

# EMC VNX2 Unified Best Practices for Performance

Applied Best Practices Guide  
VNX OE for Block 05.33.008  
VNX OE for File 8.1.8

**EMC Core Technologies Division, VNX BU**

## **Abstract**

This applied best practices guide provides recommended best practices for installing and configuring VNX2™ unified storage systems for good performance.

October 2015

Copyright © 2015 EMC Corporation. All rights reserved. Published in the USA.

Published October, 2015

EMC believes the information in this publication is accurate of its publication date. The information is subject to change without notice.

The information in this publication is provided as is. EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose. Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC2, EMC, and the EMC logo are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners.

For the most up-to-date regulatory document for your product line, go to the technical documentation and advisories section on EMC Online Support.

## **EMC VNX2 Unified Best Practices for Performance**

### **Applied Best Practices Guide**

Part Number H10938.8

<b>Chapter 1 System Configuration .....</b>	<b>7</b>
Essential guidelines .....	8
Storage Processor cache .....	8
Physical placement of drives.....	8
Hot sparing .....	8
Usage of flash drives in hybrid flash arrays .....	9
Availability and connectivity .....	9
Fibre Channel and iSCSI connectivity .....	10
NAS connectivity .....	11
<b>Chapter 2 Storage Configuration .....</b>	<b>12</b>
General considerations .....	13
Drive type.....	13
Rules of thumb.....	13
RAID level.....	14
Calculating disk IOPS by RAID type.....	14
Determining which LUN type to configure.....	15
Creating LUNs for Block access .....	16
Creating LUNs for File access.....	16
Storage pool considerations .....	17
Storage pool creation and expansion .....	17
Pool capacity considerations .....	18
Storage tiers .....	18
Tiering policies.....	19
Storage pool object considerations .....	19
Storage pool LUNs for Block access .....	19
Storage Pool LUNs for File access.....	20
Classic RAID Group LUN considerations .....	22
Drive location selection.....	22
Classic RAID Group LUNs for Block access.....	23

Classic RAID Groups for File access .....	23
<b>Chapter 3 Data Services .....</b>	<b>25</b>
FAST VP .....	26
General .....	26
Data relocation .....	26
Pool capacity utilization .....	26
Considerations for VNX OE for File .....	27
Multicore FAST Cache .....	27
General considerations .....	27
Creating FAST Cache .....	27
Enabling Multicore FAST Cache on a running system .....	27
Data @ Rest Encryption .....	28
Replication .....	28
VNX Snapshots for Block LUNs .....	28
SnapView for Block LUNs .....	29
SnapSure checkpoints for file systems .....	29
MirrorView for Block LUN replication .....	29
RecoverPoint for Block LUN replication .....	30
IP Replicator for File System replication .....	30
Deduplication and compression .....	30
Block LUN compression .....	30
Block LUN deduplication .....	30
Deduplication and compression with VNX OE for File .....	31
Anti-virus .....	32
File system CAVA .....	32
<b>Chapter 4 Application Specific Considerations .....</b>	<b>34</b>
Block application tuning .....	35
Host file system alignment .....	35
VMware ESX Server with iSCSI datastore .....	35
File application tuning .....	35
Hypervisor / Database over NFS or SMB .....	35
Bandwidth-intensive applications .....	36
<b>Conclusion .....</b>	<b>37</b>

As part of an effort to improve and enhance the performance and capabilities of its product line, EMC from time to time releases revisions of its hardware and software. Therefore, some functions described in this guide may not be supported by all revisions of the hardware or software currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, please contact your EMC representative.

**Note:** This document was accurate as of the time of publication. However, as information is added, new versions of this document may be released to EMC Online Support. Check the website to ensure that you are using the latest version of this document.

## Purpose

The *Applied Best Practices Guide* delivers straightforward guidance to the majority of customers using the storage system in a mixed business environment. The focus is on system performance and maximizing the ease of use of the automated storage features, while avoiding mismatches of technology. Some exception cases are addressed in this guide; however, less commonly encountered edge cases are not covered by general guidelines and are addressed in use-case-specific white papers.

Guidelines can and will be broken, appropriately, owing to differing circumstances or requirements. Guidelines must adapt to:

- Different sensitivities toward data integrity
- Different economic sensitivities
- Different problem sets

These guidelines contain a few “DON’T” and “AVOID” recommendations:

- DON’T means: Do not do it; there is some pathological behavior
- AVOID means: All else being equal, it is recommended not to, but it still acceptable to do it

## Audience

This document is intended for EMC customers, partners, and employees who are installing and/or configuring VNX2 unified systems. Some familiarity with EMC unified storage systems is assumed.

## Related documents

The following documents provide additional, relevant information. Access to these documents is based on your logon credentials. All of the documents can be found on <http://support.emc.com>. If you do not have access to the following content, contact your EMC representative.

- VNX2: Data at Rest Encryption
- Virtual Provisioning for the VNX2 Series - Applied Technology
- Introduction to the EMC VNX2 Series - A Detailed Review
- Introduction to EMC VNX2 Storage Efficiency Technologies
- VNX2 Multicore FAST Cache - A Detailed Review
- White Paper: VNX2 FAST VP - A Detailed Review
- White Paper: VNX2 Deduplication and Compression - Maximizing effective capacity utilization
- White Paper: VNX2 MCx - Multicore Everything
- White Paper: VNX Replication Technologies - An Overview
- White Paper: VNX Snapshots
- Host Connectivity Guide

# Chapter 1 System Configuration

This chapter presents the following topics:

Essential guidelines.....	8
Storage Processor cache .....	8
Physical placement of drives.....	8
Hot sparing.....	8
Usage of flash drives in hybrid flash arrays.....	9
Availability and connectivity .....	9

## Essential guidelines

This guide introduces specific configuration recommendations that enable good performance from a VNX2 unified storage system. At the highest level, good performance design follows a few simple rules. The main principles of designing a storage system for performance are:

- Flash First – Utilize flash storage for the active dataset to achieve maximum performance
- Distribute load over available hardware resources
- Design for 70 percent utilization (activity level) for hardware resources
- When utilizing Hard Disk Drives (HDD), AVOID mixing response-time-sensitive I/O with large-block I/O or high-bandwidth sequential I/O
- Maintain latest released VNX Operating Environment version

## Storage Processor cache

Storage Processor memory configuration is not required. Memory allocation amounts and cache page size are not configurable parameters.

