	0
IOPS limit for object	Performance	-	No limit
Disable object checksum	Performance	Yes or no	No

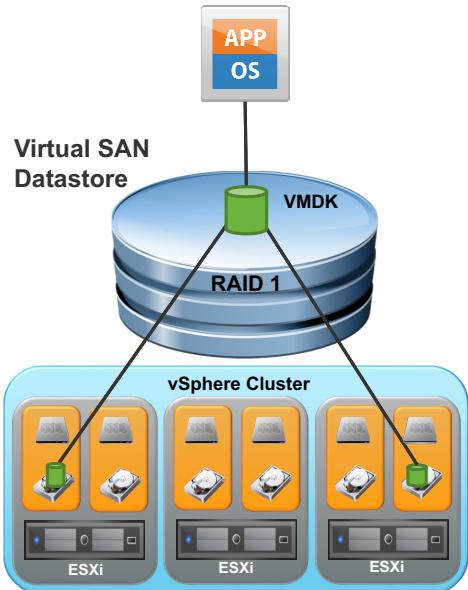
The default Virtual SAN storage policy is created and implemented when Virtual SAN is enabled on a cluster. This policy contains a rule set with all rules defined at their default values. A virtual machine with the default policy applied supports a single failure and is striped across a single disk.

# Number of Failures to Tolerate

Slide 5-11

This value defines the number of host, disk, or network failures that a storage object can tolerate:

- For  $n$  failures that are tolerated,  $n+1$  copies of the object are created.
- For  $n$  failures that are tolerated,  $2n+1$  hosts contributing storage are required.
- Default value: 1.
- Possible values: 0 through 3 (depending on the number of hosts).



This capability sets a requirement on the storage object. The storage object remains available after a specified number of failures corresponding to the number of host or disk failures in the cluster. This property specifies that configurations must contain at least Number of Failures To Tolerate + 1 replicas. The configurations can also contain an additional number of witnesses. Witnesses ensure that the object's data is available even if a specified number of host failures occur. Disk failure on a single host is treated as a failure for this metric. If Number of Failures to Tolerate is configured to 1, the object cannot persist if a disk failure occurs on one host and another host fails at the same time.

The maximum value of Number of Failures to Tolerate depends on the number of hosts or fault domains:

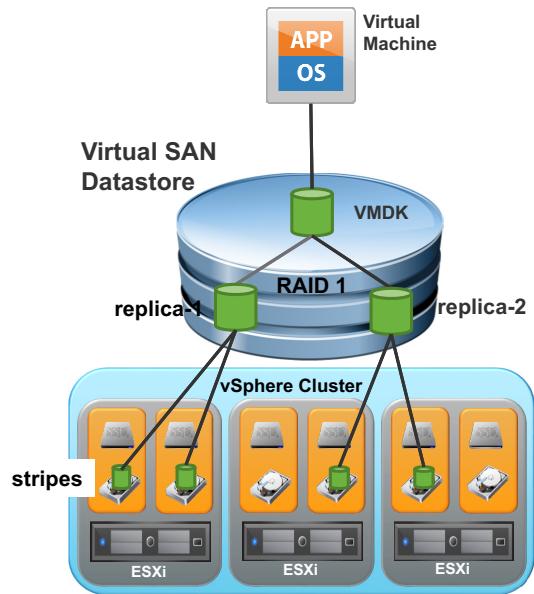
- A value of 1 requires three hosts.
- A value of 2 requires five hosts.
- A value of 3 requires seven hosts.

# Number of Disk Stripes Per Object

Slide 5-12

This value enforces the number of capacity devices across which each replica of a storage object is striped:

- Values higher than 1 can result in better performance, but can also result in higher use of system resources.
- Default value: 1.
- Possible values: 1 through 12.



This value defines the number of capacity devices across which each replica of a storage object is striped.

To understand the effect of stripe width, examine the context of write operations and read operations. All writes go to the solid-state drive (SSD) cache and the value of an increased stripe width might not improve performance. The new stripe width might use a different SSD or the new stripe might be in the same disk group and thus use the same SSD. An increased stripe width might add value if numerous writes overwhelm the SSD garbage collection routine and cause a write-through situation.

From a read perspective, an increased stripe width helps when you are experiencing many cache misses. For example, if a virtual machine consumes 4,000 read operations per second and experiences a hit rate of 90 percent, 400 read operations must be serviced directly from magnetic disks. A single hard disk drive might not be able to service those read operations, so an increase in stripe width might help.

In general, the default stripe width of 1 should meet most, if not all, virtual machine workloads. Stripe width should be changed only for isolated high-performance virtual machines.

# Failure Tolerance Method

Slide 5-13

Virtual SAN 6.2 introduces a new rule called failure tolerance method.

When RAID-5/6 (Erasure Coding) is selected:

- RAID 5 is used when the number of failures to tolerate is 1.
- RAID 6 is used when the number of failures to tolerate is 2.

## Rule-Set 1

Select rules specific for a datastore type. Rules can be based on data services provided by datastore or based on tags. The VM storage policy will match datastores that satisfy all the rules in at least one of the rule-sets.

The screenshot shows the 'Rule-Set 1' configuration for a VSAN datastore. On the left, under 'Rules based on data services', there is a dropdown menu set to 'VSAN'. Below it, the 'Number of failures to tolerate' is set to '1'. The 'Failure tolerance method' dropdown is open, showing three options: 'RAID-5/6 (Erasure Coding) - Capacity', 'RAID-1 (Mirroring) - Performance', and 'RAID-5/6 (Erasure Coding) - Capacity'. The first option is highlighted with a red box. On the right, the 'Storage Consumption Model' section provides details for a virtual disk of size 100 GB: 'Storage space' is 133.33 GB, 'Initially reserved storage space' is 0.00 B, and 'Reserved flash space' is 0.00 B.

While mirroring techniques excel in workloads where performance is the most important factor, they are expensive in terms of capacity required. RAID 5/6 (Erasure Coding) can be configured to help ensure the same levels of availability, while consuming less capacity than RAID 1 (Mirroring).

Use of erasure coding reduces capacity consumption by as much as 50% versus mirroring at the same failure tolerance level. This method of failure tolerance requires additional write overhead in comparison to mirroring as a result of data placement and parity. This additional overhead is common among any storage platform today. Because erasure coding is only supported in all-flash Virtual SAN configurations, effects to latency and IOPS are negligible in most use cases due to the inherent performance of flash devices.

## Mirroring Versus Erasure Coding

Slide 5-14

Erasure coding can provide significant capacity savings over mirroring, but erasure coding incurs additional overhead.

Failures to Tolerate (FTT)	RAID-1 (Mirroring)		RAID-5/6 (Erasure Coding)		Erasure Coding Space Savings versus Mirroring
	Minimum Hosts Required	Total Capacity Requirement	Minimum Hosts Required	Total Capacity Requirement	
0	3	1x	n/a	n/a	n/a
1	3	2x	4	1.33x	33% less
2	5	3x	6	1.5x	50% less
3	7	4x	n/a	n/a	n/a

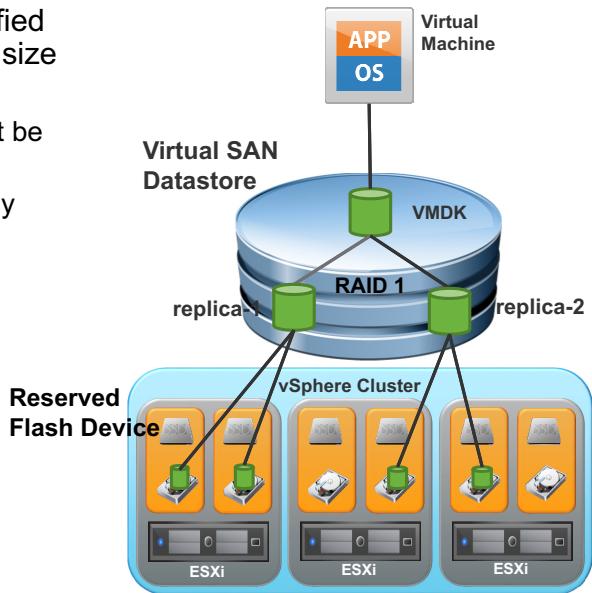
RAID 1 (Mirroring) in Virtual SAN employs a  $2n+1$  host or fault domain algorithm, where  $n$  is the number of failures to tolerate. RAID 5/6 (Erasure Coding) in Virtual SAN employs a 3+1 or 4+2 host or fault domain requirement, depending on 1 or 2 failures to tolerate respectively. RAID 5/6 (Erasure Coding) does not support 3 failures to tolerate. The table illustrates the host and capacity requirements of RAID 1 (Mirroring) versus RAID 5/6 (Erasure Coding).

# Flash Read Cache Reservation

Slide 5-15

Reserved read cache is specified as a percentage of the logical size of the object:

- Reserved flash capacity cannot be used by other objects.
- Unreserved flash is shared fairly among all objects.
- Default value: 0 percent.
- Possible values: 0 through 100 percent.



This value specifies the logical size of the storage object in percentage up to the ten-thousandth place (four decimal places). This specific unit size is needed so that administrators can express appropriate sizes as the capacity of the SSD increases. For example, in a 1 TB disk, if an administrator is limited to 1 percent increments, these increments are equivalent to cache reservations in increments of 10 GB. This value is too large for a single virtual machine in most cases. Ideally, the read cache should match the working set of the virtual disk in order to maximize the read cache hit rate.

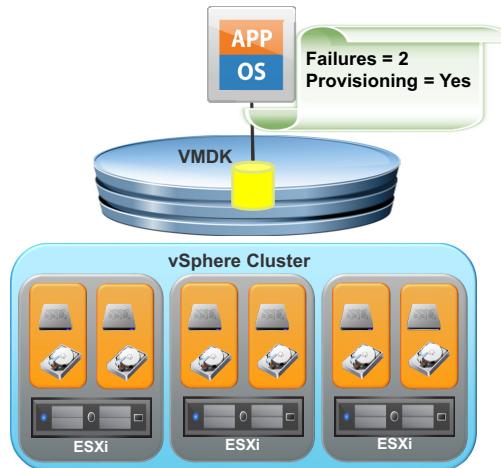
You do not have to set a reservation to get cache. You need to set a reservation only to reserve cache space for a specific virtual machine. The reservation should be set to 0 unless you are trying to solve specific use cases regarding read-intensive virtual machines.

# Force Provisioning

Slide 5-16

Force provisioning allows a virtual machine to be created with a particular storage policy, despite not having enough disks or hosts in the cluster:

- Virtual SAN makes the object compliant when additional resources are added.
- The default is No.



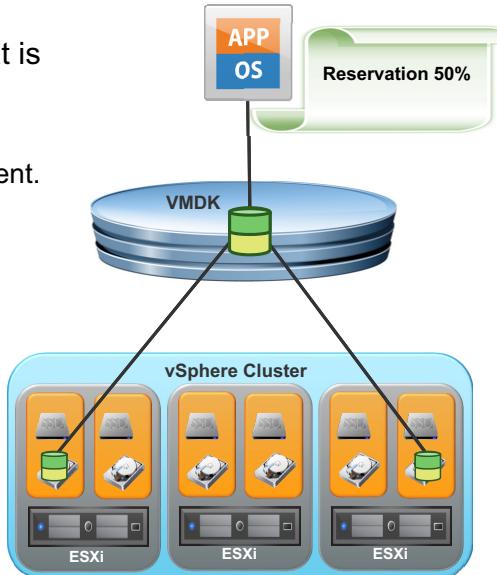
The diagram shows a virtual machine that is provisioned by using a policy in which the number of failures to tolerate is set to 2. Using the  $2n+1$  equation for the number of failures, the policy requires at least five hosts in the cluster. Using Force Provisioning, the virtual machine is deployed with a failures to tolerate of 0 and a stripe width of 1. When additional hosts are available, Virtual SAN makes the virtual machine compliant with its policy.

# Object Space Reservation

Slide 5-17

This value is the percentage of the logical size of the VMDK object that is reserved when provisioned:

- Default value: 0 percent.
- Possible values: 0 through 100 percent.



This capability defines the percentage of the logical size of the storage object that is reserved during initialization. Reserved storage is thick provisioned (lazy zero) and the remainder is thin provisioned. Lazy zero provisioning is used in calculations for total capacity but does not consume the space. The value is the minimum amount of capacity to be reserved. The virtual machine is thin provisioned but the space is reserved for that virtual machine.

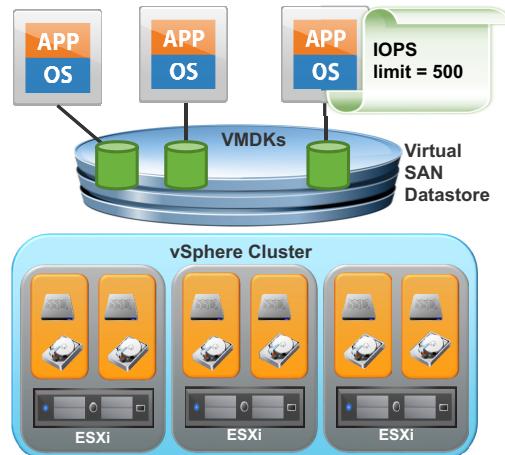
# IOPS Limit for Object

Slide 5-18

Virtual SAN 6.2 introduces a quality-of-service feature, which limits the number of IOPS an object can consume.

Use cases include the following:

- Service providers can create differentiated service offerings by using the same pool of storage.
- Administrators can mix diverse workloads while keeping workloads from impacting each other.



In underutilized configurations, limits might not be necessary because objects might have sufficient resources to effectively meet the needs of their workload. Having more than enough resources comes at a cost. Efficiently sized configurations are typically a good mix of cost and available resources. The metrics of appropriate resources for workloads can change over time, especially as utilization grows, or as workloads are added over the lifecycle of a platform.

Limiting the IOPS of one or more virtual machines might be advantageous. In environments with mix of both low and high utilization, a virtual machine with low utilization during normal operations can change its pattern and consume massive amounts of resources, preventing other virtual machines from operating properly. For example, consider those virtual machines generating reports residing on the same 4-node hybrid Virtual SAN cluster as other tier-1 applications that have stringent Service Level Agreement requirements.

This quality-of-service feature is normalized to a 32 KB block size and treats reads the same as writes. An example with an IOPS limit of 500 (regardless of block size up to 32 KB) results in 500 IOPS, while a block size of 64 KB results in 250 IOPS. You must consider the workload profile when configuring IOPS limits.

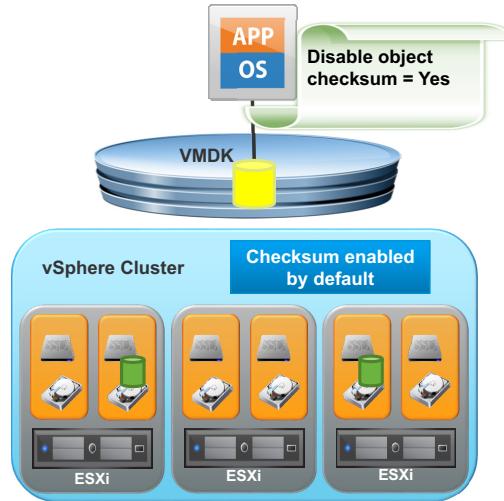
# Disable Object Checksum

Slide 5-19

Virtual SAN 6.2 introduces software checksums, used to detect data corruption that could be caused by hardware or software components:

- Automatically detects and resolves silent disk errors
- Fetches data from another copy if checksum verification failures occur
- Performs disk scrubbing in the background
- Enabled on the cluster by default
- Can be disabled per object with virtual machine storage policies

This feature should be disabled if this functionality is already included in the application.



Software checksum can detect corruptions that could be caused by hardware or software components during the read or write operations. For drives, the corruption can be of the following types:

- Latent sector errors are typically the result of a physical disk drive malfunction.
- Silent corruption can happen without warning and is typically called silent data corruption. Undetected or completely silent errors can lead to lost or inaccurate data and significant downtime. No effective means of detection exists without end-to-end integrity checking.

During the read/write operations, Virtual SAN checks for the validity of the data based on the checksum. If the data is not valid, then Virtual SAN takes the necessary steps to either correct the data or report it to the user to take action.

Virtual SAN has a disk scrubbing mechanism, which periodically checks the data on disk for errors. By default, the data is checked once a year. But this period can be modified with the advanced ESXi host setting `VSAN.ObjectScrubsPerYear`.

# Viewing Object Placement

Slide 5-20

Each virtual machine can be examined to see where its VM home and virtual disk objects and their components are physically located.

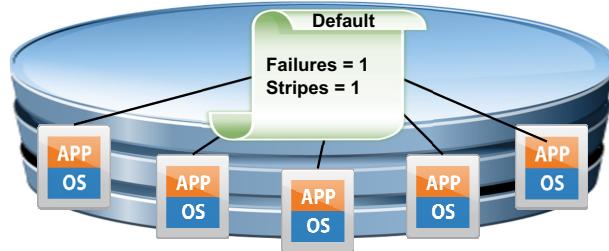
Type	Component State	Host	Capacity Disk Name
Witness	Active	sb-esxi-03.vcl...	Local VMware Disk (mpx.vmhba1:C0:T5:L0)
RAID 1			
Component	Active	sb-esxi-04.vcl...	Local VMware Disk (mpx.vmhba1:C0:T2:L0)
Component	Active	sa-esxi-01.vcl...	Local VMware Disk (mpx.vmhba1:C0:T2:L0)

By examining the properties of a virtual machine in the storage policy, the administrator can view the location of each object. The Capacity Disk Name column provides the name of the physical disk to which a particular object is deployed.

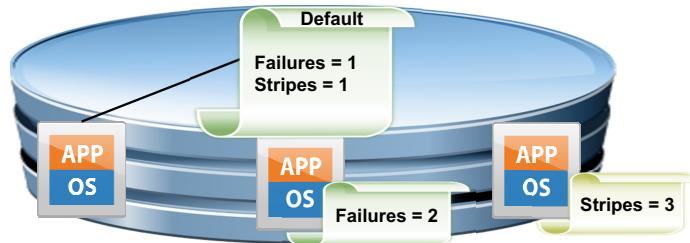
# Assigning Storage Policies

Slide 5-21

Virtual SAN datastores have a default storage policy.



Virtual machines, and even virtual disks, can have individually assigned storage policies that override the default policy.



When the Virtual SAN datastore is created, a default policy is applied. This policy tolerates a single failure and supports a single disk stripe. A datastore's default storage policy should have a rule set that applies to the widest range of virtual machines that are to be hosted on the datastore. Individual virtual machines should have a custom storage policy applied that overrides the default policy for the datastore as needed. When most virtual machines use the default datastore policy, the overhead of policy administration and compliance is minimized.

## Storage Policies and Multiple VMDKs

Slide 5-22

Each virtual disk for a specific virtual machine can have a different storage policy applied.

Name	Disk Size	VM Storage Policy	Datastore	Datastore Type
VM home		VSAN Storage Policy	vSAN Datastore-B	vsan
Hard disk 1	2.15 GB	VSAN Storage Policy	vSAN Datastore-B	vsan
Hard disk 2	60.00 GB	Database Server VSAN Storage Policy	vSAN Datastore-B	vsan

Some virtual machines with multiple virtual disks require that the data on one virtual disk be treated in a different manner than other virtual disks. Thus, the data of one virtual disk might require more safeguards than the data on the other virtual disks. For example, in a database server, the operating system of the server can be replaced with relative ease although the database is undergoing constant updates and changes. A specific storage policy that ensures that the database virtual disk is striped across more disks than the operating system virtual disk also ensures that the datastore has the required performance.

## Labs

Slide 5-23

Lab 8: Deploying Virtual Machines to Virtual SAN

Lab 9: Creating Storage Policies

VMware Confidential  
Internal Use Only

## Lab 8: Deploying Virtual Machines to Virtual SAN

Slide 5-24

Deploy a virtual machine and specify a storage policy

1. (Optional) Prepare the Environment
2. Import a Virtual Machine to the First ESXi Host
3. Migrate a Virtual Machine to a Local Datastore
4. Migrate a Virtual Machine to the Virtual SAN Datastore

VMware Confidential  
Internal Use Only

## Lab 9: Creating Storage Policies

Slide 5-25

Create and review virtual machine storage policies

1. (Optional) Prepare the Environment
2. Examine the Default Storage Policy
3. Create a Custom Policy with No Failure Tolerance
4. Assign the Custom Policy to a Virtual Machine
5. Bring the Virtual Machine into Compliance
6. Compare Virtual Machines with Different Storage Policies
7. Edit Custom Policy to Require Two Disk Stripes Per Object
8. Edit Custom Policy to Increase Failure Tolerance to One
9. Create an Invalid Storage Policy

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 5-26

You should be able to meet the following objectives:

- Explain how storage policies work with Virtual SAN
- Define and create a virtual machine storage policy
- Apply and modify virtual machine storage policies
- Change virtual machine storage policies dynamically
- Identify virtual machine storage policy compliance status

VMware Confidential  
Internal Use Only

## vsanSparse Snapshots

Slide 5-27

### Lesson 2: vsanSparse Snapshots

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 5-28

By the end of this lesson, you should be able to meet the following objectives:

- Describe how vsanSparse snapshots work
- Explain the considerations for vsanSparse snapshots
- Discuss the vsanSparse snapshot format

VMware Confidential  
Internal Use Only

## About vsanSparse

Slide 5-29

vsanSparse is a snapshot format, introduced in Virtual SAN 6.0, and has the following features:

- Improves performance
- Allows for longer snapshot retention
- Supports the full limit of 32 snapshots
- Uses an always-sparse format

The vsanSparse snapshot has the following requirements:

- VSAN FS datastores must be used.
- A virtual machine must not have any existing vmfsSparse snapshots.

vsanSparse is a performance-based snapshot format introduced in vSphere 6.0. The snapshot format uses the VSAN FS file system and the in-memory caching mechanism for updates introduced in vSphere 6.0. The system has a 512 byte block size instead of the 1 MB block size in VMFS-L in vSphere 5.5.

A vSphere administrator manages vsanSparse snapshots in the same way that previous virtual machine snapshots were managed. No selection process exists to choose the vsanSparse format. The vsanSparse format snapshots are used if the following conditions are met:

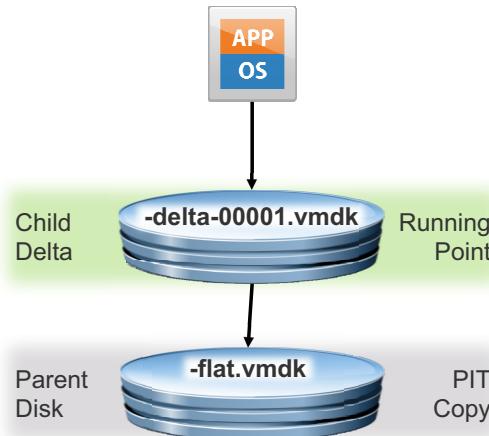
- The underlying storage is Virtual SAN.
- The on-disk format is v2 and higher.
- No older vmfsSparse/redo log format snapshots exist on the virtual machine.

## Review: VMware Snapshots

Slide 5-30

When a snapshot is created, a child delta disk is created:

- All future writes are written to the child delta disk.
- The parent disk becomes a static point-in-time copy.
- The current running state is parent disk combined with all child copies.

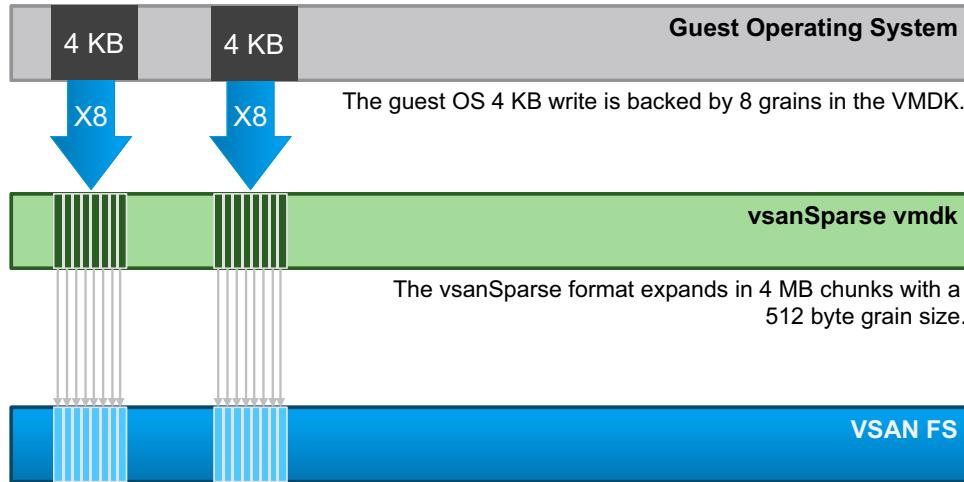


When a vsanSparse snapshot of a base disk is taken, a child delta disk is created. The base disk parent is considered a point-in-time copy. The running point of the virtual machine becomes the delta. New writes by the virtual machine go to the delta, but the base disk and the other snapshots in the chain satisfy reads. To get the current state of the disk, you can take the parent disk and redo all writes from children chain. vsanSparse format is similar to the earlier vmfsSparse format.

## vsanSparse: Always Sparse

Slide 5-31

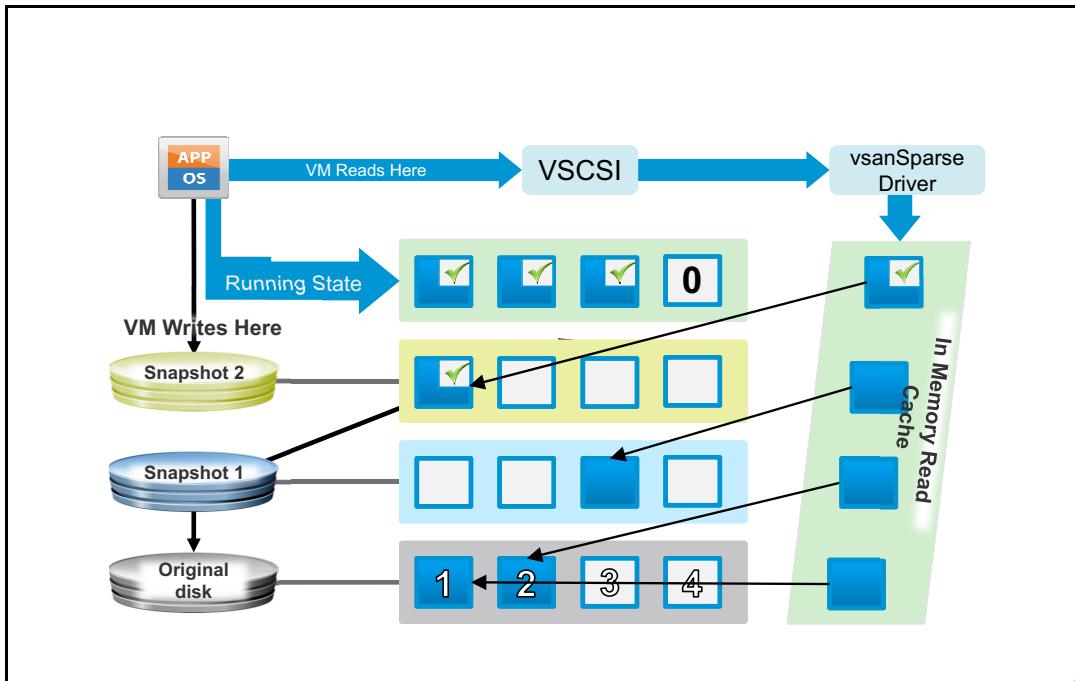
The vsanSparse snapshot format uses a 512-byte allocation unit size called a grain.



The new vsanSparse snapshot format and the underlying VSAN FS file system use a 512-byte allocation unit size. This allocation unit size allows snapshots taken with the vsanSparse format to remain sparse. Writes from the guest operating system are always placed in the same logical offsets in the object.

# vsanSparse Memory Cache

Slide 5-32



When a guest operating system sends a write to a virtual disk with snapshots, the vsanSparse driver receives the write. Writes always go to the top-most object in the snapshot chain. When the write is acknowledged, the vsanSparse driver updates its in-memory metadata cache, and confirms the write back to the guest operating system. On subsequent reads, the vsanSparse driver references the metadata cache and finds the data block on a cache hit.

The read cache is in-memory only and is never committed to persistent storage. When the virtual machine powers off, the cache is erased. The next time the virtual disk is opened, the cache is empty and fills as the virtual machine generates I/O. A read request fails to retrieve a hit from the empty cache and generates a cache miss. When a cache miss is generated, the data is retrieved and its details are cached for future requests. Cache misses increase the I/O latency.

The diagram illustrates the following write and read scenarios:

- Original disk: Writes to grain 1 and 2 before the Snapshot 1 delta object is created.
- Snapshot 1 delta object: Writes to grain 3 before the Snapshot 2 delta object is created.
- Snapshot 2 delta object: Writes to grain 1.

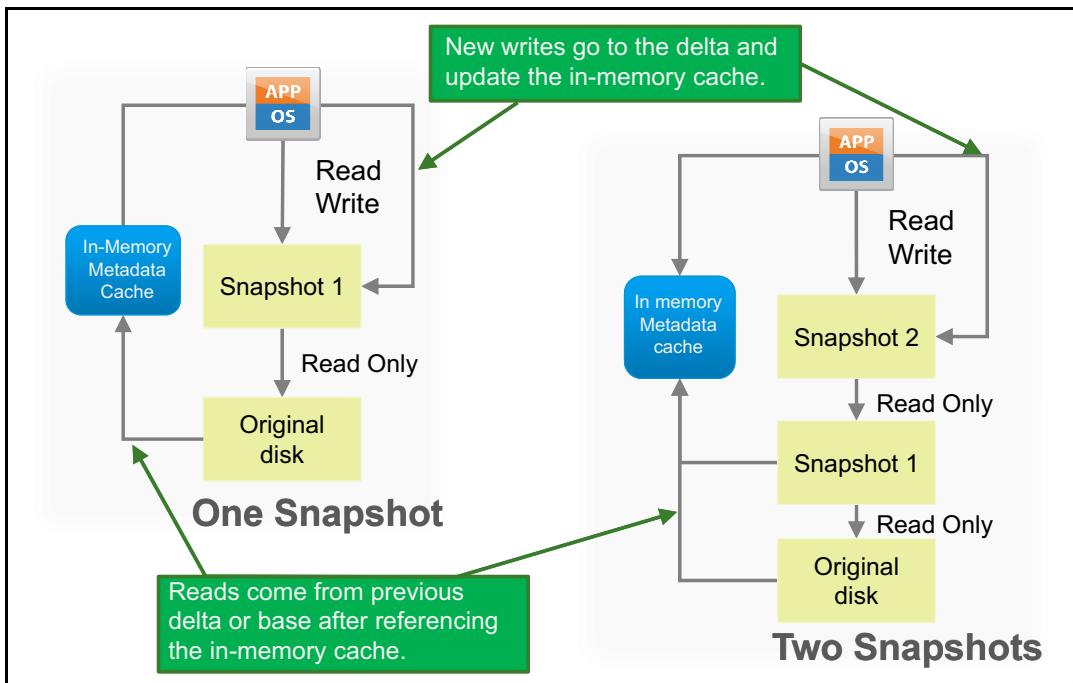
A read by the VM returns the following:

- Grain 1: Retrieved from the Snapshot 2 delta object.
- Grain 3: Retrieved from the Snapshot 1 delta object.
- Grain 2: Retrieved from the original disk.
- Grain 4: Retrieved from the original disk. A 0 value is returned because grain 4 was never written to.

VMware Confidential  
Internal Use Only

# Snapshot Flow

Slide 5-33

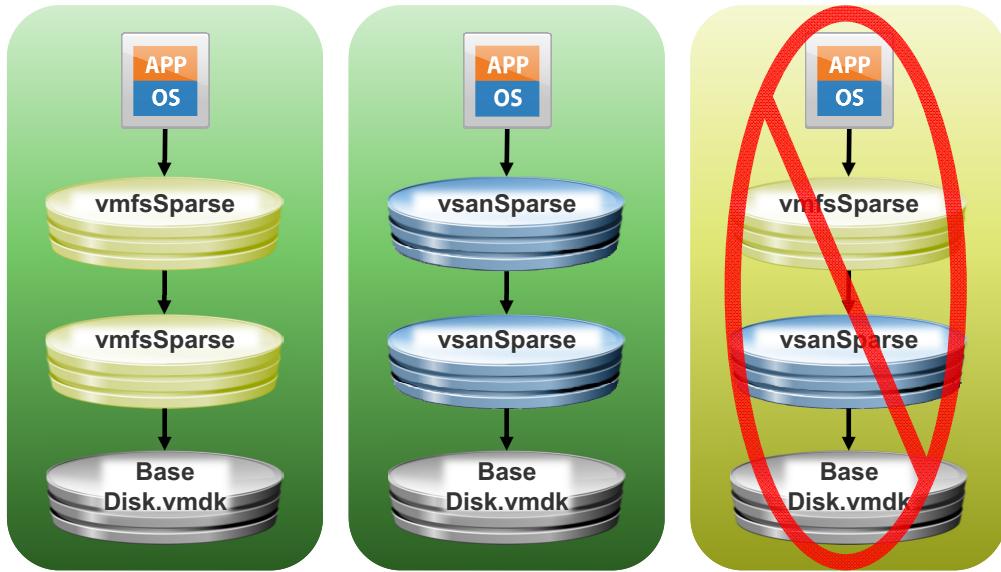


Reads are serviced from one or more of the vsanSparse deltas in the snapshot tree. The vsanSparse driver checks the in-memory metadata cache to determine which delta or deltas to read based on the parts of the data that were written in each particular snapshot level. To satisfy a read I/O request, the snapshot logic does not need to traverse through every delta of the snapshot tree. The snapshot logic can go directly to the necessary vsanSparse delta and retrieve the requested data. Reads are sent to all deltas that have the necessary data in parallel. However, on a cache miss, the vsanSparse driver must traverse each layer to fetch the latest data, similar to read requests, if the requests are sent to all layers and are parallel.

# Uniform Snapshot Format

Slide 5-34

Snapshots for a virtual machine are of a single type.



The vmfsSparse and vsanSparse snapshot formats cannot be mixed in a snapshot chain. All snapshots, except the base disk in a vsanSparse chain, must be vsanSparse. Administrators cannot create linked clones of a virtual machine with vsanSparse snapshots on datastores other than a Virtual SAN datastore with the VSAN FS on-disk format. If a virtual machine has existing vmfsSparse/redo log-based snapshots, the virtual machine continues to get vmfsSparse/redo log-based snapshots until the user consolidates and deletes all the current snapshots.

## Read Cache Considerations

Slide 5-35

The vsanSparse in-memory cache initially has unknown ranges:

- The cache is cold.

