IBM Spectrum Storage

Software Defined Data Protection

IBM Spectrum Protect™ with IBM Spectrum Scale™

Version 6.9

Agenda

- **Introduction to Spectrum Protect and Spectrum Scale**
- Why Spectrum Protect with Spectrum Scale
- Configuration options
- Positioning and Advantages

IBM Spectrum Storage™ Family

Securely 'unboxing' storage to revolutionize data economics

Family of Storage Management and Optimization Software

IBM **Spectrum Control**	Analytics-driven data management to reduce costs by up to 50 percent
IBM **Spectrum Protect**	Optimized data protection to reduce backup costs by up to 53 percent
IBM **Spectrum Archive**	Fast data retention that reduces TCO for active archive data by up to 90%
IBM **Spectrum Virtualize**	Virtualization of mixed environments stores up to 5x more data
IBM **Spectrum Accelerate**	Enterprise storage for cloud deployed in minutes instead of months
IBM **Spectrum Scale**	High-performance, highly scalable storage for unstructured data

Any Storage

FlashSystem

Private, Public or Hybrid Cloud

IBM Spectrum Protect™ – IBM Tivoli® Storage Manager (TSM)

Comprehensive backup and recovery suite for physical, virtual and cloud environments

On Premises

Microsoft Exchange · SAP HANA · DB2 · Microsoft SQL Server · ORACLE · Lotus Notes

Physical / Virtual

vmware · Microsoft Hyper-V · Linux · UNIX

Hybrid

- ✓ Backup
- ✓ Recovery
- ✓ Archive
- ✓ DR
- ✓ BaaS / DRaaS
- ✓ Object Storage

Services

- ✓ Incremental forever
- ✓ Deduplication everywhere
- ✓ Compression
- ✓ Replication
- ✓ Snapshots
- ✓ Encryption
- ✓ D2D
- ✓ D2D2T
- ✓ D2D2C
- ✓ NAS
- ✓ LAN / SAN
- ✓ WAN

Heterogeneous storage

IBM · EMC² · HITACHI DATA SYSTEMS · NetApp · hp

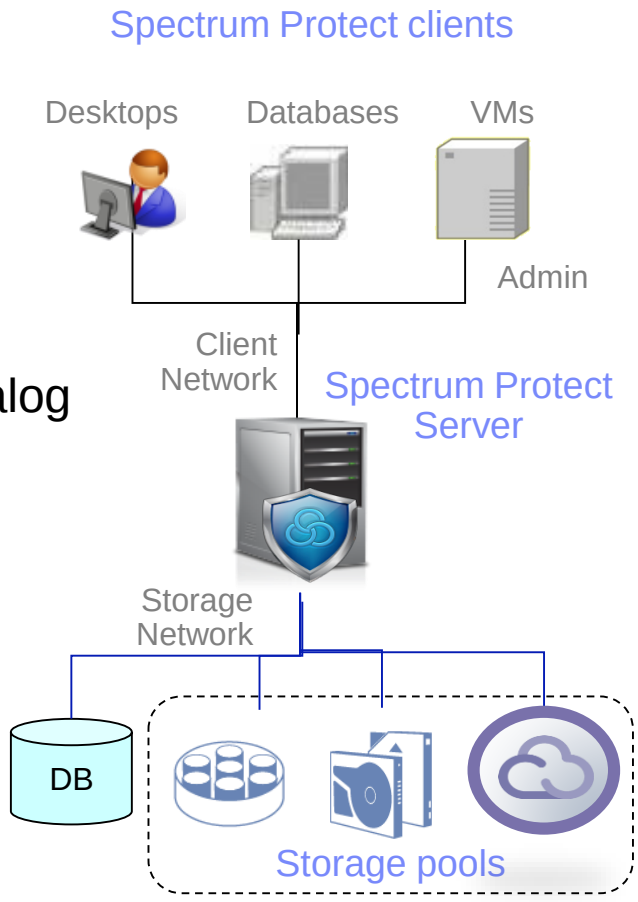
Cloud Storage

SOFTLAYER · amazon web services™ S3 · Windows Azure · Google

Over **20 years experience** protecting some of the world's largest data centers, over **20,000 active clients**

Spectrum Protect Architecture

- Spectrum Protect client – server architecture
  - Backup Client selects and backs up the data
  - Backup Server catalogs data and stores it in storage pools
- Spectrum Protect server has internal database as catalog
- Storage pools can be on a variety of storage media
  - Flash, disk, NAS, optical and tape
  - Storage pool tiering allows automated migration
  - Includes transparent migration between the pools
- Spectrum Protect server provides embedded tools for central management, monitoring and reporting



IBM Spectrum Storage™ Family

Family of Storage Management and Optimization Software

IBM **Spectrum Control**	Analytics-driven data management to reduce costs by up to 50 percent
IBM **Spectrum Protect**	Optimized data protection to reduce backup costs by up to 53 percent
IBM **Spectrum Archive**	Fast data retention that reduces TCO for active archive data by up to 90%
IBM **Spectrum Virtualize**	Virtualization of mixed environments stores up to 5x more data
IBM **Spectrum Accelerate**	Enterprise storage for cloud deployed in minutes instead of months
IBM **Spectrum Scale**	High-performance, highly scalable storage for unstructured data

