



Education

# **FIND AND SELECT THE RIGHT FILE STORAGE FOR YOUR APPLICATIONS**

Philippe Nicolas, Brocade

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - ◆ Any slide or slides used must be reproduced without modification
  - ◆ The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the Author nor the Presenter is an attorney and nothing in this presentation is intended to be nor should be construed as legal advice or opinion. If you need legal advice or legal opinion please contact an attorney.
- The information presented herein represents the Author's personal opinion and current understanding of the issues involved. The Author, the Presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.  
**NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.**

- **Title: Find and Select the Right File Storage for your Applications**
- **Abstract:** Many businesses are linked to file storage technologies as many of these new, recent and even existing applications morph now, rely and support file based data. At the same time, the volume of data explodes especially the file data type which now represents by far the larger portion of enterprise data. With the complexity and variety of market solutions, the challenge for IT buyers and storage managers is to choose and adopt the most adapted solutions aligned to their business and IT needs to address their current and future challenges with a special attention to compliance and data retention regulations. This session covers the most common deployed applications, their attributes in term of file storage needs and maps these to file storage solutions available in the industry with technologies details and advantages. Various technologies are presented in this session, among them: Clustered, SAN-based, Distributed and Parallel File System/Storage.
- **Learning objectives:** With a top-down approach, this tutorial improves file storage technologies positioning and understanding aligned to applications needs and challenges. With that survey, technologies segmentation and features matrix, it helps end-users, IT and file storage buyers to select, choose and adopt the right solution.
- **Audience:** IT & Storage Architect, IT Manager and Buyers.

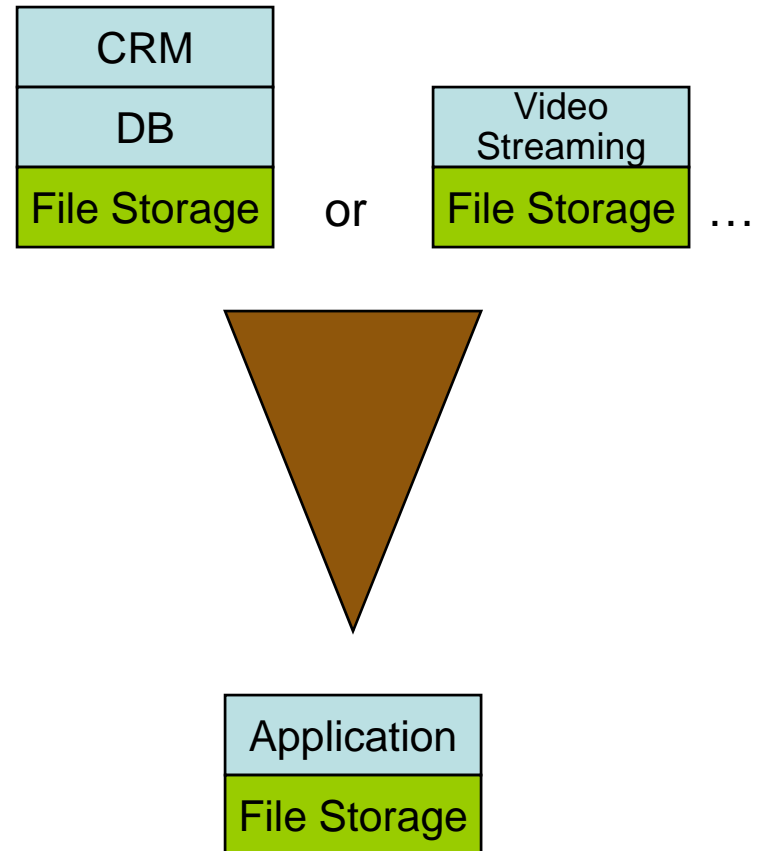
- Applications
  - ◆ Convention, Needs & Challenges
  - ◆ Type, Characteristics, I/O Patterns & Access
- File Storage
  - ◆ Clustered, SAN-based, Distributed & Parallel
  - ◆ Basic & Advanced File Services
- The Right File Storage for Each Application
- Conclusion

# Applications

**Convention, Needs & Challenges  
Type, Characteristics, I/O Patterns & Access**

## ➤ Convention

- ◆ No distinction between software services above File System/Storage logic & layer
- ◆ Everything is an Application: Database, Web Server, Computing (HPC), Archival & ILM, Video streaming... even multi-tiers or multi-layers
- ◆ **Remark - Exclusion:** Application can run on many various *local disk file systems* in a single server-storage domain but this “classical” approach is not covered in this tutorial



# End-User Needs and Examples of use

## End User Needs

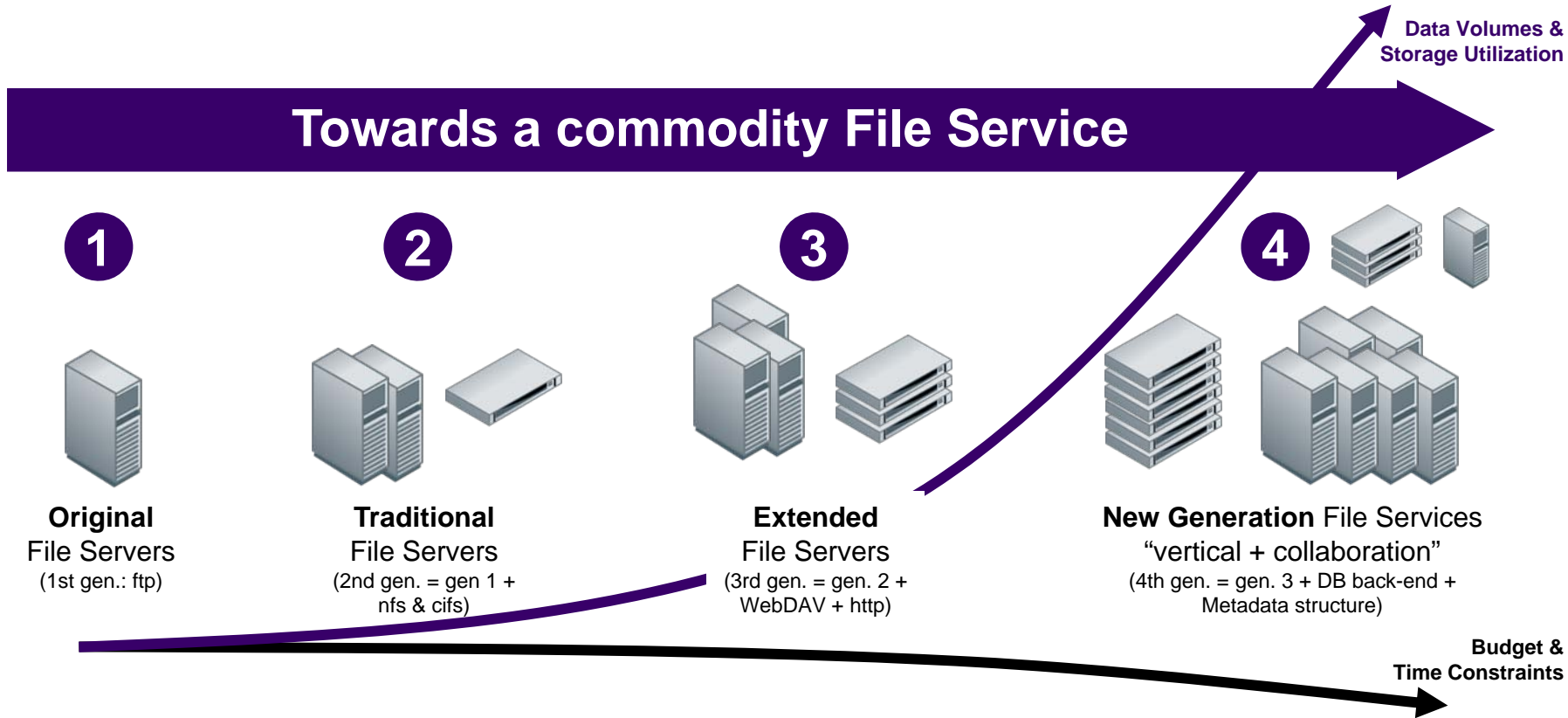
- Better scalability
  - ◆ Capacity: fast growth of volume of data, # files, filesize...
  - ◆ Performance: IOPS, BW, frame/sec...
  - ◆ Data Sharing: avoid data duplication, aggregate more servers
  - ◆ Larger server can be expensive
- More Availability, no Downtime
  - ◆ Local (clustering, failover...)
  - ◆ Remote (wide failover + data replication...)
  - ◆ Global (multi-sites)
- Easy Manageability and Administration
  - ◆ Migration and Consolidation (data movement)
  - ◆ Remote site and file server management
- Industry Standard
  - ◆ Protocols, components, COTS...
- Advanced features
  - ◆ Load Balancing, Quotas, Security, Data Protection (snapshot, replication...), ILM & Classification, Data Reduction, Encryption, Content Indexing & Search, Reporting & Statistics, XAM...
- Cost Reduction
  - ◆ \$/TB, \$/IOPS, \$/BW, \$/Transac fs op., \$/NFSops...

## Examples of use

- High Availability Clusters (local & geographic)
- Scaling applications
  - ◆ Web Servers - Read mostly/load balanced
  - ◆ Databases/OLTP/DW - Mostly use direct I/O
- Distributed and Parallel app. and fast failover
  - ◆ Data acquisition and « demanding » computing
- Systems, App. and Data Consolidation/Migration, ILM
  - ◆ Tech. Refresh
- Off-host processing
  - ◆ Based on shared file system
  - ◆ Can also use by Point-in-Time copy techniques (not related to our data sharing definition)

# Evolution of File Services

## Usage of File Servers in IT today



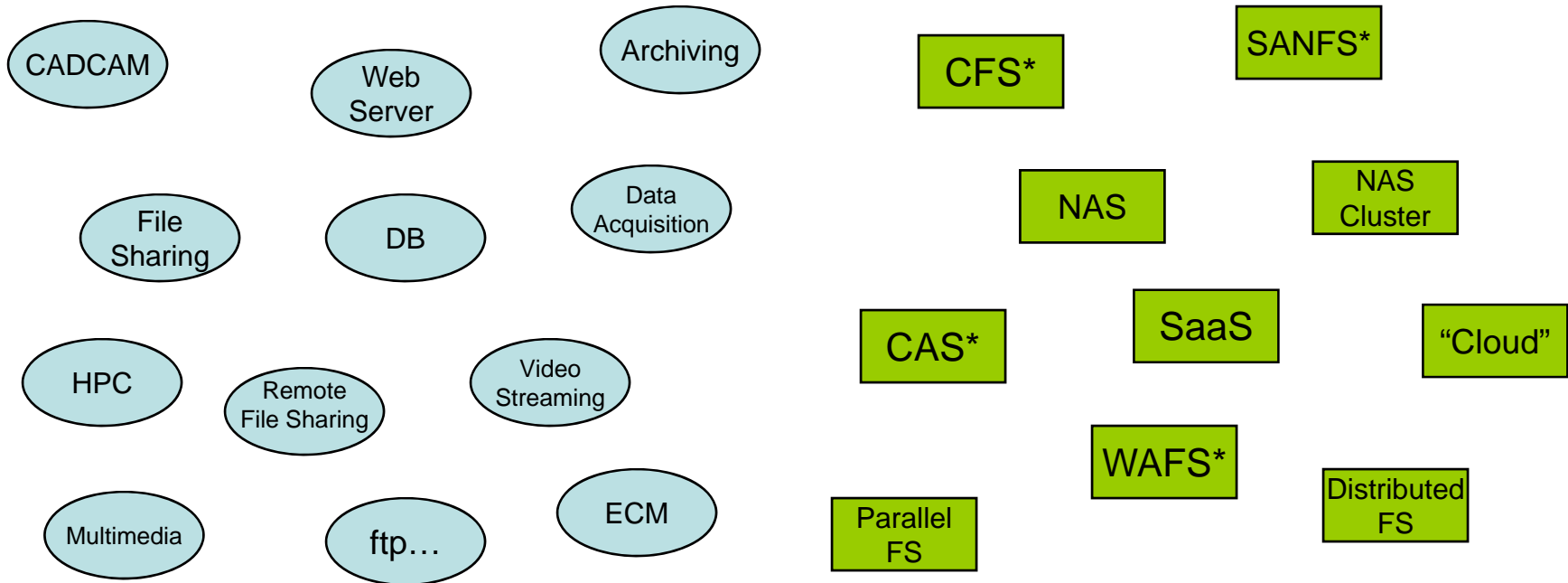
### Average disk allocation

- ▶ for Windows systems: 25-40%
- ▶ for Unix/Linux systems: 30-45%
- ▶ for iSeries and z/OS : 60-80%

### File Server/NAS considered as NAS Target

- ▶ for Backup (with DeDup) & Archiving (WORM|CAS)
- ▶ for Virtual Servers/Machines images
- ▶ for Vertical, Collaboration, HPC, Video, ...

# Applications & File Storage



➤ Based on Applications characteristics and I/O behaviors, the goal is to select the right File Storage technologies among many, many existing approaches

\* CFS: Clustered File System | SANFS: SAN File System | WAFS: Wide-Area File Service (WAN Optimization & Acceleration) | CAS: Content Addressable Storage

# Applications characteristics

<b>Workload profile</b>	<b>OLTP</b>	<b>Small Data Mart</b>	<b>Home Directory</b>	<b>Large Scale Streaming (web farm)</b>	<b>High-Frequency Meta-Data update (small file create/delete)</b>
<b>Latency sensitive</b>	High	Med	Low	Low	High
<b>Throughput</b>	High R/W	High read	Low	High read	High write
<b>Concurrent sharing</b>	High	High	Low	High read	Low
<b>Caching (re-read rate)</b>	High	High	High	Low	Low

# File Storage usage

	Processor Farms	Office Env.	Archive Data	Streaming Media	Database	Web and Content Management
Typical applications	Financial simulation, Grid computing, Asic Simulation, Oil & Gas Applications, Rich Digital Content creation – Rendering	Spreadsheet, Word processing, Presentation, Pictures editing...	Medical records & imaging, insurance policy stores, gov. records, check storage applications, video surveillance	Online content delivery Direct Pay per View News distribution online radio audio streaming social networking	Exchange, SQLServer, Oracle, other OLTP...	Web farms IIS, Apache, CAD, SW dev.
Usage of File Storage	Used as a Virtual memory	Used as a shared storage pool	Used as long retention storage for online archive of records, documents and images	Used as a large and performance storage pool	Used as alternative to “traditional” raw and local disk file system	Used as a generic repository
I/O patterns & access	Equal mix of reads and writes Sequential and large transfer sizes	Random reads and small request sizes	Large sequential access	Mixed small and large sequential transfer	Mix read/write small size	Small random reads

# File Storage

Clustered, SAN-based, Distributed & Parallel  
Basic & Advanced File Services



Check out SNIA Tutorial:  
**The File Systems Evolution**



Check out SNIA Tutorial:  
**DFS over CIFS**



Check out SNIA Tutorial:  
**Scaling NFS through pNFS**



Check out SNIA Tutorial:  
**NAS and iSCSI Technology Overview**

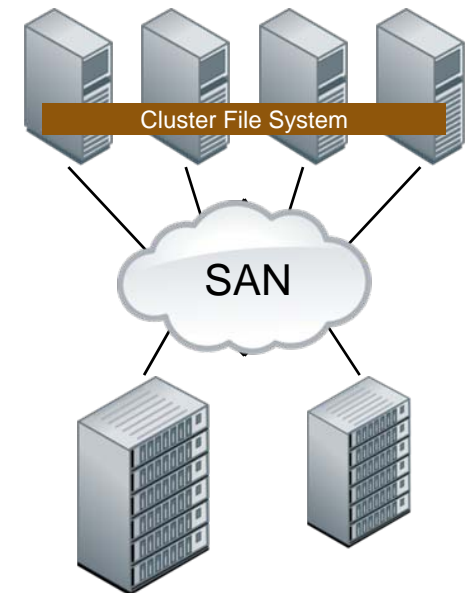


Check out SNIA Tutorial:  
**Storage Tiering for File & NAS Systems**

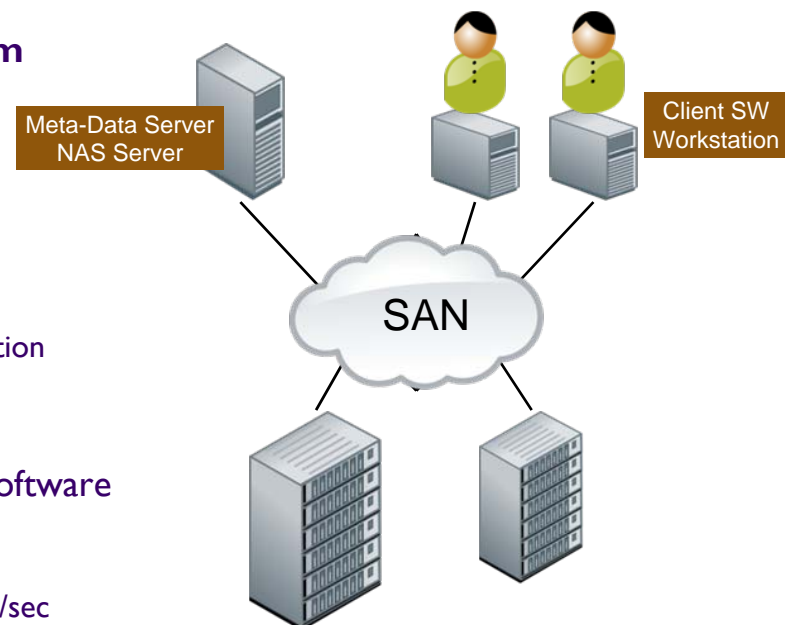


Check out SNIA Tutorial:  
**SMB2 – Big improvements  
in the Remote Filesystems Protocol**

- **Cluster File System (CFS)**, also named **Shared Data Cluster**
- A Cluster FS allows a FS and files to be shared
- Centralized (asymmetric) & Distributed (symmetric) implementation
  - ◆ Centralized uses master node for meta-data updates, logging, locking...
- All nodes understand Physical (on-disk) FS structure
  - ◆ The FS is mounted by all the nodes
  - ◆ Single FS Image (Cache Coherence)
    - › Same data view from all nodes
- Lock Mechanism
  - ◆ Distributed or Global Lock Management (DLM/GLM)
  - ◆ Granularity varies: file, record, byte...
- Usage Consideration: Concurrent vs Serial data access
  - ◆ Concurrent: multiple systems access the data simultaneously
  - ◆ Serial: one system at a time uses and access the data



- **SAN File System (SAN FS) aka SAN File Sharing System**
- A SAN FS allows files to be shared
- Client/Server or Master/Slave (aka asymmetric) model
  - ◆ I Server or a set of servers to control client access and resolves conflicts
    - › Understand, manage and use metadata on disk
    - › Use of file system even if portions of it are inaccessible
    - › Distribute block addresses to clients on request
  - ◆ Thin client software layer handles SAN device and server interaction
    - › Connection to SAN storage
    - › Avoid overhead due to Metadata management
    - › Access to data directly using blocks addresses sent by Master(s)
- Mixed role between direct data access with host based thin software and NAS access
- Flexibility of network FS at SAN speed
  - ◆ « Scaling hundreds of PetaBytes of capacity and tens of GigaBytes/sec
- Designed to support hundreds or thousands of nodes
- Lock Mechanism
  - ◆ Provided by the server at a central location
  - ◆ Various granularity: file, record, byte...
  - ◆ Some implementations use SMB/CIFS or NFS semantics
  - ◆ The server needs to be protected for SPOF reason
- Cache Coherency
  - ◆ Some implementations deliver cache coherency with traditional validate/invalidate mechanism, others don't offer cache at all



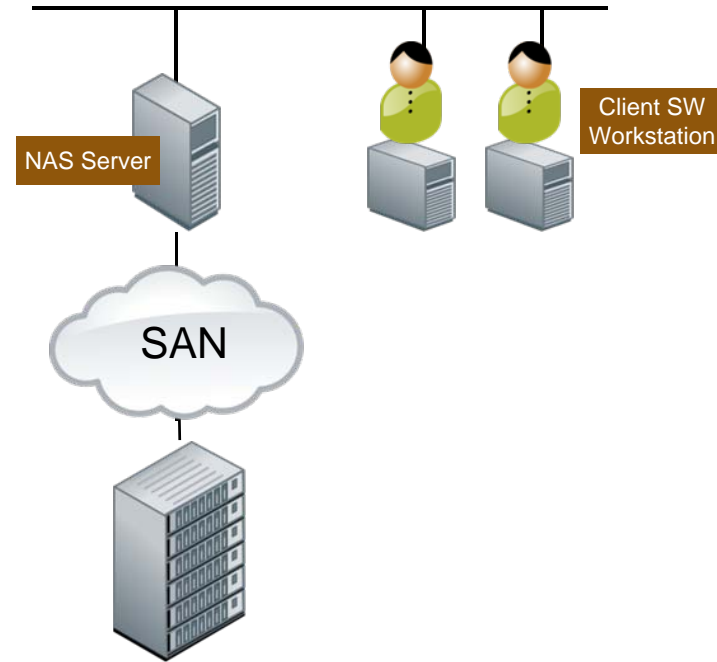
# CFS vs. SAN FS

<b>Characteristics &amp; Features</b>	<b>Cluster FS</b>	<b>SAN FS</b>
<b># of nodes</b>	Dozens	Hundreds
<b>Heterogeneous OS</b>	No	Yes
<b>Tolerance of Distance (between server and clients)</b>	Limited	Important
<b>Dedicated Meta-Data Server(s)* required</b>	No (except centralized design or configuration)	Yes, usually
<b>Physical File System layout knowledge</b>	All nodes (Cluster FS currently requires same OS)	Meta-data server only (clients may understand if same OS)
<b>Single File System Image (SFSI)/Cache coherency</b>	Yes	No

\* Meta-data Server aka Master Server

# Network File Server aka NAS

- **Distributed File System – General Characteristics**
  - ◆ Network transparency, User Mobility, Fault Tolerance, Scalability, File Mobility
  - ◆ No NAS aggregation by default
- **NFS** primarily for Unix and **CIFS** for Windows (NAS protocols)
  - ◆ Asymmetric (Client/Server) architecture
  - ◆ Uses TCP/IP (UDP for NFS in the past, NFS over RDMA)
  - ◆ De facto standards today
  - ◆ Too “chatty”/verbose for remotes file access



	NFS (v2, v3)	CIFS v1 (historically SMB)
<b>State mode</b>	<b>Stateless</b>	<b>Stateful</b>
<b>Locking</b>	<b>Advisory</b> , locking only affects applications that use locking APIs	<b>Mandatory</b> , locking affects all file access
<b>Cache Coherency (Client)</b>	<b>Weak</b> attribute expiration times plus cached data revalidation on file open	<b>Strong</b> locking and invalidation for both attributes and file data
<b>Application Impacts of coherency</b>	Stronger coherence requires application changes (e.g., close and reopen file to get current data)	Usually transparent, modifications can help with conflicts (e.g., open copy of file that's in use)
<b>Data stability</b>	Client: write-back <b>Server: write-through</b> , written data is always stable	Client: write-back <b>Server: write-back permitted</b> , written data may not be stable

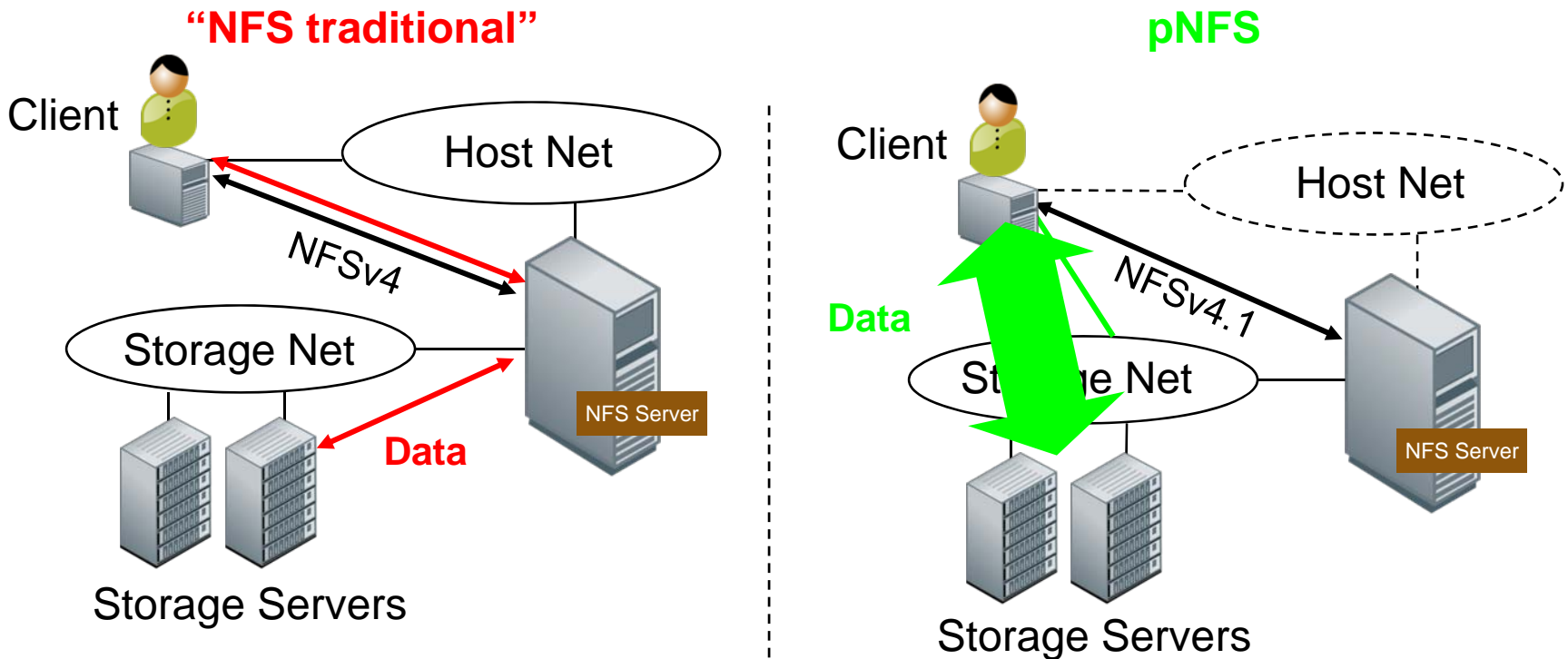
# Recent NAS Protocols Development

- CIFS and NFS based
  - ◆ SMB2/CIFS2 “WAN optimized” (v 2.002)
    - › >30 times faster compared to SMB1 over WAN and 2-10x on LAN
    - › Compound mechanism (aggregation of multiple requests – only 19 commands, reduce round trips)
    - › “durable file handles”, larger buffer sizes,, sym. links, secure and robust...
  - ◆ NFS v4, v4.1 (pNFS)
    - › Only 1 port (2049), stateful, compound operations, client caching + delegation, security (authentication + Windows ACLs), migration + replication, Unix/Linux & Window support, RDMA, TCP (only), Namespace extensions (Mirror mounts & referral) + pNFS (v4.1)...

	NFSv3	NFSv4
Personality	Stateless	Stateful
Semantics	UNIX only	Support UNIX and Windows
Authentication	Weak (AUTH_SYS)	Strong (Kerberos)
Identification	32 bit UID/GID	String based (xyz@__.com)
Permissions	UNIX based	Windows like access
Transport	UDP & TCP	TCP
Caching	Ad-hoc	Delegations

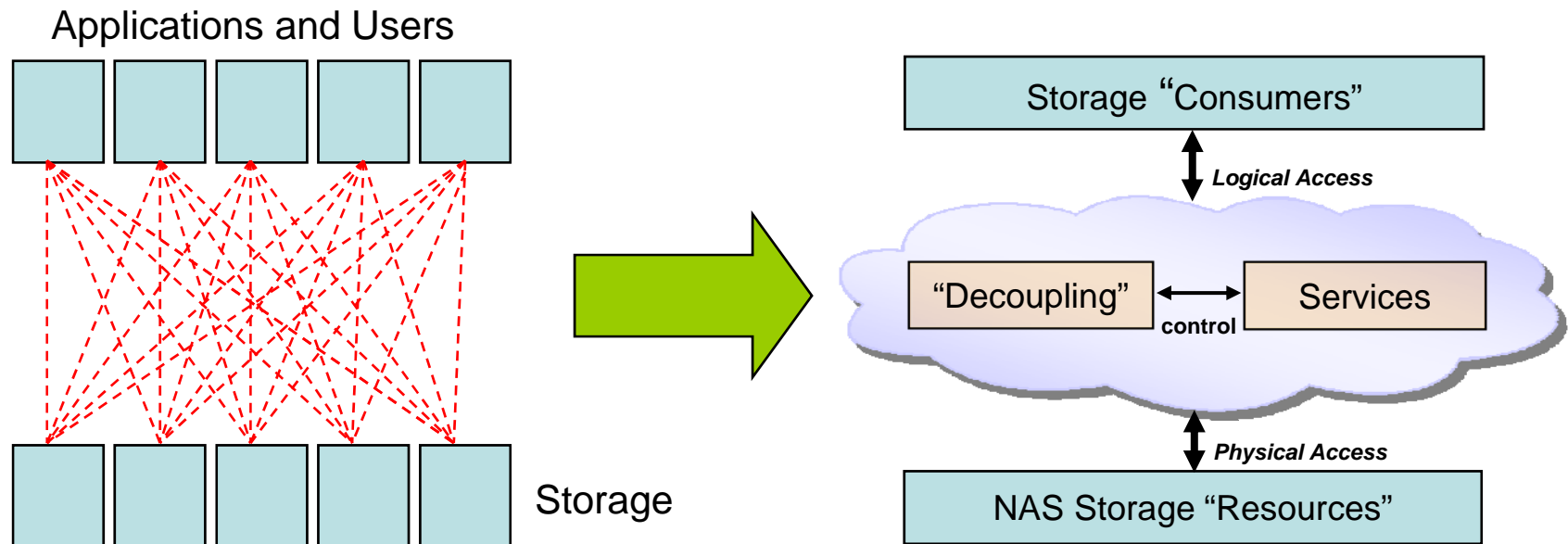
# NFS v4.1 with parallel NFS (pNFS)

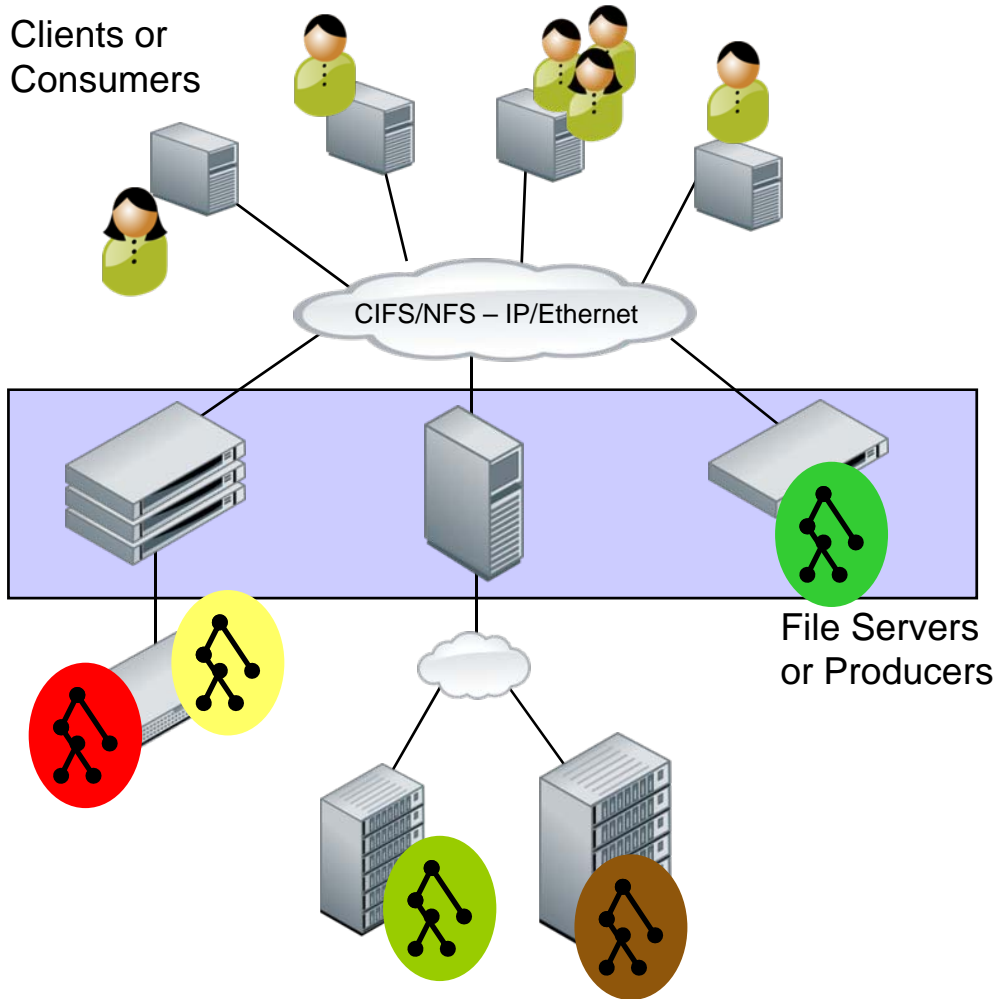
- pNFS is about scaling NFS and address file server bottleneck
  - ◆ Same philosophy as SAN FS (master/slave) and data access in parallel
- Allow NFSv4.1 client to bypass NFS server
  - ◆ No application changes, similar management model
- pNFS extensions to NFSv4 communicate data location to clients
  - ◆ Clients access data via Fibre Channel & iSCSI (block), OSD (object) or NFS (file)
- IETF standardization in progress



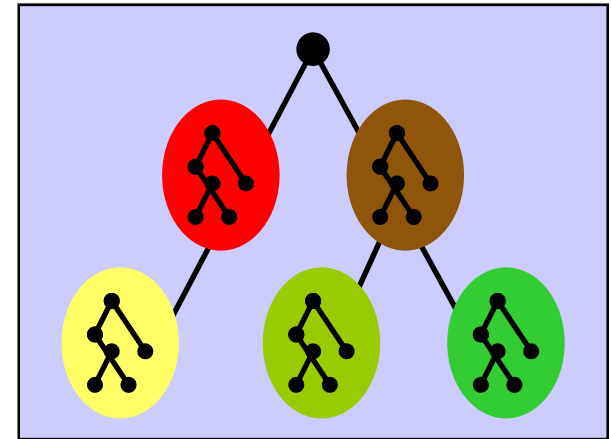
## ➤ Definition (SNIA Dictionary)

- ◆ “A **namespace-based network-oriented infrastructure** for files that includes a **decoupling layer** which **separates logical file access from physical file location**. This decoupling layer enables a **variety of services** (e.g., replication and migration) to be applied to files and filesystems”





## Namespace Aggregation



- ◆ **Shared Namespace**
  - ◆ Proprietary approach
  - ◆ “internal” aggregation of same brand/model file/storage servers
- ◆ **Global Namespace**
  - ◆ Open approach
  - ◆ “external” aggregation of individual file servers + shared namespace if any

# Decoupling Approaches

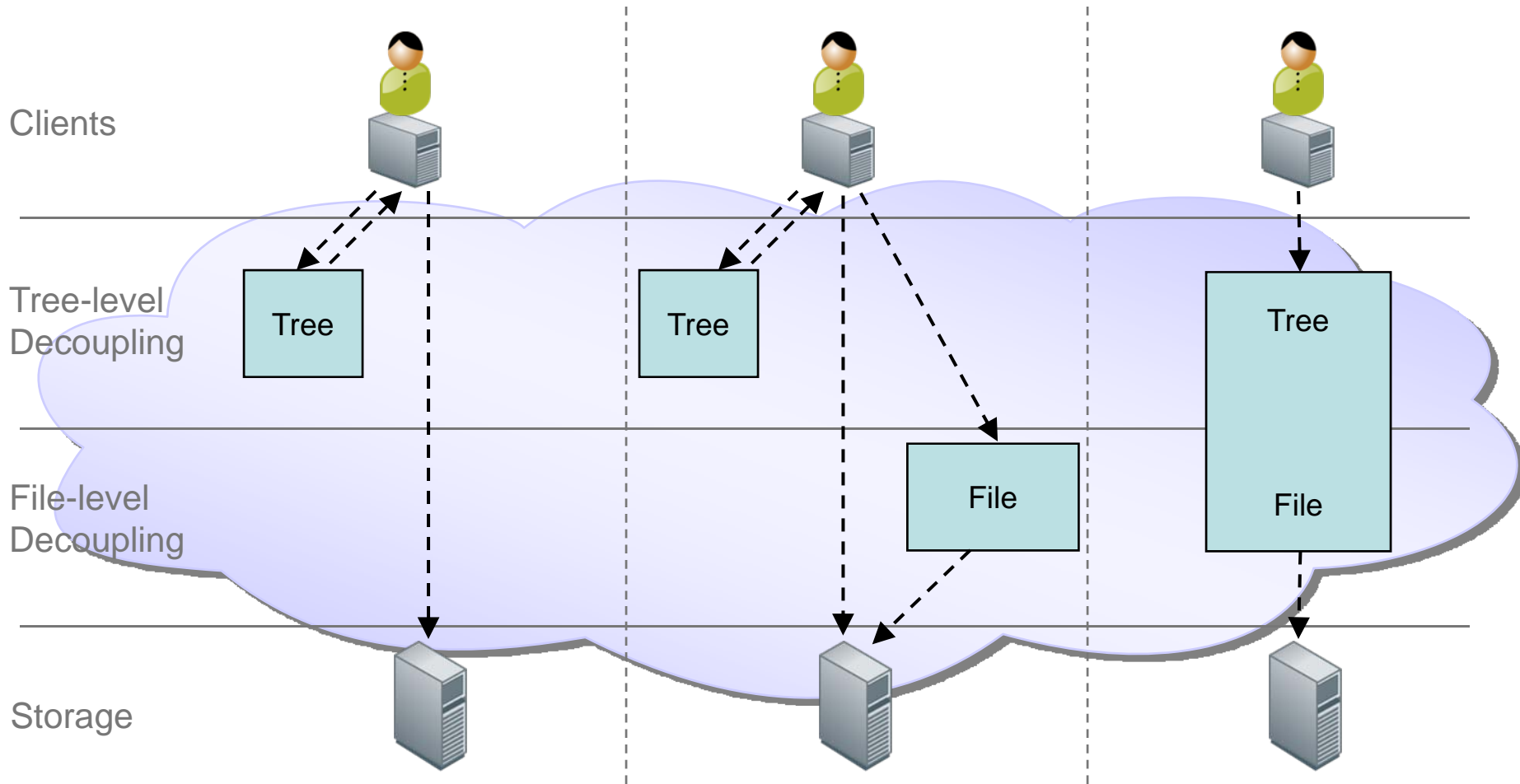
- Client-based (out-of-band)
  - ◆ OS service or agent loaded on client system
  - ◆ Tree-level granularity with asynchronous updates
  - ◆ NAS-protocol specific
- Hybrid (dual-band)
  - ◆ Combines client-based and network-based
- Network-based (in-band)
  - ◆ Continuous network-resident decoupling
  - ◆ File-level granularity and synchronous updates
  - ◆ Support for all NAS protocols
  
- Product categories are Network File Virtualization (NFV) and Network File Management (NFM)

# Decoupling Approaches

## Client-based

## Hybrid

## Network-based



## ➤ File Virtualization

- ◆ Capacity to mask physical location across (file) servers and provide logical access among them
- ◆ Seen as one logical entity
- ◆ No network, NAS protocols or client (consumer) related
- ◆ Example : Cluster File System...

## ➤ **Network** File Virtualization (NFV)

- ◆ Integration of network, NAS protocols and clients (consumers) on top of (file) servers (notion of FAN)
- ◆ Example : DFS, Automount, NFS v4 namespace extensions...

## ➤ File Management

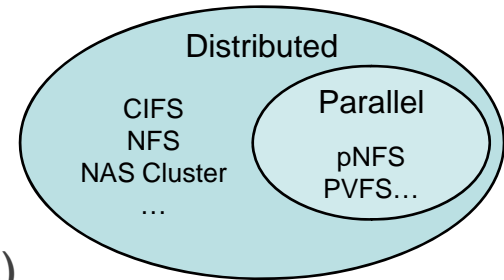
- ◆ = File Virtualization + File Services
  - › File Services: Replication, Failover, Load Balancing, Quotas, Security, Data Protection (snapshot, replication...), ILM & Classification, Data Reduction, Encryption, Content Indexing & Search, Reporting & Statistics, XAM...

