Scalable NAS for Oracle: Gateway to the (NFS) future

Dr. Draško Tomić
ESS technical consultant, HP EEM

Agenda

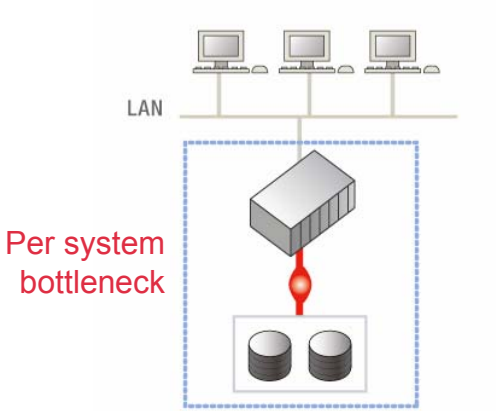
- Enterprise NAS solutions
- Oracle scalable NAS value
- HP’s scalable NAS for Oracle
- Performance metrics
- Q&A

Enterprise NAS solutions

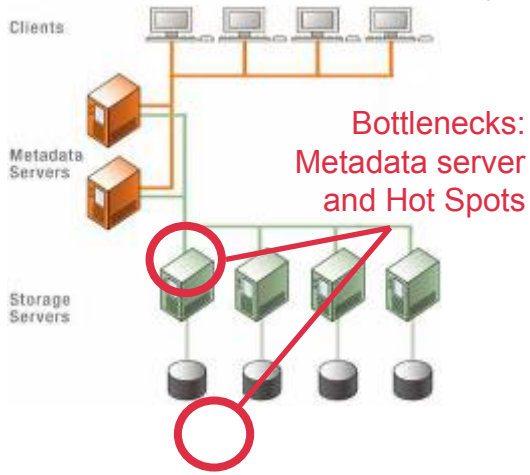
Why Cluster Design is so importan

Design Goal	Single-node NAS Designs	DFS and/or Metadata Server Designs	PolyServe Symmetric Cluster File System
Scale-out	Scales to two (2) nodes.	Scales to many nodes.	Scales to 16 nodes.
Performance	Performance degrades as scaled; limited.	Performance degrades as scaled broadly.	Performance increases linearly as scaled out.
Availability	Traditional clustering prone to error, downtime.	Requires redundant data, storage, and servers	Built into shared data and shared storage design.

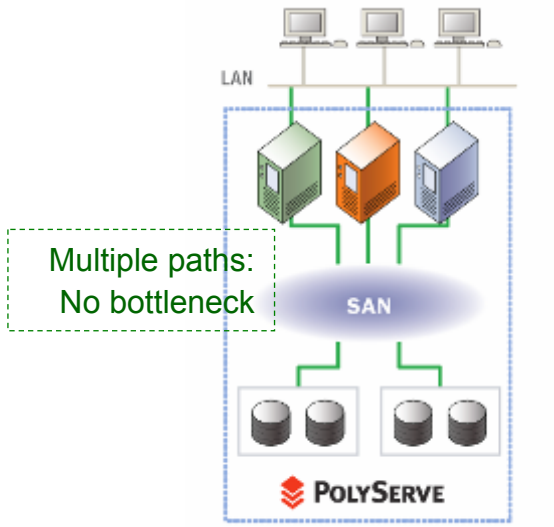
Traditional NAS / Filer design



Distributed File System and Metadata Server Designs

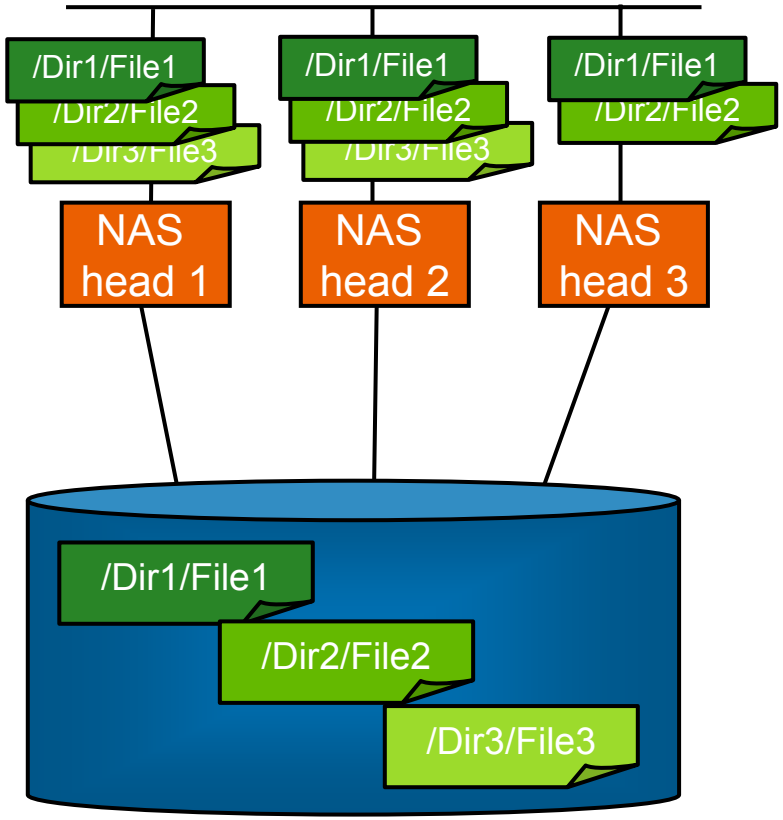


Symmetric Cluster File System



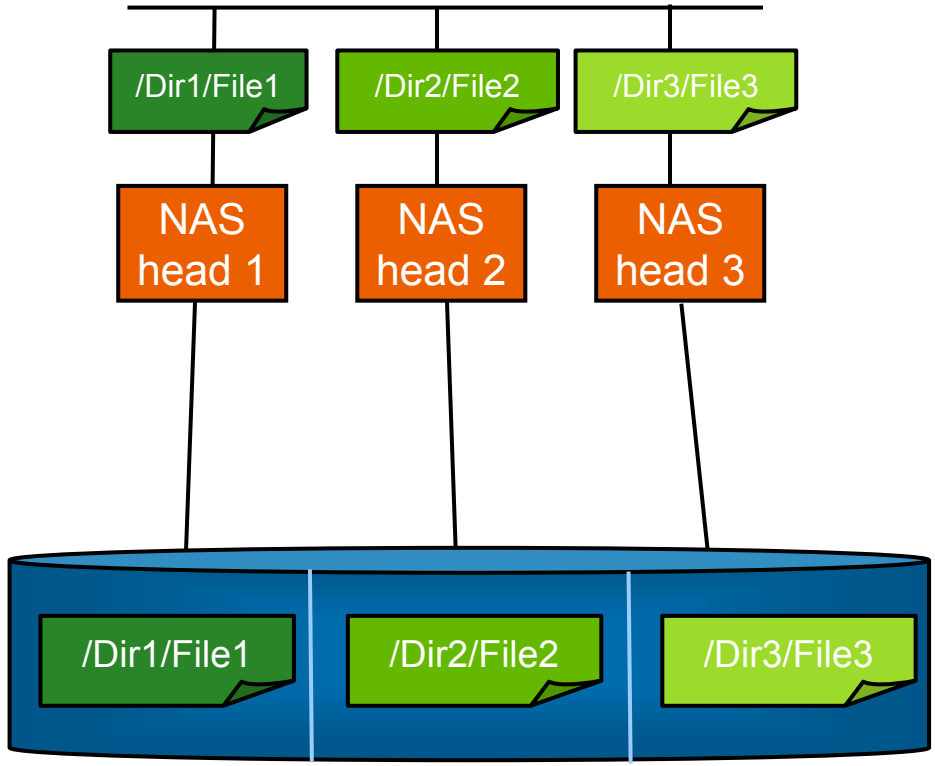
Symmetric vs. asymmetric

Symmetric—all servers see all data all the time



- All Nodes see the same storage

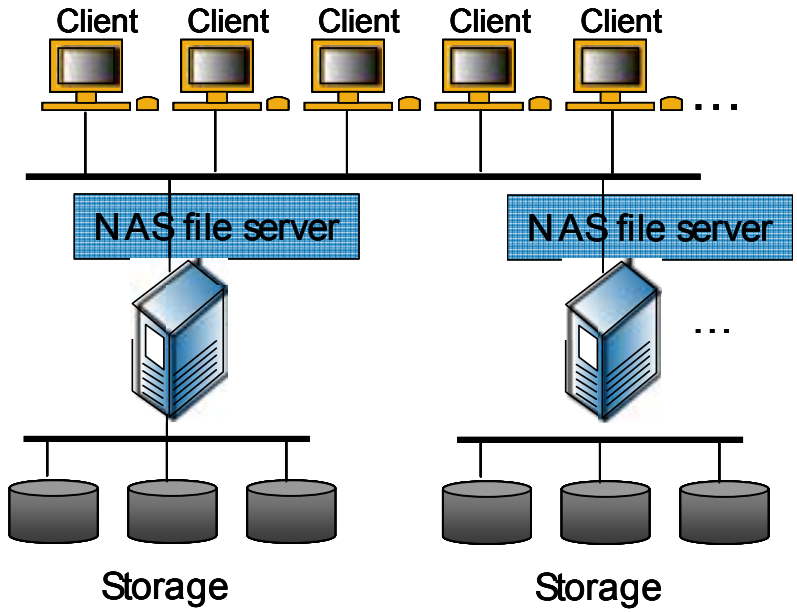
Asymmetric—a single file system can be seen from a single NAS head



- Each node has its own storage
- On failover, storage is moved to another node

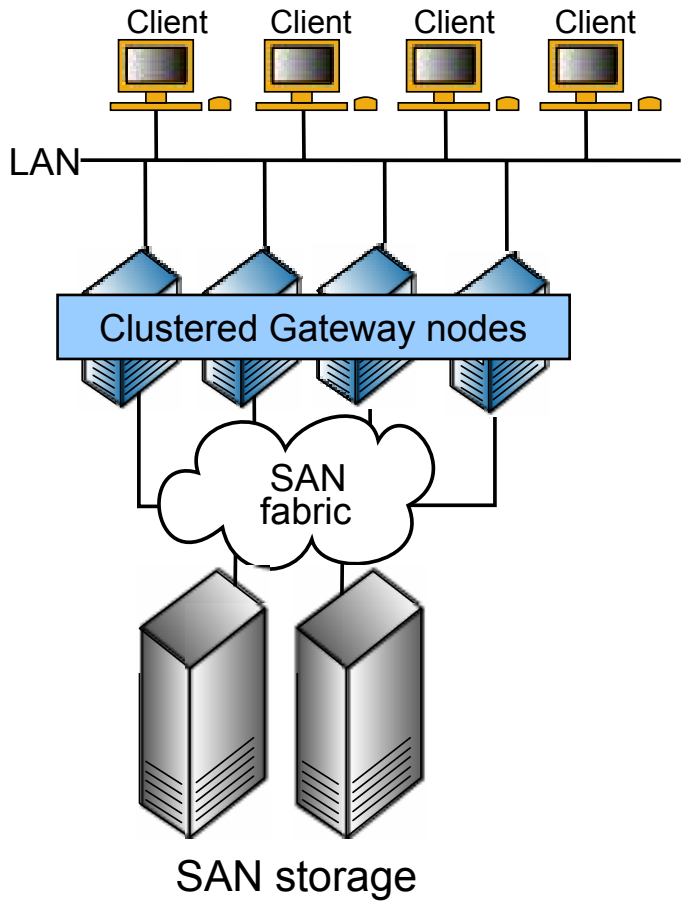
Assymetric NAS architecture

- A file system lives on one and only one filer—the file and print server owns that data
- File server is a performance bottleneck for all network traffic and I/O to that file system and Single Point of Failure (SPoF)
- Multiple file servers can create uneven work load patterns and utilization is not even across filers
- Each file server is an operational burden
  - Backed up separately
  - Updated with new patches
  - Protected against virus
  - Separated free space pool



Symmetric NAS architecture

- Eliminates file serving performance bottlenecks
- Enables mission critical availability
- Drives high storage utilization rates
- Operationally efficient



