DESIGN DOCUMENT

VMware vSAN: DELTA GSS [V7.0]

CONTENTS

Project Characteristics	2
Project Members and Responsibilities	2
Project Description	3
Design Guidelines	4
Course Outline	5
Module: Large Drive Support in vSAN	5
Lesson 1: vSAN Large Capacity Overview	5
Module: Persistent Services Platform	6
Lesson 1: Persistent Services Platform	6
Module: Marking Disks as Capacity Flash	7
Lesson 1: Marking Disks as Capacity Flash	7
Module: Object Delete Workflow Reduced to 1-3 slides	7
Lesson 1: New Object Deletion Workflow	7
Module: Metadata-Only Witness Node	8
Lesson 1: Metadata-Only Witness Node	8
Module: Stretched Cluster site failover based on imbalance (Scope TBD)	9
Lesson 1: Stretched Cluster Site Failover	9
Module: vSAN over RDMA No longer within 8.3 scope removed from training	9
Lesson 1: vSAN over RDMA	9
Module: vSAN File Services [To be scoped by Babu-as per J. Doyle]	10
Losson 1: uSAN File Services	11

PROJECT CHARACTERISTICS

This document details the project to update the VMware Site Recovery Manager™ [V8.3]: DELTA GSS course from version 8.2 to version 8.3. It includes the following:

- Project Members and Responsibilities
- Project Description
- Design Guidelines
- Course Outline

This document is preceded by a signed-off Product Requirements Document (PRD).

Revision	Description	Date
1.0	First draft, ready for review -AR	11/5/2019
1.1	Post review edits - AR	11/11/2019
1.2	Remove vSAN over RDMA, Add vSAN Files Module -AR	11/26/2019
1.3	Add Stretched Cluster Failover Module -AR	12/6/2019

PROJECT MEMBERS AND RESPONSIBILITIES

Stakeholders	Review PRD and datasheet	Jeena Moorkoth
Instructional Designers and Managers	Design and develop the course Manage development work	Dev Manager: Francis Daly Project Manager: Alannah Keena Instructional Designer: Tony Rivas Developer(s): Tony Rivas, Baburaju P Gowda
LEAD INSTRUCTORS	Review all course deliverables Conduct beta classes and provide feedback Lead TTT and create additional trainer support materials Support/lead other trainers	Javier Menendez
SMES	Review design document and beta training versions Answer content-related questions	Pete Koehler; Jesse Horne; Sergio Bernal; Bradley Mott; Sean Quinlan; Kamil Zapert; Brendan Mac Carron; Deepak Nikam; Ramanjaneya Reddy
VDC MANAGER	Manage automation work for VDC Manage Production Ready Test (PRT)	Varun Yoga Suhas B R

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

TECHNICAL EDITOR	Ensure deliverables meet corporate and department standards	Fiza Kapsi
PUBLICATION SPECIALIST	Prepare and release final files to print vendor Assign SKUs Announce course availability	Fiza Kapsi

PROJECT DESCRIPTION

PROJECT TYPE CO BUSINESS REASONS FOR TRAINING V	BD < VMware vSAN™ [V7.0]: DELTA GSS > ourse update (DELTA) Mware vSAN™ [V7.0] has major feature additions and updates. This training rovides product support training needed by GSS TSE.	
BUSINESS REASONS FOR PRINCE TRAINING VI	Mware vSAN™ [V7.0] has major feature additions and updates. This training rovides product support training needed by GSS TSE.	
BUSINESS REASONS FOR PRINCE TRAINING VI	Mware vSAN™ [V7.0] has major feature additions and updates. This training rovides product support training needed by GSS TSE.	
BUSINESS REASONS FOR pi	rovides product support training needed by GSS TSE.	
	'Mware TSE who support VMware customer environments with VMware vSAN™ V7.0].	
COURSE DESCRIPTION po	This OnDemand, hands-on training course builds skills in configuring and performing customer support tasks with vSAN 7.0 features in supported VMware vSAN environments. The course will be delivered as a OnDemand microlearning with assessments and labs.	
COURSE OBJECTIVES	fter completing this course, students should be able to: Explain Large Drive Support in vSAN Assess Large Drive Support deployment considerations Analyze how current or future deduplication configurations affect Large Drive Support Become aware of Healthcheck workflow additions Explain the need for Persistent services platform Examine Supervisor services user interface architecture and integration. Interpret vSAN health check registry for 3rd party Kubernetes objects Define Persistent Volume Claim (PVC) Affinity feature Adopt vSAN maintenance mode interaction with 3rd party Kubernetes operators Outline the need for capacity flash disk marking Explain capacityFlash Tag use Explain tag information storing Identify upgrade use case scenarios Compare differences in the Object deletion Workflow Explain Object deletion enhancements Become aware of PR Fixes for Object deletion Examine the Metadata-only Witness Node Review Metadata-only Witness Node requirements Explain the design and use cases for the Metadata-only Witness Node Contrast RDMA and vSAN over RDMA Diagram the vSAN over RDMA	