When a read request from an unknown range occurs, a cache miss is generated:

- The range is retrieved and cached for future requests.
- A cache miss increases the I/O latency.

Cache is in memory only, and is never committed to persistent storage:

- If a host failure occurs, the cache is erased.
- The next time the virtual disk is opened, the cache is cold (empty).
- The cache is updated as the virtual machine generates I/O, with the first I/O paying a latency penalty.

Intensive use of snapshots on Virtual SAN datastores in a hybrid configuration can lead to higher than normal consumption of read cache resources. Depending on the scenario, this consumption might result in the following:

- Other workloads might become temporarily cache starved.
- Snapshots might perform poorly.

# Snapshot Considerations

Slide 5-36

## Snapshot retention:

- Retain snapshots for as long as you need them.
- Check the read cache usage periodically.

## Snapshot chains:

- Maximum 32 snapshots per virtual machine using vsanSparse.
- VMware recommends not to use more than 16.
- Use the `vsan.whatif_host_failures 0` command to monitor read cache.

Resource	Usage right now	Usage after failure/re-protection
HDD capacity	67% used (1074.02 GB free)	95% used (118.93 GB free)
Components	1% used (35632 available)	1% used (26632 available)
RC reservations	0% used (521.57 GB free)	0% used (391.17 GB free)

You do not need to limit vsanSparse snapshot usage to short periods of time, for example, 24 through 72 hours. VMware supports the full maximum chain length of 32 snapshots for vsanSparse snapshots. VMware recommends remaining at 16 snapshots per chain or below, to avoid potential performance problems.

VMware recommends that you regularly check the read cache usage and the Virtual SAN datastore capacity when using snapshots extensively because the number of snapshots might contribute to performance issues. You can consolidate the number of snapshots in the snapshot chain to reduce demands on the read cache.

## Verifying Snapshot Format

Slide 5-37

Examine the snapshot descriptor file to determine the snapshot format.

```
> cat ch-vsang-desktop-000001.vmdk
# Disk DescriptorFile
version=4
encoding="UTF-8"
CID=0c4e9289
parentCID=34cbbf0c
isNativeSnapshot="no"
createType="vsanSparse"
parentFileNameHint="ch-vsang-desktop.vmdk"
# Extent description
RW 209715200 VSANSPARSE "vsan://7c110055-06d0-fd51-28dc-001517a69c72"
```

From VMware vSphere® ESXi™ Shell, use the `cat` command to examine the descriptor file to see the object type.

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 5-38

You should be able to meet the following objectives:

- Describe how vsanSparse snapshots work
- Explain the considerations for vsanSparse snapshots
- Discuss the vsanSparse snapshot format

VMware Confidential  
Internal Use Only

## Key Points

Slide 5-39

- Policy-based storage allows you to respond to changes quickly.
- You can create and update virtual machine storage requirements without maintenance windows.
- You can ensure that performance and availability requirements for virtual machines are met.
- vsanSparse snapshots enable higher performance.

Questions?

VMware Confidential  
Internal Use Only

## MODULE 6

# Managing and Operating Virtual SAN

Slide 6-1

Module 6

VMware Confidential  
Internal Use Only

# You Are Here

Slide 6-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
- 6. Managing and Operating Virtual SAN**
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 6-3

Virtual SAN aims to avoid resource contention by placing virtual machine objects and components across the disks, disk groups, and hosts in the cluster. However, on-going monitoring and analysis are needed to ensure desirable performance and availability levels.

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 6-4

By the end of this module, you should be able to meet the following objectives:

- Manage hardware storage devices
- Manage hardware device failures
- Identify vCenter Server alarms for Virtual SAN events
- Describe and configure fault domains
- Upgrade to Virtual SAN 6.2

VMware Confidential  
Internal Use Only

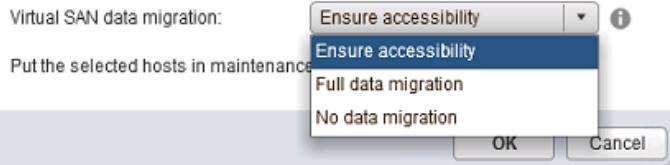
## Maintenance Mode Options

Slide 6-5

When placing a Virtual SAN host in maintenance mode, select the data migration option that is most appropriate:

- **Ensure accessibility:** Move objects to active Virtual SAN resources as needed, to ensure access.
- **Full data migration:** Move all objects to active Virtual SAN resources regardless of whether the move is needed to ensure access.
- **No data migration:** Move no objects. Some active objects might become unavailable.

Virtual SAN data might reside on the hosts in a Virtual SAN cluster. Select an option to set the migration mechanism for the Virtual SAN data that will be enforced before the hosts enter maintenance mode.

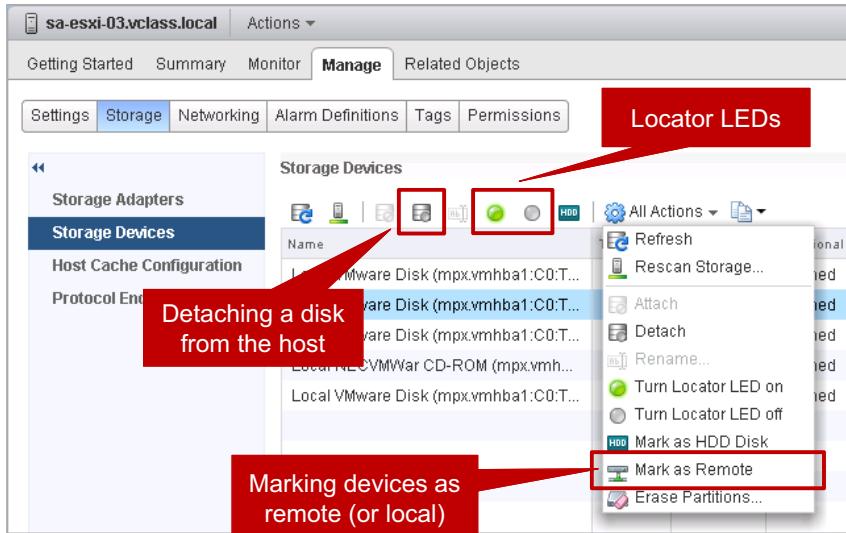


Before you shut down, reboot, or disconnect a host that is a member of a Virtual SAN cluster, you must place the host in maintenance mode. When you place a host in maintenance mode, you can select a specific evacuation mechanism. When a member node of a Virtual SAN cluster enters maintenance mode, the Virtual SAN cluster capacity is reduced because the member node no longer contributes to the cluster storage.

# Managing Hardware Storage Devices

Slide 6-6

Several additional functions are available to manage hardware devices.



Using locator LEDs, you can identify the location of storage devices during failure. When locator LEDs are enabled, an LED light glows on a failed device so that you can easily identify the device. This feature is useful when you are working with multiple hot plug and host swap scenarios. Controllers with RAID 0 mode require additional steps to enable the controllers to recognize locator LEDs. When using locator LEDs, you must use storage I/O controllers with pass-through mode.

You can detach a device from the host, which causes the host to ignore this storage device. The device must not contain a datastore, must not be used as a raw device mapping (RDM) drive by a virtual machine, or contain a diagnostic or scratch partition.

Virtual SAN might occasionally recognize a remote device as a local device or a local device as a remote device. This incorrect recognition might happen with external and shared Serial Attached SCSI (SAS) devices. An administrator can resolve these mislabeled devices using this feature.

# Resynchronizing Components

Slide 6-7

Virtual SAN might need to move data around in the background because of the following reasons:

- Policy change
- Host failure
- Long term or permanent component loss
- User triggered reconfiguration
- Maintenance mode

The screenshot shows the 'Monitor' tab selected in the vSphere Web Client. Under the 'Resyncing Components' section, it displays the following information:

Health	Resyncing components	4
Capacity	Bytes left to resync	28.84 GB
Proactive Tests	ETA to compliance	12 minutes

Below this, a detailed list of components is shown:

Name	VM Storage Policy	Host	Bytes Left to Resync	ETA
Linux-A-01	--	--	28.84 GB	12 minutes
Hard disk 1	Custom-FTT1 Strip...	--	28.84 GB	12 minutes
Component	--	sa-esxi-04.vclass.l...	7.21 GB	12 minutes
Component	--	sa-esxi-03.vclass.l...	7.21 GB	12 minutes
Component	--	sa-esxi-02.vclass.l...	7.22 GB	12 minutes
Component	--	sa-esxi-01.vclass.l...	7.21 GB	12 minutes

A service might be interrupted, for example, a hardware device, host, or network failure, or a host might be placed into maintenance mode. Virtual SAN initiates resynchronization in the Virtual SAN cluster.

The following events trigger a resynchronization operation in the cluster:

- Editing a virtual machine storage policy.
- Restarting a host.
- Recovering hosts from a long-term failure.
- Evacuating data by using the Full data migration mode before placing a host in maintenance mode.
- Exceeding the utilization of a capacity device.

## Proactive Rebalance

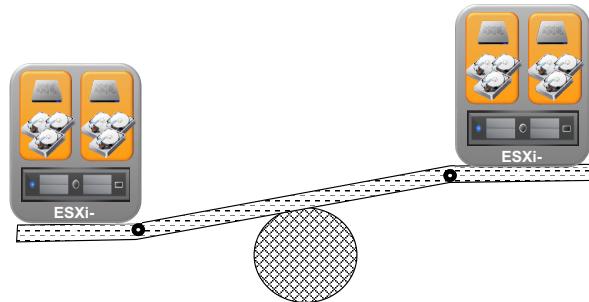
Slide 6-8

Proactive rebalance is a feature that addresses the following use cases:

- Add a new node to an existing Virtual SAN cluster or bring a node out of decommission state.
- Leverage the new nodes even if the fullness of existing disks is below 80 percent.
  - You can modify this threshold with the advanced host setting `VSAN.ClomRebalanceThreshold`.

Proactive rebalance is performed through RVC:

```
vsan.proactive_rebalance --start ~/computers/cluster
```



When any capacity device in a cluster reaches above 80 percent utilization, Virtual SAN performs a rebalance operation until the capacity utilization is below the default threshold level of 80 percent. The rebalance operation evenly redistributes resources across the cluster to maintain consistent cluster performance and availability.

Other operations that can start a rebalance operation in the cluster are the following:

- Hardware failures that are detected in the Virtual SAN cluster
- Virtual SAN hosts that are placed in maintenance mode with either the Ensure accessibility or Full data migration option set to migrate data

VMware recommends that you keep 30 percent of the Virtual SAN datastore free to provide enough space for maintenance, protection, and to minimize automatic rebalancing events in the Virtual SAN cluster.

## Understanding Failure Events (1)

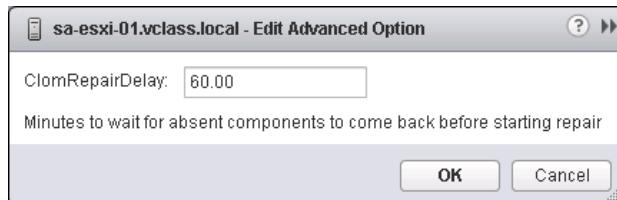
Slide 6-9

Virtual SAN recognizes the following types of hardware device events to define the type of failed scenario:

- Absent
- Degraded

Absent: The data is temporarily unavailable.

- The data component might come back soon.
- Absent events trigger 60-minute recovery operations:
  - Virtual SAN waits 60 minutes before starting the object and component recovery operations.
  - The delay is configurable with a value through the host's advanced settings.



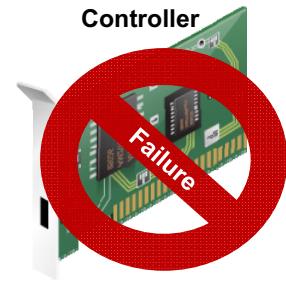
Virtual SAN generates two different types of error messages. The absent error message indicates a problem that might be remedied quickly, such as network connectivity to a host or the unplugging of a disk. By default, the Virtual SAN cluster waits 60 minutes for an absent event to resolve before starting recovery actions. The `ClomRepairDelay` advanced host setting lets you change the delay.

## Understanding Failure Events (2)

Slide 6-10

Degraded: The data copy is permanently lost.

- The data copy is known to be lost and is not expected to come back.
- This event triggers immediate recovery operations.



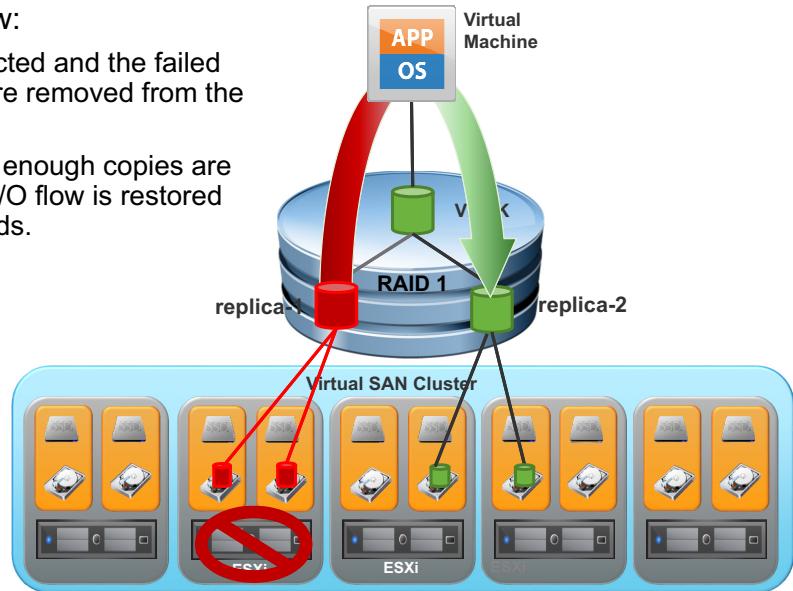
Problems like controller or disk failures start a degraded event. With a degraded event, the Virtual SAN cluster assumes that the data is lost due to the failure and begins recovery tasks immediately.

# Failure Scenario: Restoring I/O Flow

Slide 6-11

## Restore I/O flow:

- Failure is detected and the failed components are removed from the active set.
- Assuming that enough copies are available, the I/O flow is restored within 7 seconds.



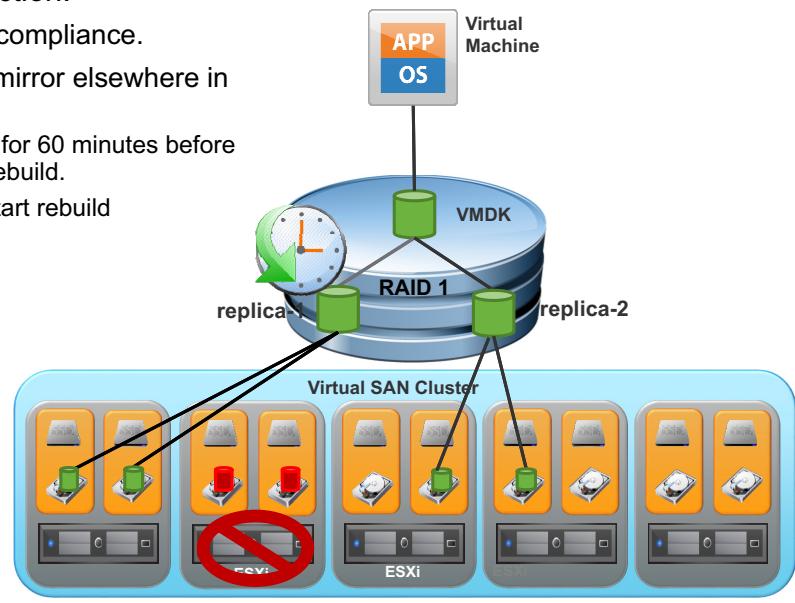
When a failure occurs and recovery operations begin, the system first removes the failed components from the Virtual SAN cluster. Virtual machines immediately change over to using their replicas on other disks to minimize impact.

## Failure Scenario: Rebuilding Components (1)

Slide 6-12

Establish protection:

- Ensure policy compliance.
- Invoke a new mirror elsewhere in the cluster:
  - Absent: Wait for 60 minutes before starting the rebuild.
  - Degraded: Start rebuild immediately.



The system attempts to bring all virtual machines into compliance with their storage policies after the following:

- The system has restored function to virtual machines
- A required wait in the case of an absent event

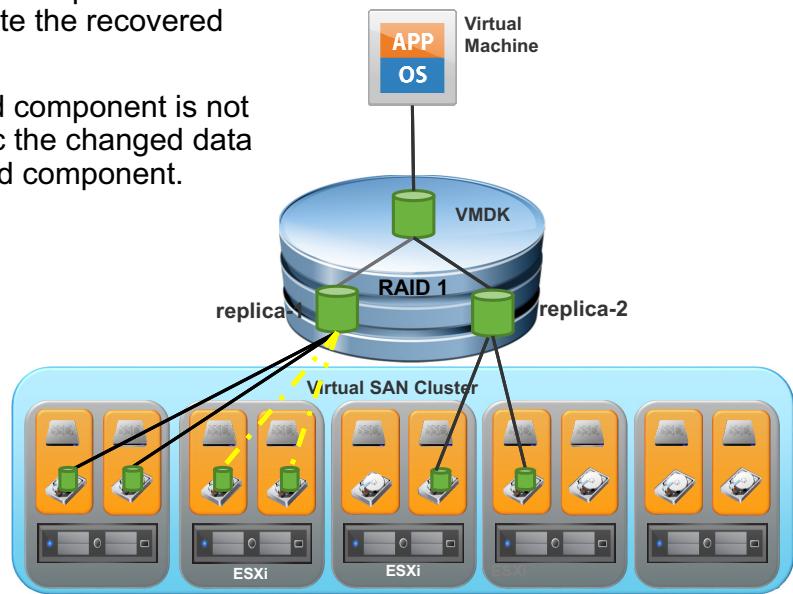
The system creates new replicas and stripes as defined by the individual machine's storage policies.

## Failure Scenario: Rebuilding Components (2)

Slide 6-13

If the recovered component is replicated, delete the recovered component.

If the recovered component is not replicated, sync the changed data to the recovered component.



If a failure is resolved and data is restored after new replicas are made, the system deletes the recovered data and continues to use the replicas that were created after the failure. If the system is restored before creating objects, changes to the current running set are copied to the restored objects and the systems continue running.

# Handling Failure

Slide 6-14

In a traditional SAN environment:

- A failed physical drive must be replaced to achieve full redundancy.
- Hot-spare disks are set aside to replace failed disks immediately.
- Both these scenarios require an immediate 1:1 disk replacement.

With a Virtual SAN cluster:

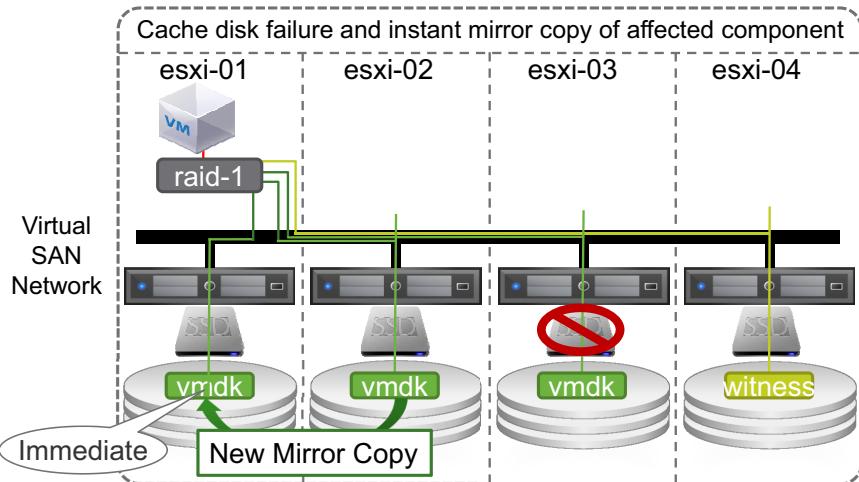
- The entire cluster is a hot-spare and redundancy is built in logically.
- During a failure, components like stripes or mirrors of objects are distributed to other resources.
- Replacement of the physical disk restores only cache or capacity resources.

When a disk fails, many small components fail, such as stripes or mirrors of objects. But new copies of these components are spread around the cluster for balancing. These copies are created as part of the resynchronization and rebalance operations that are initiated by the failures. These operations can cause resource contention depending on the size of the affected data, but Virtual SAN balances the load for fair distribution.

# Cache Device Failure: Instant Mirror Copy

Slide 6-15

Cache device failures result in immediate changes.

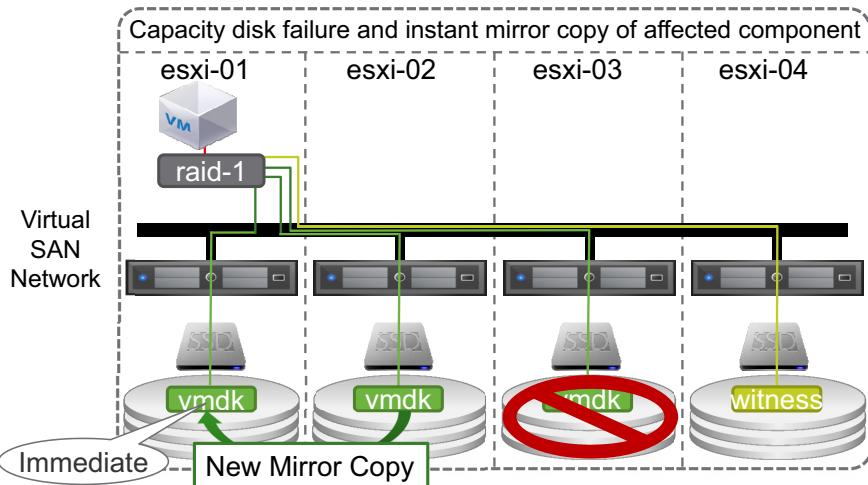


If a cache device fails, data is recreated onto other disk groups. Cache device failures have a greater effect on the cluster's overall storage capacity because the entire disk group is not accessible with a single cache device failure.

# Capacity Device Failure: Instant Mirror Copy

Slide 6-16

Capacity device failures result in immediate changes.

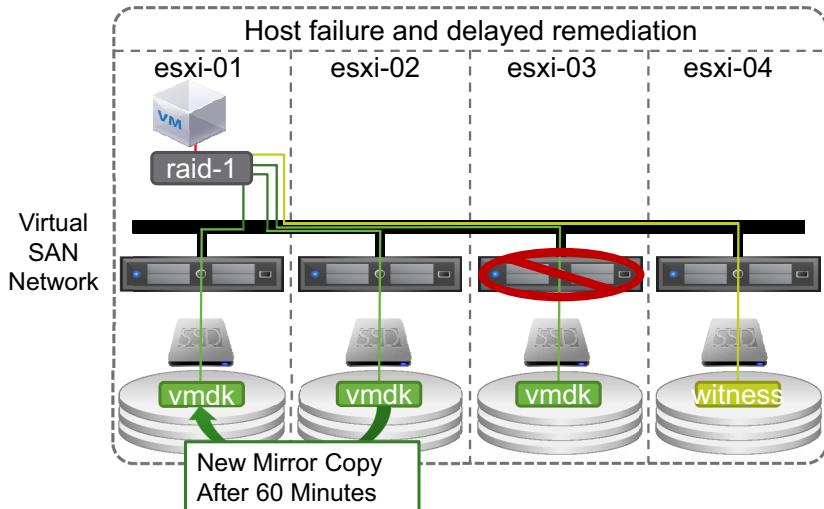


When a capacity disk fails, all the affected components on the failed device are immediately recreated on other disks. Resynchronization can take time depending on the amount of data.

## Host Failure: 60-Minute Delay

Slide 6-17

During an absent event, Virtual SAN waits before copying objects and components.

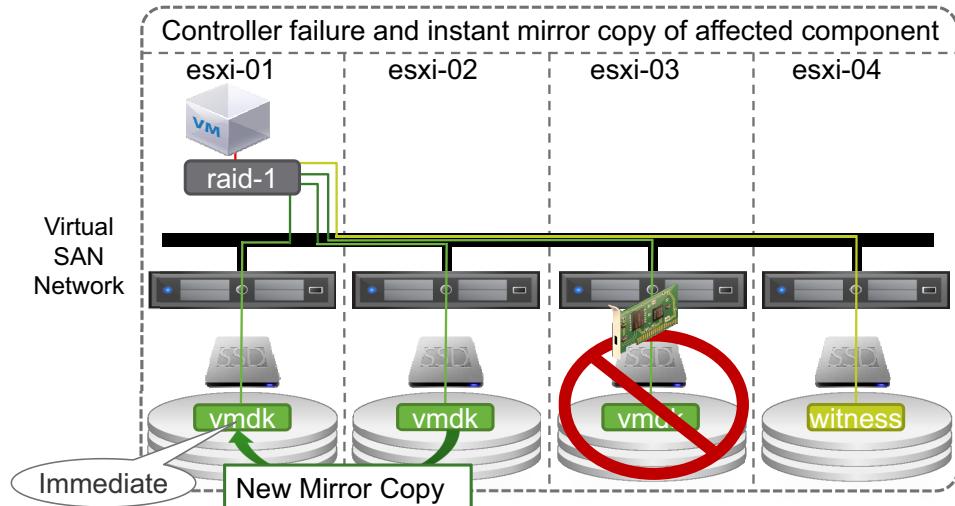


A host failure is an absent event. Virtual SAN waits before copying objects and components to other disks.

# Storage Controller Failures

Slide 6-18

During a degraded event, all affected components on a new cache device are immediately resynchronized.



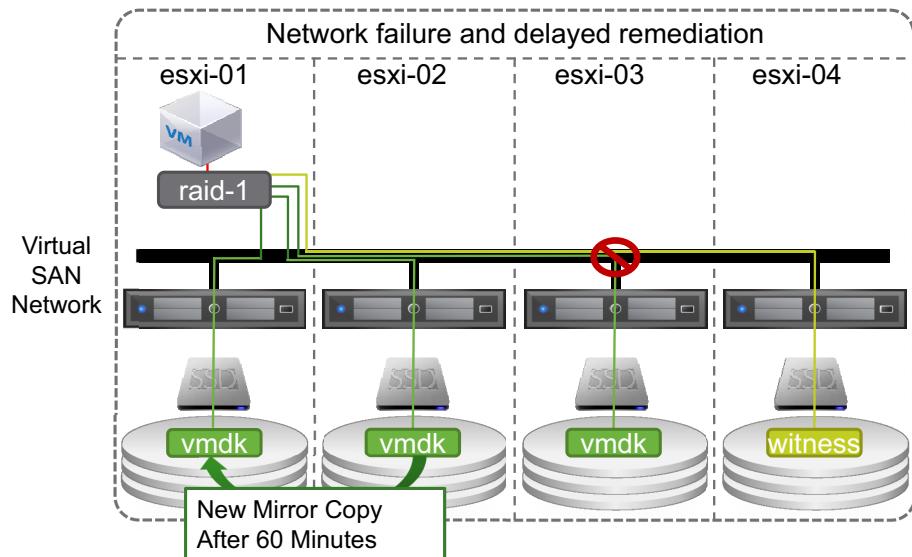
If a controller in a single controller configuration fails with a permanent failure, then every disk group on the host is affected. The behavior of Virtual SAN when a storage I/O controller fails is that all cache devices and all capacity disks appear to fail in all disk groups. Components are marked as degraded and component rebuilding is immediate.

In a single controller configuration, only one disk group might exist with one cache device in the host. You might not be able to determine whether a cache device failure or a storage I/O controller failure might occur. Both failures affect the whole disk group. The VMkernel log files on the host might be able to help you find the root cause.

# Network Failure: 60-Minute Delay

Slide 6-19

During an absent event, Virtual SAN waits before copying objects and components.



Network failures are absent errors and Virtual SAN waits before copying objects and components to other disks.

## Replace and Upgrade Devices

Slide 6-20

To replace and upgrade devices:

1. Place the host in maintenance mode and evacuate the data as needed.
2. Remove the device to be replaced from the disk group to evacuate data.
3. If the host does not support hot swappable disks, shut down the host.
4. Replace the hardware.
5. Power on or bring the host out of maintenance mode.
6. If the new hardware is not detected, rescan the storage.
7. Add the new disk to the original disk group.

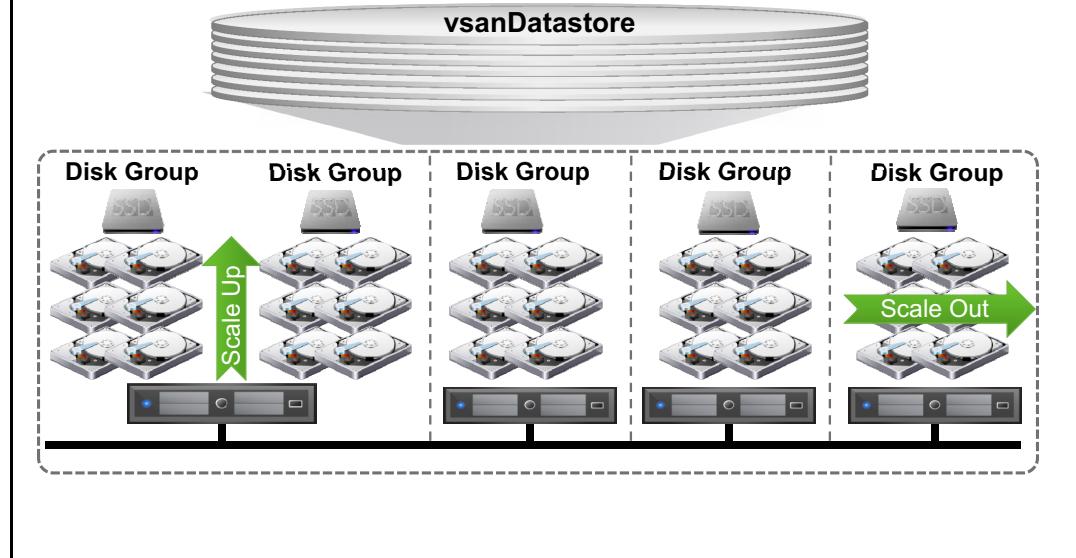
Eventually you must replace hardware components, drivers, firmware, and storage I/O controllers in the Virtual SAN cluster.

Before you physically unplug a cache device from the host, you must manually remove the device from the Virtual SAN. When you replace a cache device, the virtual machines on the disk group become inaccessible. The components on the group are marked as degraded until the cache device is replaced.

# Virtual SAN Scalable Architecture

Slide 6-21

Scale up and scale out architecture provides specific and linear storage, performance, and compute scaling capabilities.



To determine the components to upgrade or add to the Virtual SAN environment, consider the following guidelines:

- To increase capacity, add additional or larger capacity disks to the hosts in the cluster.
- To improve caching, add additional or larger cache disks to the cluster.
- To add both caching and capacity, add additional disk groups to the cluster.

Hosts can be added to the cluster without disrupting ongoing operations. New cluster members add storage and compute capacity.

# vCenter Server Alarms

Slide 6-22

Virtual SAN has many predefined alarms.

Additional alarms can be configured using VMkernel Observations.

Name	Defined In
Errors occurred on the disk(s) of a Virtual SAN host	This Object
Expired Virtual SAN license	This Object
Expired Virtual SAN time-limited license	This Object
Host flash capacity exceeds the licensed limit for Virtual SAN	This Object
Registration/unregistration of a VASA vendor provider on a Virtual S...	This Object
Virtual SAN Health Alarm 'Active multicast connectivity check'	This Object
Virtual SAN Health Alarm 'Advanced Virtual SAN configuration in sync'	This Object
Virtual SAN Health Alarm 'After 1 additional host failure'	This Object
Virtual SAN Health Alarm 'All hosts contributing stats'	This Object
Virtual SAN Health Alarm 'All hosts have a Virtual SAN vmknic config...	This Object
Virtual SAN Health Alarm 'All hosts have matching multicast settings'	This Object
Virtual SAN Health Alarm 'All hosts have matching subnets'	This Object
Virtual SAN Health Alarm 'Basic (unicast) connectivity check (norma...	This Object

The default Virtual SAN alarms are used for monitoring hosts, cluster, and overall Virtual SAN health. Several alarms are used to alert the Virtual SAN health service.

The ESXi VMkernel Observation (VOB) log contains additional system events called observations that are logged by the VMkernel. Many of these observations are specific to Virtual SAN.

Leveraging VOBs, you can quickly and easily create vCenter Server alert mechanisms for Virtual SAN implementations.

Each VOB event is associated with an ID. Before you create a Virtual SAN alarm in the vCenter Server system, you must identify an appropriate VOB ID for the Virtual SAN event.

To review the list of VOB IDs for Virtual SAN, open the `vobd.log` file that is located on your ESXi host in the `/var/log` directory. The log file contains the VOB IDs that you can use for creating Virtual SAN alarms.

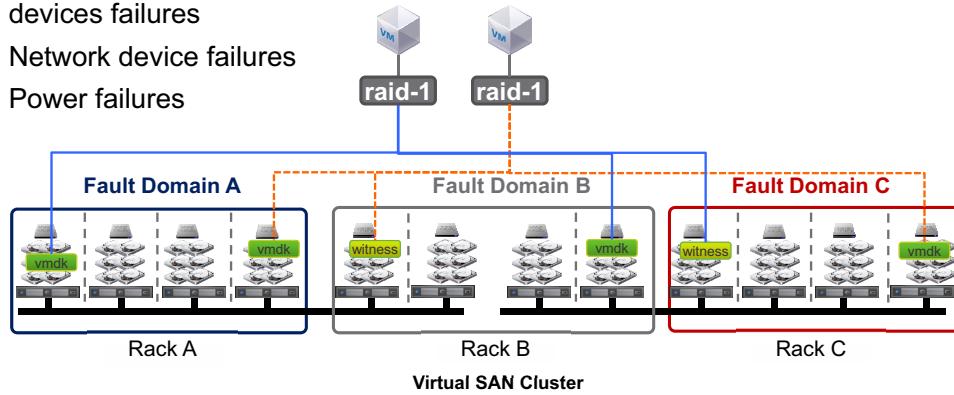
## Fault Domains

Slide 6-23

The Virtual SAN fault domains ensure that replicas of virtual machine data are distributed across defined resources.

