

# Cloud Storage with IBM Scale-out NAS (SONAS)

## GPFS in-a-box

Sep 13, 2011

**Florin MANAILA**  
Advisory Accredited Architect  
Systems and Technology Group  
[florin.manaila@ro.ibm.com](mailto:florin.manaila@ro.ibm.com)



# Agenda

Why we built SONAS

---

Policy-based Data Management

---

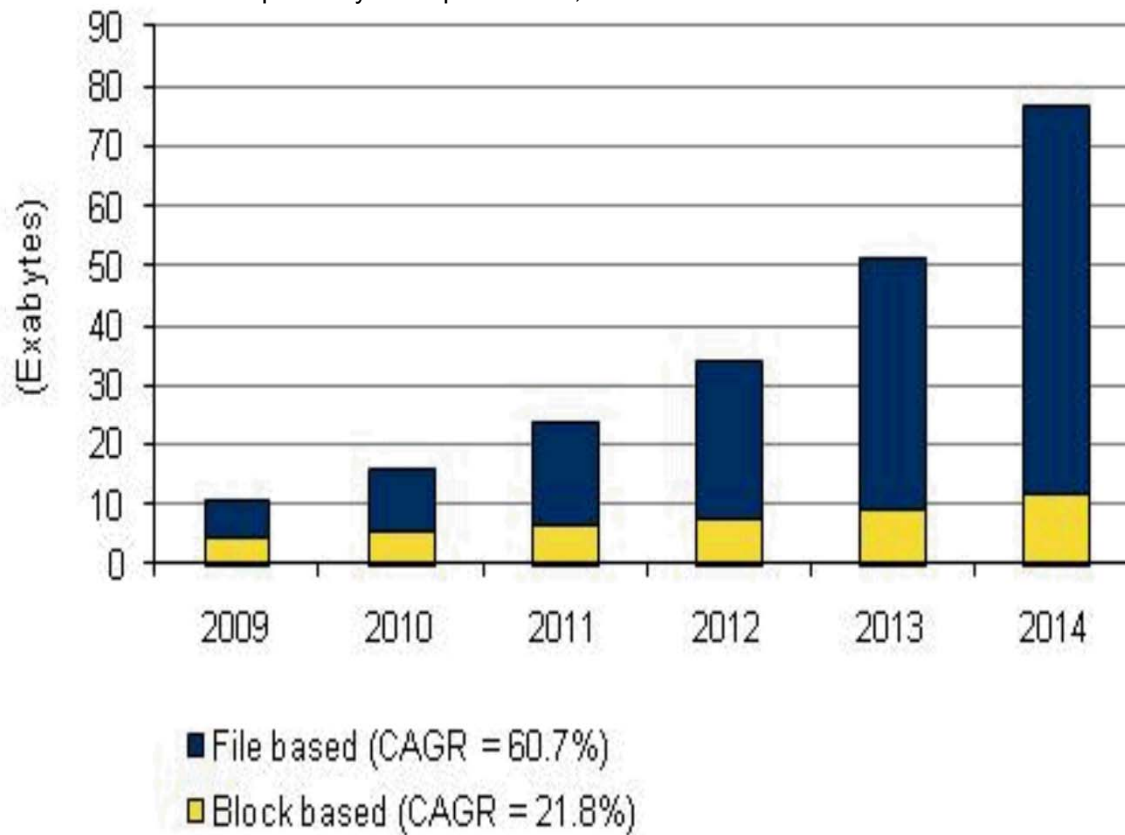
The SONAS Architecture

---

Outlook

# Storage Challenge: Demand increases faster than Harddisk Technology Gain

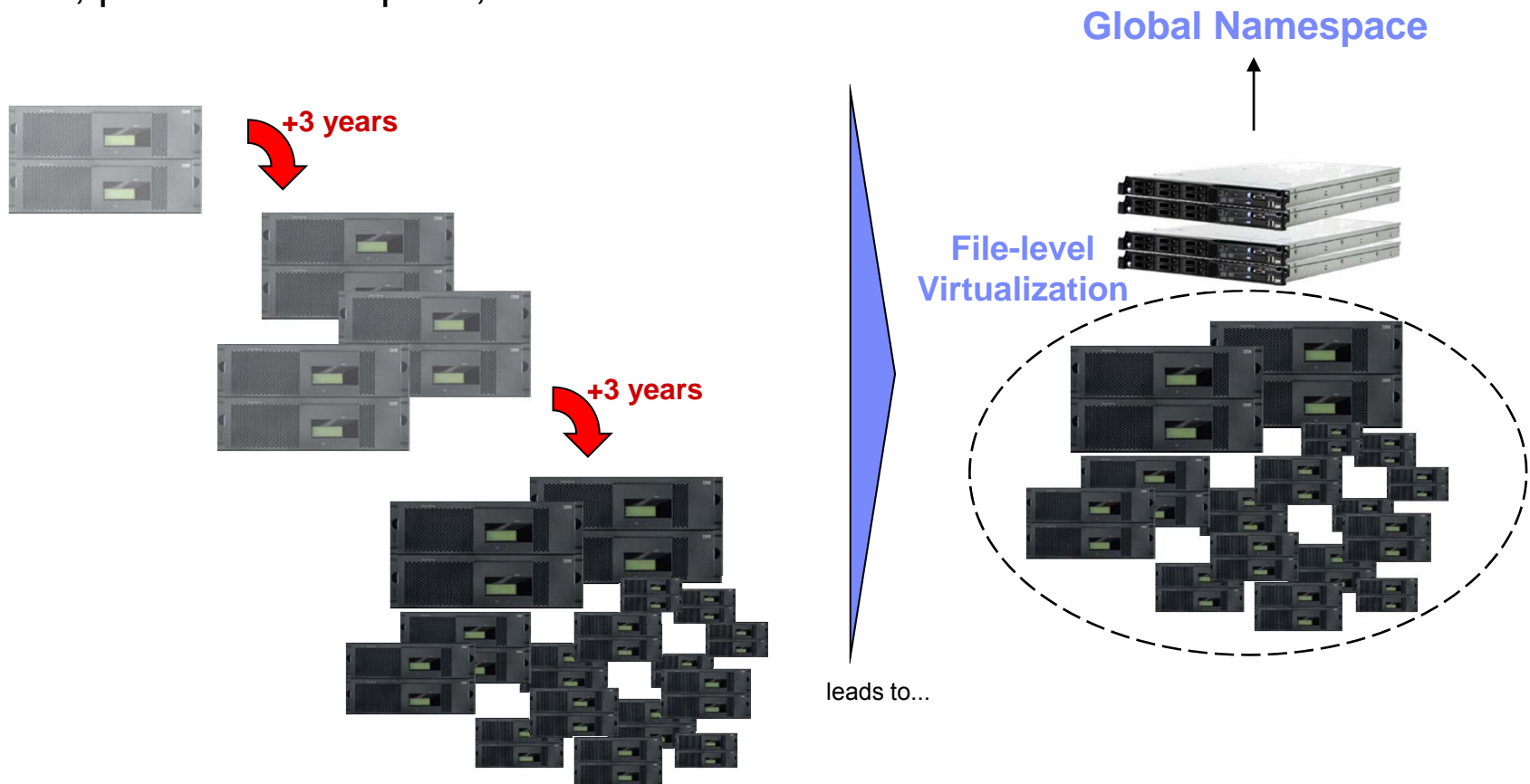
Worldwide File-Based vs Block-Based Storage  
Capacity Shipments, 2009–2014



Source: IDC's 2010 Enterprise Disk Storage Consumption Model

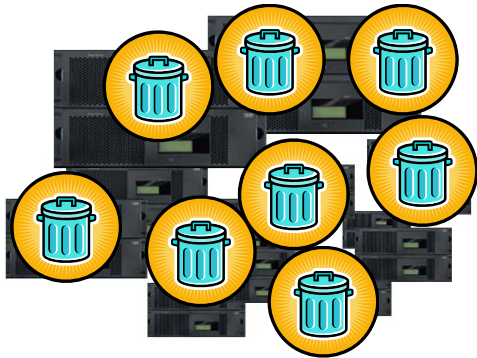
# Challenge 1 : Proliferating NAS & File Servers

- Growing # of systems, of administrative effort, power consumption, etc.

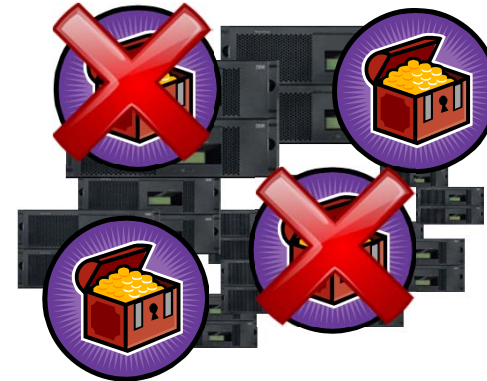


## Challenge 2 : Users don't clean up

- **Userspace** consists of 50% or more inactive data. Just which 50%?

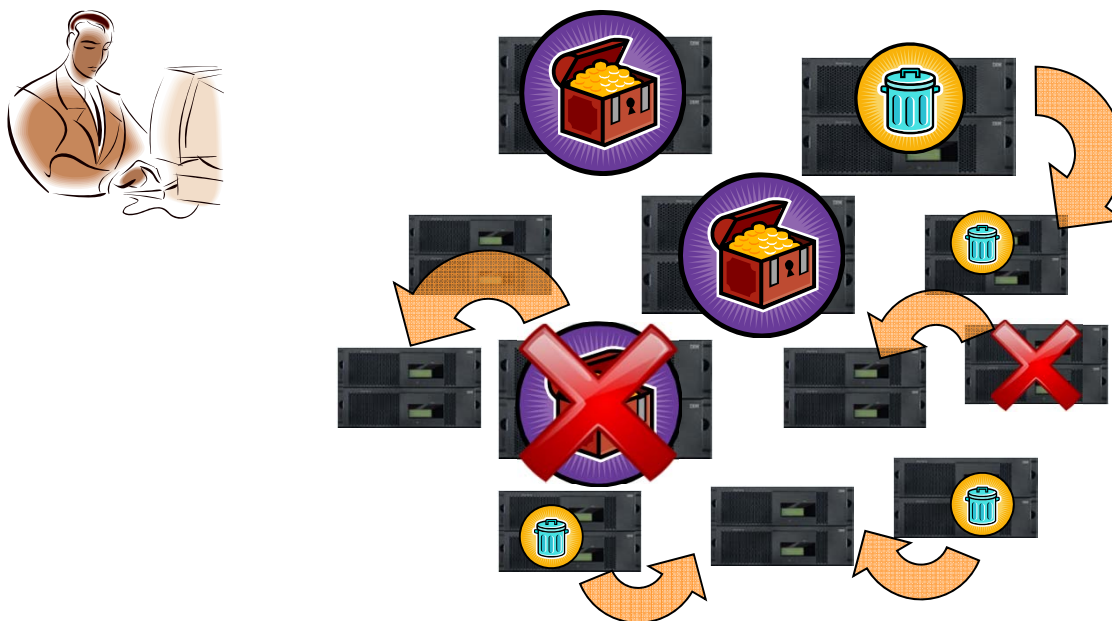


- **Tiered Storage:** Several GOLD projects have become obsolete; should turn back tier1 resources



## Challenge 3 : Manual Tiering gets cumbersome

- ...True for data lifecycle management (*how important?*)
- ... Also true for performance management (*maximize throughput*)

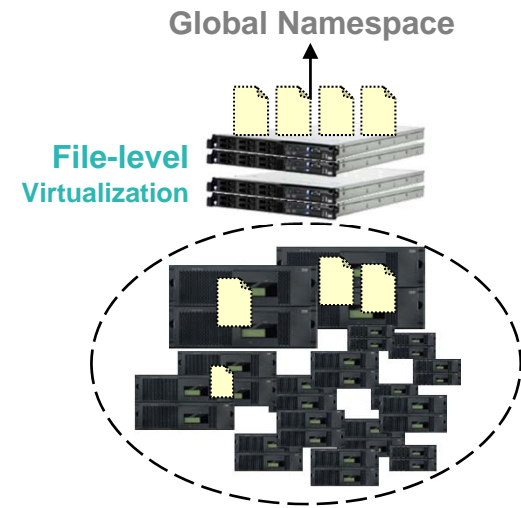


SONAS:  
Managing data **independently** from  
the amount of hardware involved

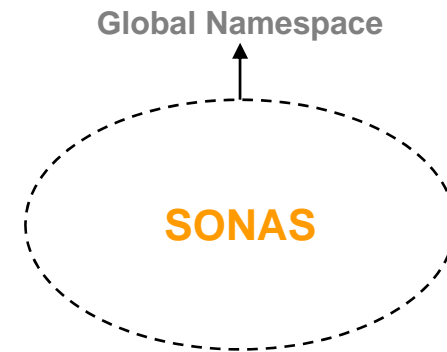


# Analogy with SAN Virtualization : Two "Religions"

## 1. External Virtualization = Flexibility



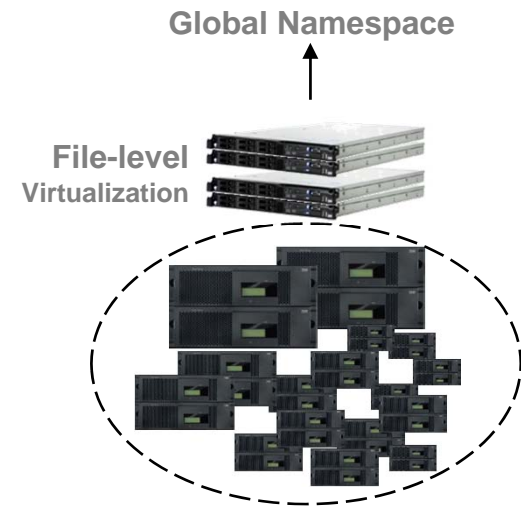
## 2. Internal "Grid" Virtualization = Availability