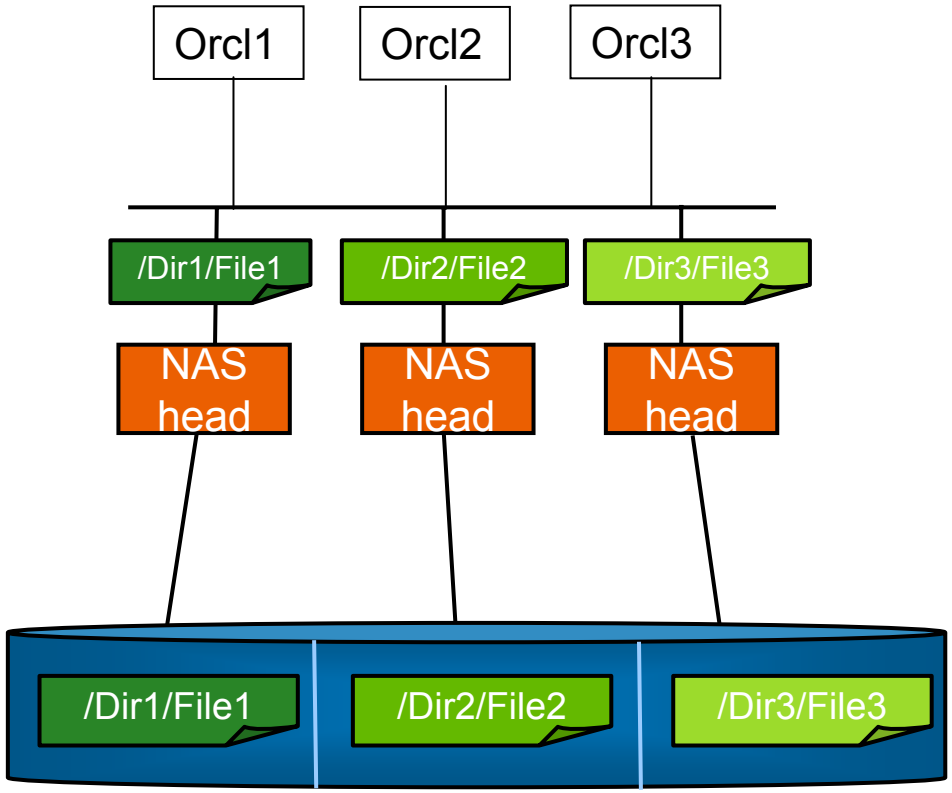
Oracle scalable NAS value

- Oracle 11g introduces Direct NFS
  - Oracle-optimized & internal NFS client
    - Lower system overhead
    - Great scalability up to 4 Ethernet paths to each NAS head
      - no NIC bonding required
    - Inherent high availability
      - Built-in multi-path I/O
    - No need for managed switches—further cost savings
    - Oracle Direct NFS supports running Oracle on Windows servers accessing databases stored in NAS devices
- Oracle prefers NAS…

Oracle non-RAC

Single instance databases on NFS Clients

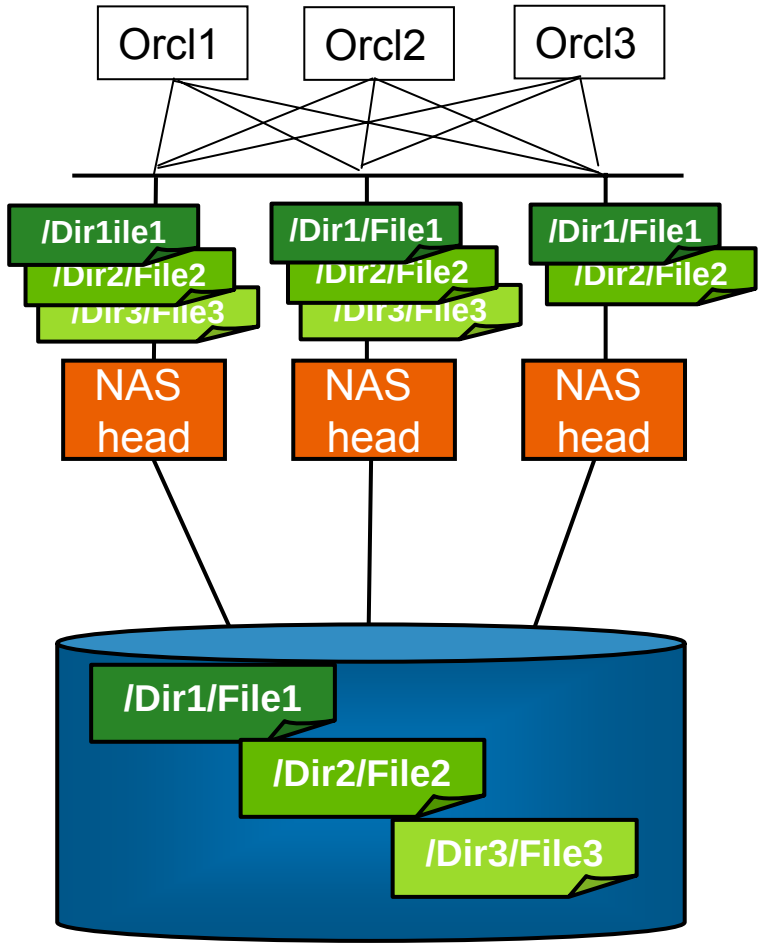
- Asymmetric multi-headed NAS
  - Oracle Servers can see data though one NAS head
  - Presenting the data through a different NAS head involves a migration or failover



Oracle non-RAC

Single instance databases on NFS Clients

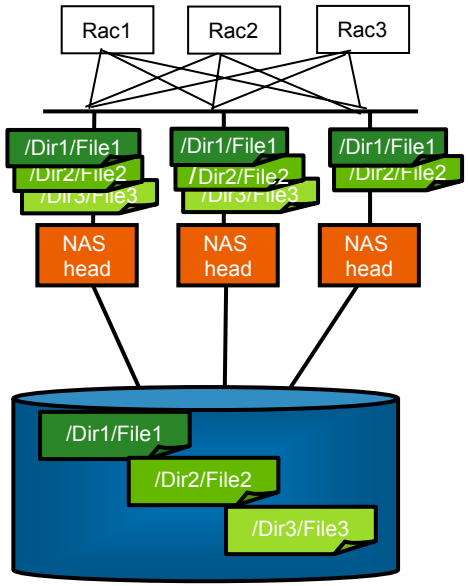
- Symmetric multi-headed NAS
  - An instance accesses the DB through any single NAS head
  - This path can be changed by pointing the client to a different NAS head rather than migrating data
    - All file systems present through all NAS heads



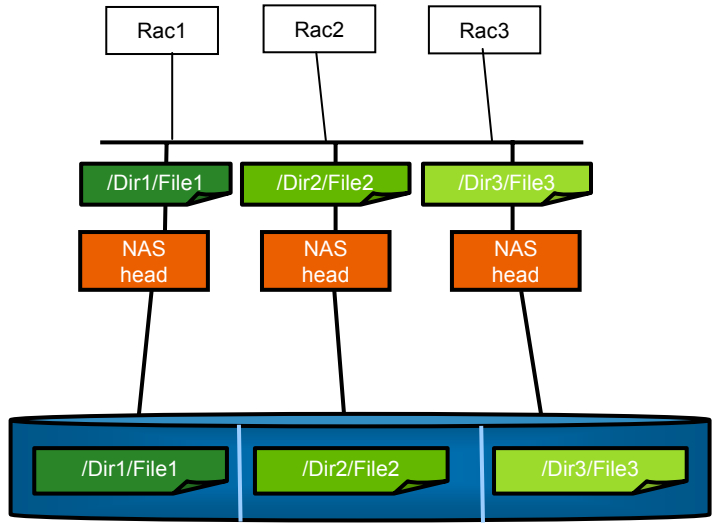
Oracle RAC

Multi-instance databases on NFS Clients

- Symmetric multi-headed NAS
  - A RAC Instance accesses the DB through any single NAS head
  - This path can be changed by pointing the client to a different NAS head rather than migrating data



- Asymmetric multi-headed NAS
  - RAC instances can see a specific file system though one NAS head
  - The database must be partitioned and each file system presented through a separate NAS head
  - Changing this configuration involves a data migration



Additional uses for symmetric NAS

- Central software provisioning
  - Shared Oracle software for legacy UNIX and Linux ports
  - Shared ORACLE_HOME, APPL_TOP, and applications tier
  - Centralized backup, patch, and upgrade
- Oracle RMAN backup target
  - Deploy scalable NAS as a backup target for existing Oracle databases
  - Single scalable storage system to grow as business grows
- General purpose grid storage provisioning
  - Easy to scale NAS for general purpose storage
  - ETL target, data warehousing, and general purpose file storage

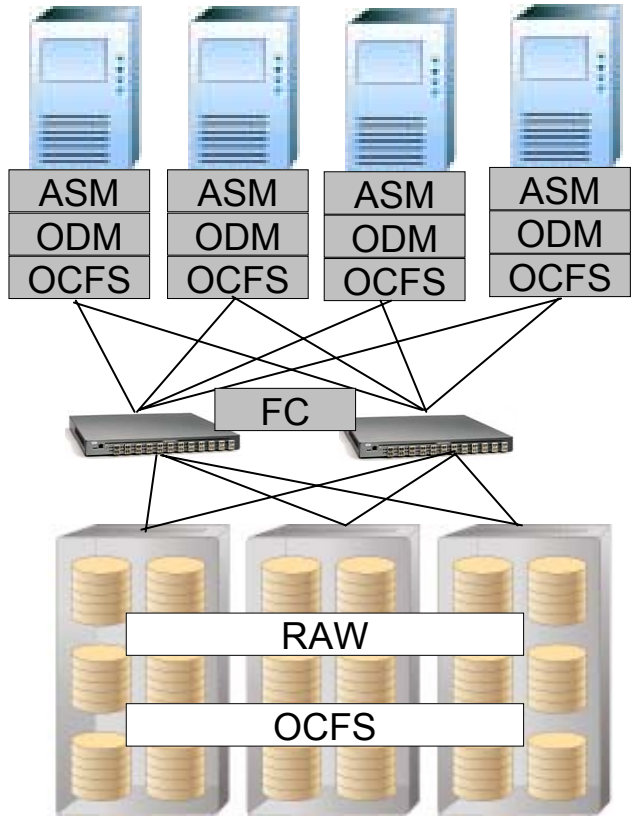
From Complex and Expensive….

Complex

- ASM, ODM, OCFS, OS File System
- Host-based volume mgnt & file system administration
  - Mix of RAW, CFS for database, Ext3/UFS for Oracle Home, etc.
- Fibre Channel SANs
  - LUN masking. Switch Zoning. MPIO.
- Difficult to increase capacity

Expensive

- Expensive Fibre Channel HBAs & switches
- Expensive to Manage—a lot of time spent managing a complex system



…To Simple and Inexpensive

Simple Provisioning & Management

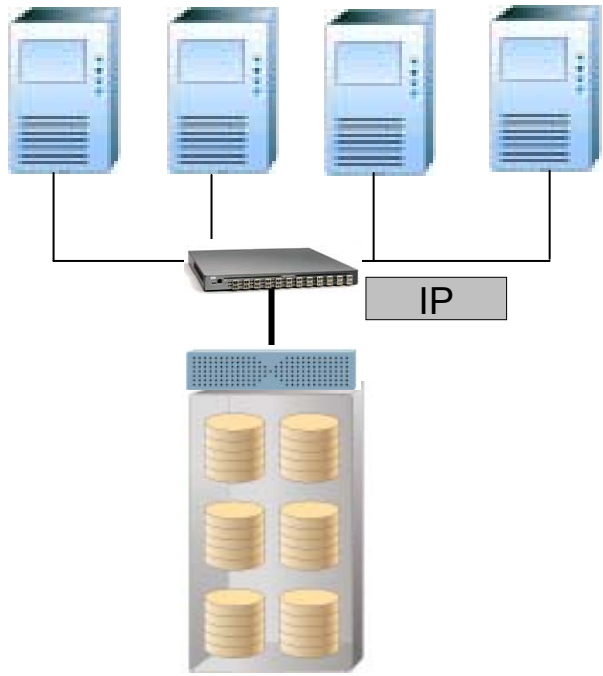
- As Easy as Ethernet
- One unified storage pool to manage
- Easy, online capacity growth

Improved Oracle Administration

- One solution for all of Oracle data
- Same model for RAC and non-RAC
- Shared storage simplifies database provisioning

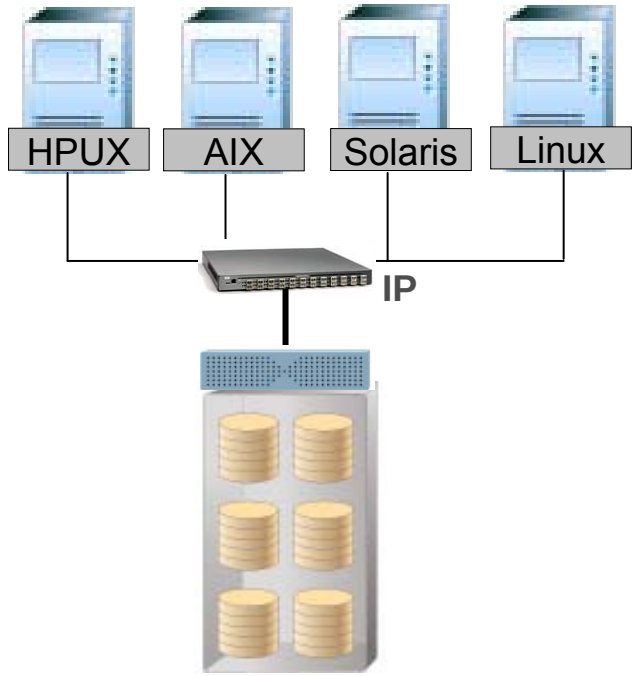
Reduce the Cost of Storage

- Inexpensive ethernet NICs & switches
- Amortize savings over many database servers