Fault domains are primarily targeted at the ability to tolerate the following:

- Rack failures
- Cache and capacity devices failures
- Network device failures
- Power failures

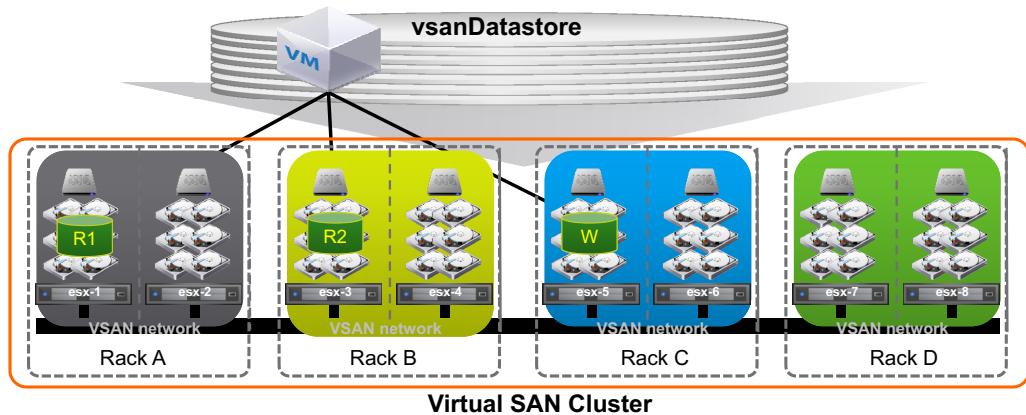


Fault domains must be used to spread replica components across servers in separate computing racks. This feature helps to protect the Virtual SAN environment from being disabled by a rack-level failure, such as a power outage.

## Example of a Fault Domain

Slide 6-24

Virtual SAN uses the fault domain feature with four racks each with two hosts.



When setting up fault domains, Virtual SAN requires a minimum of three fault domains. Each domain must contain at least a single ESXi host. VMware recommends using at least four fault domains to support all data evacuation modes and data protection configurations if a fault domain-level failure occurs.

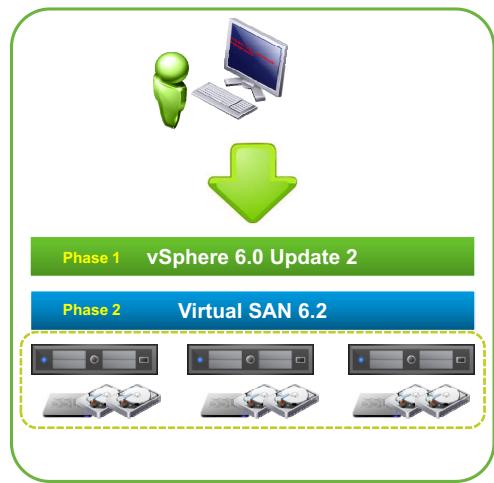
When fault domains are enabled, the storage policies are applied at the fault domain level rather than the host level. Fault domains can function as designed and redundancy exists between racks.

# Upgrade Overview

Slide 6-25

Upgrades are performed in multiple phases:

- Phase 1: Upgrade vSphere:
  - vCenter Server or vCenter Server Appliance
  - ESXi host
- Phase 2: Disk format conversion:
  - Satisfy prerequisites.
  - Perform the on-disk format upgrade using vSphere Web Client.



Virtual SAN 6.0 and later has a new on-disk format for disk groups and snapshots. Upgrading Virtual SAN from the previous version involves more than upgrading the ESXi and vCenter Server software.

Upgrading the disk format 1.0 to the new on-disk format 3.0 is optional. A Virtual SAN cluster continues to run smoothly if you choose to use disk format version 1.0.

For best results, upgrade the objects to use the new on-disk format 3.0. The new on-disk format 3.0 provides the complete feature set of Virtual SAN.

# Performing the On-Disk Format Upgrade

Slide 6-26

Upgrading from version 5.5 to 6.0 is done with Ruby vSphere Console.

Upgrading to newer Virtual SAN versions (6.1 and later) is done with vSphere Web Client.

The screenshot shows the vSphere Web Client interface for managing a Virtual SAN cluster named 'SA-VSAN-01'. The 'Manage' tab is selected. Under the 'Virtual SAN' section, the 'General' tab is active. In the 'On-disk Format Version' section, which is highlighted with a red box, the status is shown as 'Disk format version 3.0 (latest)' and 'Disks with outdated version 0 of 18'. There is also an 'Upgrade' button.

The disk format conversion phase is where the VMFS-L disk format is replaced by the VSAN FS format on all participating capacity devices. After starting the disk format conversion, a Virtual SAN node with ESXi 5.5 software is not allowed to join the Virtual SAN cluster.

## Labs

Slide 6-27

- Lab 10: Using Maintenance Mode Options
- Lab 11: Scaling Out a Virtual SAN Cluster
- Lab 12: Working with Fault Domains

VMware Confidential  
Internal Use Only

## Lab 10: Using Maintenance Mode Options

Slide 6-28

Use different maintenance mode options and observe host state changes

1. (Optional) Prepare the Environment
2. View Object Placement Across the Physical Disks in Virtual SAN
3. Put the Host in Maintenance Mode Using the No Data Migration Option
4. Put the Host in Maintenance Mode Using the Ensure Accessibility Option
5. Put the Host in Maintenance Mode Using the Full Data Migration Option

VMware Confidential  
Internal Use Only

## Lab 11: Scaling Out a Virtual SAN Cluster

Slide 6-29

Scale out the Virtual SAN cluster and perform management tasks

1. (Optional) Prepare the Environment
2. Add a Fourth Host to a Virtual SAN Cluster
3. Verify that the Fourth Host Has Flash Drives
4. Create Disk Groups for the Fourth Host
5. Test Maintenance Mode Data Evacuation with a Fourth Host
6. Demonstrate Disk Group Maintenance Without Maintenance Mode
7. Edit Custom Policy to Use RAID 5

VMware Confidential  
Internal Use Only

## Lab 12: Working with Fault Domains

Slide 6-30

Create fault domains and examine the effects on virtual machines

1. (Optional) Prepare the Environment
2. Create Three Fault Domains
3. Verify That Custom Storage Policy Works with the Second Virtual Machine
4. Enable Deduplication and Compression on the Cluster

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 6-31

You should be able to meet the following objectives:

- Manage hardware storage devices
- Manage hardware device failures
- Identify vCenter Server alarms for Virtual SAN events
- Describe and configure fault domains
- Upgrade to Virtual SAN 6.2

VMware Confidential  
Internal Use Only

## Key Points

Slide 6-32

- Resynchronization and rebalancing are important features for the day-to-day operation of Virtual SAN.
- Desirable performance and availability levels require analysis and ongoing monitoring.
- Two different types of failure events elicit recovery actions from the Virtual SAN cluster.
- Fault domains are a key feature to properly manage the Virtual SAN infrastructure.

Questions?

VMware Confidential  
Internal Use Only

## MODULE 7

# Monitoring and Troubleshooting Virtual SAN

Slide 7-1

Module 7

VMware Confidential  
Internal Use Only

# You Are Here

Slide 7-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
- 7. Monitoring and Troubleshooting Virtual SAN**
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 7-3

Virtual SAN is primarily configured and managed by using vSphere Web Client. In addition, Virtual SAN has several tools that provide methods for investigating problems and evaluating the health of the cluster.

The Virtual SAN health service is a critical component to evaluate the ongoing health and performance of Virtual SAN.

VMware Confidential  
Internal Use Only

## Module Lessons

Slide 7-4

- Lesson 1: Monitoring with vSphere Web Client
- Lesson 2: Monitoring with vRealize Operations Manager
- Lesson 3: Monitoring from the Command Line

VMware Confidential  
Internal Use Only

## Monitoring with vSphere Web Client

Slide 7-5

### Lesson 1: Monitoring with vSphere Web Client

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 7-6

By the end of this lesson, you should be able to meet the following objectives:

- Use vSphere Web Client to detect issues
- Use the health service to monitor Virtual SAN health
- Use the performance service to monitor Virtual SAN performance
- Proactively monitor and test the Virtual SAN environment

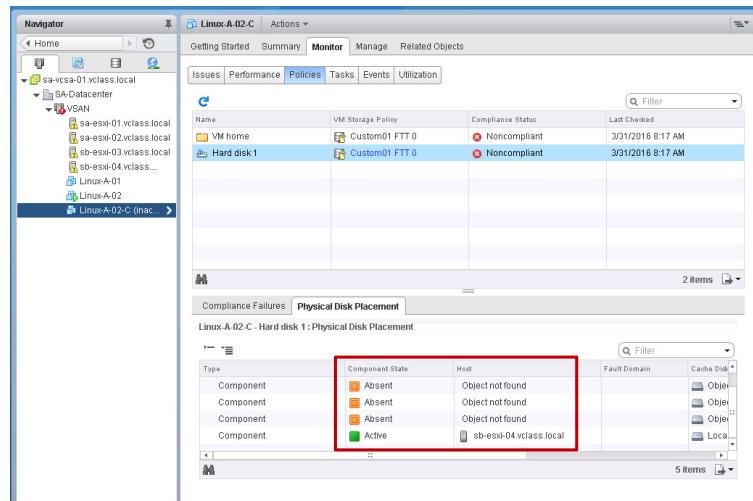
VMware Confidential  
Internal Use Only

# Monitoring with vSphere Web Client

Slide 7-7

vSphere Web Client provides diagnostic information that can help you troubleshoot Virtual SAN issues.

For example, in the **VM Storage Policies** tab, the **Physical Disk Placement** tab reports any issues regarding access to a virtual machine's components.

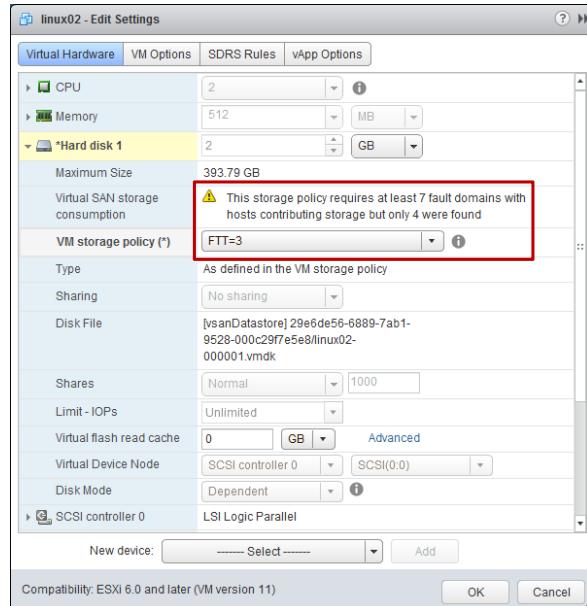


The vSphere Web Client user interface provides diagnostic information (warnings, errors, notifications) that can help you troubleshoot vSphere issues.

## Example: Storage Policy Issues

Slide 7-8

When you edit a virtual machine's settings to configure a virtual machine storage policy, vSphere Web Client verifies that the storage policy's resource requirements are met.

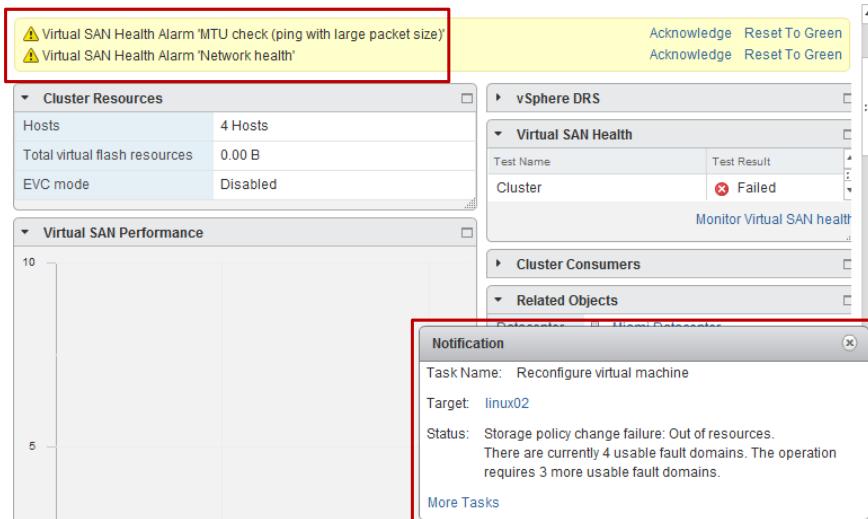


In the example, the storage policy named FTT=3 is assigned to the virtual machine named linux02. However, vSphere Web Client gives a warning that this storage policy requires at least seven fault domains, but only four fault domains were found.

## Example: Virtual SAN Health Issues

Slide 7-9

You can view errors, warnings, and notifications from the **Summary** tab of the selected vCenter Server inventory object.



The example shows the **Summary** tab of a Virtual SAN cluster. vSphere Web Client reports that two Virtual SAN health alarms were triggered. Also, a running task reported an error, which is shown in the Notification box.

# About the Health Service and Performance Service

Slide 7-10

Virtual SAN 6.2 introduces the health service and performance service:

- The health service actively tests and monitors the Virtual SAN environment.
- The performance service monitors performance-based metrics at the cluster, host, virtual machine, and virtual disk levels.

The screenshot shows the 'Manage' tab selected in the top navigation bar. Under the 'Health and Performance' section, the 'Health Service' tab is active. A red callout box points to the 'Health service status' field, which shows 'Enabled'. Another red callout box points to the 'Performance Service' section, which shows 'Performance Service is Turned OFF'.

Setting	Value
Health service status	Enabled
Health service version	6.2.0.0
Health check interval	60 minutes
HCL Database	Last updated: 20 days ago (3/15/2016)
Support Assistant	Last upload time: --
Performance Service	Turned OFF
Stats object health	
Stats object UUID	
Stats object storage policy	
Compliance status	

The Virtual SAN health service checks all aspects of a Virtual SAN cluster. The health service performs checks on a number of items, such as hardware compatibility, network connectivity, storage device health, and cluster health.

Using the health service, Virtual SAN administrators can ensure that the Virtual SAN deployment is fully supported, functional, and operational. Administrators can also receive immediate indications to a root cause if a failure occurs.

A healthy Virtual SAN environment performs efficiently. The performance service provides a number of graphs and data points that provide performance information at the cluster, host, virtual machine, and virtual disk levels.

# Health Service Tests

Slide 7-11

The health service tests a variety of items and conditions.

Test Type	Items and Conditions Tested
Cluster	Advanced configurations, deduplication/compression consistency, disk format, disk groups, CLOMD liveness, disk balance
Network	VMkernel port, subnets, multicast settings, connectivity issues, MTU check, hosts with Virtual SAN disabled
Data	Object health
Physical disk	Overall disk health, congestion, disk capacity, memory pools, metadata health
Hardware compatibility	Controller compatibility, issues retrieving hardware information, up-to-date HCL database
Limits	Host component limits, Virtual SAN component limits, disk space, current cluster situation
Performance service	Hosts contributing stats, data collection, health of stats database

The health service is quite thorough in the number of tests it performs. As an example, proper network configuration is key to a healthy Virtual SAN cluster. Eleven tests exist in the Network section of the Health pane.

## Example: Troubleshooting a Network Health Issue (1)

Slide 7-12

In this example, the Health pane shows that one or more network tests have failed.

Expand the Network category to view the individual tests.

The screenshot shows the VMware Virtual SAN Health pane. The top navigation bar includes 'SA-VSAN-01', 'Actions ▾', 'Getting Started', 'Summary', 'Monitor' (which is selected), 'Manage', and 'Related Objects'. Below this is a tab bar with 'Issues', 'Profile Compliance', 'Performance', 'Utilization', 'Tasks', 'Events', 'Resource Reservation', 'vSphere DRS', and 'Virtual SAN' (selected). On the left, a sidebar lists 'Physical Disks', 'Virtual Disks', 'Resyncing Components', 'Health' (selected), 'Capacity', and 'Proactive Tests'. The main content area is titled 'Virtual SAN Health (Last checked: Today at 4:50 PM)'. It displays a table with two columns: 'Test Result' and 'Test Name'. The table shows the following data:

Test Result	Test Name
Failed	Network
Failed	Physical disk
Failed	Cluster
Warning	Data
Warning	Hardware compatibility
Passed	Limits
Passed	Performance service

The 'Failed' row for 'Network' is highlighted with a red box.

Monitor the Health pane on a regular basis. Expand the test category to view the individual tests. Select a test to view detailed test results. You can click **Retest** (not shown in screenshot) to manually run all the tests. Otherwise, the tests are run every 60 minutes.

Identify the test categories that do not have a status of Passed. If an issue is detected in the environment, a result of Failed or Warning appears next to the test category in the Health pane. Maximizing the test category shows the specific test or tests that failed or produced a warning.

## Example: Troubleshooting a Network Health Issue (2)

Slide 7-13

One test failed and two tests have warnings.

Select the test that has failed.

Virtual SAN Health (Last checked: Today at 2:48 PM)	
Test Result	Test Name
Failed	Network
Failed	All hosts have a Virtual SAN vmknic configured
Warning	All hosts have matching subnets
Warning	MTU check (ping with large packet size)
Passed	All hosts have matching multicast settings
Passed	Basic (unicast) connectivity check (normal ping)
Passed	Hosts disconnected from VC
Passed	Hosts with connectivity issues
Passed	Hosts with Virtual SAN disabled
Passed	Multicast assessment based on other checks
Passed	Unexpected Virtual SAN cluster members
Passed	Virtual SAN cluster partition

In each of the test categories, identify the tests that did not pass. Clicking on the specific test provides more details about why the test failed or produced a warning.

## Example: Troubleshooting a Network Health Issue (3)

Slide 7-14

View the details of the test that has failed.

Click **Ask VMware** for additional help.

Virtual SAN Health (Last checked: Today at 2:48 PM) Retest

Test Result	Test Name
Failed	Network
Failed	All hosts have a Virtual SAN vmknic configured
Warning	All hosts have matching subnets

**All hosts have a Virtual SAN vmknic configured** Ask VMware

Checks if all the hosts in the Virtual SAN cluster have a configured vmknic with Virtual SAN traffic enabled.

Hosts with no Virtual SAN vmknic present

Host
sa-esxi-02.vclass.local

View the details of each test that did not pass. From the details, you might be able to identify the cause of the issue. If you cannot identify the root cause, then click **Ask VMware** for further details.

## Example: Troubleshooting a Network Health Issue (4)

Slide 7-15

The **Ask VMware** button takes you to a VMware reference.

In this example, you are taken to knowledge base article 2108062.

The screenshot shows a web browser displaying the VMware Knowledge Base. The URL in the address bar is [https://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=2108062](https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2108062). The page header includes the VMware logo and links for United States, My VMware, Partner Central, Training, and Community. Below the header, there are navigation links for Products, Cloud Services, Support, Downloads, and Configuration. The main content area is titled "Knowledge Base" and contains a search bar with the placeholder "Search the VMware Knowledge Base (KB)". A dropdown menu labeled "Products -->" is open. A red box highlights a specific search result: "Virtual SAN Health Service - Network Health - All hosts have a VSAN vmknic configured (2108062)". Below this result, a section titled "Purpose" is visible, with a note explaining the purpose of the check and providing details on why it might report an error.

Clicking the **Ask VMware** button takes you to a VMware knowledge base article that describes the test, probable causes, and advice on how to resolve the issue.

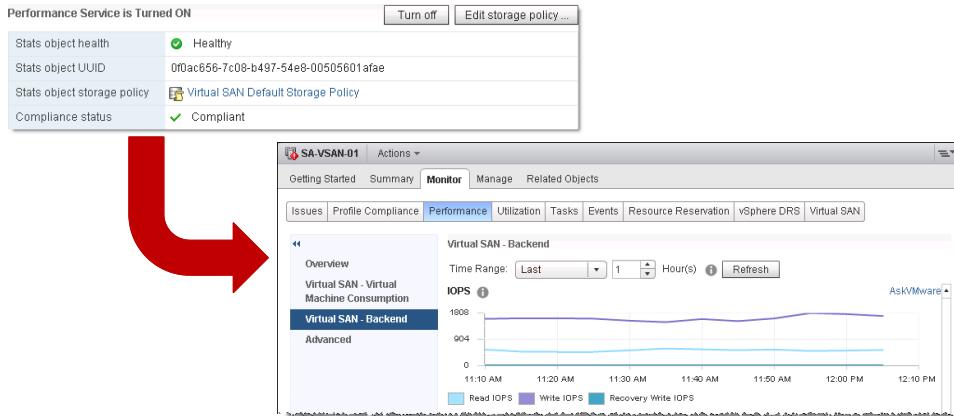
# Enabling the Performance Service

Slide 7-16

To use the performance service, the service must be enabled.

When the service is enabled, the performance history database is created and stored as an object, referred to as the Stats object.

- The Virtual SAN Default Storage Policy is assigned to the Stats object. The storage policy can be changed.



The performance history database is not a component of vCenter Server. The performance history database is stored as a Virtual SAN object, independent of vCenter Server. The Virtual SAN default storage policy is assigned to this object. If the object becomes unavailable, performance history for the cluster will be unavailable until access to the object is restored. The performance history database can consume up to 255 GB of capacity on the Virtual SAN datastore.

## Performance Views

Slide 7-17

Performance views contain metrics for monitoring Virtual SAN clusters, hosts, and virtual machines.

View Name	Description	Available For
Virtual Machine Consumption	Shows performance metrics for virtual machines running on Virtual SAN	Clusters, hosts, virtual machines
Backend	Shows what is required to deliver what is visible at the virtual machine and object level	Clusters, hosts
Disk Group	Shows how each disk group is performing, independent of other disk groups on the same host	Hosts
Disk	Shows performance metrics of a single device	Hosts, virtual machines

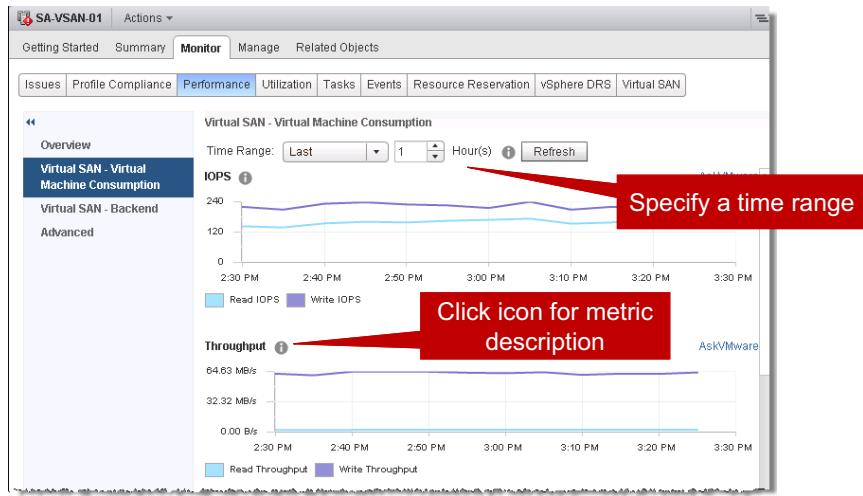
The **View Name** column lists the names of performance graphs that you can display in vSphere Web Client when the performance service is enabled on the Virtual SAN cluster. These performance graphs are located in the **Monitor > Performance** tab when you select the Virtual SAN cluster, a host in that cluster, or a virtual machine in that cluster.

# Cluster Metrics: Virtual Machine Consumption

Slide 7-18

Cluster metrics provide quick visibility to the entire Virtual SAN cluster.

This pane shows how the virtual machines and their objects as a whole are performing.



The Virtual Machine Consumption pane shows graphs for IOPS, throughput, latency, congestion, and outstanding I/Os.

The time range can be modified to show information from the last few hours or a custom date and time range.

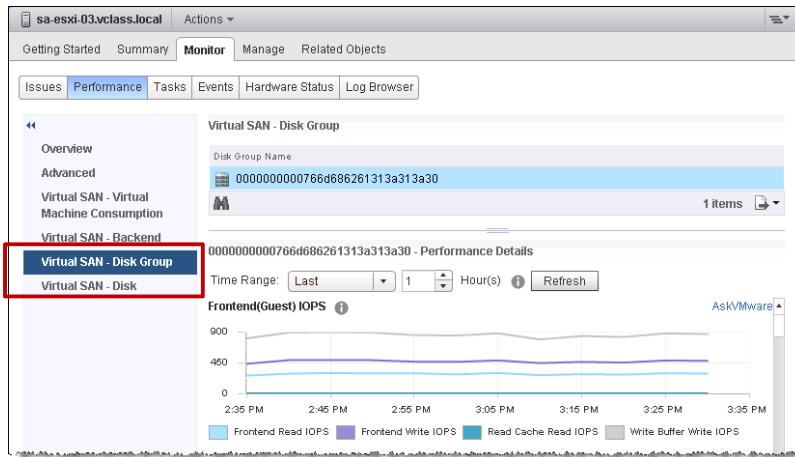
The **Information** icon provides a brief description of the metric or item in the display.

## Host Metrics: Disk Group and Disk

Slide 7-19

The Disk Group pane displays performance metrics on a per-disk group basis.

The Disk pane shows the IOPS, throughput, and latency for a physical disk in a disk group.



Virtual SAN requires at least one disk group, and can be configured with up to five disk groups per host. You must be able to distinguish how each disk group is performing, independent of any other disk groups on the same host.

The Disk Group pane has a number of graphs that enable you to monitor the performance of the cache tier and Virtual SAN internal queues. The pane also has a graph that shows the disk group capacity and usage.

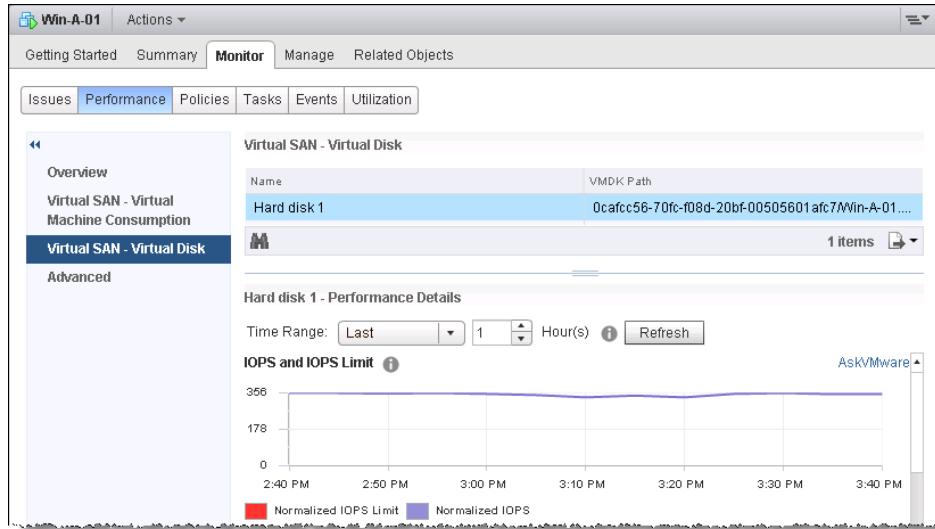
The Disk pane shows performance at the physical layer.

Use the **Information** icon next to each metric to get a description of the metric.

# Virtual Machine Metrics: Virtual Disk

Slide 7-20

The Virtual Disk pane shows metrics for each individual disk (VMDK) on the selected virtual machine.



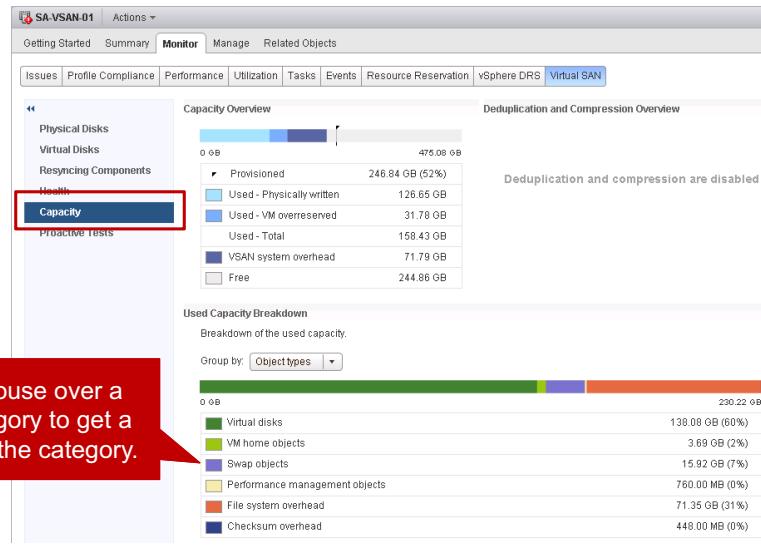
A virtual machine can be configured with one or more virtual disks. The Virtual Disk pane shows IOPS, throughput, and latency for the selected virtual disk.

Each virtual disk can be assigned a different storage policy. The storage policy's settings can contribute to the performance characteristics of the virtual disk. For example, you might create a storage policy that sets an IOPS limit for the virtual machine's object.

# Viewing Virtual SAN Capacity Details

Slide 7-21

You can use the Capacity pane to view the capacity consumption of various object types.



The Capacity pane makes it easy for administrators to understand the amount of capacity that various object types are consuming.

Performance management objects refer to objects created for storing performance metrics, such as the performance history database. Performance management objects are created when the performance service is enabled.

File system overhead refers to on-disk file system (VSAN FS) overhead, which is either attributed to deduplication, compression, or checksum overhead. When deduplication and compression is enabled, file system overhead is increased 10 times to reflect the increase in the logical size of the Virtual SAN datastore. Deduplication and compression overhead includes the associated mapping table, hash tables, and other mechanisms required by deduplication and compression.

Checksum overhead is the overhead needed to store all the checksums. When the deduplication and compression feature is enabled, checksum overhead is increased 10 times to reflect the increase in the logical size of the Virtual SAN datastore.

# Running Proactive Tests

Slide 7-22

Proactive tests can be used to check the integrity of your Virtual SAN cluster:

- Useful for checking that your Virtual SAN cluster is working properly before putting it into production.

The screenshot shows the vSphere Client interface for a cluster named 'SA-VSAN-01'. The 'Monitor' tab is selected. In the left sidebar, the 'Proactive Tests' item is highlighted with a red box. The main pane displays a table of proactive test results:

Name	Last Run Result	Last Run Time
VM creation test	Passed	2/18/2016 11:02 AM
Multicast performance test	Failed	2/22/2016 5:55 PM
Storage performance test	Failed	2/22/2016 5:58 PM

Below the table, a detailed view for the 'Multicast performance test' is shown, titled 'Multicast performance test - Details'. It lists the hosts and their performance results:

Host	Health Status	Received Bandwidth (MB/s)	Maximum Achievable Bandwidth ...
sa-esxi-04.vclass.local	Failed	13.79	125.00
sa-esxi-01.vclass.local	Failed	11.98	125.00
sa-esxi-06.vclass.local	Warning	20.90	125.00
sa-esxi-05.vclass.local	Failed	16.85	125.00

The health service contains a set of tests that you can proactively, instead of reactively, run on a Virtual SAN cluster. You can use these tests to verify that the Virtual SAN environment is functioning properly and performing as expected.

The following proactive tests are available:

- VM creation test: Test the ability to successfully create virtual machines on the Virtual SAN datastore.
- Multicast performance test: Test network connectivity and verify that the multicast speeds are suitable for Virtual SAN traffic.
- Storage performance test: Test the stability of the cluster under a particular workload. You can choose from one of several tests, such as a low stress test, a performance test of 100 percent reads and optimal read cache usage, or a performance test of 100 percent writes and optimal write buffer usage.

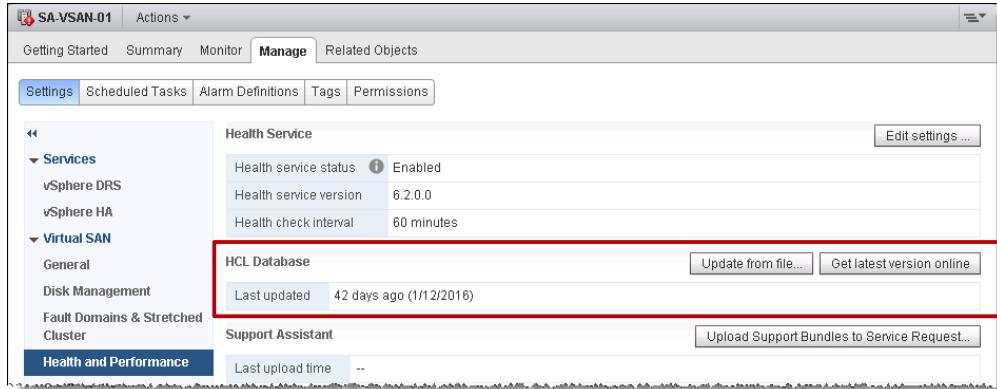
# Updating the HCL Database

Slide 7-23

The hardware compatibility list (HCL) database contains a list of certified hardware, firmware, and drivers.

You can ensure that the HCL database contains up-to-date information:

- This information is used by the health service tests.



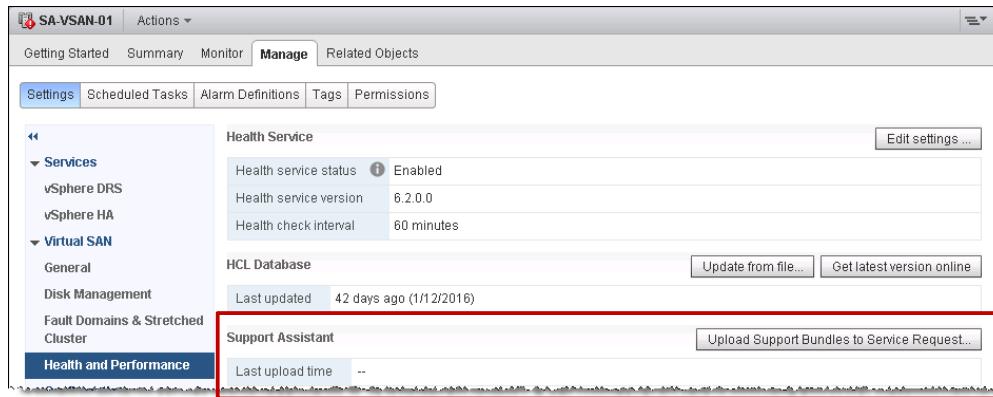
vSphere and Virtual SAN support a wide variety of hardware configurations. The list of hardware components and corresponding drivers that are supported with Virtual SAN can be found in the VMware Compatibility Guide. You must use only hardware, firmware, and drivers found in this guide to help ensure stability and performance of a Virtual SAN environment.

From the Health and Performance pane, you can easily update the information in the HCL database. If the environment has Internet connectivity, updates can be obtained directly from VMware. Otherwise, HCL updates can be downloaded as a file to enable offline updates.

# Uploading Support Bundles

Slide 7-24

If an issue arises that requires assistance from VMware technical support, you can upload support bundles to VMware technical support to help expedite the troubleshooting process.



Clicking **Upload Support Bundles to Service Request** enables an administrator to enter an existing support request number and upload the necessary logs to VMware Technical Support.