## Physical placement of drives

When initially placing drives in the array:

- Spread flash drives across all available buses, and when possible place them in the lowest-numbered enclosures
- There are no restrictions around using or spanning across Bus 0 Enclosure 0

## Hot sparing

Hot sparing is the process of rebuilding a failed drive's data onto a system-selected compatible drive. Any unbound non-system drive can be considered for sparing. When planning Hot Spares consider the following recommendations:

- Plan to reserve at least one of every 30 installed drives of a given type
  - Verify count in the GUI or CLI
    - System->Hot Spare Policy
    - `naviseccli hotsparepolicy -list`
  - **Note:** Unbound system drives (Bus 0 Enclosure 0 Disk 0 through Disk 3) cannot be used as hot spares
- Ensure that unbound drives for each drive type are available
  - SAS Flash must spare for SAS Flash
  - SAS Flash VP must spare for SAS Flash VP
  - SAS must spare for SAS (regardless of rotational speed)

- NL-SAS must spare for NL-SAS

The capacity of an unbound drive should be equal to or larger than the provisioned drives for which it will spare.

## Usage of flash drives in hybrid flash arrays

EMC recommends the use of flash drives in VNX storage systems to maximize the potential of the MCx operating environment. EMC recommends deploying flash drives in the following priority order:

- Configure Multicore FAST Cache first
  - Multicore FAST Cache is a global resource that can benefit all storage resources
  - **Note:** Multicore FAST Cache is not applicable for all-flash arrays
- Next, add a flash tier to pools containing thin LUNs
  - The flash tier can accelerate access to thin LUN metadata, improving performance
  - Configure at least 3% of pool capacity in flash, to capture metadata
  - Thin Provisioned LUNs, VNX Snapshots, Block Compression, and Block Deduplication all rely on thin LUN technology
- Then add a flash tier to pools utilizing FAST VP
  - Configure at least 10% of pool capacity, for flash acceleration of active workload
  - Configure at least 25% of pool capacity, for near-all-flash performance
- Finally, dedicate an all-flash pool to storage objects with very high performance requirements

More details on the effective use of flash drives for these purposes can be found in the relevant sections in this paper.

## Availability and connectivity

The VNX2 unified storage array offers connectivity to a variety of client operating systems, and via multiple protocols, such as FC, iSCSI, NFS, and CIFS. EMC provides connectivity guides with detailed instructions for connecting and provisioning storage via different protocols to the specific host types.

It is highly recommended that you consult connectivity documents on <http://support.emc.com> for the host types that will be connected to the array for any specific configuration options.

- Host connectivity guides cover more detail, especially for a particular operating system; reference them for host-specific connectivity guidelines

## Fibre Channel and iSCSI connectivity

Fibre Channel connectivity is facilitated via the FC I/O modules on the Block Storage Processors. iSCSI connectivity is facilitated via the iSCSI I/O modules on the Block Storage Processors.

- Use multiple I/O ports on each SP, and balance host port connections across I/O ports, as host port connections affect the preferred CPU core assignment
- If not connecting all of the available I/O ports, use the even numbered ports on each I/O module before using any odd numbered ports
- Initially skip the first FC and/or iSCSI port of the array if those ports are configured and utilized as MirrorView connections
- For the VNX8000, engage the CPU cores from both CPU sockets with Front End traffic
  - Balance the Front End I/O modules between slots 0-5 and slots 6-10
    - DON'T remove I/O modules if they are not balanced; instead, contact EMC support
  - Balance host port assignments across I/O Module slots 0-5 and 6-10
- AVOID zoning every host port to every SP port

When registering host HBAs with VNX OE for Block Storage Groups, make sure to set the appropriate failover mode based on the host type. See the Host Connectivity Guides for details.

For Fibre Channel:

- Ensure that the FC ports connect at the highest speed supported by the environment, preferably 16Gb or 8Gb
- Consider the port count in use when performance targets are required
  - 16Gb ports have a max capability of 90,000 IOPS, or 1,050 MB/s
  - 8Gb ports have a max capability of 60,000 IOPS, or 750 MB/s

For iSCSI:

- Use 10Gbps for the best performance
  - 10Gb ports have a max capability of 40,000 IOPS, or 1,200 MB/s
- Configure Jumbo Frames (MTU of 9000) on all iSCSI ports
  - **Note:** The entire network infrastructure must also support Jumbo Frames
- When possible, segregate iSCSI traffic onto dedicated storage networks

## NAS connectivity

NAS protocols (NFS and SMB/CIFS) are facilitated via I/O modules on the File data movers.

- Use 10Gbps for the best performance
- Configure Jumbo Frames (MTU of 9000) on all NAS ports
  - **Note:** The entire network infrastructure must also support Jumbo Frames
- It's recommended to use network trunking and multipathing to provide port failover and greater aggregate bandwidth for NAS connections to a single DM
  - Configure LACP across 2 or more ports on a single DM
    - Use LACP instead of EtherChannel

## Chapter 2 Storage Configuration

This chapter presents the following topics:

General considerations .....	13
Determining which LUN type to configure .....	15
Storage pool considerations .....	17
Storage pool object considerations .....	19
Classic RAID Group LUN considerations .....	22

## General considerations

### Drive type

Match the appropriate drive type to the expected workload:

Drive type	Workload type
SAS Flash	For extreme performance; these provide the best performance for transactional random workloads, and the lowest write service times. Required for Multicore FAST Cache
SAS Flash VP	For extreme performance FAST VP tier; these are a higher capacity flash option. Not for use with Multicore FAST Cache.
SAS	For general performance tier.
NL-SAS	For less active data, well-behaved streaming data, archive purposes, and backups.

### Rules of thumb

Disk drives are a critical element of unified performance. Use the rule of thumb information to determine the number of drives to use to support the expected workload.

Rule of thumb data is based on drives that are:

- Operating at or below recommended utilization levels
- Providing reasonable response times
- Maintaining overhead to handle bursts or hardware failures

These guidelines are a conservative starting point for sizing, not the absolute maximums.

