



CATE Storage and Backup Engineering
Citigroup Architecture & Technology Engineering

VMware VSI Storage

CSBE-11079-DD

Design Document

Version 1.4.3

©Citigroup Inc. 2015, All Rights Reserved.

This document contains material confidential to Citigroup. The contents are protected by specific employee and contractor agreements regarding confidentiality. Reproduction for distribution outside of Citigroup is prohibited.

1 Version Control

Version	Date	Author	Reviewed	Comments
1.0	15-Dec-2011	Eliot Wilson		Updated general NAS guidelines and added over-commit policy
1.1	22-Dec-2011	Eliot Wilson		Updated NetApp configs
1.2	9-Aug-2012	Eliot Wilson		Updated NetApp config
1.3	10-Aug-2013	Eliot Wilson		Added VNX File Configuration option for Standard Class(BAU and Dev Cloud)
1.4	20-Nov-2013	Eliot Wilson		Added VNX2 array configurations. Integrated LVMC design into this document and added XIV configurations.
1.4.2	19-Nov-2014	Eliot Wilson		Added SDDC Pilot Design. Updated VNX2 designs to use default nthreads and cached mount option.
1.4.3	1-Jun-2015	Eliot Wilson		Added T++ design. Marked NetApp NAS as retired.

2 Legal Notice

Information contained herein is for Internal Use and may be used only for business purposes authorized by Citigroup.

CITI, CITIBANK, CITICORP, CITIGROUP and the Arc Device are trademarks and service marks of Citicorp and are used and registered throughout the world. Citibank, Citicorp, Citigroup and their subsidiaries also claim rights in certain other trademarks and service marks contained in these pages.

3 Contact

Please address any comments or questions regarding the information contained in this document via email to the “*CATE Storage and Backup Engineering” distribution list or alternatively to gdse@citi.com.

4 Latest Versions

It is recommended that the latest version of the document be used from the CATE Storage and Backup Engineering website. Newer versions may include certification on additional platforms, or comments on issues discovered after deployment.

<https://catecollaboration.citigroup.net/domains/platstor/gdse/products/default.aspx>

5 Table of Contents

1	Version Control.....	2
2	Legal Notice.....	2
3	Contact	2
4	Latest Versions.....	2
5	Table of Contents	3
6	Introduction	4
7	Assumptions	4
7.1	Classes of VSI VMs	4
8	Storage Design.....	5
8.1	General Purpose Class VSI Design	5
8.1.1	EMC VNX2 File NAS Solution.....	6
8.1.2	Frequently Asked Questions Regarding the General Purpose VSI Designs	16
8.2	LVMC VSI Designs	17
8.2.1	LVMC Dedicated XIV Array Solution (Strategic)	17
8.2.2	Shared SAN storage array Solution (Legacy)	19
8.3	SDDC Pilot Class VSI Design.....	21
8.3.1	EMC VNX2 File NAS Solution.....	21
8.4	T++ Class VSI Design	25
8.4.1	ScaleIO T++ Solution	25
9	Document Version Control	28
10	Appendix A. Procedure to Increase vSphere Parallel SAN I/O	29
1.1	LUN Queue Depth	29
1.2	Disk.SchedNumReqOutstanding	29
11	Appendix B. Standard VSI NetApp Over-Commitment and Performance Metrics Policy.....	30
11.1	Over-Commitment Policy.....	30
11.1.1	General Guidelines	30
11.1.2	New Aggregates or Additional Disks.....	30
11.1.3	Additional Volume Requests	30
11.1.4	BAU Monitoring	30
11.1.5	Aggregate Assessment and Expansion	31
11.1.6	Performance Metrics.....	32
12	Appendix C. Retired Designs	33
12.1	Retired EMC VNX (1 st Gen) File NAS Solution	33
12.2	Retired EMC VNX File NAS Solution using RAID-6	42
12.3	Retired NetApp NAS Solutions.....	49

6 Introduction

This document provides an updated storage design framework for VMware VSI (Virtual Server Infrastructure) deployments. It provides use cases for selecting either SAN or NAS storage and has multiple configurations to support different levels of scalability. It supplements the standard CATE Solution Guides for the various storage products.

7 Assumptions

The assumptions used to develop this framework are based upon existing deployments throughout Citi and requirements from the global storage and virtualization teams.

7.1 Classes of VSI VMs

These classes of VSI VMs specify the type of VMware environment and indicate the storage design that should be used. General Purpose is deployed on NAS while the two classes of LVMC (Large VM Cluster) are deployed on SAN storage.

VSI Class	Per VM IO Throughput ¹ (avgP95 4KB pages/sec)	Per VM Storage Capacity (GB)	Per VM Average IOPS	Target Consolidation Ratio VM to ESX Server	Other Criteria	Storage Type
General Purpose	< 5,000	< 1,000	25	28:1		NAS
LVMC Non-DB	< 10,000	< 5,000	100	10:1	- Heavy CPU/RAM "IO latency sensitive applications"	SAN
LVMC DB	< 10,000	< 5,000	100	40:1	"IO latency sensitive applications" - SQL Server DB	SAN
SDDC Pilot	Various					NAS
T++	Various					ScaleIO

¹ These per VM 95th percentile statistics should be used to calculate the consolidated average per ESX server. Homogeneous deployments with higher per ESX server IOPS may require custom designs.

8 Storage Design

NOTE: This model presumes that MBR/Filesystem Disk Alignment has been integrated into the VSI build process for Windows 2003 (not required for Windows 2008) and Linux. Performance overhead may impact environment stability if this process has not been followed.

8.1 General Purpose Class VSI Design

The standard class includes general purpose BAU and Dev Cloud deployments.

Standard VSI VM Averages:

- OS partition allocated capacity 40GB
- OS partition used capacity 20GB (50%)

- Data partition allocated capacity 88GB
- Data partition used capacity 53GB (60%)

- Total Used Capacity 73GB

- Average IOPS per VM 25
- Average IOPS size 8K
- Read/Write ratio 40/60

- vSwap volumes per ESX cluster² 2 x 540GB

Only the [EMC VNX2 File NAS](#) is a supported solution for standard class VSI.

Storage from the legacy EMC VNX File NAS and the newer VNX2 File designs can be mixed on an ESX cluster because they used the same over-commitment policy.

² Requirement provided by Global Virtualization team

8.1.1 EMC VNX2 File NAS Solution

This section describes the standard storage designs that should be followed when deploying EMC VNX2 File NAS storage for VSI. If any deviation is necessary please contact CATE Storage Engineering to discuss the exception.

Three different configuration sizes are available depending on the expected scale of deployment.

Size	Maximum VM count	VNX2 Model	VNX2 Data Movers	Virtual Data Movers
Small NAS	160	5400	2 (1+1)	1
Medium NAS	400	5600	2 (1+1)	2
Large NAS	960	5800	3 (2+1)	6

8.1.1.1 Common VNX2 File Design Guidelines

- Follow existing Citi VNX2 File standards unless otherwise indicated in this guide
- Each EMC VNX2 File array should be dedicated for VMware
- Filers should be placed close to the VMware clusters being serviced and not more than 2 network hops away³
- Storage over-commitment will be performed at the VMware ESX layer. There will be no over-commitment on the storage array.
- All filesystems (datastores) will be thickly provisioned on VNX2 File array.
- **Assume 200% capacity over-commitment on VMware, provision only half of the total capacity requested.**(i.e. if 48TB is requested, provision only 24TB or 7 volumes 3,500GB each)
- No compression or de-duplication will be used on the VNX2 File array (EMC recommended that compression not be used with active VMs due to the potential performance impact)
- Snapshots are not created for any of the NFS volumes used with VMware
- Backups of VSI VM systems are performed within the VM itself using standard tools.
- The EMC VNX2 File storage designs have been defined to use modular expansion. The initial block supports a certain capacity and estimated VM count. Additional storage blocks can be added to increase capacity and/or VM count.
- Two VNX2 block storage pool should be used with balanced SP ownership of LUNs
- Virtual Data Movers (VDM) should be implemented to help challenges surrounding large scale migration of guests when a storage refresh is due.
 - 5 file systems per VDM
 - A maximum of 1 vSwap volume per ESX cluster per VDM to limit risk⁴

³ See [FAQ section](#) for the applicable definition of 2 network hops

⁴ The small 5400 configurations will place 2 vSwap volumes in a single VDM

- Use the default value for NFS threads (nthreads value will be based on the memory in the data mover).
- Use the cached mount (default) option with VNX OE for File 8.x.
- To capture backend performance statistics are collected, ensure that Unisphere Analyzer is licensed and enable “Periodic Archiving” without the “stop automatically” parameter checked (Statistics for Block -> Performance Data Logging)

8.1.1.2 Guidelines for Virtualization Teams

- Datatstores will be ~3446GB⁵ in size
- All VMDKs should be thinly provisioned on ESX
- For all deployments over-commitment of storage capacity should occur at the ESX layer
- To simplify management of storage over-commitment VMware Storage DRS (SDRS) should be used for initial placement and auto-balancing based on capacity
- On average each file system should have a maximum of 40 VMs provisioned to align with the performance design and the expected 200% capacity over-commitment
- All VMDK volumes need to be properly block aligned.
- Set the disk timeout value to at least 60 seconds⁶ in the guest OS

8.1.1.3 General Purpose VSI VNX2 5400 Design

The VNX2 5400 is a small entry level configuration with limited connectivity and capacity. Its scales up to support 160 VMs. Existing VNX2 File standards should be followed unless otherwise noted.

- Network connections for NFS can be 1GbE or preferably 10GbE, both with active/passive failover (aka “fail-safe networking” FSN):
 - 1GbE networks will require 8 connections
 - 4 per DM configured as 2 LACP pairs: FSN (LACP (1,2) + LACP (1,2))

Network Requirements VNX2 5400 w/ 1GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	1GbE Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	8	1	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	15	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	8	16	

⁵ Due to filesystem overhead, the usable capacity is 98.4% of the 3500GB volume size

⁶ This is set by default when VMware tools are installed

- 10GbE networks will require 4 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements VNX2 5400 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	4	1	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	15	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	4	16	

8.1.1.3.1 General Purpose VSI VNX2 5400 Storage Blocks

The VNX2 can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5400 with 2(1+1) DMs, 600GB 10K 2.5" disks, and Fast Cache							
Config	Max. ⁷ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+0 (4+4) Groups
				Vault Disks	Spare Disks	Pool Disks	
1	80	3	38	4	2	32	4
2	160	3	71	4	3	64	8

Storage Pools and Filesystems							
Config	ESX Clusters Supported	VNX2 Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	1	#1	2	4,105	1	1	98%
		#2	2	4,105	1	1	98%
2	1	#1	4	8,210	2	1	92%
		#2	4	8,210	2	1	92%

Virtual Data Mover Assignments					
Config	DM	VNX2 Block Storage Pool	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	#1	2	2
2	server_2	#1	#1	4	2

⁷ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

8.1.1.3.2 General Purpose VSI VNX2 5400 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures. The capacity expansion should be evenly split across storage pools.

