



# Washington Systems Center - Storage

## Accelerate with IBM Storage: Building and Deploying Elastic Storage Server (ESS)



Connie Rice  
IBM  
Technical Sales Specialist  
Washington Systems Center

### Spectrum Scale and Elastic Storage Server (ESS) - Agenda

---

- IBM Spectrum Storage Software
- IBM Spectrum Scale
- What makes an ESS?
- ESS Models
- ESS a software defined Spectrum Scale storage implementation
- GNR – RAID, Declustered RAID, Disk Hospital deeper dive
- Spectrum Scale with ESS Use Cases

# Washington Systems Center - Storage

## IBM Spectrum Storage Software

### Management

Monitoring  
& Control



IBM  
Spectrum  
Control

Cloud-based  
Storage  
Management  
and Support



IBM  
Storage  
Insights

Metadata-Driven  
Data Insight



IBM  
Spectrum  
Discover

Container &  
VM APIs



IBM  
Spectrum  
Connect

### Modern Data Protection

Archive



IBM  
Spectrum  
Archive

Backup



IBM  
Spectrum  
Protect

VM Data  
Availability



IBM  
Spectrum  
Protect Plus

Copy Data  
Management



IBM  
Spectrum  
CDM

### Software-Defined Storage

Scale-Out  
NAS



IBM  
Spectrum  
NAS

Virtualized  
Block



IBM  
Spectrum  
Virtualize

Scale-Out  
File



IBM  
Spectrum  
Scale

Scale-Out  
Object



IBM Cloud  
Object  
Storage

Scale-Out  
Block



IBM  
Spectrum  
Accelerate

### Infrastructure Platforms

High IOPS  
All Flash



Disk



Tape



Cloud



Servers



## IBM Spectrum Scale

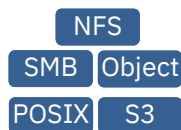


Store everywhere. Run anywhere.

- High-performance, scalable storage software for files, objects and Big Data / Analytics
- Remove data related bottlenecks
- Enable global collaboration
- Optimize cost and performance
- Ensure data availability, integrity and security



Elastic Storage Server



### Differentiation

- World-leading performance and scalability
- Multi-site caching
- Global name space
- Multi-protocol support
- Encryption at rest
- Policy based tiering
- Compression / replication

### Business Value

- Leverage no-bottleneck architecture to scale performance for extreme throughput and low-latency access
- Enable data-anywhere access that spans storage and locations to accelerate applications across the data center or around the world
- Take advantage of authentication, encryption, security and replication options to meet business and regulatory requirements

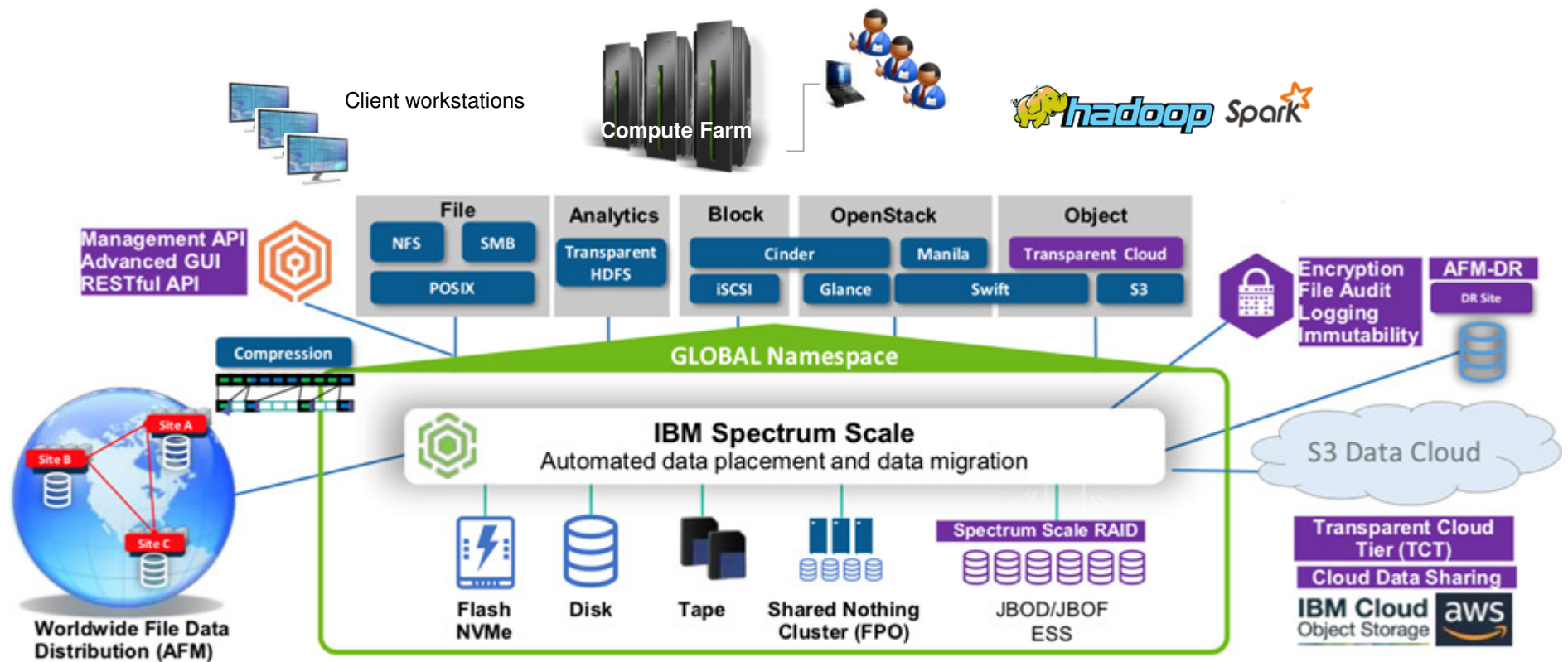
*"With IBM Spectrum Scale, we can completely eliminate unplanned downtime as hardware failures are non-disruptive."*

**Alan Malek, Director of IT**  
Cypress Semiconductor



# Washington Systems Center - Storage

## Spectrum Scale



### What is the Elastic Storage Server?

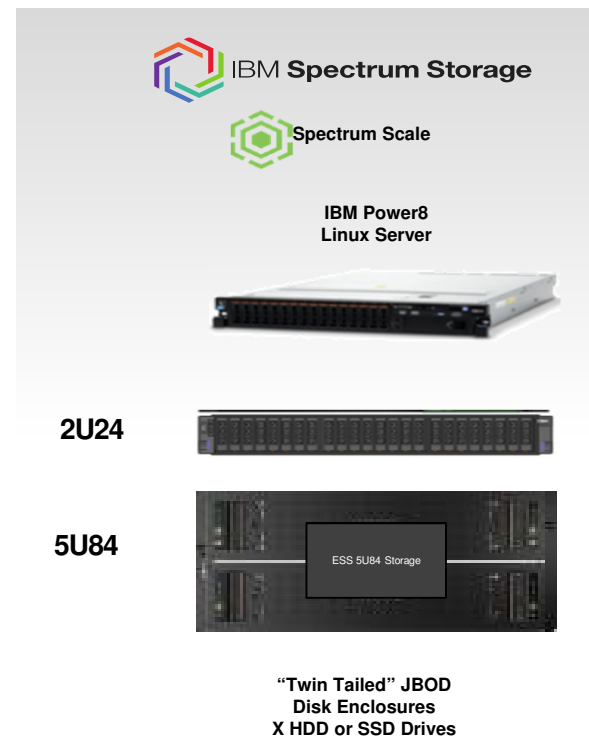
---

- ESS is a software defined Spectrum Scale storage implementation built on Power servers and commodity storage enclosures (JBODs)
- All the storage intelligence in ESS is implemented in Spectrum Scale software
- ESS eliminates risks and makes it quick to deploy and grow a Spectrum Scale cluster
  - Fully validated hardware and software stack
  - Pre-assembled and pre-configured
  - Comes with lab services for on-site deployment
- Available in high performance (All-Flash), high capacity and hybrid models

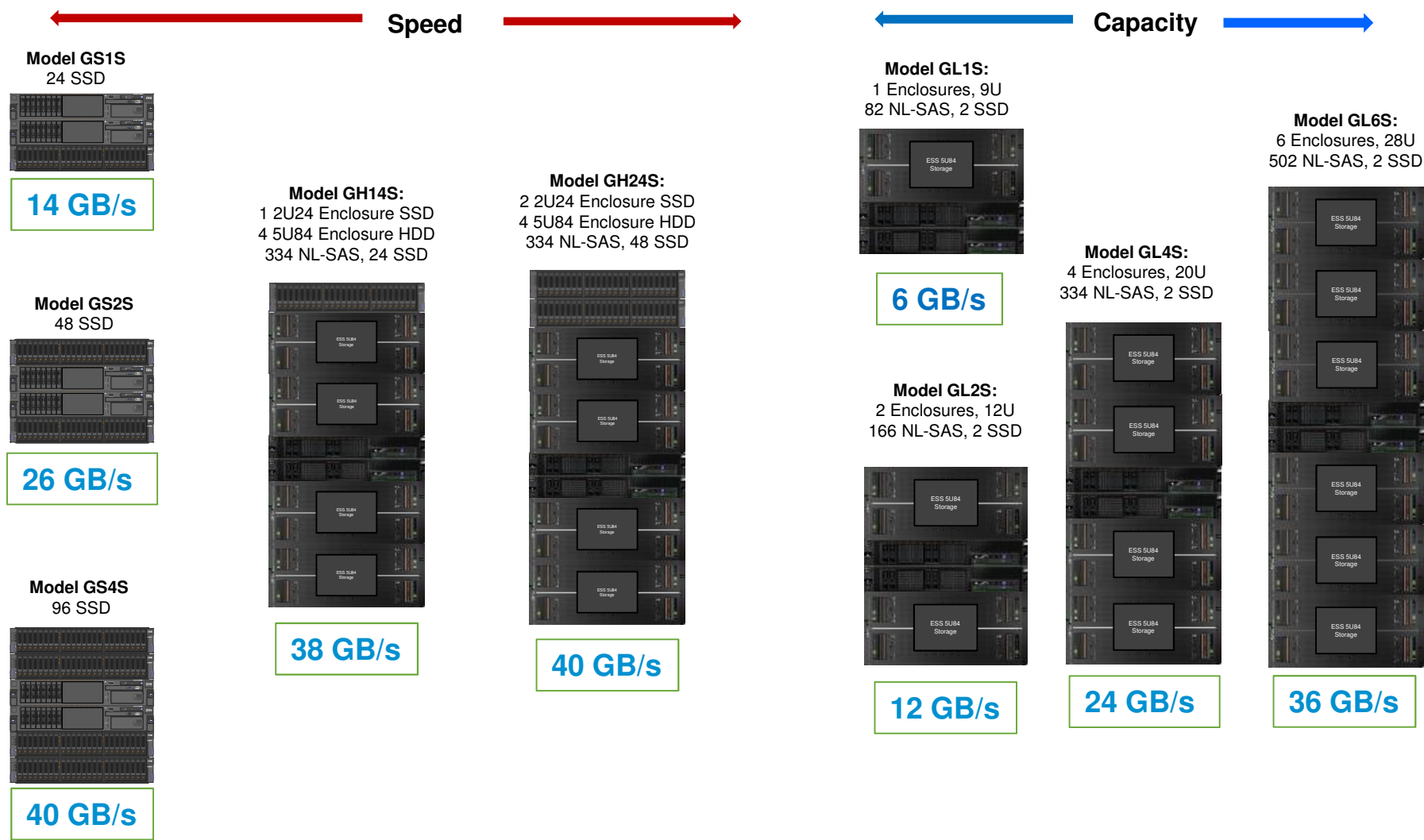
#### **Elastic Storage Server**

(ESS) is an integrated building block solution for Spectrum Scale

## IBM Elastic Storage Server Building Blocks



## IBM Elastic Storage Server (ESS) Current Family





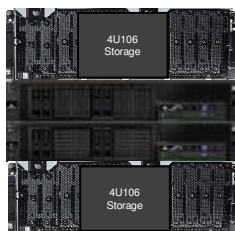
## New ESS C-Series - Maximum Density with Room to Upgrade and Grow!

**New! Model GL2C:**  
1 Enclosure, 8U  
104 NL-SAS, 2 SSD



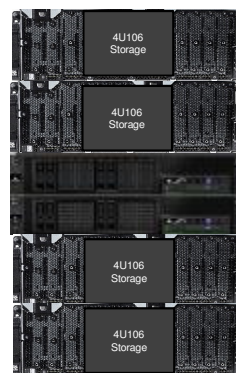
**1.0 PB Disk**

**New! Model GL2C:**  
2 Enclosures, 12U  
210 NL-SAS, 2 SSD



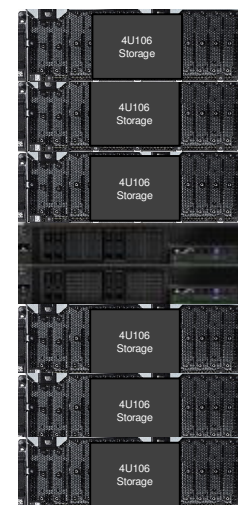
**2.0 PB Disk**

**New! Model GL4C**  
4 Enclosures, 16U  
432 NL-SAS, 2 SSD



**4.2 PB Disk**

**New! Model GL6C:**  
6 Enclosures, 28U  
634 NL-SAS, 2 SSD



**6.3 PB Disk**

## Improved storage capacity and economy

- Announcing the New ESS GLxC models
  - Uses Modern Helium drives
  - 57% Higher Storage Density per enclosure
  - 25%+ Faster per enclosure\*

| Model | 4U106<br>drawers | Drives | Raw<br>capacity |
|-------|------------------|--------|-----------------|
| GL1C  | 1                | 104    | 1.04 PB         |
| GL2C  | 2                | 210    | 2.1 PB          |
| GL4C  | 4                | 432    | 4.22 PB         |
| GL6C  | 6                | 634    | 6.34 PB         |



In a single 42U Rack

>70GB/sec\*

>8PB of storage

>789TB per Rack Unit

*Ultra-dense storage*

\* Final Benchmarks to be published

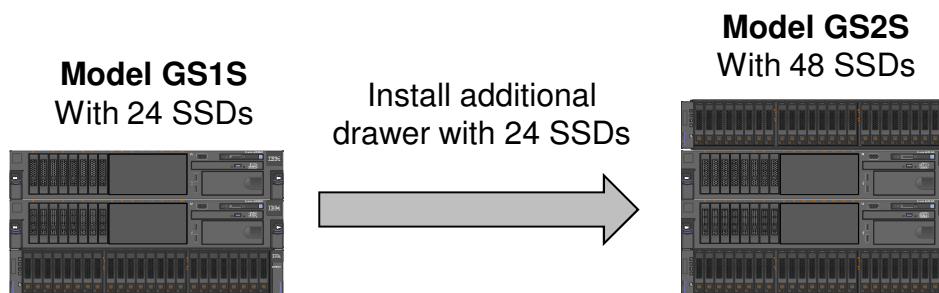
## Non-disruptive upgrades

### Simple expansion of Storage Capacity

- Spectrum Scale will automatically rebalance data in the background
- System automatically puts the new capacity to use
- No need to Archive & Restore data
- No System disruption\*

| Non Disruptive Upgrades |      |
|-------------------------|------|
| From                    | To   |
| GS1S                    | GS2S |
| GS2S                    | GS4S |
| GL1S                    | GL2S |
| GL2S                    | GL4S |
| GL4S                    | GL6S |