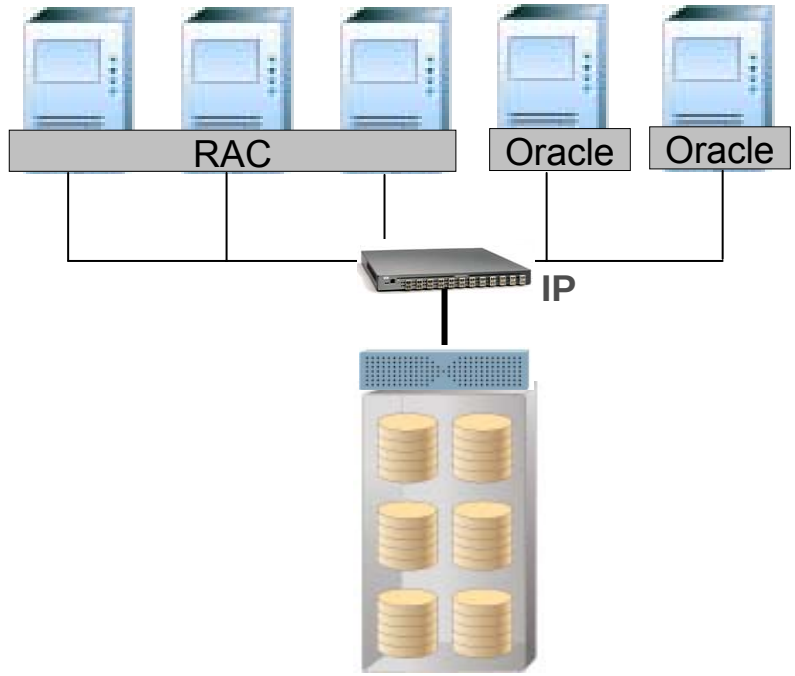
Mixed Use Environment

- Multi-platform support



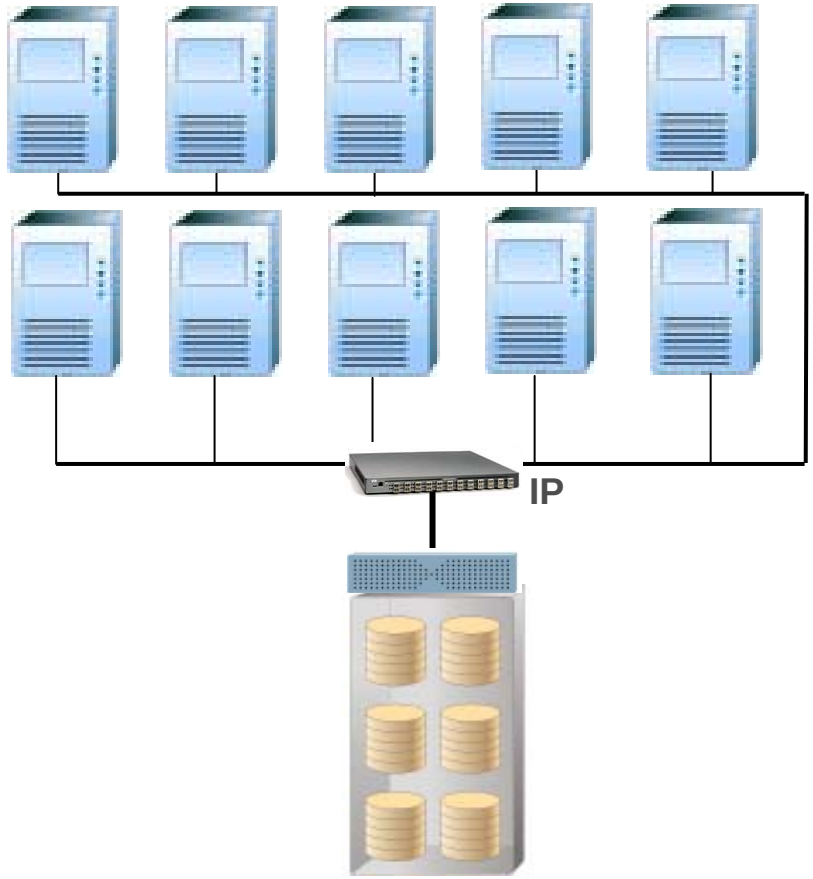
Mixed Use Environment

- Multi-platform support
- RAC and non-RAC



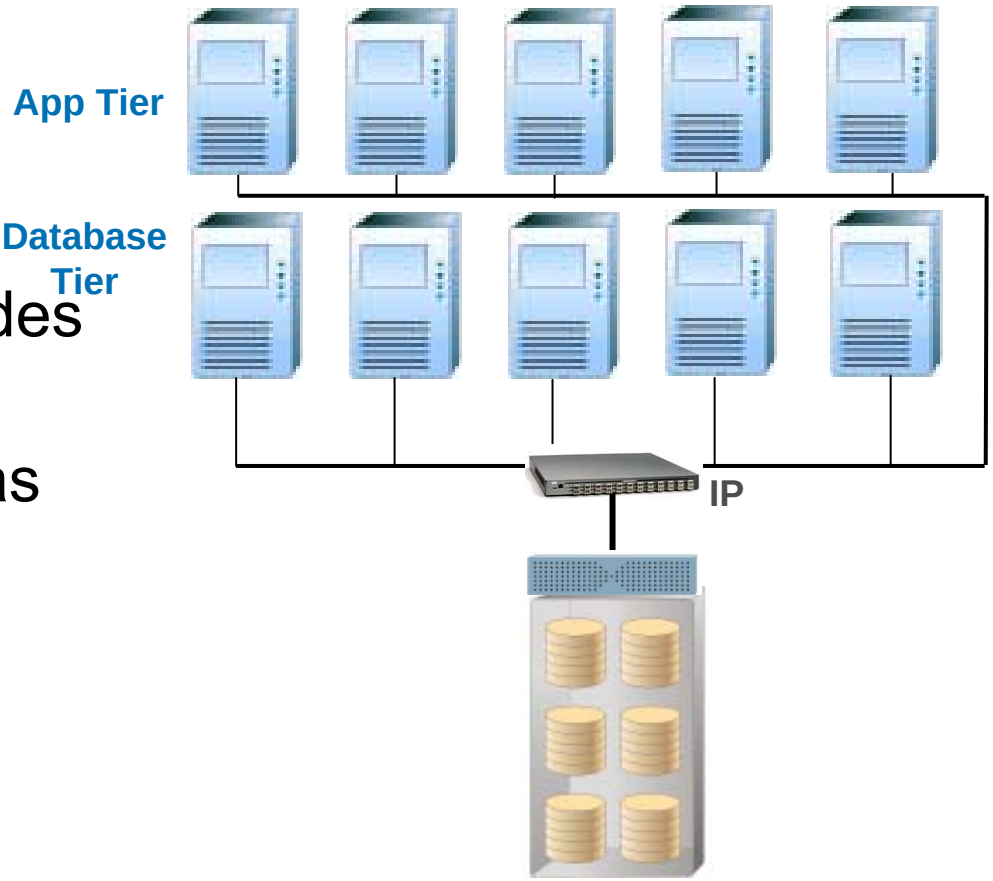
Mixed Use Environment

- Multi-platform support
- RAC and non-RAC
- Dozens of database nodes for 100's of databases



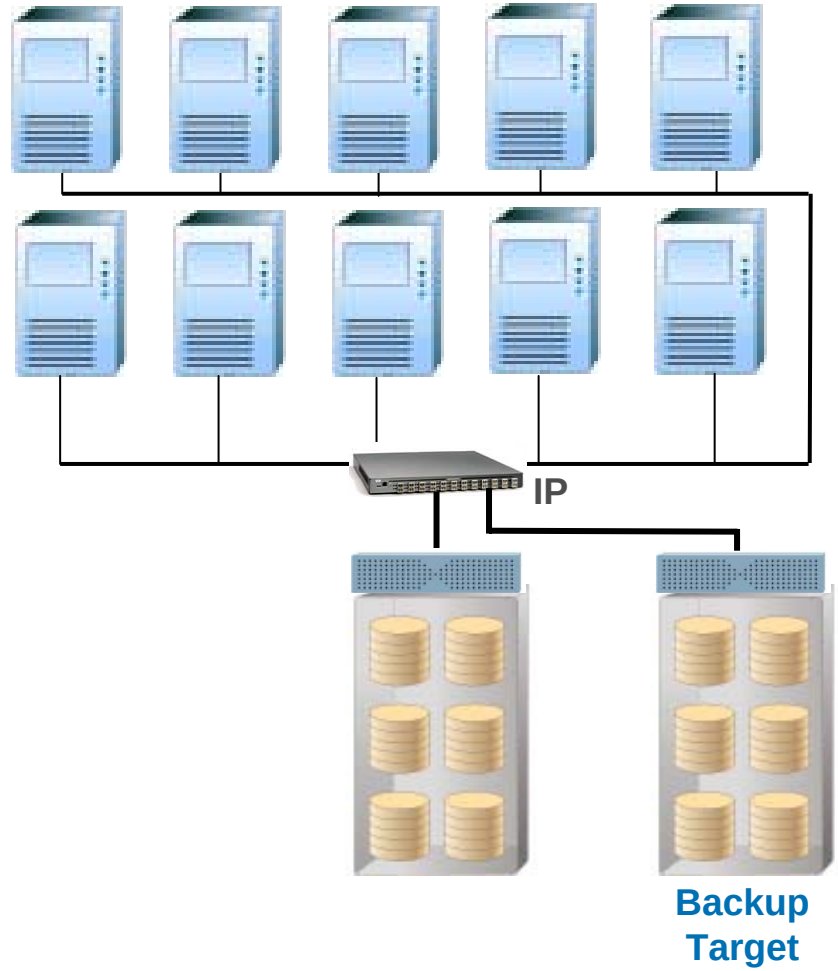
Mixed Use Environment

- Multi-platform support
- RAC and non-RAC
- Dozens of database nodes for 100's of databases
- Application tier as well as database tier



Mixed Use Environment

- Multi-platform support
- RAC and non-RAC
- Dozens of database nodes for 100's of databases
- Application tier as well as database tier
- RMAN backup target
  - Whether or not you use NAS for primary storage



Oracle symmetric scalable NAS Value

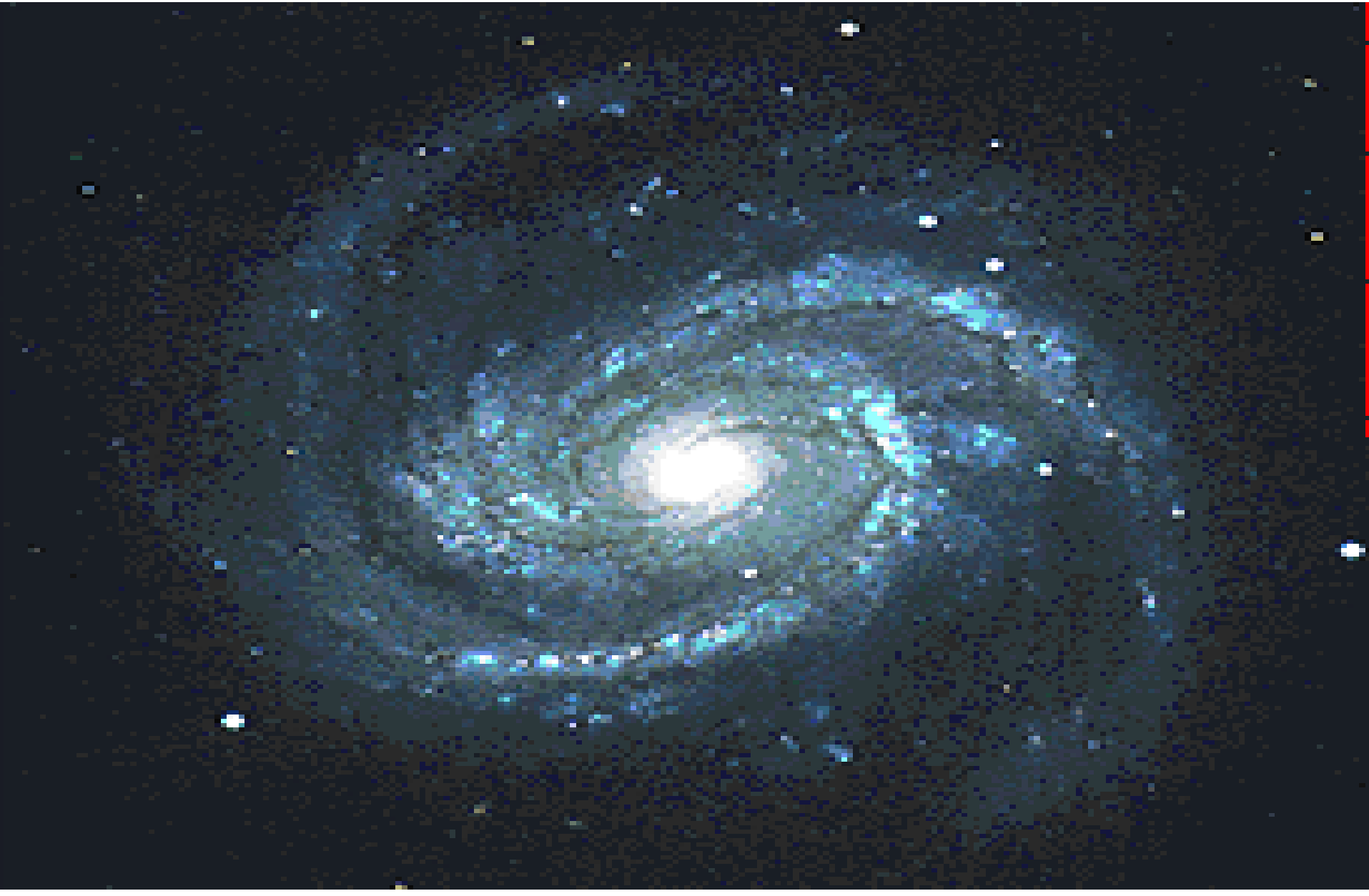
Drive complexity and cost out of Oracle deployments.