Any Storage

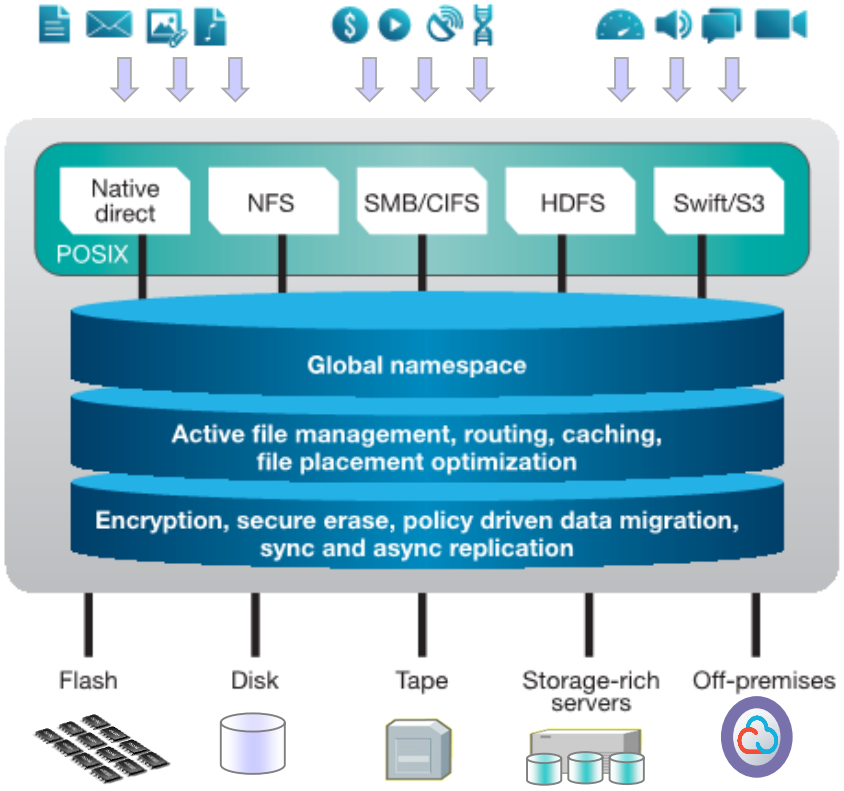
FlashSystem

Private, Public or Hybrid Cloud

IBM Spectrum Scale™

Scale out file system with comprehensive storage services

- **Global and high available name space**
  Unified file, object and block storage
- **Global collaboration**
  Connect anywhere, access any time
- **Data life cycle management**
  Tiering, encryption, compression and immutability
- **Built in protection and fault tolerance**
  Backup, Multi-site replication with site failure tolerance, Native RAID
- **Intuitive graphical user interface**
- **Unmatched reliability and scalability**
  400 GB/sec real cluster throughput



Spectrum scale integrates with different kind of storage and provides transparent placement and migration

Spectrum Scale Deployment Options

Software Only

Software license
Can be deployed on standard hardware

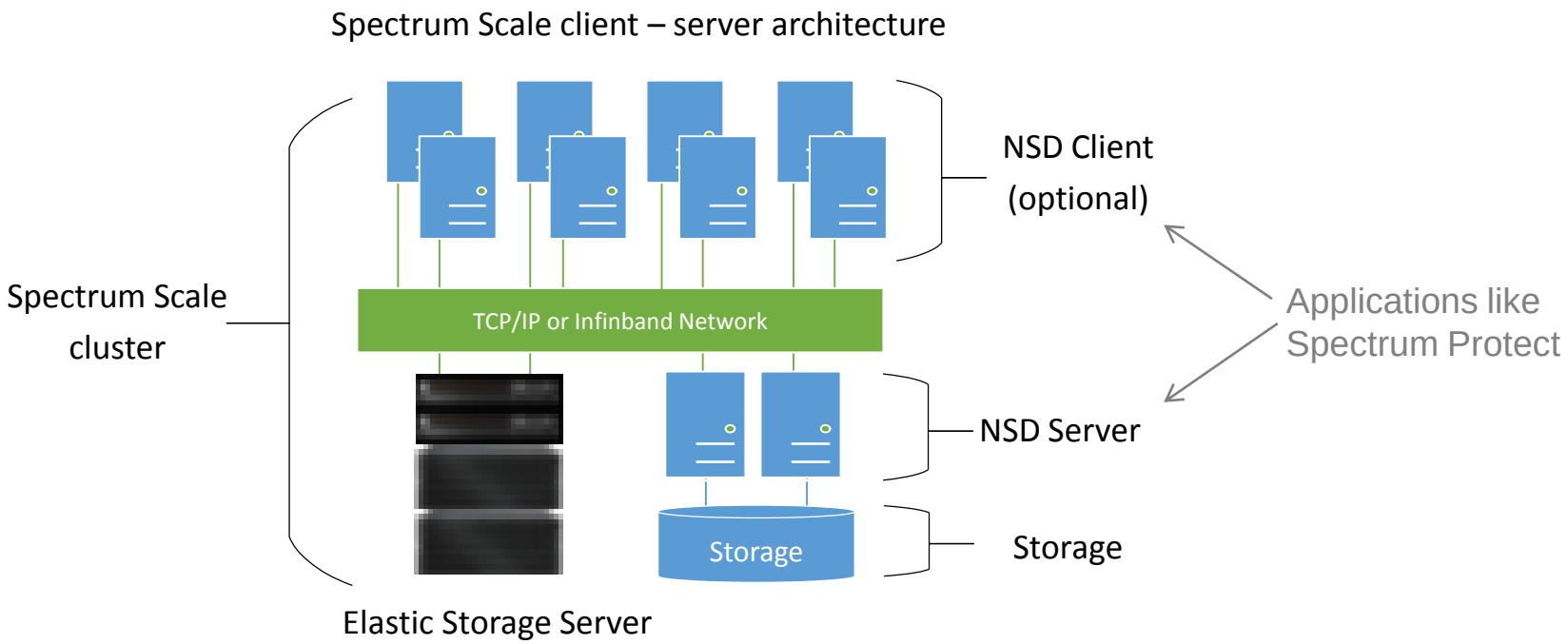
Solution Bundles

Pre-packaged with IBM Spectrum Scale Software, Spectrum Scale RAID, I/O servers, drives, support & subscription

Off-premises

Deploy Spectrum Scale in IBM Softlayer (Whitepaper)
High Performance Computing offerings with Spectrum Scale

Spectrum Scale Architecture



- Storage cluster consists of NSD server and clients (optional)
  - Spectrum Scale can also be established without clients, application can run on NSD server
- NSD servers manage the disks (NSD) and provide file system access
  - ESS contains NSD server and storage
- NSD clients access file systems directly through NSD server