Example



\*Requires space available in the rack

## ESS Family

| Model    | Base Server | Server to JBOD    | Status       | Notes                              | Drive Enclosures                  |
|----------|-------------|-------------------|--------------|------------------------------------|-----------------------------------|
| GS1S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 1 x 24 Drive SSD                  |
| GS2S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 2 x 24 Drive SSD                  |
| GS4S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 4 x 24 Drive SSD                  |
| GH14S    | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 1 x 24 Drive SSD & 4 84 Drive HDD |
| GH24S    | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 2 x 24 Drive SSD & 4 84 Drive HDD |
| GL1S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 1 x 84 Drive HDD                  |
| GL2S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 2 x 84 Drive HDD                  |
| GL4S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 4 x 84 Drive HDD                  |
| GL6S     | Power 8     | 4X Copper 6GB SAS | Released     |                                    | 6 x 84 Drive HDD                  |
| GL1C     | Power 8     | 5X Fiber 12GB SAS | Announced    | Constellation S42 Rack             | 1 x 106 Drive HDD                 |
| GL2C     | Power 8     | 5X Fiber 12GB SAS | Announced    | Constellation S42 Rack             | 2 x 106 Drive HDD                 |
| GL4C     | Power 8     | 5X Fiber 12GB SAS | Announced    | Constellation S42 Rack             | 4 x 106 Drive HDD                 |
|          |             |                   |              |                                    | * Watch this space                |
| GL6C     | Power 8     | 5X Fiber 12GB SAS | Announced    | Constellation S42 Rack             | 6 x 106 Drive HDD                 |
|          |             |                   |              |                                    | * Watch this space                |
| GL4Coral | Power 9     | 5X Fiber 12GB SAS | CORAL Actual | PCI GEN4<br>Constellation S42 Rack | 4 x 106 Drive HDD                 |

\* All ESS Systems use two Drive Slots for SSD's used for System Management

## ESS Capacity and Throughput

| Model |          | Drive    |        | Capacity (in TB) |        |      |       |       |       | Data Seq. Throughput (in GBps)(8+2P) |         |                 | Data Seq. Throughput (in GBps)(8+3P) |         |                 |
|-------|----------|----------|--------|------------------|--------|------|-------|-------|-------|--------------------------------------|---------|-----------------|--------------------------------------|---------|-----------------|
| Type  | Sub-Type | Size     | Type   | Raw (TB)         | Usable |      |       |       |       | 10G Eth GB/Sec                       | 40G Eth | FDR/EDR IB RDMA | 10G Eth GB/Sec                       | 40G Eth | FDR/EDR IB RDMA |
|       |          |          |        |                  | 8+2P   | 8+3P | 2-way | 3-Way | 4-Way |                                      |         |                 |                                      |         |                 |
| GLxS  | GL1S     | 4 TB     | NL-SAS | 328              | 245    | 222  | 153   | 102   | 76    | 4.3                                  | 5       | 6.7             | x                                    | x       | x               |
|       |          | 8 TB     | NL-SAS | 489              | 489    | 445  | 306   | 204   | 153   | 4.6                                  | 5.3     | 7.1             | x                                    | x       | x               |
|       |          | 10 TB    | NL-SAS | 820              | 612    | 556  | 382   | 255   | 191   | 4.6                                  | 5.3     | 7.1             | x                                    | x       | x               |
|       | GL2S     | 4 TB     | NL-SAS | 664              | 508    | 462  | 318   | 212   | 159   | 7.2                                  | 8.3     | 11              | 6.5                                  | 7.5     | 10              |
|       |          | 8 TB     | NL-SAS | 1328             | 1016   | 924  | 635   | 423   | 318   | 7.2                                  | 8.3     | 11              | 6.5                                  | 7.5     | 10              |
|       |          | 10 TB    | NL-SAS | 1660             | 1270   | 1155 | 794   | 529   | 397   | 7.2                                  | 8.3     | 11              | 6.5                                  | 7.5     | 10              |
|       | GL4S     | 4 TB     | NL-SAS | 1336             | 1035   | 941  | 647   | 431   | 323   | 10                                   | 14      | 22.2            | 10                                   | 14      | 20.2            |
|       |          | 8 TB     | NL-SAS | 2672             | 2070   | 1882 | 1294  | 862   | 647   | 10                                   | 14      | 22.2            | 10                                   | 14      | 20.2            |
|       |          | 10 TB    | NL-SAS | 3340             | 2587   | 2352 | 1617  | 1078  | 809   | 10                                   | 14      | 22.2            | 10                                   | 14      | 20.2            |
|       | GL6S     | 4 TB     | NL-SAS | 2008             | 1562   | 1420 | 976   | 651   | 488   | 10                                   | 14      | 33.4            | 10                                   | 14      | 30.4            |
|       |          | 8 TB     | NL-SAS | 4016             | 3123   | 2840 | 1952  | 1301  | 976   | 10                                   | 14      | 33.4            | 10                                   | 14      | 30.4            |
|       |          | 10 TB    | NL-SAS | 5020             | 3904   | 3549 | 2440  | 1627  | 1220  | 10                                   | 14      | 33.4            | 10                                   | 14      | 30.4            |
| GSxS  | GS1S     | 3.84 TB  | SSD    | 94               | 60     | 55   | 38    | 25    | 19    | 6.4                                  | 7.4     | 9.9             | 5.9                                  | 6.8     | 9               |
|       |          | 15.36 TB | SSD    | 360              | 241    | 219  | 151   | 100   | 75    | 6.4                                  | 7.4     | 9.9             | 5.9                                  | 6.8     | 9               |
|       | GS2S     | 3.84 TB  | SSD    | 188              | 132    | 120  | 83    | 55    | 41    | 10                                   | 14      | 19.8            | 10                                   | 13.5    | 18              |
|       |          | 15.36 TB | SSD    | 720              | 530    | 482  | 331   | 221   | 166   | 10                                   | 14      | 19.8            | 10                                   | 13.5    | 18              |
|       | GS4S     | 3.84 TB  | SSD    | 376              | 277    | 252  | 173   | 115   | 87    | 10                                   | 14      | 39.6            | 10                                   | 14      | 36              |
|       |          | 15.36 TB | SSD    | 1440             | 1108   | 1007 | 692   | 462   | 346   | 10                                   | 14      | 39.6            | 10                                   | 14      | 36              |

MD (Metadata) pool for ESS RAID/replication setting is 3-way (default)

MD Pool Space not included in usable space capacity calculations



## ESS Capacity and Throughput

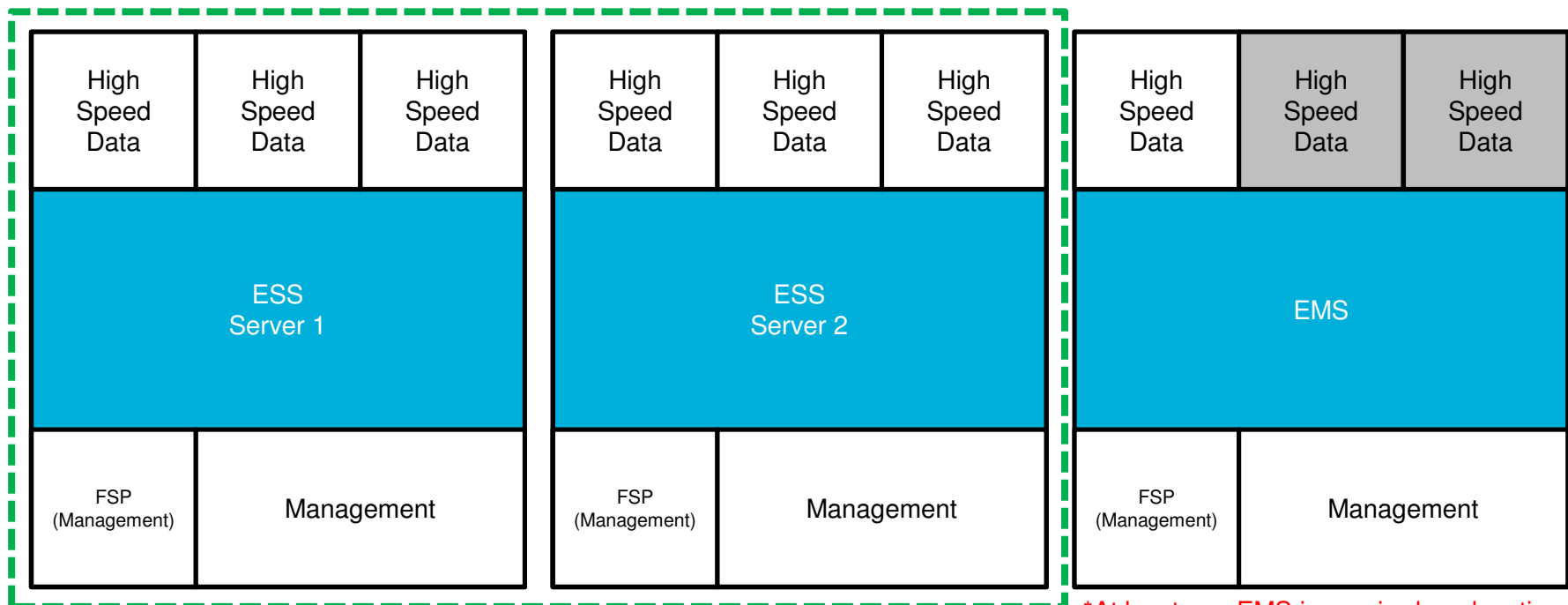
| Model |          | Drive    |        | Capacity (in TB) |        |      |       |       |       | Data Seq. Throughput (in GBps)(8+2P) |         |                 | Data Seq. Throughput (in GBps)(8+3P) |         |                 |
|-------|----------|----------|--------|------------------|--------|------|-------|-------|-------|--------------------------------------|---------|-----------------|--------------------------------------|---------|-----------------|
| Type  | Sub-Type | Size     | Type   | Raw (TB)         | Usable |      |       |       |       | 10G Eth GB/Sec                       | 40G Eth | FDR/EDR IB RDMA | 10G Eth GB/Sec                       | 40G Eth | FDR/EDR IB RDMA |
|       |          |          |        |                  | 8+2P   | 8+3P | 2-way | 3-Way | 4-Way |                                      |         |                 |                                      |         |                 |
| GLxS  | GL1S     | 4 TB     | NL-SAS | 328              | 245    | 222  | 153   | 102   | 76    | 4.3                                  | 5       | 6.7             | x                                    | x       | x               |
|       |          | 8 TB     | NL-SAS | 489              | 489    | 445  | 306   | 204   | 153   | 4.6                                  | 5.3     | 7.1             | x                                    | x       | x               |
|       |          | 10 TB    | NL-SAS | 820              | 612    | 556  | 382   | 255   | 191   | 4.6                                  | 5.3     | 7.1             | x                                    | x       | x               |
|       | GL2S     | 4 TB     | NL-SAS | 664              | 508    | 462  | 318   | 212   | 159   | 7.2                                  | 8.3     | 11              | 6.5                                  | 7.5     | 10              |
|       |          | 8 TB     | NL-SAS | 1328             | 1016   | 924  | 635   | 423   | 318   | 7.2                                  | 8.3     | 11              | 6.5                                  | 7.5     | 10              |
|       |          | 10 TB    | NL-SAS | 1660             | 1270   | 1155 | 794   | 529   | 397   | 7.2                                  | 8.3     | 11              | 6.5                                  | 7.5     | 10              |
|       | GL4S     | 4 TB     | NL-SAS | 1336             | 1035   | 941  | 647   | 431   | 323   | 10                                   | 14      | 22.2            | 10                                   | 14      | 20.2            |
|       |          | 8 TB     | NL-SAS | 2672             | 2070   | 1882 | 1294  | 862   | 647   | 10                                   | 14      | 22.2            | 10                                   | 14      | 20.2            |
|       |          | 10 TB    | NL-SAS | 3340             | 2587   | 2352 | 1617  | 1078  | 809   | 10                                   | 14      | 22.2            | 10                                   | 14      | 20.2            |
|       | GL6S     | 4 TB     | NL-SAS | 2008             | 1562   | 1420 | 976   | 651   | 488   | 10                                   | 14      | 33.4            | 10                                   | 14      | 30.4            |
|       |          | 8 TB     | NL-SAS | 4016             | 3123   | 2840 | 1952  | 1301  | 976   | 10                                   | 14      | 33.4            | 10                                   | 14      | 30.4            |
|       |          | 10 TB    | NL-SAS | 5020             | 3904   | 3549 | 2440  | 1627  | 1220  | 10                                   | 14      | 33.4            | 10                                   | 14      | 30.4            |
| GSxS  | GS1S     | 3.84 TB  | SSD    | 94               | 60     | 55   | 38    | 25    | 19    | 6.4                                  | 7.4     | 9.9             | 5.9                                  | 6.8     | 9               |
|       |          | 15.36 TB | SSD    | 360              | 241    | 219  | 151   | 100   | 75    | 6.4                                  | 7.4     | 9.9             | 5.9                                  | 6.8     | 9               |
|       | GS2S     | 3.84 TB  | SSD    | 188              | 132    | 120  | 83    | 55    | 41    | 10                                   | 14      | 19.8            | 10                                   | 13.5    | 18              |
|       |          | 15.36 TB | SSD    | 720              | 530    | 482  | 331   | 221   | 166   | 10                                   | 14      | 19.8            | 10                                   | 13.5    | 18              |
|       | GS4S     | 3.84 TB  | SSD    | 376              | 277    | 252  | 173   | 115   | 87    | 10                                   | 14      | 39.6            | 10                                   | 14      | 36              |
|       |          | 15.36 TB | SSD    | 1440             | 1108   | 1007 | 692   | 462   | 346   | 10                                   | 14      | 39.6            | 10                                   | 14      | 36              |

MD (Metadata) pool for ESS RAID/replication setting is 3-way (default)