Config	Additional Spare Disks	Additional 600GB disks	Additional RAID1+0 (4+4) Groups
2a	0	16	2

8.1.1.3.3 General Purpose VSI VNX2 5400 Space and Power

	5400 with 2(1+1) DMs		
Config	Power kVA	BTU	RU
1	1.58	4,700	11
2	1.88	5,700	13

8.1.1.4 General Purpose VSI VNX2 5600 Design

The VNX2 5600 array scales up to support 400 VMs.

- Network connections for NFS can be 1GbE or preferably 10GbE, both with active/passive failover (aka “fail-safe networking” FSN):
 - 1GbE networks will require 16 connections
 - 8 per DM configured as 2 LACP pairs: FSN (LACP (1,2,3,4) + LACP (1,2,3,4))

Network Requirements VNX2 5600 w/ 1GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	1GbE Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	16	2	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	14	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	16	16	

- 10GbE networks will require 4 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements VNX2 5600 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	4	2	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	14	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	4	16	

8.1.1.4.1 General Purpose VSI VNX2 5600 Storage Blocks

The VNX2 can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5600 with 2(1+1) DMs, 600GB 10K 2.5" disks, and Fast Cache							
Config	Max. ⁸ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+0 (4+4) Groups
				Vault Disks	Spare Disks	Pool Disks	
1	240	5	103	4	3	96	12
2	320	5	136	4	4	128	16
3	400	5	168	4	4	160	20

Storage Pools and Filesystems							
Config	ESX Clusters Supported	VNX2 Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	1	#1	6	12,315	3	1	90%
		#2	6	12,315	3	1	90%
2	2	#1	8	16,420	4	2	92%
		#2	8	16,420	4	2	92%
3	2	#1	10	20,525	5	2	91%
		#2	10	20,525	5	2	91%

Virtual Data Mover Assignments					
Config	DM	VNX2 Block Storage Pool	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	#1	3	1
		#2	#2	3	1
2	server_2	#1	#1	4	1-2
		#2	#2	4	1-2
3	server_2	#1	#1	5	1-2
		#2	#2	5	1-2

⁸ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

8.1.1.4.2 General Purpose VSI VNX2 5600 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures. The capacity expansion should be evenly split across storage pools.

Config	Additional Spare Disks	Additional 600GB disks	Additional RAID1+0 (4+4) Groups
3a	0	16	2

8.1.1.4.3 General Purpose VSI VNX2 5600 Space and Power

5600 with 2(1+1) DMs			
Config	Power kVA	BTU	RU
1	2.16	6,500	15
2	2.54	7,600	19
3	2.81	8,500	21

8.1.1.5 General Purpose VSI VNX2 5800 Design

The VNX2 5800 array scales up to support 960 VMs.

- Network connections for NFS must be 10GbE with active/passive failover (aka “fail-safe networking” FSN):
 - 10GbE networks will require 6 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements VNX2 5800 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
3 Datamovers (DM)	0	0	6	6	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	10	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	6	16	

8.1.1.5.1 General Purpose VSI VNX2 5800 Storage Blocks

The VNX2 can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5800 with 3(2+1) DMs, 600GB 10K 2.5” disks, and Fast Cache							
Config	Max. ⁹ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+5 (4+4) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
1	640	7	249	4	5	240	30
2	720	7	282	4	6	272	34
3	960	9	363	4	7	352	44

⁹ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

Storage Pools and Filesystems							
Config	ESX Clusters Supported	VNX2 Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	3	#1	15	30,788	8	3	96%
		#2	15	30,788	8	3	96%
2	3	#1	17	34,893	9	3	95%
		#2	17	34,893	9	3	95%
3	3	#1	22	45,155	12	3	97%
		#2	22	45,155	12	3	97%

Virtual Data Mover Assignments					
Config	DM	VNX2 Block Storage Pool	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	#1	4	1-2
		#1	#2	4	1
	server_3	#2	#3	4	1-2
		#2	#4	4	1
2	server_2	#1	#1	5	1-2
		#1	#2	4	1
	server_3	#2	#3	5	1-2
		#2	#4	4	1
3	server_2	#1	#1	4	1-2
		#1	#2	4	1
		#1	#3	4	0
	server_3	#2	#4	4	1-2
		#2	#5	4	1
		#2	#6	4	0

8.1.1.5.2 General Purpose VSI VNX2 5800 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures. The capacity expansion should be evenly split across storage pools.

Config	Additional Spare Disks	Additional 600GB disks	Additional RAID1+0 (4+4) Groups
3a	0	16	2

8.1.1.5.3 General Purpose VSI VNX2 5800 Space and Power

5800 with 3(2+1) DMs			
Config	Power kVA	BTU	Used RU
1	3.81	11,500	29
2	4.19	12,600	33
3	4.99	15,100	39

8.1.2 Frequently Asked Questions Regarding the General Purpose VSI Designs

1. What data was used to develop these designs?

- Requirements from the CTI Virtualization and NAS Global Teams
- CATE Storage Solution Certification documents
- Performance analysis by NetApp of existing NAM VSI deployments and FlashCache planning tools
- VMware and storage vendor best practices

2. What does “2 network hops” between ESX and Filer mean?

One hop is trace route from the NAS device and the first reply is from the ESX Cluster member itself on the same subnet and L3 switch pair.

1 Hop

```
ldnctinas501a> traceroute ldnctinassa161
traceroute to ldnctinassa161.eur.nsroot.net (169.182.241.57), 30 hops max, 40 byte packets
 1 ldnctinassa161.eur.nsroot.net (169.182.241.57) 0.000 ms 7.000 ms 6.000 ms
ldnctinas501a>
```

Two hops would involve one hop up to the router and back down again to either the same L3 switch pair on a different subnet or onto an adjacent L3 switch pair with a different subnet.

2 Hops

```
ldnctinas601a> traceroute ldnctinassa161
traceroute to ldnctinassa161.eur.nsroot.net (169.182.241.57), 30 hops max, 40 byte packets
 1 rgrbrdc-gss08-6509-vl187.eu.ssmb.com (169.182.187.4) 0.000 ms 0.000 ms 0.000 ms
 2 ldnctinassa161.eur.nsroot.net (169.182.241.57) 0.999 ms 0.000 ms 0.000 ms
ldnctinas601a>
```

8.1.2.1 EMC VNX2 File Design Specific FAQ

3. Why was over-commitment moved to the VMware layer with the VNX designs?

VNX File lacks true thin provisioning (only volume auto-grow) so it was not possible to maintain array side over-commitment. Storage DRS should be used for initial placement and capacity balancing of thin VMDKs across thick filesystems.

4. Why were larger volumes used with the VNX2 design?

The larger 3,500GB volumes were used to provide larger pools of storage on VMware and also to make the VNX2 volumes easily recognizable to ESX admins so that over-commitment can be managed properly.

5. Why weren't more capacity expansion options given for each array type?

The expanded configurations shown are close to the maximum disk quantity supported for each VNX2 model. The shift of over-commitment to the VMware layer reduces the necessity to expand the storage arrays for capacity growth.

6. Why weren't other configurations with more DMs used with the smaller arrays?

The selected configurations provide scalability of performance and capacity up to the limits of the selected VNX2 model. Additional DMs wouldn't have provided any additional scalability due to other component limitations.

8.2 LVMC VSI Designs

The LVMC Class storage designs can be used for both DB and non-DB deployments that meet these criteria:

LVMC Non-DB VSI VM Averages:

- OS partition allocated capacity 40GB
- OS partition used capacity 20GB (50%)
- Data partition allocated capacity 1500GB
- Data partition used capacity 1125GB (75%)
- Average IOPS per VM 100
- Average IOPS size 8K

LVMC DB VSI VM Averages:

- OS partition allocated capacity 40GB
- OS partition used capacity 20GB (50%)
- Data partition allocated capacity 1500GB
- Data partition used capacity 1125GB (75%)
- Average IOPS per VM 100
- Average IOPS size 8K

The current strategic design for LVMC is using dedicated XIV storage arrays on shared FC SAN fabrics. Legacy configurations using shared storage arrays (VMAX, HDS VSP, and XIV) are also still support but new deployments should use the strategic design.

NOTE: Mixing dedicated XIV storage from the strategic design and the prior legacy shared array design is not recommended because of complexities in managing and reporting on multiple storage and over-commitment policies for a single ESX cluster.

8.2.1 LVMC Dedicated XIV Array Solution (Strategic)

8.2.1.1 Prerequisites

- CATE certified versions of ESX v5.x and ESX SAN Stack
- XIV storage array dedicated to LVMC VMware cluster deployments
- Shared FC SAN Fabrics

8.2.1.2 VMware Configuration Guidelines

1. Each ESX host should have 4 HBA ports which will provide 12 paths per LUN (3:1 zoning) and the active/active round-robin path policy should be used with the standard LUN queue depth¹⁰ and matching Disk.SchedNumReqOutstanding settings.
2. Virtual disks exceeding 2TB should be provided using multiple smaller VMDKs (up to 1TB in size). The volume manager on the guest VM will concatenate the multiple disks into a single logical disk as described in Citi SAN Stack Solution Guides

¹⁰ If queue depth monitoring indicates frequent full queues, this can be raised following best practices

3. LUNs of 1.4TB (for DB transaction logs only) and 5TB (all other data) capacity should be requested based on the expected requirements
4. Dedicated vSwap volumes will not be used and swap files will be stored with each VM.
5. Storage over-commitment will be performed at the VMware ESX layer. There will be no over-commitment on the storage array.
6. LUNs must be properly cache aligned at the ESX and VM layer
7. To simplify management of storage over-commitment VMware Storage DRS (SDRS) should be used for initial placement and auto-balancing based on capacity
8. SDRS functionality with IO latency metrics will not be used for placement or load balancing.
9. A SDRS Datastore cluster should be configured using the 5TB LUNs
10. Backup/Restore infrastructure should be sized to match requirements
11. VAAI support for XIV is integrated into ESX v 5.x and will be used by default
12. Other applicable best practices from CATE Storage and Platform Engineering should be followed