	 Explain Lossless traffic configurations Detail a vSAN over RDMA Implementation 	
DELIVERY METHOD	OnDemand microlearning and labs.	
LENGTH	1 day	
LECTURE/LAB RATIO	Approximately 60% lecture and 40% lab	
RELATIONSHIP TO CERTIFICATION PROGRAMS	None	
PRE-ENROLLMENT REQUIREMENTS	Learners attending this course should have at a minimum the following VMware Infrastructure skill: Use of vSphere client and esxcli commands to view virtual machines, datastore, and network states. Previous VMware vSAN support experience.	
INSTRUCTOR REQUIREMENTS	N/A, OnDemand course.	
DESIGN GUIDELINES		
STUDENT KNOWLEDGE REQUIREMENTS	See pre-enrollment requirements.	
GENERAL GUIDELINES	Update the vSAN 6.7 GSS course.	
LAB GUIDELINES	For the on-demand session: 1 student per on-demand session.	

COURSE OUTLINE

Timing: Total course duration

(For lecture and lab timing, see the Lecture and Lab Timings Spreadsheet.)

COURSE DESCRIPTION

This OnDemand, hands-on training course builds skills in configuring and performing customer support tasks with vSAN 7.0 features in supported VMware vSAN environments. The course will be delivered as a OnDemand microlearning with assessments and labs.

COURSE LEARNING OBJECTIVES

- Explain Large Drive Support in vSAN
- Assess Large Drive Support deployment considerations
- · Analyze how current or future deduplication configurations affect Large Drive Support
- Become aware of Healthcheck workflow additions
- Explain the need for Persistent services platform
- Examine Supervisor services user interface architecture and integration.
- Interpret vSAN health check registry for 3rd party Kubernetes objects
- Define Persistent Volume Claim (PVC) Affinity feature
- Adopt vSAN maintenance mode interaction with 3rd party Kubernetes operators
- · Outline the need for capacity flash disk marking
- Explain capacityFlash Tag use
- Explain tag information storing
- Identify upgrade use case scenarios
- Compare differences in the Object deletion Workflow
- Explain Object deletion enhancements
- Become aware of PR Fixes for Object deletion
- · Examine the Metadata-only Witness Node
- Review Metadata-only Witness Node requirements
- Explain the design and use cases for the Metadata-only Witness Node
- Contrast RDMA and vSAN over RDMA
- Diagram the vSAN over RDMA Architecture
- Outline the vSAN over RDMA Workflow
- Explain Lossless traffic configurations
- Detail a vSAN over RDMA Implementation

MODULE: LARGE DRIVE SUPPORT IN VSAN

Timing: Lecture ## minutes, no lab

LESSON 1: VSAN LARGE CAPACITY OVERVIEW

LESSON LEARNING OBJECTIVES

- Explain Large Drive Support in vSAN
- Assess Large Drive Support deployment considerations
- Analyze how current or future deduplication configurations affect Large Drive Support
- Become aware of Healthcheck workflow additions

TOPICS

About Large Drive Support

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

Comment [AR1]: See vSAN Large capacity

overview:

https://confluence.eng.vmware.com/display/VSAN/ vSAN+Large+capacity+overview#vSANLargecapacity overview-TechnicalProblem