1. Less complex to configure and manage
2. Reduces infrastructure cost
3. Provides a scalable, cross platform storage solution
4. Leverages existing storage investment

Additional uses for Scalable NAS

- Scalable, central software provisioning
  - Shared Oracle software for legacy UNIX and Linux ports
  - Shared ORACLE_HOME, APPL_TOP, & applications tier
  - Centralized backup and patch/upgrade
- Oracle RMAN backup target
  - Deploy scalable NAS as a backup target for existing Oracle databases
  - Centralized nearline backup target for simplified storage management
  - Scalable performance & capacity on-demand – alleviates over provisioning risk and cost
- General purpose grid storage provisioning
  - Easy to scale NAS for general purpose storage
  - ETL target, data warehousing, general purpose file storage

HP's Scalable NAS for Oracle



HP's Scalable NAS for Oracle

Next Generation NAS ideal for Oracle NFS

- **Better Scalability**
  - Clustered-filer design provides more performance in a single system
  - 2 petabytes of data and growing
  - Support dozens of database nodes
  - Over 3 GB/s of aggregate throughput provides plenty of IO performance
  - Grow client support or storage capacity on demand
- **Better Availability**
  - Client transparent NFS failover for robust High Availability
    - Integrated, so no extra cost for HA

Oracle over NFS

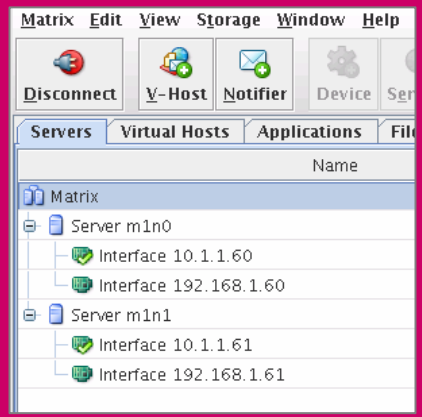
Scalable NAS for Oracle solution

- Single point of storage management across all Oracle databases and applications
- Scalable data access beyond the capabilities of traditional NAS
- Uninterrupted fault-tolerance for Oracle data
- Lower cost Oracle over NFS file serving

HP scalable NAS products

HP PolyServe software

- Matrix Server
  - SQL Consolidation
  - File Serving



EFS Clustered Gateway

- 2-16 DL380 G5s
  - Windows or Linux OS
  - Matrix Server

HP PolyServe software

The critical component in HP enterprise NAS systems

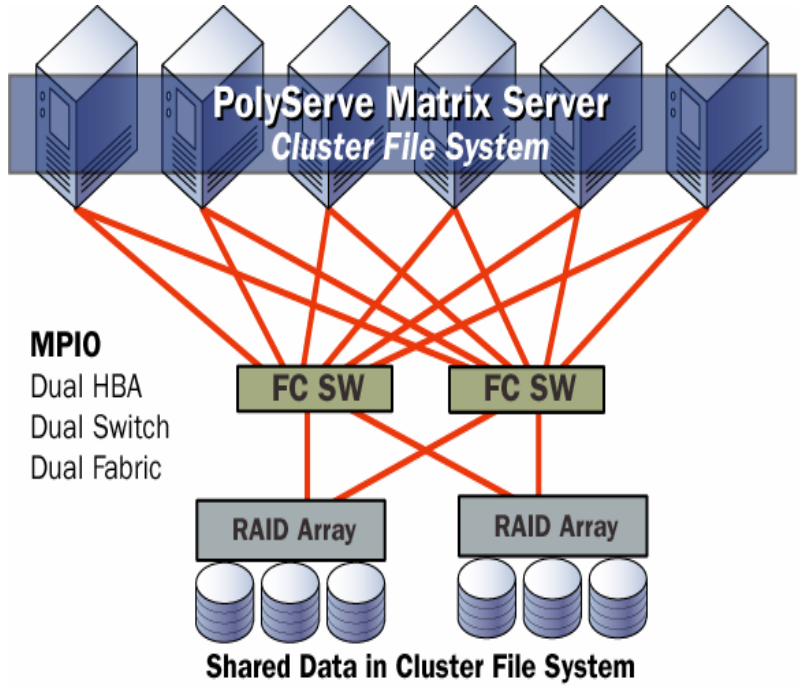
Designed without compromise for scalability and availability

Works with industry standard components

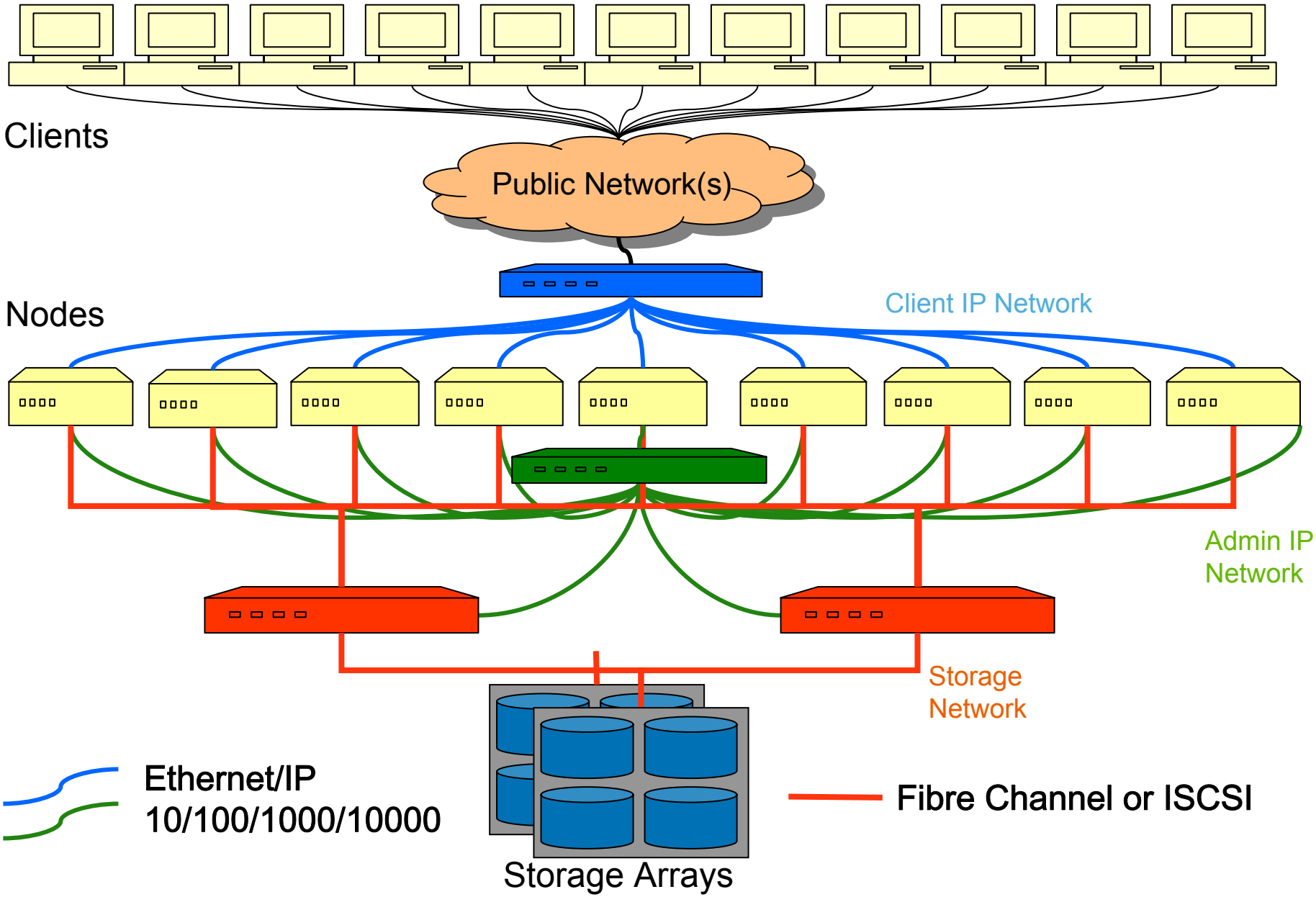
Engineering Team with 870 years of industry experience, 15+ years together

HP-PolyServe Matrix Server technology

- Cluster File System
  - All nodes can read and write ***all*** data concurrently
  - Up to 16 nodes per cluster
  - Any mix of node sizes and speeds
  - Full support for Linux and Windows
  - Certified by Oracle and Microsoft
  - Mix OS levels in one cluster
  - Mix PolyServe levels in one cluster
- Cluster Volume Manager
  - Virtualize storage across multiple LUNs or arrays
  - Striping, concatenation, online file system growth
- High-Availability Services
  - ***Fault tolerance*** using standard servers
  - Full classic clusterware features
  - Full n:1, n-m monitoring and failover

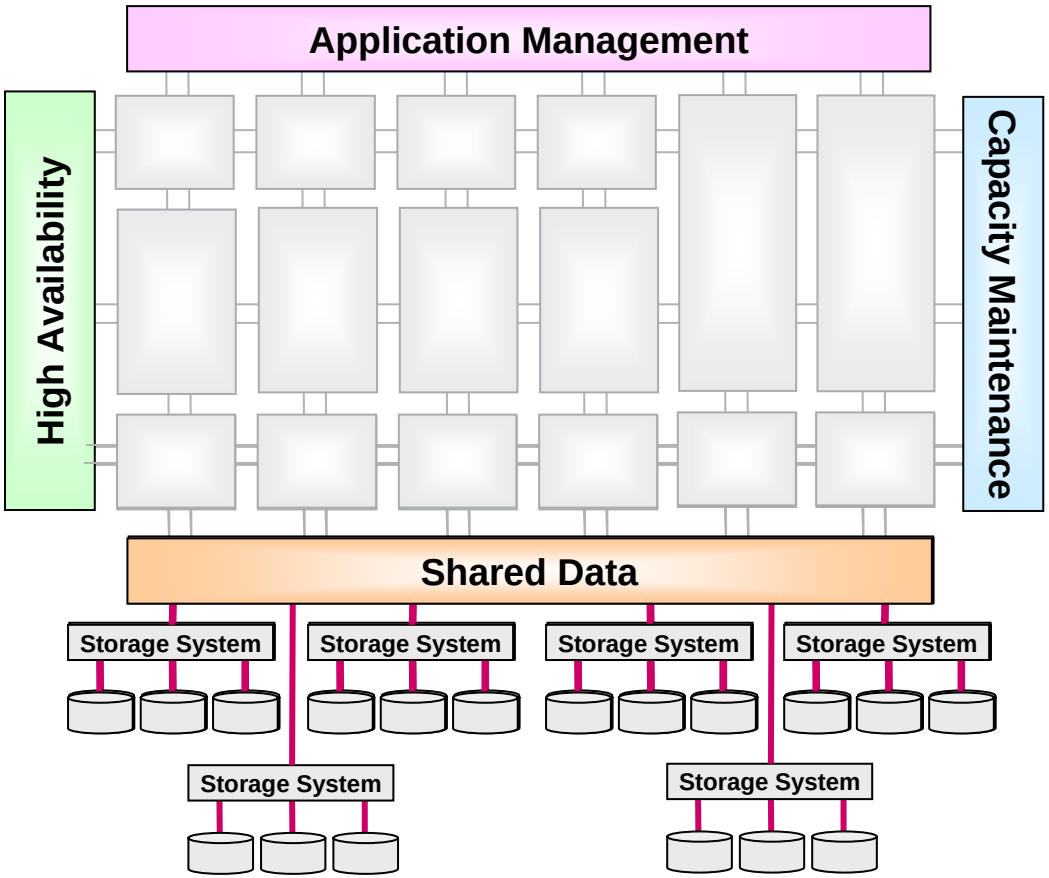


The physical structure of an HP - PolyServe matrix



HP-PolyServe delivering adaptive infrastructure On-Demand – The Always-On Computing Utility