Agenda

Introduction to Spectrum Protect and Spectrum Scale

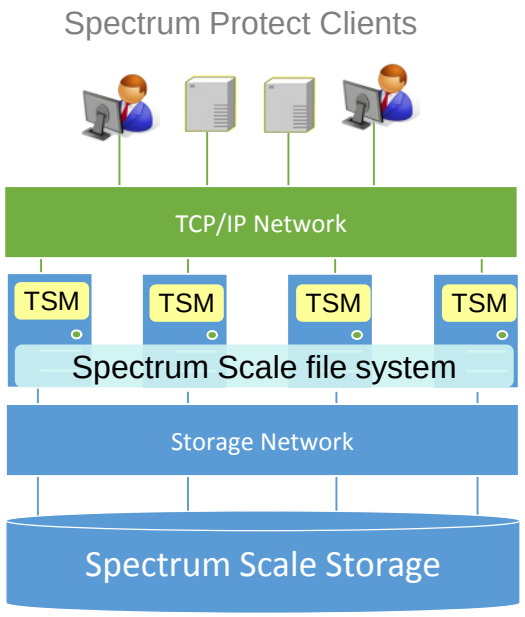
Why Spectrum Protect with Spectrum Scale

Configuration options

Positioning and Advantages

Spectrum Protect on Spectrum Scale - Overview

- **Multiple** Spectrum Protect (TSM) instances store DB and storage pools in a Spectrum Scale file system (GPFS)
  - Spectrum Scale provides global name space for all Spectrum Protect instances
  - Instances share all file system resources
- Spectrum Protect instances run on cluster nodes accessing the file system and disk directly
- Spectrum Scale file systems balances the workload and capacity for all TSM instances on disk
- Provides standardized, scalable and easy to use storage infrastructure for the multiple instances

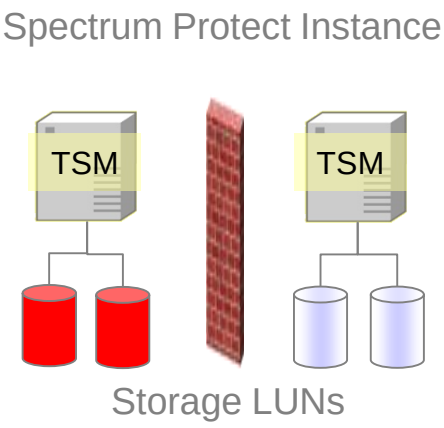


Spectrum Scale storage for Spectrum Protect

Spectrum Protect with and without Spectrum Scale

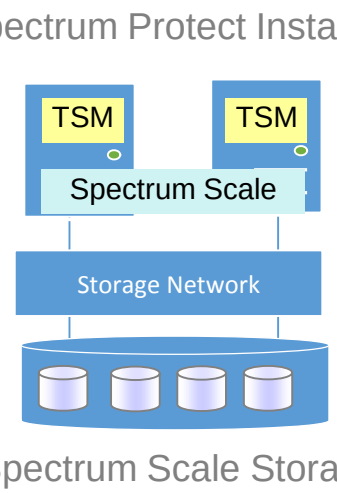
Without Spectrum Scale

- Each backup server has its dedicated LUNs
- Each backup server has its own isolated file system
- Storage islands appear with underutilized capacity
- Capacity and performance management is challenging
- Scaling and performance may impact TSM



With Spectrum Scale

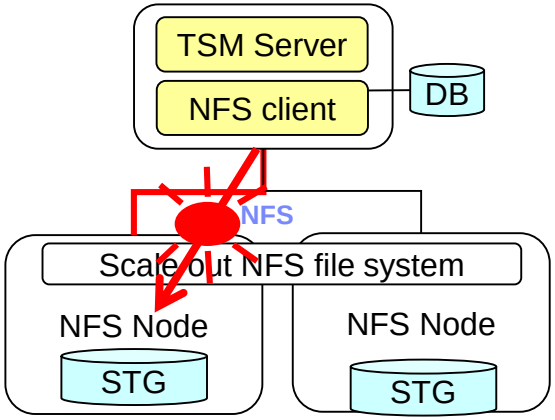
- Backup servers share LUNs and file systems
- Scale capacity and performance seamlessly and transparently to TSM under the shared file system global namespace
- File system replication is included
- Central administration of all storage



Spectrum Scale does not have the NFS bottleneck

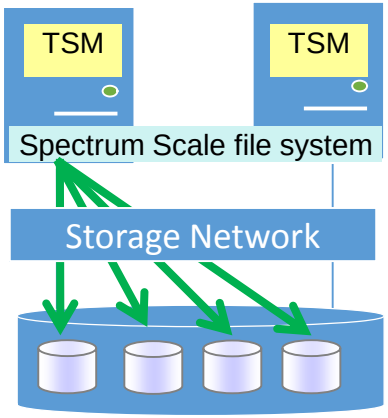
When Spectrum Protect performs I/O on single file it is bound to a **single** NFS node

- Because a NFS session is point-to-point
- I/O performance on single file volume is limited by one NFS server



With Spectrum Scale, Spectrum Protect performs I/O to **all** disk in parallel

- I/O performance on a single file volume is done to all disk in parallel



Key Advantages for Spectrum Protect with Spectrum Scale

Better storage utilization – multiple TSM server share the same storage

Better operational efficiency with one storage for all TSM server

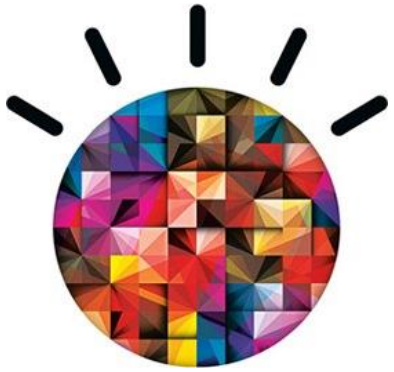
Seamless scalability of storage capacity and performance

Highly scalable performance with parallel striping across all disks

Flexible disaster protection – asynchronous, synchronous or Native RAID

Cost efficiency by utilizing standard infrastructure components

High availability in clustered file system



Agenda

Introduction to Spectrum Protect and Spectrum Scale

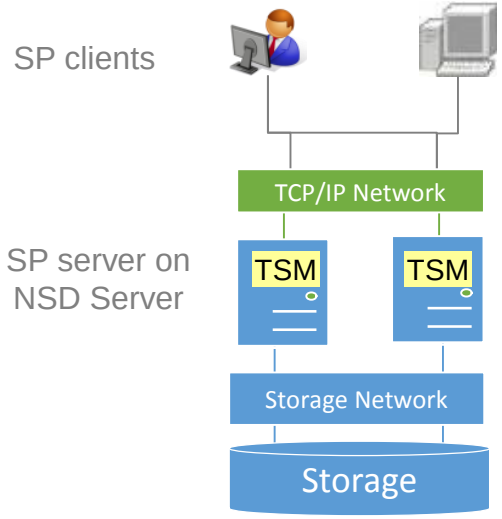
Why Spectrum Protect with Spectrum Scale

Configuration options

Positioning and Advantages

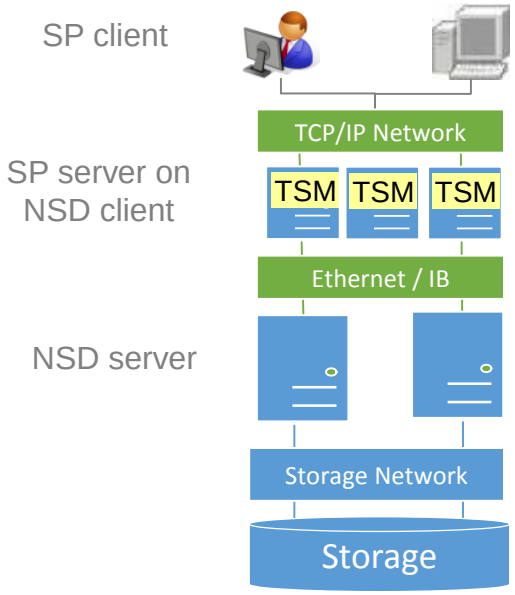
Deployment options

On NSD server



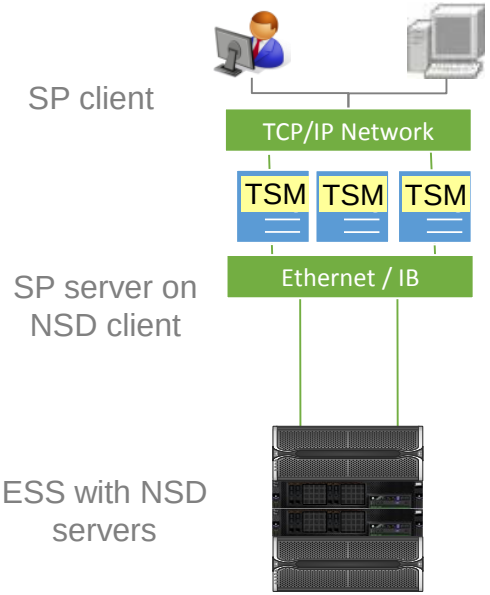
- SP server on NSD servers
- Direct SAN storage access
- Requires less infrastructure
- If Spectrum Scale is used just by SP only

On NSD client



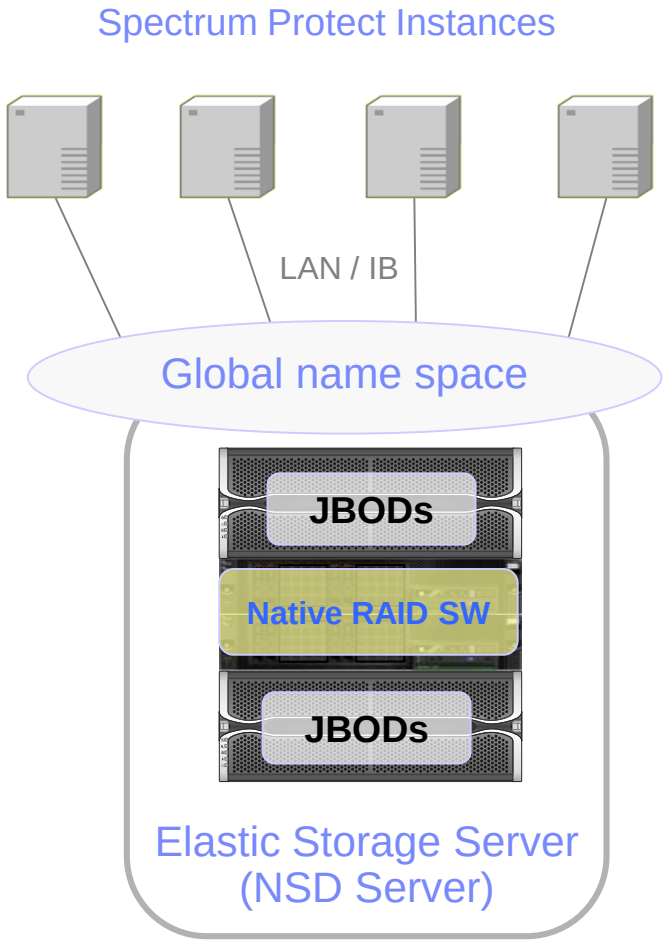
- SP runs on client connected to Scale / ESS NSD server
- SAN or LAN access from SP server to Scale/ ESS server
- Requires more infrastructure
- If Spectrum Scale is used by multiple applications

With Elastic Storage Server



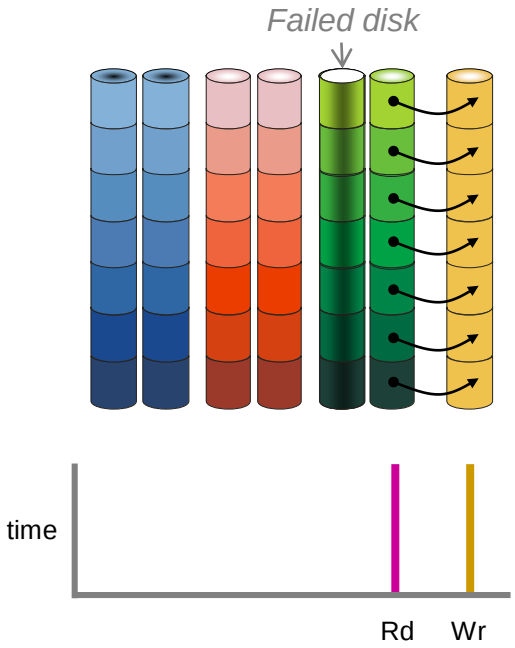
Elastic Storage Server overview

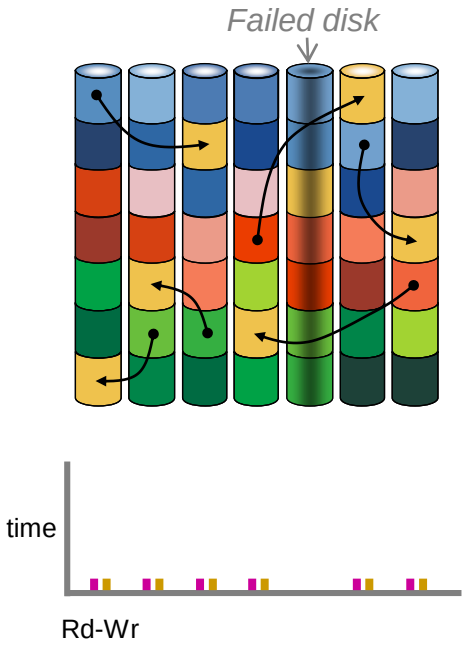
- Spectrum Scale system (pre-packaged)
  - Graphical User Interface
  - 3 Years Maintenance and Support
- Based on Spectrum Scale RAID (declustered)
  - Predictable performance
  - Faster rebuild
  - Low impact during rebuild
  - 2 and 3 fault tolerance configurable
  - End-to-end checksums
- Provides file system for Spectrum Protect
- Different models
  - GS: small and fast (2 – 125 TB)
  - GL: large and scaling (150 – 1530 TB)



Spectrum Scale RAID concept

- Data, parity and spare strips are independently distributed across disk array
  - Supports a arbitrary number of disk per array

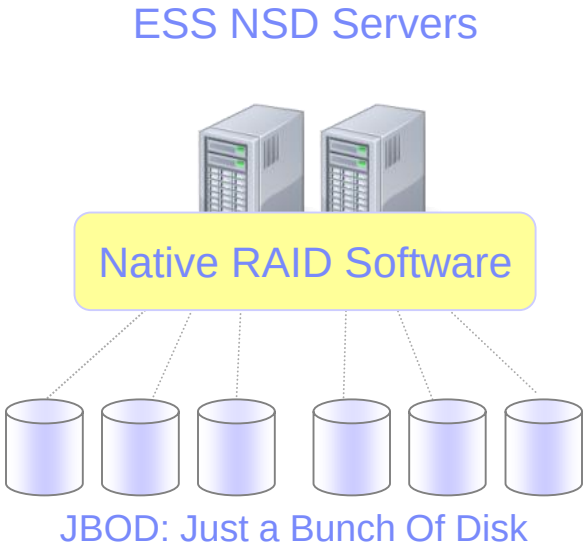




- Disk rebuilds are faster an less impacting with Spectrum Scale RAID

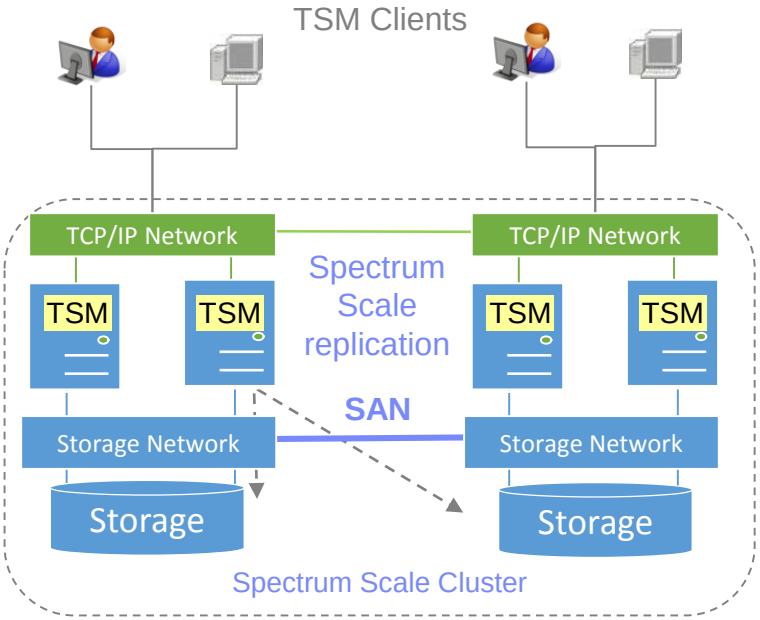
Advantages of Spectrum Scale RAID

- Use of standard and inexpensive disk drives
  - RAID software implemented in GPFS
- Faster rebuild times
  - More disks are involved during rebuild
  - Approx. 3.5 times faster than RAID-5
- Minimal impact of rebuild on system performance
  - Rebuild is done by many disks
  - Rebuilds can be deferred with sufficient protection
- Better integrity and fault tolerance
  - End to end checksum
  - Much higher mean-time-to-data-loss (MTTDL)
    - 8+2P: ~ 200 Years
    - 8+3P: ~ 200 Million Years