## ➤ **Network** File Management (NFM)

- ◆ = NFV + File Services

## ➤ Concept

- ◆ “Aggregation” of storage servers to boost Performance, Capacity and Availability
  - > NAS protocols may be available from nodes (servers)



## ➤ Features

- ◆ Notion of Namespace
  - > Cluster-wide consistent namespace (shared namespace)
- ◆ Notion of File Virtualization and Network File Virtualization
- ◆ Asymmetric (Meta-Data/Master/Mgr server) or pure Symmetric
- ◆ Parallel as a sub-segment of Distributed
  - > Parallel file access for client
- ◆ Also for some industry offering
  - > File/Data striping across I/O nodes
  - > Data redundancy between and « behind » servers (RAIN + RAID)



# Basic & Advanced File Services

## Basic

- Global Namespace (File Virtualization)
  - ◆ Organize storage in an overlay namespace
- Migration
  - ◆ Move files from one server to another
- Tiering / ILM
  - ◆ Move files via policy to the “best” storage
- Load Balancing
  - ◆ Move files to better distribute capacity or load
- Data Protection
  - ◆ Snapshot to support online data protection
  - ◆ Replication as a BC strategy
- Reporting & Statistics

## Advanced

- HA/BC/DR\*
  - ◆ Associated with data replication
- Data Classification and Optimized Placement
  - ◆ Data value and storage characteristics alignment
- Data Storage Optimization
  - ◆ DeDuplication, Reduction and Compression
- Quota Management
  - ◆ Report and enforcement
- Content Indexing & Search
  - ◆ File content to optimize file placement
  - ◆ Search for legal discovery...
- Application Acceleration
  - ◆ Local and Distributed file access
  - ◆ Notion of Distributed ILM/Tiering
- Security
  - ◆ Access Control, auditing and Encryption...

\* High-Availability/Business Continuity/Disaster Recovery

# The Right File Storage for Each Application

# Some approaches

## ➤ Top-Down

- ◆ Understand, monitor and profile application I/O pattern
- ◆ Configure volumes, LUN..., stripe size and file system (block size...)
- ◆ Tune, verify & control, adapt...

## ➤ Bottom-Up

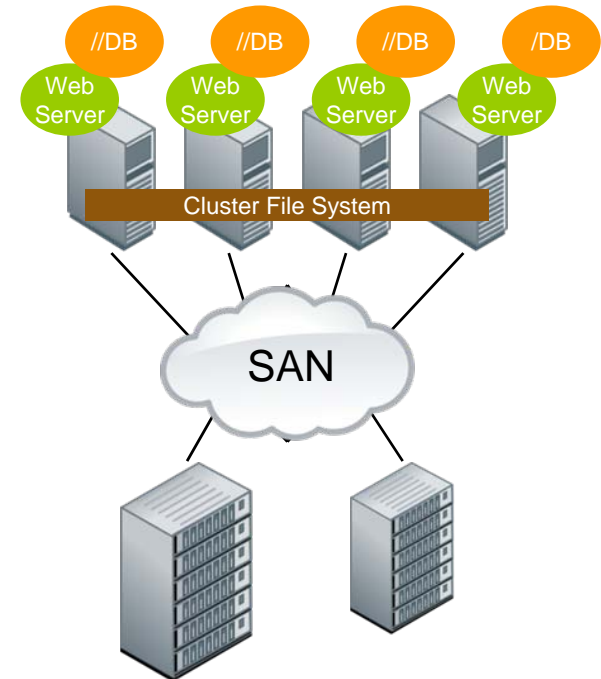
- ◆ Configure storage, LUN and volumes, stripes + file system
- ◆ Build and align application I/O size to the above one (need source)
- ◆ Tune, verify & control, adapt...

## ➤ Other

- ◆ Too many various I/O access and behaviors
- ◆ Storage and Application modification not possible

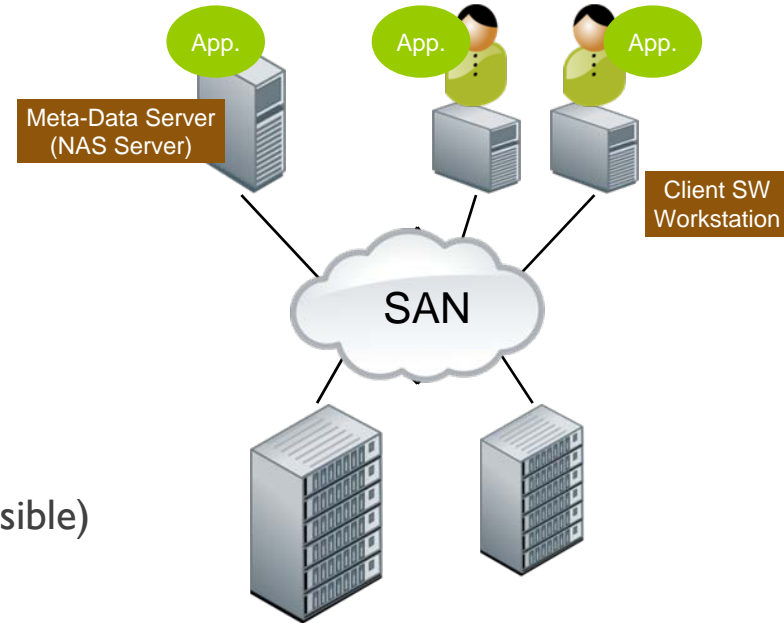
# Web Server farm – Parallel DB

- **Application**
  - ◆ Web Server farm with http, ftp... services
    - › Potential load balancer in front
  - ◆ Parallel database
- **Configuration**
  - ◆ Cluster File System
    - › From 2 to 16/32 nodes
- **Benefits**
  - ◆ COTS approach
  - ◆ Increased throughput (https requests, IOPS & db transactions)
  - ◆ Optimized failover, failure is transparent & high SLAs
  - ◆ More effective use of servers
  - ◆ Potentially 1 RW node and n RO nodes
  - ◆ Possible partition and online scalability (transparent addition of nodes)
  - ◆ SSI/SFSI



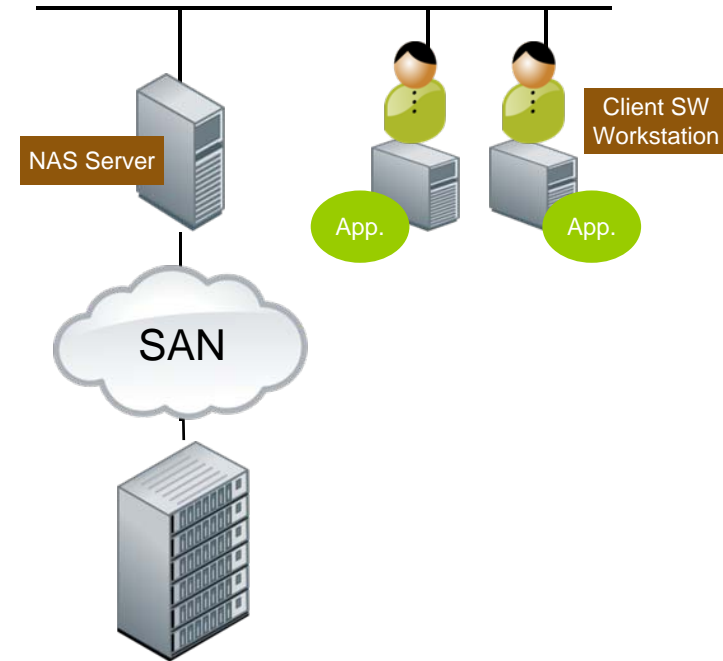
# Multimedia application

- **Application**
  - ◆ Video Streaming, Clip Editing, movie rendering...
- **Configuration**
  - ◆ **SAN File System**
    - › 1 big server (MDS) with NFS/CIFS layer
      - Multiple MDS possible for HA
    - › Server and Client SAN FS layer
    - › Hundreds of clients (heterogeneous OS possible)
- **Benefits**
  - ◆ Flexibility of NAS, Speed of SAN
  - ◆ Consolidate storage, very scalable
  - ◆ More effective use of resources
  - ◆ Concurrent or Serial file access between consumers