- o Why (Customer Issues)
- o Current support (see vSAN large capacity drive support Feature specification)
- Cause (LSOM logical addressing space limitations)
- Solution (LSOM layout is modified to support 64-bit logical space from 32 bit logical space).
- Deployment considerations (Changes to support greater than 16TB logical space needs a new on-disk layout. Furthermore, it is desirable to have same layout within a disk group.)
- Deployment Recommendations
 - (See vSAN Large capacity overview -Proposed deployment solution)
 - (See vSAN large capacity drive support Feature specification)
 - o Also available Greenfield and Brownfield deployment requirements if needed.
 - Heterogeneous disk groups (are recommended)
- Deployment Observations from the field (New format is critical only with 8+TB drives see vSAN Large capacity overview)
- Deduplication limitations (see vSAN large capacity drive support Feature specification)
- Health Check (Health check workflow depends on the Deployment choices see vSAN large capacity drive support Feature specification)
- (Background and additional information see Supporting Large Drives vSAN mgmt.)

LAB EXERCISE: NO LAB RECOMMENDED

(This feature functions dependency are hardware disk size, DG, and deduplication configuration)

Module: Persistent Services Platform

Timing: Lecture ## minutes, no lab

LESSON 1: PERSISTENT SERVICES PLATFORM

LESSON LEARNING OBJECTIVES

- Explain the need for Persistent services platform
- Examine Supervisor services user interface architecture and integration.
- Interpret vSAN health check registry for 3rd party Kubernetes objects
- Define Persistent Volume Claim (PVC) Affinity feature
- Adopt vSAN maintenance mode interaction with 3rd party Kubernetes operators

TOPICS

- Persistence Service
- Persistent services Use case
- Workflow
 - Desired Workflow
 - VmVsanObjectAffinity
 - VmVsanObjectVmVsanObjectAntiAffinity
 - VmVmAntiAffinity
 - OnEnterMaintenanceModeAction
 - OnHostRestartAction
 - o vSAN 7.0 Workflow
- UI architecture and integration
- vSAN health integration
- Shared nothing architecture
- Configuration
- Troubleshooting

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

Comment [AR2]: SEE TOI: <u>vSAN 7.0 GSS</u> TOIS - LARGE CAPACITY DRIVE

See vSAN Large capacity overview

https://confluence.eng.vmware.com/display/VSAN/vSAN+Large+capacity+overview

Comment [AR3]: See vSAN Large capacity

https://confluence.eng.vmware.com/display/VSAN/ vSAN+large+capacity+drive+support+Feature+specification

Comment [AR4]: See Heterogeneous disk

https://confluence.eng.vmware.com/display/VSAN/ Heterogeneous+disk+groups

Comment [AR5]: Supporting Large Drives - vSAN mgmt.

https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=SABU&title=Supporting+Large+ Drives+-+vSAN+mgmt

Comment [AR6]: vSAN persistence services https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=~mwiggers&title=vSAN+persist ence+services

Persistence Service (vSAN + WCP Application Platform)

https://confluence.eng.vmware.com/pages/viewpa ge.action?pageId=411631790

Additional sources:

ence+services

- $\underline{ \bullet_{Functional\ spec}} \ for\ the\ Disable DrsV motion \\ capability;$
- •<u>Functional spec</u> for the VmVmAntiAffinity capability;
- <u>Paper on compute-policy framework</u> presented at RADIO 2018;
- Paper on tags and groups and hierarchical scheduling submitted to RADIO 2019.