Replication options for Spectrum Scale and Spectrum Protect

Spectrum Scale synchronous replication

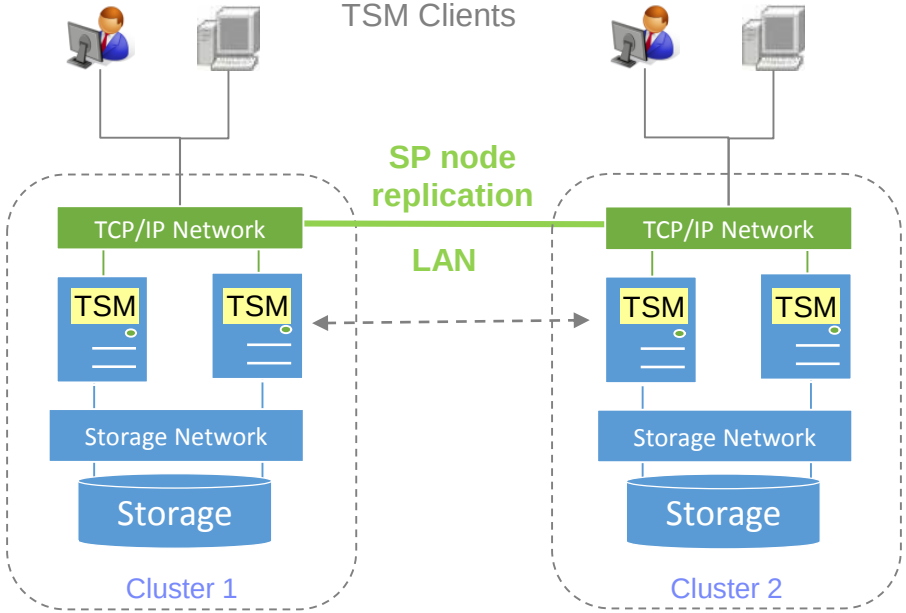


Synchronous replication via LAN or SAN

One cluster with active instances in both sites

Made for high availability and workload balancing

Spectrum Protect node replication



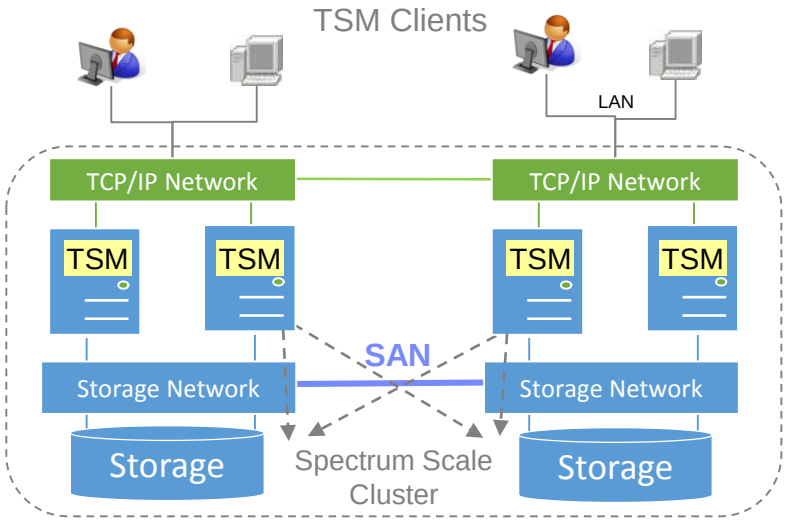
Asynchronous replication by SP via LAN

Two independent clusters with SP instances

Made for disaster recovery

Synchronous replication with Spectrum Scale and ESS

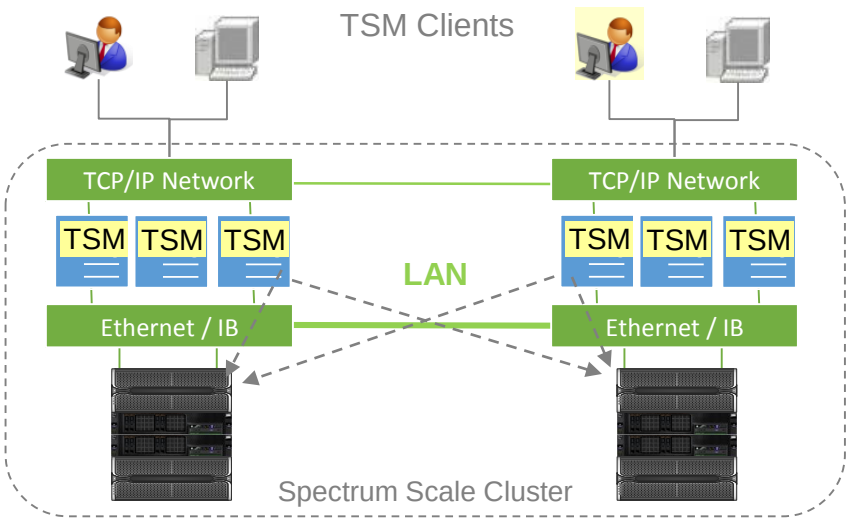
Spectrum Protect on NSD servers



Synchronous mirror over **SAN**

Spectrum Scale server on both sites have access to all storage system via SAN

Spectrum Protect with ESS



Synchronous mirror over **LAN**

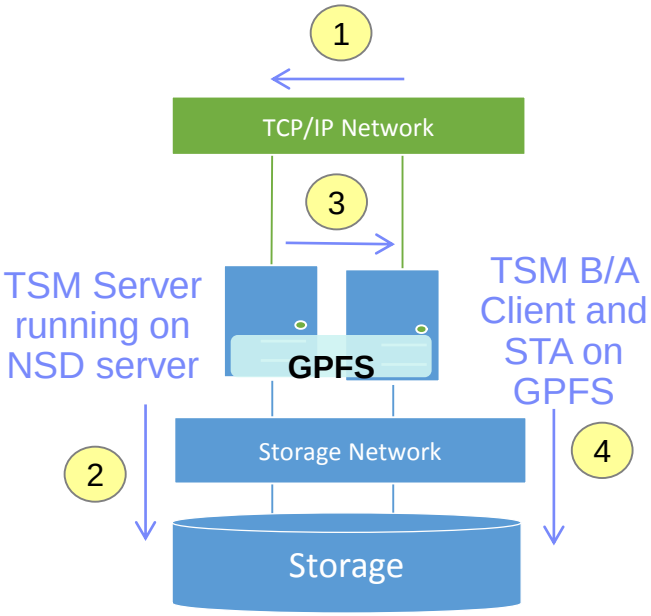
Scale clients on both sites have access to all ESS systems via LAN