## Lab 13: Using the Health and Performance Services

Slide 7-25

Monitor Virtual SAN health, performance, and capacity

1. (Optional) Prepare the Environment
2. Simulate a Host Failure to Trigger Health Alerts
3. Monitor the Health Service for Failed Tests
4. View Information About the Failed Network Tests
5. View Information About the Failed Physical Disk Tests
6. View Information About the Failed Cluster Tests
7. Resolve the Host Failure
8. Enable the Performance Service
9. View Performance Graphs
10. View Virtual SAN Capacity Details

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 7-26

You should be able to meet the following objectives:

- Use vSphere Web Client to detect issues
- Use the health service to monitor Virtual SAN health
- Use the performance service to monitor Virtual SAN performance
- Proactively monitor and test the Virtual SAN environment

VMware Confidential  
Internal Use Only

# Monitoring with vRealize Operations Manager

Slide 7-27

## Lesson 2: Monitoring with vRealize Operations Manager

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 7-28

By the end of this lesson, you should be able to meet the following objectives:

- Use VMware vRealize® Operations Manager™ dashboards to monitor Virtual SAN health and performance
- Monitor vRealize Operations Manager alerts and resolve Virtual SAN health issues

VMware Confidential  
Internal Use Only

# Overview of vRealize Operations Manager

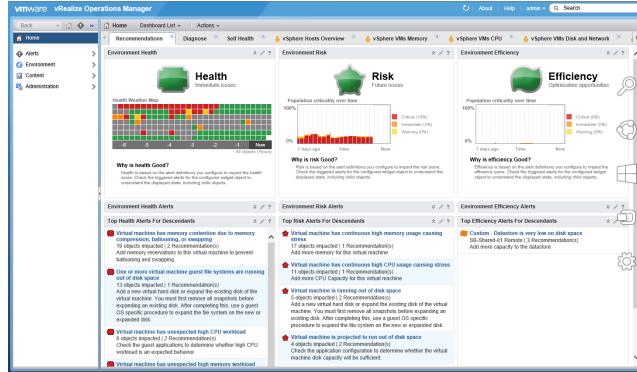
Slide 7-29

vRealize Operations Manager serves as a central point of management for your virtual infrastructure:

- Provides visibility and insights into the environment's performance, capacity, and health.

Management Pack for Storage Devices 6.0.2 extends the capabilities of vRealize Operations Manager:

- Includes predefined dashboards and alerts that provide visibility into your Virtual SAN 6.x environment.



VMware vRealize® Operations Manager™ collects performance data from each object at every level of your environment. vRealize Operations Manager stores and analyzes the data, and uses that analysis to provide real-time information about issues, or potential issues, anywhere in your environment.

vRealize Operations Manager gives you the whole picture of your environment:

- An accurate understanding of overall performance, health, and availability
- Accurate identification of and navigation to the cause
- A real understanding of normal behavior in your environment

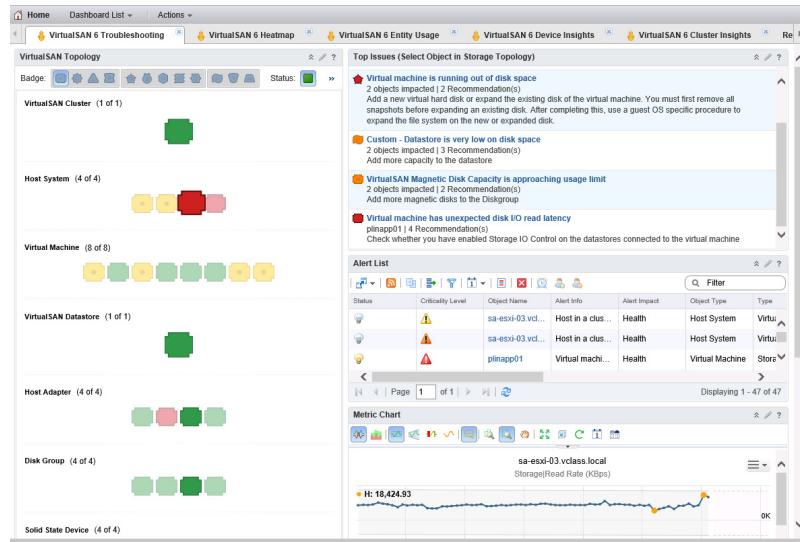
VMware vRealize® Operations Management Pack™ for Storage Devices 6.0.2 provides visibility into your Virtual SAN 6.0 environment. Predefined dashboards enable you to follow the path from a virtual machine to Virtual SAN and identify any problem that might exist along that path.

The management pack also includes a set of alert definitions that are used to alert on events occurring in the Virtual SAN cluster.

# Global View

Slide 7-30

The Virtual SAN 6 Troubleshooting dashboard provides visibility across multiple Virtual SAN clusters.



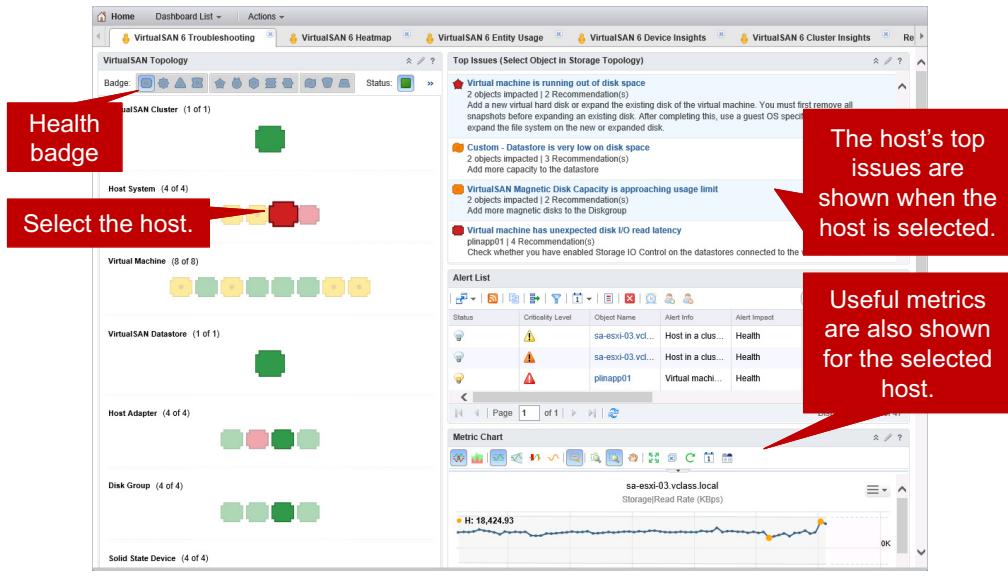
This dashboard has the following panels:

- **VirtualSAN Topology:** This panel shows a topology map of objects and their relationships with other objects in the inventory. By default, the health badge is shown. You can choose to display other badges (such as the risk badge, the efficiency badge, or any of the minor badges) in the topology map.
- **Top Issues:** This panel lists the alerts with the greatest significance for the object that is selected in the topology map. The alerts list can have an impact on health, capacity risk, or efficiency of resource usage. The issues listed are most likely to negatively affect the environment.
- **Alert List:** This panel lists all the alerts, both open and closed, which are related to the object that is selected in the topology map.
- **Metric Chart:** This panel shows the metrics that are being collected for the object that is selected in the topology map.

# Example: Troubleshooting a Health Issue (1)

Slide 7-31

The topology chart shows a host whose health is critical (red).



In this example, the topology map shows the health status for the objects in the map. The Virtual SAN cluster object contains four host systems. One of the hosts is displayed in red, which means its health is poor. This host has been selected by the user and as a result, the objects related to the selected host are highlighted. The cluster contains eight virtual machines. These virtual machines reside on the Virtual SAN datastore. Also represented in the topology map are the storage adapters for the hosts in the cluster (the adapter in the selected host is highlighted), the disk groups belonging to the cluster (the selected host's disk group is highlighted), and off the bottom of the screenshot are the solid state devices and magnetic devices belonging to the cluster.

The host has the following issues that are considered top priority:

- Virtual machine is running out of disk space: The virtual machine's disks are getting full. This issue has an impact on capacity, and is categorized as a risk-related issue (identified by the risk badge to the left of the issue).
- Custom - Datastore is very low on disk space: This issue has an impact on long-term resource usage, and is categorized as an efficiency-related issue (identified by the efficiency badge to the left of the issue).
- VirtualSAN Magnetic Disk Capacity is approaching usage limit: This issue affects the health of the Virtual SAN datastore and is categorized as a health-related issue.

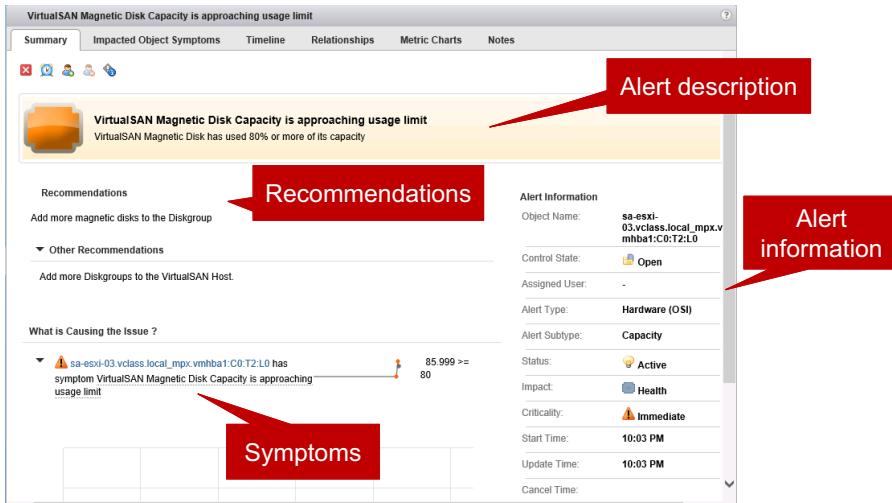
- Virtual machine has unexpected disk I/O read latency: This issue affects the health of a virtual machine.

VMware Confidential  
Internal Use Only

## Example: Troubleshooting a Health Issue (2)

Slide 7-32

By clicking an issue in the Top Issues panel, you can get details about the alert that was triggered.



In the example, clicking **VirtualSAN Magnetic Disk Capacity is approaching usage limit** takes you to the alert's **Summary** tab.

The alert's **Summary** tab provides a description of the alert. The Alert Information panel provides the name of the object on which the alert was triggered, the status of the alert, and the date on which the alert was triggered.

The **Summary** tab also lists symptoms, recommendations, and actions, if actions are configured in the alert definition.

You can use the **Summary** tab as a starting point for resolving the issue that triggered the alert.

In the example, a magnetic disk on the host named sa-esxi-03.vclass.local has used 80 percent or more of its capacity. The magnetic disk is identified as mpx.vmhba1:C0:T2:L0. The following recommendations are made:

- Add more magnetic disks to the disk group.
- Add more disk groups to the host.

## Virtual SAN Alerts

Slide 7-33

The Management Pack for Storage Devices provides a set of health alerts for monitoring the Virtual SAN environment.

Name ▾	Object Type
One or more VirtualSAN components not active	Virtual Machine
Virtual Machine is not in compliance with VirtualSAN storage profile	Virtual Machine
VirtualSAN Cluster partitioned	VirtualSAN Cluster
VirtualSAN Cluster partitioned most likely due to multicast issue in cluster	VirtualSAN Cluster
VirtualSAN Cluster partitioned most likely due to network Switch Port is down on a host in cluster	VirtualSAN Cluster
VirtualSAN Cluster partitioned most likely due to Physical NIC is down on a host in cluster	VirtualSAN Cluster
VirtualSAN Host component count limit approaching	Host System
VirtualSAN Host has a Host Adapter that has gone bad or is removed	Host System
VirtualSAN Host has misconfigured storage.	Host System
VirtualSAN Magnetic Disk Capacity is approaching usage limit	Magnetic Disk

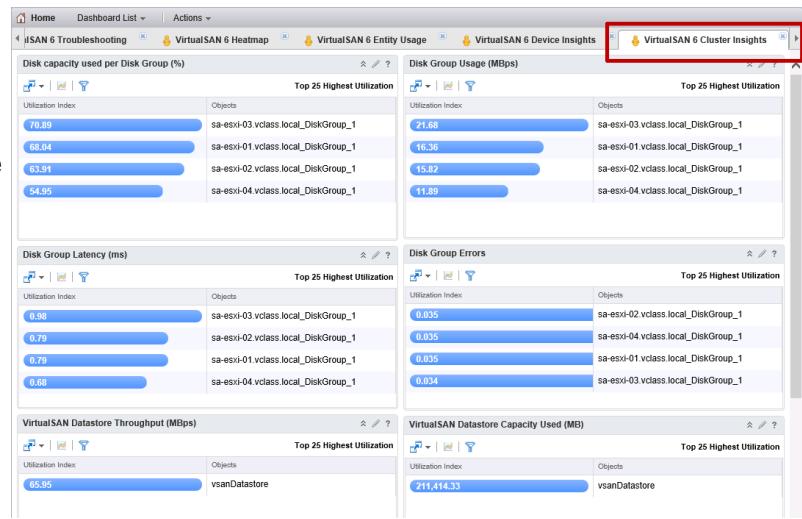
Virtual SAN alerts exist to monitor various objects in the Virtual SAN cluster. For example, an alert is triggered on a virtual machine in the Virtual SAN cluster if the virtual machine is not in compliance with its assigned storage policy.

# Monitoring the Virtual SAN Cluster

Slide 7-34

The VirtualSAN 6 Cluster Insights dashboard provides a closer look at cluster details, such as capacity used at the disk group and datastore levels.

Monitor these values over time to know what values are acceptable for your environment.



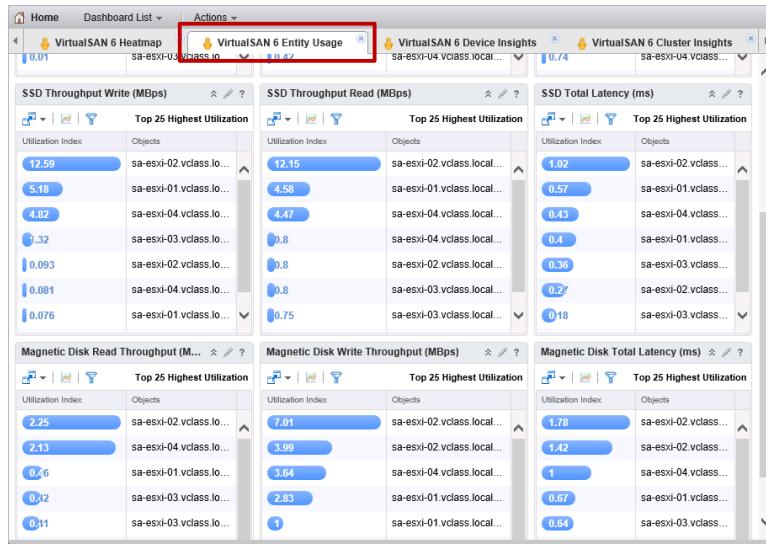
The VirtualSAN 6 Cluster Insights dashboard provides useful information about the cluster as a whole. This dashboard provides a closer look at cluster-specific details, such as the following:

- Virtual SAN datastore throughput and capacity used
- Disk group capacity, latency, and errors

# Monitoring Entity Usage

Slide 7-35

The Virtual SAN 6 Entity Usage dashboard shows throughput and latency values for SSD devices, magnetic disks, and host bus adapters.



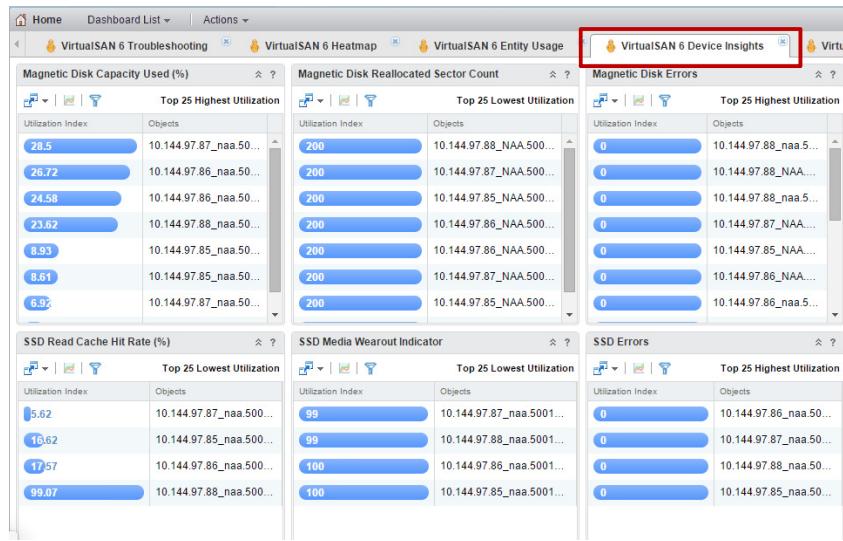
Multiple charts are available on the VirtualSAN 6 Entity Usage dashboard. These charts show the top 25 hosts with the highest utilization in several categories, such as throughput and latency.

As you monitor this dashboard and other Virtual SAN dashboards over time, you will know what values are acceptable for your environment.

# Viewing Device Information

Slide 7-36

The VirtualSAN 6 Device Insights dashboard provides information, such as read cache hit rate and SSD media wearout.



The VirtualSAN 6 Device Insights dashboard provides detailed device information. SSD Read Cache Hit Rate (%) and SSD Media Wearout Indicator are of particular interest. The Magnetic Disk Reallocated Sector Count and SSD Media Wearout Indicator metrics are displayed only if the controller provides Self-Monitoring, Analysis and Reporting Technology (SMART) metrics. If the controller does not provide SMART metrics, then the corresponding panels in the dashboard will be blank.

This dashboard also provides information on a host's CPU and memory usage. Because a host's CPU and memory are considered to be devices in a Virtual SAN cluster, host CPU and memory usage values are included on this dashboard. These numbers include overall utilization, not just the small amount of CPU cycles and memory used by Virtual SAN.

## Review of Learner Objectives

Slide 7-37

You should be able to meet the following objectives:

- Use vRealize Operations Manager dashboards to monitor Virtual SAN health and performance
- Monitor vRealize Operations Manager alerts and resolve Virtual SAN health issues

VMware Confidential  
Internal Use Only

## Monitoring from the Command Line

Slide 7-38

### Lesson 3: Monitoring from the Command Line

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 7-39

By the end of this lesson, you should be able to meet the following objectives:

- Use ESXi commands to monitor the Virtual SAN environment
- Use Ruby vSphere Console (RVC) to get detailed information about the Virtual SAN environment
- Use Virtual SAN Observer to view Virtual SAN performance

VMware Confidential  
Internal Use Only

## About vSphere ESXi Shell

Slide 7-40

vSphere ESXi Shell includes a set of commands for diagnosing and repairing ESXi hosts, such as the following:

- esxcfg-info:
  - Command that provides a view of the internal state of the VMkernel
- esxtop:
  - Text-based utility to examine real-time and historical resource usage for ESXi hosts
- esxcli:
  - Set of commands for viewing and managing ESXi hosts

Use these tools to obtain information about Virtual SAN and to troubleshoot your Virtual SAN environment.

An ESXi host includes a direct console user interface (DCUI) that enables you to start and stop the system and perform a limited set of maintenance and troubleshooting tasks. The DCUI includes vSphere ESXi Shell, which is disabled by default. You can enable vSphere ESXi Shell in the DCUI or by using vSphere Web Client.

You can enable local shell access or remote shell access:

- Local shell access enables you to log in to the shell directly from the DCUI.
- Remote shell access enables you to connect securely to the host using SSH and a client, such as PuTTY.

## esxcfg-info Command

Slide 7-41

The esxcfg-info command displays the queue depth of your controllers.

```
[root@vsan-node4:~] esxcfg-info -s | grep "==+SCSI Interface" -A 10
\==+SCSI Interface :
|----Name.....vmhba0
|----UID.....ide.vmhba0
|----Driver.....ata_piix
|----Queue Depth.....1
|----Is Virtual.....false
\==+Data Integrity Information :
|----Protection Mask.....0x00000000
|----Guard Type.....NO GUARD SUPPORT
\==+PCI Device :
|----Segment.....0x0000
--

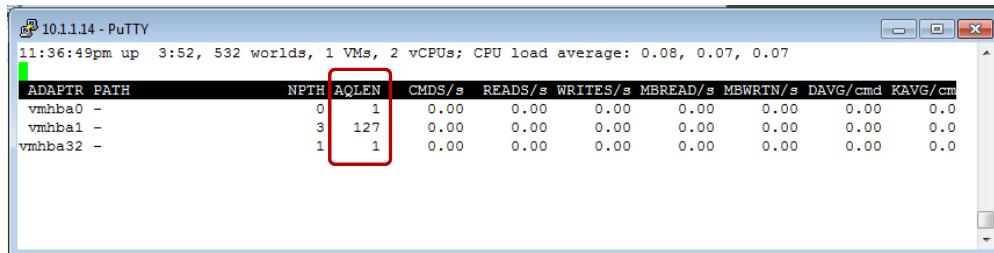
\==+SCSI Interface :
|----Name.....vmhba1
|----UID.....pscsi.vmhba1
|----Driver.....mptspi
|----Queue Depth.....127
|----Is Virtual.....false
\==+Data Integrity Information :
|----Protection Mask.....0x00000000
|----Guard Type.....NO GUARD SUPPORT
\==+PCI Device :
|----Segment.....0x0000
```

VMware Confidential  
Internal Use Only

## esxtop Command: d Option

Slide 7-42

The `d` option for `esxtop` displays the queue depth of the storage controllers.



A screenshot of a PuTTY window titled "10.1.1.14 - PuTTY". The window displays the output of the esxtop command with the "d" option. The output shows statistics for three storage adapters: vmhba0, vmhba1, and vmhba32. The columns include ADAPTR, PATH, NPORT, AQLEN, CMDS/s, READS/s, WRITES/s, MBREAD/s, MBWRTN/s, DAVG/cmd, and KAVG/cm. The AQLEN column is highlighted with a red box. The data is as follows:

ADAPTR	PATH	NPORT	AQLEN	CMDS/s	READS/s	WRITES/s	MBREAD/s	MBWRTN/s	DAVG/cmd	KAVG/cm
vmhba0	-	0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.0
vmhba1	-	3	127	0.00	0.00	0.00	0.00	0.00	0.00	0.0
vmhba32	-	1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.0

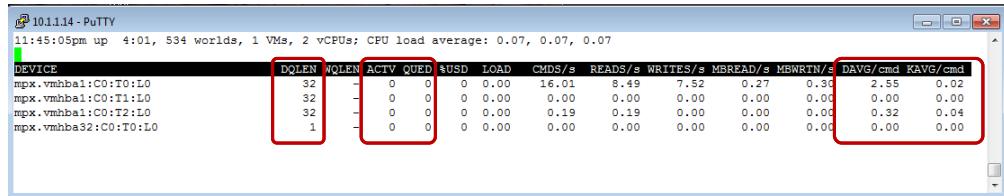
The key column to observe is AQLEN.

The AQLEN column only appears if you press `f` to add fields and select the `D` option.

## esxtop Command: u Option

Slide 7-43

The `u` option for `esxtop` displays disk-related information.



A screenshot of a PuTTY terminal window titled "10.1.1.14 - PuTTY". The window displays the output of the esxtop command with the "u" option. The output shows disk statistics for several devices, including "mpx.vmhba1:C0:T0:L0", "mpx.vmhba1:C0:T1:L0", "mpx.vmhba1:C0:T2:L0", and "mpx.vmhba32:C0:T0:L0". The columns shown are DEVICE, DQLEN, WQLEN, ACTV, QUED, %USD, LOAD, CMDS/s, READS/s, WRITES/s, MBREAD/s, MBWRIT/s, DAVG/cmd, and KAVG/cmd. Red boxes highlight the first four columns (DQLEN, WQLEN, ACTV, QUED) and the last two columns (DAVG/cmd, KAVG/cmd).

DEVICE	DQLEN	WQLEN	ACTV	QUED	%USD	LOAD	CMDS/s	READS/s	WRITES/s	MBREAD/s	MBWRIT/s	DAVG/cmd	KAVG/cmd
mpx.vmhba1:C0:T0:L0	32	-	0	0	0.00	16.01	8.49	7.52	0.27	0.30	2.55	0.02	
mpx.vmhba1:C0:T1:L0	32	-	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
mpx.vmhba1:C0:T2:L0	32	-	0	0	0.00	0.19	0.19	0.00	0.00	0.00	0.32	0.04	
mpx.vmhba32:C0:T0:L0	1	-	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

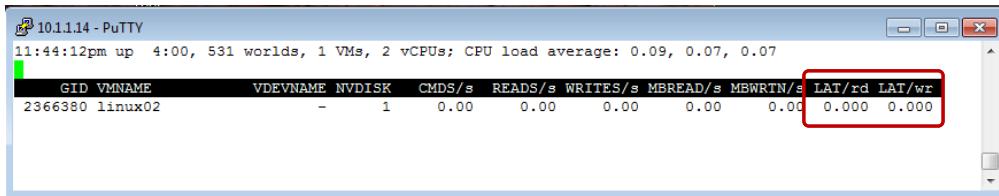
The key columns to observe are DQLEN, ACTV, QUED, DAVG/cmd, and KAVG/cmd.

VMware Confidential  
Internal Use Only

## esxtop Command: v Option

Slide 7-44

The `v` option for `esxtop` displays virtual machine disk-related information.



A screenshot of a PuTTY window titled "10.11.14 - PuTTY". The window displays the output of the esxtop command with the "-v" option. The output shows system statistics and then focuses on a single virtual machine (VMID 2366380) named "linux02". The columns shown are GID, VMNAME, VDEVNAME, NVDISK, CMDS/s, READS/s, WRITES/s, MBREAD/s, MBWRIT/s, LAT/rd, and LAT/wr. The last two columns, LAT/rd and LAT/wr, are highlighted with a red box.

GID	VMNAME	VDEVNAME	NVDISK	CMDS/s	READS/s	WRITES/s	MBREAD/s	MBWRIT/s	LAT/rd	LAT/wr
2366380	linux02	-	1	0.00	0.00	0.00	0.00	0.00	0.000	0.000

The key columns to observe are LAT/rd and LAT/wr.

VMware Confidential  
Internal Use Only

## esxtop Command: m Option

Slide 7-45

The `m` option for `esxtop` displays memory-related information.

The screenshot shows the output of the `esxtop m` command in a PuTTY window. The top part of the output provides system-level memory statistics:

```
11:40:02pm up 3:55, 533 worlds, 1 VMs, 2 vCPUs; MEM overcommit avg: 0.00, 0.00, 0.00
PMEM /MB: 8191 total: 2978 vmk,216 other, 4996 free
VMKMEM/MB: 8152 managed: 408 minfree, 5977 rsvd, 2175 ursvd, high state
PSHARE/MB: 38 shared, 33 common: 5 saving
SWAP /MB: 0 curr, 0 rclmtgt: 0.00 r/s, 0.00 w/s
ZIP /MB: 0 zipped, 0 saved
MEMCTL/MB: 0 curr, 0 target, 0 max
```

The bottom part of the output is a table showing memory usage for a specific process:

GID	NAME	MEMSZ	GRANT	CNSM	SZTGT	TCHD	TCHD W	SWCUR
2366380	linux02	512.00	5.00	0.54	13.03	0.00	0.00	0.00

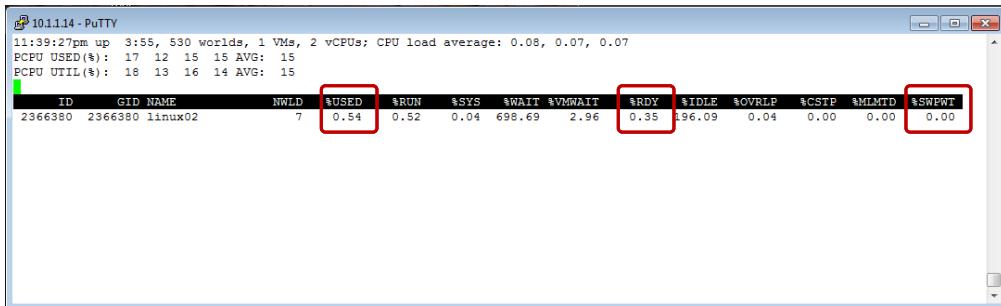
The key column to observe is SWCUR.

VMware Confidential  
Internal Use Only

## esxtop Command: c Option

Slide 7-46

The `c` option for `esxtop` displays CPU-related information.



```
10.1.1.14 - PuTTY
11:39:27pm up 3:55, 530 worlds, 1 VMs, 2 vCPUs; CPU load average: 0.08, 0.07, 0.07
PCPU USED(%): 17 12 15 15 AVG: 15
PCPU UTIL(%): 18 13 16 14 AVG: 15

ID      GID NAME      NWLD %USED   %RUN    %SYS    %WAIT   %VMWAIT   %RDY    %IDLE   %OVRLP   %CSTP   %MLMTD   %SWPWT
2366380 2366380 linux02 7 0.54    0.52    0.04   698.69   2.96    0.35   196.09   0.04    0.00    0.00    0.00
```

The key columns to observe are %USED, %RDY, and %SWPWT.

VMware Confidential  
Internal Use Only

## esxtop Command: n Option

Slide 7-47

The `n` option for `esxtop` displays network-related information.

PORT-ID	USED-BY	TEAM-PNIC	DNAME	PKTIX/s	MbTX/s	PSZTX	PKTRX/s	MbRX/s	PSZRX	%DRPTX	%DRPRX
33554433	Management	n/a	vSwitch0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33554434	vmnic0		vSwitch0	0.00	0.00	0.00	11.64	0.02	201.00	0.00	0.00
33554435	Shadow of vmnic0	n/a	vSwitch0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33554436	vmnic1		vSwitch0	2.13	0.01	314.00	9.50	0.01	176.00	0.00	0.00
33554437	Shadow of vmnic1	n/a	vSwitch0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33554438	vmk0	vmnic1	vSwitch0	2.13	0.01	314.00	1.94	0.00	60.00	0.00	0.00
33554439	304522:linux02	vmnic1	vSwitch0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50331649	Management	n/a	vSwitch1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50331650	vmnic2	-	vSwitch1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50331651	Shadow of vmnic2	n/a	vSwitch1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50331652	vmk1	vmnic2	vSwitch1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
67108865	Management	n/a	vSwitch2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
67108866	vmnic3	-	vSwitch2	3.10	0.02	678.00	15.90	0.06	492.00	0.00	0.00
67108867	Shadow of vmnic3	n/a	vSwitch2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
67108868	vmk2	vmnic3	vSwitch2	3.10	0.02	678.00	4.46	0.01	359.00	0.00	0.00

The key columns to observe are %DRPTX and %DRPRX.

VMware Confidential  
Internal Use Only

## esxtop Command: x Option

Slide 7-48

The **x** option for `esxtop` displays information related to Virtual SAN.

```
8:34:50pm up 4 days 6:22, 546 worlds, 0 VMs, 0 vCPUs; CPU load average: 0.03, 0.02, 0.02  
VSAN Enabled? Y
```

ROLE	READS/s	MBREAD/s	AVGLAT	SDLAT	WRITES/s	MBWRITE/s	AVGLAT	SDLAT	RECOWR/s	MBRECOWR/s	AVGLAT
Client	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Owner	0.4	0.0	1.8	0.1	0.4	0.0	5.0	0.4	0.0	0.0	0.0
CompMgr	0.0	0.0	0.0	0.0	0.4	0.0	3.9	0.1	0.0	0.0	0.0

The key columns to observe are RECOWR/s and MBRECOWR/s.

By default, the RECOWR/s and MBRECOWR/s columns are not displayed unless the **D** option is added.

## esxcli Command

Slide 7-49

The esxcli command displays disk-related statistics.

```
[root@vsan-node4:~] esxcli storage core device stats get | more
mpx.vmhba1:C0:T2:L0
Device: mpx.vmhba1:C0:T2:L0
Successful Commands: 1465
Blocks Read: 7513
Blocks Written: 0
Read Operations: 952
Write Operations: 0
Reserve Operations: 0
Reservation Conflicts: 0
Failed Commands: 196
Failed Blocks Read: 0
Failed Blocks Written: 0
Failed Read Operations: 0
Failed Write Operations: 0
Failed Reserve Operations: 0

mpx.vmhba1:C0:T1:L0
Device: mpx.vmhba1:C0:T1:L0
Successful Commands: 13309
Blocks Read: 100343
Blocks Written: 2296
Read Operations: 12602
Write Operations: 216
Reserve Operations: 0
Reservation Conflicts: 0
Failed Commands: 195
Failed Blocks Read: 0
Failed Blocks Written: 0
Failed Read Operations: 0
Failed Write Operations: 0
Failed Reserve Operations: 0
```

VMware Confidential  
Internal Use Only

## Ruby vSphere Console

Slide 7-50

RVC is a Linux console UI for vSphere and vCenter, used for managing and troubleshooting Virtual SAN environments.

RVC is available both on Windows vCenter Server systems and vCenter Server Appliance:

- Windows vCenter Server systems run RVC by using the **RVC.bat** file.
- RVC is built in to vCenter Server Appliance as a command shell.



Ruby vSphere Console (RVC) is a command-line utility that is used to navigate and run commands against managed entities by using an SSH connection to the vCenter Server. Ruby vSphere Console is bundled with both vCenter Server Appliance and the Windows version of vCenter Server.

# Log In to the Ruby vSphere Console

Slide 7-51

1. Connect to vCenter Server Appliance and log in as root.

```
10.1.1.15 - PuTTY
login as: root
VMware vCenter Server Appliance 6.0.0.20000
Type: vCenter Server with an embedded Platform Services Controller
root@10.1.1.15's password:
Last login: Fri Mar 11 00:06:46 UTC 2016 from 10.1.1.100 on pts/0
Last login: Fri Mar 11 00:10:43 2016 from 10.1.1.100
Connected to service

* List APIs: "help api list"
* List Plugins: "help pi list"
* Enable BASH access: "shell.set --enabled True"
* Launch BASH: "shell"

Command> shell.set --enabled True
Command> shell
```

2. Run the `rvc` command and log in as administrator.

```
10.1.1.15 - PuTTY
vsan-vcenter:~ # rvc administrator@vsphere.local@localhost
Install the "ffi" gem for better tab completion.
WARNING: Nokogiri was built against LibXML version 2.7.6, but has dynamically
aded 2.9.2
password:
0 /
1 localhost/
> [REDACTED]
```

# Navigating the vSphere and Virtual SAN Infrastructure

Slide 7-52

RVC includes commands, such as `ls` and `cd`, to navigate the vSphere infrastructure hierarchy.

```
> ls
0 /
1 localhost/
> cd localhost
/localhost> ls
0 Miami Datacenter (datacenter)
/localhost> cd "Miami Datacenter"
/localhost/Miami Datacenter> ls
0 storage/
1 computers [host]/
2 networks [network]/
3 datastores [datastore]/
4 vms [vm]/
/localhost/Miami Datacenter> cd computers
/localhost/Miami Datacenter/computers> ls
0 Miami Cluster (cluster): cpu 15 GHz, memory 9 GB
/localhost/Miami Datacenter/computers> cd 0
/localhost/Miami Datacenter/computers/Miami Cluster> ls
0 hosts/
1 resourcePool [Resources]: cpu 15.51/15.51/normal, mem 9.18/9.18/normal
/localhost/Miami Datacenter/computers/Miami Cluster> cd 0
/localhost/Miami Datacenter/computers/Miami Cluster/hosts> ls
0 10.1.1.11 (host): cpu 1*2*2.39 GHz, memory 8.00 GB
1 10.1.1.12 (host): cpu 2*1*2.39 GHz, memory 8.00 GB
2 10.1.1.13 (host): cpu 1*2*2.39 GHz, memory 8.00 GB
3 10.1.1.14 (host): cpu 2*2*2.39 GHz, memory 8.00 GB
```

Listing all hosts participating in a Virtual SAN cluster allows an administrator to run commands against individual participating hosts including disk group management and Virtual SAN participation. Each host is numerically associated for easier navigation.