Comment [AR7]: vSAN persistence services https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=~mwiggers&title=vSAN+persist ence+services

Comment [AR8]: vSAN persistence services –

https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=~mwiggers&title=vSAN+persist ence+services

Comment [AR9]: vSAN persistence services –
Desired Workflow/vSAN 7.0 Workflow
https://confluence.eng.vmware.com/pages/viewpage.action?spaceKey="mwiggers&title=vSAN+persist">https://confluence.eng.vmware.com/pages/viewpage.action?spaceKey="mwiggers&title=vSAN+persist

Comment [AR10]: Functional spec for the VmVmAntiAffinity capability

Comment [AR11]: Reference: Persistent Services Platform Project PPT

LAB EXERCISE: NO LAB RECOMMENDED

MODULE: MARKING DISKS AS CAPACITY FLASH

Timing: Lecture ## minutes, Lab ## minutes TBD

LESSON 1: MARKING DISKS AS CAPACITY FLASH

LESSON LEARNING OBJECTIVES

- Outline the need for capacity flash disk marking
- Explain capacityFlash Tag use
- · Explain tag information storing
- Identify upgrade use case scenarios

TOPICS

- Why we need to mark disks as capacity flash
 - Current support method (MPP)
 - o New support needs (PSA/capacityflash tags)
- capacityFlash Tag Use
 - o UI update
 - o Escxli
- When capacityFlash tag is removed?
- Behavior changes
- PSA tag information storing
 - o esx.conf
- vSAN tag information storing
 - o ESX ConfigFile
 - Config Store
- Feature switch state
 - VSAN_CapacityFlashTag
- Upgrade Use Case
 - See UpgradeVsanDisksTagInfo

LAB EXERCISE: TBD

Possible lab scenario: See Verification with capacityFlashTag and configstore on

MODULE: OBJECT DELETE WORKFLOW REDUCED TO 1-3 SLIDES

Timing: Lecture ## minutes, no lab

LESSON 1: NEW OBJECT DELETION WORKFLOW

LESSON LEARNING OBJECTIVES

- Compare differences in the Object deletion Workflow
- Explain Object deletion enhancements
- Become aware of PR Fixes for Object deletion

TOPICS

About Object Deletion

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

Comment [AR12]: Before this change, the capacityFlash tag information is removed when marking a SSD disk as Non-SSD, as Explaind in section 2.2.3. In our new solution, we changed this behavior. Since the tag is automatically added for internal use when creating disk group, we are supposed not to keep it after disk group has been decommissioned. It's more reasonable to remove the tag during disk group decommissioning.

Comment [AR13]: For disks marked as capacity flash with this change, their capacity flash information is kept in esx.conf / configstore. But for all flash disks consumed by old version vSAN software bits, when they go through software upgrade, such information will be lost. As a result, we should rebuild for them. One solution is we extract tag information from disk groups and persist then in config store. We will have a UpgradeVsanDisksTagInfo for this purpose in UpgradeVsanConfig.

Comment [AR14]: Verification with capacityFlashTag and configstore on https://confluence.eng.vmware.com/display/~bang/huil/Verification+with+capacityFlashTag+and+configstore+on

Comment [AR15]: Reduced to 1-3 slides as per J Doyle

Note from Engineering: [Richard Nazarath] Delete changes mainly are VSAN internal code improvements.

There is no UI or customer noticeable enhancements which would need any documentation.

I suggest we treat this change as incremental enhancement rather than a standalone feature.

Comment [AR16]: See Object Deletion - Introduction

https://confluence.eng.vmware.com/display/VFB/Object+Deletion

Current Behavior Overview

- Delete Modes
 - Forced Delete Mode
 - Normal Delete Mode
 - Actual deletion and discarded component entries
 - Owner Transfer
 - Entry Persistence Daemon (EPD)
- Current Implementation Issues and Bugs (PRs)
- Object Deletion Improvements
 - o New Delete Workflow
 - Forced Delete Mode
 - Normal Delete Mode
 - Actual deletion and discarded component entries
 - New Owner, Old Components
 - New Component, Old Owner
 - Owner Transfer
 - Entry Persistence Daemon (EPD) Improvements
- New Workflow Bug Fixes
- Future Enhancements

LAB EXERCISE: NO LAB SEE COMMENT

MODULE: METADATA-ONLY WITNESS NODE

Timing: Lecture ## minutes, no lab

LESSON 1: METADATA-ONLY WITNESS NODE

LESSON LEARNING OBJECTIVES

- Examine the Metadata-only Witness Node
- Review Metadata-only Witness Node requirements
- Explain the design and use cases for the Metadata-only Witness Node