MD Pool Space not included in usable space capacity calculations

# Washington Systems Center - Storage

## ESS GxxS Network Options

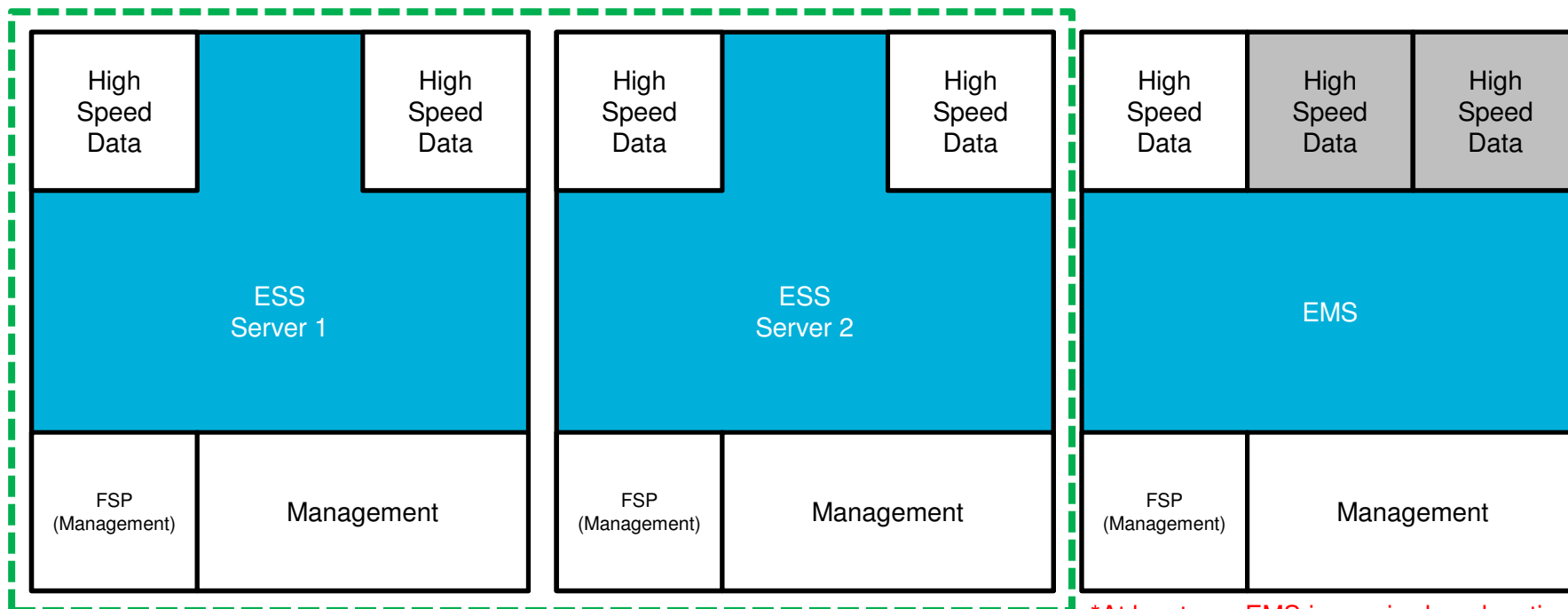


\*At least one EMS is required per location

| Purpose         | Per server (2 servers per ESS)  | Interface Type | Ports |
|-----------------|---|----------------|-------|
| FSP             | Built In – 1 per server   | 1 GbE          | 1     |
| Management      | 1 per server  | 1 GbE          | 4     |
| High Speed Data | <ul style="list-style-type: none"> <li>Up to three per server</li> <li>Must be the same in both ESS Servers</li> <li>EMS must have ability to connect to all networks – redundancy not required.</li> </ul> | 10 GbE         | 2     |
|                 |   | 40 GbE         | 2     |
|                 |   | 100 GbE        | 2     |
|                 |   | 56 Gb IB       | 2     |
|                 |   | 100 Gb IB      | 2     |

# Washington Systems Center - Storage

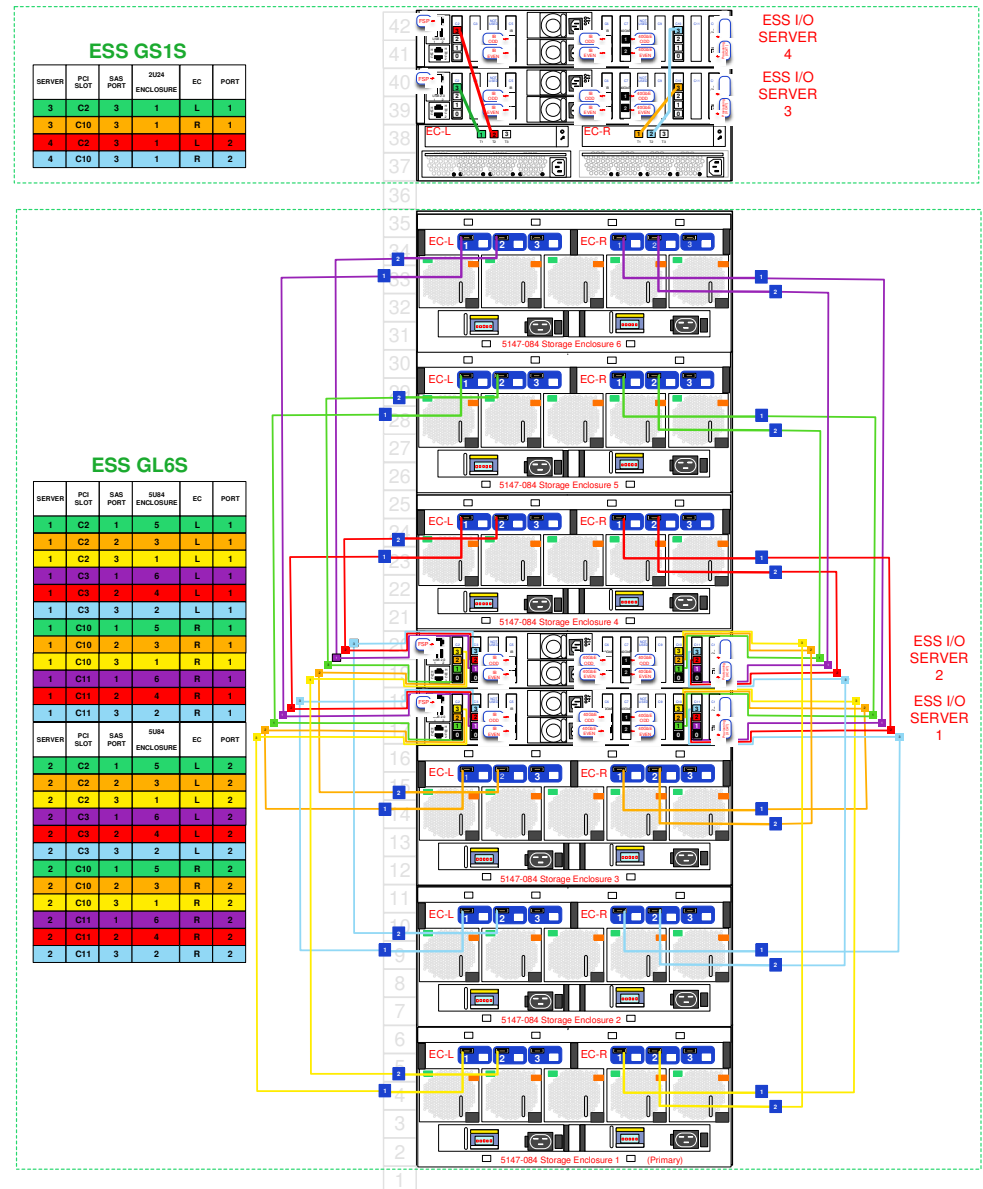
## ESS GxxC Network Options



| Purpose         | Per server (2 servers per ESS)  | Interface Type | Ports |
|-----------------|---|----------------|-------|
| FSP             | Built In – 1 per server   | 1 GbE          | 1     |
| Management      | 1 per server  | 1 GbE          | 4     |
| High Speed Data | <ul style="list-style-type: none"> <li>Up to two per server</li> <li>Must be the same in both ESS Servers</li> <li>EMS must have ability to connect to all networks – redundancy not required.</li> </ul> | 10 GbE         | 2     |
|                 |   | 40 GbE         | 2     |
|                 |   | 100 GbE        | 2     |
|                 |   | 56 Gb IB       | 2     |
|                 |   | 100 Gb IB      | 2     |

## Example Rack Configuration

- Real life deployment example
- Rack contains ESS GS1S and ESS GL6S
- EMS is in a different rack
- This client chose to provide network infrastructure
  - No Top of Rack Switch Included
  - Client wanted IB network & 40 GbE to serve two different workloads

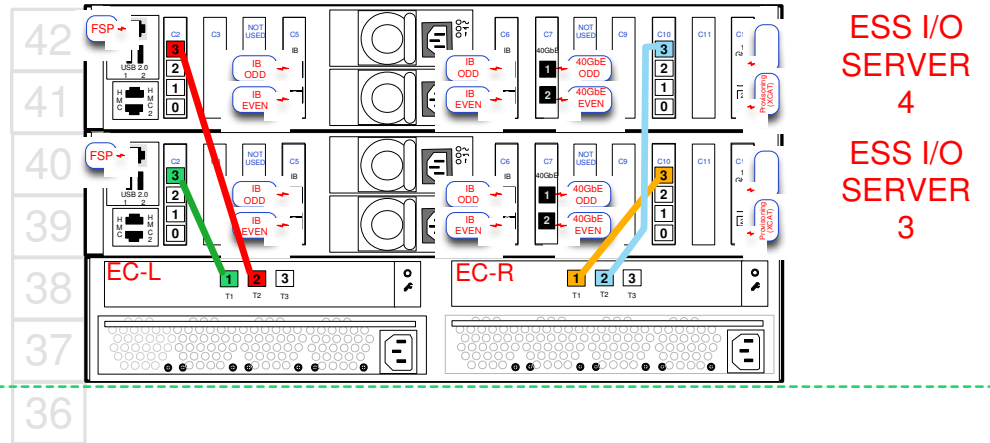


# Washington Systems Center - Storage

## Example Rack Configuration

### ESS GS1S

| SERVER | PCI SLOT | SAS PORT | 2U24 ENCLOSURE | EC | PORT |
|--------|----------|----------|----------------|----|------|
| 3      | C2       | 3        | 1              | L  | 1    |
| 3      | C10      | 3        | 1              | R  | 1    |
| 4      | C2       | 3        | 1              | L  | 2    |
| 4      | C10      | 3        | 1              | R  | 2    |

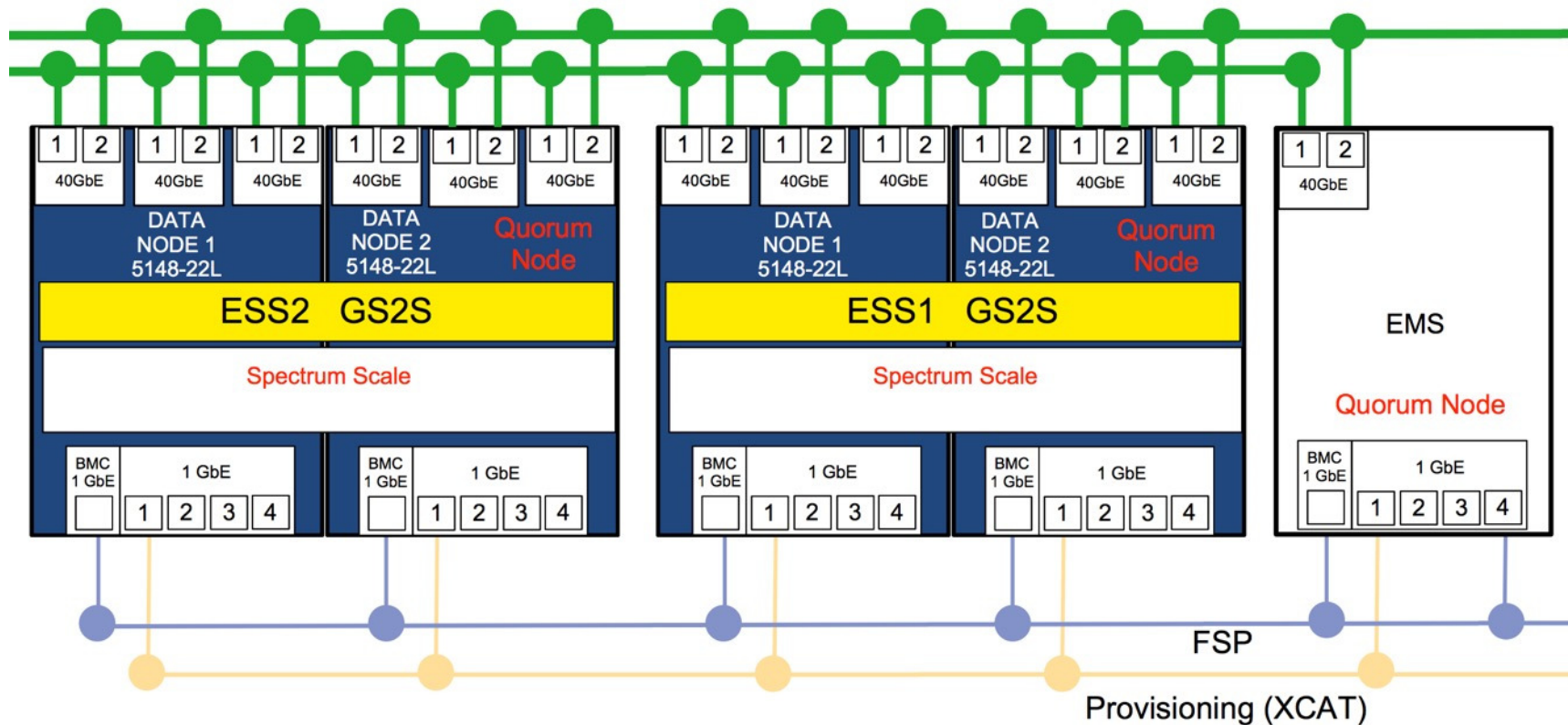


| Location | Interface Type   | Purpose                         |
|----------|------------------|---------------------------------|
| Built In | 1 GbE BMC        | FSP – Field Service Processor   |
| C2       | SAS              | Server to JBOD                  |
| C3       | NOT USED         |                                 |
| C4       | NOT USED         |                                 |
| C5       | 2 ports 56 Gb IB | High Speed Data                 |
| C6       | 2 Ports 56 Gb IB | High Speed Data                 |
| C7       | 2 Ports 40 GbE   | High Speed Data                 |
| C8       | NOT USED         |                                 |
| C9       | NOT USED         |                                 |
| C10      | SAS              | Server to JBOD                  |
| C11      | NOT USED         |                                 |
| C12      | 4 Ports 1 GbE    | Management (XCAT, Campus, etc.) |

- Real life deployment example
- ESS GS1S
- Client wanted IB network & 40 GbE to serve two different workloads