Powered by shared data

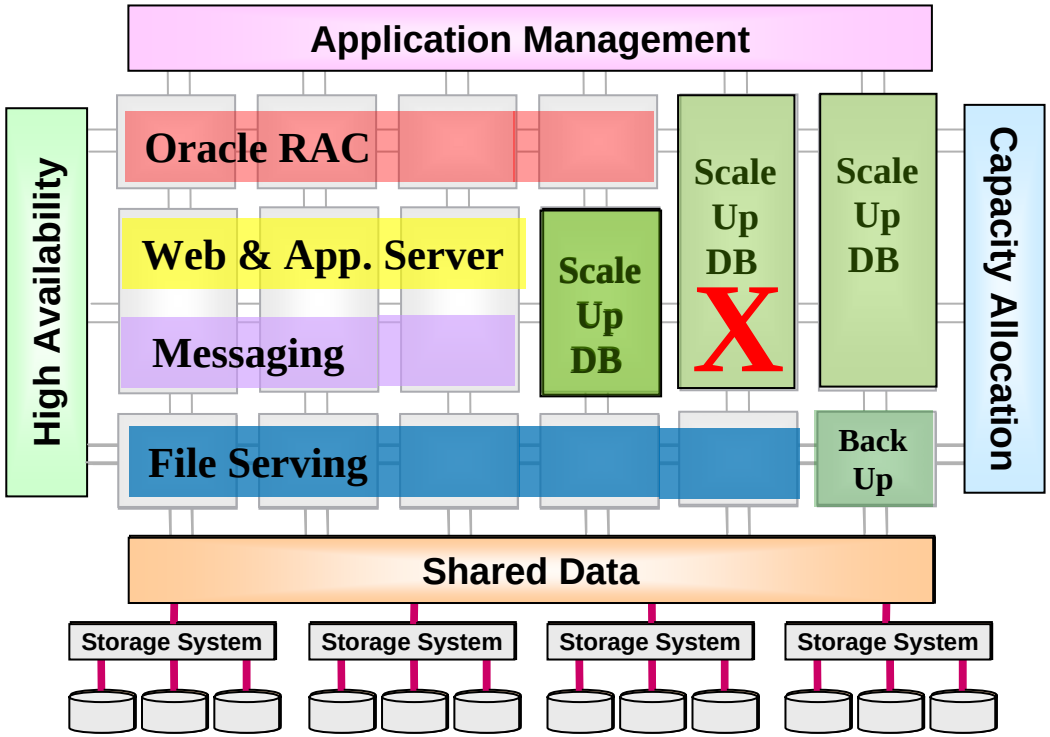


- Shared Data
  - Allows all servers to “own” all data with speed & integrity
- High Availability
  - **Easy**: Scale-out data access, data transition
  - **Affordable**: Mix server brands, CPU counts, CPU speeds, and OS versions
- Capacity Maintenance
  - Insert & upgrade servers and storage as needed, with no service disruption
- Application Management
  - Match applications to required capacity - dynamically

Mission-Critical Application Deployment Platform

- adaptive datacenter infrastructure for consolidation

HP-PolyServe - On-Demand, Always-on Capacity



- Scale-Out Applications
  - File Serving
  - Media Streaming
  - Oracle RAC
  - Web & Application Servers
  - Messaging – MQseries,Tibco
  - Back-up and Restore
- Scale-up Databases
  - SQL Server
  - DB2
  - Oracle8i, 9i & 10G
  - My SQL
  - Sybase
  - Etc.

HP EFS Clustered Gateway

Scalability and performance

- Linear scalability up to 16 nodes
- 128 TB file systems up to 2 PB total storage
- Throughput over 3 GB/s

Availability

- Fully transparent failover – preserving client state information

Storage Utilization

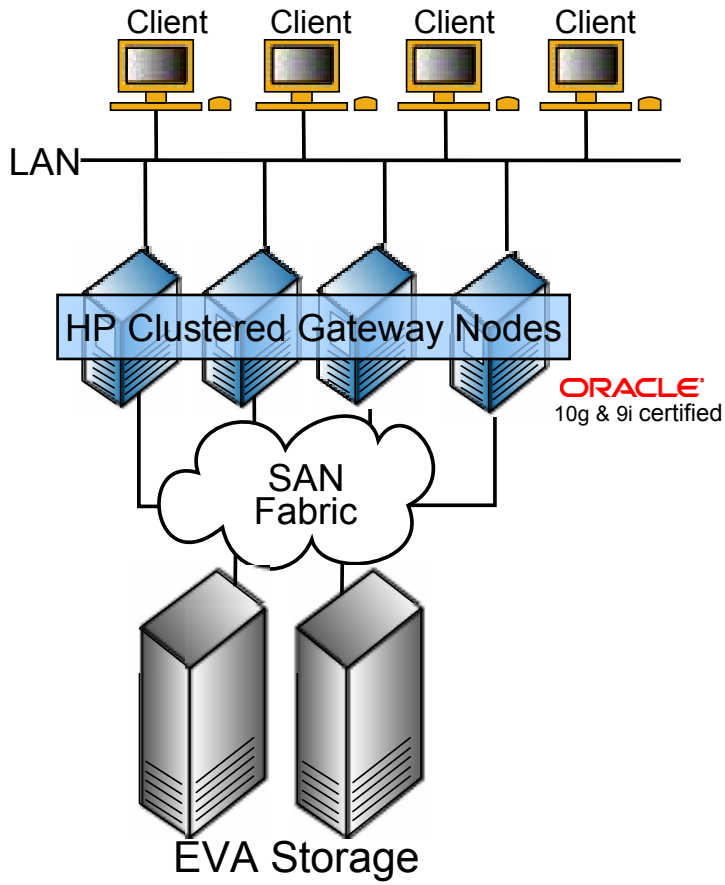
- Create a single pool of storage
- Virtualization across heterogeneous storage

Manageability

- Manage the cluster from anywhere
- Utilize standard HP management tools
- Standard OS integrates into data center

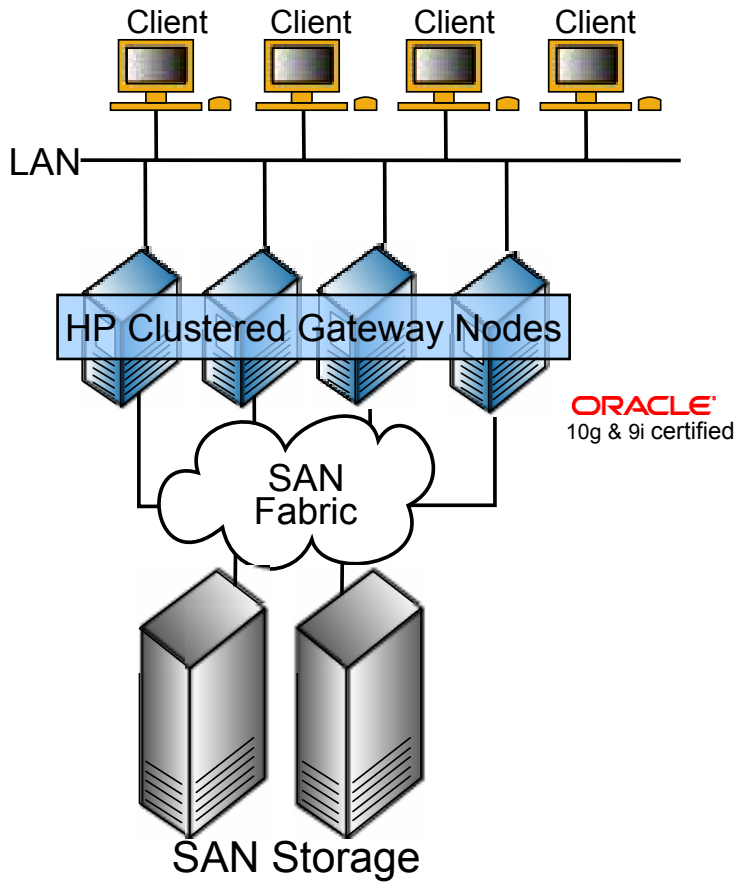
Value

- Industry leading price performance
- Industry standard components

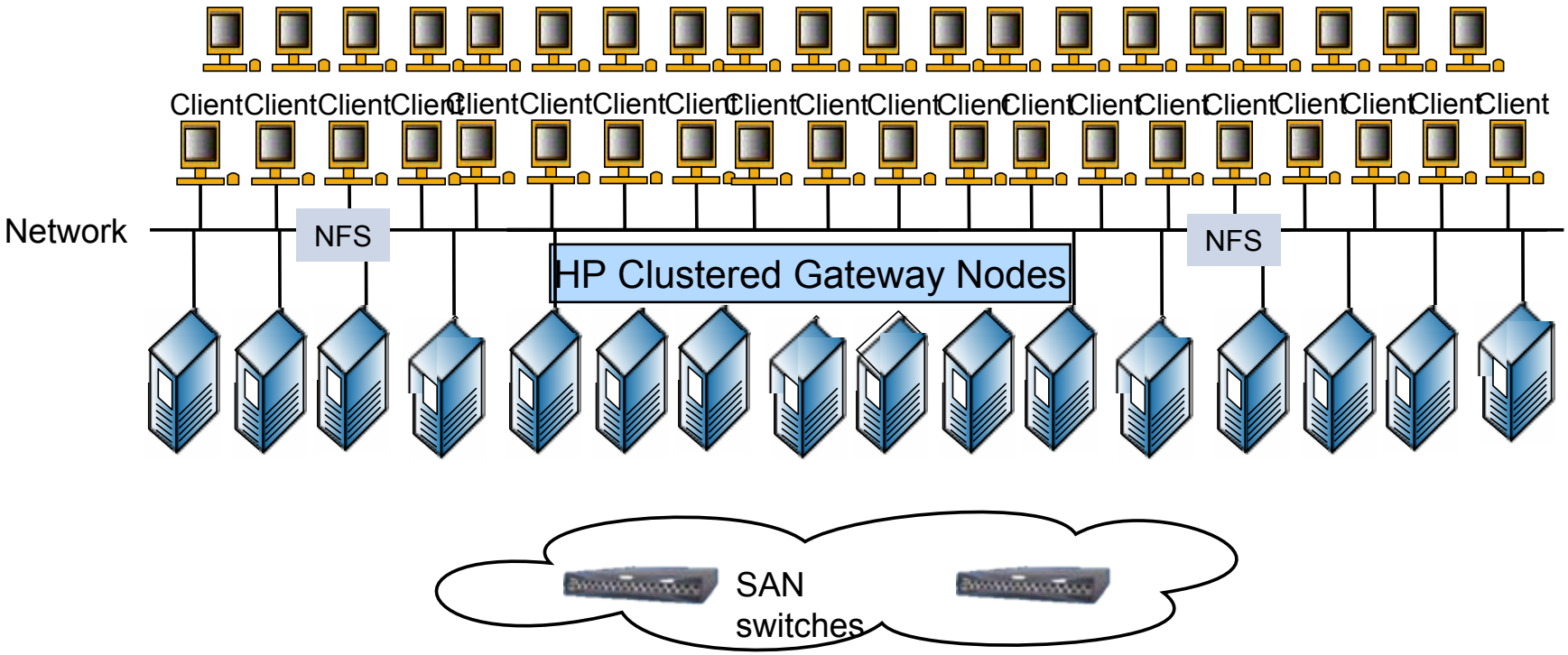


Building Blocks

- HP EFS Clustered File System Software
- Standard Proliant Servers as Cluster Nodes
  - Rack or Blades
- Shared Storage: HP or Legacy
  - SAN and/or iSCSI
  - HP Storage:
    - MSA (value focused)
    - EVA (optimal ease-of-use)
    - XP (maximum performance)
  - Third Party Storage: HDS, EMC Clariion, EMC Symmetrix