One Spectrum Scale cluster across sites

One Spectrum Protect instance is active at one site

LAN free backup to Spectrum Scale

- TSM server has storage pools in GPFS file system
- TSM client with storage agent (STA) has direct access to GPFS storage
- How it works
  - 1 TSM client request volume from TSM server in GPFS file system
  - 2,3 TSM server mounts volume in file system and gives handle to TSM client
  - 4 TSM client writes directly to volume in GPFS file system



- LAN free backup not possible with ESS – but server free
- LAN free backup not possible with container pools (inline dedup)
  - Deduplication only possible with legacy dedup on file volumes

Agenda

Introduction to Spectrum Protect and Spectrum Scale

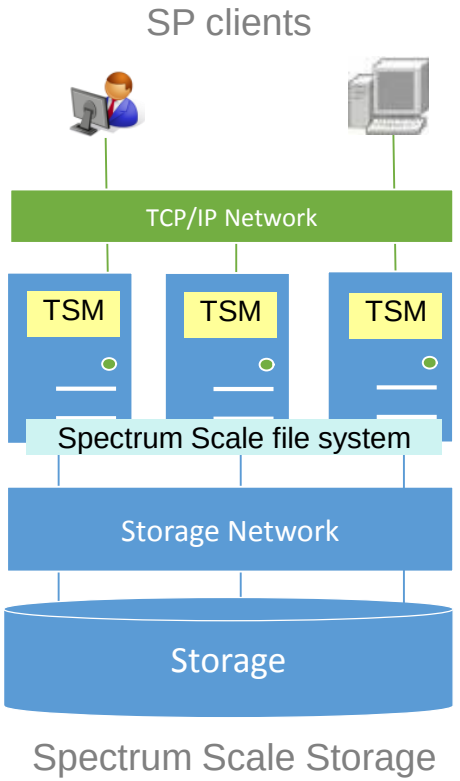
Why Spectrum Protect with Spectrum Scale

Configuration options

Positioning and advantages

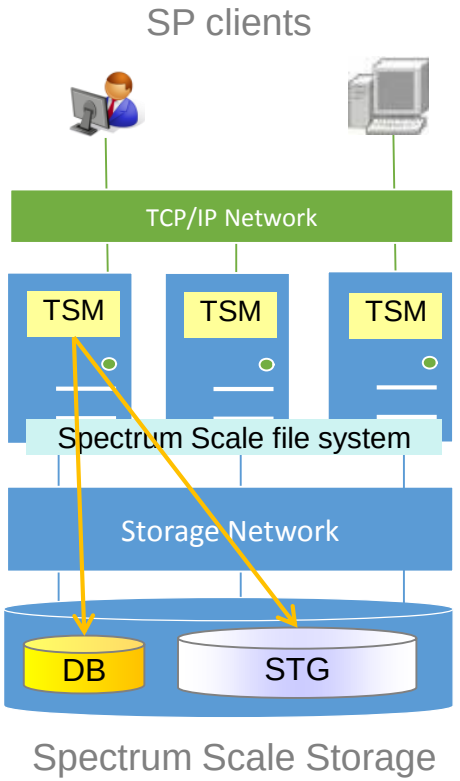
Better operational efficiency

- Single point of management
  - Storage resources for all Spectrum Protect instances are centrally managed
- Always-on
  - Adding storage capacity is transparent to Spectrum Protect instances
  - Spectrum Scale can be design highly resilient
- Transparent resource balancing
  - Instances requiring more resources can benefit from instances requiring less



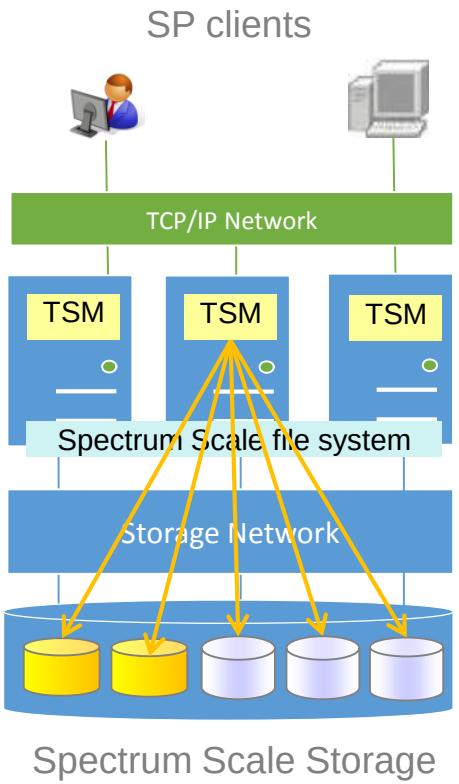
Lower cost

- No extra storage resources for DB required
  - All Spectrum Protect instances store DB and storage pools in one Spectrum Scale cluster
- Better storage utilization
  - Storage capacity is provided to all Spectrum Protect instances in global namespace
- No extra infrastructure required
  - Multiple Spectrum Protect instances can run on one Spectrum Scale node
- Use and reuse standard Infrastructure