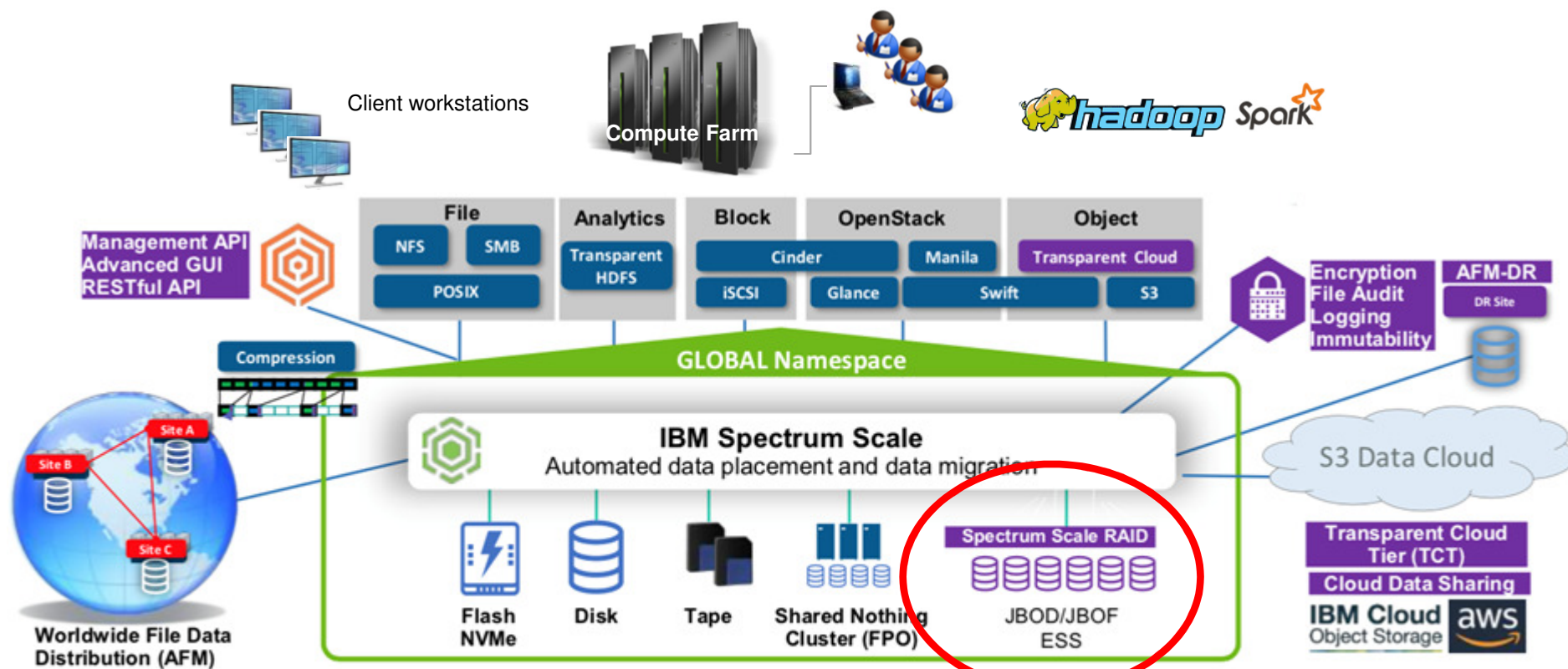
## Example Spectrum Scale Deployment with ESS



- Real life deployment example
- Two ESS GS2S systems (SSD's)
- 40 GbE High Speed Data Network
- First-on-site EMS system
- FSP and Provisioning(XCAT) Networks

# Washington Systems Center - Storage

## Building a Global Namespace – It's case of “and” rather than “or”

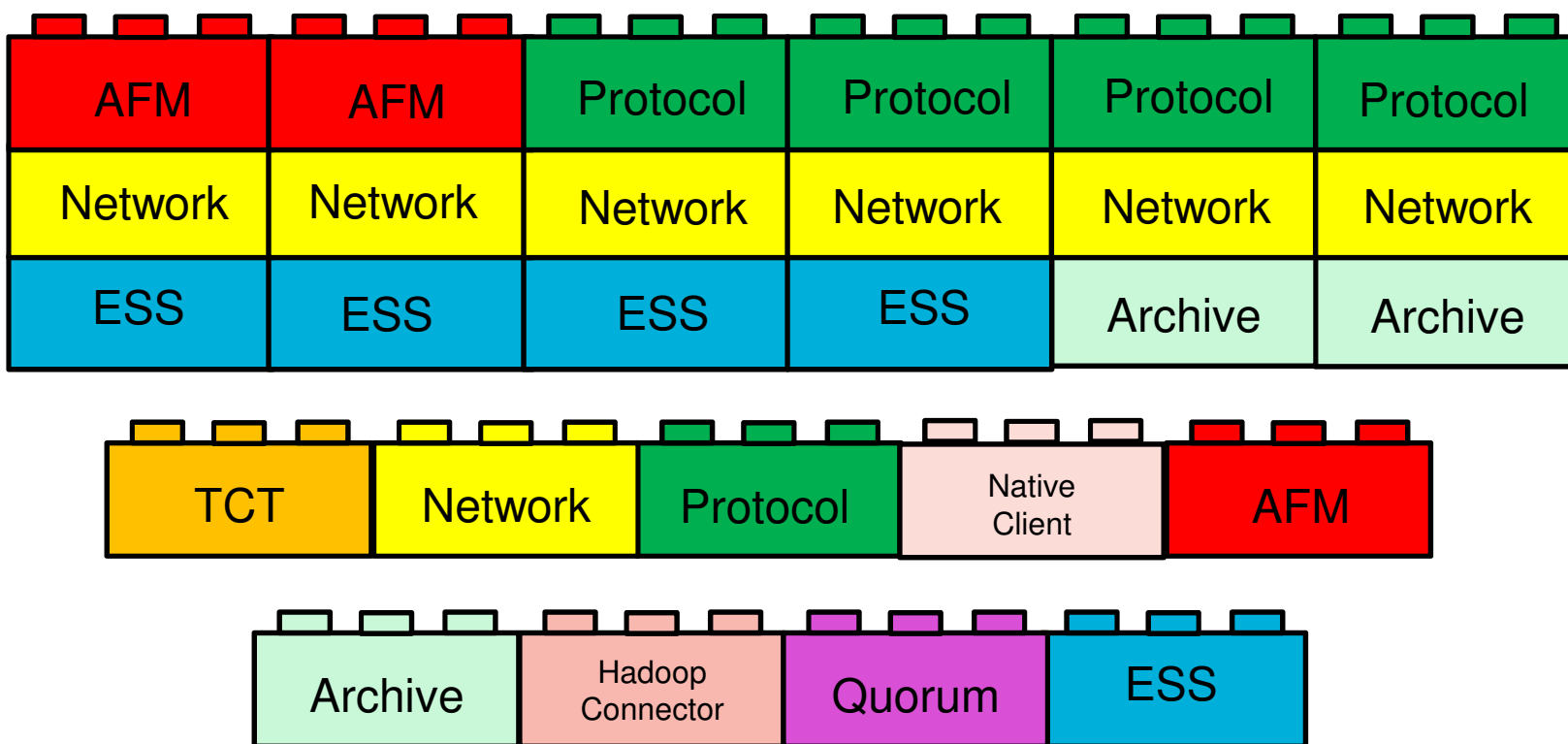


- AFM Nodes for caching and distribution
- AFM-DR Nodes for non-synchronous DR
- ISKLM for Encryption Key Management
- Protocol Nodes for Object, NFS and SMB access
- Transparent Cloud Tiering (TCT) Nodes
- Hadoop Connector lives in Hadoop Cluster
- Archive via Spectrum Archive
- Native Spectrum Scale File system access

Where ESS fits in the overall solution

## Spectrum Scale Building Blocks

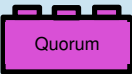
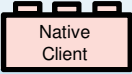
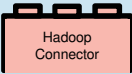






Build and the File System solution you need by adding storage and functionality as required



Add Spectrum Scale Building Blocks to add capacity, speed, and functionality.

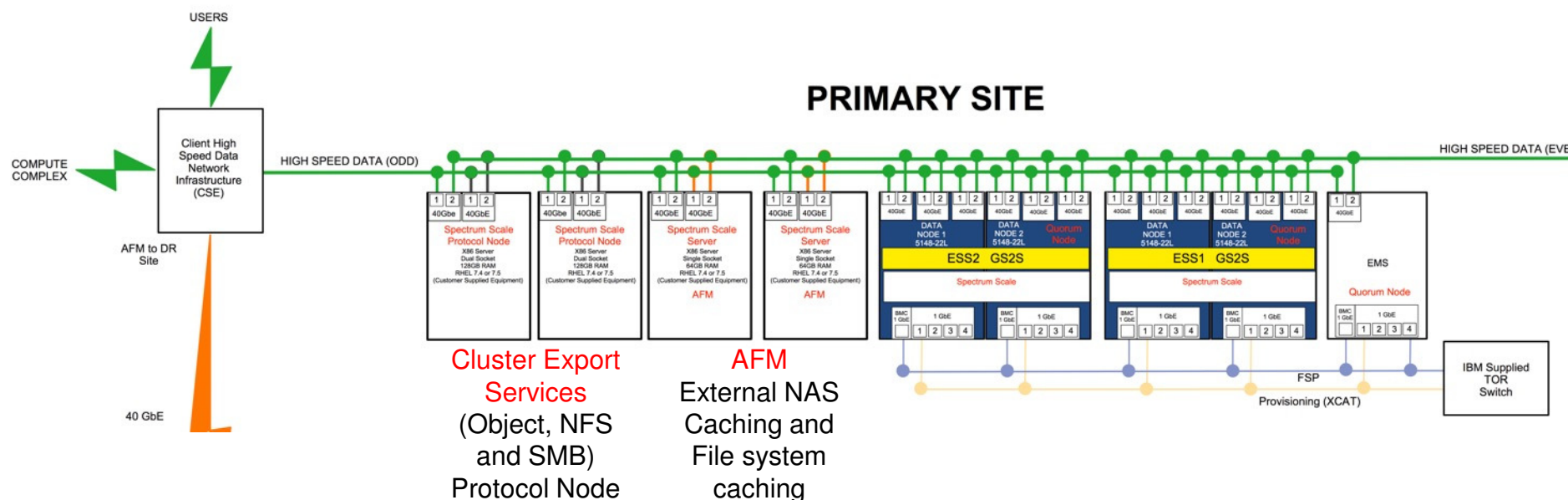
# Washington Systems Center - Storage

## Spectrum Scale Building Blocks

| Building Block   | Function                                       | Description  | Requirements  |
|--|--|--|---|
|  Quorum           | Spectrum Scale Cluster Quorum Nodes            | Spectrum Scale is a clustered parallel file system. Like any cluster, a set of quorum nodes is required to monitor health and availability   | Three per Spectrum Scale Cluster<br>Choose stable systems with stable network.  |
|  Native Client    | Spectrum Scale Client                          | Users and compute nodes with native Spectrum Scale file access. These nodes see the most functionality and best performance.   | Spectrum Scale Client software loaded on client servers and workstations  |
|  Hadoop Connector | Hadoop Cluster HDFS / Spectrum Scale Connector | The connector lives in some number of servers within a HADOOP cluster and allow members of that cluster to access the Spectrum Scale File system using Hadoop File System (HDFS) protocol and commands   |   |
|  Protocol         | Cluster Export Services<br>SMB, NFS, & Object  | Spectrum Scale file system access for NFS (Ganesha), SMB (Samba), and Object (Swift & Keystone). Files are stored using Spectrum Scale formatting, but may be accessed via any of these protocols.   | z, POWER, x86: RHEL 7.1+<br>x86: SLES 12, 12 SP2<br>z, x86: SLES 12 SP3<br>x86: Ubuntu 16.04, 16.04.03, 18.04.1 All protocol nodes must be same OS and hardware platform.<br>Object not supported on z.<br>Max of 32 Protocol nodes in all NFS systems<br>Max of 16 Protocol nodes if CIFS/SMB is present |
|  AFM              | Active File Management                         | Active File Management (AFM) lets Spectrum Scale extend over geographic distances. AFM tolerates unreliable, high-latency networks (like a WAN).<br>Caches copies of data from a remote file system into the local Spectrum Scale cluster. Cached files have the same read and write performance as other local files. | Gateways are required, and must be Linux system Dedicated gateway nodes, 128 GiB RAM ms   |
|  TCT            | Transparent Cloud Tiering                      | Extend Spectrum Scale to private or public cloud.<br>IBM Cloud Object Storage, Amazon S3.<br>Private Cloud or on premises: IBM Cloud Object Storage, OpenStack Swift (and Swift3). Transparent to end users using Spectrum Scale   | Server: RHEL 7.1+ on x86 or RHEL 7.2+ on POWER8 LE.<br>Client on x86: RHEL 7.1+, SLES 12, Debian 7/7<br>Client on POWER8 LE: RHEL 7.2+  |
|  Archive        | Integration with Spectrum Archive              | Spectrum Archive provides direct, intuitive and graphical access to data stored in IBM tape drives and libraries by incorporating the Linear Tape File System™ (LTFS) format standard for reading, writing and exchanging descriptive metadata on formatted tape cartridges.   | Offered as three software solutions for managing your digital files with the LTFS format: Single Drive Edition, Library Edition and Enterprise Edition. Spectrum Archive eliminates the need for additional tape management and software to access data.  |
|  Network        | Management and High Speed Data                 |  | IB requires compatible firmware and OFED levels on all systems and switches. Mellanox IB Multiple 1, 10, 40,& 100 GbE   |
|  ESS            | Elastic Storage Server                         | ESS is a software defined Spectrum Scale storage implementation built on Power servers and commodity storage enclosures (JBODs)  | Power 8 & Redhat  |

# Washington Systems Center - Storage

## Example Spectrum Scale Deployment with ESS



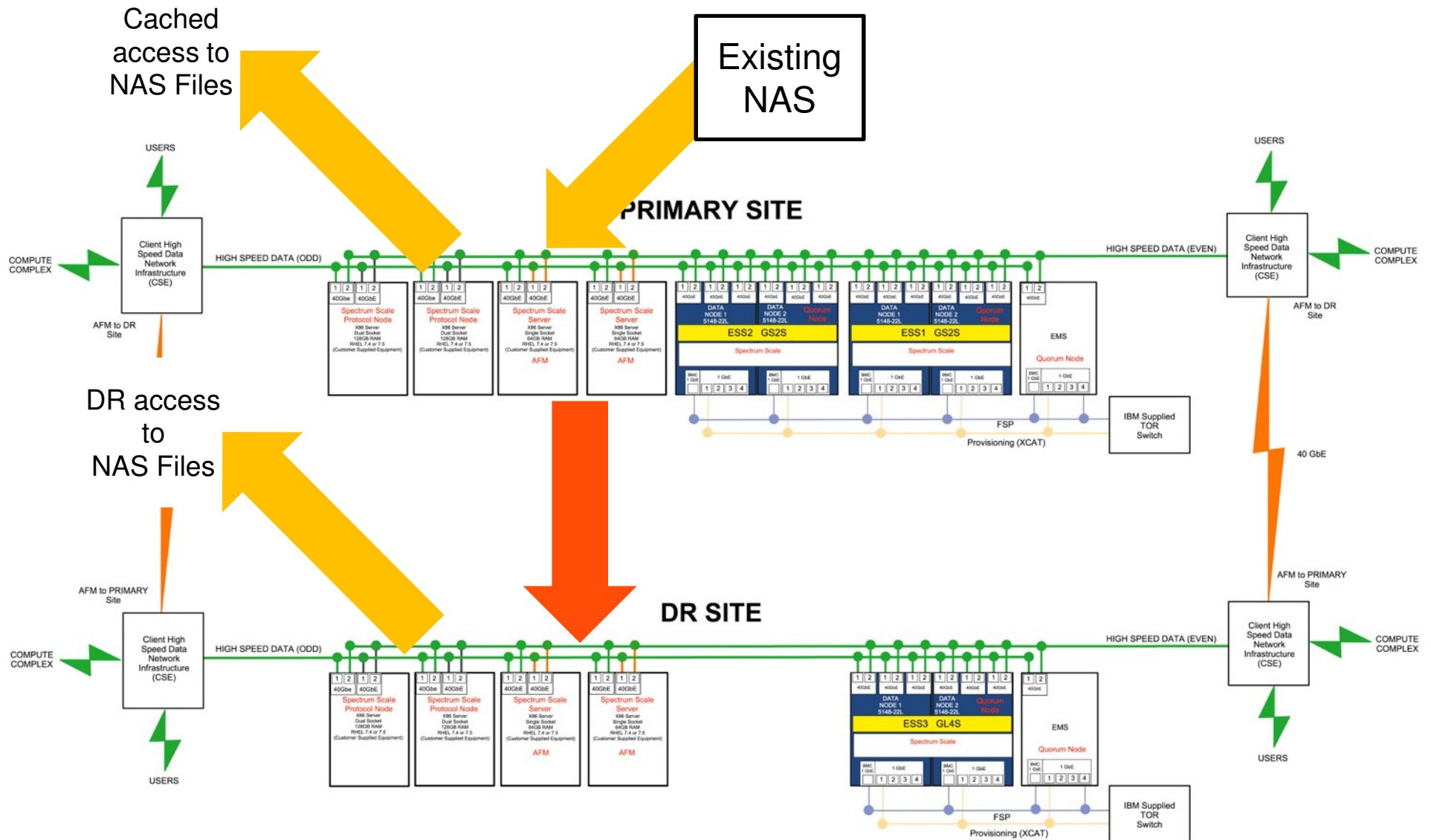
## Making the high performance file system available outside the cluster

- Protocol Nodes provide Cluster Export Services
- NFS
- SMB
- Object
- Active File Management (AFM)
- AFM Servers Allow migration from existing NAS
- Access to NAS is immediately available as cache from the Spectrum Scale File System
- Think of it as Snapping between disparate systems
- AFM also allows files to be cached at remote site for workload distribution and DR



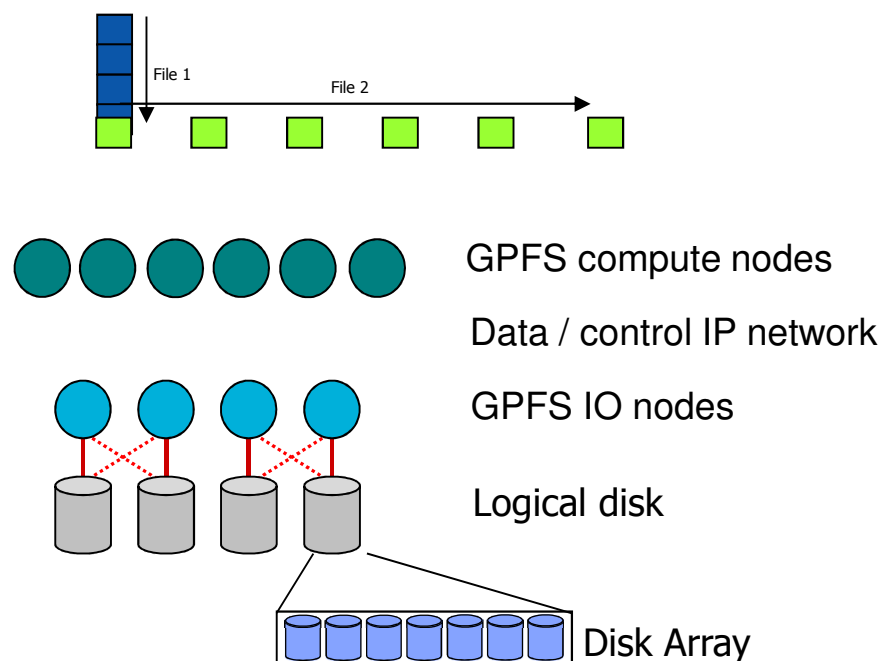
# Washington Systems Center - Storage

## Example Spectrum Scale Deployment with ESS



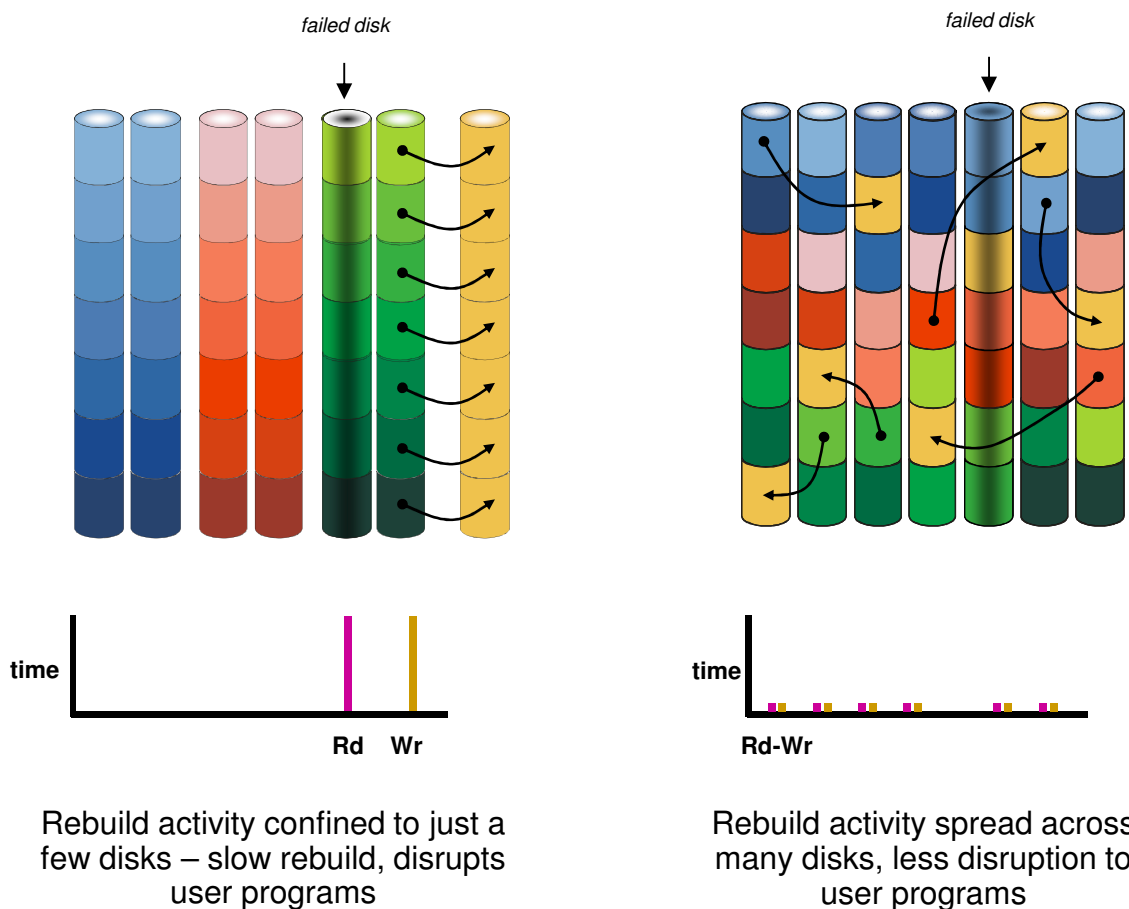
## Spectrum Scale - GPFS – High Throughput

- Wide Striping
  - Both data and metadata striped across many disks
- Files striped block by block across all disks
  - ... for throughput and load balancing



YouTube: IBM Spectrum Scale RAID for petabyte storage:  
<https://www.youtube.com/watch?v=2g5rx4gP6yU>

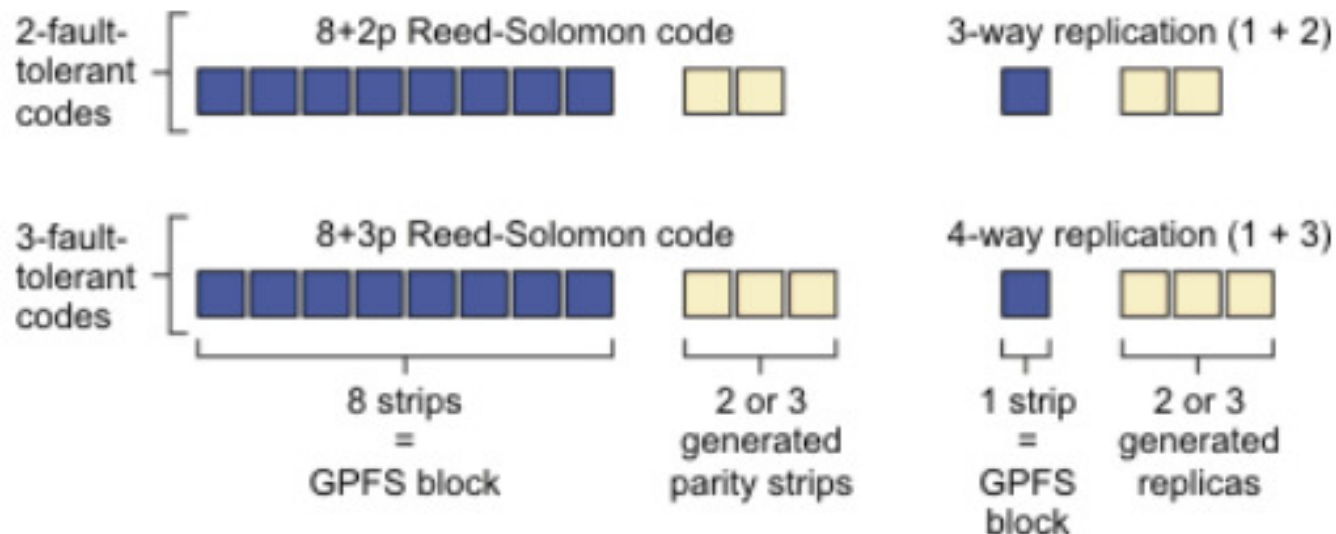
## Spectrum Scale RAID (GNR) Rebuild Overhead Reduction Example



Rebuild overhead reduced by 3.5x

## Elastic Storage Server – GNR Supported Redundancy Codes

- **Reed-Solomon Encoding**
  - 8 Data Strips + 2 or 3 Parity Strips
  - Stripe width 10 or 11 strips
  - Storage efficiency 80% or 73% respectively\*
- **3-way or 4-way replication**
  - Strip size is GPFS Data Block size
  - Storage efficiency 33% or 25% respectively



- **\*Excluding user-configurable spare space for rebuilds**

## IBM Spectrum Scale RAID (GNR) – value add on Spectrum Scale

---

### **Spectrum Scale RAID Gives you**

- **Software RAID in a dual ported JBOD Array**
- **Automatic Recovery**
- **Disk scrubbing – detect & repair latent sector errors in the background**
- **Journaling – small write data to NVRAM / SSD**
- **Asynchronous error diagnosis while affected IOs continue**
- **Support for live replacement of disks**
- **Plus more...**



### IBM Spectrum Scale RAID (GNR) – end-to-end checksum

---

- When an NSD client writes data, an **8 byte checksum** is calculated and appended to the data for transport over the network. On reception, Spectrum Scale RAID verifies the checksum and stores it with data, and logs a version number in metadata
- When Spectrum Scale RAID reads data from disk, checksum and version number are validated. If validated, the data & checksum are transported to the NSD client. If not validated, the data is reconstructed along with the checksum.

### IBM Spectrum Scale RAID (GNR) – Disk Hospital

---

- **Asynchronously diagnoses errors and faults in the storage subsystem**
- **On pdisk I/O operation timeout, I/O error, or a checksum mismatch, the suspect pdisk is immediately admitted into the disk hospital**
- **First determination is whether the error is pdisk caused or path to pdisk**
- **Spectrum Scale RAID uses vdisk redundancy codes to recover data**
- **Disk Hospital tracks internal health metrics for each pdisk e.g. Relative Performance and logs internal remapping events**
- **Below threshold, pdisk marked “slow” and prepped for replacement**
- **Records FRU and physical hardware location and turns on identity light on storage drawer and drive to be replaced**

### ESS Unique Advantages beyond GNR

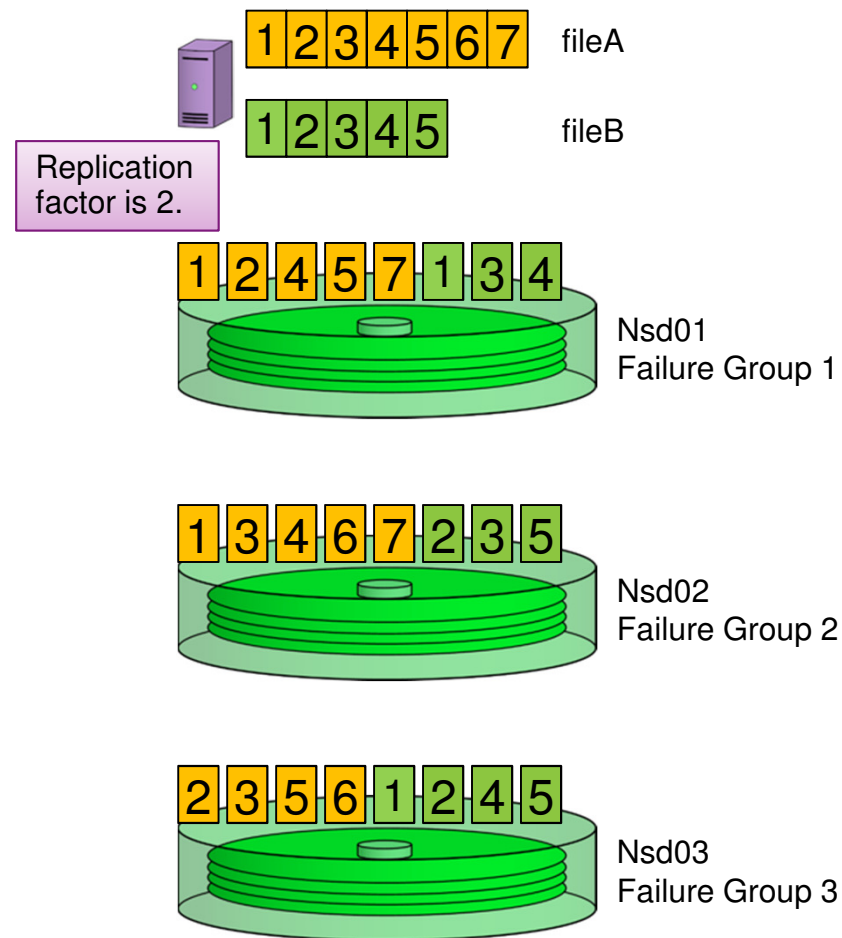
---

- Scripts to ease deployment, upgrade & testing
  - xCat
  - Firmware updates
  - HMC
  - Drivers (e.g. SAS, Networking)
  - Kernel and other OS patches
- GUI & Sensor
- ESS “samples” scripts
  - Client config
  - GNR Callback