# Data Acquisition, File Sharing

- **Application**
  - ◆ Data Acquisition, File Serving/Sharing, File Storage Consolidation, Data/File repository, Office application... even remote file access
- **Configuration**
  - ◆ Distributed File System
    - › I/N file servers/NAS with NFS/CIFS logic
    - › Hundreds of clients (heterogeneous OS possible)
  - ◆ Aggregation of file servers/NAS (FAN)
    - › Network File Virtualization + Network File Management
      - File oriented ILM, dataflow, Archiving
  - ◆ Last NAS protocols (SMB2) for remote access
- **Benefits**
  - ◆ Very easy Deployment and Management
    - › Very common solution
  - ◆ Consolidate storage, very scalable
  - ◆ More effective use of resources
  - ◆ Concurrent or Serial file access between consumers



# Data intensive & HPC application

## ➤ Application

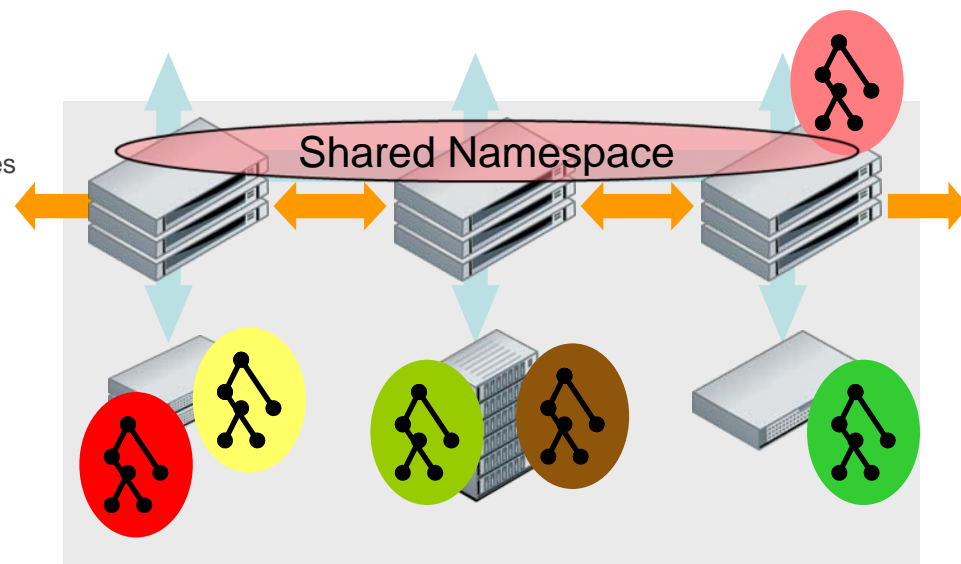
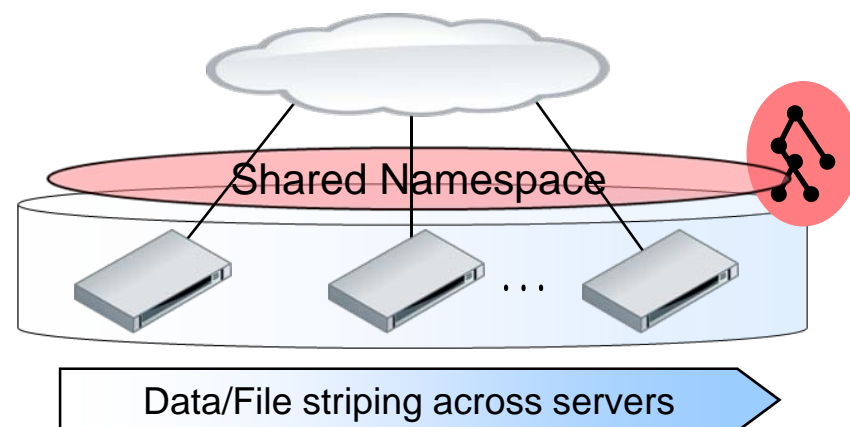
- ◆ Data intensive application, High Performance Computing...

## ➤ Configuration

- ◆ Distributed & Parallel File System
  - › N file server/NAS based on COTS hardware with NFS/CIFS logic
    - Some RAIN + RAID configurations
    - Parallel file access (pNFS...)
    - Potential stripe at the file level
  - › Hundreds of computing clients (heterogeneous OS possible)
    - Servers can be themselves compute nodes

## ➤ Benefits

- ◆ Consolidate storage, very scalable
- ◆ Increased throughput
- ◆ More effective use of servers



# Applications/File Storage Matrix

	HPC	DB OLTP	Web farm	Office Environment	Multimedia Video	Data Acquisition	Archiving	ILM	Other Vertical
<b>CFS</b>		x	x		x	x		x	x
<b>SANFS</b>	x				x	x			x
<b>NAS (+ FAN)</b>	x	x	x	x	x	x	x	x	x
<b>NAS Cluster</b>	x				x	x			x
<b>Distributed &amp; Parallel</b>	x				x	x			x
<b>WAFS (WAN Opt. &amp; Acc.)</b>				x					x
<b>CAS/Worm (with NAS access)</b>							x	x	

# Conclusion

- File System and File Storage technologies are key for current and next IT and Information, Data and Storage Challenges
  - ◆ Allows growth Control and Online Scalability
  - ◆ Delivers high level Performance, but HA/BC/DR is a must
  - ◆ Uses more and more Commodity hardware
  - ◆ Think standard (de facto and industry) as many of them evolve
  - ◆ Enables easy Deployment and Management (NFV & NFM)
  - ◆ Reduces Complexity
  
- Many approaches and philosophies in the industry
  - ◆ There is no single solution that is superior in all cases BUT these approaches deliver real applications and business benefits for different applications needs
  - ◆ Study and choose the one which delivers the best value for you

- Please send any questions or comments on this presentation to SNIA
  - ◆ [trackfilemgmt@snia.org](mailto:trackfilemgmt@snia.org) (File Systems & File Management)

**Many thanks to the following individuals  
for their contributions to this tutorial.**

**- SNIA Education Committee**

**Philippe Nicolas**