TOPICS

- About the metadata-only Witness Node
- Metadata-only Witness Node Requirements
- Metadata-only Witness Node Design
- Design Details and Architecture
 - o DOM, CMMDS, and CLOM
- vSAN Management and DFC Workflow
 - Workflows
 - Deployment of a Metadata node
 - Scale-out Workflow
 - Scale-in Workflow
 - VMC Workflow
- metadata-only Witness Node Use Cases
- Implementation
 - vMODL APIs
 - No UI for this feature

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

Comment [AR17]: See Re-work Object Deletion

- Overview of current behavior

https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=VSAN&title=Rework+Object+Deletion

Comment [AR18]: Lab requested by F. Daly Lab request refuted by J. Doyle

Comment [AR19]: See Metadata Witness Node Functional Spec

https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=SABU&title=Metadata+Witness +Node+Functional+Spec

Comment [AR20]: See Metadata Witness Node Functional Spec section 1

Comment [AR21]: See Metadata Witness Node Functional Spec section 2

Comment [AR22]: See Metadata Witness Node Functional Spec section 3

 $\begin{tabular}{ll} \textbf{Comment [AR23]:} See Metadata Witness Node \\ \textbf{Functional Spec } 3.2.1-3.2.3 \\ \end{tabular}$

https://confluence.eng.vmware.com/pages/viewpa ge.action?spaceKey=SABU&title=Metadata+Witness +Node+Functional+Spec

(DOM local read won't be enabled for data nodes in cluster with metadata node, as there is no plan to support multiple geo locations for a vSAN cluster with metadata node.)

Comment [AR24]: See Metadata Witness Node Functional Spec Section 4 Workflows

Comment [AR25]: See Metadata Witness Node Functional Spec section 5

Comment [AR26]: See Metadata Witness Node Functional Spec section 6

- o CMMDS Data Structures
- ESXCLI changes
- Interoperabilities

Lab Exercise: Deploy the Metadata-Only Witness Node

In this lab, you deploy a metadata-only Witness Node

- 1. (Optional) Prepare the Environment
- 2. X
- 3. X
- 4. X

MODULE: STRETCHED CLUSTER SITE FAILOVER BASED ON IMBALANCE (SCOPE TBD)

Timing: Lecture ## minutes, Lab ## minutes TBD

LESSON 1: STRETCHED CLUSTER SITE FAILOVER

LESSON LEARNING OBJECTIVES

- Outline
- Explain
- Explain
- Identify

TOPICS

- XXX
 - o XXX
- XXX
 - o XXX
 - XXX
- XXX
- XXX

LAB EXERCISE: TBD Possible lab scenario:

Module: vSAN over RDMA No longer within 8.3 scope removed from training

Timing: Lecture ## minutes, no lab

LESSON 1: VSAN OVER RDMA

LESSON LEARNING OBJECTIVES

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

Comment [AR27]: See Metadata Witness Node Functional Spec Section 8

https://confluence.eng.vmware.com/display/SABU/ Metadata+Witness+Node+Functional+Spec

Comment [AR28]: <u>Stretched Cluster site</u> failover based on imbalance https://jira.eng.vmware.com/browse/VSANCORE-599

Currently, Stretched Cluster will only failover VMs to the other site when all hosts at the problem site are offline.
Customers are asking now if we can also failover VMs if capacity fills up at one site (and sufficient capacity is available at the other site). This can occur if there is an imbalanced cluster,

due to Site Affinity VMs as an example.