LVMC DB ESX Cluster Specific Guidelines

13. DB vendor best practice guidelines for separate disks should be interpreted as a requirement for separate VMDK virtual disks.
14. Separate SDRS Datastore Clusters should be created with the 1.4TB and 5TB LUNs to provide isolation of DB transaction logs.
15. Database VMs should have multiple VMDKs allocated:
 - For DB VMs, SDRS VMDK Anti-Affinity Rules must be specified for the VM indicating that the DB, TempDB, and Backup Target VMDKs should not be placed on the same datastore.
 - Each DB VM must be provisioned at least 5 VMDK virtual disks as required and the DBA must utilize them as identified for:

Virtual Disk	Placement
OS	5TB Datastore Cluster
DB	5TB Datastore Cluster
TempDB	5TB Datastore Cluster
Transaction Logs	1.4TB Datastore Cluster
Backup Target (Recovery)	5TB Datastore Cluster

16. Other applicable best practices from CATE Storage, Platform and DB Engineering should be followed

8.2.1.3 Storage Guidelines

- XIV storage arrays dedicated for LVMC VMware clusters on shared FC SAN fabrics will be used. LUNs should be provided in two different sizes depending on the requirements:
 - ~5TB LUNs:
 - Base10: 5497 GB
 - Base2: 5128 GB
 - ~1.4TB LUNs only for DB Transaction Logs:
 - Base10: 1546 GB
 - Base2: 1442 GB
 - ESX hosts should be cabled across multiple edge SAN switches to avoid potential ISL contention and to limit the failure domain
- ESX hosts will have 4 HBA ports which should be zoned 3:1 with storage array ports. Consult the [XIV Solution Guide](#) Section 11.5 *VMware Zoning* for specific zoning guidelines.
- A single large XIV storage pool that is defined as a “regular pool” (not thin) should be created.

- Storage over-commitment will be performed at the VMware layer and there should be no over-commitment of capacity on the XIV storage array.
- Other applicable best practices from CATE Storage, Platform and DB Engineering should be followed
- Consult the [XIV Solution Guide](#) for details on configuring an XIV array including use of the VMware vCenter storage plug-in to permit VMware admins to provision storage LUNs.

8.2.2 Shared SAN storage array Solution (Legacy)

8.2.2.1 Prerequisites

- CATE certified versions of ESX and ESX SAN Stack
- CATE certified shared SAN arrays (EMC VMAX, HDS VSP, or IBM XIV)
- Shared FC SAN Fabrics

8.2.2.2 VMware Configuration Guidelines

1. Each ESX host should have 4 HBA ports –an exception exists for general purpose class VSI servers with lower IO workloads which can have just 2 HBA ports
2. An active/active round-robin path policy should be used with **a custom LUN queue depth¹¹ and matching Disk.SchedNumReqOutstanding setting**. See [Appendix A](#) for the recommended LUN queue depth.
3. Virtual disks exceeding 500GB should be provided using multiple smaller VMDKs (up to 500GB in size). The volume manager on the guest VM will concatenate the multiple disks into a single logical disk as described in Citi SAN Stack Solution Guides.
4. LUNs of 1.4TB and 500GB capacity should be requested based on the expected requirements. Either size can also be used for ESX vSwap depending on the requirement.
5. Over commitment of thinly provisioned storage capacity will be done and managed on the storage array only
6. LUNs must be properly cache aligned at the ESX and VM layer
7. Other applicable best practices from CATE Storage, Platform and DB Engineering should be followed
8. Backup/Restore infrastructure should be sized to match requirements
9. Database VMs should have multiple VMDKs allocated:
 - DB vendor best practice guidelines for separate disks should be interpreted as a requirement for separate VMDK virtual disks.
 - Each DB VM must be provisioned at least 5 VMDK virtual disks as required and the DBA must utilized them as identified for:
 1. OS
 2. DB
 3. TempDB¹²
 4. Transaction Logs
 5. Backup Target (Recovery)
 - VMDKs must be laid out on datastores/LUNs following these guidelines:
 1. Transaction Logs must be placed on a 500GB LUN from the “transaction log” pool
 2. OS, DB, and TempDB must be placed on 1 or more 1.4TB LUNs from the “general” pool
 3. Backup Target placed on a different 1.4TB LUN than the DB
 - ESX Clusters for DB consolidation should be limited to 10 ESX hosts and an absolute maximum storage capacity of 320TB (192 x 1.4TB and 64 x 500GB LUNs) .

¹¹ If queue depth monitoring indicates frequent full queues, this can be raised following best practices

¹² Optional if the DBA indicates that the existing layout includes TempDB placed on the same drive as SQL DB

8.2.2.3 Storage Guidelines

- Backend SAN storage will be Enterprise Silver or Gold¹³ tier (EMC Symmetrix, HDS Tagmstore, or IBM XIV) implemented on a widely striped thin pools adhering to CATE standards. LUNs will be provided in 2 different sizes depending on the requirements:
 - ~1.4TB LUNs - spread widely across segments
 - EMC Symmetrix: 3025923840¹⁴ block 6-way meta-LUN
 - HDS Tagmstore: 1477501 MB
 - IBM XIV: 3025935360 blocks
 - ~500GB LUNs for DB Logs - spread widely across segments
 - EMC Symmetrix: 1008641280¹⁵ block 2-way meta-LUN
 - HDS Tagmstore: 492500 MB
 - IBM XIV: 1008645120 blocks
 - For EMC and HDS, LUNs should be provisioned as widely as possible across front end port pairs (segments)
 - ESX hosts should be cabled across multiple edge SAN switches to avoid potential ISL contention and minimize the failure domain
- SAN Zoning should provide a maximum of 4 paths per LUN to maximize storage scalability with the 1024 path limit per ESX server:

HBA Ports per ESX Server	Storage Ports per provisioning segment	Zoning
4	2	2:1
4	4	1:1
2	2	1:1
2	4	1:2

- For IBM XIV, each ESX host should be zoned to 4 array ports on 4 different Interface Modules. With 4 HBA ports per ESX cluster this will be a 1:1 zoning configuration (1:2 with 2 HBA ports). Consult the [XIV Solution Guide VMware Zoning](#) section for specific zoning guidelines.
- Over commitment of thinly provisioned storage capacity will be done and managed on the storage array only
- Other applicable best practices from CATE Storage, Platform and DB Engineering should be followed

¹³ Should be considered for critical highly latency sensitive applications

¹⁴ If auto-metavolume is not used to create the volume then the size is 3025935360 blocks

¹⁵ If auto-metavolume is not used to create the volume then the size is 1008645120 blocks

8.3 SDDC Pilot Class VSI Design

The SDDC Pilot class is for pilot deployments of SDDC architecture.

This class includes two different tiers of storage which can be selected during deployment of VMs within the environment.

NOTE: This design is only for the initial SDDC Pilot deployment and future production deployments will use a different design.

SDDC Pilot VSI VM Averages:

- OS partition allocated capacity 40GB
- OS partition used capacity 20GB (50%)

- Data partition allocated capacity 88GB
- Data partition used capacity 53GB (60%)

- Total Used Capacity 73GB

- Average IOPS size 8K
- Read/Write ratio 40/60

- Silver Storage Tier
 - Average IOPS per VM 25

- Archive Storage Tier
 - Average IOPS per VM 5

8.3.1 EMC VNX2 File NAS Solution

This section describes the standard storage designs that should be followed when deploying EMC VNX2 File NAS storage for VSI. If any deviation is necessary please contact CATE Storage Engineering to discuss the exception.

Three different configuration sizes are available depending on the expected scale of deployment.

Size	Maximum VM count	VNX2 Model	VNX2 Data Movers	Virtual Data Movers
5800 Design	640	5800	3 (2+1)	6

8.3.1.1 Common VNX2 File Design Guidelines

- Follow existing Citi VNX2 File standards unless otherwise indicated in this guide
- Each EMC VNX2 File array should be dedicated for VMware
- Filers should be placed close to the VMware clusters being serviced and not more than 2 network hops away¹⁶

¹⁶ See [FAQ section](#) for the applicable definition of 2 network hops

- Storage over-commitment will be performed at the VMware ESX layer There will be no over-commitment on the storage array.
- All filesystems (datastores) will be thickly provisioned on VNX2 File array.
- No compression or de-duplication will be used on the VNX2 File array (EMC recommended that compression not be used with active VMs due to the potential performance impact)
- Snapshots are not created for any of the NFS volumes used with VMware
- Backups of VSI VM systems are performed within the VM itself using standard tools.
- VNX2 block storage pools should be used with balanced SP ownership of LUNs (The silver SAS disks should be configured into 2 storage pools, NL-SAS into a single pool due to disk qty limitation)
- Virtual Data Movers (VDM) should be implemented to help challenges surrounding large scale migration of guests when a storage refresh is due.
 - Up to 5 file systems per VDM
- Use the default value for NFS threads (nthreads value will be based on the memory in the data mover).
- Use the cached mount (default) option with VNX OE for File 8.x.
- To capture backend performance statistics are collected, ensure that Unisphere Analyzer is licensed and enable "Periodic Archiving" without the "stop automatically" parameter checked (Statistics for Block -> Performance Data Logging)

8.3.1.2 Guidelines for Virtualization Teams

- vCenter and vSphere 5.5 should be used
- Datastores will be sized at ~3,446GB¹⁷ for silver tier and ~9TB¹⁸ for Archive tier
- Datastores will be provisioned to and shared by all SDDC ESX clusters managed by the same vCenter
- All VMDKs should be thinly provisioned on ESX
- Over-commitment of storage capacity should occur at the ESX layer
- To simplify management of storage over-commitment VMware Storage DRS (SDRS) should be used:
 - Initial placement and auto-balancing based on capacity and IO Latency
 - Separate Datastore Clusters (DSC) should be created for each storage tier (silver and archive)
 - With vCenter 5.5, for each DSC set the advanced option
SetSpaceLoadToDatastoreUsedMB=0
- All VMDK volumes need to be properly block aligned.
- Set the disk timeout value to at least 60 seconds¹⁹ in the guest OS

¹⁷ Due to filesystem overhead, the usable capacity is ~98.4% of the 3500GB volume size

¹⁸ Due to filesystem overhead, the usable capacity is ~98.4% of the 9.15TB volume size

¹⁹ This is set by default when VMware tools are installed

8.3.1.3 SDDC Pilot VNX2 5800 Design

The VNX2 5800 array scales up to support 640 VMs.