# Using RVC to Get Help

Slide 7-53

Enter **help vsan** to get a list of all available RVC commands related to Virtual SAN administration and management.

```
> help vsan
Namespaces:
  health
  perf
  sizing
  stretchedcluster
  vsanmgmt

Commands:
  apply_license_to_cluster: Apply license to VSAN
  check_limits: Gathers (and checks) counters against limits
  check_state: Checks state of VMs and VSAN objects
  clear_disks_cache: Clear cached disks information
  cluster_change_autoclaim: Enable/Disable autoclaim on a VSAN cluster
  cluster_change_checksum: Enable/Disable VSAN checksum enforcement on a cluster
  cluster_info: Print VSAN config info about a cluster or hosts
  cluster_set_default_policy: Set default policy on a cluster
  cmmds_find: CMMDS Find
  disable_vsan_on_cluster: Disable VSAN on a cluster
  disk_object_info: Fetch information about all VSAN objects on a given physical disk
  disks_info: Print physical disk info about a host
  disks_stats: Show stats on all disks in VSAN
  enable_vsan_on_cluster: Enable VSAN on a cluster
  enter_maintenance_mode: Put hosts into maintenance mode
  Choices for vsan-mode: ensureObjectAccessibility, evacuateAllData, noAction
```

The **help** command displays each of the available commands with a brief description of that command.

You can use the **help** command in the following ways:

- **help command\_name**: Gives a description of the command and its syntax  
For example: `help vsan.vm_object_info`
- **help namespace\_name**: Lists the commands that are part of the namespace and a short description of each command  
For example: `help vsan.health`

# Using RVC to List Virtual SAN Commands

Slide 7-54

Enter **vsan.** and press the Tab key twice to list all the Virtual SAN RVC commands and namespaces that are available.

```
> vsan.  
vsan.apply_license_to_cluster      vsan.lldpnetmap  
vsan.check_limits                 vsan.obj_status_report  
vsan.check_state                  vsan.object_info  
vsan.clear_disks_cache           vsan.object_reconfigure  
vsan.cluster_change_autoclaim    vsan.observer  
vsan.cluster_change_checksum     vsan.observer_process_statsfile  
vsan.cluster_info                 vsan.perf.  
vsan.cluster_set_default_policy  vsan.proactive_rebalance  
vsan.cmds_find                   vsan.proactive_rebalance_info  
vsan.disable_vsan_on_cluster     vsan.purge_inaccessible_vsdp_objects  
vsan.disk_object_info            vsan.reapply_vsan_vmknic_config  
vsan.disks_info                  vsan.recover_spbm  
vsan.disks_stats                 vsan.resync_dashboard  
vsan.enable_vsan_on_cluster      vsan.scrubber_info  
vsan.enter_maintenance_mode    vsan.sizing.  
vsan.fix_renamed_vms            vsan.stretchedcluster.  
vsan.health.  
vsan.host_claim_disks_differently vsan.support_information  
vsan.host_consume_disks          vsan.v2_ondisk_upgrade  
vsan.host_evacuate_data          vsan.vm_object_info  
vsan.host_exit_evacuation        vsan.vm_perf_stats  
vsan.host_info                   vsan.vmdk_stats  
vsan.host_wipe_non_vsan_disk    vsan.vsanmgmt.  
vsan.host_wipe_vsan_disks       vsan.whatif_host_failures  
> █
```

The Virtual SAN RVC commands are useful for collecting general information regarding the health of a Virtual SAN cluster.

## Special Objects and Commands

Slide 7-55

RVC has some command objects that allow faster navigation:

- Enter ~1 to represent the first object of a listing (ls).
- Enter ~ to refer to the data center.
- Enter ~~ to always point to the previous directory.
- Enter **mark object\_path** to create a quick reference for an object.

```
/localhost/Miami Datacenter/computers> mark cluster ~/computers/"Miami Cluster"
/localhost/Miami Datacenter/computers> cd ~cluster
/localhost/Miami Datacenter/computers/Miami Cluster> █
```

When navigating in RVC, you can use the standard Linux command shortcuts like the arrow and Tab keys. RVC also includes some command objects that allow faster navigation.

The **mark** command sets a placeholder for a longer or cumbersome object.

# Viewing Virtual Machine Object Information

Slide 7-56

vSphere Web Client gives you component information for the following virtual machine objects:

- VM Home Namespace
- VMDKs

Name	1 VM Storage Policy	Compliance Status
VM home	Virtual SAN Default Storage Policy	Compliant
Hard disk 1	Virtual SAN Default Storage Policy	Compliant

RVC enables you to find component information for all other virtual machine objects:

- Snapshot deltas
- VM swap
- VM memory

vSphere Web Client is used for most of the tasks needed to monitor and administer Virtual SAN. For example, when you select a virtual machine in the vCenter Server inventory, you can view the **Monitor > Policies** tab to view the component placement for VMDK objects and the VM Home Namespace object.

However, for the other objects included in the virtual machine, you must use RVC if you want more information about the objects' components.

# Displaying Snapshot Delta Object Information

Slide 7-57

The `vsan.vm_object_info` command displays specific information about the snapshot delta objects.

```
[localhost:SA-Datacenter/computers] vsan.vm_object_info ~vm
VM Win-A-03:
  Namespace directory
    DOM Object: bdb1cc56-e0e4-f845-f800-00505601afc3 (v4, owner: sa-esxi-01.vclass.local, policy: forceProvisioning = 0, hostFailuresToTolerate = 1, spbmProfileId = aa6d5a82-1c88-45da-85d3-3d74b91a5bad, proportionalCapacity = [0, 100], spbmProfileGenerationNumber = 0, cacheReservation = 0, stripeWidth = 1)
      RAID_1
        Component: bdb1cc56-54a3-fb46-585e-00505601afc3 (state: ACTIVE (5), host: sa-esxi-02.vclass.local, md: mpx.vmhba1:CO:T1:L0, ssd: mpx.vmhba1:CO:T4:L0,
          votes: 1, usage: 0.4 GB)
        Component: bdb1cc56-d4aa-fd46-63ec-00505601afc3 (state: ACTIVE (5), host: sa-esxi-05.vclass.local, md: mpx.vmhba1:CO:T3:L0, ssd: mpx.vmhba1:CO:T1:L0,
          votes: 1, usage: 0.4 GB)
        Witness: bdb1cc56-60ed-fe46-a85a-00505601afc3 (state: ACTIVE (5), host: sa-esxi-06.vclass.local, md: mpx.vmhba1:CO:T3:L0, ssd: mpx.vmhba1:CO:T1:L0,
          votes: 1, usage: 0.0 GB)
      Disk Backing: [vsanDatastore] bdb1cc56-e0e4-f845-f800-00505601afc3/Win-A-03-000001.vmdk
        DOM Object: 30dbcc56-972d-65e1-2e4c-00505601a712 (v4, owner: sa-esxi-01.vclass.local, policy: forceProvisioning = 0, hostFailuresToTolerate = 1, spbmProfileId = aa6d5a82-1c88-45da-85d3-3d74b91a5bad, proportionalCapacity = [0, 100], spbmProfileGenerationNumber = 0, cacheReservation = 0, stripeWidth = 1)
          RAID_1
            Component: 30dbcc56-e2c3-30ed-18cc-00505601a712 (state: ACTIVE (5), host: sa-esxi-04.vclass.local, md: mpx.vmhba1:CO:T2:L0, ssd: mpx.vmhba1:CO:T1:L0,
              votes: 1, usage: 3.6 GB)
            Component: 30dbcc56-a8ef-33ed-cc6c-00505601a712 (state: ACTIVE (5), host: sa-esxi-02.vclass.local, md: mpx.vmhba1:CO:T2:L0, ssd: mpx.vmhba1:CO:T4:L0,
              votes: 1, usage: 3.6 GB)
            Witness: 30dbcc56-53bf-35ed-a2e7-00505601a712 (state: ACTIVE (5), host: sa-esxi-01.vclass.local, md: mpx.vmhba1:CO:T2:L0, ssd: mpx.vmhba1:CO:T1:L0,
              votes: 1, usage: 0.0 GB)
      Disk Backing: [vsanDatastore] bdb1cc56-e0e4-f845-f800-00505601afc3/Win-A-03.vmdk
        DOM Object: cab1cc56-44d0-803b-0204-00505601afc3 (v4, owner: sa-esxi-01.vclass.local, policy: forceProvisioning = 0, hostFailuresToTolerate = 1, spbmProfileId = aa6d5a82-1c88-45da-85d3-3d74b91a5bad, proportionalCapacity = [0, 100], spbmProfileGenerationNumber = 0, cacheReservation = 0, stripeWidth = 1)
```

The `vsan.vm_object_info` command displays specific information about the VM Home Namespace, VMDK objects, and snapshot delta objects. With this command, you can view information about the object's components and witnesses, such as the hosts that they reside on and the number of votes that they have.

# Displaying Swap File Object Information

Slide 7-58

First, view the .vswp file to retrieve the object ID of the VM swap object.

Use vsan.object\_info to display information for that object ID.

```
[root@sa-esxi-01:/vmfs/volumes/vsan:522b5862ac4c988d-99006  
-f845-f800-00505601afc3] cat Win-A-03-441b2ed2.vswp  
# Object DescriptorFile  
version = "1"  
  
objectID = "vsan:/41d4cc56-6053-352e-2036-00505601a712"  
object.class = "vmswap"  
  
swapObj.lock = "/vmfs/volumes/vsan:522b5862ac4c988d-990060  
  
/localhost/SA-Datacenter/computers> ls  
0 SA-VSAN-01 (cluster): cpu 18 GHz, memory 10 GB  
/localhost/SA-Datacenter/computers> vsan.object.info 0 41d4cc56-6053-352e-2036-00505601a712  
DOM Object: 41d4cc56-6053-352e-2036-00505601a712 (v4, owner: sa-esxi-01.vclass.local, policy: hostFailuresToTolerate = 1, forceProvisioning = 1, proportionalCapacity = 100)  
    RAID_1  
        Component: 41d4cc56-18dd-783a-6277-00505601a712 (state: ACTIVE (5), host: sa-esxi-03.vclass.local  
, md: mpx.vmhba1:CO:T2:L0, ssd: mpx.vmhba1:CO:T1:L0,  
            votes: 1, usage: 2.0 GB)  
        Component: 41d4cc56-6c3b-7e3a-4c03-00505601a712 (state: ACTIVE (5), host: sa-esxi-04.vclass.local  
, md: mpx.vmhba1:CO:T2:L0, ssd: mpx.vmhba1:CO:T1:L0,  
            votes: 1, usage: 2.0 GB)  
    Witness: 41d4cc56-a09d-823a-c0b2-00505601a712 (state: ACTIVE (5), host: sa-esxi-01.vclass.local, md  
: mpx.vmhba1:CO:TS:L0, ssd: mpx.vmhba1:CO:T1:L0,  
            votes: 1, usage: 0.0 GB)  
    Extended attributes:  
        Address space: 2147483648B (2.00 GB)  
        Object class: vmswap  
        Object path: /vmfs/volumes/vsan:522b5862ac4c988d-99006079595f2d58/bdb1cc56-e0e4-f845-f800-0050560  
1afc3/Win-A-03-441b2ed2.vswp  
        Object capabilities: NONE  
  
/localhost/SA-Datacenter/computers>
```

## To display information about the VM swap object

- Get the object ID for the VM swap object.

You get this information from the .vswp file, which is an object descriptor file that contains the object ID of the VM swap object.

- Change to the computer's directory under your data center directory and run the `ls` command.

By changing the directory you can get the identifier for the Virtual SAN cluster. In the example, the cluster is identified as 0.

- Run the `vsan.object_info` command, where the first argument is the cluster identifier and the second argument is the VM swap object ID.

With this command, you can view information about the object's components and witnesses, such as the hosts that they reside on and the number of votes that they have. You can also see that `hostFailuresToTolerate` is 1, which means that VM swap object uses RAID 1. Also, `proportionalCapacity` is 100, which means that the space needed for the VM swap object is fully reserved.

## Viewing Host-Specific Information

Slide 7-59

The `vsan.host_info` command displays information about hosts participating in the Virtual SAN cluster.

```
/localhost/Miami Datacenter/computers/Miami Cluster/hosts> vsan.host_info 10.1.1.1
1
2016-03-11 00:30:05 +0000: Fetching host info from 10.1.1.11 (may take a moment) .
..
Product: VMware ESXi 6.0.0 build-3626945
VSAN enabled: yes
Cluster info:
  Cluster role: backup
  Cluster UUID: 52f273bd-ca0c-b285-a95b-81695681de82
  Node UUID: 56dec41f-5287-bea3-ba3c-000c29e60e5b
  Member UIDs: ["56dec61d-b125-352a-e807-000c29f7e5e8", "56dec41f-5287-bea3-ba3c-000c29e60e5b", "56dec7ea-6d6d-64d8-4456-000c295de7d4", "56dec9b4-bf13-4d44-826b-000c298bea43"] (4)
  Node evacuated: no
Storage info:
  Auto claim: no
  Disk Mappings:
    SSD: Local VMware Disk (mpx.vmhba1:C0:T1:L0) - 10 GB, v3
    MD: Local VMware Disk (mpx.vmhba1:C0:T2:L0) - 100 GB, v3
FaultDomainInfo:
  Not configured
NetworkInfo:
  Adapter: vmk2 (10.1.3.11)
```

You can get information about the cluster role, UUIDs of the cluster and members, disk mapping, and network adapters.

VMware Confidential  
Internal Use Only

# Viewing Host Disk Information

Slide 7-60

The `vsan.disks_info` command displays disk information for a specific host.

```
/localhost/Miami Datacenter/computers/Miami Cluster/hosts> vsan.disks.info 10.1.1.11
2016-03-11 00:32:41 +0000: Gathering disk information for host 10.1.1.11
2016-03-11 00:32:45 +0000: Done gathering disk information
Disks on host 10.1.1.11:
+-----+-----+-----+
| DisplayName | isSSD | Size | State
+-----+-----+-----+
| Local VMware Disk (mpx.vmhba1:C0:T2:L0) | SSD | 100 GB | inUse
| VMware Virtual disk | | | VSAN Format Version: v3
+-----+-----+-----+
| Local VMware Disk (mpx.vmhba1:C0:T1:L0) | SSD | 10 GB | inUse
| VMware Virtual disk | | | VSAN Format Version: v3
+-----+-----+-----+
```

The `vsan.disks_info` command displays disk-by-disk and per-host information. The formatting of the table wraps because of the limitation of the screen resolution. This command informs the administrator whether a disk is in use and is eligible for participation in a disk group. This command also provides reasons for the ineligibility of a disk, for example, an existing partition table or disk capacity is not sufficient for Virtual SAN.

## Viewing Host Resource Statistics

Slide 7-61

The `table` command displays resource statistics in an organized table:

- `table -f name -f state.connection -f num.vms -f cpuusage -f memusage *`

<code>name</code>	<code>state</code>	<code>num</code>	<code>cpuusage</code>	<code>memusage</code>
	<code>connection</code>	<code>vms</code>		
10.1.1.11	connected	0	0.00 %	0.00 %
10.1.1.12	connected	1	0.00 %	0.00 %
10.1.1.13	connected	1	5.91 %	37.73 %
10.1.1.14	connected	1	27.29 %	39.57 %

The `table` command can be used to see a view that is similar to information that is available through vSphere Web Client. This command shows all hosts by name and includes the following details:

- State of the connection
- Number of virtual machines running
- CPU usage
- Memory usage

# Viewing Data Center Disk Statistics

Slide 7-62

The `vsan.disks_stats Object_Number` command displays site-wide disk statistics.

```
/localhost/Miami Datacenter/computers> vsan.disks_stats 0
2016-03-11 00:39:42 +0000: Fetching VSAN disk info from 10.1.1.13 (may take a moment) ...
2016-03-11 00:39:42 +0000: Fetching VSAN disk info from 10.1.1.14 (may take a moment) ...
2016-03-11 00:39:42 +0000: Fetching VSAN disk info from 10.1.1.12 (may take a moment) ...
2016-03-11 00:39:42 +0000: Fetching VSAN disk info from 10.1.1.11 (may take a moment) ...
2016-03-11 00:39:48 +0000: Done fetching VSAN disk infos

+-----+-----+-----+-----+-----+-----+-----+
| | | Num | Capacity | | | Status
| | | Host | isSSD | Comp | Total | Used | Reserved | Health
+-----+-----+-----+-----+-----+-----+-----+
| mpx.vmhba1:CO:T1:L0 | 10.1.1.11 | SSD | 0 | 10.00 GB | 0.00 % | 0.00 % | OK (v3)
|
| mpx.vmhba1:CO:T2:L0 | 10.1.1.11 | MD | 7 | 98.99 GB | 0.92 % | 0.16 % | OK (v3)
|
+-----+-----+-----+-----+-----+-----+-----+
| mpx.vmhba1:CO:T1:L0 | 10.1.1.12 | SSD | 0 | 10.00 GB | 0.00 % | 0.00 % | OK (v3)
|
| mpx.vmhba1:CO:T2:L0 | 10.1.1.12 | MD | 7 | 98.99 GB | 1.24 % | 0.16 % | OK (v3)
|
+-----+-----+-----+-----+-----+-----+-----+
| mpx.vmhba1:CO:T1:L0 | 10.1.1.13 | SSD | 0 | 10.00 GB | 0.00 % | 0.00 % | OK (v3)
|
| mpx.vmhba1:CO:T2:L0 | 10.1.1.13 | MD | 8 | 98.99 GB | 1.41 % | 0.67 % | OK (v3)
|
+-----+-----+-----+-----+-----+-----+-----+
```

The output provides details about the following:

- Number of components per disk
- Disk capacity
- Percentage used and percentage reserved for hard-disk drive and solid-state drive
- Performance metrics including device latencies

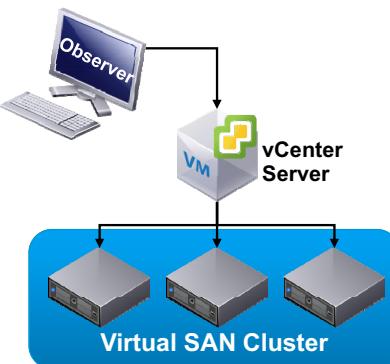
# Activating Virtual SAN Observer

Slide 7-63

Virtual SAN Observer is a browser-based tool for viewing deep-level performance charts.

Virtual SAN Observer is a part of RVC and is accessed with a web browser:

- Supports the same list of browsers as vSphere Web Client
- Viewed from a connection to a production vCenter Server that manages a Virtual SAN cluster



Virtual SAN Observer is a monitoring and troubleshooting tool for Virtual SAN. The tool is launched from RVC and can be utilized for monitoring performance statistics for Virtual SAN live mode or offline. When running in live mode, a web browser can be pointed at vCenter Server to see live graphs related to the performance of Virtual SAN.

Virtual SAN Observer either outputs to log files or uses a built-in Web server that is accessed using an HTTP connection. The RVC console creates a connection to the vCenter Server system regardless of where the command is issued.

Virtual SAN Observer runs from memory when adjusting runtime settings. By default, Virtual SAN Observer polls for information every 60 seconds and runs for 2 hours. You can use Virtual SAN Observer from another vCenter Server system and connect to the vCenter Server system that is managing the Virtual SAN cluster. Thus, you can prevent aggressive timings from overloading a production vCenter Server.

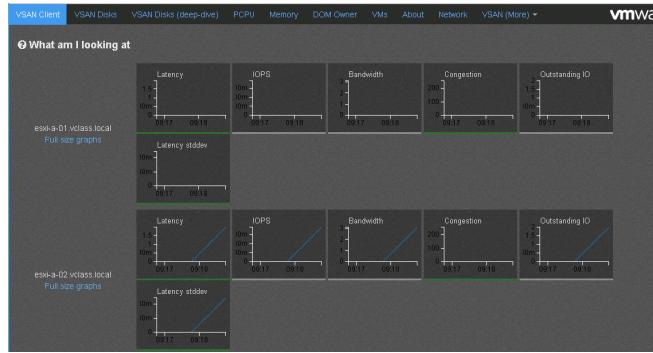
# Virtual SAN Observer Web Server

Slide 7-64

Virtual SAN Observer is launched from the vCenter Server system by using RVC:

- cd /localhost/*data\_center\_name*
- mark cluster ~/computers/"*cluster\_name*"
- vsan.observer ~cluster --run-webserver --force

To access Virtual SAN Observer, go to [https://vcenter\\_server\\_ip:8010/](https://vcenter_server_ip:8010/).

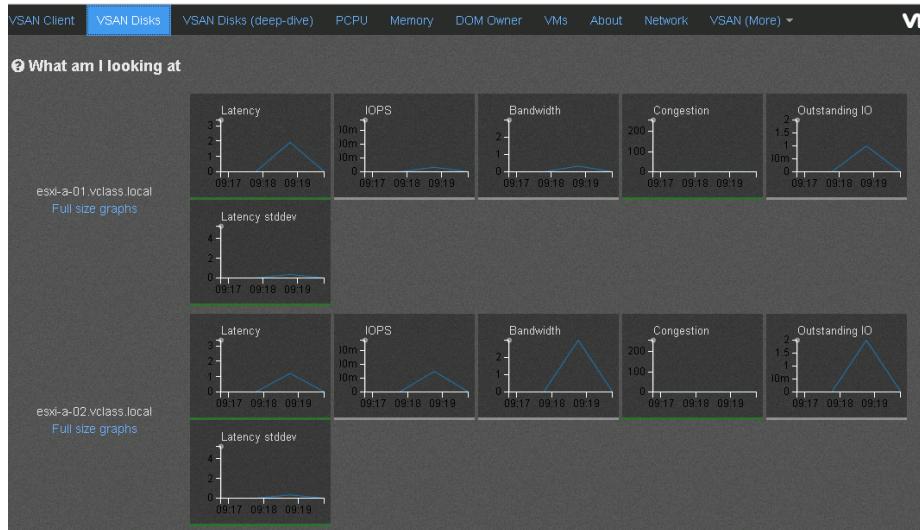


The **VSAN Client** tab provides an overview of the level of service that virtual machines are currently getting from Virtual SAN. Every host in a Virtual SAN cluster consumes storage that is distributed across all other hosts in the cluster. For example, a performance issue on esxi-a-01.vclass.local might be due to the overloaded disks on esxi-a-02.vclass.local.

## VSAN Disks Tab

Slide 7-65

The **VSAN Disks** tab displays the Virtual SAN physical disk layer.

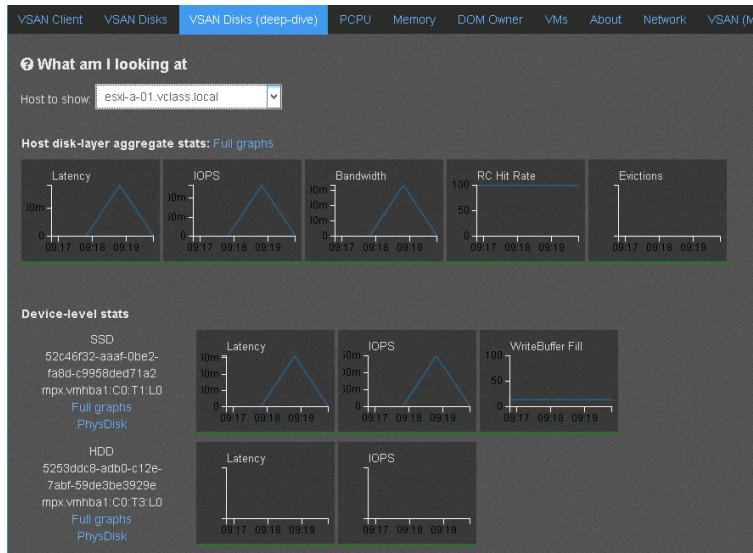


The **VSAN Disks** tab provides the ability to look at Virtual SAN from the physical disk layer and determine how storage contributing nodes are performing while servicing I/O from their local disks.

## VSAN Disks (Deep-Dive) Tab

Slide 7-66

On the **VSAN Disks (deep-dive)** tab, you can navigate into the Virtual SAN disks layer per host.

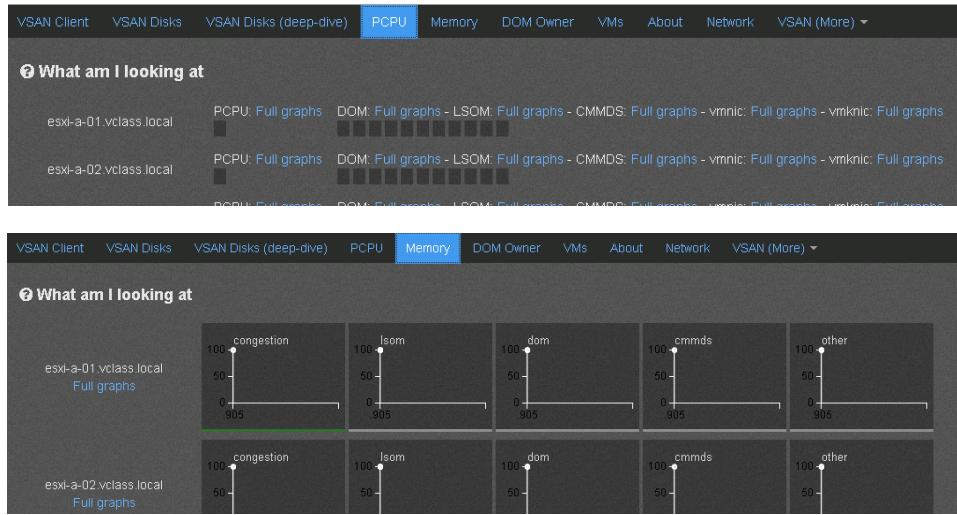


The **VSAN Disks (deep-dive)** tab navigates into the Virtual SAN disks layer per host and provides information to each individual cache and capacity disk per host. This layer does the I/O to cache and capacity disks. This tab provides insight into how Virtual SAN is splitting the I/O work between the cache and capacity disks.

# PCPU and Memory Tabs

Slide 7-67

CPU and memory consumption have dedicated tabs in the Virtual SAN Observer.



The **PCPU** tab shows the overall and per-component CPU usage statistics of Virtual SAN. This view shows CPU usage from an overall host perspective and from the view of individual Virtual SAN and networking processes.

The **Memory** tab displays consumption of various Virtual SAN memory pools. The pools tracked under congestion directly affect performance if they are above 75 percent utilization. A high utilization triggers the congestion mechanism of Virtual SAN, which imposes I/O delays at the Virtual SAN client.

This tab is critical because you can see the cache hit rate. If the cache hit rate is too low, consider reserving flash capacity or striping to critical virtual machines.

## VMs Tab

Slide 7-68

The **VMs** tab displays performance per virtual machine or per virtual disk.



The **VMs** tab also provides visibility into the per-disk policy settings applied when using virtual machine storage profiles.

## Lab 14: Using Ruby vSphere Console and ESXi Commands

Slide 7-69

Use RVC, Virtual SAN Observer, and ESXi commands to monitor Virtual SAN

1. (Optional) Prepare the Environment
2. Determine the Target Host
3. Create a Virtual Machine Snapshot
4. Log In to Ruby vSphere Console
5. Configure RVC Marks
6. Use vsan Commands to Monitor Virtual SAN
7. Access the Virtual SAN Observer Web Site and Monitor Performance Metrics
8. Use the esxtop Command to Monitor Virtual SAN

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 7-70

You should be able to meet the following objectives:

- Use ESXi commands to monitor the Virtual SAN environment
- Use RVC to get detailed information about the Virtual SAN environment
- Use Virtual SAN Observer to view Virtual SAN performance

VMware Confidential  
Internal Use Only

## Key Points

Slide 7-71

- vSphere Web Client is the main tool for monitoring the health and performance of your Virtual SAN cluster.
- The performance service monitors the performance of Virtual SAN clusters, hosts, virtual machines, and virtual disks.
- The health service actively tests and monitors the Virtual SAN environment.
- vRealize Operations Manager, with the Management Pack for Storage Devices, provides dashboards and alerts that give you visibility into your Virtual SAN 6.x environment.
- RVC can be used to view information about your Virtual SAN environment, which you cannot get in vSphere Web Client.

Questions?

VMware Confidential  
Internal Use Only

# Stretched Clusters and Two-Node Clusters

Slide 8-1

Module 8

VMware Confidential  
Internal Use Only

# You Are Here

Slide 8-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
- 8. Stretched Clusters and Two-Node Clusters**
9. Interoperability with vSphere Features
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 8-3

A stretched cluster is a solution that is implemented whenever disaster avoidance or a swift disaster recovery is important.

A two-node cluster gives you the benefits provided around manageability, performance, and availability in a traditional Virtual SAN cluster implementation, without the minimum requirement of three nodes.

VMware Confidential  
Internal Use Only

## Module Lessons

Slide 8-4

Lesson 1: Stretched Clusters and Two-Node Clusters

Lesson 2: Stretched Cluster Failure Scenarios

VMware Confidential  
Internal Use Only

## Stretched Clusters and Two-Node Clusters

Slide 8-5

### Lesson 1: Stretched Clusters and Two-Node Clusters

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 8-6

By the end of this lesson, you should be able to meet the following objectives:

- Describe the architecture for stretched clusters and two-node clusters
- Create a stretched cluster

VMware Confidential  
Internal Use Only

# About Virtual SAN Stretched Cluster

Slide 8-7

Introduced in Virtual SAN 6.1, stretched clusters can be used in environments where disaster and downtime avoidance is a key requirement:

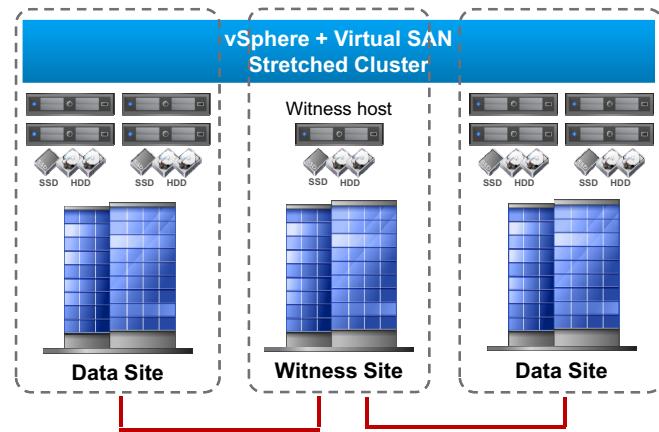
- Protects virtual machines across data centers, not just racks

A stretched cluster extends across three sites:

- Two active-active data sites
- One witness site

The witness site contains the witness host:

- Provides cluster quorum during a site failure



A Virtual SAN stretched cluster refers to a deployment where you set up a Virtual SAN cluster with two active-active data sites. An identical number of hosts is distributed evenly between the two sites. The sites are connected with a high bandwidth, low latency link.

The third site hosts the Virtual SAN witness host, which is connected to both of the active-active data sites. The sole purpose of the witness host is to provide cluster quorum during data site failure events. A special witness appliance is available for use as the witness host. Connectivity between the witness site and each of the data sites can be a low bandwidth, high latency link.

The hosts should contain 50 percent of the maximum virtual machines supported in a standard Virtual SAN cluster. If a site failure occurs, the virtual machines can receive the same level of service on the surviving site.

# Stretched Cluster Use Cases

Slide 8-8

A stretched cluster has the following use cases.

Planned Maintenance	Disaster Avoidance	Automated Recovery
<ul style="list-style-type: none"><li>Planned maintenance of one site without any service downtime</li><li>Transparent to application owners and end users</li><li>Ability to migrate applications back after maintenance is complete</li></ul>	<ul style="list-style-type: none"><li>Prevent service outages before an impending disaster (such as hurricane, rising flood levels)</li><li>Avoid downtime, not recover from it</li></ul>	<ul style="list-style-type: none"><li>Automated initiation of virtual machine restart or recovery</li><li>Very low RTO for majority of unplanned failures</li><li>Allows users to focus on application health after recovery, not how to recover virtual machines</li></ul>

A stretched cluster can be used for planned and unplanned downtime of data sites. For planned downtime, you can migrate the virtual machines from one data site to the other data site. You can then perform maintenance on the evacuated data site without causing any virtual machine downtime.

Also, if one data site might be affected by an impending disaster, such as a hurricane, virtual machines running at this site can be migrated to the other data site and can continue to run.

If a failure occurs at one of the data sites, the stretched cluster can automatically restart virtual machines on the other data site.

# Stretched Cluster Architecture

Slide 8-9

A stretched cluster is composed of only three sites:

- Two active data sites
- One witness site

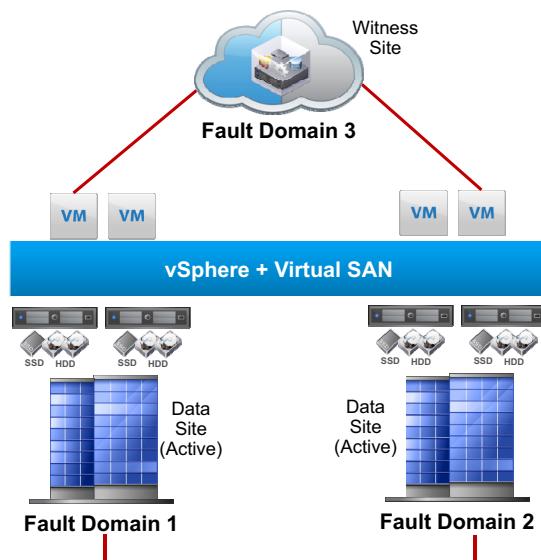
Each site is its own fault domain:

- Supports only FTT = 1

Across the fault domains, the stretched cluster can have a minimum of three hosts and a maximum of 31.