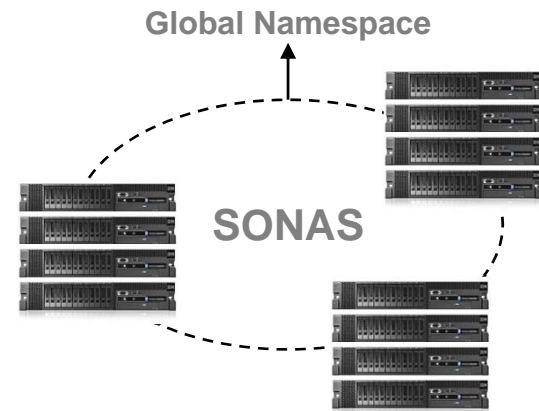


# Classic Design and "New Generation" Design

## 1. Redundant Component Pairs

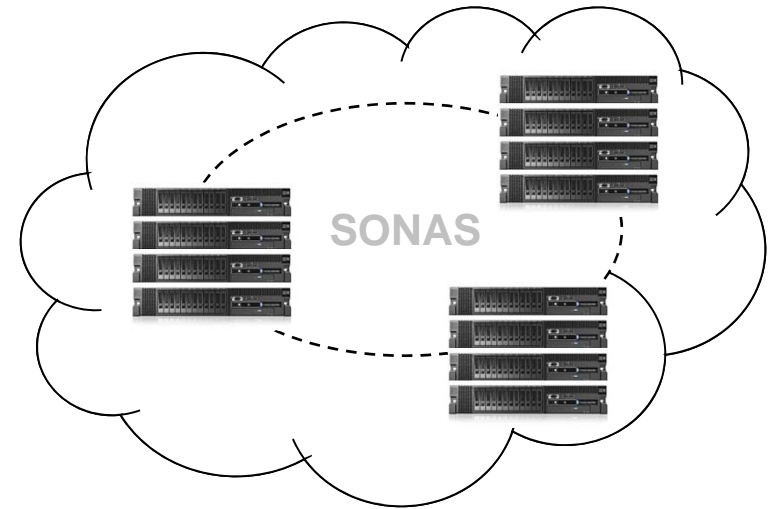
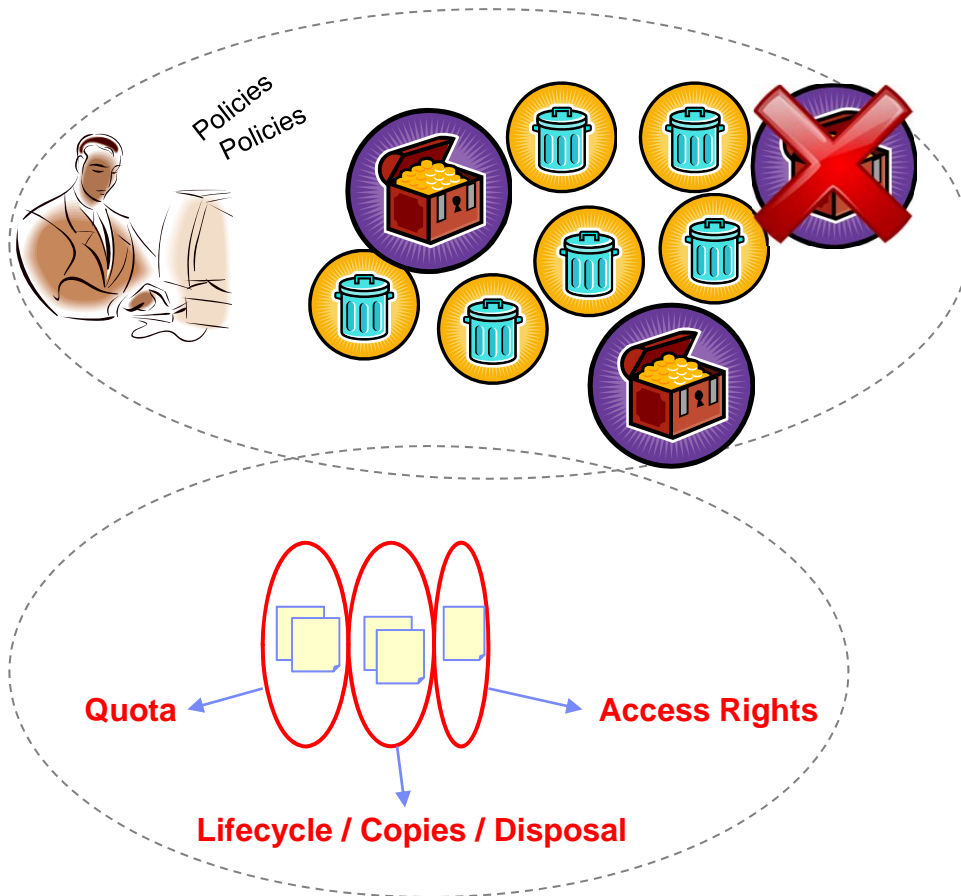


## 2. Distributed Services (no pairs)



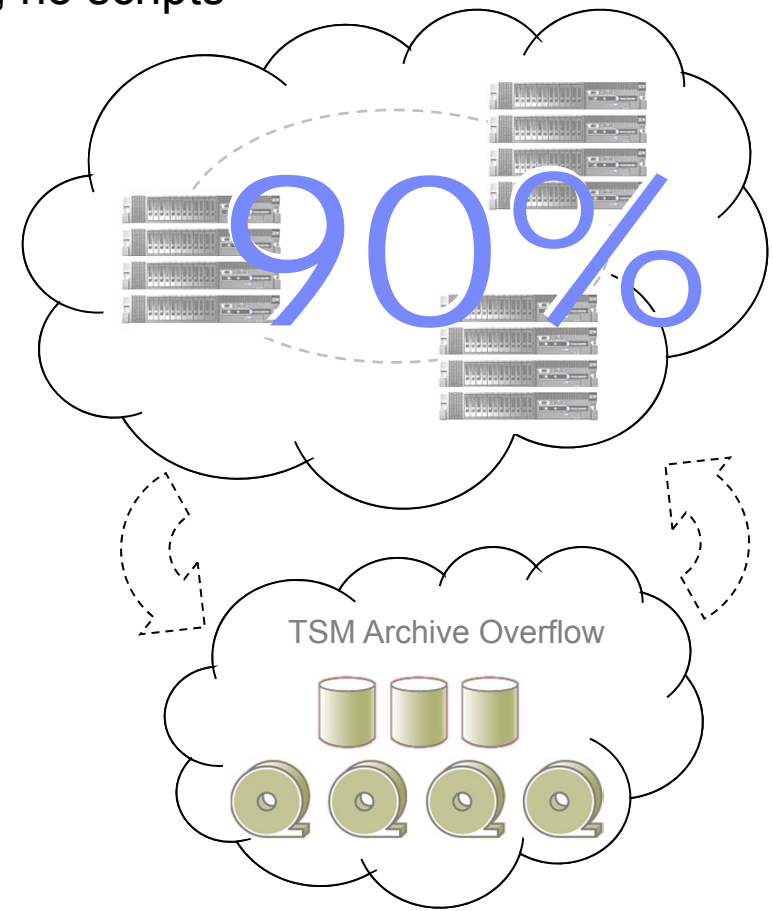
# SONAS: Manage Data independently from involved Hardware

- driven by content metadata policies (name, type, date, age, frequency of use...)
- infinitely scalable



# Use "Overflow" Policy to achieve 90% Fill Grade

- Maintain high fill grade on gold tier while minimizing risk
- Filesystem **Policy** Automation – no daemons, no scripts

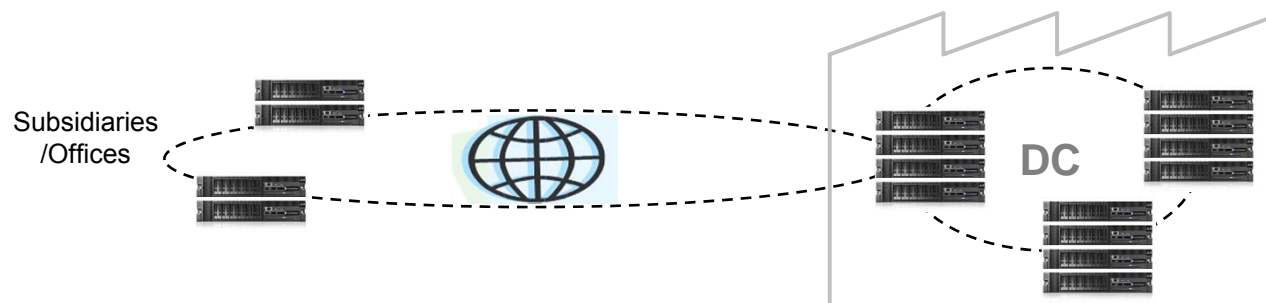


## Policy-based Overflow Management : Example

90%

- ① **Placement** policies, evaluated at file creation, example
  - rule rogersfiles set pool gold for fileset rogersfileset
  - rule otherfiles set pool silver
- ② **Migration** policies, evaluated periodically
  - rule cleangold migrate from pool gold threshold (90,85) to pool silver
  - rule cleansilver when day\_of\_week() = monday migrate from pool silver to pool pewter where access\_age > 30 days
- ③ **Deletion** policies, evaluated periodically
  - rule purgepewter when day\_of\_month() = 1 delete from pool pewter where access\_age > 365 days
- ④ **Exclusion** rules, exempts files from being chosen as candidates

## Policy-based Data Management: Geo-Distribution & Caching



- Reduce admin effort in subsidiaries – install empty SONAS caching instances
- Keep copies of subsidiaries data in the DC, or copies of DC data in subsidiaries
- Maintain coherency over long distance





# Scientific Background of GPFS Geo-Distribution Caching

## Panache: A Parallel File System Cache for Global File Access

Marc Eshel    Roger Haskin    Dean Hildebrand    Manoj Naik    Frank Schmuck  
Renu Tewari

*IBM Almaden Research*

{eshel, roger, manoj, schmuck}@almaden.ibm.com, {dhildeb, tewarir}@us.ibm.com

### Abstract

Cloud computing promises large-scale and seamless access to vast quantities of data across the globe. Applications will demand the reliability, consistency, and performance of a traditional cluster file system regardless of the physical distance between data centers.

Panache is a scalable, high-performance, clustered file system cache for parallel data-intensive applications that require wide area file access. Panache is the first file system cache to exploit parallelism in every aspect of its design—parallel applications can access and update the cache from multiple nodes while data and metadata is pulled into and pushed out of the cache in parallel. Data is cached and updated using pNFS, which performs parallel I/O between clients and servers, eliminating the single-server bottleneck of vanilla client-server file access protocols. Furthermore, Panache shields applications from fluctuating WAN latencies and outages and is easy to deploy as it relies on open standards for high-performance file serving and does not require any proprietary hardware or software to be installed at the remote cluster.

In this paper, we present the overall design and imple-

Traditionally, NFS (for Unix) and CIFS (for Windows) have been the protocols of choice for remote file serving. Originally designed for local area access, both are rather “chatty” and therefore unsuited for wide-area access. NFSv4 has numerous optimizations for wide-area use, but its scalability continues to suffer from the “single server” design. NFSv4.1, which includes pNFS, improves I/O performance by enabling parallel data transfers between clients and servers. Unfortunately, while NFSv4 and pNFS can improve network and I/O performance, they cannot completely mask WAN latencies nor operate during intermittent network outages.

As “storage cloud” architectures evolve from a single high bandwidth data-center towards a larger multi-tiered storage delivery architecture, e.g., Nirvanix SDN [7], file data needs to be efficiently moved across locations and be accessible using standard file system APIs. Moreover, for data-intensive applications to function seamlessly in “compute clouds”, the data needs to be cached closer to or at the site of the computation. Consider a typical multi-site compute cloud architecture that presents a virtualized environment to customer applications running at multiple sites within the cloud. Applications run inside a virtual machine (VM).

# GPFS has Field Experience

- Supercomputing basis technology

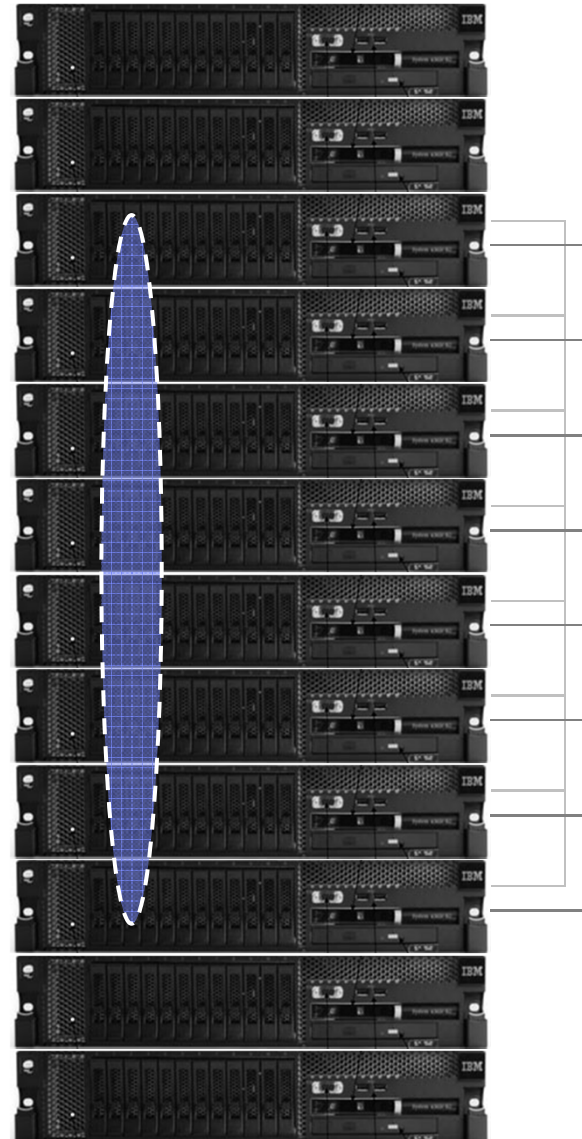


~50% of supercomputing capacity

- **Parallelization**  
= high scalability  
without Hotspots



- **Virtualization**  
= best yield

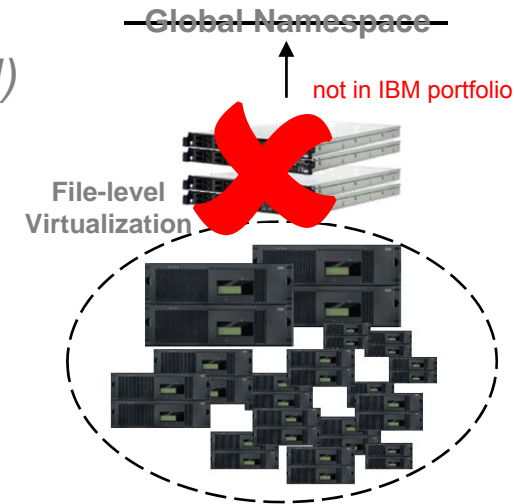




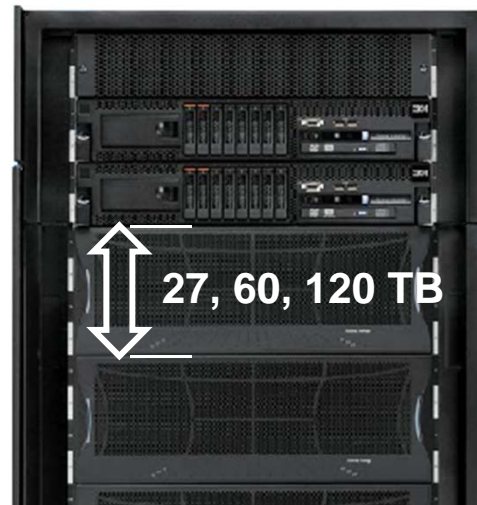
# SONAS and the N series Portfolio

# N Series (NetApp OEM) Cooperation has been reaffirmed

- **IBM N series** midrange and enterprise (*non-global*)



- **IBM SONAS** large/global enterprise



**Global Namespace** ✓

**high density:** 1 Rack = 960 TB\*

# SONAS Architecture

GPFS



SoFS



SONAS



# SONAS Architecture

**Solves the storage problem**

**3 basic components**

- ✓ **Interface Nodes = how fast**
- ✓ **Storage Pods = how big**
- ✓ **Management Node**