- Network connections for NFS must be 10GbE with active/passive failover (aka “fail-safe networking” FSN):
 - 10GbE networks will require 6 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements VNX2 5800 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
3 Datamovers (DM)	0	0	6	6	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	10	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	6	16	

8.3.1.3.1 General Purpose VSI VNX2 5800 Storage Blocks

The VNX2 can have the following configurations to support the indicated number of VMs

5800 with 3(2+1) DMs, 600GB 10K 2.5” disks, 3TB 7.2K 3.5” disks and Fast Cache							
Config	Max. ²⁰ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+0 (4+4) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
SDDC	640	7	249	4	5	240	30
Config	Max. VMs Supported		3TB Disks	3TB Disk Assignment			Qty of RAID6 (14+2) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
SDDC	n/a		18	n/a	2	16	1

²⁰ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

600GB Storage Pools and Filesystems – “Silver Tier”						
Config	ESX Clusters Supported	VNX2 Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	Storage Pool Capacity Utilization
SDDC	3	#1	15	30,788	8	96%
		#2	15	30,788	8	96%

3TB Storage Pools and Filesystems – “Archive Tier”						
Config	ESX Clusters Supported	VNX2 Block Storage Pool	Qty of RAID6 (14+2) Groups in Pool	Usable Capacity (GB)	9.15TB file systems	Storage Pool Capacity Utilization
SDDC	n/a	#3	1	38,486	4	99%

Virtual Data Mover Assignments					
Config	DM	VNX2 Block Storage Pool	VDM	3,500GB file systems	9.15TB file systems
SDDC	server_2	#1	#1	4	
		#1	#2	4	
		#3	#3		2
	server_3	#2	#4	4	
		#2	#5	4	
		#3	#6		2

8.4 T++ Class VSI Design

The T++ class is for deployments of T++ pods with ScaleIO storage.

This class includes three different tiers of storage which can be selected during deployment of VMs within the environment.

T++ Storage Tiers:

For planning, consider these performance specifications of each storage tier:

- SSD Performance Storage Tier
 - Average IOPS per TB 25,000
- SAS Standard Storage Tier
 - Average IOPS per TB 500
- NL-SAS Capacity Storage Tier
 - Average IOPS per TB 30

8.4.1 ScaleIO T++ Solution

This section describes the standard storage designs that should be followed when deploying EMC ScaleIO storage on T++ pods for CitiCloud VMware. If any deviation is necessary please contact CATE Storage Engineering to discuss the exception.

8.4.1.1 ScaleIO Design Guidelines

- Follow existing Citi ScaleIO T++ standards unless otherwise indicated in this guide
 - [ScaleIO Product Page](#)
- SAS storage pools used for VMware workloads should be dedicated to VMware
- SSD and NL-SAS pools may be shared with non-VMware workloads
- Storage over-commitment will be performed at the VMware ESX layer. There will be no over-commitment on the storage array.
- All LUNs (datastores) will be thickly provisioned on ScaleIO.

8.4.1.2 Guidelines for VMware Teams

- vCenter and vSphere 5.5 or higher should be used
- Each ESX host should have two 10Gb connections dedicated for ScaleIO storage and configured as:
 - The connections should be on different adapters
 - Each vmnic should be assigned to its own standard switch and VMkernel port group with unique IP subnet
 - Each VMkernel adapter should have MTU set to 8800
 - To verify the IP settings and MTU use these commands:

```
# esxcli <conn_options> network ip interface ipv4 get
# esxcli <conn_options> network ip interface list
```

- Confirm each storage network connection is up by ping'ing the gateway/router and confirming "Packet Lost" is zero. **Do install or configure the ScaleIO SDC client until this is verified.**

- For example, the initial GTDC T++ deployment:

GTDC Compute Pod #1 hosts use:

```
# esxcli <conn_options> network diag ping -W=1 -c=1 -H 10.123.0.1
# esxcli <conn_options> network diag ping -W=1 -c=1 -H 10.123.64.1
```

GTDC Compute Pod #2 hosts use:

```
# esxcli <conn_options> network diag ping -W=1 -c=1 -H 10.123.0.65
# esxcli <conn_options> network diag ping -W=1 -c=1 -H 10.123.64.65
```

- Create a static route for the corresponding storage supernet across each connection's gateway/router. **Do not install or configure the ScaleIO SDC client until the correct static routes are in place.**

- For example, the initial GTDC T++ deployment:

GTDC Compute Pod #1 hosts use:

```
# esxcli <conn_options> network ip route ipv4 add -g 10.123.0.1 -n
10.123.32.0/19
# esxcli <conn_options> network ip route ipv4 add -g 10.123.64.1 -n
10.123.96.0/19
```

GTDC Compute Pod #2 hosts use:

```
# esxcli <conn_options> network ip route ipv4 add -g 10.123.0.65 -n
10.123.32.0/19
# esxcli <conn_options> network ip route ipv4 add -g 10.123.64.65 -n
10.123.96.0/19
```

- To display the host routing table to confirm the entries use:

```
# esxcli <conn_options> network ip route ipv4 list
```

- Install the ScaleIO ESX SDC client as described in the [ScaleIO v1.31 Solution Guide](#)
- LUNs/Datatstores sized and configured as:

Tier	LUN Size	VMware SIOC	ScaleIO IOPS Throttling per TB ²¹	ScaleIO Bandwidth Throttling per TB
SSD Performance	2.5TB (2,560GB)	Enabled@5ms	50,000 IOPS	1,600 MB/sec
SAS Standard	5TB (5,120GB)	Enabled@30ms	1,000 IOPS	32 MB/sec
NL-SAS Capacity	11TB (11,264GB)	Enabled@50ms	60 IOPS	8 MB/sec

²¹ IOPS and bandwidth throttling will be enabled on the ScaleIO cluster when LUNs are provisioned

- Datastores will be provisioned to and shared by all ESX clusters in a Mobility Domain
 - Members of a Mobility Domain must be managed by the same vCenter
 - Due to VMFS limits, the maximum number of hosts per Mobility Domain is 64
- Datastore names should include the storage tier number (see below) to identify them
- All VMDKs should be thinly provisioned on ESX
- Over-commitment of storage capacity should occur at the ESX layer
- To simplify management of storage over-commitment VMware Storage DRS (SDRS) should be used:
 - Initial placement and auto-balancing based on capacity and IO Latency
 - Separate Datastore Clusters(DSC) should be created for each storage tier in each Mobility Domain
 - Up to 64 datastores per datastore cluster
 - Name each DSC with the tier following this template:
 - Tier[tier number]_DSC[datastore cluster number] (e.g. Tier2_DSC1)
 - ScaleIO Tier numbers are (0=SSD, 2=SAS, 4=NL-SAS)
 - With vCenter 5.5, for each DSC set the advanced option
SetSpaceLoadToDatastoreUsedMB=0
- All VMDK volumes need to be properly block aligned.
- Set the disk timeout value to at least 60 seconds²² in the guest OS
- Each provisioned ScaleIO LUN must have its ESX queue depth increased to 256 using this command:


```
# esxcli storage core device set -d <device name> -O 256
```

 - Verify the device queue depth with this command:


```
# esxcli storage core device list -d <device name>
```

²² This is set by default when VMware tools are installed

9 Document Version Control

This section contains a detailed list of the updates made during each document release. Section 1 is purely an overview.

Version	Changes
0.5	Initial draft – general peer review
0.6	Revised NetApp Common Guideline equation for calculating volume count
1.0	1. Removed old references to thick 270GB vs wap volumes 2. Corrected volume count calculation 3. Modified vfiler count to indicate up to 120 VMs per and the minimum count 4. Added Appendix B. with over-commitment and performance metric sections
1.1	Modified 3240AE expansion configs to show extra spare. Flash Cache uses default config. Dedicated filers.
1.2	Changes all occurrences of 3240 to 3240AE Updated table on page 4 to include ESX server average IOPS implied in other sections
1.3	Sect 8.1.2 VNX Design added
1.3.1	Added space/power sections for VNX Clarified dev-cloud deployments have storage over-commitment with VNX Re-worded NetApp vFiler guidelines Corrected hot spare count for VNX 5300 config 3
1.3.3	Various changes made based on RFC feedback Added power and space requirements for NetApp Added 2-hop maximum recommendation and associated FAQ entry Clarified that only VNX or NetApp should be provisioned to a ESX cluster Remove VADP from VNX section because it is not yet supported
1.3.4	Added general guidelines for DB VMs sect 8.3.1
1.3.5	Modified VNX configuration to use RAID1+0 instead of RAID-6 Added BTU ratings for all configurations Added large configuration comparison table
1.3.6	Clarified that VNX uses 2 storage pools Fixed network port count in VNX 5300 table. Clarified that VNX expansion should be equally spread across storage pools
1.3.7	Added Appendix C Retired Designs to include VNX RAID-6 design that was deployed on a few arrays
1.4.0	Changed 'Standard' class name to 'General Purpose' to match CTI terminology. Change 'Heavy' class name to 'LVMC' to match CTI terminology. Moved VNX (1 st Gen) to retired design section. Added VNX2 File design. Added dedicated XIV design for LVMC classes. Integrated LVMC and DB docs.
1.4.2	Added SDDC Pilot Design Added advanced setting for SDRS with vCenter 5.5 Clarified Dev Cloud can use SDRS and over-commit since IBM TSAM is retired. Clarified VNX volumes have ~1.6% capacity overhead
1.4.3	Added T++ design. Marked NetApp NAS as retired.

10 Appendix A. Procedure to Increase vSphere Parallel SAN I/O

These two settings increase the maximum number of parallel I/Os that will be dispatched to any single SAN LUN. A reboot of the ESX host is required after LUN Queue Depth is changed.

Citi Cluster Type	Max Hosts/ Cluster	lun_queue_depth	Disk.SchedNumReqOutstanding
SQL Server Consolidation with Shared SAN Arrays	10	64	64
LVMC with Shared SAN Arrays	16	96	96
LVMC with Dedicated XIV Arrays	16	Default - 30	Default - 30

1.1 LUN Queue Depth

Follow the procedure in the applicable ESX SAN Stack Solution Guide to change the `lpfc_lun_queue_depth` parameter.