A stretched cluster does not support these features:

- vSphere Fault Tolerance
- RAID 5/6 Erasure Coding



Each site in the stretched cluster is configured as a fault domain. Fault domains are used to spread redundancy components across servers. In a traditional Virtual SAN cluster, redundant components are spread across servers in separate computing racks, and as a result, can tolerate rack failures, cache and capacity device failures, network device failures, or power failures.

When used in a stretched cluster, fault domains spread redundancy components across sites, and therefore can tolerate the failure of an entire data site.

Because a stretched cluster has three fault domains, the number of failures to tolerate (FTT) policy setting has a maximum value of 1. Virtual SAN cannot comply with FTT values that are greater than 1 in a stretched cluster configuration. Also, because a stretched cluster cannot have more than three fault domains, RAID 5/6 is not supported because this failure tolerance method requires more than three fault domains.

The minimum number of hosts in a stretched cluster is three: one host in each data site plus the witness host in the witness site.

The maximum number of hosts in a stretched cluster is 31: Fifteen hosts in each data site plus the witness host in the witness site.

# Single-Site Versus Stretched Clusters

Slide 8-10

	Single-Site Fault Domains	Stretched Cluster Fault Domains
Main use case	Protection against rack and chassis failures	Protection against site failures
Number of fault domains	3 to 32	3
Node configuration	All fault domains must have the same resources in terms of memory, CPU, and storage	Nodes must be evenly distributed between the two data sites.
Storage and CPU	All fault domains contribute to storage and CPU.	Only the data sites contribute to storage and CPU. The witness does not run virtual machines.
Network configuration	All fault domains must satisfy very low latency and high bandwidth requirements (< 5 msec and 10 Gbps).	The data sites should have very low latency (<= 5 msec) and high bandwidth. Witness requirements are much lower.

The table compares the characteristics of fault domains in a single-site Virtual SAN cluster and fault domains in a stretched cluster, which spans three sites.

## About the Witness Host

Slide 8-11

The witness host is located at the witness site and is used to store witness components for virtual machine objects.

VMware provides a Virtual SAN witness appliance that is downloadable from the VMware Web site:

- Is an ESXi instance
- Does not run virtual machines
- Requires less capacity and bandwidth than regular ESXi hosts
- Is packaged with its own license

Each stretched cluster must have its own witness host.

A physical ESXi host can also serve as the witness host:

- Requires a vSphere license

The Virtual SAN witness virtual appliance (referred to as witness appliance or witness host) is uniquely designed with the sole purpose of providing cluster quorum services during failure events and to store witness objects and cluster metadata information. The witness appliance does not contribute to compute and storage capacities.

The use of a virtual appliance as a witness eliminates the need to deploy a third physical server, which reduces the overall cost of the solution without sacrificing the benefits of shared storage.

The witness appliance is optimized to receive minimal amount of traffic when compared to scenarios with a traditional Virtual SAN cluster. At a steady state, there is barely any communication between the two data sites and the witness. Read and write operations do not require any communication to the witness appliance because the traffic to the witness appliance is mostly limited to create, delete, reconfigure, and change policy operations.

The witness appliance is available as an Open Virtual Appliance (OVA) from VMware. The witness appliance must reside on a physical ESXi host and requires a special networking configuration. This appliance comes with its own license. You do not need to consume vSphere licenses for the witness appliance.

The witness appliance never runs virtual machines in a stretched cluster configuration.

# Sizing the Witness Host

Slide 8-12

Size the witness host based on the maximum number of witness components that it will hold.

Witness Size	Max VMs	Max Components	RAM	vCPUs	Storage
Tiny	<= 10	750	8 GB	2	ESXi boot disk: 8 GB Cache tier: 10 GB Capacity tier: one 15 GB disk
Normal	<= 500	22,000	16 GB	2	ESXi boot disk: 8 GB Cache tier: 10 GB Capacity tier: one 350 GB disk
Large	> 500	45,000	32 GB	2	ESXi boot disk: 8 GB Cache tier: 10 GB Capacity tier: three 350 GB disks

If you use the witness appliance, the size is dependent on the configuration type (tiny, normal, or large). The configuration type is determined during the deployment process.

The purpose of the witness host is to store witness components for virtual machine objects. Because a single magnetic disk supports approximately 21,000 components, and the maximum components supported on the witness host are 45,000, a minimum of three magnetic disks are required on the witness host to support the maximum number of components.

If you use a physical server as the witness host, the minimum ESXi host requirements meet the needs of a witness host.

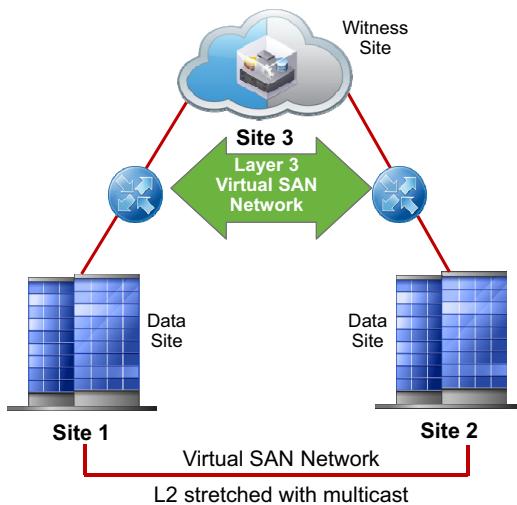
If you use a physical ESXi host, a single physical disk can support a maximum of 21,000 components. Each witness component in a Virtual SAN stretch cluster requires 16 MB of storage. Therefore, to support 21,000 components on a magnetic disk, VMware recommends a disk of approximately 350 GB in size.

# Networking Requirements

Slide 8-13

A stretched cluster has the following network requirements:

- Management connectivity to all 3 sites:
  - L2 stretched or L3 (routed)
- Virtual machine network connectivity between the data sites
  - VMware recommends L2 stretched
- VMware vSphere® vMotion® network connectivity between the data sites
  - L2 stretched or L3 (routed)
- Virtual SAN network connectivity to all 3 sites
  - VMware recommends L2 stretched between the two data sites, and L3 (routed) between the data sites and the witness site



Both Layer 2 (same subnet) and Layer 3 (routed) configurations are used in a recommended stretched cluster deployment:

- VMware recommends that Virtual SAN communication between data sites be over stretched L2. Data site to data site network refers to the communication between nonwitness sites, that is, sites that run virtual machines and hold virtual machine data.
- VMware recommends that Virtual SAN communication between the data sites and the witness site is over L3.

Virtual SAN traffic between data sites is multicast. Witness traffic between a data site and the witness site is unicast.

# Network Latency and Bandwidth Recommendations

Slide 8-14

Virtual SAN supports sites up to 100 km apart, if network requirements are met.

Network	Traffic Type	Latency (or RTT) Recommendations	Bandwidth Recommendations
Data site to data site	Layer 2, multicast	<= 5 msec	Workload dependent: Minimum of 10 Gbps for most workloads
Data site to witness site	Layer 3, unicast	Stretched cluster: <=200 msec (for <= 10 + 10 + 1);  <=100 msec (for >=10 + 10 + 1)	Rule of thumb: 2 Mbps for every 1000 objects

For stretched clusters, geographical distances are not a support concern. The key requirement is the actual latency numbers between sites.

For data site to data site communication, latency or round trip time (RTT) between sites hosting virtual machine objects should not be greater than 5 msec (< 2.5 msec one-way).

Bandwidth between data sites is workload dependent. For most workloads, VMware recommends a minimum of 10 Gbps or greater bandwidth between sites. In use cases such as two-node configurations for Remote Office/Branch Office (ROBO) deployments, dedicated 1 Gbps bandwidth will be sufficient with less than 10 virtual machines.

The latency from the data sites to the witness site depends on the number of objects in the cluster. VMware recommends that on stretched cluster configurations up to 10+10+1, a latency of less than or equal to 200 milliseconds is acceptable. For configurations that are greater than 10+10+1, VMware recommends a latency of less than or equal to 100 milliseconds.

In typical two-node configurations, such as ROBO deployments, the latency or RTT is supported up to 500 milliseconds.

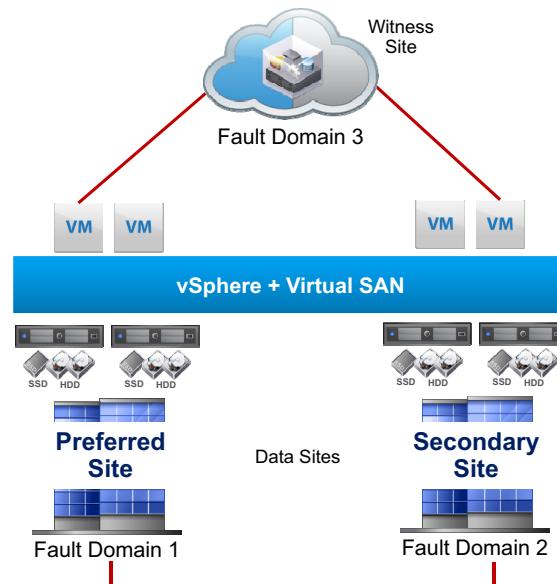
Bandwidth between the data sites and the witness site is dependent on the number of objects residing on Virtual SAN. Size the bandwidth appropriately for both availability and growth.

## About the Preferred Site

Slide 8-15

The preferred site, or preferred fault domain, is the data site that remains active when a network partition occurs between the two data sites:

- The other site is known as the secondary or nonpreferred site.
- The witness site can still communicate with both data sites.



When you create a stretched cluster, you must designate one site as the preferred site. The other site becomes the secondary or nonpreferred site. The preferred site is used for the case where a network partition has occurred between the two data sites, yet the witness site can still communicate with both data sites.

If the network connection fails between the two data sites, the witness host and the preferred site continue to service storage operations, and keep data available. When the network connection returns, the two active sites are resynchronized.

If the preferred site becomes isolated from both the secondary site and the witness, the witness host forms a cluster using the secondary site. When the preferred site is online again, data is resynchronized to ensure that both sites have the latest copies of all data.

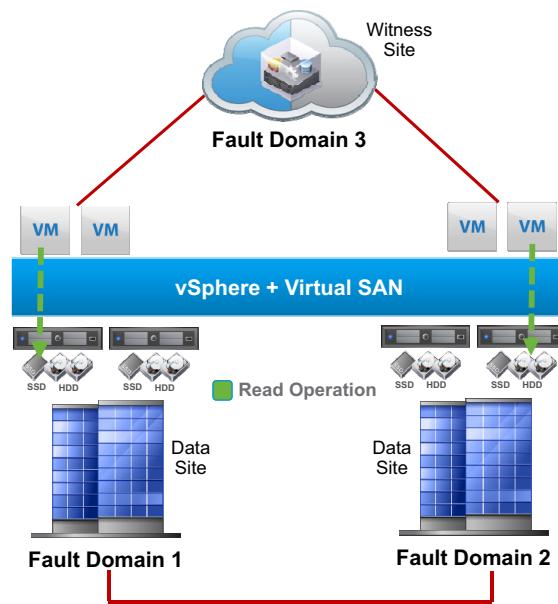
# Read Locality

Slide 8-16

Virtual SAN uses a read locality algorithm to read 100 percent from the data copy on the local site:

- The local site is the same site where the compute resides.

Read locality reduces the latency incurred on read operations.



In traditional Virtual SAN clusters, a virtual machine's read operations are distributed across all replica copies of the data in the cluster. For example, the policy setting of `NumberOfFailuresToTolerate = 1` results in two copies of the data. Therefore, 50 percent of the reads come from the first replica and 50 percent of the reads come from the second replica.

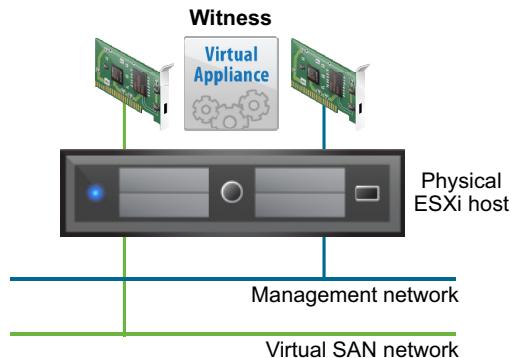
In a stretched cluster, Virtual SAN avoids increased latency caused by reading across the intersite link. Virtual SAN uses a read locality mechanism to ensure that 100 percent of reads occur at the site in which the virtual machine resides. If the virtual machine moves to the other site (for example, by vSphere HA failover, or vSphere vMotion, or a power off/power on cycle), then reads are served from the (consistent) copy of the data at the new site.

# Deploy the Witness Host

Slide 8-17

To deploy the Virtual SAN witness appliance:

1. Download the OVA template from the VMware Web site.
2. Deploy the OVA template to an ESXi host:
  - Choose the size (tiny, normal, large) of the deployment.
3. Configure the management network.
4. Add the witness appliance to the vCenter Server inventory.
5. Configure the Virtual SAN network.
6. Implement static routes between the data sites and the witness site.



When you deploy the witness appliance from the OVA template, have the following information handy:

- Witness name (your choice)
- Size of deployment (tiny, normal, or large)
- Datastore on which to deploy the appliance
- Management network name
- Root password

The witness appliance contains two network adapters. One network adapter connects to the management network and the other network adapter connects to the Virtual SAN network. Therefore, you must have two virtual machine networks created on the physical ESXi host so that the witness appliance can communicate with the management network and the Virtual SAN network.

Implement static routes on the ESXi hosts and the witness hosts to ensure that the hosts residing in the data sites can reach the witness host's Virtual SAN network and vice versa. L3 is required to reach the Virtual SAN network of the witness appliance. ESXi hosts have a single default TCP/IP

stack, and thus a single default gateway. As a result, no route exists to the Virtual SAN networks from these hosts. Static routes are not required for the data hosts on different sites to communicate with each other over the Virtual SAN network. The Virtual SAN network is a stretched L2 broadcast domain between the data sites as per VMware recommendations.

For detailed steps on deploying the witness appliance, see *Virtual SAN 6.1 Stretched Cluster Guide* at <http://www.vmware.com/files/pdf/products/vsan/VMware-Virtual-SAN-6.1-Stretched-Cluster-Guide.pdf>.

VMware Confidential  
Internal Use Only

## Create a Stretched Cluster

Slide 8-18

To deploy the Virtual SAN stretched cluster:

1. Create a Virtual SAN cluster.
2. Configure the stretched cluster.
  - a. Assign hosts to the preferred and secondary sites (fault domains).
  - b. Select a witness host and disk group.
3. Configure DRS affinity groups and rules.

VMware Confidential  
Internal Use Only

# Creating a Virtual SAN Cluster

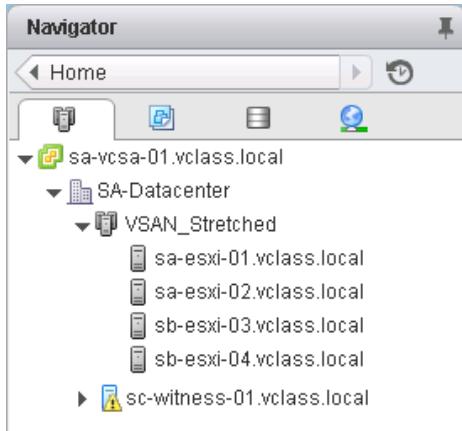
Slide 8-19

Create a Virtual SAN cluster using the normal procedure.

In this example, a Virtual SAN cluster named VSAN\_Stretched is created with four ESXi hosts:

- Two ESXi hosts from site A
- Two ESXi hosts from site B

The witness host is in site C and is not part of the cluster.



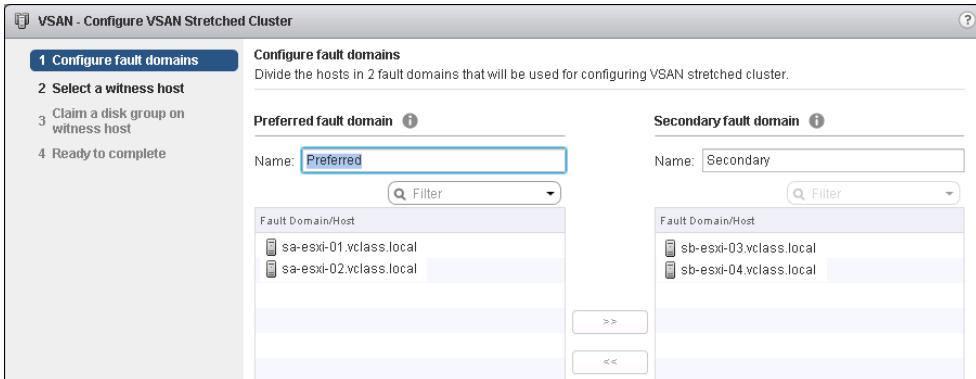
In this example, four nodes are available: sa-esxi-01.vclass.local, sa-esxi-02.vclass.local, sb-esxi-03.vclass.local, and sb-esxi-04.vclass.local. All four hosts reside in a cluster called VSAN\_Stretched. The fifth host, sc-witness-01.vclass.local, is the witness host. The witness host is on its own site and must not be added to the cluster.

# Configuring the Stretched Cluster

Slide 8-20

Use vSphere Web Client to configure the stretched cluster:

- Add hosts to the preferred fault domain and the secondary fault domain.
- Select the witness.
- Create a disk group on the witness host.



You configure the stretched cluster from the Manage > Settings > Fault Domains & Stretched Cluster panel.

Hosts must be added to the fault domains. The names Preferred and Secondary have been preassigned, but you can rename the domains. The preferred site is the site that will run the virtual machines if a network partition occurs.

In the example, sa-esxi-01.vclass.local and sa-esxi-02.vclass.local are assigned to the preferred site. sb-esxi-03.vclass.local and sb-esxi-04.vclass.local are assigned to the secondary site.

After adding hosts to the sites, you select the witness host. This host does not reside in the cluster.

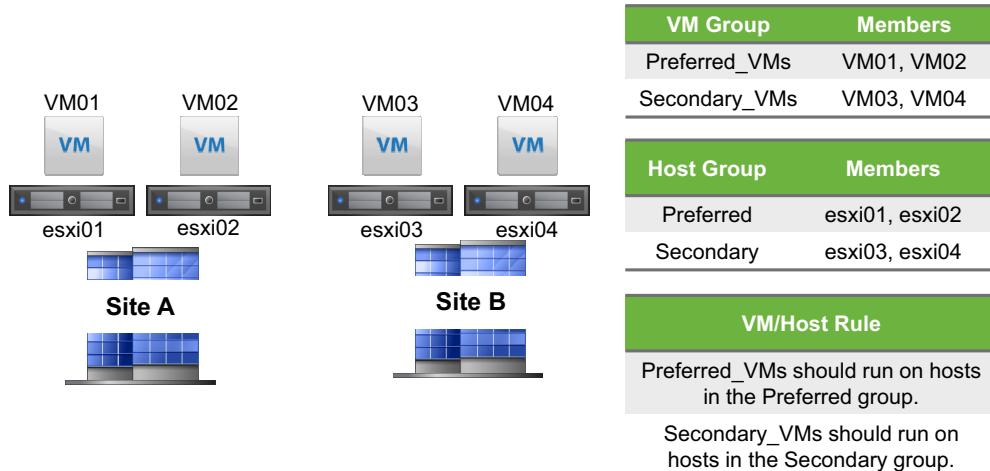
When the witness is selected, a flash device and a magnetic disk must be chosen to create a disk group. These disks are already available in the witness appliance.

# Configuring DRS Affinity Groups and Rules

Slide 8-21

VMware recommends enabling VMware vSphere® Distributed Resource Scheduler™ on stretched clusters:

- Use DRS affinity rules and groups to pin virtual machines to a specific data site.



VMware recommends enabling DRS to allow the creation of Host-VM affinity rules. These rules help to do the initial placement of the virtual machines and to avoid unnecessary vSphere vMotion migrations of virtual machines between sites. These rules also help to avoid any negative impact on read locality.

VMware recommends the following configuration:

- Create two Host groups: Name them Preferred and Secondary to match the Virtual SAN naming conventions.
- Create two VM groups: Name them Preferred\_VMs and Secondary\_VMs.
- Create two VM/Host rules: These rules state that the Preferred\_VMs group should run on the hosts in the Preferred group, and the Secondary\_VMs group should run on the hosts in the Secondary group. Select **Should run on hosts in group**. By selecting this rule, the rule can be violated by vSphere HA in the case of a full site outage.

For most cases, you want the virtual machine to reside on the set of hosts in the selected host group. However, in the event of a full site failure, you want the virtual machines to be restarted on the surviving site.

## Interoperability with vSphere HA

Slide 8-22

If a data site is partitioned away or fails, all Virtual SAN objects become inaccessible in that partition:

- vSphere HA fails over the virtual machines running on that data site to the other data site.

vSphere HA Setting	Recommended Value
Host Monitoring	Enabled
Host Hardware Monitoring – VM Component Protection	Disabled (default)
Virtual Machine Monitoring	Customer Preference – Disabled by default
Admission Control	Reserved failover CPU capacity: 50% Reserved failover memory capacity: 50%
Host Isolation Response	Power off and restart VMs
Datastore Heartbeats	Disable datastore heartbeats: Use datastores only from the specified list, but do not add datastores to the list
Host Isolation Addresses	Two isolation addresses: one per data site

Certain vSphere HA behaviors have been modified especially for Virtual SAN. Virtual SAN checks the state of the virtual machines per virtual machine. vSphere HA decides whether a virtual machine should be failed over based on the number of components belonging to a virtual machine that can be accessed from a particular partition.

vSphere HA and DRS do not use the witness host as a target because the witness is a standalone host in vCenter Server.

# Management and Maintenance

Slide 8-23

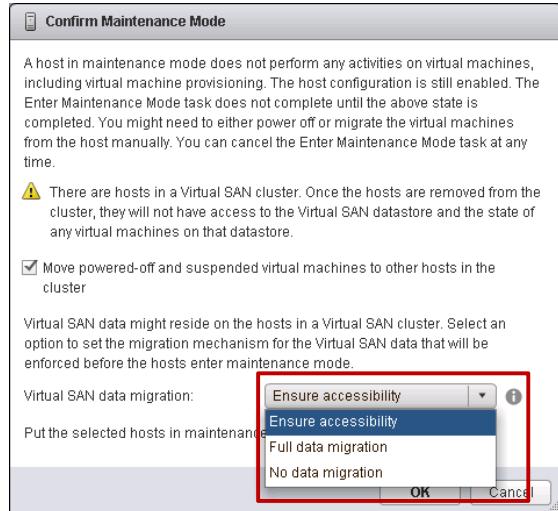
In a stretched cluster, you can use maintenance mode on a data site host and maintenance mode on the witness host.

For maintenance mode on a data site host:

- All maintenance modes are supported:
  - Ensure accessibility
  - Full data migration
  - No data migration

For maintenance mode on the witness host:

- The data migration option is not configurable.
- Witness objects are not migrated.



For hosts in the data sites, all data migration options are available when placing a host in maintenance mode. However, to do a full data migration, you must ensure that enough resources are available in the same site to facilitate the rebuilding of components on the remaining hosts on that site.

Maintenance mode on the witness host should be an infrequent event, because the witness host does not run virtual machines. Before doing maintenance on the witness host, you should check that all virtual machines are in compliance and no failures are occurring.

## About the Two-Node Cluster

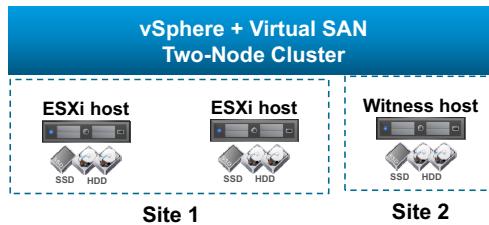
Slide 8-24

Virtual SAN can be implemented with two physical ESXi nodes in a cluster.

A third virtual node exists, which serves as the witness host:

- The special witness appliance is intended for use as the witness host.

The nodes can be located at the same site or at different sites.



With the two-node model, the required failure zones are based on three nodes: two physical nodes and a witness virtual appliance. The use of the witness appliance eliminates the requirement for a third physical node.

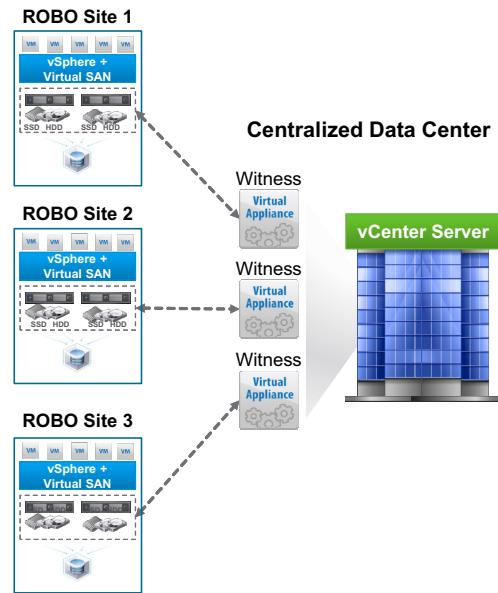
The two-node model enables you to take advantage of the manageability, performance, and availability benefits that a Virtual SAN cluster offers, without the minimum requirement of three nodes.

## Two-Node Cluster Use Case

Slide 8-25

The two-node architecture is ideal for Remote Office/Branch Office (ROBO) use cases:

- Remote offices are centrally managed by one vCenter Server instance.
- Each two-node cluster has its own witness.



The two-node model is ideal for smaller environments such as ROBO. You can deploy multiple two-node Virtual SAN clusters, located in branch offices, and centrally manage them from a single vCenter Server instance from a centralized data center.

## Two-Node Cluster and Stretched Cluster

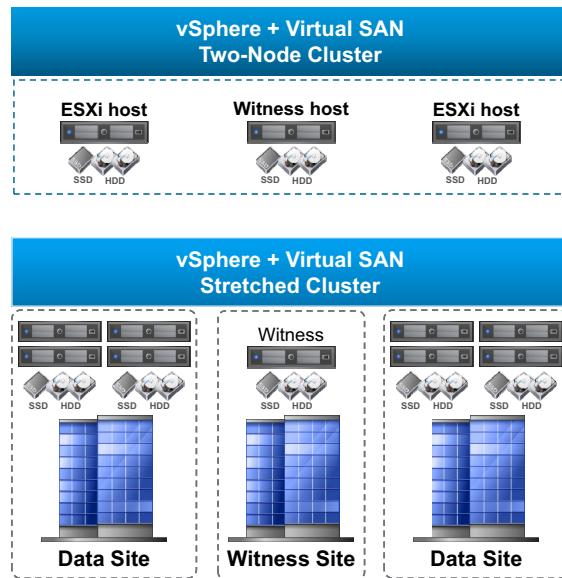
Slide 8-26

The two-node cluster and stretched clusters have the same architecture, based on fault domains.

The witness host is only supported for these two configurations.

Both hybrid and all-flash configurations are supported.

You configure a two-node cluster in the same way that you configure a stretched cluster.



The same procedure is used for setting up the witness host, networking, and fault domains for both two-node clusters and stretched clusters. When you configure either a two-node cluster or a stretched cluster, you assign hosts to the preferred domain and the secondary domain, and identify the witness host. In a two-node cluster, the preferred domain and the secondary domain each contain only one host.

## Lab 15: Creating a Stretched Cluster

Slide 8-27

Configure the witness appliance and create a stretched cluster

1. (Optional) Prepare the Environment
2. Add the Witness Appliance to the Data Center
3. Configure the Virtual SAN Network for the Witness Appliance
4. Create Static Routes on the ESXi Hosts
5. Remove the Existing Fault Domains
6. Configure a Stretched Cluster
7. Check the Health of the Stretched Cluster
8. View Virtual Machine File Placement in the Stretched Cluster

VMware Confidential  
Internal Use Only

## Review of Learner Objectives

Slide 8-28

You should be able to meet the following objectives:

- Describe the architecture for stretched clusters and two-node clusters
- Create a stretched cluster

VMware Confidential  
Internal Use Only

## Stretched Cluster Failure Scenarios

Slide 8-29

### Lesson 2: Stretched Cluster Failure Scenarios

VMware Confidential  
Internal Use Only

## Learner Objective

Slide 8-30

By the end of this lesson, you should be able to meet the following objective:

- Discuss the behavior of a stretched cluster when various failures occur

VMware Confidential  
Internal Use Only

# Stretched Cluster Heartbeats

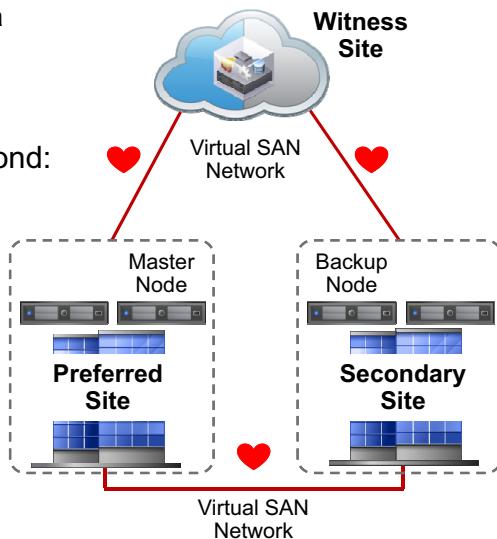
Slide 8-31

Stretched clusters use heartbeats to detect site failures:

- Virtual SAN designates a master node on the preferred site and a backup node on the secondary site to send and receive heartbeats.

Heartbeats are sent every second:

- Between the master node and the backup node
- Between the master node and the witness host
- Between the backup node and the witness host



When a stretched cluster is created, an ESXi host on the preferred site is designated as the master node. A master node is always selected from an available node on the preferred site. Similarly, an ESXi host on the secondary site is designated as the backup node. A backup node is always selected from an available node on the secondary site.

If communication is lost for five consecutive heartbeats (5 seconds) between the master node and the backup node, due to an issue with the backup node, the master node chooses a different ESXi host as a backup node on the secondary site. This process is repeated until all hosts on the secondary site are checked. If a complete site failure occurs, the master node selects a backup node from the preferred site. A similar scenario arises when the master node has a failure.

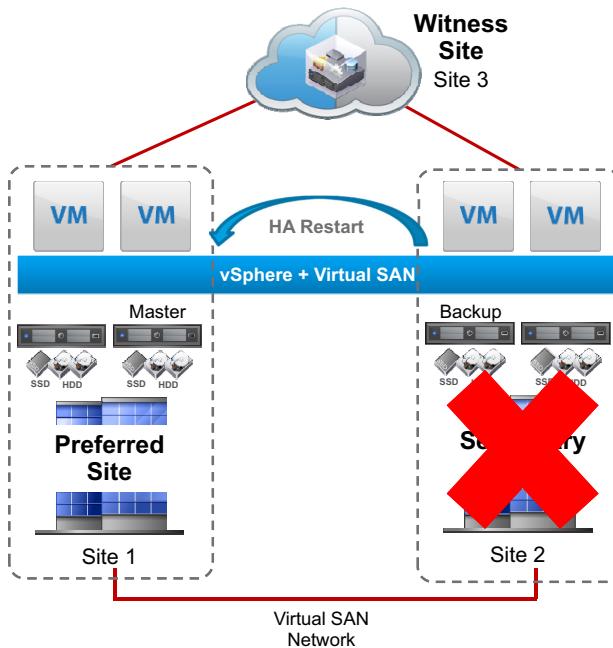
If a node rejoins the empty preferred site after a preferred site failure, then the master node migrates back to that site. If a node rejoins the empty secondary site after a secondary site failure, then the backup node migrates back to that site.

If communication is lost for five consecutive heartbeats (5 seconds) between the master and the witness, the witness is deemed to have failed. If the witness has suffered a permanent failure, a new witness host can be configured and added to the cluster.

## Data Site Failure

Slide 8-32

In the event of a complete site failure, vSphere HA restarts the virtual machines from the failed site to the other site.



If either of the data sites fails, then the data site that is still running creates a quorum with the witness site. The virtual machines' files and objects continue to be accessible on the Virtual SAN datastore. The reason is because a full copy of the data is available on the remaining data site and the witness components are available on the witness host, together accounting for more than 50 percent of the objects' components.

If the ESXi host, which holds the compute resources of the virtual machine is unaffected by this failure, then vSphere HA takes no action and no virtual machine downtime is experienced.

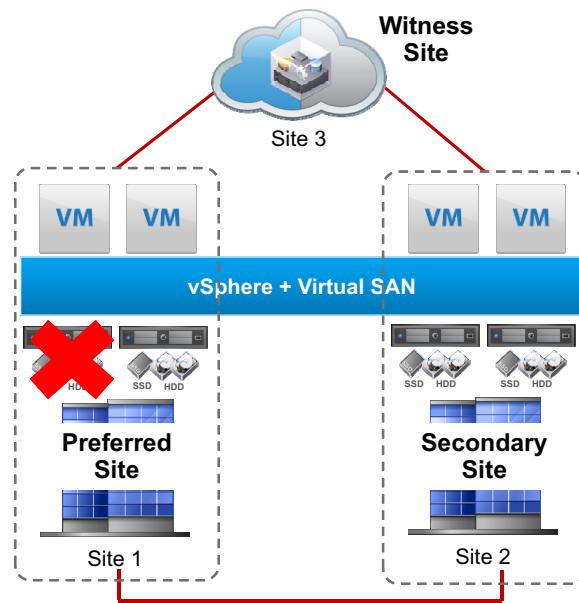
While the data site is down, you can continue to create and run virtual machines. However, the virtual machines will not be compliant with their storage policies. Remember that the number of failures to tolerate for a stretched cluster is one. After the failed data site is back online, the two data sites are resynchronized and the virtual machines become compliant again.

## Failure of Single Host in Data Site

Slide 8-33

If multiple hosts exist in a data site, then vSphere HA fails over the virtual machines on the failed host to other hosts in the same site.

If the site contains a single host, then vSphere HA fails over the virtual machines to a host on the remaining data site.



If each data site has multiple hosts, then a host failure on one of the data sites allows vSphere HA to restart virtual machines on other hosts on the same site. If the data site has only one host, and that host fails, then vSphere HA restarts the virtual machines on the other site.