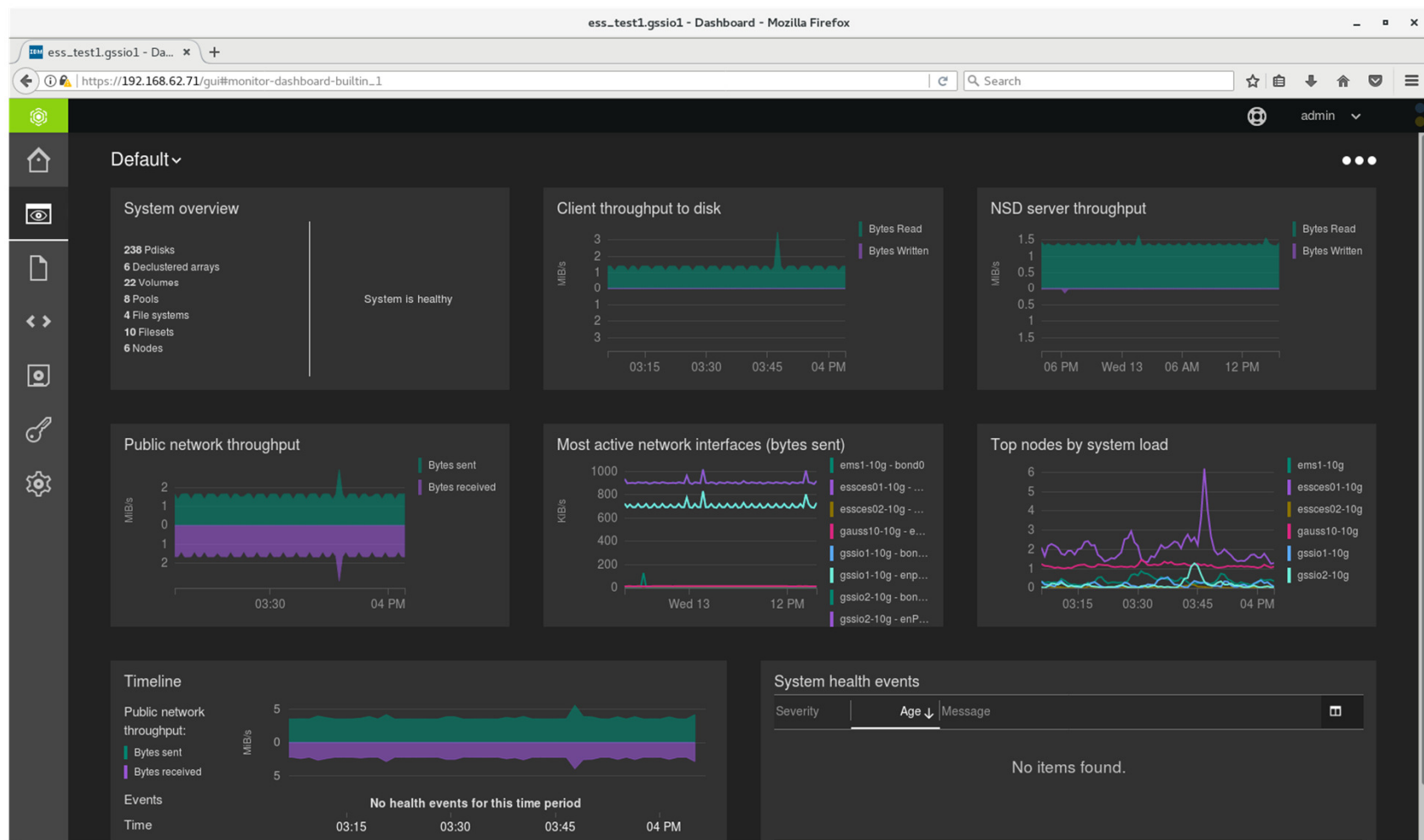
## Pools vs. failure groups

- A **pool** is a class of storage device. Every NSD is in a pool.
- A **failure group** indicates the failure domain of an NSD (often linked to the location). Every NSD is in a failure group.
- Every file has a replication factor associated with it:
  - The replication factor,  $r$ , may be 1, 2, or 3.
  - Every block of every file has  $r$  replicas, each in the same pool, but different failure groups.
- Every block of a file is in the same pool.
- Policies can set and change the replication factor and the pool of a file, but can not control which failure groups are used.



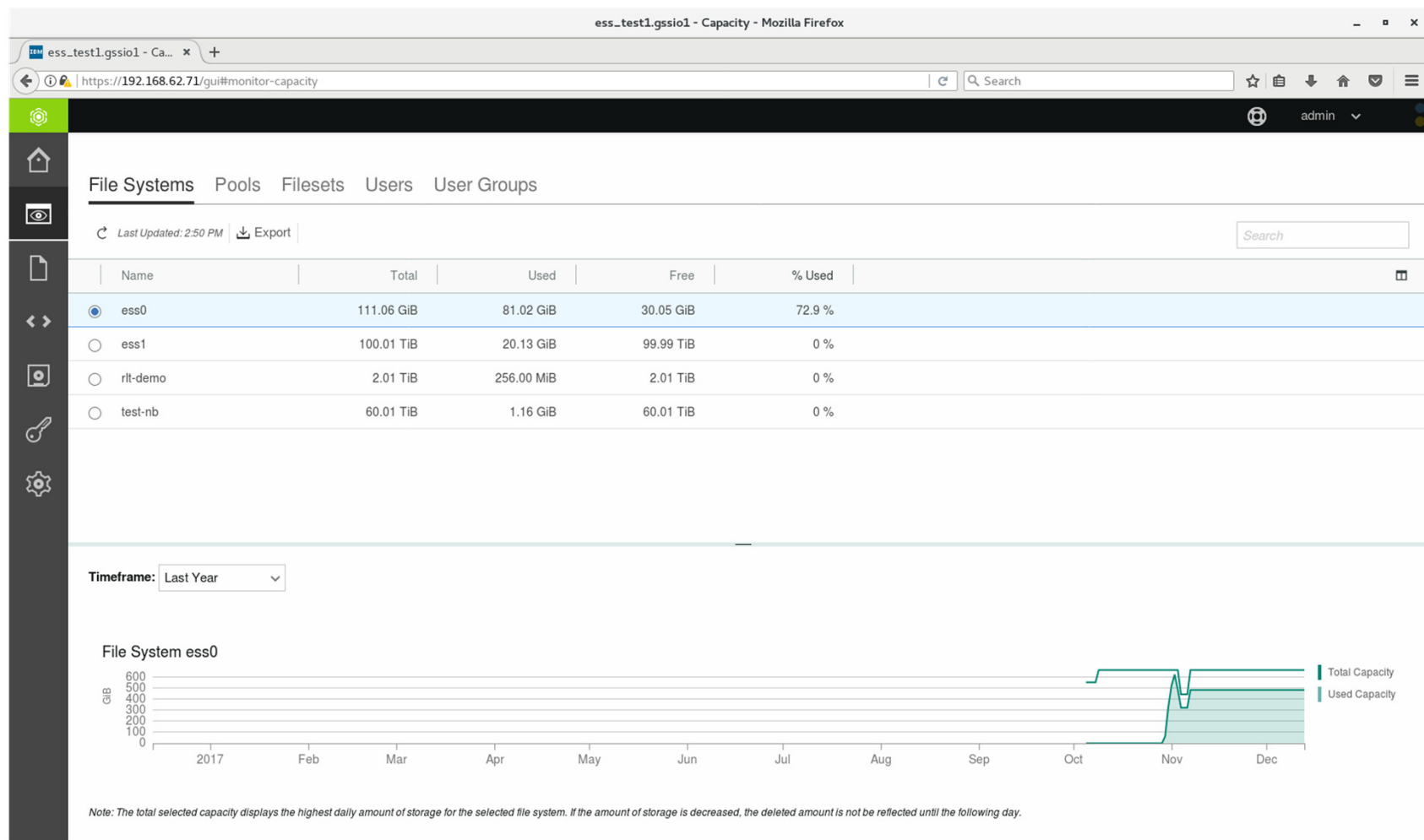
# Washington Systems Center - Storage

## GUI: Monitoring → Dashboard



# Washington Systems Center - Storage

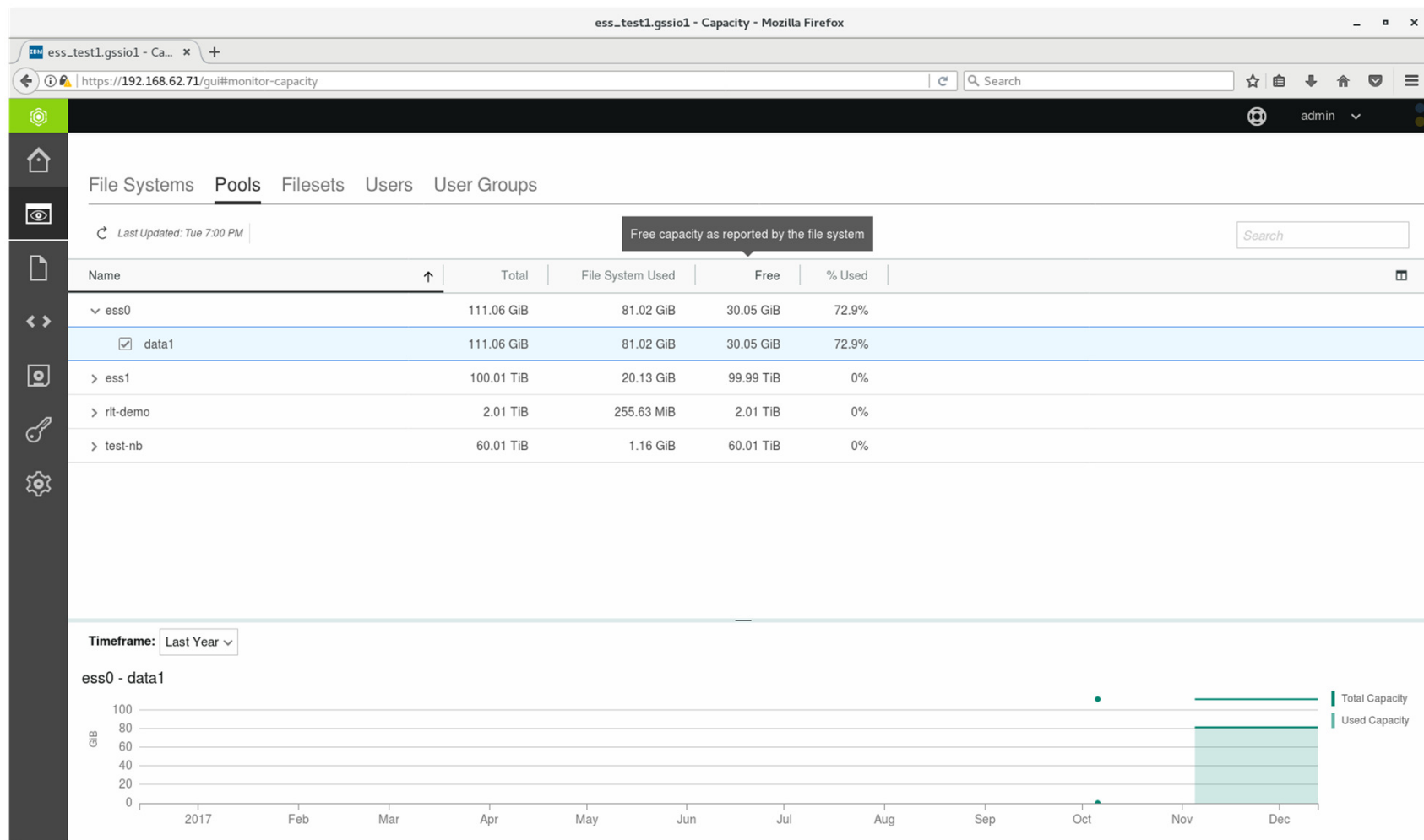
## GUI: Monitoring → Capacity (File Systems)





# Washington Systems Center - Storage

## GUI: Monitoring → Capacity (Pools)



### Advantages of ESS Spectrum Scale RAID

---

- **Better fault tolerance**
  - Data tracks/stripes are distributed as widely as possible across fault zones to provide better fault tolerance
  - ESS provide enclosure/drawer and disk level fault tolerance
- **End to end checksum**
  - Spectrum Scale RAID maintains checksum of data blocks from the client to the blocks on the disk and validates at every point, eliminating the chances of silent data corruption or data loss
- **Asynchronous error diagnosis while Application IOs continue**
  - Special feature called “Disk Hospital” which monitor health of disks
  - If media error - Verify and restore if possible
  - If path problem - Attempt alternate paths
  - If unresponsive or misbehaving disk- Power cycle disk
  - Monitor for slow disk in Declustered Array and take corrective action



## Priority Use Cases for IBM Spectrum Scale

---

***Delivering the highest levels of scalability and performance***

### **Resiliency**

1. Back-up / Restore

### **Lifecycle Management**

1. Archive
2. Information Life Cycle Management

### **Transformation**

1. “Data Ocean” Unified Storage
2. Data-intensive Technical Computing
3. Big Data and Analytics
4. Artificial Intelligence (AI)
5. Selected Solutions
  - ISV Solutions
  - Industry Solutions



## Resiliency: Back-up and Restore Use Case

---

### Representative Customer request:

- Help completing backups within a backup window(RPO)
- Assistance completing the restores within a restore window(RTO)
- Improved management and utilization of back-up application silos
- Guaranteed backup data integrity

### IBM Solution:

- IBM Spectrum Scale Client for Spectrum Protect Server
- IBM Elastic Storage Server as centralized data pool

### Customer benefit:

- Faster back-up and restore of their business data (within a limited time period) based upon parallel access to multiple Spectrum Protect Servers
- More cost-effective/higher utilization backup storage target
- Low operational load on the I/T resource team (automated, policy driven tiering)
- Seamless migration of data without downtime; lower operational risks
- All the data across all the tiers including Tape available in one filespace / namespace making data recall (recovery) and compliance-driven data discovery fast and efficient. It can be the difference between passing and failing a legal or financial compliance audit.