<https://catecollaboration.citigroup.net/domains/platstor/gdse/products/techstdshmpgs/SANStacks.aspx>

1.2 Disk.SchedNumReqOutstanding

Set the `Disk.SchedNumReqOutstanding` value per host to match the `lun_queue_depth` from the table above. This ensures fair access to the LUNs from multiple ESX hosts.

VirtualCenter -> Configuration Tab -> Advanced Settings -> Disk -> `Disk.SchedNumReqOutstanding`

11 Appendix B. Standard VSI NetApp Over-Commitment and Performance Metrics Policy

11.1 Over-Commitment Policy

11.1.1 General Guidelines

- Standard 1.5TB volumes will be provisioned for VMware
- VM team will allocate at most 1.5TB per volume with no over-commitment
- VM team provides allocation report per volume **monthly**
- NAS team reviews capacity utilization quarterly and is alerted anytime it exceeds 50%
- **Aggregates used primarily for COB might require special handling and coordination with the VM team due the potential for rapid growth**

Over-commitment = Aggr_Committed%

Capacity utilization = Aggr Used%

Growth = Increase in Aggr Used% over time period

11.1.2 New Aggregates or Additional Disks

For new capacity, add volumes up to 200% over-commitment and wait 60 days before considering adding additional volumes.

For example, a new 29.92TB aggregate with 2 vSwap volumes would have **34** 1.5TB volumes provisioned
 $(29.92 * 1024 * 0.9 - 540 - 540) * 2 / (1.5 * 1024) = 34.49$

The addition of a new 17 disk (15+2) Raid Group to an existing aggregate would provide 7,844GB of additional capacity and assuming no vSwap volumes would have **9** 1.5TB volumes provisioned
 $(7844 * 0.9) * 2 / (1536) = 9.19$

11.1.3 Additional Volume Requests

If over commitment is < 300% and:

- 1) **Initial volumes were created > 60 days ago and it has been confirmed with VM team that existing volumes are fully allocated and,**
 - 2) Performance Metrics indicate additional volumes can be supported
 - 3) If capacity utilization is < 50%
- Then add volumes up to 300% over-commit (1.5TB minimum)

If over commitment is < 400% and:

- 1) If capacity utilization is < 50% and
 - 2) Performance Metrics indicate increased workload can be supported
 - 3) growth has been static for 90 days (< 2% growth)
- Then add volumes up to 400% over-commit and no more than 25% of current aggr capacity (1.5TB minimum)

For all other cases, see Aggregate Assessment and Expansion

11.1.4 BAU Monitoring

Capacity utilization exceeds 50%

- Notify VM team
- No additional volumes will be provisioned from the aggregate

Capacity utilization exceeds 60%

- Notify VM team
- Begin monthly monitoring of aggregate utilization

Capacity utilization exceeds 70%

- Notify VM team
- If capacity grew from 60% to 70% in less than 60 days, begin Aggregate Assessment and Expansion

Capacity utilization exceeds 80%

- Notify VM team
- Escalate Aggregate Assessment and Expansion

Capacity utilization exceeds 90%

- Notify VM team
- Request VM team assistance to reduce capacity below 90%
- Escalate Aggregate Assessment and Expansion

11.1.5 Aggregate Assessment and Expansion

For new requests if:

1. disk shelves are being added, the aggregate size will not increase beyond 70% of the maximum (3270AE=70TB, 3240AE/3210A=50TB) and
 2. the filer is < 2 years old and
 3. disks of the same size and speed are available from the vendor and
 4. Performance Metrics indicate increased workload can be supported
- Then additional disk shelves can be added to the filer and aggregate for new volumes

For BAU growth plan to reduce capacity utilization below 70%:

If the filer is > 2 years old **or aggregate capacity cannot be increased (70TB/50TB limits)**, request VM team assistance with migration to lower utilization

If capacity utilization > 90%, escalate acquisition of additional disk shelves as severity 1 issue for resolution ASAP

If capacity utilization > 80%, escalate acquisition of additional disk shelves as severity 2 issue for resolution in less than 15 days

If capacity utilization > 70%, begin acquisition of additional disk shelves for resolution within 30-60 days

11.1.6 Performance Metrics

These 3 performance criteria should be validated before adding additional volumes to an existing filer. If any of the criteria have been exceeded then the filer's disks should be considered performance saturated. If criteria #3 has not been exceeded then adding additional disks can be considered.

Only criteria #3 should be validated before adding additional disks to an existing filer. If criteria #3 has been exceeded then the filer should be considered performance saturated.

1. "1Week WAFL 4K blocks/sec" should be less than the max shown in the table to be eligible for adding the indicate number of additional 1.5TB volumes

Rational:

- a) Assumes that WAFL 4k blocks/sec is equivalent to IOPS going to disks
- b) Expected IOPS shouldn't exceed 75% of maximum, to maintain low response time
- c) Expect organic growth to increase by 25%
- d) Maximum of 200 IOPS per disk
- e) Assumes each additional volume may add 500 4K blocks/sec

	Aggr Data Disks	Max. "1W WAFL 4K blocks/sec" to add 1 volume	Max. "1W WAFL 4K blocks/sec" to add 2 volumes
<u>3270AE Config</u>			
1	16	1420	920
2	32	3340	2840
3	48	5260	4760
4	64	7180	6680
5	79	8980	8480
<u>3240AE Config</u>			
1	16	1420	920
2	32	3340	2840
3	48	5260	4760
<u>3210A Config</u>			
1	12	940	440
2	24	2380	1880
3	36	3820	3320

2. 1 Day and 1 Week: "NFSv3 Read Latency" should be < 30ms and "NFSv3 Write Latency" < 10ms

Rational:

- a) Matches threshold in NetApp Solution Guide

3. 1 Day and 1 Week: "CPU Busy %" sum of both clustered filer heads should be < 110

Rational:

- a) Assumes busiest CPU is Kahuna domain and limits at 55% to avoid impact to latency during failover event

12 Appendix C. Retired Designs

12.1 Retired EMC VNX (1st Gen) File NAS Solution

This section describes the standard storage designs that should be followed when deploying EMC VNX File NAS storage for VSI. If any deviation is necessary please contact CATE Storage Engineering to discuss the exception.

Three different configuration sizes are available depending on the expected scale of deployment.

Size	Maximum VM count	VNX Model	VNX Data Movers	Virtual Data Movers
Small NAS	160	5300	2 (1+1)	1
Medium NAS	400	5500	2 (1+1)	2
Large NAS	960	5700	3 (2+1)	6

12.1.1.1 Common VNX File Design Guidelines

- Follow existing Citi VNX File standards unless otherwise indicated in this guide
- Each EMC VNX File array should be dedicated for VMware
- Filers should be placed close to the VMware clusters being serviced and not more than 2 network hops away²³
- Storage over-commitment will be performed at the VMware ESX layer (except for Dev Cloud deployments) There will be no over-commitment on the storage array.
- All filesystems (datastores) will be thickly provisioned on VNX File array.
- **For non-dev cloud deployments, assume 200% capacity over-commitment on VMware, provision only half of the total capacity requested.**(i.e. if 48TB is requested, provision only 24TB or 7 volumes 3,500GB each)
 - **For Dev Cloud deployments storage over-commitment will not be possible on ESX or the array so the full requested capacity should be procured.**
- No compression or de-duplication will be used on the VNX File array (EMC recommended that compression not be used with active VMs due to the potential performance impact)
- Snapshots are not created for any of the NFS volumes used with VMware
- Backups of VSI VM systems are performed within the VM itself using standard tools.
- The EMC VNX File storage designs have been defined to use modular expansion. The initial block supports a certain capacity and estimated VM count. Additional storage blocks can be added to increase capacity and/or VM count.
- Two VNX block storage pool should be used with balanced SP ownership of LUNs

²³ See [FAQ section](#) for the applicable definition of 2 network hops

- Virtual Data Movers (VDM) should be implemented to help challenges surrounding large scale migration of guests when a storage refresh is due.
 - 5 file systems per VDM
 - A maximum of 1 vSwap volume per ESX cluster per VDM to limit risk²⁴
- Set the # of NFS threads to 1024 on each Data Mover. The nthreads parameter can be set by using a command like this for each DM:
 - `Server_param server_2 -facility nfs -modify nthreads -value 1024`
 - Reboot the Data Mover for the change to take effect
- Use the VNX “uncached” mount option as it improves overall VM performance. The “uncached” option can be set using a command like this for every file system on each DM:
 - `server_mount server_2 -option uncached fs1 /fs1`
- To capture backend performance statistics are collected, ensure that Unisphere Analyzer is licensed and enable “Periodic Archiving” without the “stop automatically” parameter checked (Statistics for Block -> Performance Data Logging)

12.1.1.2 Guidelines for Virtualization Teams

- Datstores will be 3,500GB in size
- All VMDKs should be thinly provisioned on ESX
- For non-Dev cloud deployments over-commitment of storage capacity should occur at the ESX layer
- To simplify management of storage over-commitment VMware Storage DRS (SDRS) should be used for initial placement and auto-balancing based on capacity
 - With vCenter 5.5, for each Datastore Cluster, set the advanced option
SetSpaceLoadToDatastoreUsedMB=0
- Dev Cloud will not be able to use SDRS auto-balancing due to TSAM restrictions and storage over-commitment will not be possible due to the operational overhead of manual rebalancing.
- On average each file system should have a maximum of 40 VMs provisioned to align with the performance design and the expected 200% capacity over-commitment
- All VMDK volumes need to be properly block aligned.
- Set the disk timeout value to at least 60 seconds²⁵ in the guest OS

12.1.1.3 General Purpose VSI VNX 5300 Design

The VNX 5300 is a small entry level configuration with limited connectivity and capacity. Its scales up to support 160 VMs. Existing VNX File standards should be followed unless otherwise noted.

- Network connections for NFS can be 1GbE or preferably 10GbE, both with active/passive failover (aka “fail-safe networking” FSN):
 - 1GbE networks will require 8 connections
 - 4 per DM configured as 2 LACP pairs: FSN (LACP (1,2) + LACP (1,2))

Network Requirements for VNX 5300 w/ 1GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	1GbE Connections	NFS IP Qty	

²⁴ The small 5300 configurations will place 2 vSwap volumes in a single VDM

²⁵ This is set by default when VMware tools are installed

Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	8	1	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	15	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	8	16	

- 10GbE networks will require 4 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements for VNX 5300 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	4	1	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	15	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	4	16	

12.1.1.3.1 General Purpose VSI VNX 5300 Storage Blocks

The VNX can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5300 with 2(1+1) DMs, 600GB 10K 2.5" disks, and Fast Cache							
Config	Max. ²⁶ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+0 (4+4) Groups
				Vault Disks	Spare Disks	Pool Disks	
1	80	3	38	4	2	32	4
2	160	3	71	4	3	64	8

Storage Pools and Filesystems							
Config	ESX Clusters Supported	VNX Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization

²⁶ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

1	1	#1	2	4,105	1	1	98%
		#2	2	4,105	1	1	98%
2	1	#1	4	8,210	2	1	92%
		#2	4	8,210	2	1	92%

Virtual Data Mover Assignments					
Config	DM	VNX Block Storage Pool	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	#1	2	2
2	server_2	#1	#1	4	2

12.1.1.3.2 General Purpose VSI VNX 5300 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures. The capacity expansion should be evenly split across storage pools.