### Rules of thumb (RoT) for drive bandwidth (MB/s)

Bandwidth	NL-SAS	SAS 10K	SAS 15K	Flash (All)
RoT per drive, Sequential Read	15 MB/s	25 MB/s	30 MB/s	90 MB/s
RoT per drive, Sequential Write	10 MB/s	20 MB/s	25 MB/s	75 MB/s

This chart gives the expected per-drive bandwidth of the different drive types when servicing sequential workloads. Disk drives deliver optimal bandwidth when the workload consists of:

- Large-block I/O (128KB or larger)
- Multiple concurrent sequential streams

EMC recommends the use of parity RAID (RAID-5 or RAID-6) for predominantly sequential workloads. When sizing for bandwidth with RoT, do not include parity drives in the calculations.

- For example, to estimate the MB/s of a 4+1 RAID group, multiply the appropriate value from the chart by 4 (the number of non-parity drives)
  - SAS 15K, RAID-5 4+1, with sequential write:  $4 * 25 \text{ MB/s} = 100 \text{ MB/s}$

## Rules of thumb (RoT) for drive throughput (IOPS)

Throughput	NL-SAS	SAS 10K	SAS 15K	SAS Flash VP (eMLC)	SAS Flash (SLC)
Per drive RoT	90 IOPS	150 IOPS	180 IOPS	3500 IOPS	5000 IOPS

This chart gives the expected per-drive IOPS of the different drive types when servicing multi-threaded random workloads. Disk drives deliver optimal IOPS when the workload consists of:

- Small-block I/O (64KB or smaller)
- Multiple parallel workload threads, sending concurrent activity to all drives

When drives are combined with RAID protection, additional drive I/O is needed to service random writes from the host.

- To size for host IOPS, you must include the RAID overhead as described in the section [Calculating disk IOPS by RAID type](#)

System drives (Bus 0 Disk 0 through Bus 0 Disk 3) have reduced performance expectations due to the management activities they support; rules of thumb for these drives are adjusted accordingly

- **Note:** The system drives cannot be included in storage pools in the VNX2

## RAID level

For best performance from the least number of drives, match the appropriate RAID level with the expected workload:

RAID level	Expected workload
RAID 1/0	Works best for heavy transactional workloads with high (greater than 30 percent) random writes, in a pool with primarily HDDs
RAID 5	Works best for medium to high performance, general-purpose and sequential workloads
RAID 6 for NL-SAS	Works best with read-biased workloads such as archiving and backup to disk RAID 6 provides additional RAID protection to endure longer rebuild times of large drives

## Calculating disk IOPS by RAID type

Front-end application workload is translated into a different back-end disk workload based on the RAID type in use.

For reads (no impact of RAID type):

- 1 application read I/O = 1 back-end read I/O

For random writes:

- RAID 1/0 - 1 application write I/O = 2 back-end write I/O
- RAID 5 - 1 application write I/O = 4 back-end disk I/O (2 read + 2 write)
- RAID 6 - 1 application write I/O = 6 back-end disk I/O (3 read + 3 write)

Example for calculating disk IOPS from host IOPS:

Host IOPS required = 3000, with a read to write ratio of 2 to 1, using RAID 5.

2 out of every 3 host I/O is a read.

$$\text{Disk reads} = (2 * (3000 / 3)) = 2000$$

RAID 5 requires 4 disk I/O for every host write. 1 out of every 3 host I/O is a write.

$$\text{Disk writes} = 4 * (1 * (3000 / 3)) = 4000$$

$$\text{Total disk IOPS} = 2000 + 4000 = 6000$$

If looking to support that required workload with 15K rpm SAS drives, one would simply divide the rule of thumb into the required backend IOPS:

$$6000 / 180 = 33.3, \text{ so round up to 35 to align with a preferred drive count of the RAID 5 option.}$$

## Determining which LUN type to configure

The VNX2 storage system supports multiple types of LUNs to meet the demands of different workloads and support different features. In general, Thin Pool LUNs are required for Block space efficiency features. Pool LUNs in general (either Thin or Thick) are required for FAST VP tiering. Classic RAID Group LUNs do not provide support for advanced features.

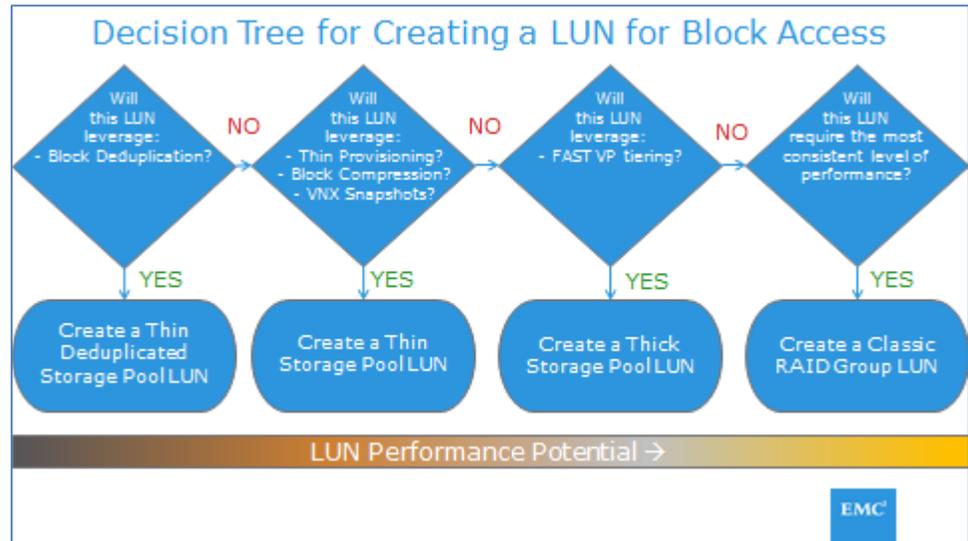
In terms of performance:

- Thin LUNs provide good performance for most workloads
- Thick LUNs can provide higher performance than Thin, given the same platform and drive complement, by removing the CPU and IOPS load of certain features
- Classic RAID Group LUNs provide the most consistent performance levels, for environments where variance in performance cannot be tolerated.

Use the following charts to determine which is appropriate for your environment.

## Creating LUNs for Block access

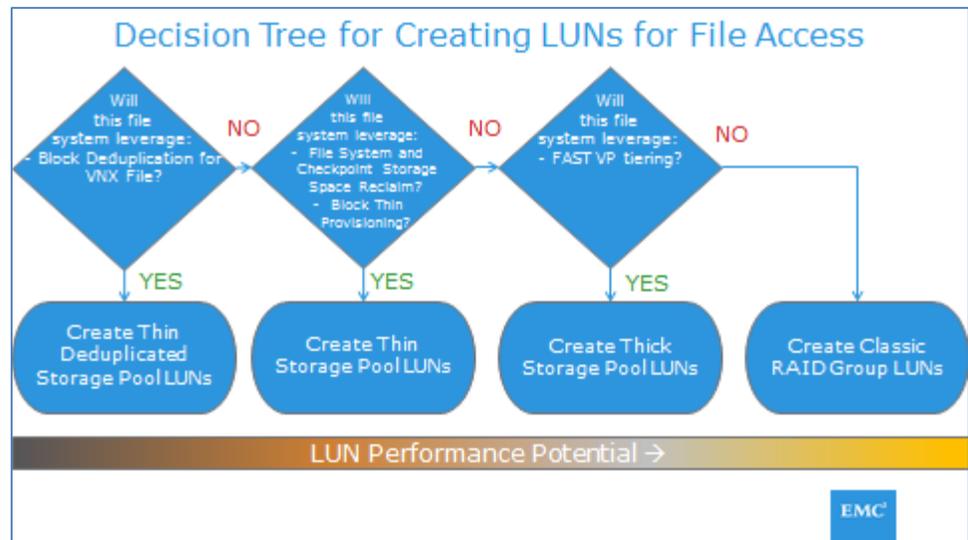
When creating a Block LUN, determine the desired feature set for this LUN from the chart below, and then create the appropriate LUN type.



See the appropriate sections in this document for best practice recommendations for configuring the LUN type and features selected.

## Creating LUNs for File access

VNX OE for File builds file systems using Block LUNs. When creating Block LUNs for File, determine the desired feature set for from the chart below, and then create the appropriate LUN type.



See the appropriate sections in this document for best practice recommendations for configuring the LUN type and features selected.

## Storage pool considerations

### Storage pool creation and expansion

Create multiple pools in order to:

- Separate workloads with different I/O profiles
  - Predominantly sequential workloads should be placed in dedicated pools or Classic RAID Groups
- Dedicate resources, when you have specific performance goals
- Vary pool parameters, such as Multicore FAST Cache enabled/disabled
- Minimize failure domains
  - Although unlikely, loss of a private RAID group in the pool compromises the total capacity of that pool; it may be desirable to create multiple smaller pools rather than configure the total capacity into a single pool

Storage pools have multiple RAID options per tier for preferred type and drive count

- Consider the following rule of thumb for tier construction:
  - Extreme performance flash tier: 4+1 RAID 5
  - Performance SAS tier: 4+1 or 8+1 RAID 5
  - Capacity NL-SAS tier: 6+2 or 8+2 RAID 6
  - **Note:** Classic RAID Groups pools have different recommended preferred drive counts as described in the section on Classic RAID Group creation
- Use RAID 5 with a preferred drive count of 4+1 for the best performance versus capacity balance
  - Using 8+1 improves capacity utilization at the expense of reduced availability
- Use RAID 6 for NL-SAS tier
  - Preferred drive counts of 6+2, 8+2, or 10+2 provide the best performance versus capacity balance
  - Using 14+2 provides the highest capacity utilization option for a pool, at the expense of slightly lower availability and performance
- Use RAID 1/0 when a high random write rate (> 30%) is expected with HDD
  - For best possible performance with RAID 1/0, use the largest available preferred drive count (i.e., 4+4 > 3+3 > 2+2, etc.)

Recommendations for creating and expanding storage pools:

- When creating a pool, it is best to specify a multiple of the preferred drive count for each tier you configure
  - For example, when using RAID 5 4+1, specify a drive count of 5, 10, 15, etc.
- It is best to maintain the same capacity and rotational speed of all drives within a single tier of a given pool

- For example, *AVOID* mixing 600GB 10K SAS drives in the same pool with 300GB 15K SAS drives; instead, split them into 2 different pools
- Within a given pool, use all of the same flash technology for the extreme performance tier
- When expanding pools, use a multiple of the preferred drive count already in use for the tier being expanded

## Pool capacity considerations

It is recommended to leave approximately 10% free space in the storage pool, to accommodate data services.

**Note:** The pool can still be over-subscribed above 100% of actual capacity; the 10% referenced here refers to actual free capacity that is not used in the pool.

- FAST VP requires free space to perform efficient relocations; it attempts to keep 10% free per tier
- VNX Snapshots requires at least 5% free to buffer snapped writes
- Block Deduplication uses free space to buffer write-splits
- Maintaining a total of 10% free will meet the requirements of all features

**Note:** By default, the VNX2 will begin issuing alerts when more than 70% of available capacity has been subscribed.

*AVOID* over-subscribing pools which contain thin LUNs for VNX File.

- VNX File requires free space in the storage pool for normal functioning

## Storage tiers

The number of tiers required in a storage pool is influenced by the following:

- Performance requirements
- Capacity requirements
- Knowledge of the skew between active and inactive capacity

The capacity required for each tier depends on expectations for skew, which is the locality of active data within the total storage capacity. Best performance is achieved when the entire active dataset can be contained within the capacity of the Extreme Performance (flash) and Performance (SAS) tiers.

As a starting point, consider capacity per tier of 10 percent flash, 20 percent SAS, and 70 percent NL-SAS. This works on the assumption that less than 30 percent of the used capacity will be active and infrequent relocations from the lowest tier will occur.

If the active capacity is known, the capacity per tier should be sized accordingly. Best performance is achieved when the active capacity fits entirely in the top tier.