All nodes are clustered for availability

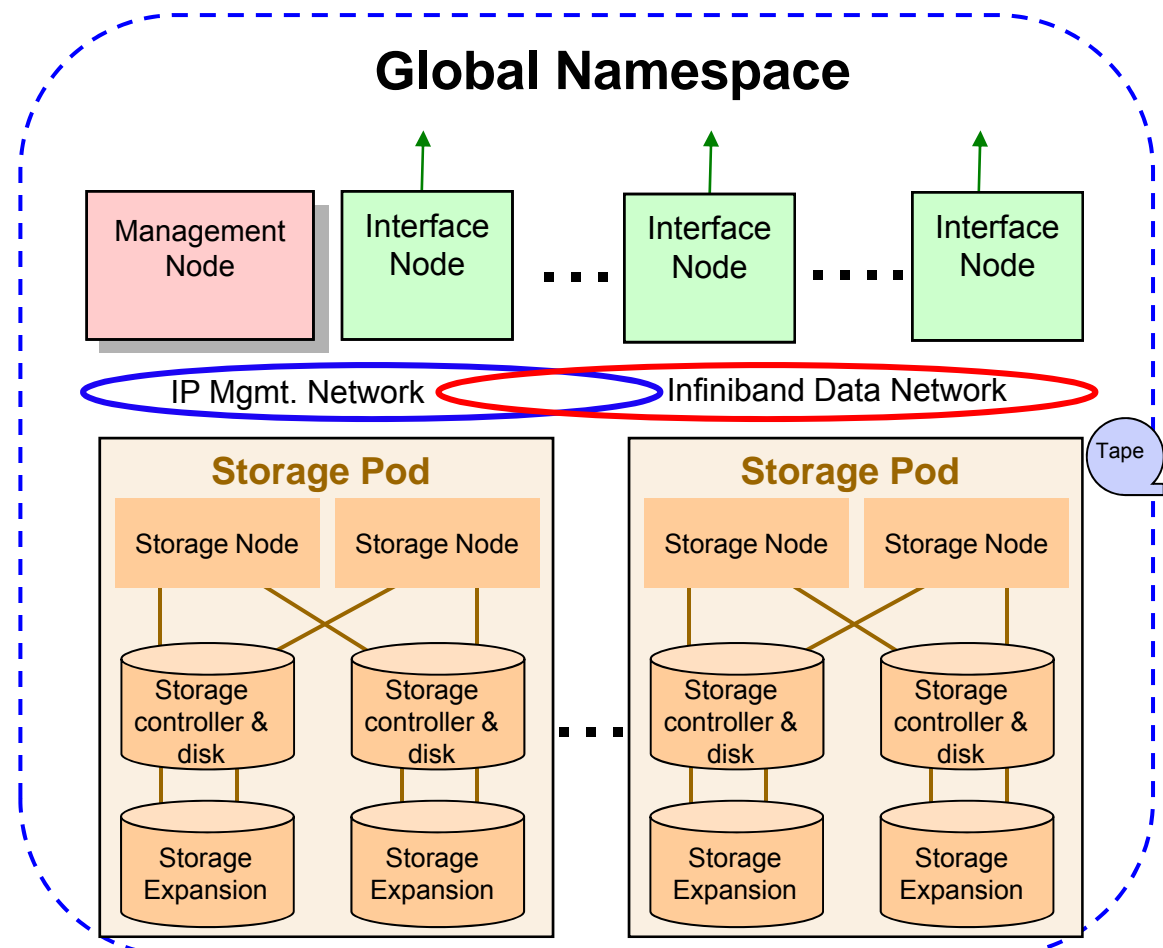
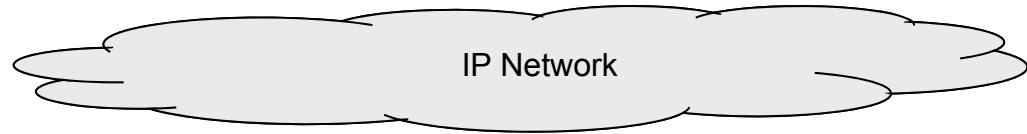
- Users connected through 1GbE or 10GbE
- All nodes are connected through private Infiniband network

## Parallel Grid Architecture

- Massive linear scalability
- High performance
- High availability & redundancy

## SONAS Software runs on all nodes

- Policy automation
- Global file system
- GUI and operating system

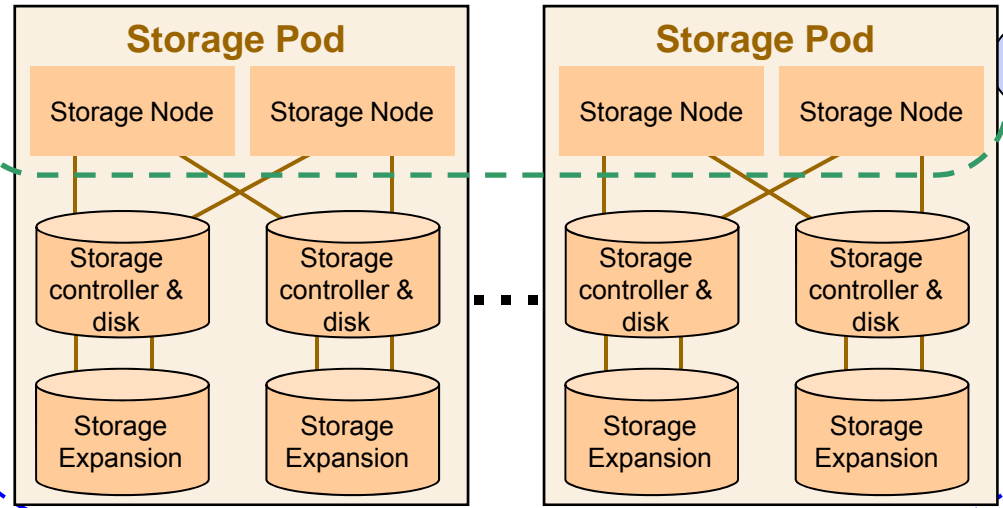
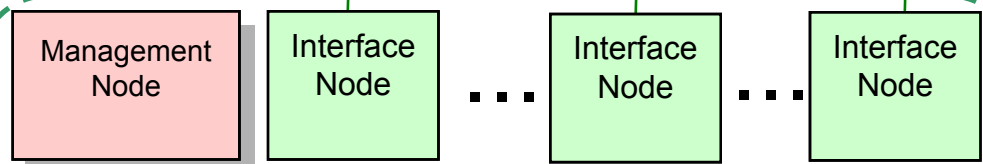


# SONAS Architecture



This is the internal GPFS parallel file system global cluster

## Global Namespace



External Storage

- All nodes share workload equally in global SONAS cluster
- Nodes provide high availability dynamic failover for each other
- Global cluster can be dynamically upgraded / changed by adding or removing nodes and/or storage, file system stays up while doing this
- Massive scalability: max interface nodes = 30, max storage nodes = 60

All interconnect components in SONAS global cluster are duplexed

# Building Blocks

2851-RXA



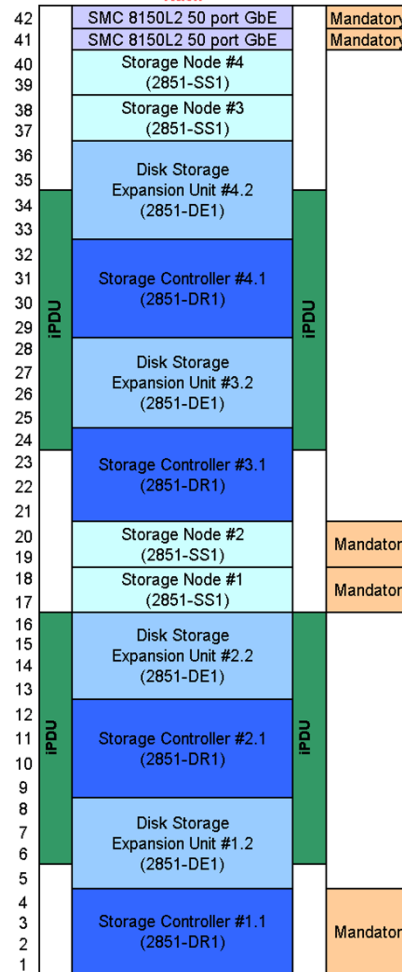
First Rack = Mixed type of interface nodes und storage

Extension Racks = either interface nodes or storage

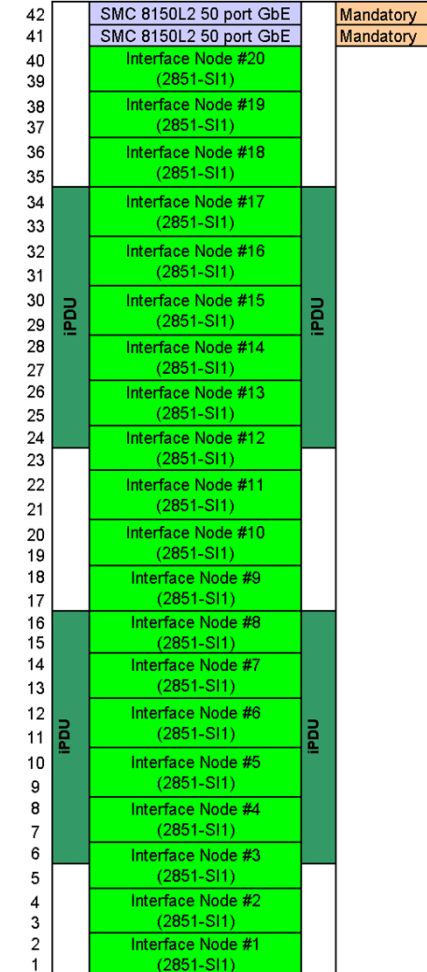
InfiniBand switch size in first rack limits the maximum non-disruptive scalability (in R1)

## Expansion Racks

**2851-RXB**  
SONAS Storage Expansion Rack



**2851-RXC**  
SONAS Interface Expansion Rack





## SONAS / GPFS Architectural Scalability

<b>Max. Capacity (in 2010)</b>	14,4 PB
<b>Max. GPFS Design Capacity</b>	134217728 Yobibytes ( $2^{107}$ Bytes)
<b>Max. Files × File systems × Snaps</b>	2 billion ( $2^{31}$ ) × 256 × 256
<b>Max. single File Size</b>	16 Exibytes ( $2^{64}$ Bytes)

Lawrence Livermore National Lab



126GB/s *single file* write performance

# What SONAS architecture (GPFS) is capable of

## File system requirements

“Balanced” capacity and performance

(100 PB file system, 6 TB/s file I/O)

Reliability in the presence of localized failures

Support for full-up system (~64K nodes)

One trillion files to a single file system

32K file creates per second

Streaming I/O at 30GB/s full duplex (for data capture)

**5 Years of iTunes music  
in 32 minutes**

**1 PB of metadata**

## Storage system future requirements

Geographically Dispersed Storage

Solve issues of very high reliability at PB level

2-fault-tolerant RAID-6 (8+2P): MTTDL about 100 years

Assuming max PERCS system, 600-GB SAS drives with  
MTTF=600khrs, hard error rate = 1E-15, and uncorrelated failures,  
MTTDL ~ 100 years

3-fault-tolerant RAID (8+3P): MTTDL is 6 orders of magnitude  
better

Same assumptions above, MTTDL ~ 130 million years

Additional 10% overhead

This is what IBM  
GPFS is capable  
of today

[Click here for  
more GPFS info](#)

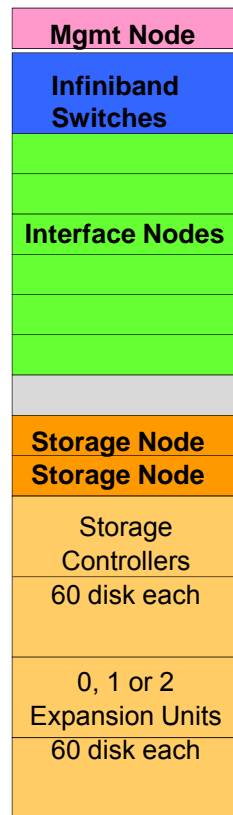
IBM GPFS is  
foundation for future  
capabilities

# IBM Cloud Storage - SONAS R1.2

# IBM Scale Out Network Attached Storage – Appliance

## Fully Integrated Appliance

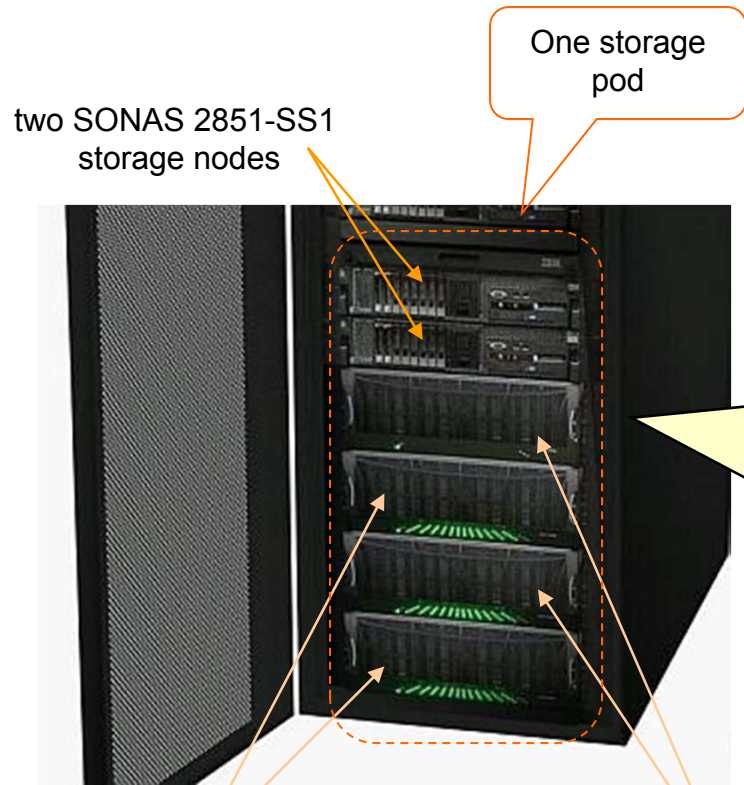
Scale Out Network Attached Storage (NAS) system with differentiated, world-proven General Parallel File System



- Global file system namespace
- File system snapshots, quotas, async replication
- Automated Tiered Storage
  - Integrated Information Lifecycle Management (ILM)\*
  - HSM support, using external Tivoli Storage Manager (TSM) server
- Integrated System Health Center for HW monitoring
- World-class scale and performance
  - High performance scan engine

Today, in IBM-supplied SONAS rack.  
Customer-supplied rack is a known requirement

# SONAS Integrated Storage



Disk Drives available today are:  
600GB SAS, 15K RPM, RAID-6\*  
or  
2TB Nearline SAS, 7.2K RPM, RAID-6

60-drive drawer must be all SAS or all Nearline SAS

Can intermix SAS and Nearline SAS drawers  
in a storage pod

Usable storage = conservative average is 76% of raw storage

two SONAS 2851-DR1  
storage controllers

two SONAS 2851-DE1  
disk expansion units

## SONAS Storage Pod

Photograph of high density advanced SONAS Storage Pod



- **Fast**
  - 800 MB/sec sequential for this drawer if SAS drives
  - 3200 MB/sec sequential for full storage pod (4 drawers)
- **Ultra-Dense**
  - 60 Drives in just 4U
- **Highly Reliable**
  - Active/Active Failover
  - New: default is RAID 6 for all drives
  - RAID-5 is RPQ option for SAS drives
  - Redundancy Throughout
  - Battery Backed Cache

# IBM Scale Out Network Attached Storage Expansion Racks

## Node Expansion

- Up to 30 interface nodes
- High speed extremely low latency (20 Gbps) private Infiniband cluster data network

## Storage Expansion

- Up to 30 storage pods, each with up to 240 HDD's
- In single system using 2 TB drives, scalable to 14.4 PB (decimal) of raw storage
  - Equal to 10.6 PB usable (decimal)
- Max of 7,200 hard drives

Switches
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node
Interface Node

Interface node  
Expansion rack

Switches
Storage Node
Storage Node
60 Disks
60 Disks
60 Disks
60 Disks
Storage Node
Storage Node
60 Disks
60 Disks
60 Disks
60 Disks
60 Disks

Storage Expansion  
Rack



# IBM SONAS Scale Out NAS expansion example - middle

Switches	Switches	Switches	Switches	Switches	Switches	Switches	Switches	Switches
Interface Node	Mgmt Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node
Interface Node		Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node
Interface Node	Infiniband switch	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node		60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Infiniband switch	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node		60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Monitor/KVM	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Interface Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node
Interface Node	Interface Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node	Storage Node
Interface Node	Interface Node	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Interface Node	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Interface Node	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Interface Node	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Interface Node	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks
Interface Node	Interface Node	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks	60 Disks