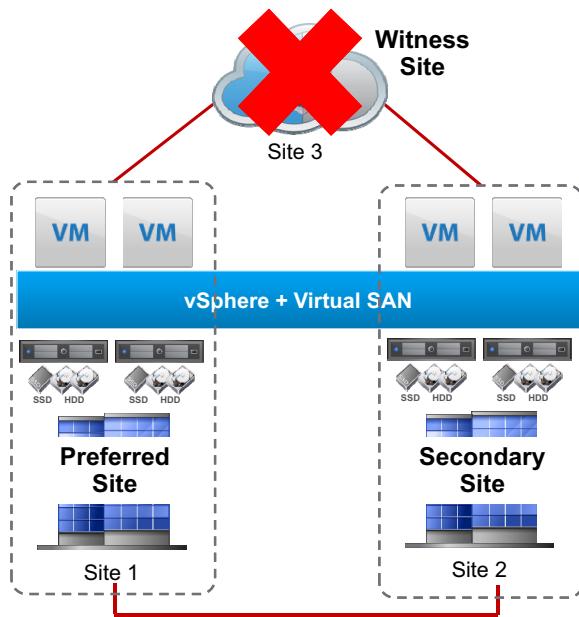
# Witness Host Failure or Loss of Network Connection

Slide 8-34

If communication is lost for five consecutive heartbeats between the witness host and the data sites, then the witness has failed.

Virtual machines continue to run without interruption:

- The two data sites continue to have a quorum.



If communication is lost for five consecutive heartbeats (5 seconds) between the master and the witness host, the witness host is deemed to have failed. Loss of communication can happen if the witness host fails or becomes isolated from the preferred site or the secondary site. Objects become noncompliant but the run state of the virtual machines is not impacted. One full copy of the virtual machines' data is available, and more than 50 percent of the objects' components are available. But the witness components residing on the witness host appear as Absent.

If the witness host has suffered a permanent failure, then you can replace the failed witness appliance with a new witness appliance. The stretched cluster configuration must be modified to point to the new witness appliance.

After the witness appliance is back online, the metadata for the objects in the cluster is resynchronized.

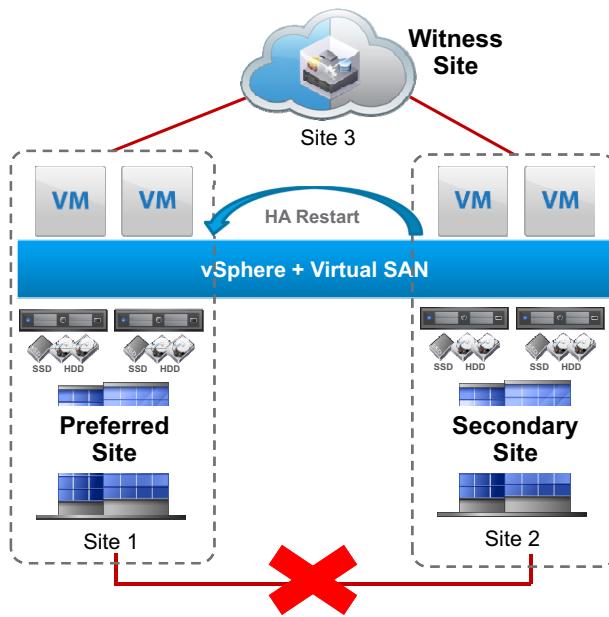
# Network Failure Between Data Sites

Slide 8-35

If the network fails between the data sites, then the preferred site is selected as the surviving site.

vSphere HA restarts all the virtual machines from the secondary site on the preferred site.

After the network is available, DRS might move the virtual machines based on VM-to-host affinity rules.



A network failure between data sites occurs because five consecutive heartbeats from the master node have not been received by the backup node.

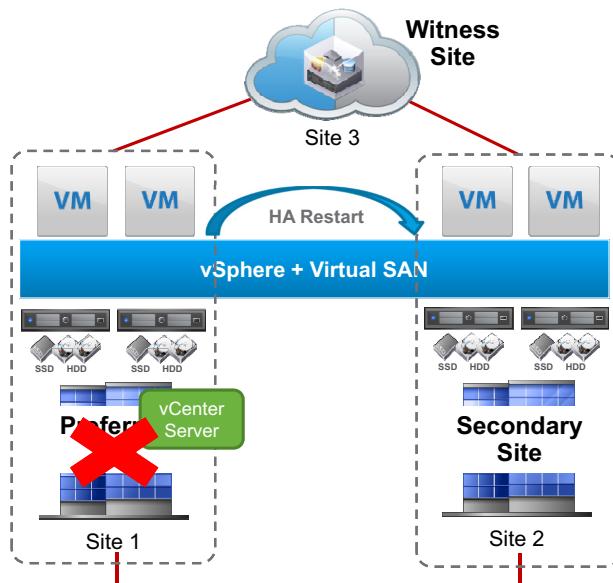
When this type of failure occurs, virtual machines can run on either of the two data sites. Therefore, if network connectivity is lost between the two sites, but both sites still have connectivity to the witness site, the preferred site survives. The components on the preferred site remain active, while the storage of the secondary site is marked as down and the components on that site are marked as absent.

## Site Failure Where vCenter Server Is Hosted

Slide 8-36

If the data site that holds vCenter Server fails, then the other site creates a quorum with the witness.

vSphere HA restarts all the virtual machines from the failed site to the other site.



If the site that holds the vCenter Server instance fails, then the other site creates a quorum with the witness site. vSphere HA restarts the virtual machines from the failed site on the other site.

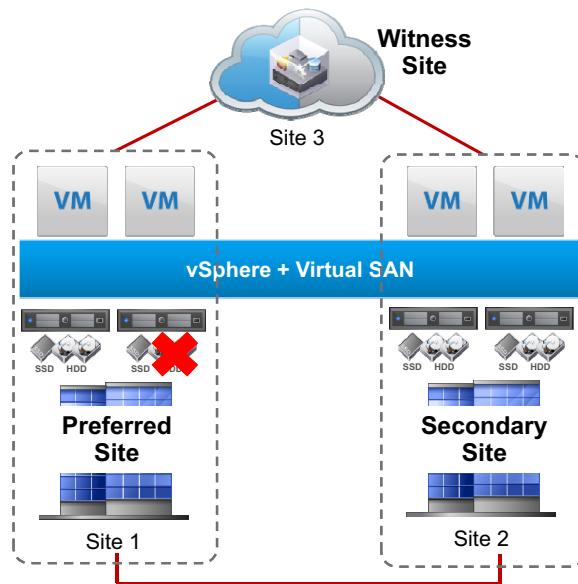
vSphere HA might restart the vCenter Server system as a virtual machine that is part of a cluster. If the vCenter Server system is not part of a cluster, then the vCenter Server system must be recovered.

## Disk Failure

Slide 8-37

If a disk fails on one of the hosts in a data site, then the virtual machines continue to run without interruption:

- One full copy of the data is still available.
- More than 50 percent of the components making up the object are available.



This disk that fails contains one of the components belonging to an object that is part of the virtual machine. However, the running virtual machine is not impacted because one full copy of the data is still available, and more than 50 percent of the components included in the object are available. The missing data component appears as absent in vSphere Web Client.

Similar behavior occurs if a disk fails in a disk group on the witness host.

## Lab 16: Configuring vSphere HA and DRS in the Stretched Cluster

Slide 8-38

Configure vSphere HA and DRS in the stretched cluster and simulate failure scenarios

1. (Optional) Prepare the Environment
2. Enable vSphere HA on the Cluster
3. Create DRS Affinity Rules and Groups
4. Simulate a Site Failure
5. Observe the Virtual Machine State
6. Resolve the Site Failure
7. (Optional) Recreate Static Routes on the ESXi Hosts

VMware Confidential  
Internal Use Only

## Review of Learner Objective

Slide 8-39

You should be able to meet the following objective:

- Discuss the behavior of a stretched cluster when various failures occur

VMware Confidential  
Internal Use Only

## Key Points

Slide 8-40

- Two-node clusters and stretched clusters have the same architecture and are configured using the same procedure.
- The witness appliance does not run virtual machines and is only supported with two-node clusters and stretched clusters.
- Two-node clusters and stretched clusters have network latency and bandwidth requirements that should be followed.
- Configuring a two-node cluster or stretched cluster includes adding hosts to the preferred and secondary fault domains, and identifying a witness host.
- Heartbeats, sent every second between the data sites and witness site, are used to detect failures.

Questions?

VMware Confidential  
Internal Use Only

VMware Confidential  
Internal Use Only

# Interoperability with vSphere Features

Slide 9-1

Module 9

VMware Confidential  
Internal Use Only

# You Are Here

Slide 9-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
- 9. Interoperability with vSphere Features**
10. Designing a Virtual SAN Deployment

VMware Confidential  
Internal Use Only

## Importance

Slide 9-3

Virtual SAN includes native support for vSphere vMotion, vSphere HA, DRS, and other key features.

Other VMware products, such as VMware Horizon® View™, VMware vRealize® Automation™, and Site Recovery Manager are able to leverage Virtual SAN. Third-party solutions that add functionality to Virtual SAN are also supported.

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 9-4

By the end of this module, you should be able to meet the following objectives:

- Identify vSphere features and VMware products that interoperate with Virtual SAN 6.2
- Describe how Virtual SAN 6.2 interoperates with third-party products and solutions

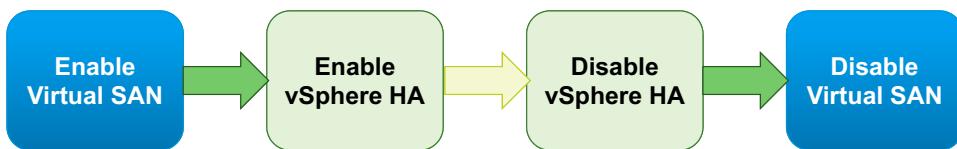
VMware Confidential  
Internal Use Only

## Virtual SAN and vSphere HA

Slide 9-5

Cluster requirements for using vSphere HA on a Virtual SAN cluster include the following:

- ESXi hosts in the cluster must be version 5.5 U1 or later.
- Virtual SAN and vSphere HA must be configured in a specific order:
  - Virtual SAN must be enabled before vSphere HA is enabled.
  - vSphere HA must be disabled before Virtual SAN is disabled.



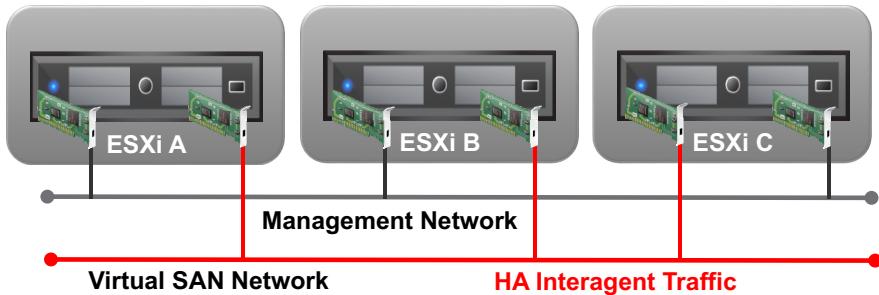
vSphere HA and Virtual SAN can be enabled on the same cluster and provide vSphere HA protection for virtual machines that reside on the Virtual SAN datastore. The features for vSphere HA must be enabled and disabled while Virtual SAN is active. Enabling vSphere HA before Virtual SAN on a cluster prevents Virtual SAN from being enabled until vSphere HA is disabled.

## vSphere HA Networking Differences with Virtual SAN

Slide 9-6

Consider the following when enabling vSphere HA and Virtual SAN on the same cluster:

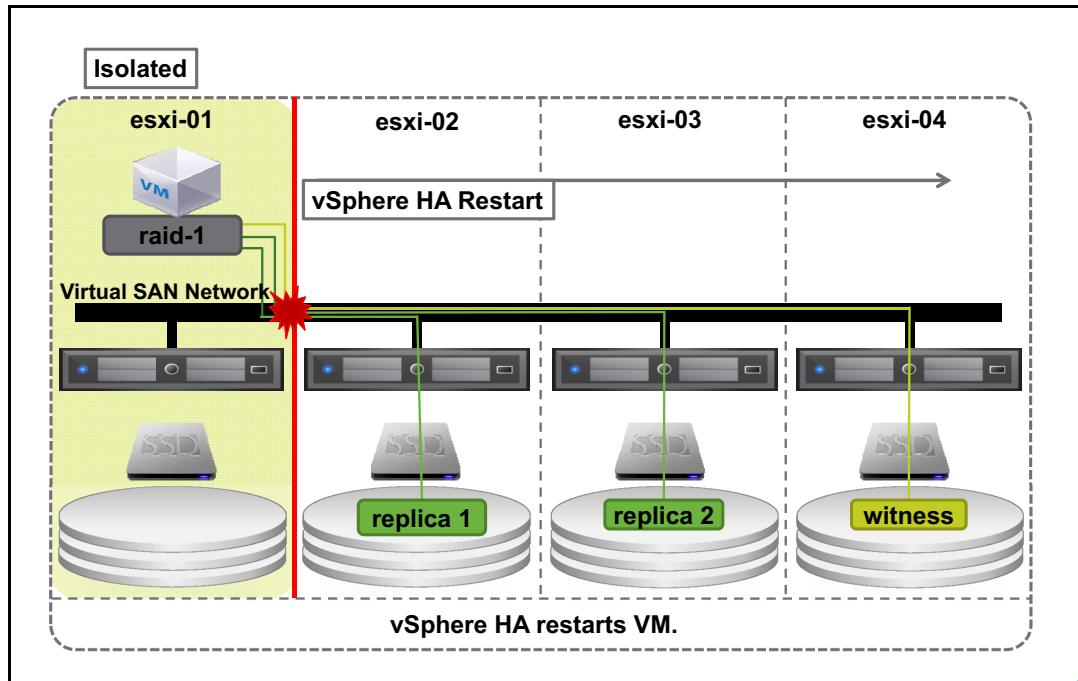
- vSphere HA interagent traffic traverses the Virtual SAN network rather than the management network.
- vSphere HA traffic migrates back to the management network if Virtual SAN is disabled.



Virtual SAN uses its own logical network. When Virtual SAN and vSphere HA are enabled for the same cluster, the vSphere HA interagent traffic flows over the Virtual SAN network instead of the management network.

## One Host Isolated: vSphere HA Restarts Virtual Machine

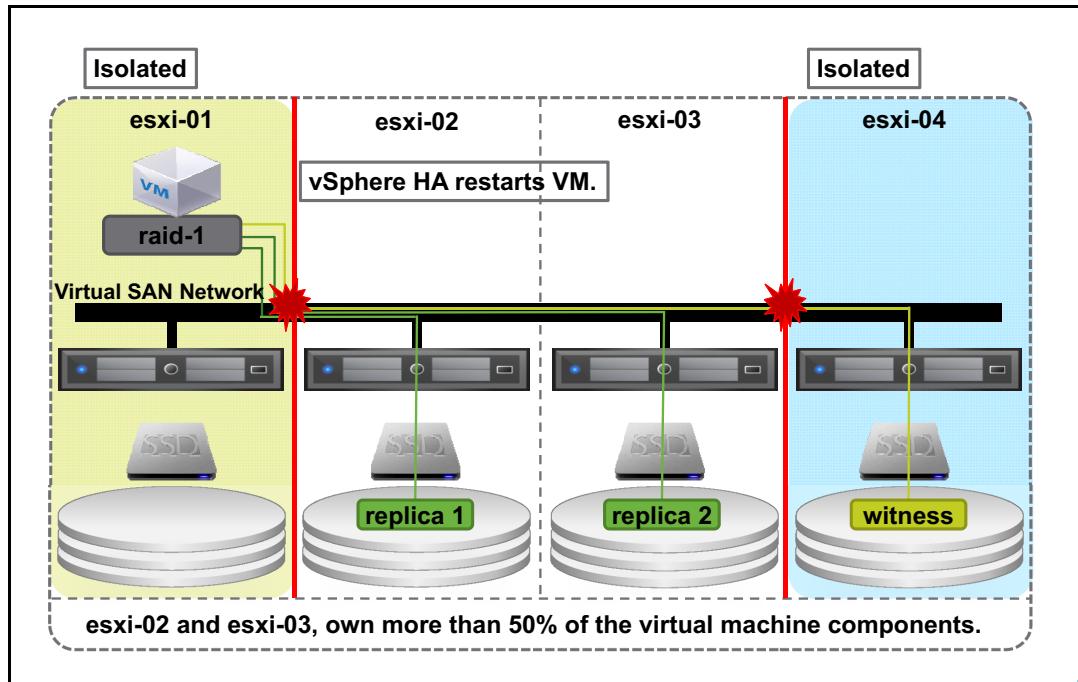
Slide 9-7



In the example, esxi-01 is isolated from the other hosts. The virtual machine that is hosted on esxi-01 is unavailable when esxi-01 becomes isolated. Because the files that are included in the virtual machine are hosted on storage available to the other ESXi hosts, vSphere HA restarts the virtual machine on one of the other hosts.

## Two Hosts Isolated: vSphere HA Restarts Virtual Machine

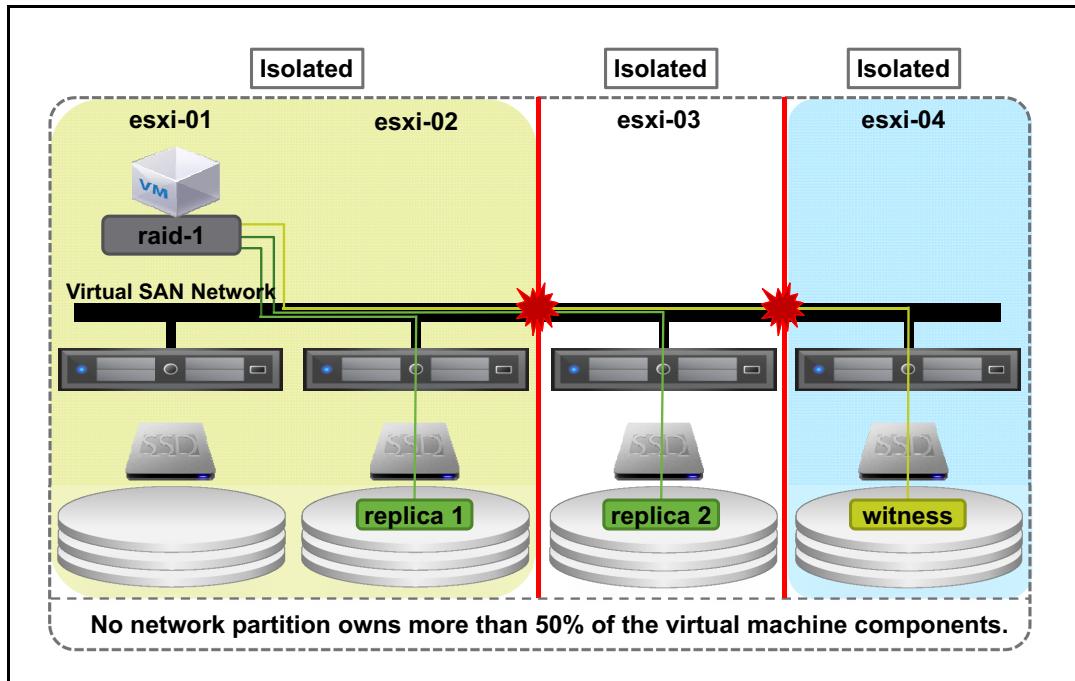
Slide 9-8



In the example, both esxi-01 and esxi-04 are isolated from the cluster. vSphere HA restarts the virtual machine on esxi-02 or esxi-03 because these hosts include more than half of the components for the virtual machine and a full replica.

## Hosts Isolated: vSphere HA Fails to Restart Virtual Machine

Slide 9-9

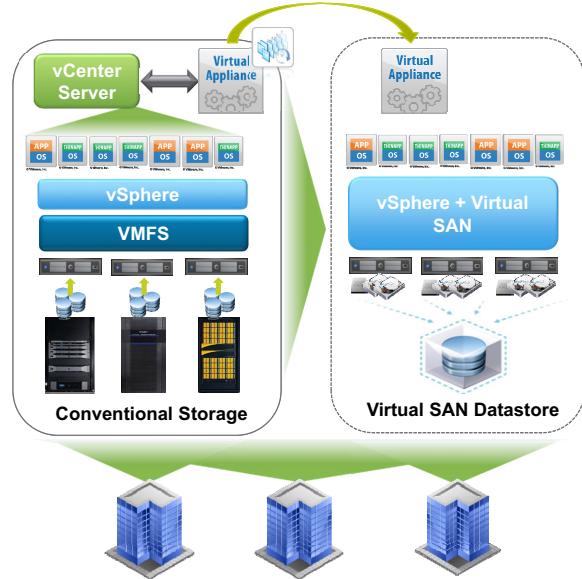


In the example, all hosts that include a component of the virtual machine are isolated from each other. Because none of these hosts own more than 50 percent of the virtual machine components, the virtual machine will be powered off and cannot be restarted.

# vSphere Data Protection

Slide 9-10

Virtual SAN is compatible with VMware vSphere® Data Protection™.



Virtual SAN supports VMware vSphere® Data Protection™. The vSphere Data Protection appliance can reside on vsanDatastore and back up virtual machines residing on the Virtual SAN datastore. The full complement of restore operations is available, such as overwrite and create new.

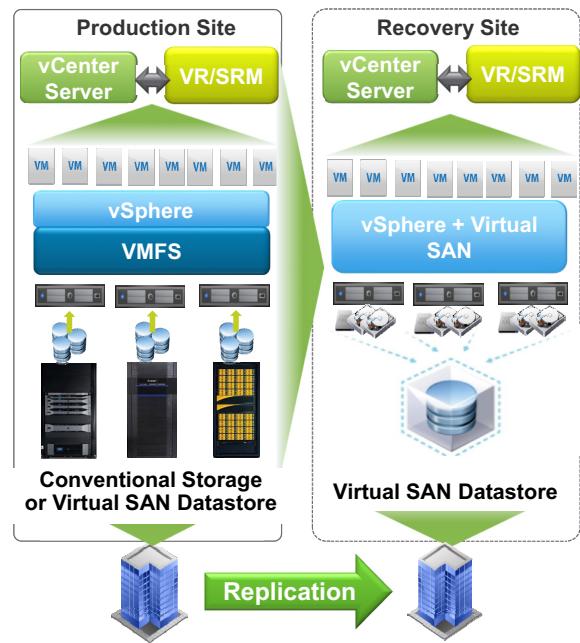
Each vSphere Data Protection appliance supports up to 400 virtual machines and up to 20 vSphere Data Protection appliances can be deployed per vCenter Server instance. After the appliance is deployed, management is performed by using vSphere Web Client with a supported Web browser.

# vSphere Replication and Site Recovery Manager

Slide 9-11

Fully automated recovery operations and orchestration procedures are supported:

- Planned migrations
- Automated failover
- Automated failback
- Nondisruptive disaster recovery testing



Virtual SAN fully supports vSphere Replication. vSphere Replication is a host-based asynchronous replication engine. vSphere Replication is deployed as one or more virtual appliances and managed by using vSphere Web Client. vSphere Replication is used to replicate virtual machines between different types of storage, such as from a traditional SAN to Virtual SAN. vSphere Replication supports storage policy-based management (SPBM) and a storage policy for the recovered virtual machine can be specified when configuring vSphere Replication.

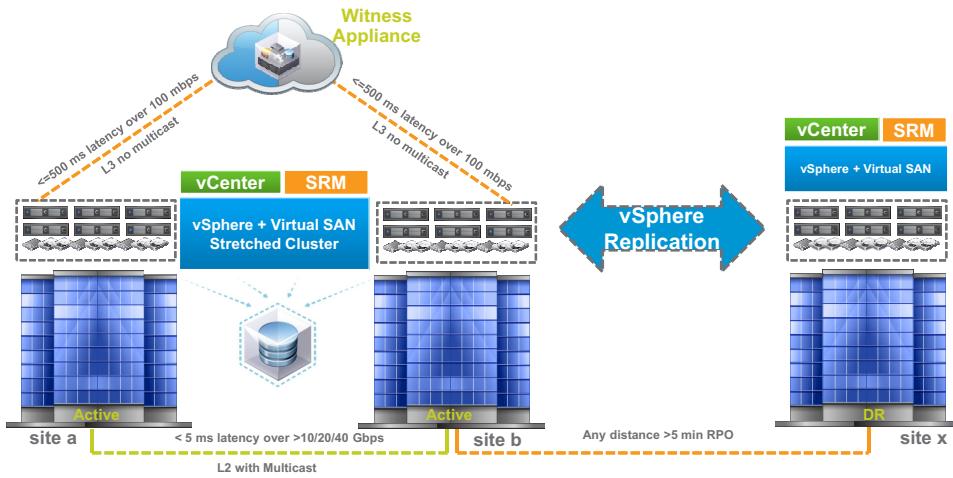
With vSphere Replication, virtual machines are recovered one at a time. This method works well when only few virtual machines must be recovered. To recover many virtual machines, Site Recovery Manager is recommended. Site Recovery Manager orchestrates the recovery of large numbers of virtual machines. Site Recovery Manager automates IP address changes during failover and provides automated failback. Site Recovery Manager can be used with array replication and vSphere Replication for faster, more reliable data center migrations and disaster recoveries.

# Stretched Cluster with vSphere Replication and SRM

Slide 9-12

Stretched clusters can be used with VMware vSphere® Replication™ and VMware Site Recovery Manager™:

- Replication between Virtual SAN datastores enables RPOs as low as 5 minutes.



vSphere Replication and Site Recovery Manager support Virtual SAN stretched clusters. Live migrations and automated vSphere HA restarts can occur between stretched cluster sites. Lower RPO values are achievable due to the efficient vsanSparse snapshot mechanism of Virtual SAN.

Site Recovery Manager is designed to protect virtual machines residing in datastores on replicated storage at the protected site. If a storage array failure or a complete site failure occurs, virtual machines can be failed over to a remote data center (also called the recovery site). The recovery virtual machines continue to operate while the protected site is unavailable.

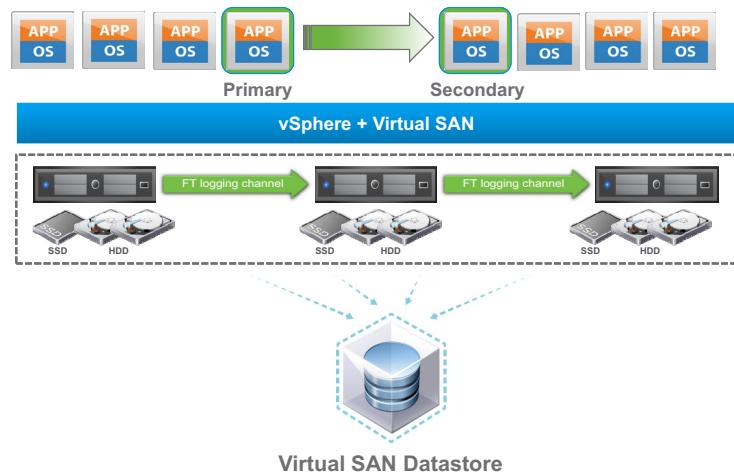
Site Recovery Manager requires vCenter Server to be installed at the protected site and at the recovery site. Therefore, Site Recovery Manager does not support a standalone Virtual SAN cluster with one vCenter Server system.

# vSphere Fault Tolerance

Slide 9-13

Virtual machines enabled for VMware vSphere® Fault Tolerance are supported in a Virtual SAN cluster:

- vSphere Fault Tolerance is not supported for stretched cluster configurations.



vSphere Fault Tolerance protects mission-critical, high-performance applications regardless of the operating system used. vSphere Fault Tolerance provides continuous availability for such a virtual machine by creating and maintaining another virtual machine that is identical and continuously available to replace it in the event of a failover situation.

The protected virtual machine is called the Primary VM. The duplicate virtual machine, the Secondary VM, is created and runs on another host. The Secondary VM's execution is identical to that of the Primary VM and it can take over at any point without interruption, thereby providing fault tolerant protection.

VMware vSphere® Fault Tolerance can accommodate symmetric multiprocessor (SMP) virtual machines with up to four vCPUs.

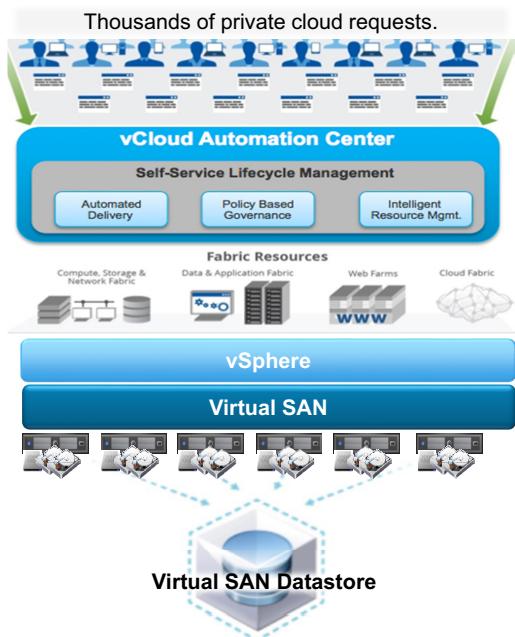
# vRealize Automation

Slide 9-14

vRealize Automation supports Virtual SAN as a storage platform.

vRealize Automation provides centralized cloud management through the following:

- Customizable portal
- Cloud service storage costing models
- Self-service consumption capabilities



vRealize Automation provides some key benefits to Virtual SAN:

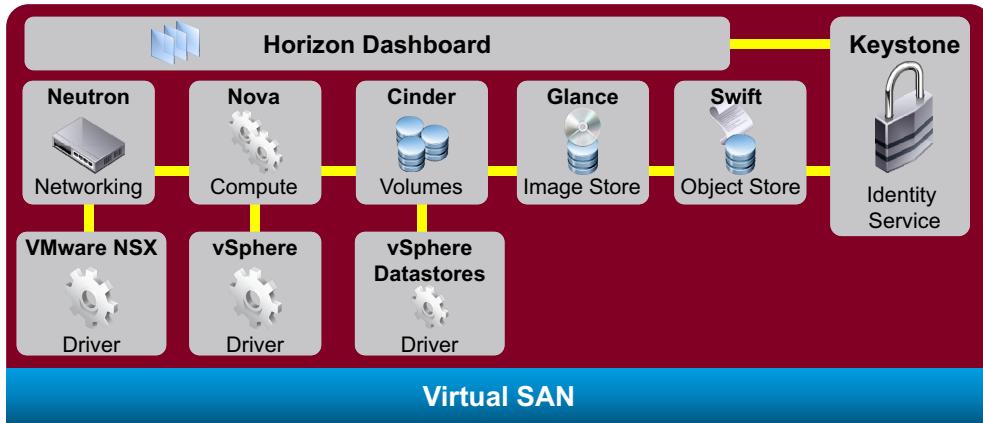
- Centralized provisioning and management
- Self-service consumption
- Entitlement, compliance monitoring, and enforcement
- Leverage of existing business processes and tools
- Delegation for the control of resources

# OpenStack Framework

Slide 9-15

The OpenStack framework supports Virtual SAN as a storage platform:

- Cloud-ready application to hypervisor-converged solution
- Leverage the use of flash-optimized storage in OpenStack
- Resiliency for legacy and cloud-ready applications
- vSphere Web plug-in for OpenStack UI



Virtual SAN supports interoperability with the OpenStack framework by using Cinder through the vSphere datastore driver.

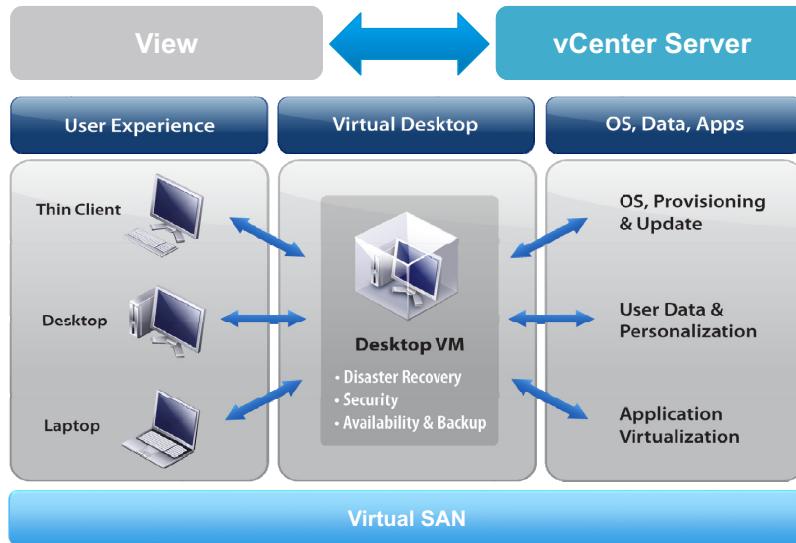
Virtual SAN provides several benefits to OpenStack:

- Hybrid disk solutions with flash-based cache devices and magnetic disks for capacity
- Performance disk solutions with flash-based cache devices and flash disks for capacity
- High performance and distributed RAID architecture
- Built-in application availability
- Policy-based storage management
- Dynamic and scalable storage capabilities

# View

Slide 9-16

View supports Virtual SAN as a storage platform.



The scale-out nature of Virtual SAN and the performance that is provided by the solid-state drive (SSD) layer makes Virtual SAN a low-cost and high-performance solution for View.

When used with additional vSphere features, such as vSphere Storage Accelerator, desktops deployed on Virtual SAN and View provide an excellent combination for customers interested in VDI.

Virtual SAN does not support the SE Sparse Disk format on View. Thus, only desktops that are deployed as full clones or as linked clones using the vmfsSparse format are supported.

# vSphere PowerCLI Cmdlets

Slide 9-17

VMware vSphere® PowerCLI™ 6.0 delivers a set of cmdlets for managing Virtual SAN:

- The new cmdlets are the following:

- Export-SpbmStoragePolicy
- Get-SpbmCapability
- Get-SpbmCompatibleStorage
- Get-SpbmEntityConfiguration
- Get-SpbmStoragePolicy
- Get-VsanDisk
- Get-VsanDiskGroup
- Import-SpbmStoragePolicy
- New-SpbmRule
- New-SpbmRuleSet
- New-SpbmStoragePolicy
- New-VsanDisk
- New-VsanDiskGroup
- Remove-SpbmStoragePolicy
- Remove-VsanDisk
- Remove-VsanDiskGroup
- Set-SpbmEntityConfiguration
- Set-SpbmStoragePolicy



For additional details on these commands, see vSphere PowerCLI Documentation at <https://www.vmware.com/support/developer/PowerCLI/> and the VMware blogs Web site.