Config	Additional Spare Disks	Additional 600GB disks	Additional RAID1+0 (4+4) Groups
2a	0	16	2

12.1.1.3.3 General Purpose VSI VNX 5300 Space and Power

5300 with 2(1+1) DMs			
Config	Power kVA	BTU	RU
1	1.36	3,900	12
2	1.67	4,900	14

12.1.1.4 General Purpose VSI VNX 5500 Design

The VNX 5500 array scales up to support 400 VMs.

- Network connections for NFS can be 1GbE or preferably 10GbE, both with active/passive failover (aka “fail-safe networking” FSN):
 - 1GbE networks will require 16 connections
 - 8 per DM configured as 2 LACP pairs: FSN (LACP (1,2,3,4) + LACP (1,2,3,4))

Network Requirements for VNX 5500 w/ 1GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	1GbE Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	16	2	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	14	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	16	16	

- 10GbE networks will require 4 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements for VNX 5500 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
2 Datamovers (DM)	0	0	4	2	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	14	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	4	16	

12.1.1.4.1 General Purpose VSI VNX 5500 Storage Blocks

The VNX can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5500 with 2(1+1) DMs, 600GB 10K 2.5" disks, and Fast Cache							
Config	Max. ²⁷ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+0 (4+4) Groups
				Vault Disks	Spare Disks	Pool Disks	
1	240	5	103	4	3	96	12
2	320	5	136	4	4	128	16
3	400	5	168	4	4	160	20

Storage Pools and Filesystems							
Config	ESX Clusters Supported	VNX Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	1	#1	6	12,315	3	1	90%
		#2	6	12,315	3	1	90%
2	2	#1	8	16,420	4	2	92%
		#2	8	16,420	4	2	92%
3	2	#1	10	20,525	5	2	91%
		#2	10	20,525	5	2	91%

Virtual Data Mover Assignments					
Config	DM	VNX Block Storage Pool	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	#1	3	1
		#2	#2	3	1
2	server_2	#1	#1	4	1-2
		#2	#2	4	1-2
3	server_2	#1	#1	5	1-2
		#2	#2	5	1-2

²⁷ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

12.1.1.4.2 General Purpose VSI VNX 5500 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures. The capacity expansion should be evenly split across storage pools.

Config	Additional Spare Disks	Additional 600GB disks	Additional RAID1+0 (4+4) Groups
3a	0	16	2

12.1.1.4.3 General Purpose VSI VNX 5500 Space and Power

5500 with 2(1+1) DMs			
Config	Power kVA	BTU	RU
1	1.97	5,900	18
2	2.36	7,000	22
3	2.65	7,900	24

12.1.1.5 General Purpose VSI VNX 5700 Design

The VNX 5700 array scales up to support 960 VMs.

- Network connections for NFS must be 10GbE with active/passive failover (aka “fail-safe networking” FSN):
 - 10GbE networks will require 6 connections
 - 2 per DM: FSN (1 + 1)

Network Requirements for VNX 5700 w/ 10GbE	Mgmt Interfaces		NFS Interfaces		Notes
	100Mb RJ45 Connections	Mgmt IP Qty	10GbE optical SFP+ Connections	NFS IP Qty	
Dual Control Stations(CS) and Service Processors(SP)	2	4	0	0	Mgmt interfaces can be on different subnet from Datamover NFS interfaces
3 Datamovers (DM)	0	0	6	6	1 IP per vDM , Used for NFS data access
Tech Refresh Reserve	-	-	-	10	Sized so that NFS IP total is consistent with the NetApp config
Total	2	4	6	16	

12.1.1.5.1 General Purpose VSI VNX 5700 Storage Blocks

The VNX can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5700 with 3(2+1) DMs, 600GB 10K 2.5” disks, and Fast Cache							
Config	Max. ²⁸ VMs Supported	100GB Fast Cache SSDs	600GB Disks	600GB Disk Assignment			Qty of RAID1+5 (4+4) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
1	640	7	249	4	5	240	30
2	720	7	282	4	6	272	34
3	960	9	363	4	7	352	44

Storage Pools and Filesystems							
Config	ESX Clusters Supported	VNX Block Storage Pool	Qty of RAID1+0 (4+4) Groups in Pool	Usable Capacity (GB)	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	3	#1	15	30,788	8	3	96%
		#2	15	30,788	8	3	96%
2	3	#1	17	34,893	9	3	95%
		#2	17	34,893	9	3	95%
3	3	#1	22	45,155	12	3	97%
		#2	22	45,155	12	3	97%

²⁸ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

Virtual Data Mover Assignments					
Config	DM	VNX Block Storage Pool	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	#1	4	1-2
		#1	#2	4	1
	server_3	#2	#3	4	1-2
		#2	#4	4	1
2	server_2	#1	#1	5	1-2
		#1	#2	4	1
	server_3	#2	#3	5	1-2
		#2	#4	4	1
3	server_2	#1	#1	4	1-2
		#1	#2	4	1
		#1	#3	4	0
	server_3	#2	#4	4	1-2
		#2	#5	4	1
		#2	#6	4	0

12.1.1.5.2 General Purpose VSI VNX 5700 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures. The capacity expansion should be evenly split across storage pools.

Config	Additional Spare Disks	Additional 600GB disks	Additional RAID1+0 (4+4) Groups
3a	0	16	2

12.1.1.5.3 General Purpose VSI VNX 5700 Space and Power

5700 with 3(2+1) DMs			
Config	Power kVA	BTU	Used RU
1	3.96	11,700	31
2	4.35	12,800	35
3	5.14	15,300	41

12.2 Retired EMC VNX File NAS Solution using RAID-6

This section describes the standard storage designs that should be followed when deploying EMC VNX File NAS storage for VSI. If any deviation is necessary please contact CATE Storage Engineering to discuss the exception.

Three different configuration sizes are available depending on the expected scale of deployment.

Size	Maximum VM count	VNX Model	VNX Data Movers	Virtual Data Movers
Small NAS	160	5300	2 (1+1)	1
Medium NAS	440	5500	2 (1+1)	3
Large NAS	960	5700	3 (2+1)	6

12.2.1.1 Common VNX File Design Guidelines

- Follow existing Citi VNX File standards unless otherwise indicated in this guide
- Each EMC VNX File array should be dedicated for VMware
- Filers should be placed close to the VMware clusters being serviced and not more than 2 network hops away²⁹
- 300GB 15K 3.5" disks are not certified for general use with VNX File but are included in this design because they are required to meet the performance requirements
- Storage over-commitment will be performed at the VMware ESX layer (except for Dev Cloud deployments) There will be no over-commitment on the storage array.
- All filesystems (datastores) will be thickly provisioned on VNX File array.
- **For non-dev cloud deployments, assume 200% capacity over-commitment on VMware, provision only half of the total capacity requested.**(i.e. if 48TB is requested, provision only 24TB or 7 volumes 3,500GB each)
 - **For Dev Cloud deployments storage over-commitment will not be possible on ESX or the array so the full requested capacity should be procured.**
- No compression or de-duplication will be used on the VNX File array (EMC recommended that compression not be used with active VMs due to the potential performance impact)
- Snapshots are not created for any of the NFS volumes used with VMware
- Backups of VSI VM systems are performed within the VM itself using standard tools.
- The EMC VNX File storage designs have been defined to use modular expansion. The initial block supports a certain capacity and estimated VM count. Additional storage blocks can be added to increase capacity and/or VM count.

²⁹ See [FAQ section](#) for the applicable definition of 2 network hops

- A single VNX block storage pool should be used with balanced SP ownership of LUNs
- Virtual Data Movers (VDM) should be implemented to help challenges surrounding large scale migration of guests when a storage refresh is due.
 - 5 file systems per VDM
 - A maximum of 1 vSwap volume per ESX cluster per VDM to limit risk³⁰
- Set the # of NFS threads to 1024 on each Data Mover. The nthreads parameter can be set by using a command like this for each DM:
 - `Server_param server_2 -facility nfs -modify nthreads -value 1024`
 - Reboot the Data Mover for the change to take effect
- Use the VNX “uncached” mount option as it improves overall VM performance. The “uncached” option can be set using a command like this for every file system on each DM:
 - `server_mount server_2 -option uncached fs1 /fs1`
- To capture backend performance statistics are collected, ensure that Unisphere Analyzer is licensed and enable “Periodic Archiving” without the “stop automatically” parameter checked (Statistics for Block -> Performance Data Logging)

12.2.1.2 Guidelines for Virtualization Teams

- Datatstores will be 3,500GB in size
- All VMDKs should be thinly provisioned on ESX
- For non-Dev cloud deployments over-commitment of storage capacity should occur at the ESX layer
- To simplify management of storage over-commitment VMware Storage DRS (SDRS) should be used for initial placement and auto-balancing based on capacity
- Dev Cloud will not be able to use SDRS auto-balancing due to TSAM restrictions and storage over-commitment will not be possible due to the operational overhead of manual rebalancing.
- On average each file system should have a maximum of 40 VMs provisioned to align with the performance design and the expected 200% capacity over-commitment
- All VMDK volumes need to be properly block aligned.
- Set the disk timeout value to at least 60 seconds³¹ in the guest OS

12.2.1.3 Standard VSI VNX 5300 Design

The VNX 5300 is a small entry level configuration with limited connectivity and capacity. Its scales up to support 160 VMs. Existing VNX File standards should be followed unless otherwise noted.

- Network connections for NFS can be 1GbE or preferably 10GbE, both with active/passive failover (aka “fail-safe networking” FSN):
 - 1GbE networks will require 8 connections
 - 4 per DM configured as 2 LACP pairs: FSN (LACP (1,2) + LACP (1,2))
 - 10GbE networks will require 4 connections
 - 2 per DM: FSN (1 + 1)

12.2.1.3.1 Standard VSI VNX 5300 Storage Blocks

The VNX can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

³⁰ The small 5300 configurations will place 2 vSwap volumes in a single VDM

³¹ This is set by default when VMware tools are installed

5300 with 2(1+1) DMs, 300GB 15K 3.5" disks, and Fast Cache							
Config	Max. ³² VMs Supported	100GB Fast Cache SSDs	300GB Disks	300GB Disk Assignment			Qty of RAID-6 (14+2) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
1	80	5	54	4	2	48	3
2	120	5	71	4	3	64	4
3	160	5	87	4	3	80	5

File systems					
Config	Usable Capacity ³³ (GB)	ESX Clusters support	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	7,000	1	2	2	77%
2	10,500	1	3	2	83%
3	14,000	1	4	2	86%

Virtual Data Mover Assignments				
Config	DM	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	2	2
2	server_2	#1	3	2
3	server_2	#1	4	2

12.2.1.3.2 Standard VSI VNX 5300 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures.

Config	Additional Spare Disks	Additional 300GB disks	Additional RAID-6 (14+2) Groups
3a	0	16	1

12.2.1.3.3 Standard VSI VNX 5300 Space and Power

5300 with 2(1+1) DMs		
Config	Power kVA	RU
1	1.77	17
2	2.12	23
3	2.18	26

12.2.1.4 Standard VSI VNX 5500 Design

³² Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

³³ Excluding vSwap capacity

The VNX 5500 array scales up to support 440 VMs.

- Network connections for NFS can be 1GbE or preferably 10GbE, both with active/passive failover (aka "fail-safe networking" FSN):
 - 1GbE networks will require 16 connections
 - 8 per DM configured as 2 LACP pairs: FSN (LACP (1,2,3,4) + LACP (1,2,3,4))
 - 10GbE networks will require 4 connections
 - 2 per DM: FSN (1 + 1)

12.2.1.4.1 Standard VSI VNX 5500 Storage Blocks

The VNX can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5500 with 2(1+1) DMs, 300GB 15K 3.5" disks, and Fast Cache							
Config	Max. ³⁴ VMs Supported	100GB Fast Cache SSDs	300GB Disks	300GB Disk Assignment			Qty of RAID-6 (14+2) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
1	240	9	119	4	3	112	7
2	320	9	152	4	4	144	9
3	440	9	217	4	5	208	13

Filesystems					
Config	Usable Capacity ³⁵ (GB)	ESX Clusters support	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	21,000	1	6	2	90%
2	28,000	1-2	8	2-4	83%-86%
3	38,500	1-2	11	2-4	87%-89%

Virtual Data Mover Assignments				
Config	DM	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	3	1
		#2	3	1
2	server_2	#1	4	1-2
		#2	4	1-2
3	server_2	#1	4	1-2
		#2	4	1-2
		#3	3	0

³⁴ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

³⁵ Excluding vSwap capacity

12.2.1.4.2 Standard VSI VNX 5500 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures.

Config	Additional Spare Disks	Additional 300GB disks	Additional RAID-6 (14+2) Groups
3a	0	16	1

12.2.1.4.3 Standard VSI VNX 5500 Space and Power

5500 with 2(1+1) DMs		
Config	Power kVA	RU
1	2.92	32
2	3.44	38
3	4.60	55

12.2.1.5 Standard VSI VNX 5700 Design

The VNX 5700 array scales up to support 960 VMs.

- Network connections for NFS must be 10GbE with active/passive failover (aka “fail-safe networking” FSN):
 - 10GbE networks will require 6 connections
 - 2 per DM: FSN (1 + 1)

12.2.1.5.1 Standard VSI VNX 5700 Storage Blocks

The VNX can have the following configurations to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

5700 with 3(2+1) DMs, 300GB 15K 3.5” disks, and Fast Cache							
Config	Max. ³⁶ VMs Supported	100GB Fast Cache SSDs	300GB Disks	300GB Disk Assignment			Qty of RAID-6 (14+2) Groups in Storage Pool
				Vault Disks	Spare Disks	Pool Disks	
1	680	13	315	4	7	304	19
2	760	13	347	4	7	336	21
3	960	13	445	4	9	432	27

Filesystems					
Config	Usable Capacity ³⁷ (GB)	ESX Clusters support	3,500GB file systems	540GB vSwap file systems	Storage Pool Capacity Utilization
1	59,500	2-3	17	4-6	93%-94%
2	66,500	2-3	19	4-6	89%-91%
3	84,000	2-3	24	4-6	91%-92%

Virtual Data Mover Assignments				
Config	DM	VDM	3,500GB file systems	540GB vSwap file systems
1	server_2	#1	4	1-2
		#2	4	1
		#3	1	0
	server_3	#4	4	1-2
		#5	4	1
2	server_2	#1	4	1-2
		#2	4	1
		#3	3	0
	server_3	#4	4	1-2
		#5	4	1
3	server_2	#1	4	1-2
		#2	4	1
		#3	4	0
	server_3	#4	4	1-2
		#5	4	1
		#6	4	0

³⁶ Maximum VM count based on performance design and expected 200% capacity over-commitment on VMware

³⁷ Excluding vSwap capacity

12.2.1.5.2 Standard VSI VNX 5700 Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures.

Config	Additional Spare Disks	Additional 300GB disks	Additional RAID-6 (14+2) Groups
3a	0	16	1

12.2.1.5.3 Standard VSI VNX 5700 Space and Power

5700 with 3(2+1) DMs		
Config	Power kVA	Used RU
1	6.66	75
2	7.18	81
3	8.86	92

12.3 Retired NetApp NAS Solutions

This table highlights solution differences for the large configuration:

Solution	VSI VMs Supported	Storage Over-Commitment	Raw HDD Capacity	Disk Qty	Rack RU	Power (kVA)	Cost
NetApp 3270AE NAS	960	On NAS Filer	112.5 TB	192	38	5.44	Consult with Procurement
EMC VNX2 File 5800 NAS	960	On ESX	177.1 TB	372	39	4.99	Consult with Procurement

This section describes the standard storage designs that should be followed when deploying NetApp NAS storage for VSI. If any deviation is necessary please contact CATE Storage Engineering to discuss the exception.

NOTE: These configurations are based on the agreement with NetApp to provide additional discounts on 512GB FlashCache adaptors with the Citi VMware VPA for 3240AE and 3270AE filers (35% and 60% respectively).

Three different configuration sizes are available depending on the expected scale of deployment. The planned VM count should be used to select the size for initial deployment. Performance analysis of the initial deployment may or may not permit additional VM provisioning up to the maximum VM count and it should not be used for planning purposes.

Size	Maximum VM count	NetApp Model	vFilers (up to 120 VM each)
Small NAS	360	3210A	3 (min)
Medium NAS	660	3240AE	4 (min)
Large NAS	960	3270AE	6 (min)

12.3.1.1 Common NetApp Design Guidelines

- NetApp filer clusters are dedicated for VMware
- Filers should be placed close to the VMware clusters being serviced and not more than 2 network hops away³⁸
- Storage over-commitment will be conducted at the NetApp NAS layer. VMware teams should be directed not to over-commit the NFS volumes they are provisioned.
- All volumes (data-stores) will be thinly provisioned on NetApp NAS
- Deduplication should be run on all NFS volumes (data-stores) with the exception of volumes used for VMware vSwap. These de-dup savings estimates were used in the design:
 - OS partition: 50%
 - Data partition: 10%
 - De-dup Buffer per aggregate: 10%
- Snapshots are not created for any of the NFS volumes used with VMware
- Backups of VSI VM systems are performed within the VM itself using standard tool tools or with Netbackup VADP.
- The NetApp storage designs have been defined to use modular expansion. The initial block supports a certain capacity and estimated VM count. Additional storage blocks can be added to increase capacity and/or VM count.

³⁸ See [FAQ section](#) for the applicable definition of 2 network hops

- vFilers should be implemented to help challenges surrounding large scale migration of guests when a storage refresh is due.
 - To include roughly 120 VMs, every 5 volumes (1.5TB) should be grouped into a vFiler. If a small grouping is preferred they can be split into multiple vFilers but the quantity of vFilers per physical filer should not exceed 20.
 - A maximum of 1 vSwap volume should be included per vFiler to limit the risk
- Capacity Over-Commitment Contingency - The configuration supports the best practice of not over-committing capacity beyond what can physically be installed on a storage array should either Thin Provisioning or De-Duplication fail to optimize. It does not cover failure of both. In the future if capacity thresholds indicate unexpected growth of used space or IOPS, the aggregate can be expanded with additional disk shelves to meet the new requirement and remain below the current ONTAP 8 64-bit aggregate maximum size limits of 70TB and 50TB for 3270AE and 3240AE/3210A respectively.
- New capacity will be provisioned into 1.5TB volumes over-committing available storage by 200% after capacity for 10% dedup buffer and vSwap volumes are accounted for. For example, a 29.92TB aggregate with 2 vSwap volumes would have **34** 1.5TB volumes provisioned

$$(29.92 * 1024 * .9 - 540 - 540) * 2 / (1.5 * 1024) = 34.49$$
- The NetApp Capacity Planning procedures should be followed to determine when additional volumes can be added to increase over-commitment.

12.3.1.2 Guidelines for Virtualization Teams

- Volumes will be 1.5TB in size
- Over-commitment of storage capacity should not occur at the ESX layer. Only 1.5TB should be allocated to each volume.
- On average each volume should have a maximum of 12 VMs provisioned to align with the performance design. A performance analysis of the filer should be conducted before exceeding this guideline.
- All VMDK volumes need to be properly block aligned.