HP EFS Clustered Gateway



1. Shared storage
2. All nodes active on all files and disks
3. Linear performance growth
4. No failover necessary

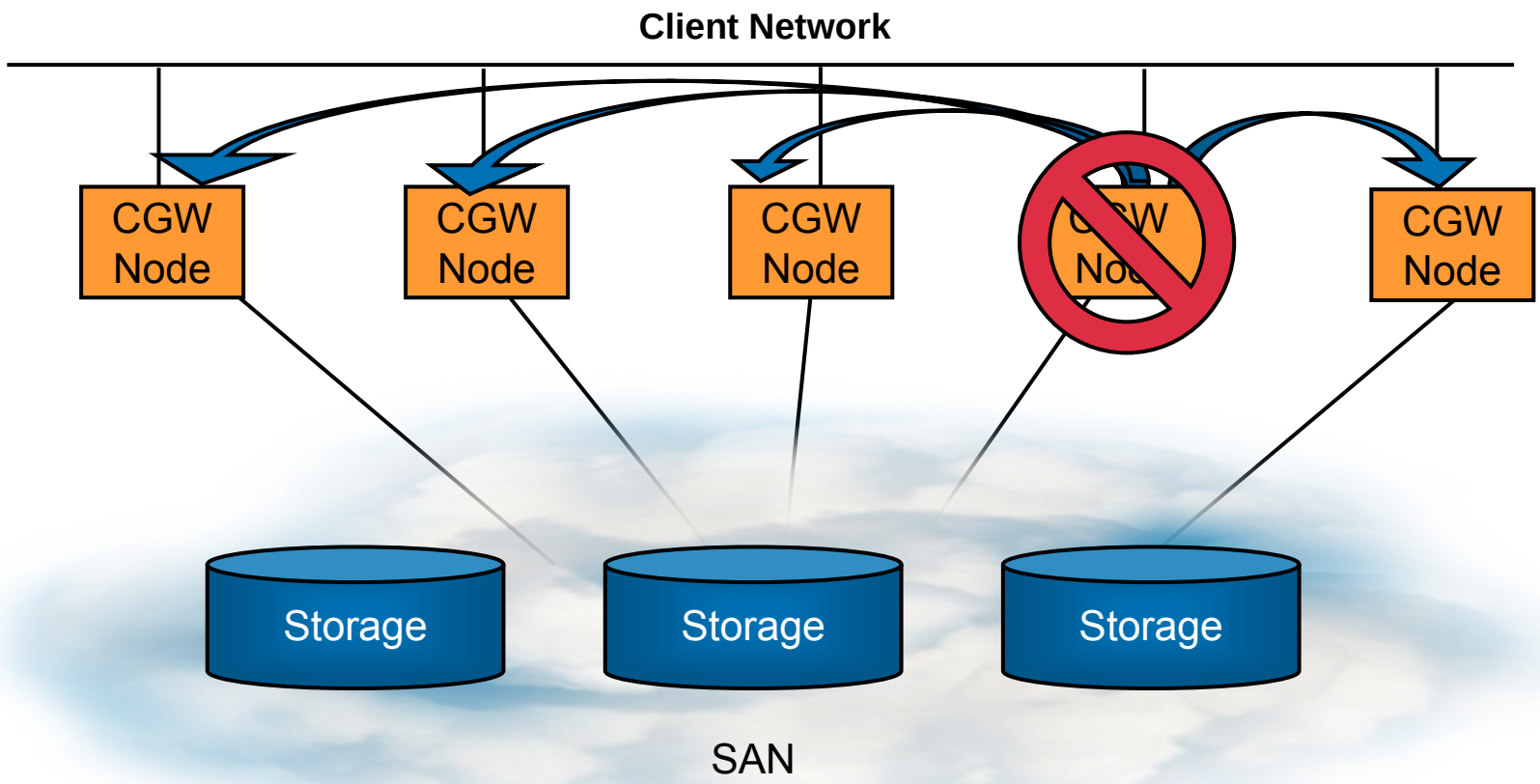
External Storage

1. Single point of backup
2. Single point remote mirror
3. Array not JBOD

EFS Clustered Gateway specs

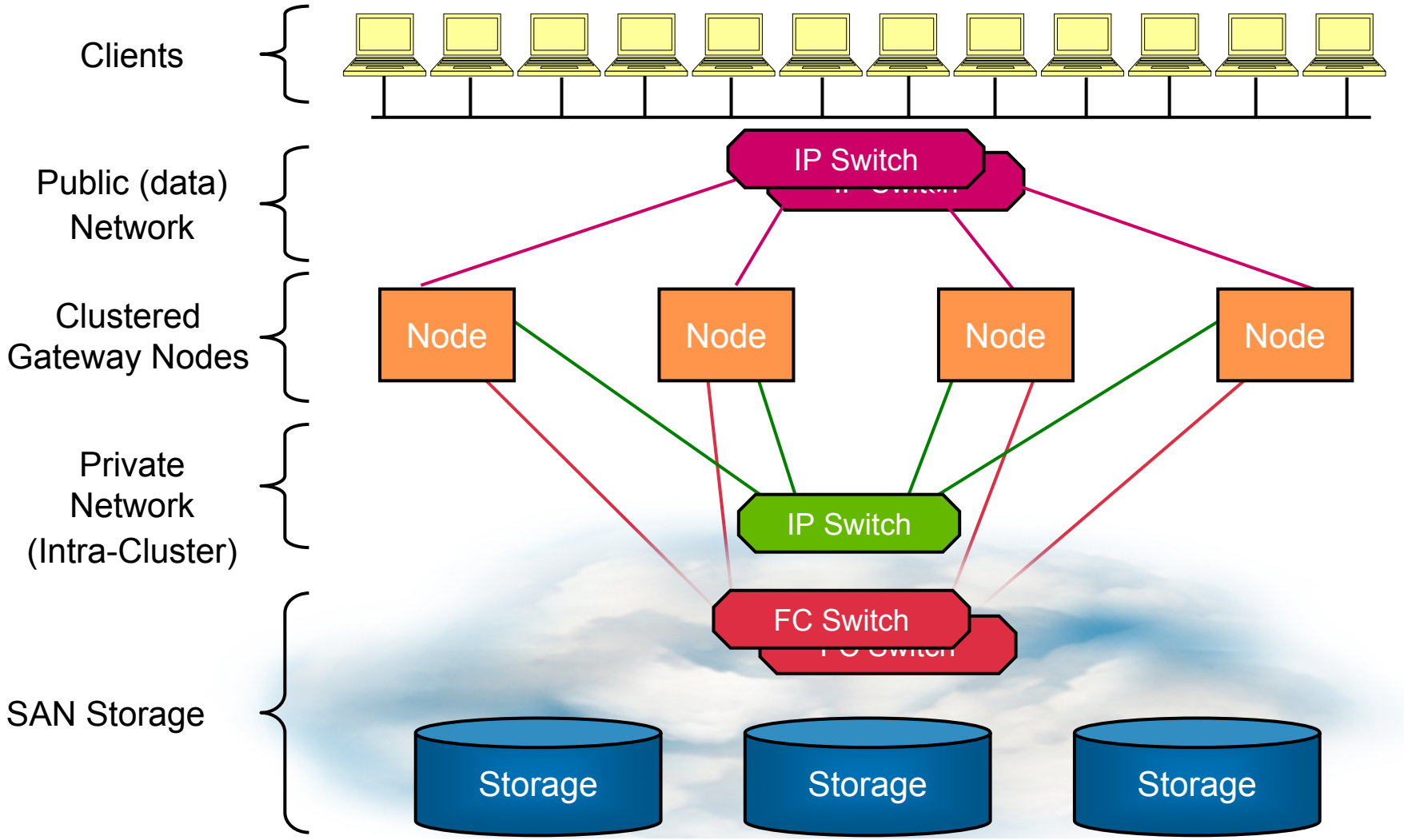
- It allows up to 16 Windows or Linux servers to read and write to the same storage pool
- It therefore allows them to share data and break the limits of what a single server can do
- This enables -
  - Scalability (e.g. file serving)
  - Manageability (e.g. database consolidation)
  - Availability (everything!)

HP EFS Clustered Gateway Symmetrical File System

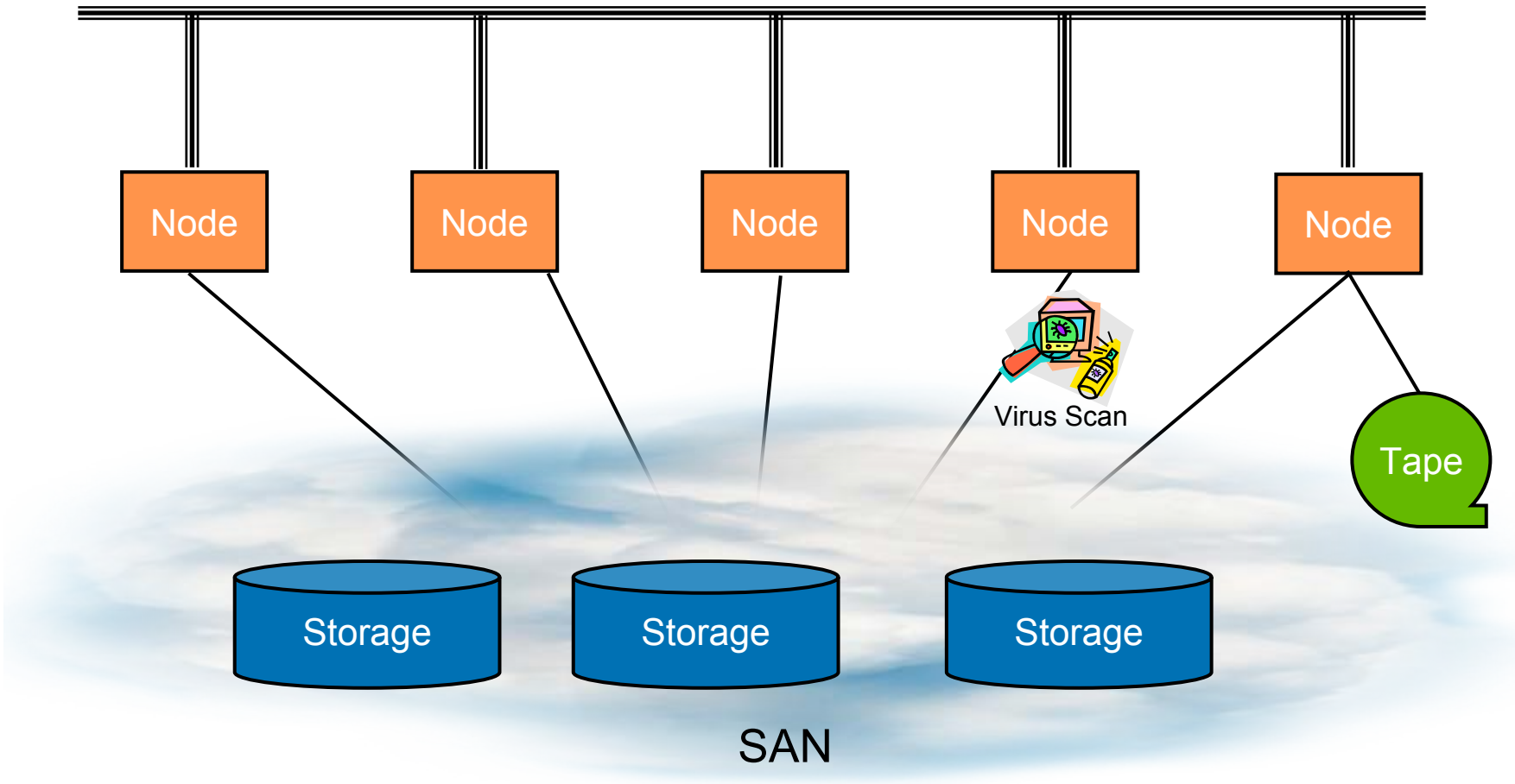


- Each node sees the complete file system
- Any node can fail over for any other
  - More efficient use of node hardware
- Eliminates hot spot & load balancing issues
- Distributed Lock Manager – scales with cluster