Feature List

https://VMware.zoom.us/recording/share/yGooz QSSKVsQ1crGuVT50L MISkOgAojjg5NjaCR0 FawlumekTziMw

Owner: VSrinivas Vinnakota Developers meeting recording about SC Failover https://wmware.zoom.us/recording/play/_luXzKzVbJ zgeKQ2KCOKG82aujTLVa0nNgUHewZ_csWHJLVe_V_ MaOJupC2xLVAvg?continueMode=true

Comment [AR29]:

Comment [AR30]: VSAN over RDMA

- Contrast RDMA and vSAN over RDMA
- Diagram the vSAN over RDMA Architecture
- Outline the vSAN over RDMA Workflow
- Explain Lossless traffic configurations
- Detail a vSAN over RDMA Implementation

TOPICS

- About RDMA
- vSAN over RDMA
- vSAN over RDMA Architecture
 - o RDMA in vSphere -Architecture
 - o RDMA Transport in RDT Layer
- vSAN over RDMA Workflow
 - RDMA Connections
 - o RDMA Upgrade
 - vmkinc configuration for vSAN
- Usability
 - o NIC Teaming changes
 - Network Configuration
 - Lossless Layer 2 Network
 - Lossless Layer 3 Network
 - DCBx Health Checks
- Implementation
 - o RDMA Detection
 - o Establishing RDMA Connections
 - o Receive Buffer Memory Allocations
 - Network Requirements
 - o vSAN reconfigure
 - o Enable RDMA
 - o Disable RDMA
 - o Node joining and Node leaving
 - o RDMA Monitoring
- Interoperabilities
 - o vSAN Stretched Cluster and ROBO interoperability
 - o iSCSI support
 - UI Changes
 - Addition to the health UI
 - Addition to vSphere/VSAN UI
- Network Setup Requirements
 - 1. RDMA capable NICs
 - 2. Configure DCB or global pause at the switch ports and in the ESX host.
 - 3. Configure VLAN id for the vSAN network.
 - 4. RoCE Switches

LAB EXERCISE: NO LAB RECOMMENDED

MODULE: vSAN FILE SERVICES [TO BE SCOPED BY BABU-AS PER J. DOYLE]

Timing: Lecture ## minutes, Lab ## minutes TBD

Design Document

VMware vSAN™ [V7.0]: DELTA GSS

Comment [AR31]: See Enable RDMA for better performance & workload consolidation – Business Case

https://confluence.eng.vmware.com/pages/viewpage.action?pageId=298998344

See VSAN RDMA Support Functional Spec –

https://confluence.eng.vmware.com/display/SABU/ VSAN+RDMA+Support+Functional+Spec

Comment [AR32]: See VSAN over RDMA https://confluence.eng.vmware.com/display/SABU, VSAN+over+RDMA

Comment [AR33]: See VSAN RDMA Support Functional Spec – Implementation – RDMA Architecture in vSphere section 5.1 https://confluence.eng.vmware.com/display/SABU/ VSAN+RDMA+Support+Functional+Spec

See vSAN over RDMA Architecture https://confluence.eng.vmware.com/display/SABU/ vSAN+over+RDMA+Architecture

Comment [AR34]: See VSAN RDMA Support Functional Spec – Workflow section 3 https://confluence.eng.vmware.com/display/SABU/ VSAN+RDMA+Support+Functional+Spec

Comment [AR35]: See VSAN RDMA Support Functional Spec – Usability section 4 https://confluence.eng.vmware.com/display/SABU/ VSAN+RDMA+Support+Functional+Spec

Comment [AR36]: See VSAN RDMA Support Functional Spec – Implementation section 5 https://confluence.eng.vmware.com/display/SABU/ VSAN+RDMA+Support+Functional+Spec

Comment [AR37]: See VSAN RDMA Support Functional Spec – Usability section 8 https://confluence.eng.vmware.com/display/SABU/ VSAN+RDMA+Support+Functional+Spec

Comment [AR38]: See VSAN RDMA Support Functional Spec – Setup Requirements section 9.1 * The absence of virtual RDMA driver for ESX means we need a physical setup with RDMA capable nics to do any testing. https://confluence.eng.wmware.com/display/SABU/VSAN+RDMA+Support+Functional+Spec

LESSON 1: VSAN FILE SERVICES

LESSON LEARNING OBJECTIVES

- Outline
- Explain
- Explain
- Identify

TOPICS

- Architecture
 - o XXX
 - o XXX
- Configuring Files Services
 - o XXX
 - o XXX
- Troubleshooting
- Activities

LAB EXERCISE: TBD Possible lab scenario: