VNX - THE STORAGE [R]EVOLUTION A COMPARATIVE APPROACH WITH OTHER STORAGE PLATFORMS

Siva Shankari Chandrasekaran
Systems Engineer
Tata Consultancy Services

PROVEN PROFESSIONAL

EMC²

Table of Contents

Introduction

The pace of information growth is astonishing. Managing, sharing, processing, backing up, and maintaining ever-growing information is a very challenging task. Today's IT industry provides a number of products for data storage, retrieval, processing, and replication. Figuring out the right technology, architecture, and the right product per business requirements is tricky. Inevitably, factors such as security, performance, efficiency, scalability, manageability, and accessibility become vital.

The issue becomes determining the best way to gather all the necessary factors. Interestingly, the feel good factor is that the IT industry keeps revamping its strategies, evolving constantly to provide the best possible output. Thus, we get hybrid products which amazingly suit our requirements thereby drastically reducing otherwise hectic administrative tasks.

One such evolved concept is 'Unified Storage'. In this article, we will study:

- various types of data access solutions available such as block level, file level, and object level
- different network architectures available that support these various access solutions, with suitable examples
- process of Unified Storage Evolution

As a reference, we will take the industry's standard products which are already available, to study this new [R]evolution.

Various Data Access Solutions

There are different data access levels for different sets of requirements. The following topics briefly discuss the various access levels and their applications, advantages, and disadvantages.

Block Level Access

Block level access is the basis for accessing (read/write) data in the disk regardless of the raw or formatted access. Applications such as DBMS have the potential to access the raw disk and access data in the form of blocks. They directly access the storage device or LUN thereby bypassing the file system. Block level access is provided by Storage Area Network (SAN) systems. Common protocols are iSCSI and Fibre Channel. Block level access is the best option where high I/O throughput and performance are the key requirements.

File Level Access

File level access enables sharing of files with different users which is achieved by sharing the metadata across multiple clients. File level access is usually provided by a Network Attached Storage (NAS) system and is cost effective when compared with a Block level system. Common protocols are Network File System (NFS) and Common Internet File System (CIFS). Thus, it can provide access to heterogeneous operating systems such as UNIX, Linux, Windows, and so on. Administrative tasks such as authentication, authorization, file permission, access control, and so forth can be easily managed in this type of access.

Object-Based Access

Data objects are comprised of user data including data attributes and metadata. Object-based storage has the intelligence to associate the actual data and its metadata. The storage disk plays a main role in storing and retrieving objects. The object storage disk also takes care of allocating space and providing secure access. Object Identifier is the method used to retrieve an object. The storage determines the size of the data; the method of retrieval with respect to individual objects thereby enables flexibility and manageability. The advantages are: enhanced space utilization, security, scalability, robustness, and improved data access methods.

With respect to different data access levels, we have dedicated network architectures. We will have a concise study on the various prominent network architectures available.

Various Storage Network Architectures

Direct Attached Storage

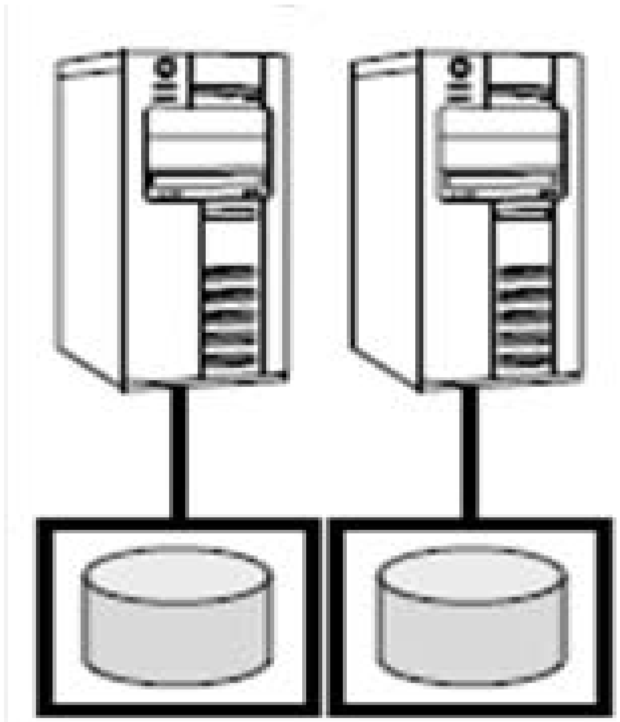


Figure 1: Direct Attached Storage
[Server has its own internal/external storage]

Direct attached storage (DAS) is the traditional storage design wherein the storage is directly attached to the server either internally or externally. Every server will have its own dedicated storage. The storage can be expanded as per the requirement. DAS also provides high bandwidth and good I/O throughput. It is cost-efficient and management tasks are relatively easier. It is a good solution for small businesses.

A major disadvantage of DAS is that the storage cannot be shared across servers, resulting in underutilization of resources.

Network Attached Storage

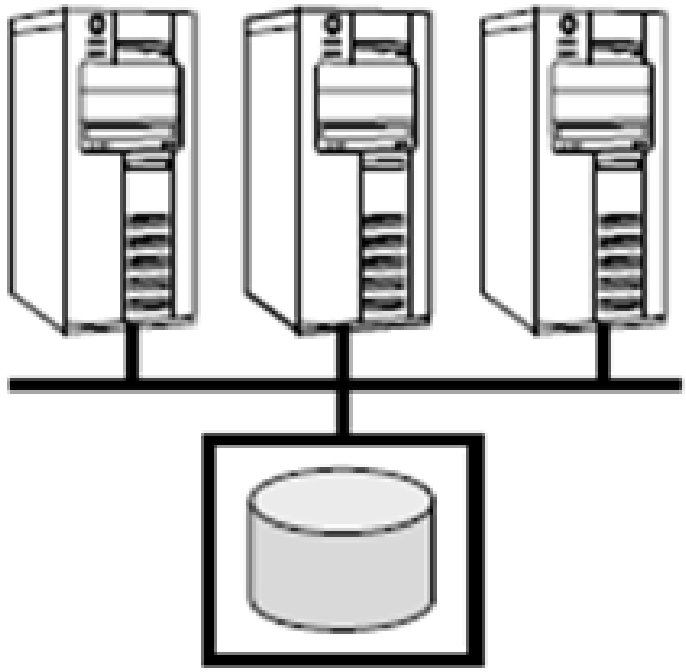


Figure 2: Network Attached Storage
[Server connected over a network to the storage (file based)]

Because direct attached storage resulted in islands of data and increased complexity in managing the network components separately, NAS evolved to meet the demand for consolidated and shared use of data in heterogeneous systems. NAS is based upon a dedicated file server which will handle only the file serving task, thus offloading data management and maintenance tasks. NAS can be accessed by multiple heterogeneous clients or other servers. Scaling the storage as per requirements for growth is also easier in NAS environments. Since NAS generally implements a specialized NAS device which will handle the file serving operations, scaling of storage or upgrading NAS components can be done independently. As storage is shared, the utilization rate ranges up to 95%. With respect to the network structure, we have two types of NAS: Integrated and Gateway.

Integrated NAS

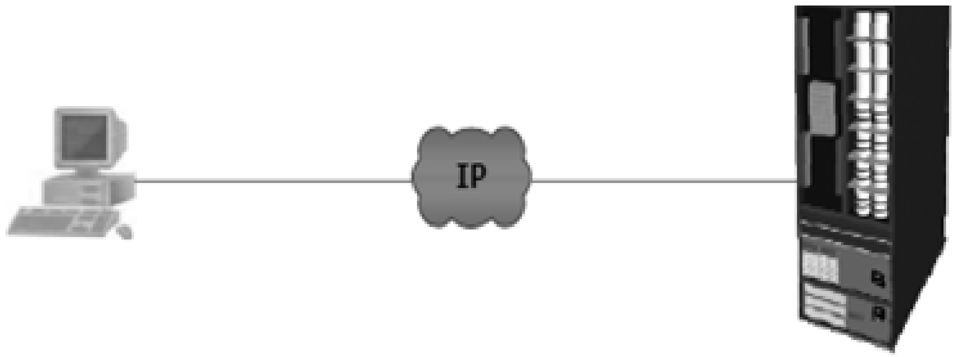


Figure 2.1 Integrated NAS Architecture
[NAS device encloses storage, NAS component]

In the integrated NAS set up, the storage, along with the NAS components, is enclosed in a single framework. It can also have an externally connected storage array. The NAS component connects to the network and caters the client file I/O requests.

Gateway NAS

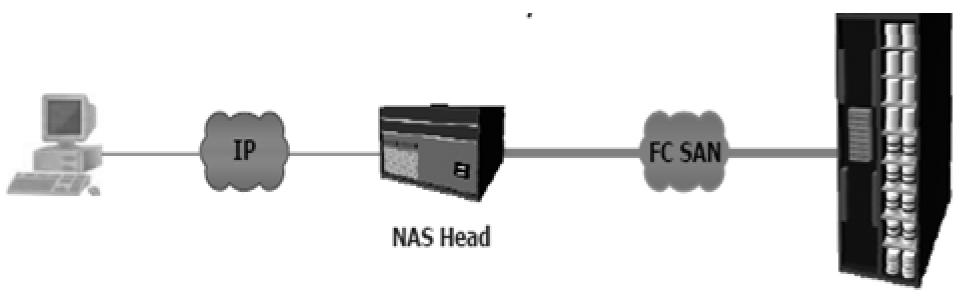


Figure 2.2 Gateway NAS Architecture
[NAS Head connects to the external storage, which in turn is shared by other applications]

In the gateway NAS set up, the storage is shared with other applications such as block level access applications other than the NAS environment. Gateway NAS architecture and management is complex when compared with Integrated NAS. The file I/O operations are the same as in Integrated NAS.

Benefits of NAS

- Centralization
- Better storage utilization
- Optimized management
- Cost effective
- Better collaboration

Shortcomings of NAS

- Suitable only for file-based applications
- Not suitable for high bandwidth applications as the network is shared

Let us consider EMC Celerra® as an example and study its architecture and features closely.

EMC Celerra

Celerra is EMC's NAS-based product. It supports NAS file access protocols such as CIFS, NFS, and FTP and block access protocols iSCSI, Fibre Channel, and so on.

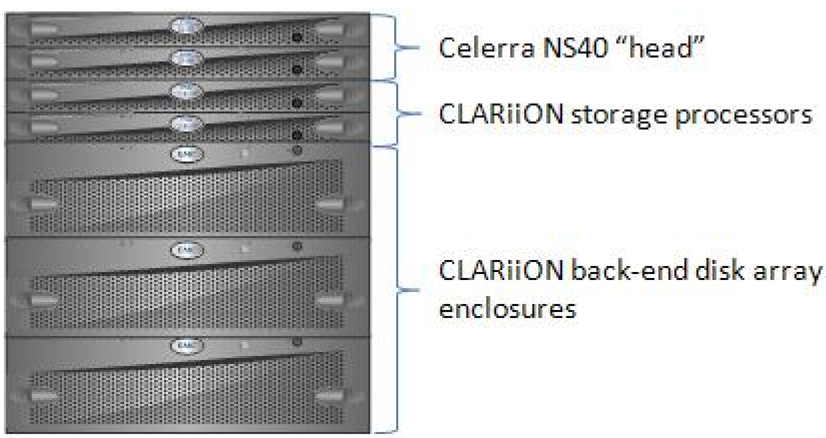


Figure 2.3 EMC Celerra NS40 Model
[Control Station and Data Mover embedded in NAS Head]

Celerra Components

- Data Mover/Blade
- Control Station
- DART
- Backend storage

Data Movers/Blades

Blades (also called Data Movers) are responsible for moving data between the storage and the client over the network. They are housed in a cabinet and act as independent storage devices. Blades are high-end servers containing memory, ports for network, storage connectivity, and other server components. Blades boot and get their storage from other storage arrays. Blade configuration and management is done via the Control Station. Blades cannot be accessed directly.

Control Station

Control Stations may exist in pairs for redundancy. EMC-specialized Linux OS is installed on the Control Station for installation, management, and configuration of blades and also to monitor the performance of all components and other additional features. It is used to manage volumes and file systems, configure network components, file system creation, extension, and maintenance. A blade doesn't depend upon Control Station for normal operations such as data flow. But, when a blade fails, failover occurs, only through Control Station. Control Station can be managed via GUI or CLI.

DART

Data Access in Real Time (DART) is the NAS operating system responsible for data movement between storage and the network. Installed in each blade, DART provides efficient data access and high throughput. It separates the data and control paths in a Celerra environment.

Backend Storage

Storage arrays such as Symmetrix® or CLARiiON® can be used as the backend storage for Celerra. The storage protection is taken care of by the array itself by mirroring, striping, parity protection, and RAID mechanisms. In an integrated NAS configuration, backend is CLARiiON array whereas in a Gateway configuration, it can be Symmetrix and/or CLARiiON.

Data Replication Features

Let us study the various replication and recovery features available in Celerra.

Celerra SnapSure

SnapSure™ is the point-in-time copy of the file system which does not contain the complete actual data; only the portions that are changed with respect to a time frame. The timeline-based views—referred to '*Checkpoints*', which are read-only—are created at regular intervals. This takes relatively less time than a complete backup since it is only a copy of the modified data. Celerra SnapSure supports both CLARiiON and Symmetrix® array backends.

TimeFinder/FS

TimeFinder®/FS creates a complete image of the file system. It can also be done dynamically. In that case, continuous mirroring of the file system will occur, during which the mirror copy will be read/write protected. The mirror image created by this replication feature can be used as a backup or for testing purposes.

Celerra SRDF/S

Celerra SRDF/S is a synchronous remote replication feature which is provided in the Celerra Gateway products with Symmetrix array as its backend. The file system is mirrored synchronously in the remote array which can be used as a disaster recovery solution. The distance limitation is 60km between the source and remote arrays.

Celerra SRDF/A

Celerra SRDF®/A is an asynchronous remote replication feature provided by the Celerra Gateway products with Symmetrix backend. Replication is done in asynchronous manner. There is no distance limitation in this mode.

Celerra MirrorView/S

Celerra MirrorView™/S provides the same functionality as Celerra SRDF/S except that the backend array is a CLARiiON array.

Celerra MirrorView/A

Celerra MirrorView/A provides the same functionality as Celerra SRDF/A except the backend array is Clariion array.

Celerra Replicator

Celerra Replicator is an-IP based replication which replicates within the same Celerra or can replicate to a remote system. It follows the 'checkpoint' mechanism as in SnapSure to ensure availability of the most recent copy. Initially, the replication will be a full backup followed by differential backup.

Storage Area Network

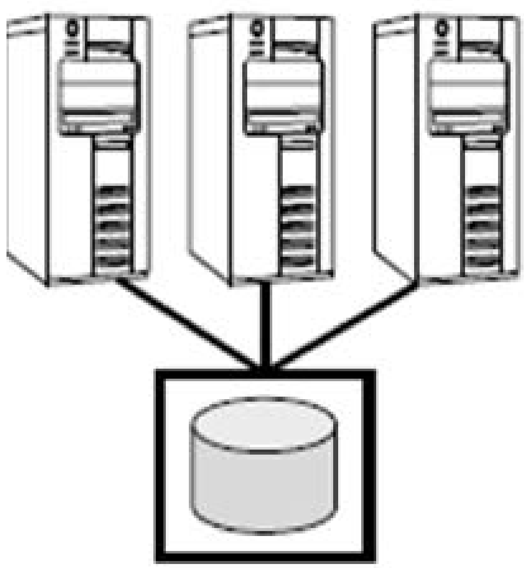


Figure 3: Storage Area Network
[Each server is directly connected to the storage (block based)]

File level data access requires additional processing capabilities, as data will be first accessed in the form of blocks and have to be mapped into file format. This operation requires additional hardware or software components. However, an application capable of accessing data in the block format bypasses these overheads and utilizes data more efficiently. Examples are database applications, enterprise resource planning applications, online transaction processing applications, and so on. They provide better backup and disaster recovery solutions.

Benefits of SAN

- Very efficient
- Provides great support for backup and disaster recovery processes
- High availability
- Enhanced resource utilization

Limitations of SAN

- Expensive to implement and manage
- Not suitable for small-sized business applications; though requirements would be less, expense would be huge

We will study EMC's popular SAN product, Symmetrix [Direct Matrix Architecture] with respect to its architecture and replication features.

EMC Symmetrix

EMC Symmetrix evolved from the Intelligent Storage Architecture, which aims to provide I/O optimization, intelligent caching, better storage capacity utilization, heterogeneous operating systems support, and enhanced replication features. Though the requirements are complex, management, scalability, security, and performance should be optimal. EMC Symmetrix supports protocols such as Fibre Channel [FC], iSCSI, Fibre Channel over Ethernet [FCoE] for open systems, as well as FICON and ESCON for mainframes. Symmetrix storage processors act as the backbone for the Symmetrix architecture, handling the I/O processing.

Symmetrix Components

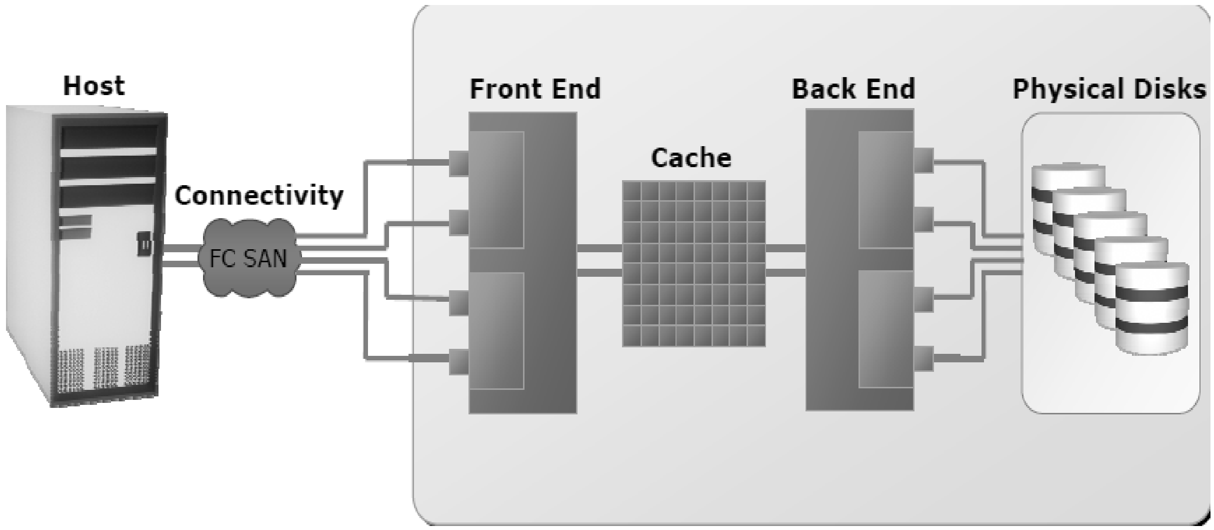


Figure 3.1 EMC Symmetrix Architecture

Front-End Directors

The front-end directors are responsible for connecting with the hosts through a switch for I/O operations. Each front-end director has 4 processors and two ports 0 and 1 for host connectivity. The processors have mezzanine cards which provide different emulations for different type of hosts. Front-end directors are paired to provide failover.

Cache

Cache memory is shared across all the directors, which provides better performance by reducing the time required to fetch data from physical disks. Frequently accessed data will be kept in the cache memory for further access. Also, the data to be written will be first written into the cache memory which will be slowly de-staged to the physical disks. This is a background process. Cache memory is mirrored for high availability.

Back-End Directors

The back-end directors attach to the physical disks and fetch the data. They communicate with the front-end directors, get the request for data, and provide it in the cache memory for host access through front-end directors.

Physical Disks

Symmetrix provides various RAID-level protections for the physical data. Also, we have various device types for different purposes such as business continuance, local and remote replication, virtual provisioning, and so forth. The physical disks are formatted as 'Logical Volumes' which will be visible to the host.

Every component in Symmetrix has failover capability to avoid data unavailability or data loss.

Data Replication Features

Local Replication

In local replication, the *replica*—the volume that contains the copy of data—exists in the same Symmetrix array. EMC's TimeFinder family of local replication products differ in functionality from one another.

TimeFinder/Mirror

TimeFinder/Mirror involves a special volume type called 'Business Continuity Volume' [BCV] for replication. The BCV attaches itself with the standard volume in a replication session. Data is synchronously mirrored from the standard to the BCV. Once the session is complete, the BCV can be used by other applications.

TimeFinder/Clone

TimeFinder/Clone doesn't require a special volume like TimeFinder/Mirror. It can establish with another standard volume itself. TimeFinder/Clone provides a *Point-in-Time* copy of the complete data. Once the replication session is activated, the clone can be accessed by other applications unlike TimeFinder/Mirror, where session termination is required.

TimeFinder/Snap

TimeFinder/Snap creates *Point-in-Time* snapshots, not the complete data, as in TimeFinder/Clone, thus saving storage space. When a session is established, it tracks the changes in data after activation and creates the copy of only the modified data.

Remote Replication

With the Remote Replication feature, data is copied to a remote storage array which is physically apart from the primary site. Symmetrix Remote Replication Feature [SRDF] has a family of products which are responsible for the remote replication of the primary Symmetrix Array.

Symmetrix Remote Replication Feature

SRDF has several product lines, i.e. SRDF/Synchronous [SRDF/S], SRDF/Asynchronous [SRDF/A], and SRDF/Star. Each of these products differ by their mode of data transfer. In SRDF/S, data is transferred synchronously, whereas in SRDF/A, data is transferred asynchronously. In SRDF/Star, a tertiary site is involved. From primary to secondary site, data transfer is synchronous, and to tertiary, it is asynchronous. Synchronous transfer imposes a distance limit, but is highly efficient for recovery and, thus, is suitable for business-critical applications; asynchronous transfer doesn't have a distance limitation, making it suitable for applications where a small amount of data loss is acceptable.

Data Migration Feature

Symmetrix Data Migration feature, not generally used as a backup or recovery solution, is meant for data migration from one Symmetrix array to another Symmetrix array. Migration can be done from third party arrays also.

SANCopy

SANCopy™, also known as *Open Replicator for Symmetrix* [ORS], is a highly efficient data migration feature of Symmetrix. It utilizes the Symmetrix FA ports' bandwidth for the data migration and thereby does not impose any overhead on the host side.

Unified Storage Infrastructure – A Quick Overview

Unified Storage is a convergence of SAN and NAS, which provides the benefits of both block- and file-level access in a unified manner. Thus, it increases storage utilization drastically, enhances performance, and provides unified administration and maintenance of otherwise chaotic storage networks.

Features anticipated from a unified storage infrastructure include:

- Intelligent management software support
- Support for variety of storage devices and storage tiering
- Flash drive support
- Thin/Virtual provisioning feature
- Cost efficient, less foot print
- Optimal data storage
- Advanced monitoring and reporting capabilities
- Efficient backup and recovery features

Unified storage is much more than just putting together the file and block access components in a single enclosure. The infrastructure should possess the following characteristics:

- Scalability
- Security
- Reliability
- Incomparable performance
- Cost and power saving efficiency

Let's look at how EMC's Unified Storage Infrastructure product, VNX®, encompasses the features of SAN and NAS network infrastructures and features which we studied in the previous topics.

What is VNX?

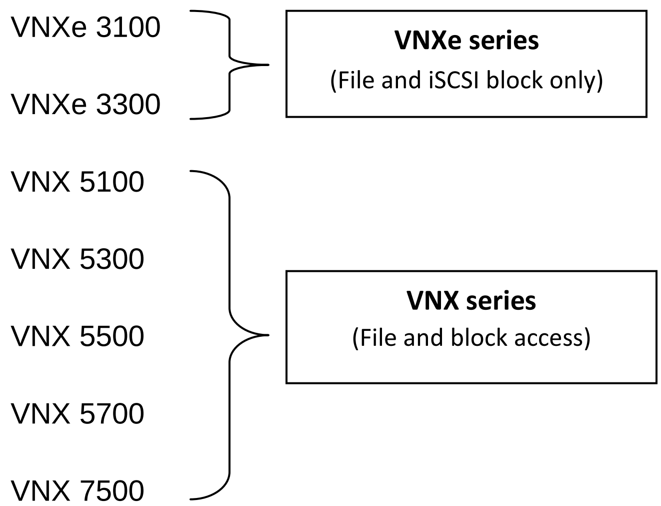
VNX stands for:

V – Optimized for **V**irtualization

N – stands for Celerra [**N**S20, **N**S40, etc.] – NAS Architecture platforms

X – stands for CLARiiON [C**X**3, C**X**4, etc.] – SAN Architecture platforms

VNX has a series of products that provide file and block level access to the data. It is highly unified in managing and retrieving the data, unified User Interface for administration, scalability, and packaging. The series of products are:



We will consider the high range product of VNX throughout this article, which has support for all types of file and block sharing capabilities. It also provides very good support for replication, security, and application features.

Unified Storage Architecture

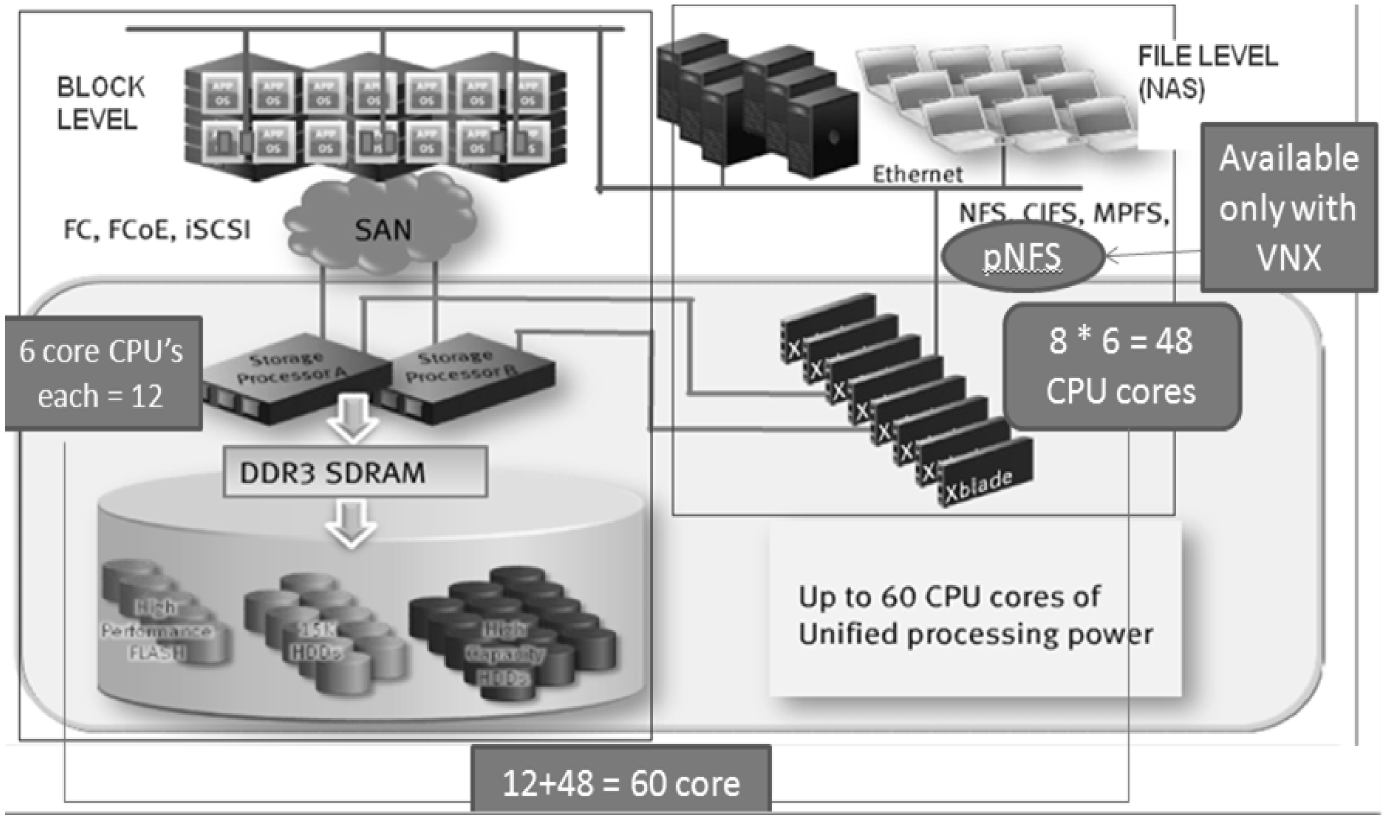


Figure 4: Unified Storage Architecture

The Unified Storage Architecture has the following set of components for file and block level access to the storage data.

Components of VNX

- Gateway Components (File-level access)
- Control Stations (File-level access)
- X-Blades (File-level access)
- Storage Processors (Block-level access)
- Operating Environment for Block and File
- Backend Storage Array

Table 1 shows the various components Unified Storage inherited from its parent concepts, NAS and SAN.

NAS Array [EMC Celerra]	**SAN Array [EMC CLARiiON/Symmetrix]**	**Unified Storage Array [VNX]**	
		File Level	**Block Level**
NAS Head	---	Gateway Component [VG2/VG8]	---
---	*Storage Processor*	---	2 Storage Processors [for high availability]
Control Station	---	Control Station [one or more – for high availability]	---
Data Movers	---	2 to 8 X-blades	---
DART Operating Environment	Enginuity/FLARE Operating Environment	OE for File	OE for Block

Table 1: VNX Components
[In comparison with SAN and NAS Array products]

Gateway Components

Gateway Component VG2/VG8, which comprises X-Blades and Control Station, channelizes the data and control flow through X-Blades and Control Stations, respectively. We can configure 2-8 X-Blades for high availability and failover. VG2 supports a maximum of two X-Blades and VG8 supports a maximum of 8 X-Blades. VG2/VG8 supports 1 or 2 Control Stations.

Control Station

Control Station, which runs on EMC's specialized Red Had Linux Operating system, provides a secure, unique GUI for accessing the various VNX components. Since X-Blades cannot be directly accessed, Control Stations are used for administrative tasks such as X-Blade installation, configuration, and maintenance. Also, for file system related tasks such as creation, extension, maintenance, and so on, Control Station also manages other network components.

Dual control stations can be configured for failover purposes.

X-Blades

X-Blades are responsible for moving data from disks to the network. They run a specialized Operating Environment for file sharing. X-Blades cannot be accessed directly, and are highly secured. Control Stations are autonomous and independent and don't depend on Control Stations for standard data processing operations. Thus, a Control Station failure or other X-Blade failure doesn't impact an X-Blade's performance. As each blade has 6 processors, a fully configured system can have 8 * 6 = 48 processors.

Storage Processors

For high availability, VNX has two storage processors [SP], which take care of block-level data access. SP runs a specialized Operating Environment for Block. As each SP has 6 processing units, there are 12 processing units for block-level access processing.

VNX Unified Features

- **Data Deduplication and Compression**

  VNX performs deduplication and compression processes as a background activity. It also ensures that the CPU utilization rate is in check so that the file/data transfer activity is not hampered. It establishes criteria such as data size, usage, age, and so on in selecting the data to deduplicate. Data which is not accessed by the clients are termed as 'inactive data'. Only this data will be taken into consideration.

- **Thin Provisioning**

  Generally, space utilized is below 50% than the space claimed. With this perspective, VNX provides a 'thin provisioning' feature to enhance capacity utilization. Timely reports on available capacity, utilized space, and claimed space will help the administrator to predict capacity requirements and take necessary actions. VNX provides features such as Automatic File Extension. Dynamic LUN Extension does this dynamically.

- **Intelligent Management Software**

  Unisphere® is web-based intelligent management software that allows us to configure, monitor, manage, and administer the VNX series. It also provides utilities such as analytic and report tools to diagnose the current situation of the resources, state of the hardware and software components, and predict future requirements. It also provides a User Interface to access and manage replication, security, and application features.

Local Replication Features

The VNX series inherited local replication features from Celerra and CLARiiON/Symmetrix. It has specific replication features for both file system and block data.

EMC Celerra	**EMC CLARiiON/Symmetrix**	**EMC VNX Series**
Celerra SnapSure	---	VNX SnapSure
Celerra Replicator	---	VNX Replicator
---	Clariion SnapView™	VNX SnapView
---	RecoverPoint CDP	VNX RecoverPoint CDP

Table 2: VNX Local Replication Features
[In comparison with SAN and NAS Array products]

VNX SnapSure

VNX SnapSure creates Point-in-Time snapshots of production data with respect to time frames called *'Checkpoints'.* It is file-based replication; it does not replicate a complete copy of the data.

VNX Replicator

VNX Replicator is an IP-based replication solution; known as a 'LANless solution'. It provides read-only local copy of the production data. This is a replication solution for file systems.

VNX SnapView

VNX SnapView is similar to VNX SnapSure which is for block level data. VNX inherited this feature from EMC CLARiiON. Similar to SnapSure, SnapView creates point-in-time snapshots of production data.

VNX RecoverPoint Continous Data Protection [CDP]

RecoverPoint utilizes a special appliance called 'Write Splitter' for the replication process. While a write operation initiated from the host, the Write Splitter creates a local replication copy. This is a replication solution for block-level data.

Remote Replication Features

EMC Celerra	EMC CLARiiON/Symmetrix	EMC VNX Series
Celerra Replicator (Remote)	---	VNX Replicator (Remote)
---	Clariion MirrorView	VNX MirrorView
---	RecoverPoint CRR	VNX RecoverPoint CRR

Table 3: VNX Remote Replication Features
[In comparison with SAN and NAS Array products]

VNX MirrorView

This is a remote replication solution inherited from EMC CLARiiON. This is for block-level data which can is replicated both synchronously and asynchronously. It provides failover/failback options. When the primary site crashes, production can be failed over to the remote site.

VNX Security Features

- **File level Retention**

  In VNX, each file can be provided with a retention period for modification/deletion. The specific file cannot be modified/deleted until that retention period expires.

- **Antivirus protection**

  To protect data from threats such as malware, virus, and so forth, antivirus solutions are required. They ensure data reliability, security, and integrity. Tied to the VNX Event Enabler [VEE] to get alerts on threats, they also help manage the antivirus servers.

- **File locking**

  To ensure the integrity of a file, a file locking mechanism is followed. Thus, when a file is being modified, it cannot be accessed concurrently by another application. The VNX Operating Environment takes care of this.

- **Host-Based Encryption**

  In host-based encryption, data is encrypted at LUN level on the host side. It is not applied to the complete array, thus reducing complexity in maintenance. Host-based encryption protects data from unauthorized access.

Benefits of VNX

- Increased performance, scalability, maintainability, and efficiency
- Unified management and replication features
- Unified Operating Environment and user interface
- Enhanced security features
- Excellent packaging
- Efficient hardware and software components/configurations
- Support for storage tiering, thin provisioning, etc.
- Advanced analytic and reporting tools