### Customer Example

*Healthcare  
Company*

**5:1 cost  
advantage** over  
Data Domain

**Performance  
Scalability  
Extensibility**

# Lifecycle Management: Archive Use Case

---

### Representative Customer request:

- Need for new and smarter storage system
- Data management for over 5TB per day per vessel of raw seismic data plus 10-100 times more volume of analyzed data
- Requirement per Norwegian regulations for all generations and formats of seismic data to be stored in a national repository

### IBM Solution:

- Fully redundant solution supported by one vendor
- IBM Spectrum Scale and Spectrum Archive software along with tape and disk hardware including the IBM Elastic Storage Server

### Customer benefit:

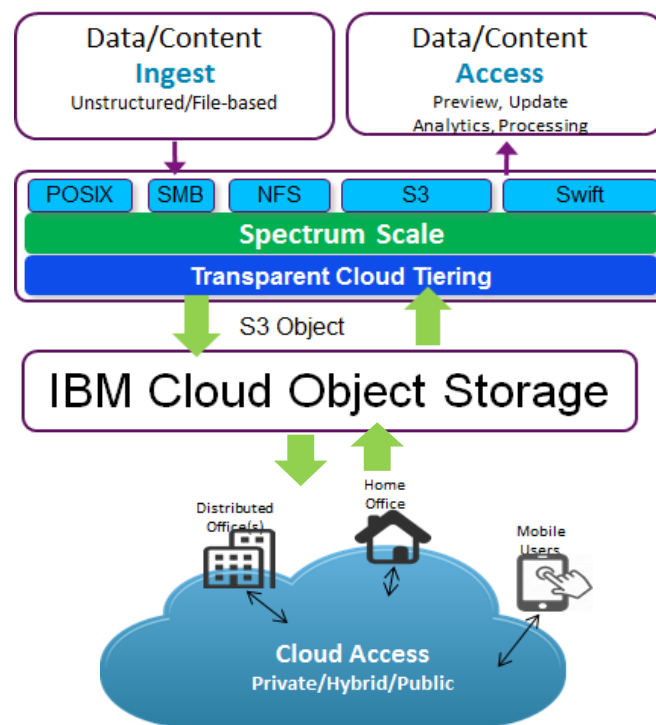
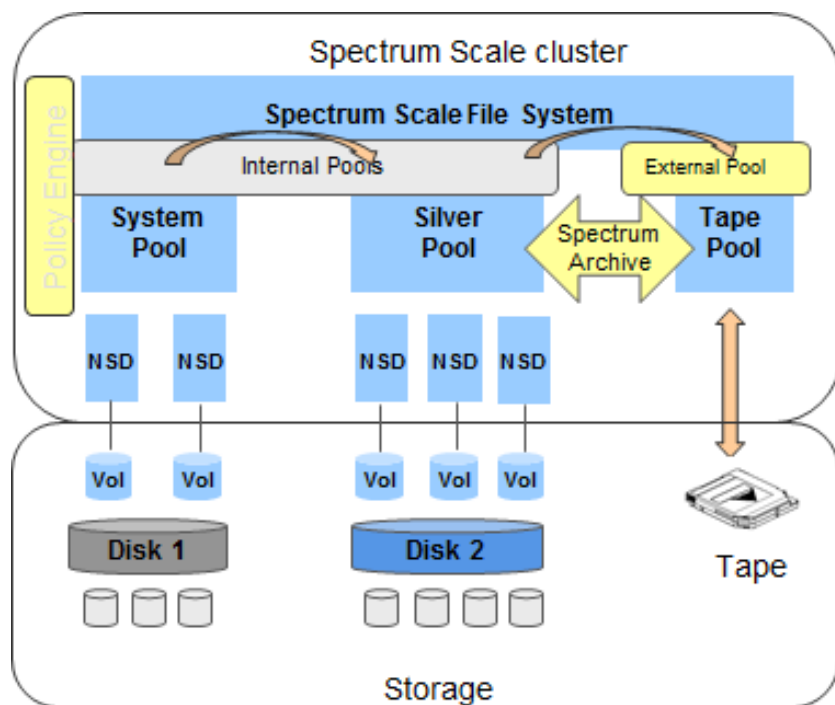
- High performance due to read/write streaming being faster with tape than disk
- Low cost due to tape characteristics including \$/TB and environmental factors
- High levels of automation based on data management policies
- High availability including non-disruptive upgrades

### Customer Example



**Unlimited  
Scalability** to  
accommodate  
growing data  
volumes  
**Lowered  
storage  
costs**

## Spectrum Scale integrates with Spectrum Archive and IBM COS



- Spectrum Archive can be configured as tape tier
- IBM Cloud Object Storage (COS) can be configured as an object / cloud tier
- Spectrum Scale policies can be used to automatically migrate files from disk to tape or to IBM COS
- After migration, the file or object remains visible in Spectrum Scale and can be recovered



# Transform: “Data Ocean” / Unified Storage Use Case

---

### Representative Customer request:

- For my new generation workloads, I have a data storage, data integration, application data workflow problem
- I can't implement modern competitive advantage applications, because I'm blocked by issues with application data integration, data accessibility
- My data is stored in too many different places, with no integration, no common ability to manage
- I need a enterprise class, truly global shared data infrastructure

### IBM Solution:

- Creation of a Federated Data Model (aka Data Ocean) infrastructure
- Based on Spectrum Scale software and Elastic Storage Server (ESS) hardware
- To provide high performance, low-cost, scalable centralized storage

### Customer benefit:

- Construct a Data Optimization strategy for unstructured file and object data for the enterprise
  - ...that enables low latency big data analytics for new generation cloud, mobile and converged applications
  - ...by implementing a continuous data access, shared enterprise data infrastructure for multiple applications, with premium performance, scale, automated data management efficiency and effectiveness

### Customer Example

*Major Financial  
Services Company*

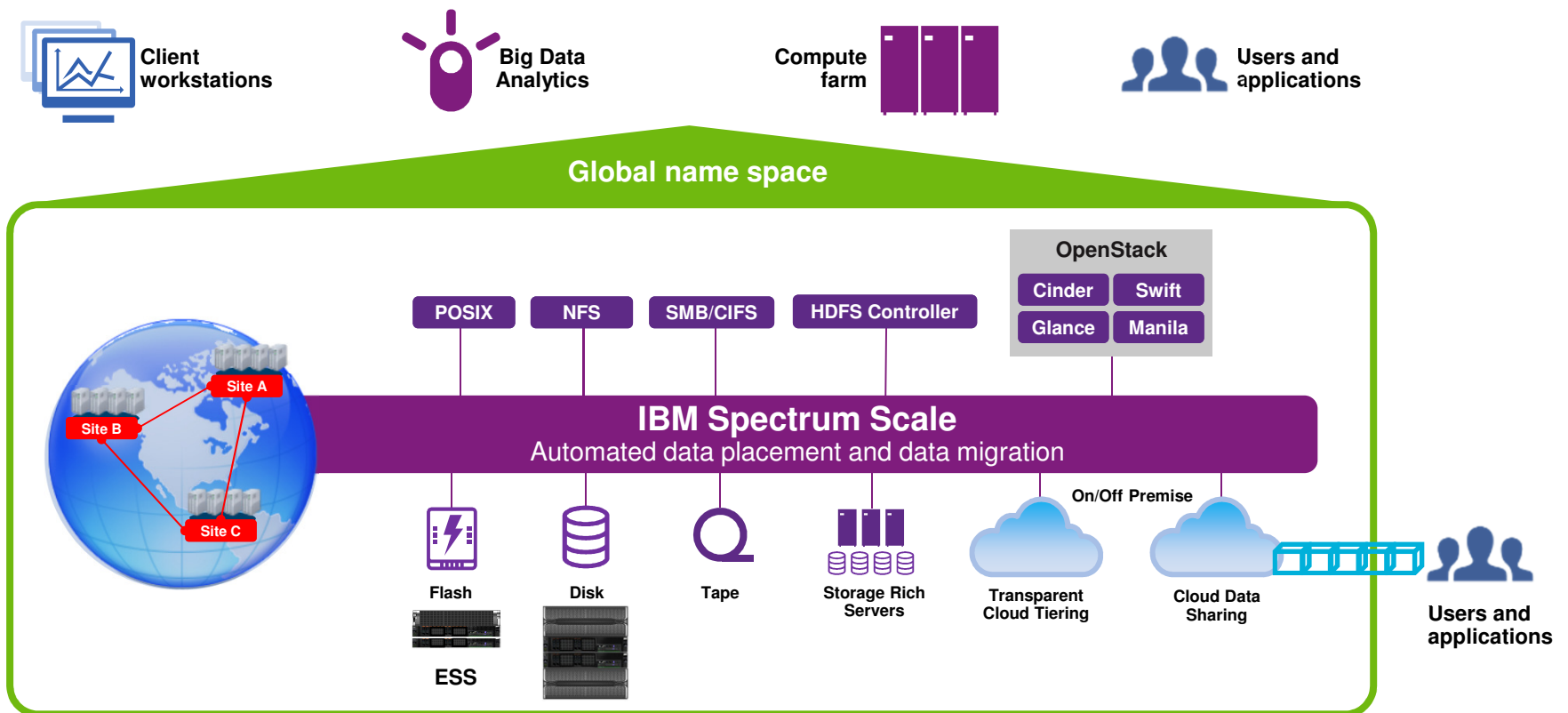
**Up to \$18 M in  
storage cost  
savings**

over 5 years  
versus incumbent

**Break even in  
9 months**

## “Data Ocean” / Unified Storage

IBM Spectrum Scale and ESS enable existing and next generation workloads to leverage a “data ocean”.



# Transform: Data-intensive Technical Computing Use Case

---

### Representative Customer request:

- High performance sequential throughput
- For leading edge and emerging Industry Analytics applications
  - such as oil/gas, healthcare, genomics, physics research, defense/intelligence, EDA and more
- Requiring up to multi-petabyte data scale, low cost, extreme performance, data manageability

### IBM Solution:

- Spectrum Scale software provides high performance enterprise shared data infrastructure, accessible via multiple protocols,
- Proven for 15 years in the most demanding supercomputers
- #1 HPC Storage Mgmt Software - "The Top of All Things in HPC" per Intersect360 Jan 2017
- Elastic Storage Server as packaged storage solution for petascale supercomputing

### Customer benefit:

- Combine diverse structured and unstructured data types for high performance, low latency analytics for business insight
- Premium performance, premium scale, premium manageability
- ESS with Spectrum Scale Native RAID provides premium data reliability, premium data integrity, consistent persistent performance

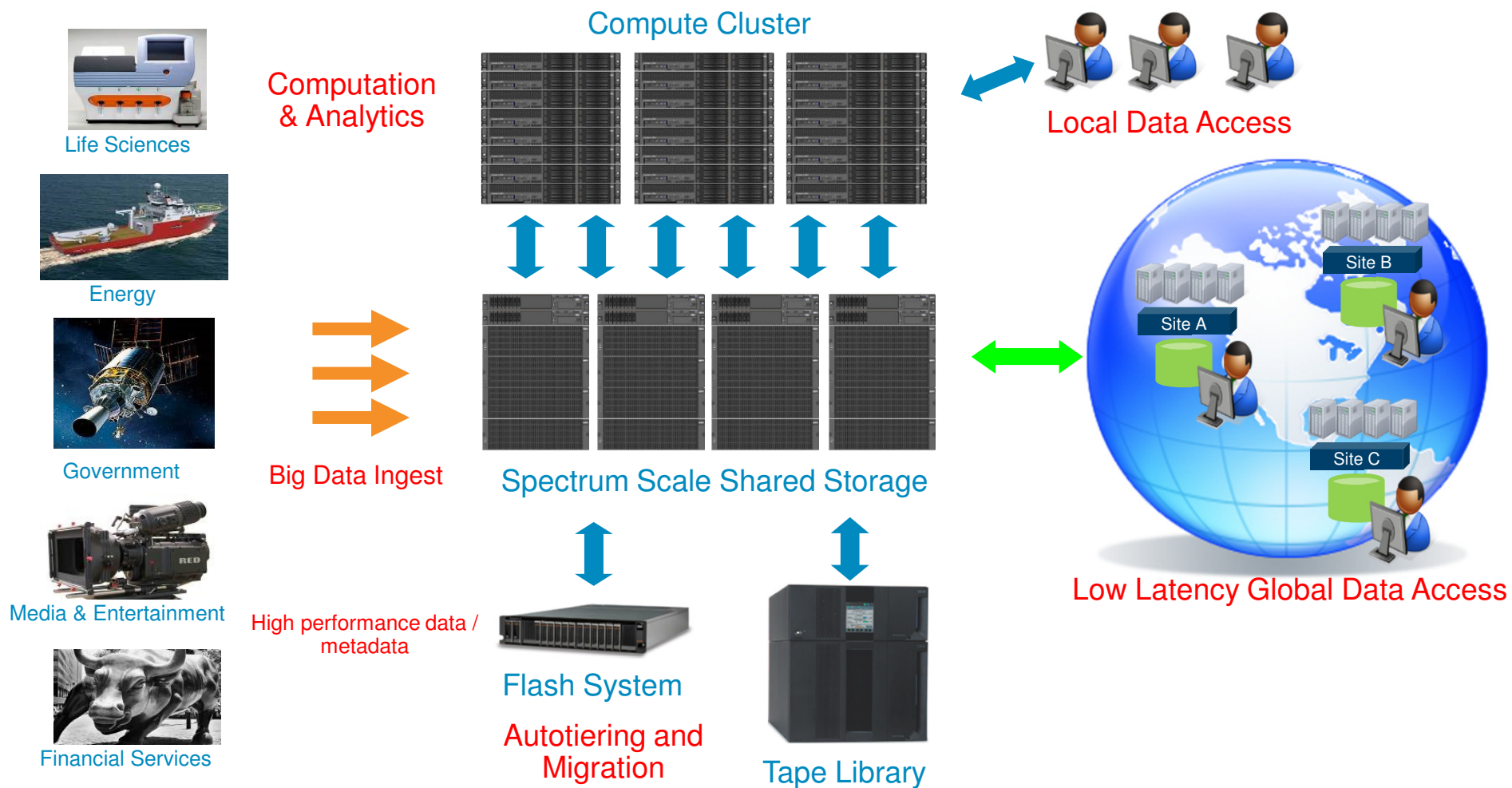
### Customer Example



**10x better  
performance**  
(on the same  
hardware)

**Total Cycle  
Time  
reduction**

## Data-intensive Technical Computing



# Transform: Big Data and Analytics Use Case

---

### Representative Customer request:

- Typically, Hadoop/HDFS setups are inefficient in terms of data-copies (sometimes up to 8 copies of data!)
- Lack of redundancy/enterprise RAS
- Usability (Ease-of-Use)

### IBM Solution:

- IBM Spectrum Scale instead of HDFS (Hadoop Filesystem)
- Integration of existing Hadoop-nodes in one common architecture

### Customer benefit:

- One file system with a global namespace for easy-scalability and very efficient Meta Data mechanisms mean very fast analysis on very large volumes of data = shorter time to better results
- Integrated backup and automated policy driven tiering: less Administrative expense
- Elimination of ETL to save time and money in supporting workflow of applications
- IBM proven RAS
- No need to change API/program interfaces (Java binding) - no hidden costs
- Lower \$\$ investment in raw storage capacity because more efficient data protection (including distributed RAID) than traditional Hadoop

### Customer Example

*Major Automobile  
Company*

**70% reduction**

in Hadoop infrastructure  
required

**Independent scaling  
of storage & compute**

**Analytics in place**

# Transform: Artificial Intelligence (AI) Use Case

---

### Representative Customer request:

- Maintain sufficient data throughput to feed multiple GPUs per compute node
  - to avoid having these expensive resources being idle the majority of the time
  - to speed time to results
- Enable low latency for inference – to quickly read random and sometimes small data
- Enable distributed accelerated cache for Training – to read the same data over and over again (i.e. 'reference' data)

### IBM Solution:

- IBM Spectrum Scale including Client code and LROC (Local Read Only Cache)
- ESS delivering by far the fastest per HDD performance in the industry, and ensuring data availability and continued performance even in the face of drive failures

### Customer benefit:

- High throughput and low latency
- Scale-out with virtually no limits
- Flexible deployment models
- Geo-dispersed option to support collaboration
- Efficiency and Security
- Flexibility to support diverse applications and access methods