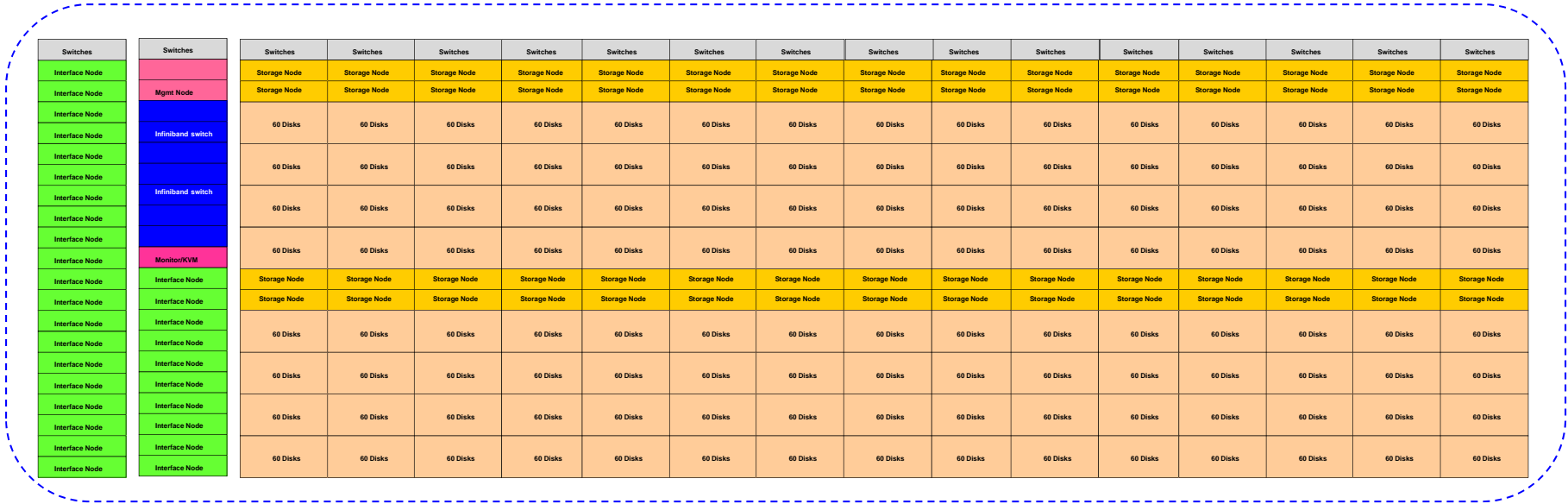
**30 interface nodes, 7 disk expansion racks, 3360 disks**

**With 2TB Nearline SAS drives, this is:**

**6.7 PB raw, 5.1 PB usable (decimal)**

# IBM SONAS Scale Out NAS expansion example - maximum

Can be single file system, or up to 256 file systems



**30 interface nodes, 15 disk expansion racks, 7200 disks**

**With 2TB Nearline SAS drives, this is:  
14.4 PB raw, 10.6 PB usable (decimal)**

Only reason we stop here is because we used up all of the ports on 96-port Infiniband switch

# IBM SONAS Node Components — x3650M2

## Interface Node



The interface node provides the connections to the customer's IP network for attaching to the SONAS system for network file serving capabilities (CIFS, NFS, HTTP, FTP, SCP).

- **Processor:**
  - Dual Quad Core Intel® Xeon® X5530
  - 2.26GHz, 8MB L2 cache, 80W
- **Memory:**
  - 32GB, 64GB, 128GB DDR3 RAM
- **Storage:**
  - 2 – 300G SAS 10K (1-RAID1 pair)
- **Network Interfaces:**
  - 1 GbE or 10 GbE
  - For customer network

## Management Node



The management node provides the user interface for configuring, administering and monitoring the SONAS system.

- **Processor:**
  - Dual Quad Core Intel® Xeon® X5530
  - 2.26GHz, 8MB L2 cache, 80W
- **Memory:**
  - 32GB DDR3 RAM
- **Storage:**
  - 2 – 300G SAS 10K (1-RAID1 pair)
  - 1 – 300G SAS 10K (log/trace collection)
- **Network Interfaces:**
  - GbE
    - 2 Customer Network
    - 2 Management Network

## Storage Node



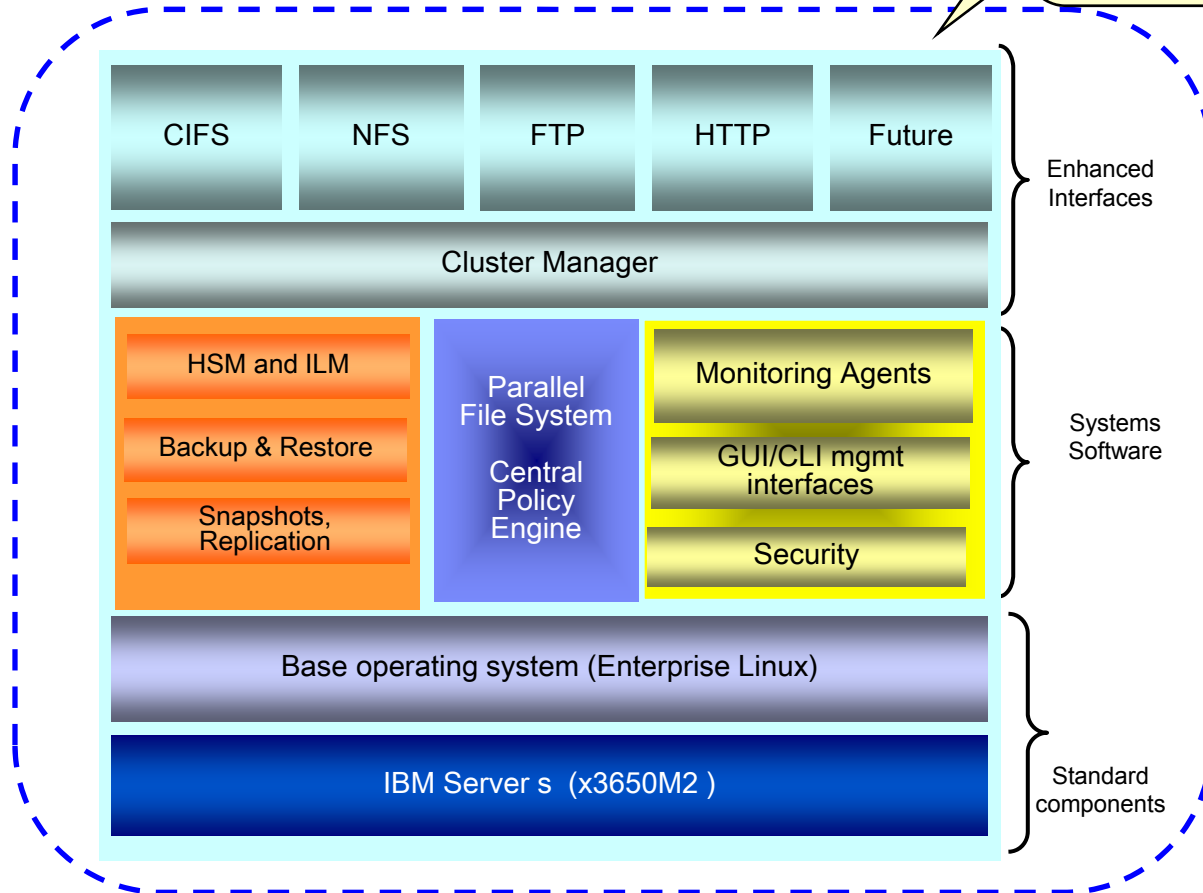
The storage node provides the Infiniband connection to the InfiniBand cluster interconnect and direct fibre-channel attachment to the SONAS RAID controller. Storage nodes must be configured in High-availability pairs.

- **Processor:**
  - Dual Quad Core Intel® Xeon® X5530
  - 2.26GHz, 8MB L2 cache, 80W
- **Memory:**
  - 8GB DDR3 RAM
- **Storage:**
  - 2 – 300G SAS 10K (1-RAID1 pair)
- **Network Interfaces:**
  - Four 1 Gbps NICs
    - 2 Management Network
  - Two single port 4X DDR Infiniband Host Channel Adapters
  - Two dual-port 8Gbps Fibre Channel Host Bus Adapters (HBA)

# SONAS Software

One software product supported by IBM

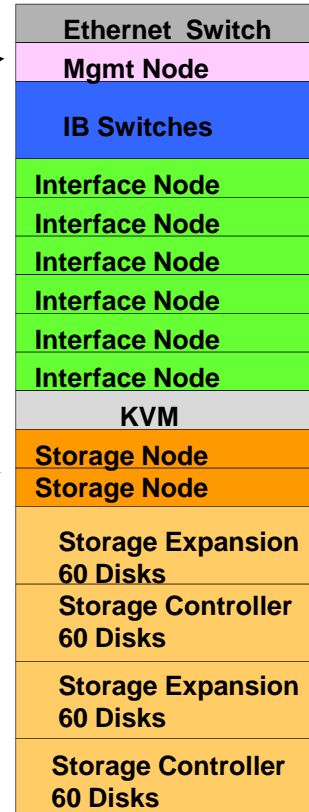
IBM SONAS Software (5639-SN1)  
One copy per node



# Case Study SONAS Configuration

SONAS starter configuration (480TB Nearline SAS)

- Management Node
- Infiniband Switches
- Six interface nodes
- Two Storage Nodes
- Two Storage controllers
- Two Disk Storage Expansions

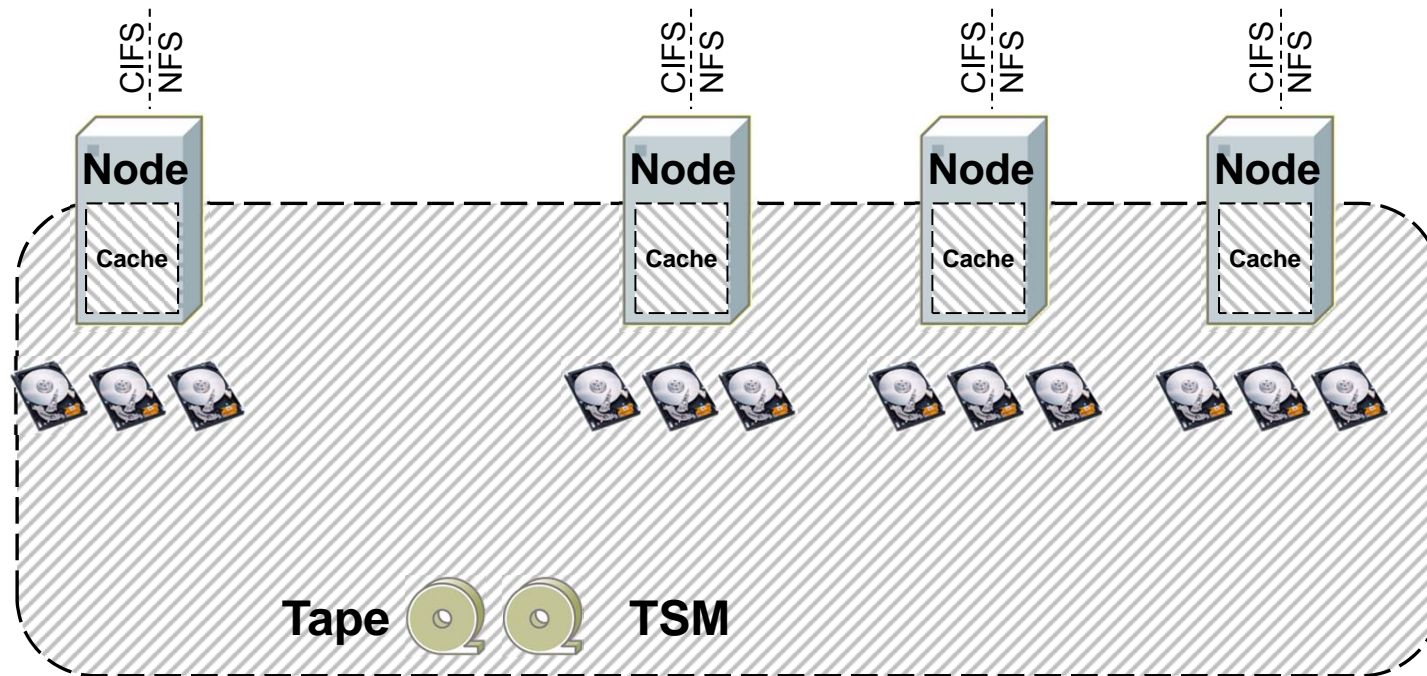


# SONAS Storage Continuum

# Cache, Disks and Tape : *One Storage Continuum*

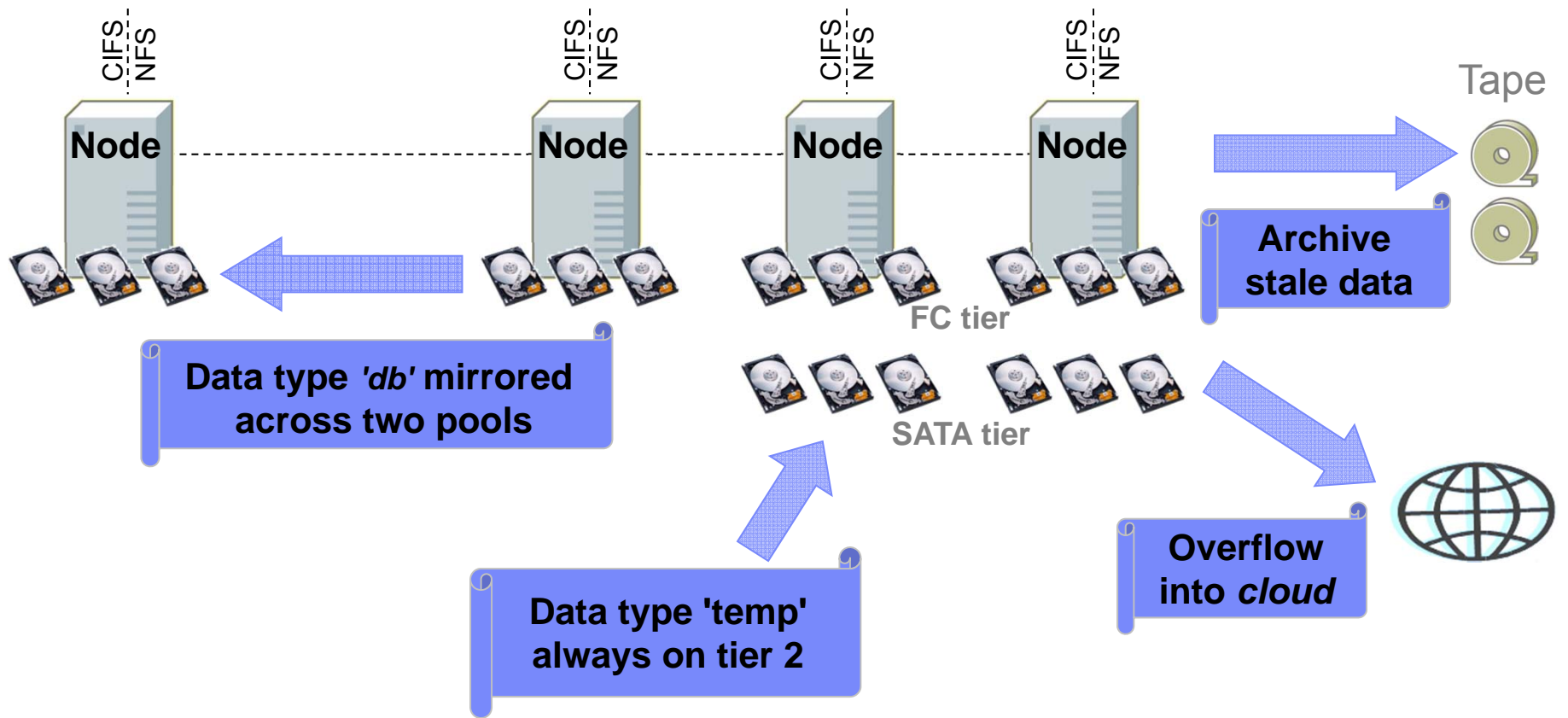
- Parallel Windows® access
- Uniform Windows/Unix view
- Differential Snapshots
- Wide Data Striping
- Quick Restore**

The whole storage stack including cache (RAM) and backup/archive (tape) is integrally managed.