12.3.1.3 General Purpose VSI NetApp 3210A Design

The 3210A is a small entry level configuration with limited frontend connectivity. The 3210A filer cluster scales up to 360 VMs with 2.5 DS4243 disk shelves per filer head. The DS4243 disk shelves should be configured as specified in the [FAS3200 Series Release Notes](#) with additional disk adaptors as necessary for a HA config. Other design specifications:

- One quad-port GigE adapters will supplement built-in connections to provide a 2+2 configuration for frontend connectivity per head as shown in the FAS3200 Release Notes.
- FlashCache cannot be used because of the limited adaptor slots in the 3210A.
- Disks will be **450GB** (SAS-3.5-15K-450) to minimize quantity and cost (If an initial usable capacity is required that is 50% greater than in the tables below then 600GB disks could be used instead)
- A single 64-bit aggregate (aggr1) is used for all VMware volumes to balance load and maximize scalability

12.3.1.3.1 General Purpose VSI 3210A Storage Blocks

Each filer head can have the following configuration of 450GB disk shelves to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The **green** highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

Config	VMs @ 200% OC ³⁹	Max. ⁴⁰ VMs Supported	Shelves of Disks	Aggr0 Disks	Spare Disks	Unused Disks	Aggr1 Disks (all RAID-DP)
1	36	60	1.0	3 (1+2)	2	5	14 (rg0=14)
2	96	120	1.5	3 (1+2)	2	3	28 (rg0=14/rg1=14)
3	156	180	2.0	3 (1+2)	2	1	42 (rg0=14/rg1=14/rg2=14)

Note: Per best practice, 4243 disk shelves cannot be shared between filer heads so the configuration with 1.5 shelves per head would require a total of 4 shelves total per cluster.

Config	Usable Cap. ⁴¹	1.5TB Volumes @ 200% OC	Est. Aggr1 Utilization ⁴²
1	3.7 TB	3	48%
2	7.5 TB	8	65%
3	11.3 TB	13	70%

12.3.1.3.2 General Purpose VSI 3210A Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures.

Config	Shelves of Disks	Additional Spare Disks	Unused Disks	Additional Aggr1 Disks (all RAID-DP)
3a	2.5	0	1	12 (rg3=12)
3b	3.0	0	1	24 (rg3=12/rg4=12)
3c	3.5	0	1	36 (rg3=12/rg4=12/rg5=12)
3d	4.0	1	0	48 (rg3=12/rg4=12/rg5=12/rg6=12)

12.3.1.3.3 General Purpose VSI 3210A Space and Power

Config	Power kVA	BTU	RU
1	1.82	6,183	11
2	2.49	8,088	19
3	3.17	10,272	19

³⁹ Estimated VM count when accounting for 2 x 540Gb vSwap volumes and storage is overcommitted by 200%

⁴⁰ Absolute maximum VM count based on 3210A scaling limitations

⁴¹ Approximate usable storage capacity after 10% WAFL formatting overhead and 10% free space buffer but before vSwap volumes and over-commitment

⁴² Aggr utilization increases to between 70%-80% if vSwap volumes are fully utilized.

12.3.1.4 General Purpose VSI NetApp 3240AE Design

The 3240AE filer cluster scales up to 660 VMs with 2.5 DS4243 disk shelves per filer head. The DS4243 disk shelves should be configured as specified in the [FAS3200 Series Release Notes](#) with additional disk adaptors as necessary for a HA config. Other design specifications:

- Two quad-port GigE adapters (4+4) or two 10GbE (1+1) can be used for frontend connectivity per head as shown in the FAS3200 Release Notes
- Each filer head will use a 512GB Flash Cache adapter to reduce the backend load and permit the use of fewer denser disks. Default Flash Cache settings should be used for the “normal” mode of operation.
- Disks will be 600GB (SAS-3.5-15K-600) to minimize quantity and cost
- A single 64-bit aggregate (aggr1) is used for all VMware volumes to balance load and maximize scalability

12.3.1.4.1 General Purpose VSI 3240AE Storage Blocks

Each filer head can have the following configuration of 600GB disk shelves to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

Config	VMs @ 200% OC ⁴³	Max. ⁴⁴ VMs Supported	Shelves of Disks	Aggr0 Disks	Spare Disks	Unused Disks	Aggr1 Disks (all RAID-DP)
1	84	108	1.0	3 (1+2)	3	0	18 (rg0=18)
2	192	217	2.0	3 (1+2)	3	6	36 (rg0=18/rg1=18)
3	300	330	2.5	3 (1+2)	3	0	54 (rg0=18/rg1=18/rg2=18)

Note: Per best practice, 4243 disk shelves cannot be shared between filer heads so the configuration with 2.5 shelves per head would require a total of 6 shelves total per cluster.

Config	Usable Cap. ⁴⁵	1.5TB Volumes @ 200% OC	Est. Aggr1 Utilization ⁴⁶
1	6.5 TB	7	64%
2	13.5 TB	16	73%
3	20 TB	25	76%

⁴³ Estimated VM count when accounting for 2 x 540Gb vSwap volumes and storage is overcommitted by 200%

⁴⁴ Estimated maximum VM count based on post-FlashCache 22 IOPS per VM

⁴⁵ Approximate usable storage capacity after 10% WAFL formatting overhead and 10% free space buffer but before vSwap volumes and over-commitment

⁴⁶ Aggr utilization increases to ~80% if vSwap volumes are fully utilized.

12.3.1.4.2 General Purpose VSI 3240AE Capacity Expansion

If capacity expansion beyond Config 3 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures.

Config	Shelves of Disks	Additional Spare Disks	Unused Disks	Additional Aggr1 Disks (all RAID-DP)
3a	3.5	1	6	18 (rg3=18)
3b	4.0	0	0	36 (rg3=18/rg4=18)
3c	5.0	0	6	54 (rg3=18/rg4=18/rg5=18)
3d	5.5	0	0	72 (rg3=18/rg4=18/rg5=18/rg6=18)

12.3.1.4.3 General Purpose VSI 3240AE Space and Power

Config	Power kVA	BTU	RU
1	1.91	6,188	14
2	3.06	9,918	22
3	4.21	13,648	30

12.3.1.5 General Purpose VSI NetApp 3270AE Design

The 3270AE filer cluster scales up to 960 VMs with 4 DS4243 disk shelves per filer head. The DS4243 disk shelves should be configured as specified in the [FAS3200 Series Release Notes](#) with additional disk adaptors as necessary for a HA config. Other design specifications:

- Two 10GbE (1+1) can be used for frontend connectivity per head as shown in the FAS3200 Release Notes
- Each filer head will use a 512GB Flash Cache adapter to reduce the backend load and permit the use of fewer denser disks. Default Flash Cache settings should be used for the “normal” mode of operation.
- Disks will be 600GB (SAS-3.5-15K-600) to minimize quantity and cost
- A single 64-bit aggregate (aggr1) is used for all VMware volumes to balance load and maximize scalability

12.3.1.5.1 General Purpose VSI 3270AE Storage Blocks

Each filer head can have the following configuration of 600GB disk shelves to support the indicated number of VMs and can be incrementally expanded up to the maximum configuration shown. The green highlight indicates the recommended config to deploy if the VM requirements are unknown and may vary from the design assumptions:

Config	VMs @ 200% OC ⁴⁷	Max. ⁴⁸ VMs Supported	Shelves of Disks	Aggr0 Disks	Spare Disks	Unused Disks	Aggr1 Disks (all RAID-DP)
1	84	108	1.0	4 (2+2)	2	0	18 (rg0=18)
2	192	217	2.0	4 (2+2)	2	6	36 (rg0=18/rg1=18)
3	300	326	2.5	4 (2+2)	2	0	54 (rg0=18/rg1=18/rg2=18)
4	408	435	3.5	4 (2+2)	2	6	72 (rg0=18/rg1=18/rg2=18/rg3=18)
5	480	480	4.0	4 (2+2)	3	0	89 (rg0=18/rg1=18/rg2=18/rg3=18/rg4=17)

Note: Per best practice, 4243 disk shelves cannot be shared between filer heads so the configuration with 2.5 shelves per head would require a total of 6 shelves total per cluster and the 3.5 would require 8.

Config	Usable Cap. ⁴⁹	1.5TB Volumes @ 200% OC	Est. Aggr1 Utilization ⁵⁰
1	6.5 TB	7	64%
2	13.5 TB	16	73%
3	20 TB	25	76%
4	27 TB	34	77%
5	33 TB	42	77%

⁴⁷ Estimated VM count when accounting for 2 x 540Gb vSwap volumes and storage is overcommitted by 200%

⁴⁸ Estimated maximum VM count based on post-FlashCache 22 IOPS per VM

⁴⁹ Approximate usable storage capacity after 10% WAFL formatting overhead and 10% free space buffer but before vSwap volumes and over-commitment

⁵⁰ Aggr utilization increases to ~80% if vSwap volumes are fully utilized.

12.3.1.5.2 General Purpose VSI 3270AE Capacity Expansion

If capacity expansion beyond Config 5 is required the following increments should be used. This should not be used to increase the number VMs beyond the maximum supported limits above or that indicated by capacity planning procedures.

Config	Shelves of Disks	Additional Spare Disks	Unused Disks	Additional Aggr1 Disks (all RAID-DP)
5a	5.0	0	6	18 (rg5=18)
5b	5.5	0	0	36 (rg5=18/rg6=18)
5c	6.5	0	6	54 (rg5=18/rg6=18/rg7=18)
5d	7.0	0	0	72 (rg5=18/rg6=18/rg7=18/rg8=18)

12.3.1.5.3 Standard VSI 3270AE Space and Power

Config	Power kVA	BTU	RU
1	1.99	6,448	14
2	3.14	10,178	22
3	4.29	13,908	30
4	4.87	15,773	38
5	5.44	17,638	38

12.3.1.6 NetApp Design Specific FAQ

7. Why were 600GB disks used for the 3240AE and 3270AE instead of 450GB? Won't that lower utilization?

The use of 600GB disks provided the most cost efficient configuration given the workload requirements and increased capacity requirements per VM that are expected. Capacity requirement estimates provided by the global virtualization team increased to 128GB average per VM.

Using 450GB drives with 3210A configuration provided smaller building blocks for those smaller deployments.

8. Will utilization rates be higher with this configuration?

If the estimated average VM size is achieved then utilization rates will be higher at near 70%. The use of a single 64-bit aggregate per filer head provides a shared free space pool for all volumes and should help increase overall utilization.

9. Are additional SAS controllers required?

The release notes for the filer model should be consulted for appropriate adaptor requirements to provide a HA configuration.

10. What is the smallest configuration possible?

A 3210A cluster with 1.5 disk shelves in an active/passive configuration would support between 36-80 VMs.

11. Why weren't 256GB FlashCache adaptors used with the 3240AE?

The limited reduction in IOPS provided by the 256GB adaptor fails to eliminate any disk from the final configuration.

12. Why wasn't a different RAID group size used to reduce the number of unused disks?

The use of RAID groups sized at 16+2 with the 3270AE/3240AE and 14+2/10+2 with the 3210A were the most efficient choices that also maintained a reasonably large group size and adhered to NetApp best practices.

13. Why weren't the larger 62xx filer models used?

They were considered but would have increased the failure domain and had a higher cost per VM.