In summary, follow these guidelines:

- When Multicore FAST Cache is available, use a 2-tier pool comprised of SAS and NL-SAS and enable Multicore FAST Cache as a cost-effective way of realizing flash performance without dedicating flash to this pool
  - Flash tier can be added later if Multicore FAST Cache is not fully capturing the active data

- For a 3-tier pool, start with 10 percent flash, 20 percent SAS, and 70 percent NL-SAS for capacity per tier if skew is not known
  - Tiers can be expanded after initial deployment to effect a change in the capacity distribution if needed
- For a 2-tier pool, combining flash and SAS is an effective way of providing consistently good performance.
  - The SAS tier provides a buffer for active data not captured in the flash tier; the SAS tier still provides modest performance, as well as quicker promotion to flash when relocations occur
  - NL-SAS can be added later if capacity growth and aged data require it
  - AVOID using a 2-tier pool of flash and NL-SAS if there is uncertainty about the active data fitting in the flash tier
- Add a flash tier to a pool with thin LUNs so that metadata is promoted to flash and overall performance is improved
  - Consider approximately 3GB of flash capacity to capture each TB of active thin LUN capacity
  - Thick LUNs employing VNX Snapshots, Compression, or Deduplication will convert to thin LUNs and therefore also require metadata consideration

## Tiering policies

When creating LUNs in tiered pools:

- DON'T use auto-tier for LUNs with low-skew random workloads where the active dataset would not fit in the highest tier
  - This might cause excessive tier relocations that may not benefit the active data
- AVOID using highest available when the LUN capacity exceeds 90% the highest tier capacity
  - This will affect the overall efficiency of the highest tier to service active data for LUNs running in auto-tier mode, and also remove the capability to hold metadata to help thin LUNs
- AVOID using lowest available with thin, deduplicated, or compressed LUNs
  - This will force the metadata to the lowest tier, which can negatively impact performance

## Storage pool object considerations

### Storage pool LUNs for Block access

Storage pool LUNs can be created as deduplicated thin, or thin (virtually provisioned), or thick (fully allocated).

### Deduplicated Thin LUNs for Block

When planning to use Block Deduplication, it is recommended to start with a Deduplicated Thin LUN.

When creating Deduplicated Thin LUNs, assign the same SP owner to all deduplicated LUNs in a given pool.

- All Deduplicated Thin LUNs in a pool reside in a single Deduplication Container, which is managed by a single SP
- Match the LUN ownership to the pool's "Optimal Deduplicated LUN SP Owner"

Balance SP utilization as follows when creating Deduplicated Thin LUNs:

- In a single pool, assign deduplicated LUNs to one SP, and non-deduplicated LUNs to the other SP
- With multiple pools, balance the Deduplication Containers across the SPs
  - For instance, Pool 1 has deduplicated LUNs on SPA, and Pool 2 has deduplicated LUNs on SPB

### Thin LUNs for Block

Thin LUNs are recommended when planning to implement Snapshots or Compression on Block LUNs

- Thin LUNs are recommended when storage efficiency requirements outweigh performance requirements
  - Thin LUNs maximize ease-of-use and capacity utilization at some cost to maximum performance
    - This does not always mean you will get slower response time or fewer IOPS, but the potential of the drives and SPs to deliver IOPS to the host is less
  - When using thin LUNs, adding a flash tier to the pool can improve performance
    - Thin LUN metadata can be promoted to the flash tier when FAST VP is enabled

### Thick LUNs for Block

Thick LUNs (without Snapshots) are recommended for the highest level of pool-based performance

- A thick LUN's performance can be better than the performance of a thin LUN

## Storage Pool LUNs for File access

### General guidelines for storage pool LUNs for File

In general, when creating pool LUNs for use with File:

- Pre-provision space from the pool; create and assign LUNs to VNX OE for File, so that VNX OE for File has available space for file system creation and extension, snapshots, etc.
- Create approximately 1 LUN for every 4 drives in the storage pool
- Create LUNs in even multiples of 10
  - Number of LUNs = (number of drives in pool divided by 4), rounded up to nearest multiple of 10
- Make all LUNs the same size

- Balance LUN ownership across SPA and SPB
- Apply the same tiering policies to all LUNs in the storage pool

File-based space efficiency features are generally recommended over Block LUN-based space efficiency features:

- If Virtual Provisioning™ is required for VNX OE for File, use a thin-enabled file system on classic or thick LUNs
- If compression is required for VNX OE for File, use VNX OE for File Deduplication and Compression
  - DON'T use compressed LUNs with VNX OE for File
    - Block Compression is not supported with VNX OE for File LUNs

If snapshots or checkpoints are required for VNX OE for File, use SnapSure.

- DON'T create VNX Snapshots on LUNs used by VNX OE for File

### Deduplicated Thin LUNs for File

If planning to implement Block Deduplication for VNX File, deduplicated thin LUNs are recommended. Block Deduplication must also be enabled on these LUNs from VNX File.

When planning to use Block Deduplication for VNX File:

- Create deduplicated thin LUNs for use by VNX File
- Create all deduplicated thin LUNs in a given pool with the same SP owner
  - Match the LUN ownership to the pool's "Optimal Deduplicated LUN SP Owner"
- Balance SP ownership by assigning to the other SP any LUNs in this pool that will not be deduplicated
- Once the LUNs are visible to VNX File, create a VNX File user-defined storage pool for LUNs that will be deduplicated; enable `fixed_block_dedup` on the user-defined storage pool
  - DON'T create a file system on the LUNs until `fixed_block_dedupe` has been enabled from VNX File

### Thin LUNs for File

Thin LUNs are required when planning to utilize File System and Checkpoint Storage Space Reclaim.

Follow the General Guidelines when creating Thin LUNs for File

## Thick LUNs for File

Thick LUNs are generally recommended for use with File when performance needs outweigh Block space efficiency requirements.

Follow the General Guideline when creating Thick LUNs for File. In addition:

- Allow LUNs to complete the prepare process (thick LUN slice allocation) before adding to the File storage group; use this command to display the status of the prepare operation:
  - `naviseccli lun -list -opDetails`

## Creating file systems from storage pool LUNs

When creating file systems from Block storage pool virtual LUNs, consider the following:

- Use Automatic Volume Management (AVM) to create file systems
- When creating volumes manually:
  - Stripe across 5 LUNs from the same pool
  - Use a stripe size of 262144 (256KB)
  - Balance SP ownership of LUNs in each stripe
- Create file systems with “split” log type
  - “common” log type is only recommended when replicating to or from other common log file systems
- Use thin enabled file systems for optimum capacity utilization

## Storage pool LUN ownership considerations

AVOID changing the ownership of a pool LUN after initial creation.

- Use LUN migration to move LUNs to their peer SP if required to change ownership

## Classic RAID Group LUN considerations

Classic RAID Group LUNs generally provide the most consistent level of performance. Classic RAID Group LUNs are recommended for workloads where performance variation cannot be tolerated. Classic RAID Group LUNs also provide the highest performance with sequential workloads.

## Drive location selection

When selecting the drives to use in a Classic RAID group, drives can be selected either from a DAE or DAEs on the same bus (horizontal positioning), or from multiple DAEs on different buses (vertical positioning). There are now almost no performance or availability advantages to using vertical positioning. Therefore:

- Use the default horizontal positioning method of drive selection when creating RAID groups.
- If a single DAE does not have enough available drives to create the RAID group, selecting drives from a second DAE is acceptable.

- The SAS buses are fast enough to support drives distributed across enclosures belonging to the same RAID group.

## Classic RAID Group LUNs for Block access

### Drive count

Classic RAID groups can have a maximum count of 16 drives. For parity RAID, higher disk count offers higher capacity utilization, but with higher risk to availability.

For predominantly random workloads, use the drive type rule of thumb values and the RAID level IOPS calculations to determine the number of drives needed to meet the expected application workload.

Sequential optimization occurs when the array can perform full-stripe operations normally associated with large-block I/O. Certain drive counts are more likely to enable this behavior, though these optimizations can also occur with non-preferred drive counts.

With a predominance of large-block sequential operations, the following applies:

- RAID 5 has a preference of 4+1 or 8+1
- RAID 6 has a preference of 8+2
- RAID 1/0 has a preference of 4+4

**Note:** Storage pools have different recommended preferred drive counts as described in the section on Storage Pool Creation.

### MetaLUNs or host striping

A Classic RAID Group can contain a maximum of 16 drives. If Classic RAID Group LUNs are preferred, but more than 16 drives are required for a single LUN, VNX MetaLUNs can be used, or host striping can be employed.

If using host striping:

- Use a stripe element size equal to or a multiple of the RG full-stripe-width
- Stripe across LUNs from different SPs

### Large element size

When creating a 4+1 RAID 5 Classic RAID group for Block access, you can select a 1024 block (512KB) element size.

- Use large element size when the predominant workload is large-block random read activity (such as data warehousing)
  - AVOID using large element size with any other workloads
- The default element size of 128 blocks (64KB) is preferred for most workloads

## Classic RAID Groups for File access

### Drive count

When creating LUNs for VNX OE for File from Classic RAID groups, consider the following:

- Preferred RAID group sizes:
  - RAID 1/0: 1+1
  - RAID 5: 4+1, 8+1

- RAID 6: 8+2, 4+2
- Create 1 LUN per RAID Group, up to the maximum LUN size of 16TB
  - Create multiple LUNs when necessary to keep LUN size less than 16TB
- Balance SP ownership of LUNs from multiple RAID Groups

### Creating file systems from Classic RAID Group LUNs

Guidelines for creating file systems from Classic RAID group LUNs:

- Use Automatic Volume Management (AVM) to create file systems
  - AVM creates well-designed file system layouts for most situations
- Manual Volume Management can be used to create file systems with specific attributes for specialized workloads
  - For metadata-heavy workloads, stripe across an odd number of disk volumes (dVols)
  - Match VNX OE for File stripe size to the Block LUN full-stripe width for best sequential write performance
    - AVOID wide striping across a large number of dVols
- When creating striped volumes manually:
  - Stripe across LUNs from different RAID Groups
  - Use a stripe size of 262144 (256KB)
  - Balance SP ownership of selected dVols

## Chapter 3 Data Services

This chapter presents the following topic:

FAST VP .....	26
Multicore FAST Cache.....	27
Data @ Rest Encryption.....	28
Replication .....	28
Deduplication and compression.....	30
Anti-virus.....	32

# FAST VP

## General

FAST VP moves data between tiers in a pool based on the performance needs of the data. Construct the pool such that each tier will provide consistent performance.

- Use consistent drive technology for each tier within a single pool
  - Same flash drive technology and drive size for the extreme performance tier
  - Same SAS RPM and drive size for the performance tier
  - Same NL-SAS drive size for the capacity tier

## Data relocation

Relocation is the process of moving pool data slices across tiers or within the same tier, to move hot data to higher performing drives or to balance underlying drive utilization. Relocation can occur as part of a FAST VP scheduled relocation, as an automated relocation after a storage pool expansion, or as a result of manually requesting relocation.

- Enable FAST VP on a pool, even if the pool only contains a single tier, to provide ongoing load balancing of data across available drives based on slice temperature and capacity utilization
- Schedule relocations for off-hours, so that relocation activity does not contend with the primary workload.
- Schedule relocations to run before or during backup windows, so that the relocations are based on the primary workload activity
- Enable FAST VP on a pool before expanding the pool with additional drives.
  - With FAST VP enabled, slices rebalance according to slice temperature and capacity utilization of pool resources.

## Pool capacity utilization

FAST VP requires unallocated space within the pool to accommodate data relocations.

- It is recommended to leave about 10% free space in storage pools with FAST VP enabled
  - Relocation will attempt to reclaim 10% free per tier
  - Free space is used to optimize relocation operations
  - Free space is used for new allocations to thin LUNs
  - Free space is used to support Snapshot schedules

## Considerations for VNX OE for File

By default, a VNX OE for File system-defined storage pool is created for every VNX OE for Block storage pool that contains LUNs available to VNX OE for File. (This is a “mapped storage pool.”)

- All LUNs in a given VNX OE for File storage pool should have the same FAST VP tiering policy
- Create a user-defined storage pool to separate VNX OE for File LUNs from the same Block storage pool that have different tiering policies

When using FAST VP with VNX OE for File, use thin enabled file systems for increased benefit from FAST VP multi-tiering.

## Multicore FAST Cache

Multicore FAST Cache is best for small random I/O where data has skew. The higher the locality, the greater the benefits of using Multicore FAST Cache. Multicore FAST Cache also adapts quickly to changes in locality.

### General considerations

EMC recommends first utilizing available flash drives for Multicore FAST Cache, which can globally benefit all LUNs in the storage system. Then supplement performance as needed with additional flash drives in storage pool tiers.

Preferred application workloads for Multicore FAST Cache:

- Small-block random I/O applications with high locality
- High frequency of access to the same data, not entirely serviced from system cache
- Systems where current performance is limited by HDD capability, not SP capability

AVOID enabling Multicore FAST Cache for pools that are not expected to benefit, such as when:

- The primary workload is sequential
- The primary workload is large-block I/O
- The primary workload is small-block sequential, like database logs, circular logs, or VNX OE for File SavVol (snapshot storage)

### Creating FAST Cache

Choose drives from multiple buses when possible.

### Enabling Multicore FAST Cache on a running system

When adding Multicore FAST Cache to a running system, it is recommended to enable Multicore FAST Cache on a few LUNs at a time, and then wait until the LUNs have reached steady state in Multicore FAST Cache before enabling more.

**Note:** For storage pools, Multicore FAST Cache is a pool-wide feature so you have to enable/disable at the pool level (for all objects in the pool).

Multicore FAST Cache can improve overall system performance if the current bottleneck is drive-related, but boosting the IOPS will result in greater CPU utilization on the SPs. On an existing system, check the SP CPU utilization of the system, and then proceed as follows:

- Less than 60 percent SP CPU utilization – enable a few LUNs or one pool at a time; let it reach steady state in Multicore FAST Cache, and ensure that SP CPU utilization is still acceptable before enabling Multicore FAST Cache on more LUNs/pools
- 60-80 percent SP CPU utilization – scale in carefully; enable Multicore FAST Cache on one or two LUNs, or one pool with the smallest capacity, and verify that SP CPU utilization does not go above 80 percent
- CPU greater than 80 percent – DON'T activate Multicore FAST Cache

## Data @ Rest Encryption

Enable Data @ Rest Encryption (D@RE) before populating the storage system with host data, to avoid any performance impact from the data-in-place encryption process.

- Order the storage system with D@RE factory enabled
- For systems already in-house, enable D@RE before creating storage pools or LUNs
  - NOTE: Multicore FAST Cache must be disabled before D@RE can be enabled which might further impact performance

## Replication

### VNX Snapshots for Block LUNs

VNX Snapshots are used to take point-in-time checkpoints of pool LUNs.

- Start with thin LUNs if planning to use Snapshots
  - Thick LUNs are eventually converted to thin LUNs once a VNX Snapshot is created on them; all new writes and overwrites require thin
- Plan the deletion of snapshots
  - Whenever possible, schedule the deletion of Snapshots during non-peak hours of operation
    - If snapshots must be deleted during peak periods of array activity, lessen the impact by reducing the number of concurrent Snapshot deletes (for example, stagger the delete operations over several hours, instead of all at once)
  - DON'T delete the last snapshot of a Thick LUN, if you intend to create another snapshot immediately after deleting the last snapshot
    - Create the new snapshot before deleting the older snapshot

- Deleting the last snapshot of a Thick LUN will undo the thin conversion, which would then be reconverted for the new snapshot
- For additional technical information on VNX Snapshots, refer to *EMC VNX Snapshots* at EMC Online Support.

## SnapView for Block LUNs

For SnapView™:

- Use SAS drives for reserve LUN pool (RLP) configurations, with write cache-enabled LUNs
- Match the secondary side RLP configuration to the primary side.
- AVOID configuring RLP on the same drives as the primary and secondary LUNs to avoid drive contention
- DON'T enable Multicore FAST Cache on RLP LUNs
  - RLP exhibits multiple small-block sequential streams that are not suited for Multicore FAST Cache
- DON'T enable Multicore FAST Cache on clone private LUNs (CPL)

## SnapSure checkpoints for file systems

SnapSure Checkpoints are used to take point-in-time checkpoints of file systems.

If using SnapSure™ to create user snapshots of the primary file system:

- SnapSure sizing:
  - Size the disk layout for the primary file system (PFS) to include copy-on-first-write activity
  - Include one additional read I/O from the PFS, and one additional write I/O to the SavVol, for every host write I/O
  - Size the layout for SavVol based on expected user load to the snapshot file systems
- DON'T disable SmartSnap traversal

## MirrorView for Block LUN replication

For MirrorView™:

- AVOID enabling Multicore FAST Cache on MirrorView secondary LUNs
  - MV/S secondary LUNs replicate only writes from the source and serviced well by SP cache
  - MV/A secondary LUNs replicate writes during updates and incur copy-on-first-write activity; this can incur additional Multicore FAST Cache promotions that do not lead to performance gain
- DON'T enable Multicore FAST Cache on Write Intent Log (WIL)
- With MirrorView/A, follow RLP guidelines in SnapView section

## RecoverPoint for Block LUN replication

For RecoverPoint:

- DON'T enable Multicore FAST Cache for RecoverPoint journal LUNs
  - Journals exhibit primarily large-block sequential activity not suited for Multicore FAST Cache use

## IP Replicator for File System replication

If using VNX File IP Replicator:

- Size disk layout for the primary file system to include SnapSure copy-on-first-write and replication transfer activity
- Place the secondary file system on the same drive type as the primary file system
- It is usually acceptable to place SavVol on NL-SAS for replication
  - If user snapshots are also enabled on the primary file system, then consider the user load from the snapshots to determine whether NL-SAS is still be adequate for SavVol
- Use 1GbE links for Replicator interconnects when traversing WAN links
  - 10GbE is typically not necessary for replication across a WAN
- When using high-latency networks with Replicator, use a WAN-accelerator to reduce latency

## Deduplication and compression

### Block LUN compression

Start with thin LUNs if planning to use Block compression.

- Classic or thick LUNs convert to thin LUNs when compressed

Manage the processes of Block Compression to minimize impact to other workloads.

- Pause or change the compression rate to Low at the system level when response-time critical applications are running on the storage system.
  - The option exists to pause compression at the system level to avoid compression overhead during spikes in activity
  - Pausing the compression feature ensures background compression, or associated space reclamation operations, do not impact host I/O

### Block LUN deduplication

Block deduplication works best in storage environments which include multiple copies of the same data that is read-only and remains unchanged.