# SONAS Automated Lifecycle Management

More Examples – Policy-based, transparent for the user

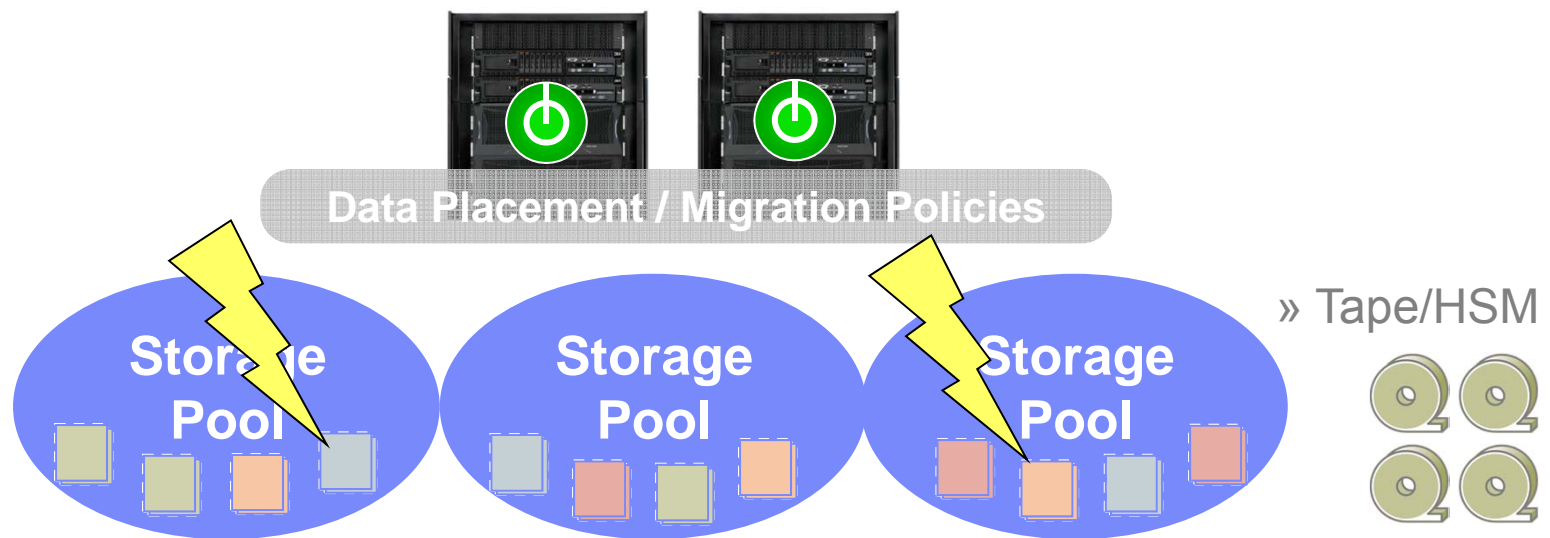




# Quick Restore from HSM in worst case scenario (R1.1.1)

up to the Petabyte capacity

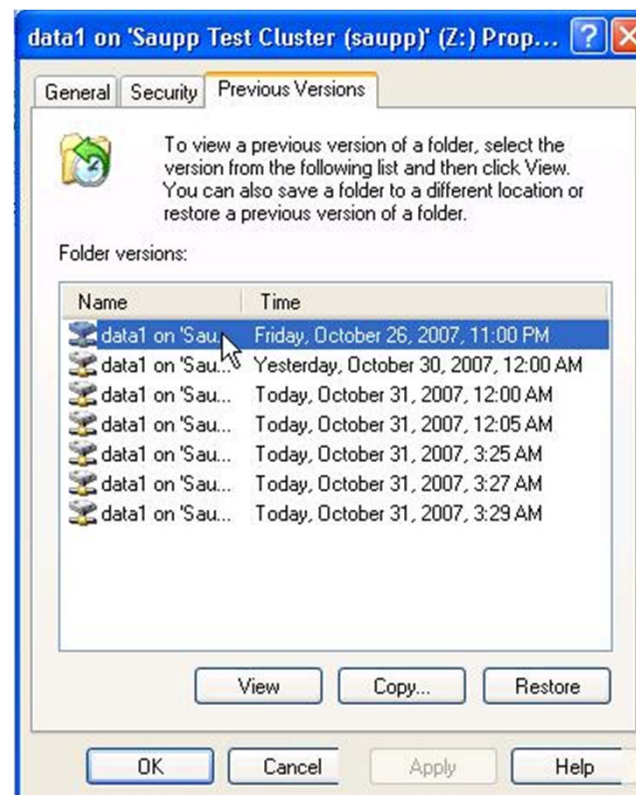
- **Lifecycle Management** connects all storage tiers, including tape
- Fast TSM backup: Scanning all changed files takes mere seconds
- Quick Restore: Metadata first, active data second, (pre)migrated data on first request



# SONAS Data Protection

## Data Protection/High Availability Features

- **Snapshots**
  - Space efficient, differential snapshots
  - Includes Microsoft Windows VSS integration
- **Synchronous Replication**
  - File, Set of files or entire file system
  - Single site in initial release
- **Asynchronous Replication (Release 1.1.1)**
  - Any file system sub-tree from one cluster to another
  - Batched based, hub and spoke
- **High Availability (HA) Features**
  - Redundant Interface Nodes: allows access to data by users
  - Redundant Storage Nodes: allows access to storage
  - Redundant private 1GbE internal management network
  - Redundant private DDR Infiniband data network
  - RAID 5/6: Protects against individual disk failures

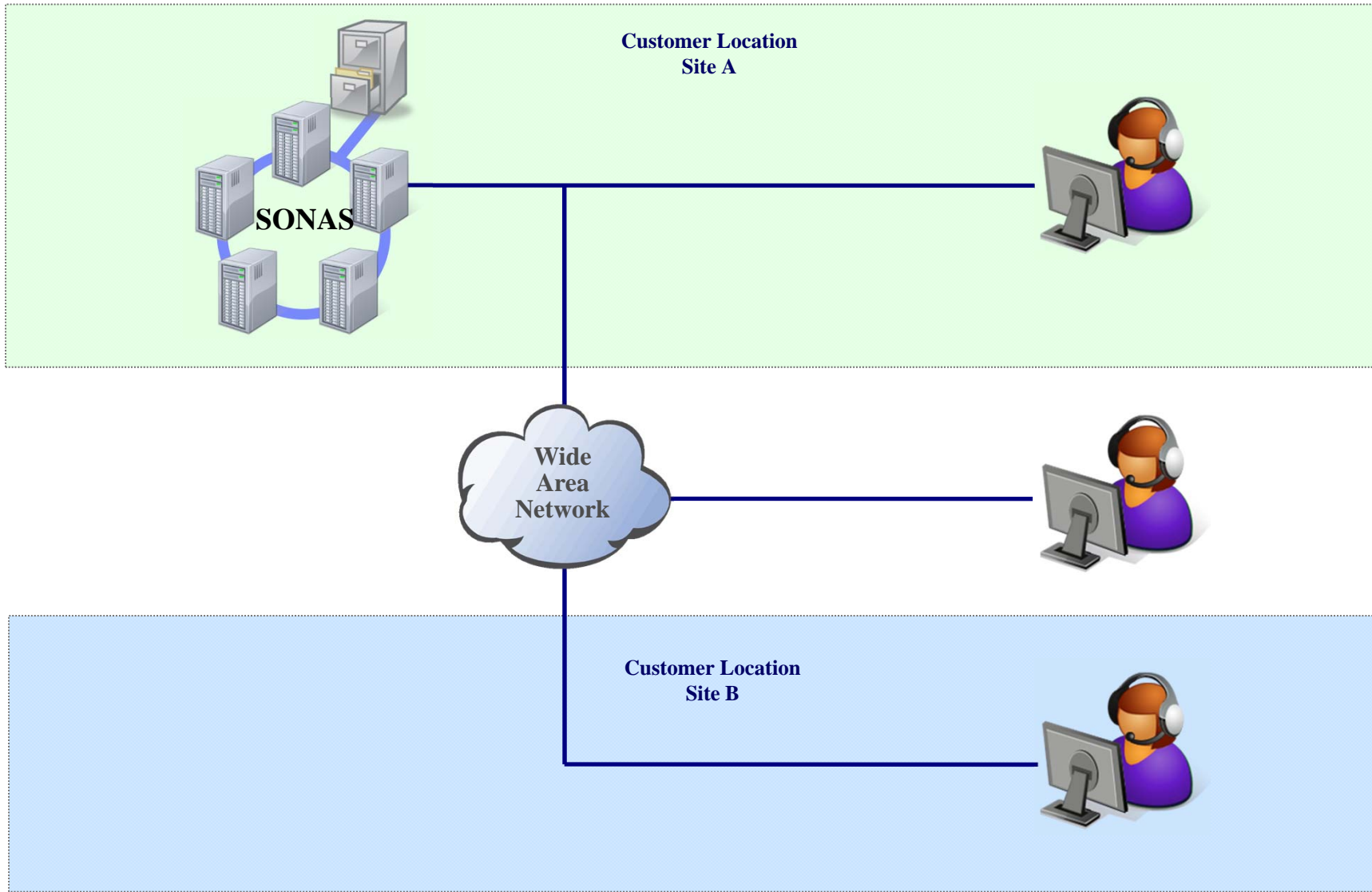


**Snapshots** Integrated into Windows Explorer using the Volume Shadow Copy Services (VSS)

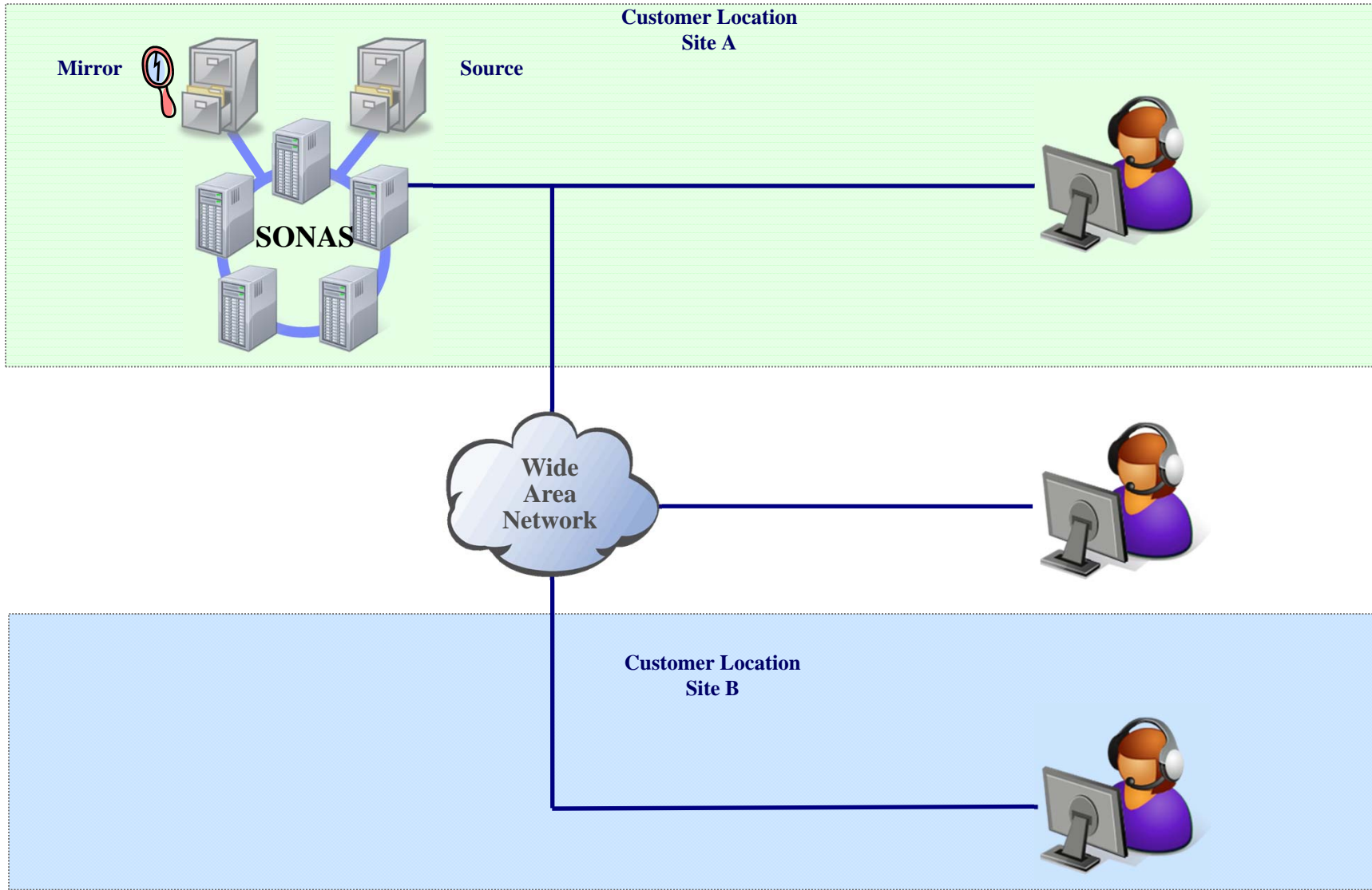
**Snapshots** in Linux/Unix → `/.snapshots`