EFS Clustered Gateway interconnect

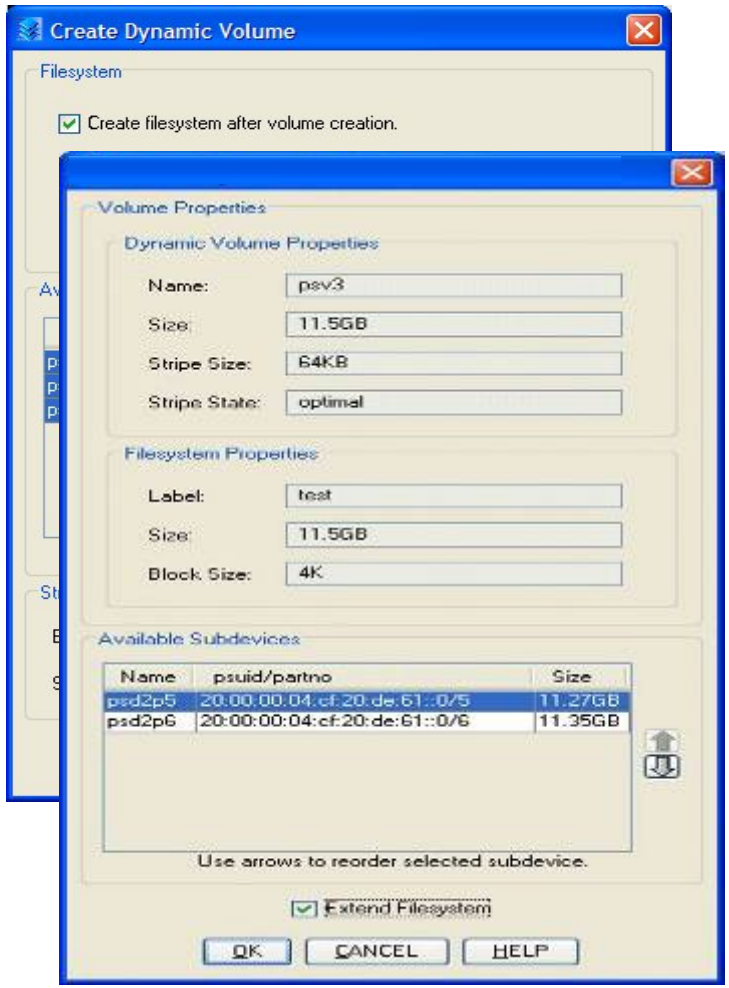


Special purpose nodes

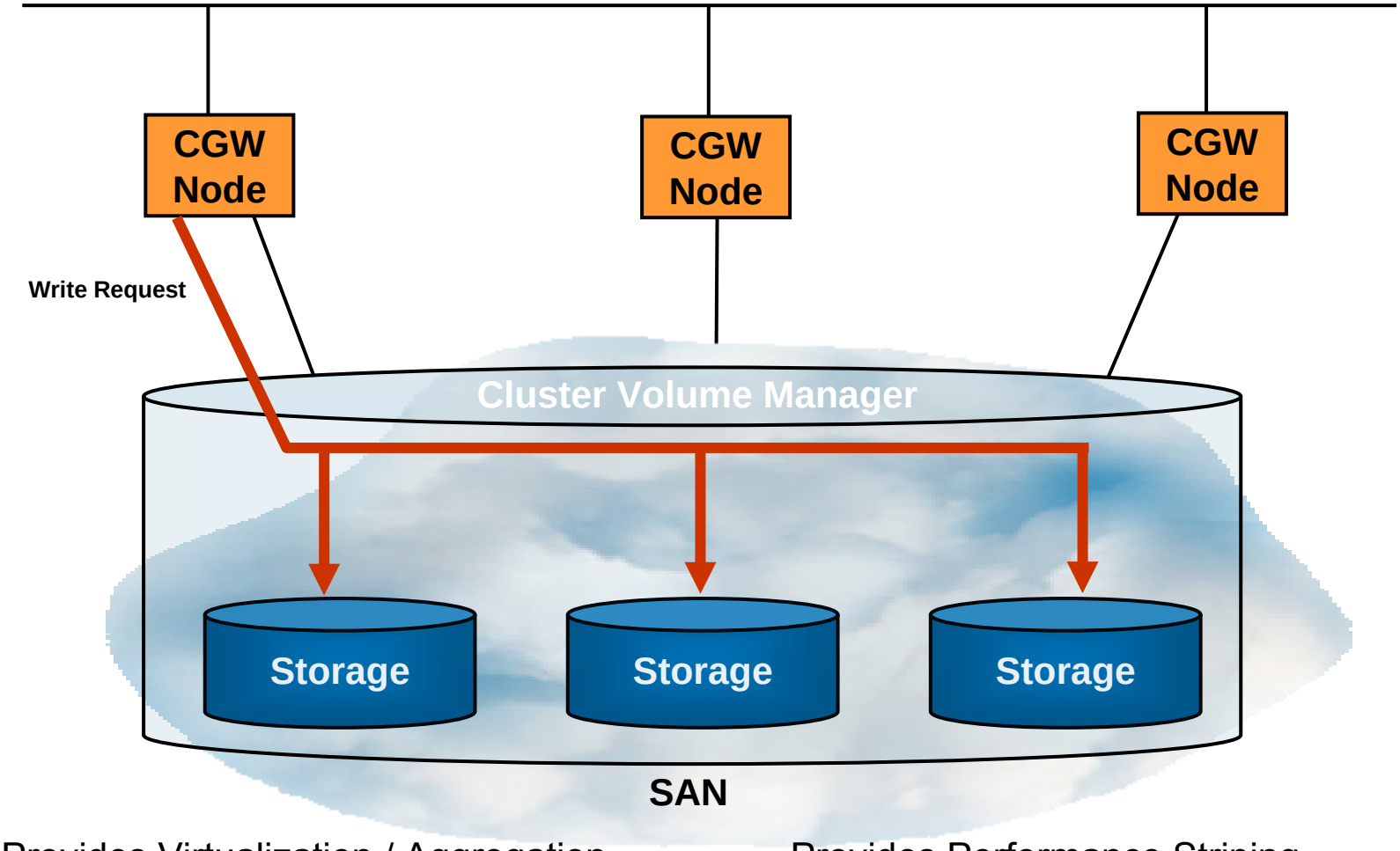


HP Clustered Gateway Cluster Volume Manager

- Improve storage utilization across cluster
- Optimize servers and storage for price-performance
- Flexibly manage storage across your business
- Configurable striping – optimize for price and performance
- Stripe across LUNS within an array or LUNs spanning multiple arrays

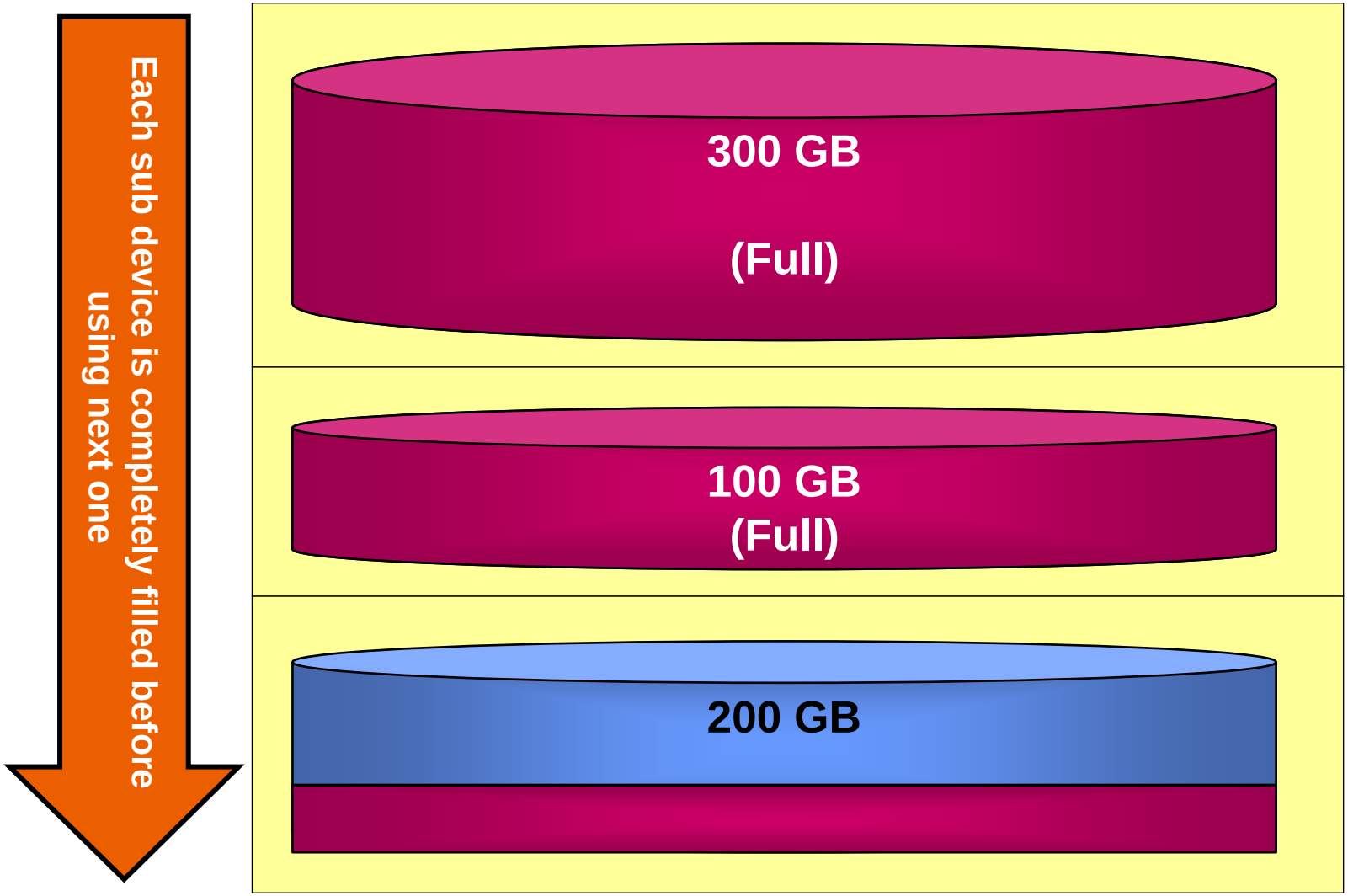


HP Clustered Gateway
Cluster Volume Manager

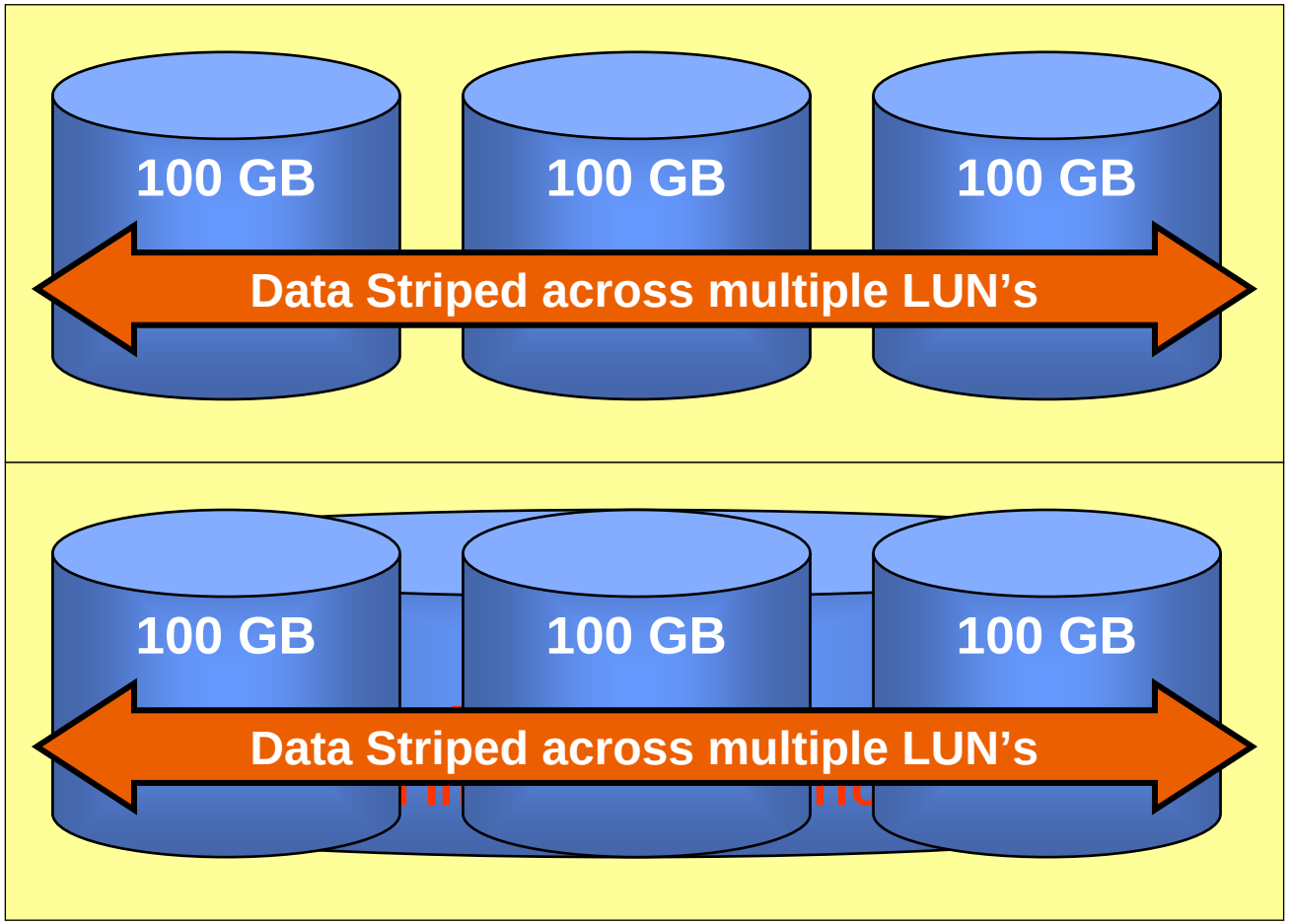


- Provides Virtualization / Aggregation
- Provides Performance Striping

Cluster Volume Manager
Growing Volumes (concatenation)