- Evaluate the I/O profile of the workload and data content to determine if deduplication is an appropriate solution
  - AVOID using Block Deduplication on data that does not have a large amount of duplicated data; the added metadata and code path can

impact I/O response time and outweigh any advantage seen by deduplication

- In instances where portions of data on a single LUN are a good fit for Block Deduplication while other portions are not, consider separating the data onto different LUNs when possible
- AVOID deploying deduplication into a high write profile environment; this action either causes a cycle of data splitting/data deduplication or data splitting inflicting unnecessary overhead of the first deduplication pass and all future I/O accesses.
  - Each write to a deduplicated 8 KB block causes a block for the new data to be allocated, and updates to the pointers to occur; with a large write workload, this overhead can be substantial
- AVOID deploying deduplication into a predominantly bandwidth environment
- Use Multicore FAST Cache and/or FAST VP with Deduplication
  - Optimizes the disk access to the expanded set of metadata required for deduplication
  - Data blocks not previously considered for higher tier movement or promotion can now be hotter when deduplicated
- Start with Deduplication Enabled thin LUNs if preparing to use Block Deduplication, to avoid background LUN migration
- Use the default deduplication rate of Medium
  - Be aware that setting Deduplication Rate to High can impact CPU utilization
  - Set Deduplication Rate to Low during periods of heavy workload
- Use the Force Deduplication option sparingly
  - Recommended only if a large amount of new or changed data is added to the LUNs
  - Normal deduplication process is preferred for all other scenarios

## Deduplication and compression with VNX OE for File

Space efficiency for File is available via either VNX OE for File Deduplication and Compression, or Block Deduplication for VNX File.

Utilize VNX OE for File Deduplication and Compression when:

- Space savings must be available within the target file system
  - File Deduplication and Compression returns the freed blocks to the file system
- The target file system contains highly compressible data
  - File Deduplication and Compression uses compression, which is generally more effective with File data
- The target file system is sensitive to changes in performance

- File Deduplication and Compression can be utilized with file systems built on Classic RAID Group LUNs, or Thick or Thin storage pool LUNs

Utilize VNX Block Deduplication for VNX File when:

- Space savings from the target file system must be available to other Block LUNs in the Block storage pool
  - The deduplicated file system will not see the space savings, but other LUNs in the pool can use the space
- Multiple target file systems contain duplicate copies of data
  - Block Deduplication for VNX File can deduplicate data across multiple file systems if they are in the same Block storage pool

See specific recommendations for each solution below.

### VNX OE for File Deduplication and Compression

If using VNX OE for File deduplication and compression:

- Target deep compression at inactive files.
- DON'T enable CIFS compression on busy Data Movers.
  - CIFS compression occurs in-line with the host write and can delay response time if the compression activity is throttled.

### Block Deduplication for VNX File

If using Block Deduplication for VNX File:

- Plan to place all file systems with duplicate data in the same block storage pool
- Create file systems with “split” log type
- Ensure that all LUNs from the storage pool that will be used by File are owned by a single SP
- After diskmarking the LUNs on VNX File, enable Fixed-Block Deduplication via the File storage pool properties
  - Note: All LUNs and file systems in the File storage pool will participate in Block Dedupe

## Anti-virus

### File system CAVA

If using Common Anti-Virus Agent:

- Always consult with the antivirus vendor for their best practices
- Ensure CIFS is completely configured, tested, and working before setting up virus checker
- Ensure the antivirus servers are strictly dedicated for CAVA use only
- Ensure that the number CIFS threads used are greater than virus checker threads.

- Exclude real-time network scanning of compressed and archive files
- Set file mask to include only file types recommended for scanning by the antivirus provider
  - DON'T include \*.\*
- Disable virus scanning during migrations

## Chapter 4 Application Specific Considerations

This chapter presents the following topics:

Block application tuning .....	35
File application tuning .....	35

## Block application tuning

### Host file system alignment

File system alignment is covered in detail in “Host Connectivity Guide” documents on <http://support.emc.com>. In general:

- Windows Server 2008 and later automatically align
- Recent Linux operating systems automatically align
- When provisioning LUNs for older Windows and Linux operating systems that use a 63-block header, the host file system needs to be aligned manually

Alignment practices:

- Use host-based methods to align the file system
- EMC recommends aligning the file system with a 1 MB offset

### VMware ESX Server with iSCSI datastore

Disable “Delayed Ack” for iSCSI storage adapters and targets.

- For further detail, see VMware Knowledge Base article 1002598
  - <http://kb.vmware.com/kb/1002598>

Update ESX to avoid TCP Offload Engine Chimney issues

- For further detail, see VMware Knowledge Base article 2099293
  - <http://kb.vmware.com/kb/2099293>

## File application tuning

### Hypervisor / Database over NFS or SMB

Recommendation is to use cached mount in 8.x

- VNX OE for File cached I/O path supports parallel writes, especially beneficial for Transactional NAS.
  - Prior to VNX OE for File 8.x, file systems required the Direct Writes mount option to support parallel writes.
  - DON'T use Direct Writes
    - Direct Writes mount option (GUI); uncached mount (CLI).
- VNX OE file systems use cached mount by default, which provides buffer cache benefits.
  - Buffer cache ensures that files are 8KB aligned.
  - Allows read cache hits from data mover.

For VMWare ESX Server, install the vSphere plug-in and use it to provision storage

- Available from [support.emc.com](http://support.emc.com)

## Bandwidth-intensive applications

For bandwidth-sensitive applications over NFS:

- Increase the value of param **nfs v3xfersize** to **262144** (256KB).
- Negotiate 256KB NFS transfer size on client with:
  - `mount -o rsize=262144,wsiz=262144`

Single Data Mover bandwidth has a default nominal maximum of 1600 MB/s (unidirectional); this is due to the 2x 8Gb FC connections from the Data Movers to the Storage Processors.

- Scale VNX OE for File bandwidth by utilizing more active Data Movers (model permitting), or by increasing the number of FC connections per Data Mover from 2x to 4x.

Increasing the FC connectivity to 4x ports requires utilizing the 2x AUX FC ports; as a result, FC-based NDMP backups are not allowed directly from Data Mover. An RPQ is required for this change.

## Conclusion

This best practices guide provides configuration and usage recommendations for VNX2 unified storage systems in general usage cases.

For detailed discussion of the reasoning or methodology behind these recommendations, or for additional guidance around more specific use cases, see the documents the [related documents](#) section.