# SONAS Remote Mirroring

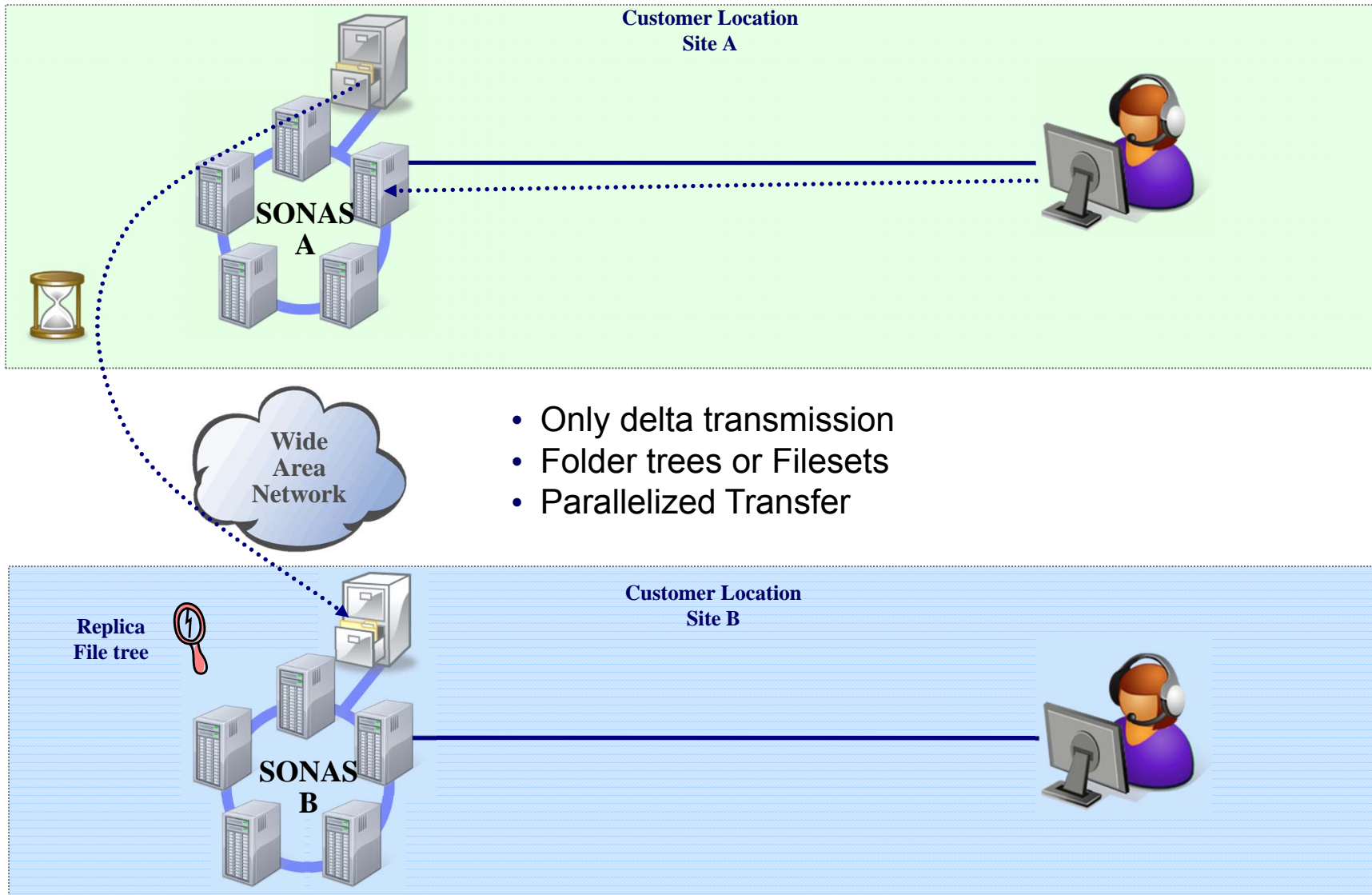
# Local SONAS Installation



# SONAS with locally mirrored data pools

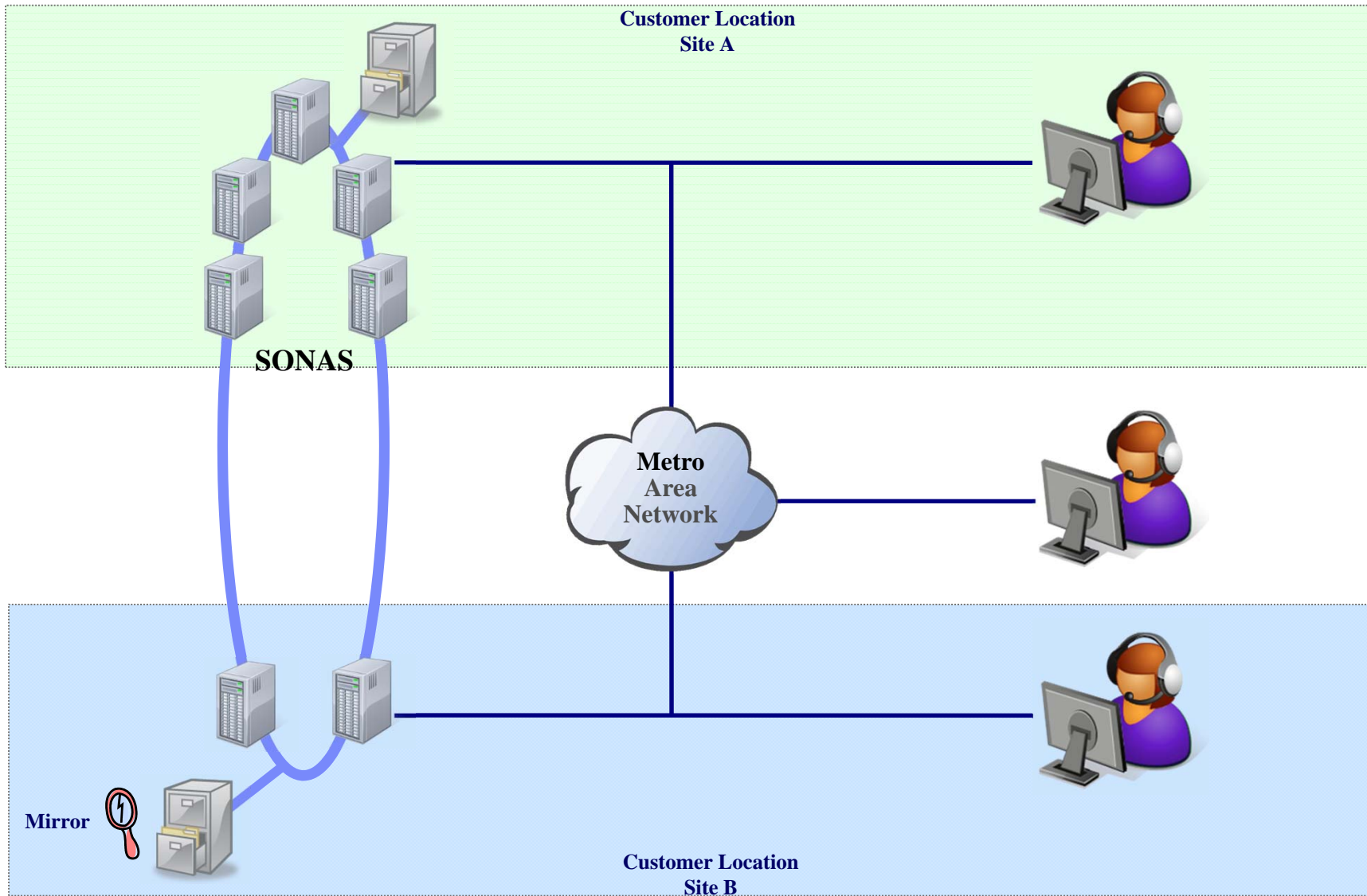


# SONAS with asynchronous Mirror (R1.5)

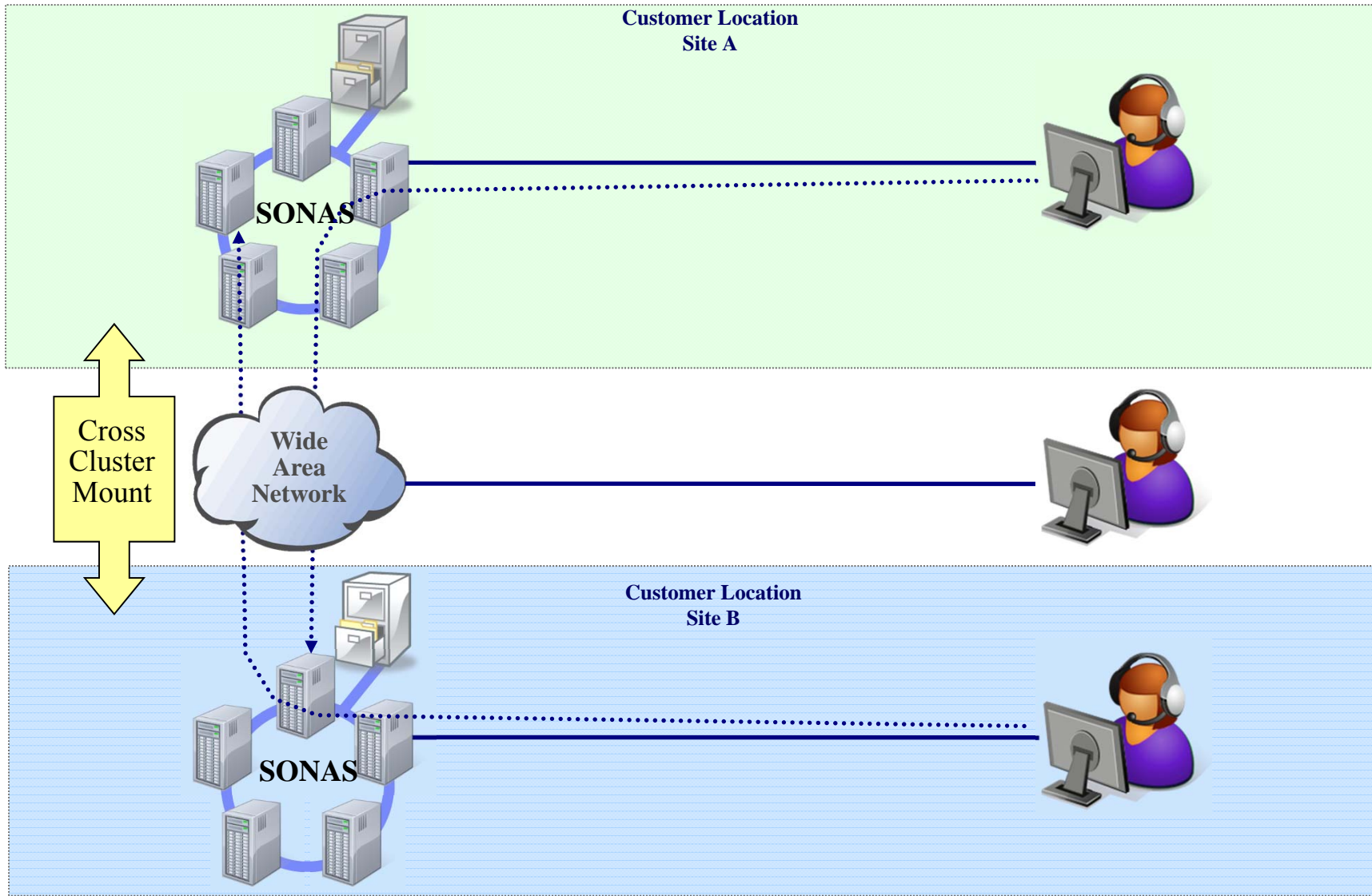




# SONAS across campus distance (R1.x)



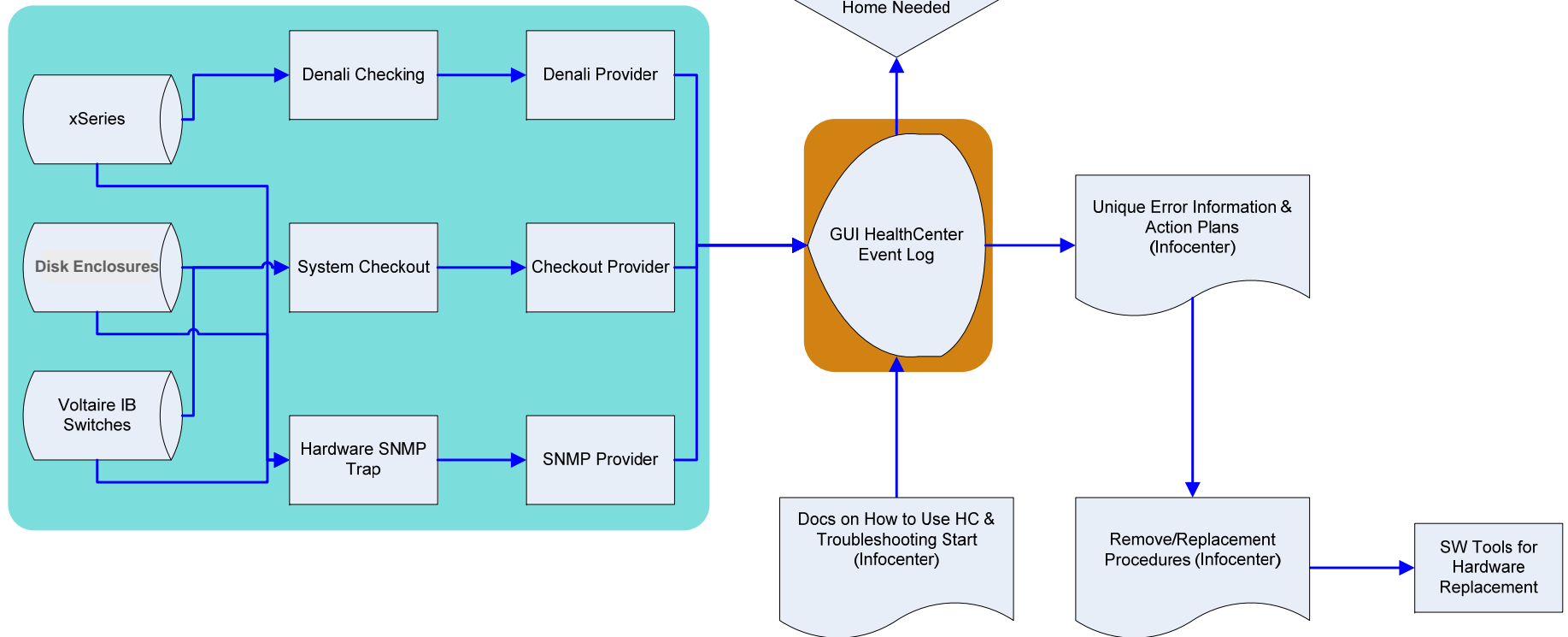
# SONAS with long-distance cross-cluster mount



# SONAS RAS and health check

# SONAS Integrity Check with Call-Home

Every status information or error message is collected in **one log**.



NOTE: This diagram represents end to end error flow within a cluster including documentation.