### Customer Example

*Major Technology  
Supplier*

### **Extreme Performance**

for benchmarking of  
AI/ML/DL workloads

**High read throughput  
per individual client  
node**

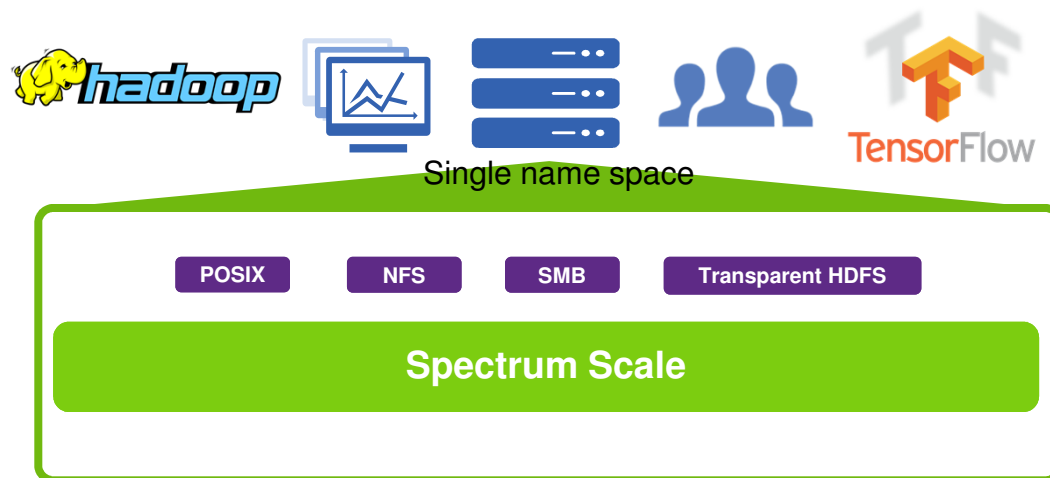
**Low latency for random  
I/O operations**



## Transform: Artificial Intelligence (AI) Use Case

Spectrum Scale  
Client to read data  
at 24GB/sec with  
two EDR IB ports vs.  
Lustre at ~2GB/sec,  
Isilon at ~1GB/sec, &  
NetApp at ~0.7GB/sec

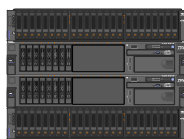
Spectrum Scale  
Local Read Only  
Cache) (LROC) can  
speed training by up to  
800%



Spectrum Scale has  
filesystem overhead  
of ~15usec and  
communication code  
overhead of ~15usec so  
end-to-end latency of  
~30usec (excluding  
media latency) for a  
remote 4k random I/O  
vs. Lustre at ~500usec,  
Isilon at ~700usec, and  
NetApp at ~200usec

All Flash ESS with ESS 5.3 software  
delivers random I/O performance of  
almost 1 million 8k random reads/sec

High Performance  
All Flash ESS  
GSxS Models







High Capacity  
GLxS Models

ESS delivers 105 MB/sec per HDD for  
a 100% random full block workload vs.  
DDN at ~70 MB/sec, and Netapp or  
Isilon in the 30MB/sec range per disk

# Washington Systems Center - Storage

## Selected ISV Solution Use Cases

| Example ISV  | Solution  | Benefit to Client  |
|--|---|--|
|                               | Enterprise File Synch and Share with Spectrum Scale and ownCloud                        | 100,000s of Users: Private or Hybrid Clouds, Secure, Easy to Manage at Scale   |
| <br>The Open Platform Company | Digital Video Surveillance: Spectrum Archive + Milestone xProtect Corporate             | Digital Video Archives Requiring PBs of Storage. Achieve up to 26X Lower TCO than Spinning Disk  |
|                               | Managed Data Stores for Sharepoint Blobs with Spectrum Scale and Storwize V7000 Unified | Break through the Capacity Limits of Sharepoint with Avepoint & IBM Spectrum Storage   |
|                             | SAS Analytics (SAS Grid) Running on Spectrum Scale with XIV & FlashSystem               | Performance at Scale for SAS Environments: Spectrum Scale provides a flexible virtual storage option for distributed SAS applications. |



# Washington Systems Center - Storage

## Industry Workload Examples



Chip Design – Cypress Semiconductor:  
<https://www.youtube.com/watch?v=Kn90TkMvkfw>  
<https://www.youtube.com/watch?v=TfYDeP1F-CY>



Finance - Viewpointe:  
<https://www.youtube.com/watch?v=SgJc0finOsg&t=52s>



Real time analytics – Red Bull Racing:  
[https://www.youtube.com/watch?v=r\\_DD65KnYO8](https://www.youtube.com/watch?v=r_DD65KnYO8)



IBM  
Spectrum  
Scale



SuperComputing 2017:  
<https://www.youtube.com/watch?v=xVzPkUIVDh4>



Healthcare - SIDRA: <https://www.youtube.com/watch?v=BdoU5naRiyo>  
<https://www.youtube.com/watch?v=BdoU5naRiyo&t=8s>



NUANCE

Speech and Imaging - Nuance Communications:  
<https://www.youtube.com/watch?v=Cq7daARviak>



Media/Entertainment – new Atlanta stadium:  
<https://www.youtube.com/watch?v=GMarZjIviYs>  
<https://www.youtube.com/watch?v=n3sdQypcxok>  
[https://www.youtube.com/results?search\\_query=ibm+atlanta+falcons+stadium](https://www.youtube.com/results?search_query=ibm+atlanta+falcons+stadium)

- **Banking and Financial Services** – fraud prevention, check truncation, regulatory compliance
- **Insurance** – Policies & underwriting
- **Media and Entertainment** – radio and television, Remote DVR
- **Petroleum** – Seismic data ingest and analytics
- **Smarter cities** – Video surveillance
- **Automotive** – Crash test recording, autonomous driving, etc.
- **Defense / Military** – flight recording, threat assessment
- **Public / Government** – Law enforcement, recording of events, personnel cameras, satellite Image archival
- **Telecom** – Call Detail Records
- **Life Sciences and Pharmaceuticals** – Genomics
- **Hospitals** – Medical images

# What is CORAL?



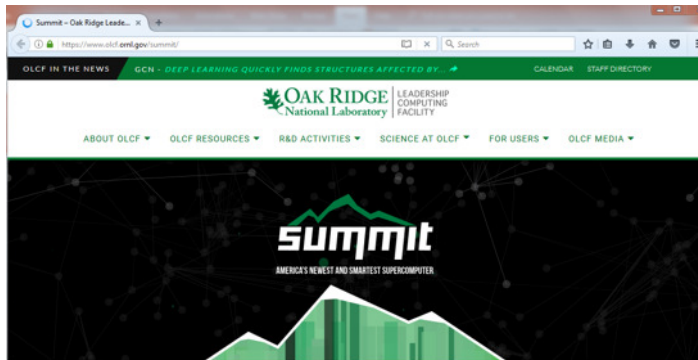
- Collaboration of Oak Ridge, Argonne, and Livermore (U.S. Department of Energy National Laboratories)
- Joint procurement activity to build state-of-the-art high-performance computing technologies
- DoE supercomputers run the most complex energy, science, and security mission applications
- DoE investments advance HPC technologies to demonstrate continued U.S. leadership in the field
- New systems will deliver much greater capabilities and energy efficiency
- Key milestone in capacity and efficiency toward exascale
- Extreme scale computing aims to deliver 20-40x the capabilities of today's computers with a similar size and power footprint

Last day of #SC17 Exhibit Hall. Don't miss video of US #DoE #Summit SuperComputer deployment at #IBMSC17 booth 1525 :-)) - at Colorado Convention Center



<https://www.youtube.com/watch?v=OoaiYVQulhA>  
<https://www.youtube.com/watch?v=J6oE5Knk8K8>

## Spectrum Scale V5 and ESS: Performance engineering matters



<https://www.olcf.ornl.gov/summit/>

**Imagine you need to deliver  
the following Spectrum Scale goals :**


- **2.5 TB/sec single stream IOR**  
as requested from ORNL
  - **1 TB/sec 1MB sequential read/write**  
as stated in CORAL RFP
  - **Single Node 16 GB/sec sequential read/write**  
as requested from ORNL
  - **50K creates/sec per shared directory**  
as stated in CORAL RFP
  - **2.6 Million 32K file creates/sec**  
as requested from ORNL
- 
- **What innovations in Storage would result from this?**


## IBM High Performance Analytics Portfolio

### CORAL

- **Pioneering** Exascale HPC *meets Artificial Intelligence*
- IBM **delivering** largest supercomputer on Earth
- OpenPOWER **Collaboration**  
→ mini-CORAL *for everyone*



- 4600 AC922 P9 nodes
- 200+ Petaflops
- ~2.5 TB/sec
- Greenest 


 Lawrence Livermore  
National Laboratory  
129+ Petaflops



### Technology Leadership

Breaking barriers with open **data-centric innovation**



**Heterogeneous** and hybrid **Cloud** ready 

With a future paved by **IBM Research**



5nm



TrueNorth

**IBM Q**

### Architected for AI

Record-breaking **performance & scalability** for DL training

**PowerAI**



Meets **enterprise-grade tooling** for end-to-end AI

 **AI Vision**

 **DL Insight**

With a **comprehensive portfolio** designed for data **workflows**.





## Spectrum Scale/ESS – Leading Edge Scale / Performance

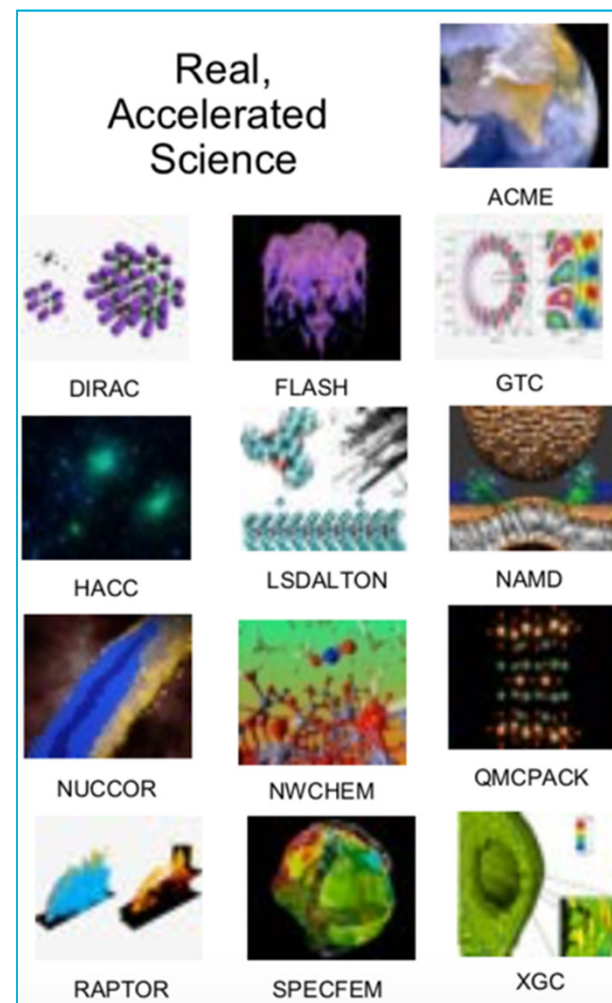


### Summit System

- 4608 nodes, each with:
  - 2 IBM Power9 processors
  - 6 Nvidia Tesla V100 GPUs
  - 608 GB of fast memory
  - 1.6 TB of NVMe memory
- 200 petaflops peak performance for modeling and simulation
- 3.3 ExaOps peak performance for data analytics and AI
- **Spectrum Scale V5**
- **77 Elastic Storage Servers**



**2.5 TB/sec**  
Throughput to storage  
architecture  
**250 PB**  
HDD storage capacity



### A new level of storage performance and efficiency

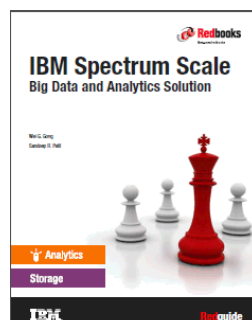
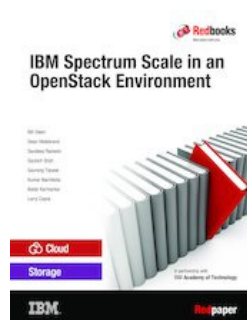
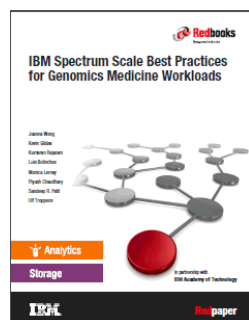
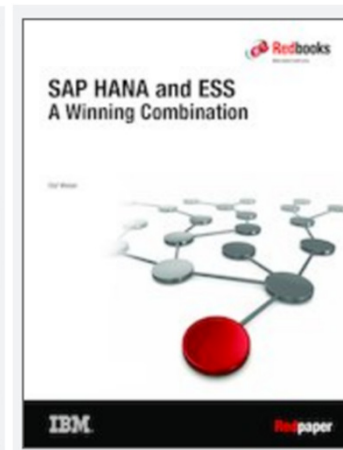
---

- **Dramatic improvements in I/O performance**
- **Support for newest low-latency, high bandwidth hardware such as NVMe**
- Significantly reduced communication latency between nodes
- **Improved performance, space efficiency for mixed workloads**
  - Small and large block size workloads running simultaneously in same file system
  - Optimize large block performance via new 4MB default block
  - Simultaneously optimize small file space efficiency with variable sub-block size
- **Improved IOP/sec and metadata performance**
  - IOP/s can improve 3x to 5x over previous releases\*



\* Performance numbers are estimates based on IBM internal lab tests and are subject to verification

## Elastic Storage Server Redbooks



Redbook Search <http://www.redbooks.ibm.com/redbooks.nsf/searchsite?SearchView>

## Accelerate with IBM Storage Webinars

---

### The Free IBM Storage Technical Webinar Series Continues in 2019...

*Washington Systems Center – Storage* experts cover a variety of technical topics.

Audience: Clients who have or are considering acquiring IBM Storage solutions. Business Partners and IBMers are also welcome.

To automatically receive announcements of upcoming Accelerate with IBM Storage webinars, Clients, Business Partners and IBMers are welcome to send an email request to

[accelerate-join@hursley.ibm.com](mailto:accelerate-join@hursley.ibm.com).

Located in the Accelerate with IBM Storage Blog:

<https://www.ibm.com/developerworks/mydeveloperworks/blogs/accelerate/?lang=en>

Also, check out the WSC YouTube Channel here:

[https://www.youtube.com/channel/UCNuks0go01\\_ZrVVF1jgOD6Q](https://www.youtube.com/channel/UCNuks0go01_ZrVVF1jgOD6Q)



### 2019 Webinars:

**January 24** - Discover how IBM Storage Insights and Storage Insights PRO can transform your IBM Storage Support Experience

**January 31** - Elastic Storage Server (ESS) Architecture and use cases

**February 28** - Data Protection and High Availability in Spectrum Virtualize

**Register Here:** <https://ibm.webex.com/ibm/onstage/g.php?MTID=ebe25786da3342f1336b1135c443225f7>

**Thank You!**



## Trademarks and Disclaimers

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries. IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce. Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office. UNIX is a registered trademark of The Open Group in the United States and other countries. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom. Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Other product and service names might be trademarks of IBM or other companies. Information is provided "AS IS" without warranty of any kind.

The customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Prices are suggested U.S. list prices and are subject to change without notice. Starting price may not include a hard drive, operating system or other features. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Photographs shown may be engineering prototypes. Changes may be incorporated in production models.

© IBM Corporation 2018. All rights reserved.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Trademarks of International Business Machines Corporation in the United States, other countries, or both can be found on the World Wide Web at <http://www.ibm.com/legal/copytrade.shtml>.

ZSP03490-USEN-00