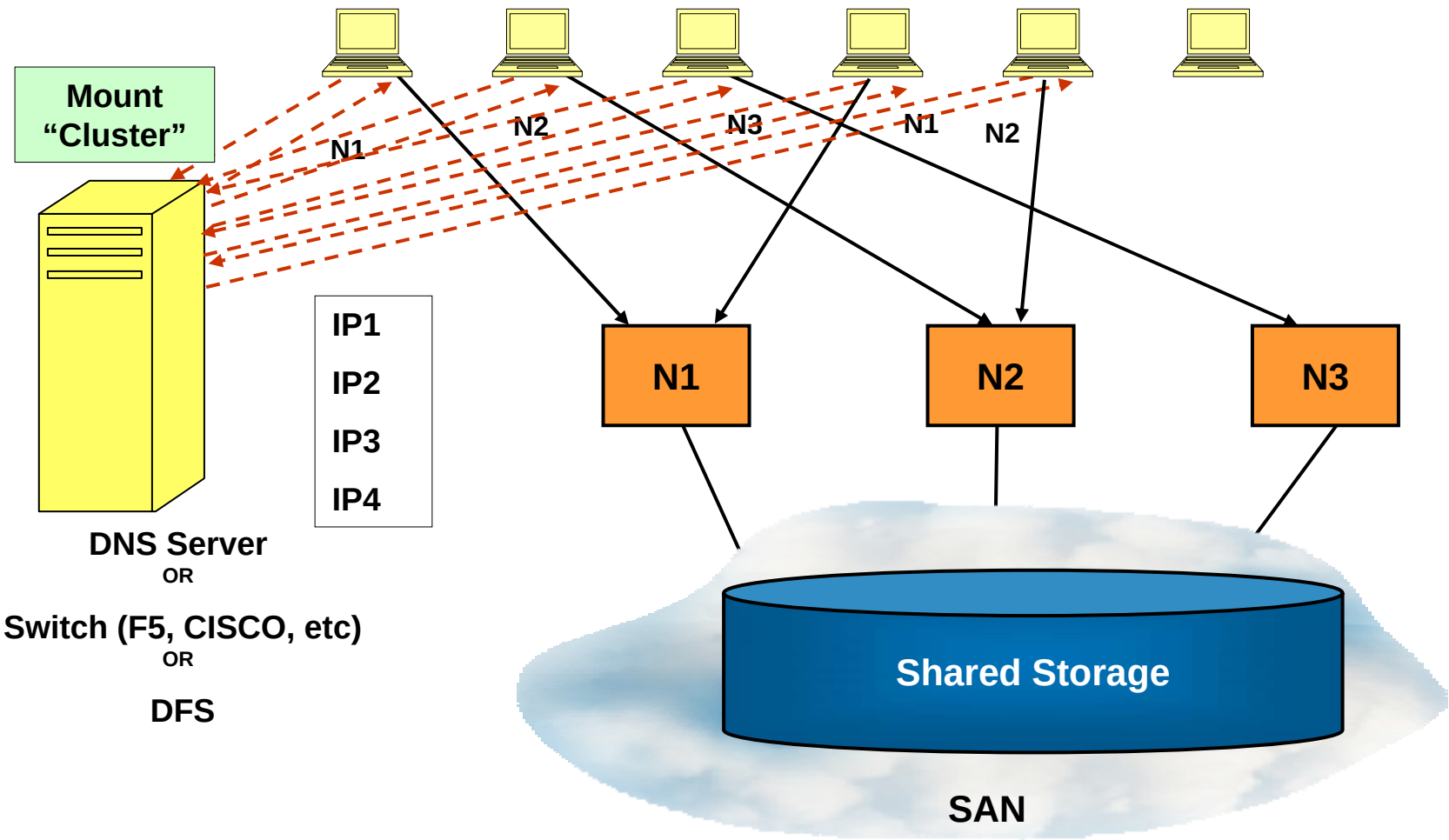


Cluster Volume Manager
Growing Striped Volumes

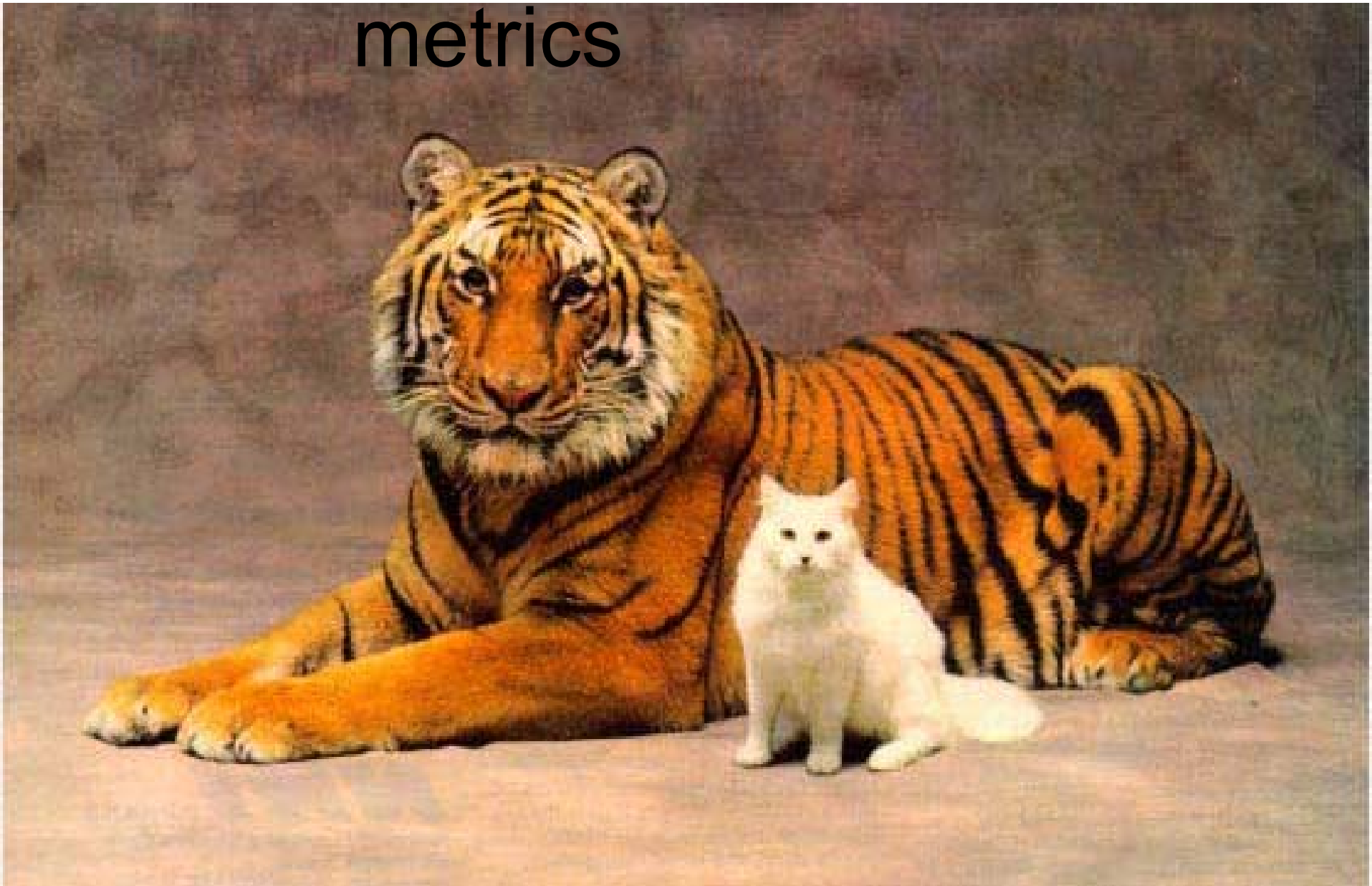


- **Grow striped volumes by adding uniform stripe sets for consistent performance**

Client Load Balancing

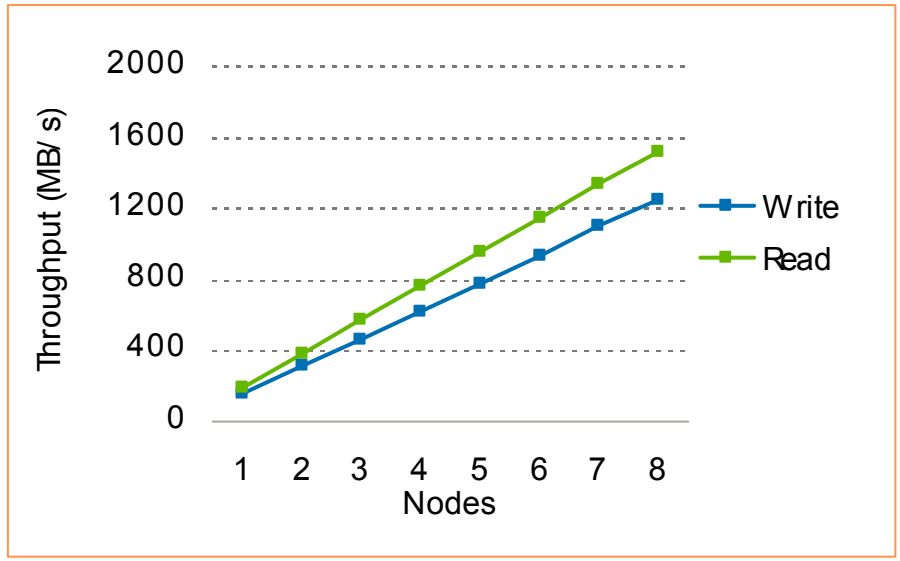


Performance metrics



Scalable performance – CIFS

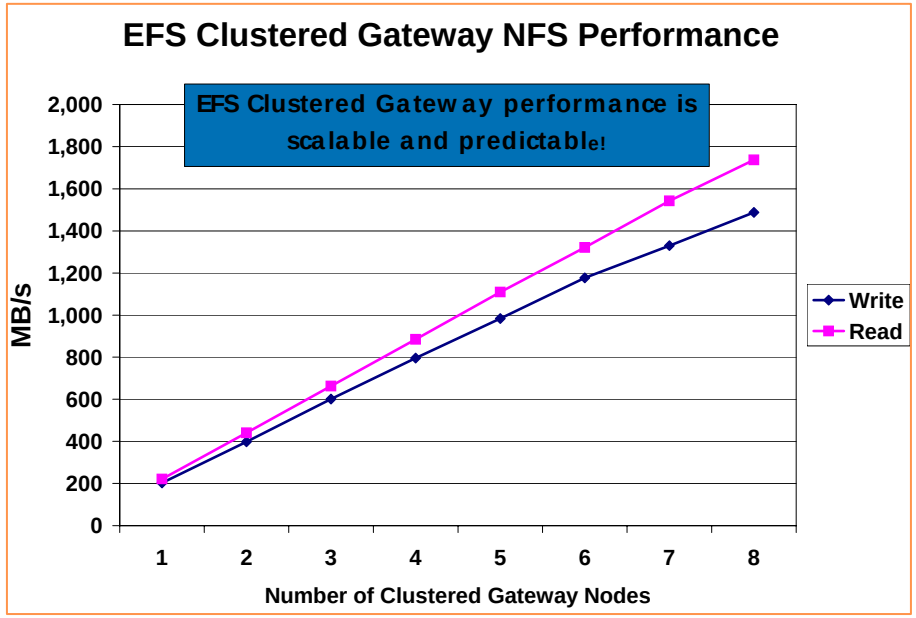
- Incrementally Scalable > 3GB/s
  - Supports 16 nodes for huge bandwidth
  - Can refresh old hardware with new—no need to match hardware specs in cluster
- Client Traffic is Load-Balanced
  - Client mounts are spread across all nodes
  - All nodes can cache and serve all files—no hotspots



# Servers	Mbytes/Sec.	Scale Factor	Scaling Coefficient
1	157	1.00	100%
2	310	1.97	99%
3	465	2.96	99%
4	620	3.95	99%
5	775	4.94	99%
6	935	5.96	99%
7	1,098	6.99	100%
8	1,243.0	7.92	99%

Scalable performance – NFS

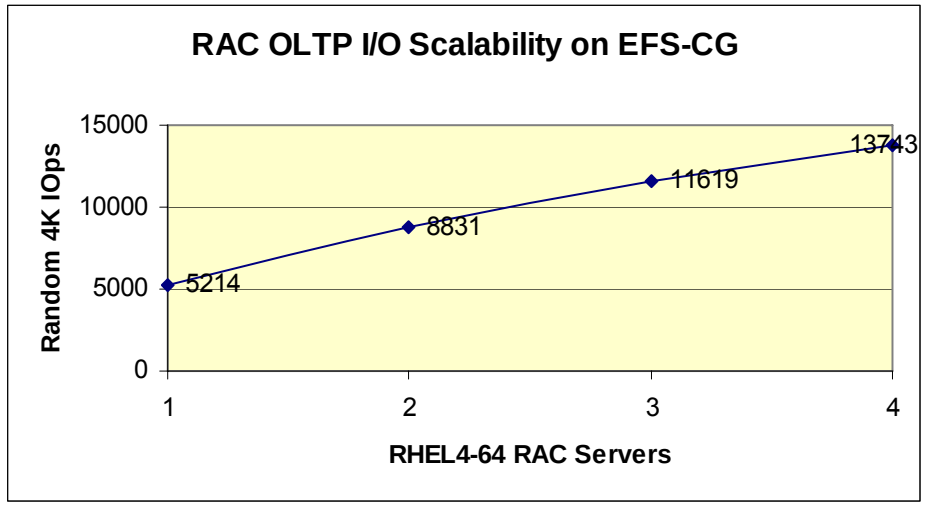
- High degree of scalability
  - Scales to over 3 GB/s
  - Supports 16 nodes
  - Can use mix of different node types
- Client Traffic is Load-Balanced
  - Clients are spread across nodes
  - Can use round-robin assignment or load balancer

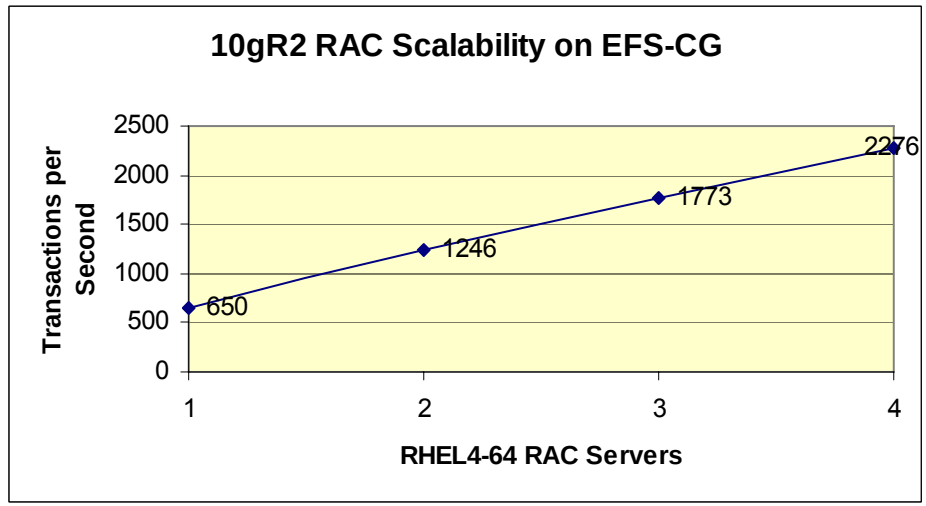


# Servers	Megabytes/second Write	Megabytes/second Read	Scaling Factor Write	Scaling Factor Read	Scaling Coefficient Write	Scaling Coefficient Read
1	202.663	221.559	1.00	1.00	100%	100%
2	398.124	441.045	1.96	1.99	98%	100%
3	601.358	662.925	2.97	2.99	99%	100%
4	795.857	885.327	3.93	4.00	98%	100%
5	983.787	1109.29	4.85	5.01	97%	100%
6	1176.74	1321.05	5.81	5.96	97%	99%
7	1329.72	1542.49	6.56	6.96	94%	99%
8	1487.96	1737.46	7.34	7.84	92%	98%

Scalable performance for Oracle

- Linear scaling performance
- Add Capacity on Demand
- Performance optimized for database transactions





For a full report

Scalable, Fault-Tolerant NAS for Oracle—The Next Generation

http://h71028.www7.hp.com/ERC/downloads/4AA0-4746ENW.pdf

Q/A