# Preventive Check of Structural Integrity (*checkout* process)

```

Check Status File: /opt/IBM/sonas/ras/config/rsSnScStatusComponent.xml

Summary of NON-OK Statuses:
Warnings: 1
Degrades: 2
Failures: 0
Offlines: 0

Ethernet Switch status:
Verify Ethernet Switch Configuration (Frame:1, Slot:41) OK
Verify Ethernet Switch Hardware (Frame:1, Slot:41) OK
Verify Ethernet Switch Firmware (Frame:1, Slot:41) OK
Verify Ethernet Switch Link (Frame:1, Slot:41) DEGRADED*

*DEGRADED: 80000840 8 links on this switch is up but 9 links should be on this switch.

Verify Ethernet Switch Configuration (Frame:1, Slot:42) OK
Verify Ethernet Switch Hardware (Frame:1, Slot:42) OK
Verify Ethernet Switch Firmware (Frame:1, Slot:42) OK
Verify Ethernet Switch Link (Frame:1, Slot:42) OK

InfiniBand Switch status:
Verify InfiniBand Switch Configuration (Frame:1, Slot:35) OK
Verify InfiniBand Switch Hardware (Frame:1, Slot:35) OK
Verify InfiniBand Switch Firmware (Frame:1, Slot:35) OK
Verify InfiniBand Switch Link (Frame:1, Slot:35) DEGRADED*

*DEGRADED: A00A0A00 7 links on this switch is up but 8 links should be on this switch.

Verify InfiniBand Switch Configuration (Frame:1, Slot:36) OK
Verify InfiniBand Switch Hardware (Frame:1, Slot:36) OK
Verify InfiniBand Switch Firmware (Frame:1, Slot:36) OK
Verify InfiniBand Switch Link (Frame:1, Slot:36) OK

Node status:
Verify Node General WARNING*

*WARNING: 2010A002 Invalid IP address 'N/A' was detected for the RSA IP of int001st001 (Frame:1, Slot:23).
*WARNING: 2020A002 Invalid IP address 'N/A' was detected for the RSA IP of int002st001 (Frame:1, Slot:25).
*WARNING: 2030A002 Invalid IP address 'N/A' was detected for the RSA IP of int003st001 (Frame:1, Slot:27).
    
```

```

Verify Disk Enclosure Hardware (Frame:1, Slot:1) OK
Verify Disk Enclosure Firmware (Frame:1, Slot:1) OK
Verify Array in Disk Enclosure (Frame:1, Slot:1) FAILED*

*FAILED: 60150014 Array OID=0x188a001d in DDN controller 1 is in state 'DEGRADED'

FibreChannel HBA status:
Verify Fibre Channel HBA Configuration (Frame:1, Slot:19, Instance:0) OK
Verify Fibre Channel HBA Firmware (Frame:1, Slot:19, Instance:0) OK
Verify Fibre Channel HBA Link (Frame:1, Slot:19, Instance:0) OK
Verify Fibre Channel HBA Configuration (Frame:1, Slot:19, Instance:1) OK
Verify Fibre Channel HBA Firmware (Frame:1, Slot:19, Instance:1) OK
Verify Fibre Channel HBA Link (Frame:1, Slot:19, Instance:1) OK

strg002st001

[root@RASborg.mgmt001st001 ~]# cnrsshc --status

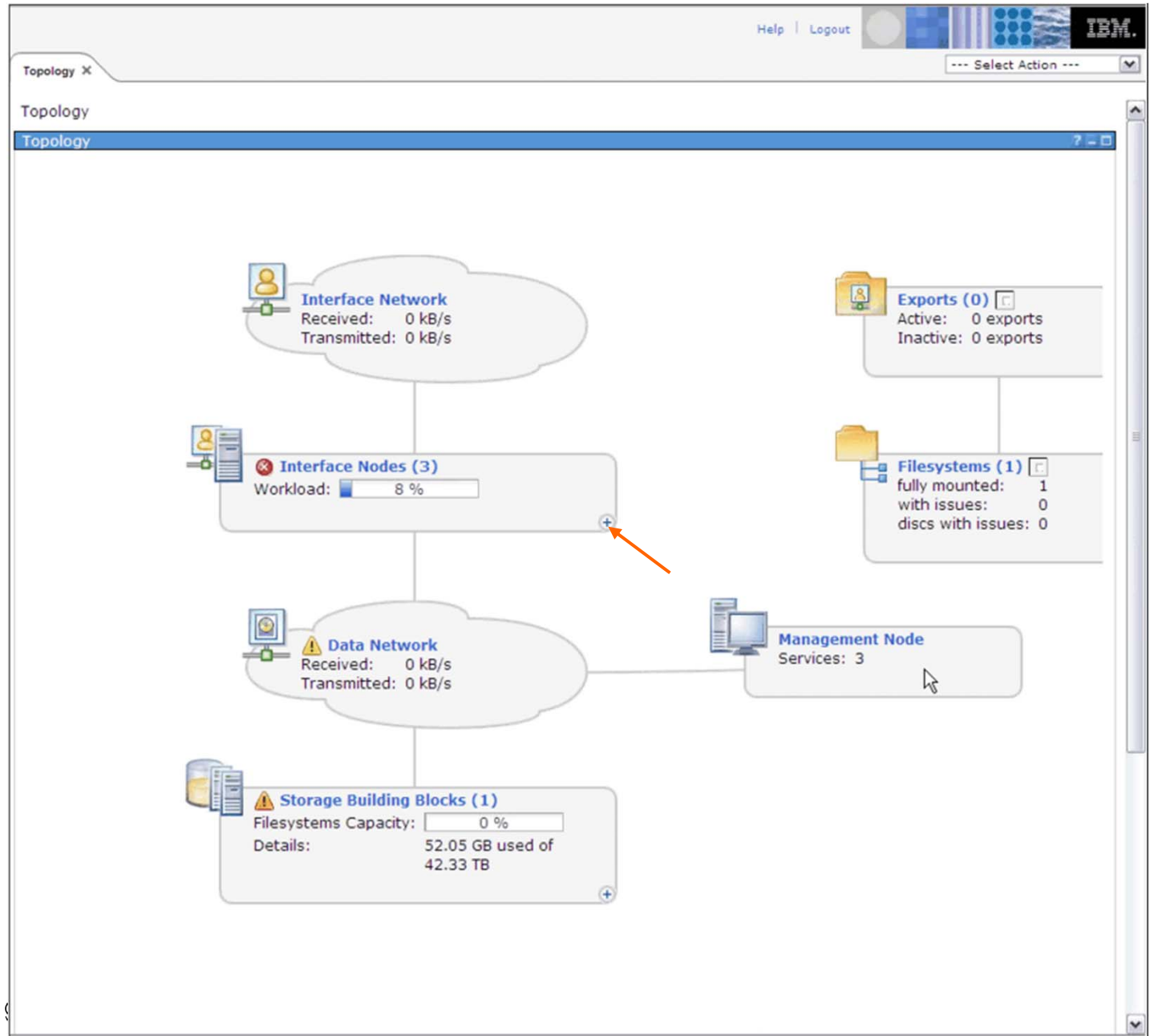
Status on mgmt001st001:
Verify sofsGUI is running: OK
Verify cimserverd is running: OK
Verify status of CIM provider modules: OK
Verify CIM indication from Director is subscribed: OK
Verify CIM indication from IBMNAS is subscribed: OK
Verify CIM indication from LSI is subscribed: OK

Status on int001st001:
Verify cimserverd is running: OK
Verify status of CIM provider modules: OK
Verify CIM indication from Director is subscribed: OK
Verify CIM indication from IBMNAS is subscribed: OK
Verify CIM indication from LSI is subscribed: OK

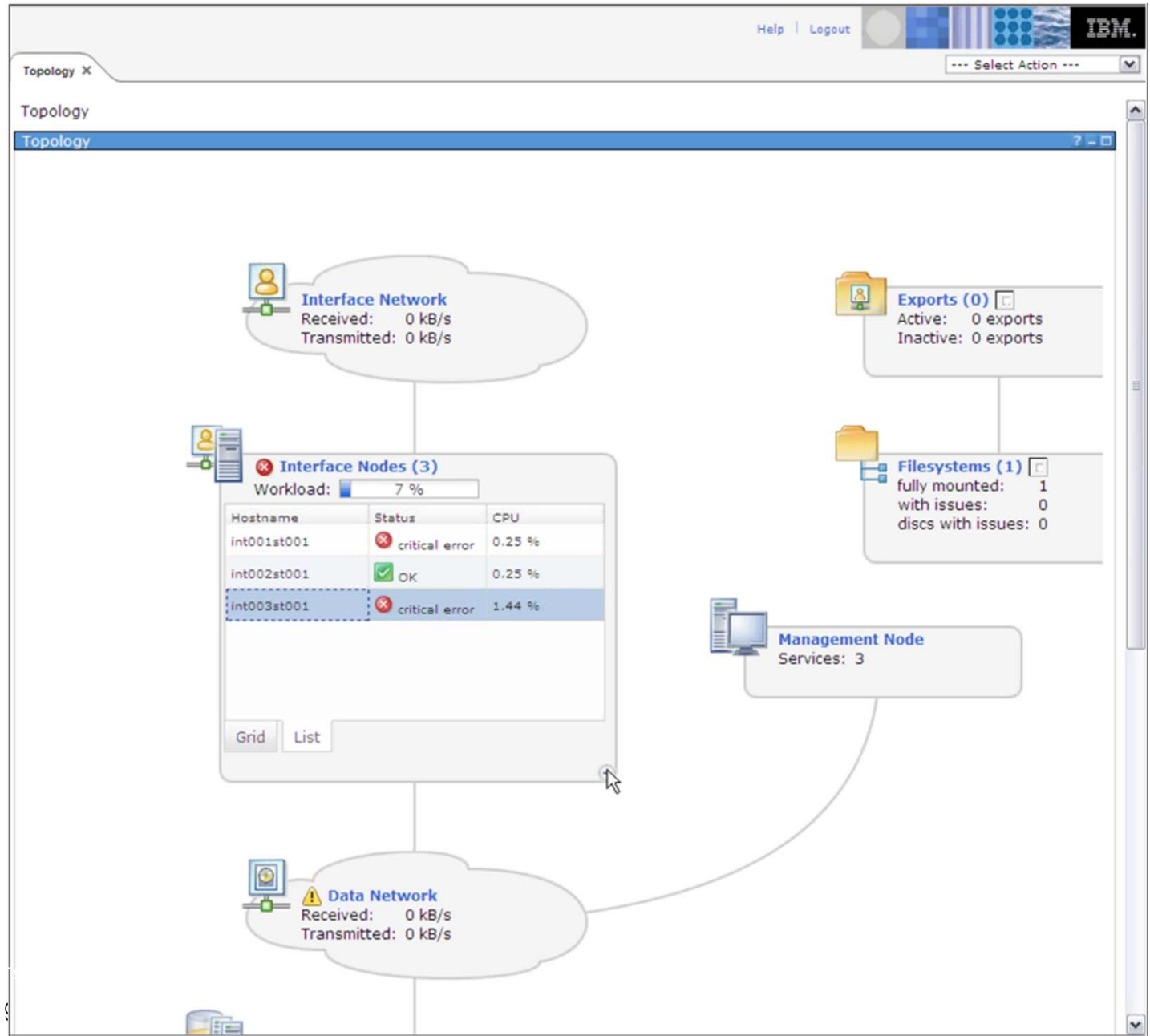
Status on int002st001:
Verify cimserverd is running: OK
Verify status of CIM provider modules: OK
Verify CIM indication from Director is subscribed: OK
Verify CIM indication from IBMNAS is subscribed: OK
Verify CIM indication from LSI is subscribed: OK

Status on int003st001:
Verify cimserverd is running: OK
Verify status of CIM provider modules: OK
    
```

# SONAS Health Center

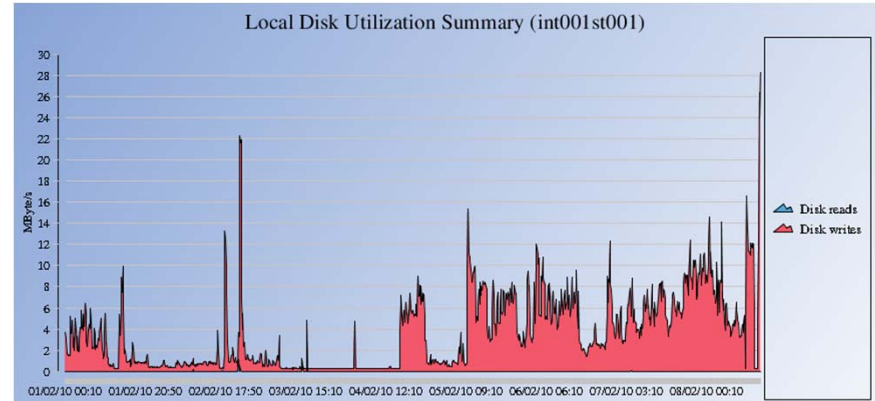
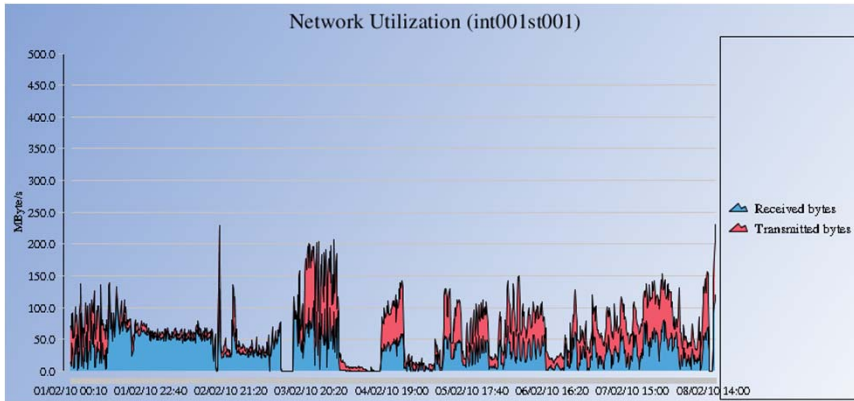
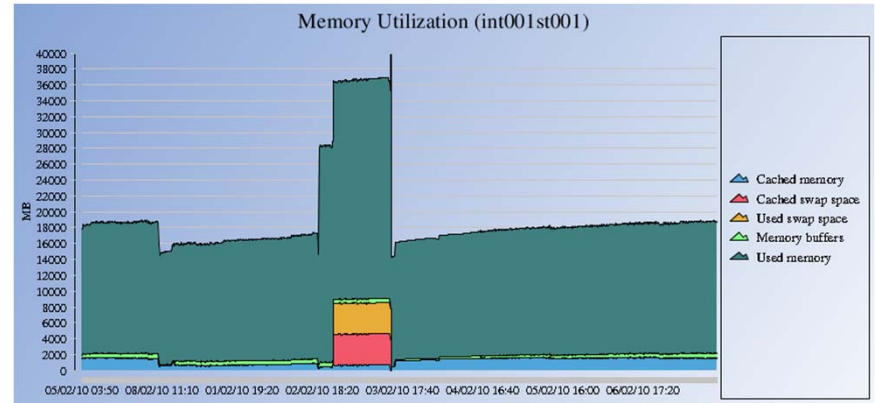
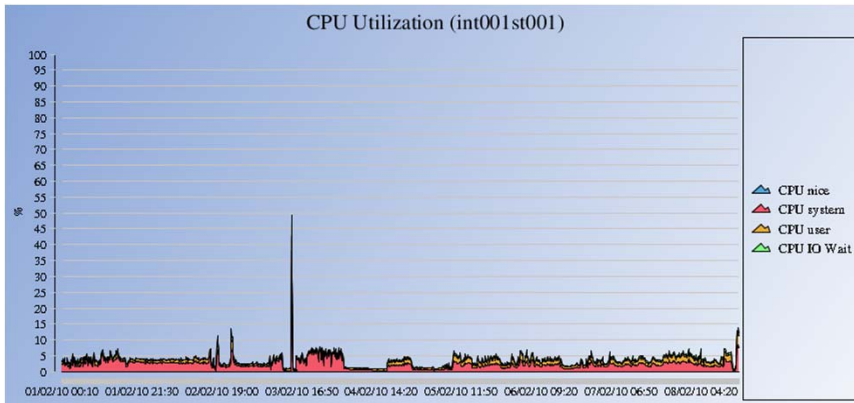


# SONAS Health Center





# SONAS System Utilization Monitor



# SONAS Scalability & Summary

# SONAS SPECsfs® Performance Benchmark



Feb 22, 2011: SONAS = Maximum Throughput:

403,000 IOPS (\*)

Sets a new World Record for performance per file system, based on the SPECsfs benchmark

See: <http://www.spec.org/sfs2008/results/sfs2008.html>

What makes this SPECsfs benchmark special is that it proves SONAS provides true scale out by combining:

**Capacity and a single file system and Leadership in Performance**

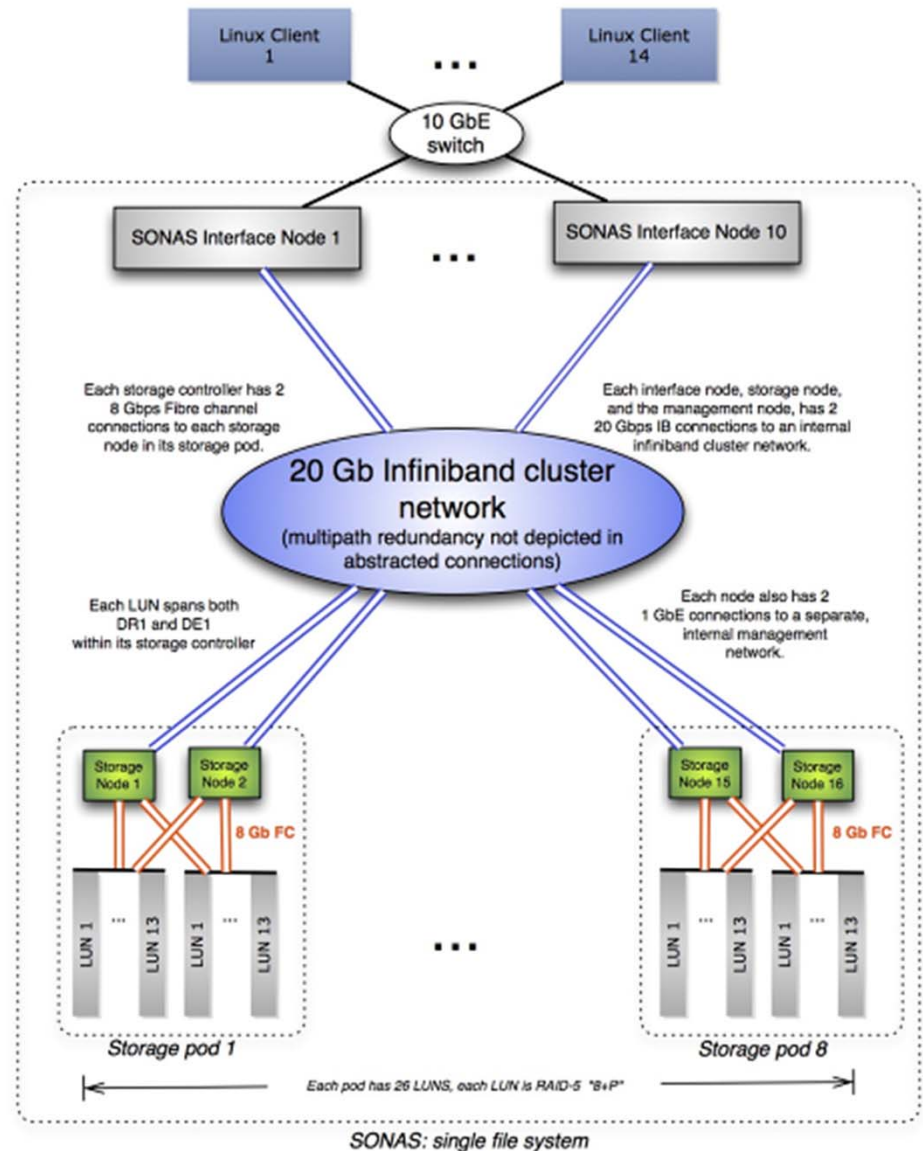
Combining both small block random I/O and large sequential  
Into a single file system **See SONAS SPECsfs deliverables on SONAS Sales Kit**