# File Services with NexentaConnect

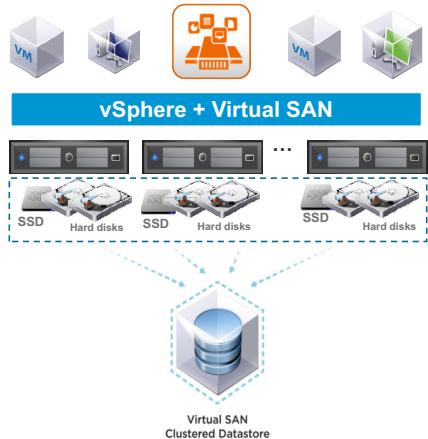
Slide 9-18

NexentaConnect complements the Virtual SAN simplified operating and storage consumption models:

- Adds file services
- Provides ease of management capabilities
- Leverages storage policy-based management
- Is used for storing files while Virtual SAN is for virtual machine storage

NexentaConnect offers flexibility and benefits to vSphere administrators, such as:

- Abstracted pool of file services
- High-performance NFS and SMB shares
- Live monitoring capabilities
- Disaster recovery planning capabilities



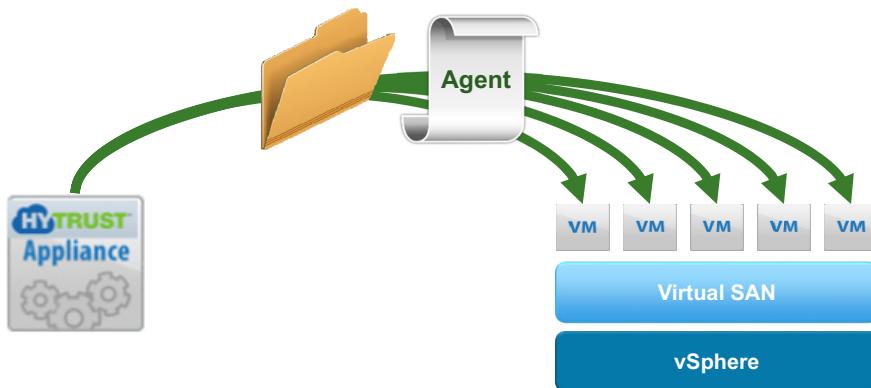
NexentaConnect is a third-party application that complements a Virtual SAN deployment. NexentaConnect provides file services for Virtual SAN that adds NFS and SMB access in addition to the existing Virtual SAN datastore. The software is managed through vSphere Web Client and saves capacity by using inline compression and deduplication.

# Virtual SAN and Hytrust DataControl

Slide 9-19

Hytrust DataControl offers the following benefits:

- Encryption functionality
- Easy deployment
- Infrastructure independence
- Operational transparency



Hytrust DataControl is offered by Hytrust as a supported encryption solution for Virtual SAN. Hytrust is a member of the VMware partner ecosystem that is focused on data security services for cloud infrastructures. Hytrust DataControl is a software-based solution that is designed to protect virtual machines and their data. Hytrust DataControl is easy to deploy and manage. These capabilities comply with one of the main principles of Virtual SAN, which is simplicity and ease of management. Hytrust DataControl virtual machine edition is based on a software agent that encrypts data from the Windows or Linux operating system of a virtual machine and ensures protection.

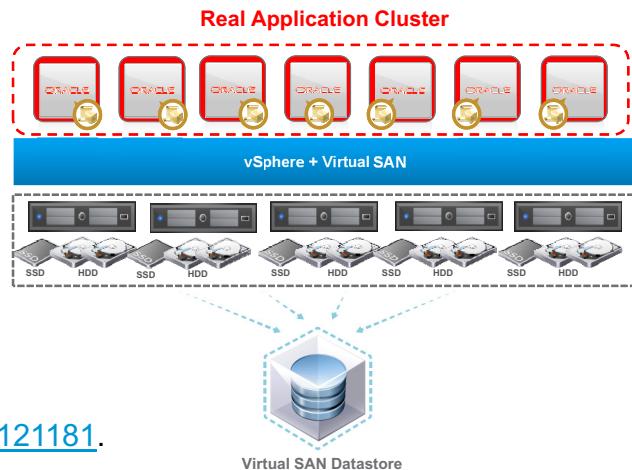
# Virtual SAN and Oracle RAC

Slide 9-20

Virtual SAN supports Oracle Real Application Clusters (RAC):

- Supported on all Virtual SAN deployment types, including stretched cluster configurations.
- VMDK must be in eager-zeroed thick format.
- VMDK must be enabled with the multi-write option.

For setup requirements, see VMware knowledge base article 2121181 at <http://kb.vmware.com/kb/2121181>.



Oracle Real Application Clusters (RAC) allows multiple virtual machines to run Oracle RDBMS software simultaneously while accessing a single database, thus providing clustering.

As with VMFS and NFS datastores, Virtual SAN prevents multiple virtual machines from opening the same VMDK file in read-write mode. Thus, the data stored on the virtual disk is prevented from corruption caused by multiple writers on the non-cluster-aware filesystems used by most guest operating systems.

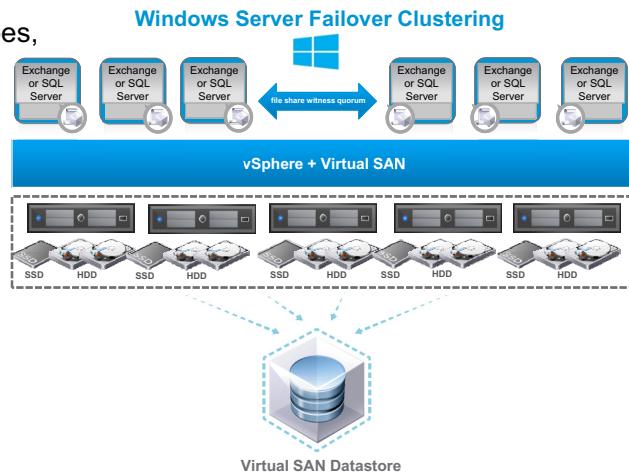
To use in-guest shared-storage clustering solutions, such as Oracle RAC, multiple virtual machines must be allowed to access the same VMDK files in read-write mode. To enable this distributed write capability, you must explicitly enable multi-writer support for all applicable virtual machines and VMDK files.

# Virtual SAN and Windows Server Failover Clustering

Slide 9-21

Virtual SAN supports Microsoft Windows Server Failover Clustering:

- Supported on all Virtual SAN deployment types, including stretched cluster configurations
- Supports file share witness quorum with:
  - Exchange Data Availability Groups (DAG)
  - SQL Server AlwaysOn Availability Groups (AAG)
- Does not support DAG or SQL Server AlwaysOn with Failover Cluster Instances (FCI)



Windows Server Failover Clustering can be enabled in a cluster to guard against application and service failures. With Windows Server Failover Clustering, you can run multiple SQL Server instances or multiple Exchange servers that access the Virtual SAN datastore.

Currently, Virtual SAN supports Windows Server Failover Clustering with only Microsoft Exchange and SQL Server, and only with the limitations stated above.

## Review of Learner Objectives

Slide 9-22

You should be able to meet the following objectives:

- Identify vSphere features and VMware products that interoperate with Virtual SAN 6.2
- Describe how Virtual SAN 6.2 interoperates with third-party products and solutions

VMware Confidential  
Internal Use Only

## Key Points

Slide 9-23

- Virtual SAN integrates with critical vSphere features like vSphere HA, DRS, and vSphere vMotion.
- Virtual SAN supports vRealize Automation and View.
- Virtual SAN supports a variety of third-party software.
- Virtual SAN easily integrates with available third-party encryption software.

Questions?

VMware Confidential  
Internal Use Only

VMware Confidential  
Internal Use Only

# Designing a Virtual SAN Deployment

Slide 10-1

Module 10

VMware Confidential  
Internal Use Only

# You Are Here

Slide 10-2

1. Course Introduction
2. Storage Fundamentals
3. Introduction to Virtual SAN
4. Virtual SAN Configuration
5. Virtual SAN Policies and Virtual Machines
6. Managing and Operating Virtual SAN
7. Monitoring and Troubleshooting Virtual SAN
8. Stretched Clusters and Two-Node Clusters
9. Interoperability with vSphere Features
- 10. Designing a Virtual SAN Deployment**

VMware Confidential  
Internal Use Only

## Importance

Slide 10-3

You must understand the features and capabilities of Virtual SAN so that you can design a solution that fulfills the workload and scalability requirements. You must follow a design process to develop an appropriate design.

VMware Confidential  
Internal Use Only

## Learner Objectives

Slide 10-4

By the end of this module, you should be able to meet the following objectives:

- Understand Virtual SAN design considerations
- Plan and design Virtual SAN clusters
- Identify the design and sizing tools for Virtual SAN
- Describe Virtual SAN use cases

VMware Confidential  
Internal Use Only

# Following the Compatibility Guide

Slide 10-5

The most important part of designing a Virtual SAN cluster is ensuring that the proposed hardware is compatible.

Consider using the Virtual SAN Ready Node Configurator.

The screenshot shows a search interface for 'Virtual SAN' compatibility guides. A red box highlights the text 'Need Help? Try out the [Virtual SAN Ready Node Configurator](#).'. Below it, instructions say 'STEP 1: Refer to the "Virtual SAN Hardware Quick Reference Guide" for guidance on how to build a Virtual SAN Ready Node.' and 'STEP 2: To build a Virtual SAN Ready Node: Select your Virtual SAN Ready Node of choice based on following certified Ready Nodes.' The interface includes dropdown menus for 'Ready Node Types' (All), 'Ready Node Vendors' (All, Cisco, DELL, Fujitsu, Hewlett Packard Enterprise, Hitachi), 'Ready Node Profile' (All, HY-2 Series, HY-4 Series, HY-6 Series, HY-8 Series, AF-6 Series), 'Ready Node Generation' (All, Gen1 - 6G, Gen2 - 12G), 'Ready Node Server Type' (All, Blade, Rackmount), 'Keyword' (empty), 'Posted Date Range' (All), and 'Raw Storage Capacity (TB)' (All). At the bottom are 'Update and View Results' and 'Reset' buttons.

The most important aspect of storage design is ensuring that the components under consideration appear in the VMware Compatibility Guide. Always verify that VMware supports any hardware components that are used for a Virtual SAN deployment. This online tool is regularly updated to ensure that you have the latest compatibility guidance from VMware.

There are two ways to build a Virtual SAN cluster:

- Build your own Virtual SAN cluster based on certified components
- Choose from a list of Virtual SAN Ready Nodes

A Virtual SAN Ready Node is a validated server configuration in a tested, certified hardware form factor for Virtual SAN deployment, jointly recommended by the server OEM and VMware.

# Cluster Design Considerations (1)

Slide 10-6

Design Consideration	Description
Three-node configurations	<ul style="list-style-type: none"><li>Limited recovery</li><li>Limited data migrations</li><li>Low failure tolerance</li></ul>
Two-node configurations	<ul style="list-style-type: none"><li>Limited recovery</li><li>Ideal for small environments</li><li>Requires the use of a witness appliance</li></ul>
Fault domains	<ul style="list-style-type: none"><li>Provides rack awareness</li><li>Functions like multiple hosts</li></ul>
Stretched clusters	<ul style="list-style-type: none"><li>Enterprise availability and data protection</li><li>Provides site awareness</li></ul>
vSphere HA	<ul style="list-style-type: none"><li>Provides high availability for virtual machines</li><li>Fully compatible</li></ul>

Virtual SAN fully supports two-node and three-node configurations but these configurations behave differently than configurations with four or more nodes. A single failure can be tolerated for a virtual machine (FTT = 1). Also, you cannot migrate all data from a node during maintenance. Consider four or more nodes for the Virtual SAN cluster design for maximum availability.

To calculate the number of fault domains required to tolerate failures, use the same equation as you use for hosts.

Stretched clusters work across two data sites and one witness site. The maximum number of fault domains is three, and the number of failures to tolerate is one.

With vSphere HA, Virtual SAN provides a highly available solution for virtual machines. When designing your Virtual SAN deployment, consider the needs for vSphere HA.

Consider the number of hosts needed in the cluster to meet the NumberOfFailuresToTolerate policy requirements. Also, consider if enough resources are available on the remaining hosts to handle the amount of data that must be migrated from the host that is placed into maintenance mode.

## Cluster Design Considerations (2)

Slide 10-7

Design Consideration	Description
Maintenance mode	<ul style="list-style-type: none"><li>• Necessary for some tasks</li><li>• Multiple data migration options</li><li>• Requires similar resource as failures</li></ul>
Deduplication and compression	<ul style="list-style-type: none"><li>• Available for all-flash configurations only</li><li>• Can reduce the amount of physical storage consumed by as much as 7x</li><li>• Increases latency and CPU usage</li><li>• Good for environments with highly-redundant data</li></ul>

If enough resources are not available to do a full data migration when entering maintenance mode, the host fails to enter maintenance mode with the following message: Failed to enter maintenance mode in the current Virtual SAN data migration mode due to insufficient nodes or disks in the cluster. Retry the operation in another mode or after adding more resources to the cluster.

Enabling deduplication and compression can reduce the amount of physical storage consumed as much as 7 times, resulting in a lower total cost of ownership. Environments with highly-redundant data, such as full-clone virtual desktops and homogeneous server operating systems benefit the most from deduplication. Deduplication and compression results vary based on the types of data stored in an all-flash Virtual SAN environment. For example, compression offers more favorable results with data that compresses well, such as text, bitmap, and program files. Data that is already compressed, such as certain graphics formats and video files, as well as files that are encrypted, yields little or no reduction in storage consumption from compression.

# Using All-Flash Architectures

Slide 10-8

Virtual SAN supports both hybrid and all-flash architectures:

- A hybrid architecture delivers up to 40K IOPs per host.
- An all-flash architecture delivers up to 90K IOPs per host, with predictable low latencies.

You must use an all-flash configuration if you want to use the following new features in Virtual SAN 6.2:

- Deduplication and compression
- RAID-5/6 (Erasure Coding)

Virtual SAN 6.2 is optimized for modern all-flash storage, delivering space efficiency capabilities, such as near-line deduplication, compression, and erasure coding.

At the initial launch of Virtual SAN in March 2014, flash was still relatively expensive. Hybrid Virtual SAN with 10K RPM and 7.2K RPM drives offered a lot of value by enabling cost effective capacity and performance. Since then, flash has significantly closed in on the price advantage of traditional magnetic disks while still offering significantly better IOPS and latency than many workloads.

Storage trends are becoming more predictable. Hybrid configurations continue to provide excellent value for specific workloads. However, falling flash prices and data reduction technologies will allow many customers to make all-flash Virtual SAN configurations the primary deployment method.

## Disk Group Design

Slide 10-9

Each disk group has one cache device:

- Create multiple disk groups to leverage additional cache devices.
- The more cache to capacity a disk group has, the more cache is available to virtual machines for accelerated performance.

VMware recommends multiple disk groups. Multiple disk groups typically mean better performance and smaller fault domains, but might consume additional disk slots.

Disk group design affects availability, performance, and capacity. With additional disks and disk groups, components are distributed across more devices, which decreases risk. Using an increased number of flash devices and stripes has several performance benefits.

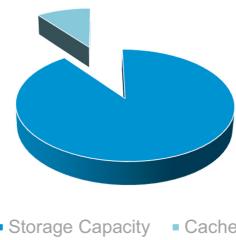
An individual disk group constitutes a single failure domain. If a cache disk fails, all capacity disks in that particular disk group become unusable and the total storage capacity provided by the affected disk group is unavailable to the cluster. Therefore, VMware recommends creating multiple disk groups to protect against a cache device failure.

## Cache Capacity Sizing

Slide 10-10

VMware recommends having at least a 10 percent flash cache to consumed capacity ratio.

Measurement Requirements	Values
Projected VM space usage	20 GB
Projected number of VMs	1000
Total projected consumption per VM	$20 \text{ GB} \times 1,000 = 20,000 \text{ GB} = 20 \text{ TB}$
Target flash capacity percentage	10%
Total flash capacity required	$20 \text{ TB} \times .10 = 2 \text{ TB}$



The general recommendation for sizing the Virtual SAN cache size is that the cache device must be a minimum of 10 percent of the anticipated consumed storage capacity before the Number of Failures To Tolerate parameter is considered. This configuration is a general recommendation and might not be adequate for some designs.

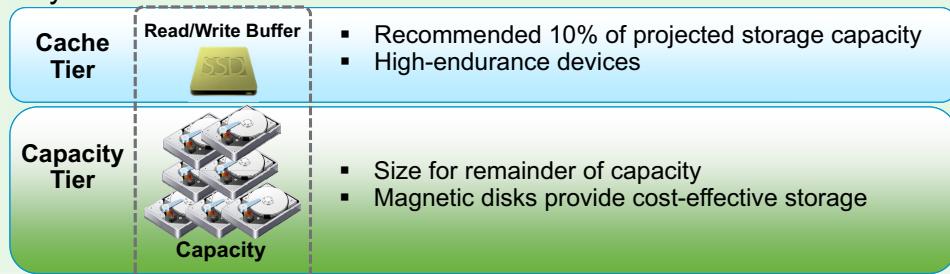
# Choosing Devices for Cache and Capacity Tiers

Slide 10-11

## All-Flash Architecture



## Hybrid Architecture



When choosing a device for the cache tier, VMware recommends devices that offer at least 2 or more terabyte writes per day for a 5-year rating of 3,650 or higher. For clusters with write-oriented workload characteristics, an even higher terabyte write per day value might be required. A small device that supports a high daily write per day rating might be equal in the long term to a large device with a lower write per day rating. So a 5-year terabyte per day write rating is a better calculation for sizing than strictly using the terabyte write per day or Program/Erase cycles.

Generally, given the same terabyte write per day rating, fewer but higher capacity devices yield better longevity because each individual write buffer location has greater capacity to hold component writes without needing to destage writes to disk. However, more cache devices and disk groups offer easier scaling and device replacement. Business requirements should drive the architecture decision between fewer but larger cache devices versus more but smaller cache devices.

When choosing devices for the capacity tier, you must consider that the capacity serves predominantly as a read layer. Thus, the primary consideration for sizing is the cost per gigabyte of storage capacity. The capacity tier should be sized to the total required space and not the estimated used space. When using flash for capacity, the device choice should be based on the cost per gigabyte of storage. The write per day value is not of particular concern and any supported device with a terabyte write per day value of 0.2 or higher is acceptable.

# Capacity Planning Considerations

Slide 10-12

Design requirements and constraints can affect the following:

- Total indicated cluster resources
- Number of hosts per cluster
- Individual component considerations:
  - Availability requirements
  - Potential bottlenecks
  - Security and workload isolation requirements
  - Anticipated future growth

The availability requirements of the virtual machines hosted on Virtual SAN affect design considerations. The number of hosts and amount of capacity required are affected if a large percentage of the virtual machines require more than a single failure or space reservations.

Plan for potential bottlenecks. Stripe virtual machines that require more disk I/O across more disks than virtual machines with less I/O intensive operations.

Plan the infrastructure to accommodate the security and isolation requirements for your environment.

The growth of the virtual machine infrastructure must also be considered. If fast growth is expected, then you must design an infrastructure to support growth.

# Establishing Baseline Capacity Requirements (1)

Slide 10-13

The following considerations are significant when planning for capacity:

- Count of virtual machines
- Average size of virtual machines:
  - Average virtual CPU count per virtual machine
  - Average memory consumption per virtual machine
  - Virtual disk size and utilization totals

Establishing baseline capacity requirements relies on virtual machine capacity estimates and host sizing limits. You must consider all the factors when looking at host sizing. Planning for Virtual SAN to support 1,000 virtual machines should also include the CPU and memory allocations.

## Establishing Baseline Capacity Requirements (2)

Slide 10-14

You can establish baseline capacity requirements of cluster CPU, memory, and storage by using these formulas:

- **Cluster CPU baseline:**

$VMs \times Average\_vCPU\_perVM / Target\_VCPU\_to\_PCPU$

- **Cluster memory baseline:**

$VMs \times Memory\_per\_VM\_in\_GB$

- **Cluster persistent storage baseline (assuming no overcommit):**

$VMs \times (VMDK\_Provisioned\_Size + (Memory\_Per\_VM \times 2))$

- **Minimum hosts per cluster:**

$VMs / Max\_VMs\_Per\_Host$

Using these formulas, establish a minimum set of requirements for the cluster. These formulas include baseline values and might need to be adjusted to adequately account for failure scenarios or growth at this stage in the design process.

The formulas in the slide include the following references:

- $VMs$  represents the total anticipated count of virtual machines to be run in the cluster.
- $Average\_vCPU\_perVM$  represents the number of processors anticipated per virtual machine over the entire set.
- $Target\_VCPU\_to\_PCPU$  represents the target number of virtual CPUs per physical processor.
- $Memory\_per\_VM\_in\_GB$  represents the total memory consumption, including overhead, per virtual machine.
- $Max\_VMs\_Per\_Host$  represents the desired maximum number of virtual machines per host.
- $VMDK\_Provisioned\_Size$  represents the estimated provisioned size of virtual machine disks per virtual machine.

# Virtual SAN TCO and Sizing Calculator

Slide 10-15

The Virtual SAN TCO and Sizing Calculator gives you specific information on the best strategy for your Virtual SAN deployment.

<http://vsantco.vmware.com>

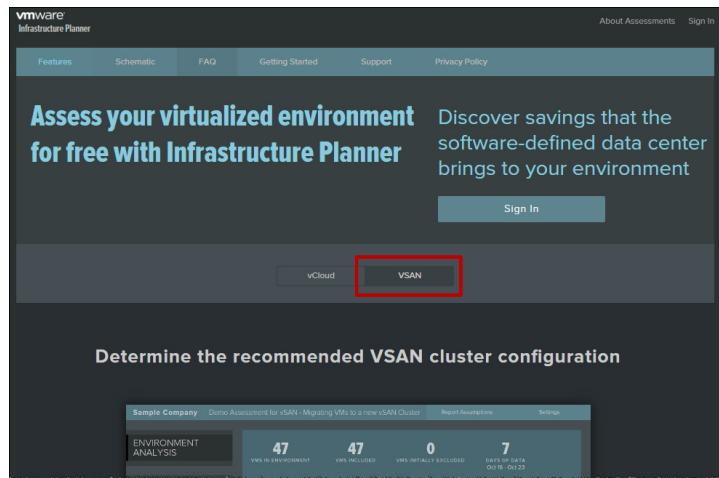
The Virtual SAN TCO and Sizing Calculator gives you specific information on the best strategy for your Virtual SAN deployment. See the Virtual SAN TCO and Sizing Calculator at <https://vsantco.vmware.com/>.

# VMware Infrastructure Planner

Slide 10-16

VMware Infrastructure Planner™ is a virtual appliance that gathers data on a virtual environment and displays potential savings when using a software-defined data center.

<http://vip.vmware.com>



VMware Infrastructure Planner™ accepts input data on a virtual environment and displays a summary of the specific resources that can be saved if deploying VMware vCloud Suite® and other software-defined data center products. These reports are segmented in easy-to-understand categories like compute, storage, and networking, and are backed up by more detailed reports. VMware Infrastructure Planner also provides a high-level estimate of the financial benefits from deploying vCloud Suite.

# Network

Slide 10-17

Consider the following guidelines for the Virtual SAN network:

- Use 10 Gb networks for large environments or all-flash configurations.
- Separate the different traffic types onto different networks.
- In data centers where jumbo frames are already enabled in the network infrastructure, jumbo frames are recommended for the Virtual SAN network.
- Use distributed switches instead of standard switches to take advantage of advanced features such as VMware vSphere® Network I/O Control.

When planning the components of the Virtual SAN cluster, you must consider that the Virtual SAN network activities can saturate a 1 Gb network during rebuild and synchronization operations. A 10 Gb network is recommended for larger Virtual SAN deployments or a Virtual SAN cluster with flash capacity devices.

Separating the traffic types like management, Virtual SAN, and IP storage onto different networks and using shares as a quality of service mechanism sustains the performance expected during possible contention scenarios.

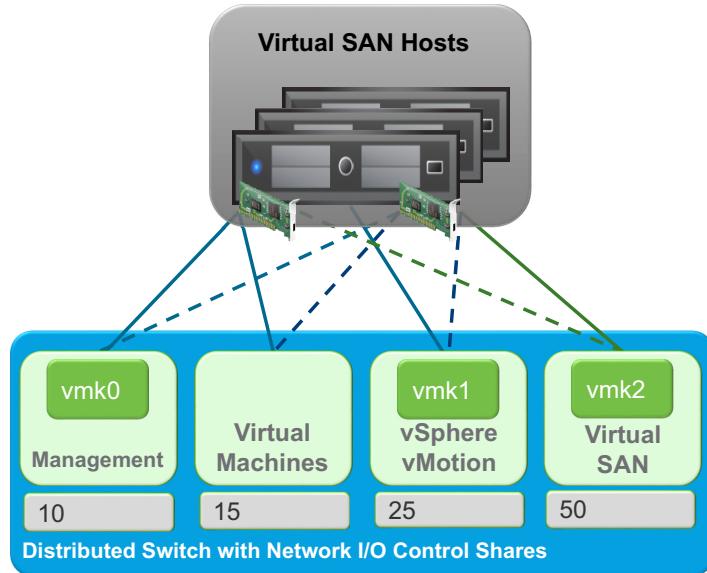
Virtual SAN requires IP multicast to be enabled on the network segment that is used for Virtual SAN communication. Multicast performance is also important. So ensure that a high-quality enterprise switch is used. If a lower-end switch is used, then explicitly test the switch for multicast performance, because unicast performance is not an indicator of multicast performance.

If a data center does not use jumbo frames in the network infrastructure, then jumbo frames are not recommended. The operational cost of configuring jumbo frames throughout the network infrastructure could outweigh the performance benefits.

# Network I/O Control

Slide 10-18

Virtual SAN networking on distributed switches can leverage Network I/O Control for proportional share-based allocation of network bandwidth.



Network bandwidth performance affects host evacuation and rebuild times more than workload performance. Virtual SAN traffic might use 10 Gb physical network adapters that are shared with other system traffic types, such as vSphere vMotion traffic, vSphere HA traffic, virtual machine traffic, and so on. You can use Network I/O Control in vSphere Distributed Switch to guarantee the amount of bandwidth that is required for Virtual SAN.

In Network I/O Control, you can configure reservation and shares for the Virtual SAN outgoing traffic:

- Set a reservation so that Network I/O Control guarantees that minimum bandwidth is available on the physical adapter for Virtual SAN.
- Set shares so that certain bandwidth is available when the physical adapter assigned for Virtual SAN becomes saturated to Virtual SAN and to prevent Virtual SAN from consuming the entire capacity of the physical adapter during rebuild and synchronization operations.

# Virtual SAN Storage Policy Considerations

Slide 10-19

You must understand how storage capabilities affect consumption of storage capacity:

- Use the Storage Consumption Model to understand the impact that certain policy settings have on storage consumption.

**Rule-Set 1**  
Select rules specific for a datastore type. Rules can be based on data services provided by datastore or based on tags.  
The VM storage policy will match datastores that satisfy all the rules in at least one of the rule-sets.

Rules based on data services	VSAN
Number of disk stripes per object	1
Flash read cache reservation (%)	0.0000
Number of failures to tolerate	1
Failure tolerance method	RAID-1 (Mirroring) - Performance
IOPS limit for object	0
Disable object checksum	No
Force provisioning	No
Object space reservation (%)	0

**Storage Consumption Model**

A virtual disk with size 100 GB would consume:  
Storage space  
200.00 GB  
Initially reserved storage space  
0.00 B  
Reserved flash space  
0.00 B

Storage policy rules affect the planning of a Virtual SAN deployment. The virtual machines intended for the Virtual SAN datastore should be evaluated to determine the storage policy rules that are applied to each virtual machine. After determining the storage policy rules, the planning of the Virtual SAN can include the projected capacity usage due to policies and the raw virtual machine storage requirements.

From the Storage Consumption Model, you can review the virtual disk size available for use and the corresponding flash cache and storage capacity. The values shown by this model include the reserved storage space that your virtual machines would potentially consume when you apply the specified storage policies.

# Host Design Considerations: CPU and Memory

Slide 10-20

Host CPU considerations include the following:

- Desired sockets per host
- Desired cores per socket
- Desired number of VMs and thus how many virtual CPUs (vCPUs) are required
- Desired vCPU-to-core ratio
- Provide for a 10 percent CPU overhead for Virtual SAN

Memory considerations include:

- Desired memory for virtual machines
- Number of disks and disk groups to be supported

Compute-only hosts are not recommended as members of a Virtual SAN cluster.

Memory requirements for Virtual SAN are defined based on the number of disks and disk groups that are managed by ESXi. If vSphere hosts have greater memory configurations than 32 gigabytes of RAM, the vSphere hosts can support the maximum disk group and disks configuration supported in Virtual SAN. Virtual SAN is designed to introduce no more than 10 percent of CPU overhead for each host. You must consider this fact in Virtual SAN implementations with high consolidation ratios and CPU-intensive applications requirements.

VMware recommends that you use uniformly configured hosts for Virtual SAN deployments. Although compute-only hosts can exist in a Virtual SAN environment and consume storage from other hosts in the cluster, VMware recommends avoiding unbalanced cluster configurations.

## Host Design Considerations: Storage Considerations

Slide 10-21

Virtual SAN 6.x supports USB, SD, and SATADOM as supported ESXi boot devices.

VMware recommends redirecting logging and traces to persistent storage.

Blade servers are limited in drive slots due to their low profile:

- Virtual SAN 6.x supports external storage options to prevent this limitation.
- Verify if any external storage devices are listed in the VMware Compatibility Guide.

Virtual SAN 6.0 introduced SATADOM as a supported ESXi boot device. When SATADOM devices are used for boot devices, the logs and traces reside in RAM disks that are not persisted during reboots:

- Consider redirecting logging and traces to persistent storage when these devices are used as boot devices.
- VMware does not recommend storing logs and traces on the Virtual SAN datastore. These logs might not be retrievable if Virtual SAN has an issue, which affects access to the Virtual SAN datastore. This issue can hamper any troubleshooting effort.
- To redirect Virtual SAN traces to a persistent datastore, the `esxcli vsan trace set` command can be used. For information, see vSphere Command-Line Interface Documentation at <https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-6-pubs.html>.

To support customers that use blade servers as hosts, Virtual SAN 6.x includes support for approved external storage devices. For details about supported devices, see VMware Compatibility Guide at <http://www.vmware.com/resources/compatibility/search.php>.

## Use Cases

Slide 10-22

vSphere with Virtual SAN supports a wide array of use cases including:

- Virtual Desktop Infrastructures (VDI)
- Management Cluster
- Private Cloud for Testing and Development

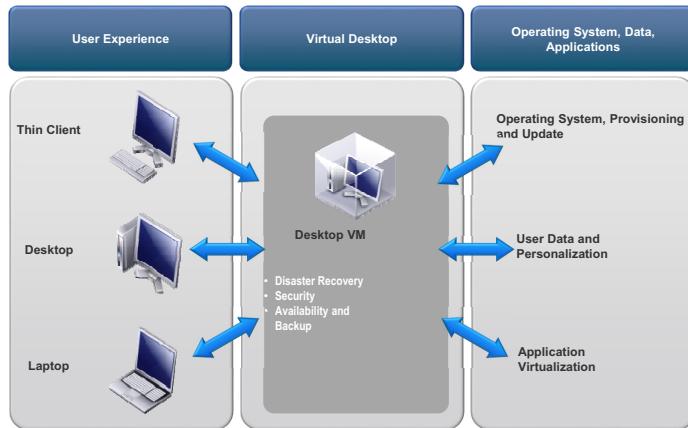
VMware Confidential  
Internal Use Only

# Use Case: Virtual Desktop Infrastructures

Slide 10-23

Virtual SAN provides an efficient and scalable storage platform for VDI:

- Handles peak performance operations, such as boot, login, read/write storms
- Provides seamless and detailed scaling without large upfront investments
- Supports high VDI density

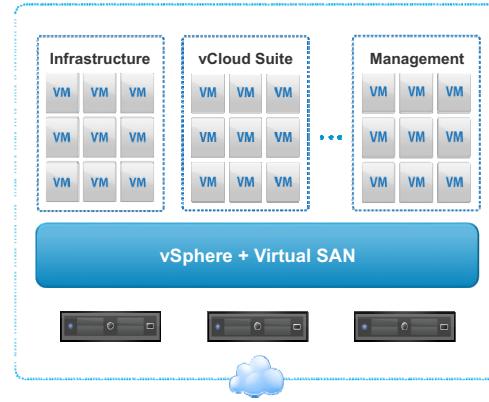


## Use Case: Management Cluster

Slide 10-24

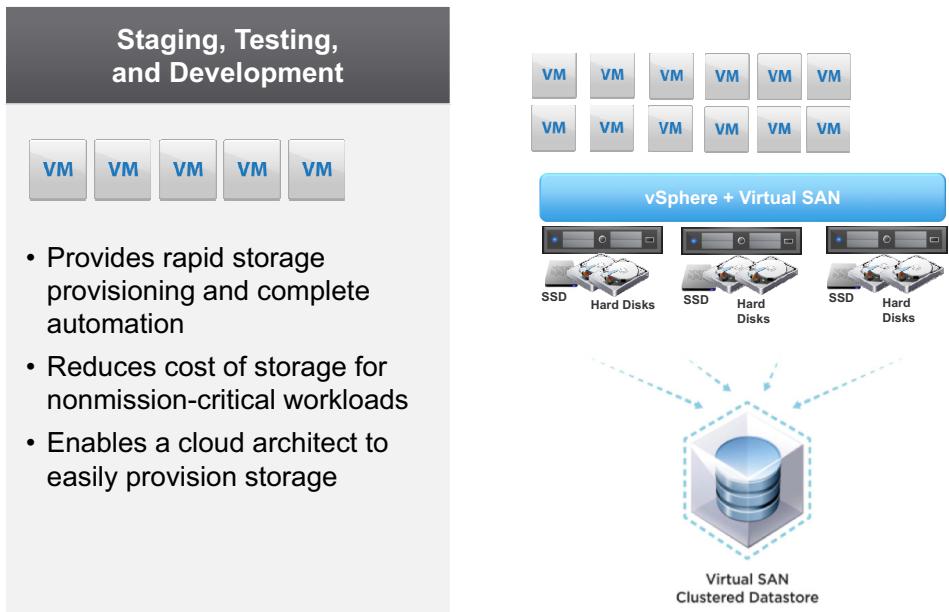
All vSphere management components can be run inside a dedicated Virtual SAN cluster:

- Can be administered by vSphere administrators
- Reduces storage design complexity
- Provides performance and availability capacities to infrastructure-related systems



# Use Case: Private Cloud for Testing and Development

Slide 10-25



## Review of Learner Objectives

Slide 10-26

You should be able to meet the following objectives:

- Understand Virtual SAN design considerations
- Plan and design Virtual SAN clusters
- Identify the design and sizing tools for Virtual SAN
- Describe Virtual SAN use cases

VMware Confidential  
Internal Use Only

## Key Points

Slide 10-27

- You understand requirements of the intended Virtual SAN deployment.
- You follow a design process to develop an appropriate design.
- You understand the different hardware considerations.

Questions?

VMware Confidential  
Internal Use Only

VMware Confidential  
Internal Use Only