More information:

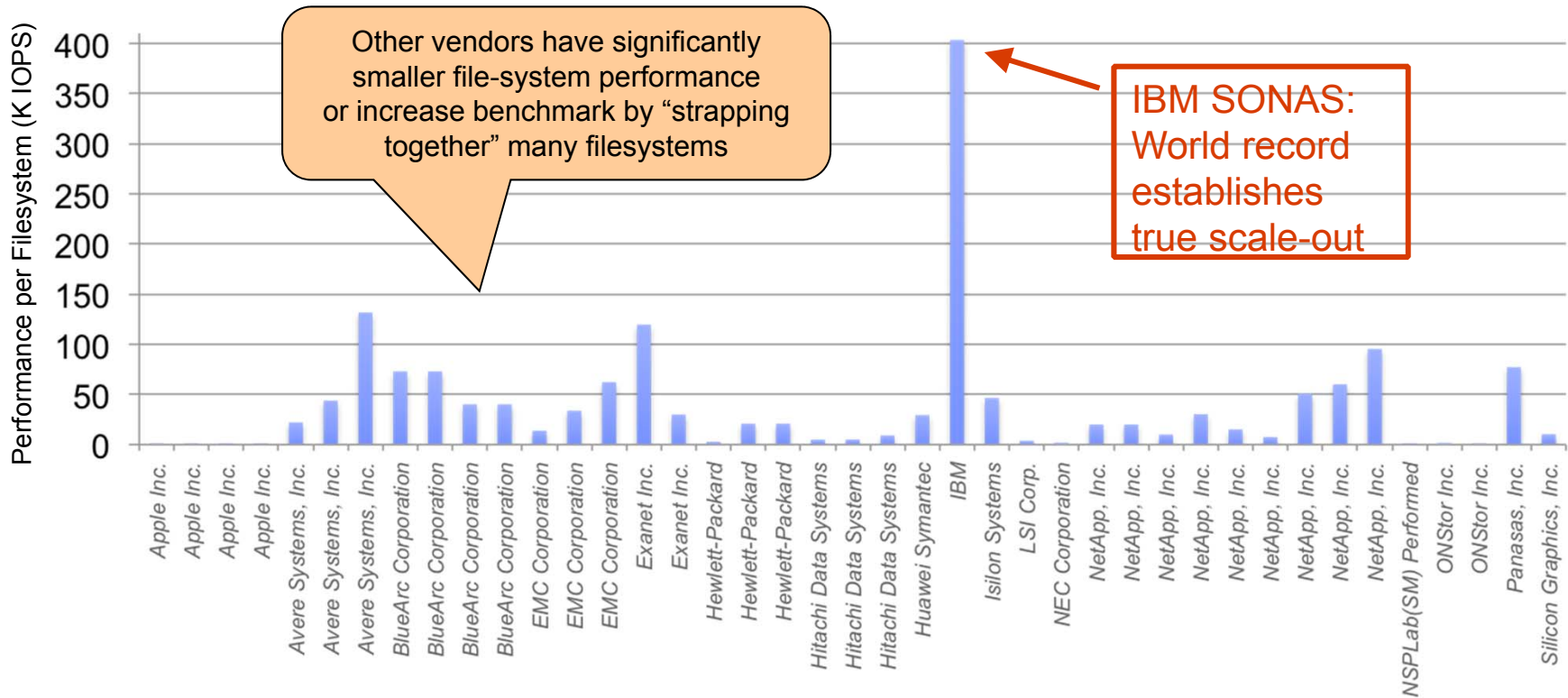
(\*) Based on 403,326 SPECsfs2008 nfs v3 ops per second with an overall response time of 3.23 ms

# SONAS Configuration used for SPECsfs

- **SONAS configuration for SPECsfs:**
  - **10 Interface Nodes** (x3650M3 with max 144 GB RAM)
  - Two **10GbE** ports per Interface Node, only one port active,
  - **8 Storage Pods**; each with 2 Storage nodes and 240 drives presenting 26 LUNs per pod (208 total)
  - Drive type: **15K RPM SAS** hard drives
  - Data Protection: the drives were configured in 208 **RAID-5** arrays (“8+P”)
- **Single File System**
- **This represents no more than 1/3 of the max. configuration**
  - Max configuration still supported with single file system
- **Benchmark used: SPECsfs2008\_nfs.v3**

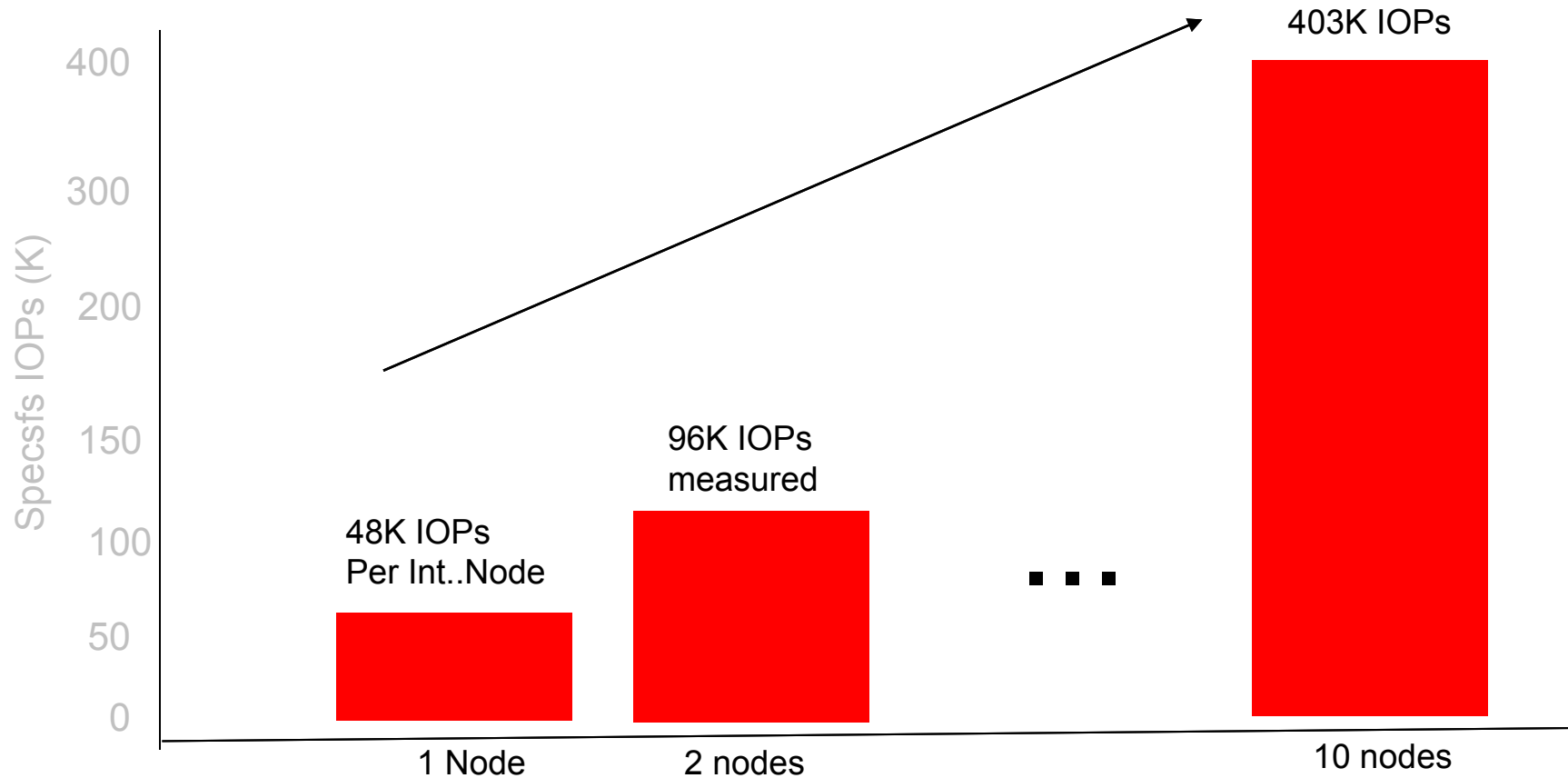


# Performance *per File-System*, by Vendor, based on all publications



The graph shows the maximum throughput per file-system, in thousands of IOPS, based on all SPECsfs2008\_nfs.v3 publications, by vendor. Data as of February 22, 2011  
 Source: <http://www.spec.org/sfs2008/results/sfs2008.html>

# SONAS SPECsfs2008 – Linear Performance



## SONAS R1 Features at a glance

- **Network File Serving**
  - NFS v2/v3/v4\*, CIFS, FTP, HTTP, SCP
  - CIFS ACL mapping into NFSv4 ACL
  - Coherent file locking between NFS and CIFS
- **Clustered parallel file system**
  - Up to 256 file systems
  - Up to 2 billion files per file system
  - Maximum 2PB per file system
- **Quota**
  - User, group and fileset level quotas
  - Soft limits, hard limits, grace periods
- **User Authentication/Authorization**
  - Microsoft® Active Directory
  - Lightweight Directory Access Protocol (LDAP) / with Kerberos
  - Samba primary domain controller (PDC)
- **Data Protection**
  - File system Snapshots, up to 256 per file system
  - Synchronous replication of file system metadata and file data
  - Integrated TSM V6.1 Backup/Archive (B/A) client
- **Centralized Management and Administration**
  - Both Graphical User Interface and Command Line Interface
  - Centralized alert log and event log
  - Event notifications via email or SNMP
- **Integrated Solution Packaging**
  - Single software product, multiple expandable hardware
  - All components integrated into rack(s), cabled, fully tested
  - Updates/patches via centralized SONAS patch management
- **Scalability and Performance**
  - ✓ Up to 30 interface nodes for I/O performance (Release 1)
  - ✓ High Density packaging of HDDs
  - ✓ Support for high performance 15K SAS disk drives and high capacity 7.2K SATA disk drives
  - ✓ Up to 7200 HDDs in single system (14.4PB using 2TB SATA)
- **RAS**
  - ✓ Centralized integrity monitoring via System Health Center
  - ✓ Call home and remote service features
  - ✓ Fully redundant capability in all components for HA
- **Information Lifecycle Management (Release 1.1.1)**
  - ✓ Policy driven file placement, movement, migration and deletion of files over their entire lifetime
  - ✓ Storage tiering, support for SAS and SATA HDD's
  - ✓ Integrated TSM V6.1 HSM (space management) client for migration of inactive files to external TSM server
- **Disaster Recovery (Release 1.1.1)**
  - ✓ Asynchronous replication to another SONAS system
- **Public documentation / users manual / help center**
  - <http://publib.boulder.ibm.com/infocenter/sonasic/sonas1ic/index.jsp>





Publicly available User's Manual (R1.2)
Country/region [ select ]    Terms of use

Home    Products    Services & solutions    Support & downloads    My account

Search:      Search scope: All topics

**Content**

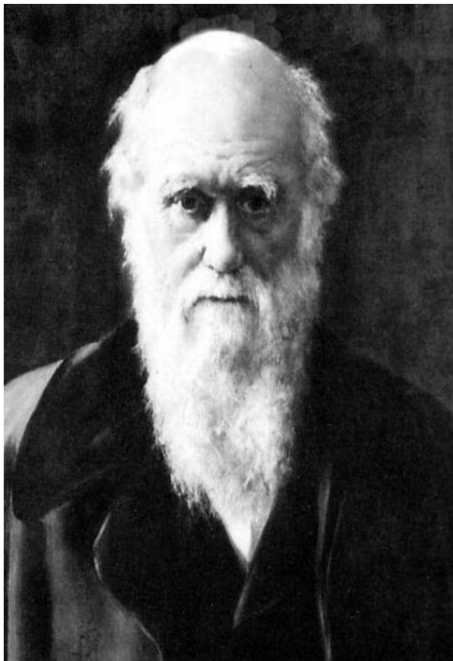
- Welcome
- Product overview
  - What is SoNAS?
  - Warranty Information
- Notices
  - Accessibility
  - Concepts
- Learning and tutorials
- Planning
- Administering
- Using
- Reference
- Troubleshooting
- Service provider information
- Publications
- Glossary
- Feedback
- ibm.com: About IBM - Privacy - C

The SONAS system supports petabytes (PB) of storage and billions of files in a single large file system (up to 2 PB). You can have as few as eight file systems in a fully configured 14.4 PB SONAS system or as many as 256 file systems.

Table 1. Features and benefits

Traditional NAS features in SONAS	SONAS-supported access protocols	Additional SONAS features
<ul style="list-style-type: none"> <li>Snapshots (instantaneous copies of file systems), with up to 256 snapshots per file system</li> <li>User-level, group-level, and fileset-level quotas</li> <li>Synchronous replication of data</li> <li>Integration with user directory servers— Microsoft® Active Directory (AD) and Lightweight Directory Access Protocol (LDAP)</li> <li>Command-line interface (CLI) and browser-based graphical user interface (GUI)</li> </ul>	<p>Clients access data using industry-standard network-file protocols that include:</p> <ul style="list-style-type: none"> <li>Network File System (NFS)</li> <li>Common Internet File System (CIFS)</li> <li>File Transfer Protocol (FTP)</li> </ul>	<ul style="list-style-type: none"> <li>Automated policy-based file management that controls backups and restores, snapshots, and remote replication</li> <li>A single global namespace with logical paths that do not change because of physical data movement</li> <li>Support for Serial Attached SCSI (SAS) and Serial Advanced Technology Attachment (SATA) drives</li> <li>Support for the cloud environment</li> </ul> <p>A controlled set of end users, projects, and applications can:</p> <ul style="list-style-type: none"> <li>Share files with other users within one or more file spaces</li> <li>Control access to their files using access control lists (Microsoft Windows® clients) and user groups</li> <li>Manage each file space with a browser-based tool</li> </ul> <p>The SONAS system also provides:</p> <ul style="list-style-type: none"> <li>Superior performance per price</li> <li>High-availability and load-balancing</li> <li>Centralized management</li> <li>Centralized backup</li> <li>An interconnected cluster of file-serving and network-interfacing nodes in a redundant high-speed data network</li> <li>Virtually no capacity limits</li> </ul>

## A Final Health Warning



**“It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change.”**



IBM

# Disclaimer

No part of this document may be reproduced or transmitted in any form without written permission from IBM Corporation.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The performance data contained herein was obtained in a controlled, isolated environment. Actual results that may be obtained in other operating environments may vary significantly. While IBM has reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customer experiences described herein are based upon information and opinions provided by the customer. The same results may not be obtained by every user.

Reference in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead. It is the user's responsibility to evaluate and verify the operation on any non-IBM product, program or service.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (e.g. IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.