Scaling parallel performance

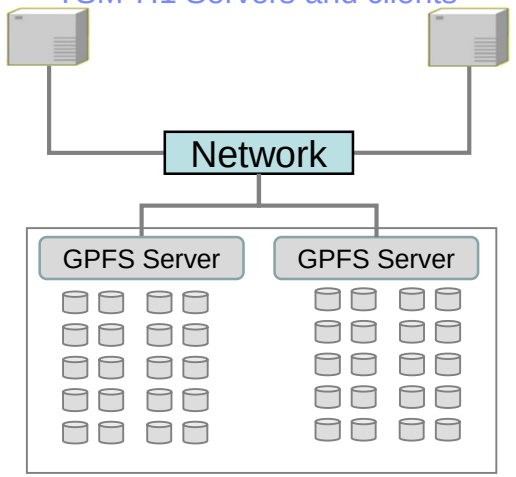
- Direct access to Spectrum Scale storage
  - Spectrum Protect stores directly in Spectrum Scale storage, no overhead for conversion of files to blocks (like with NFS)
- Parallel usage of all available disk
  - Spectrum Protect writes in parallel to all disk
  - No dedicated connection or session between TSM server and GPFS storage (no NFS bottleneck)
- TSM workload specific storage configuration
  - Variable file system block size allow to adopt specific TSM workloads for DB and storage pools
- Efficient caching for data and metadata in
  - Spectrum Scale cache improves sequential storage pool workloads



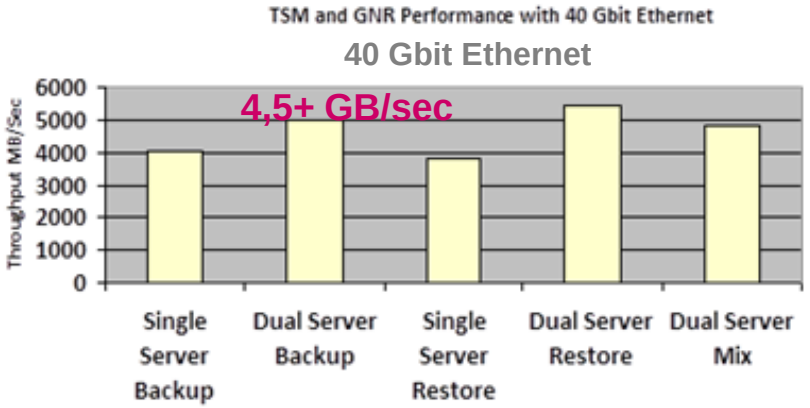
How fast can a backup with Spectrum Protect on Spectrum Scale be?

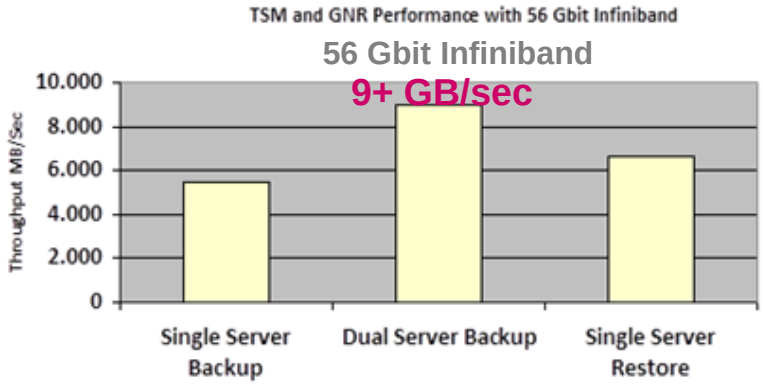
Test setup

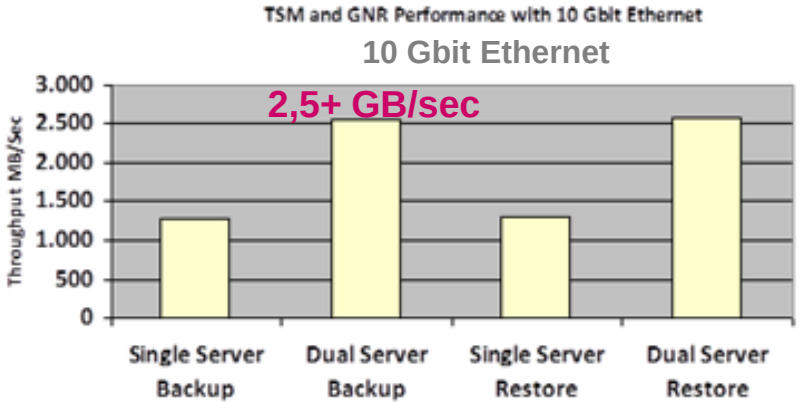
TSM 7.1 Servers and clients



GPFS Native RAID system with NL-SAS disk

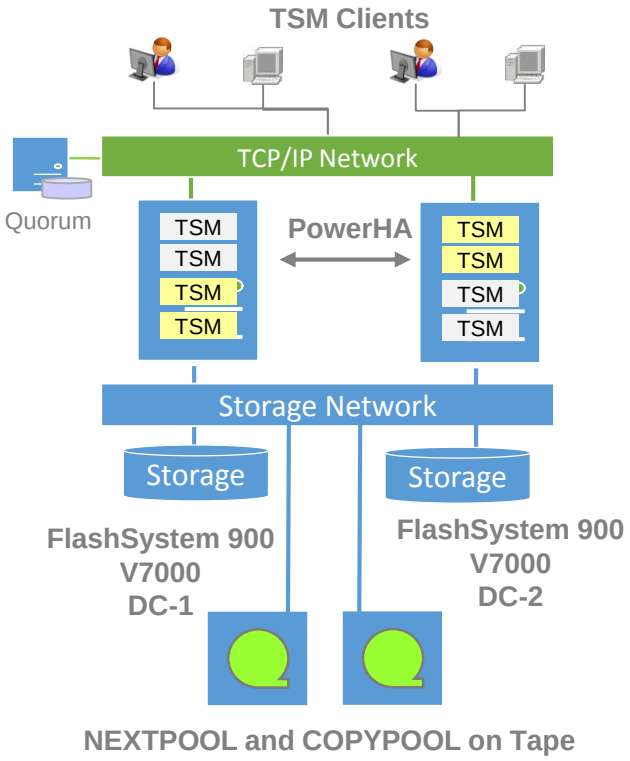






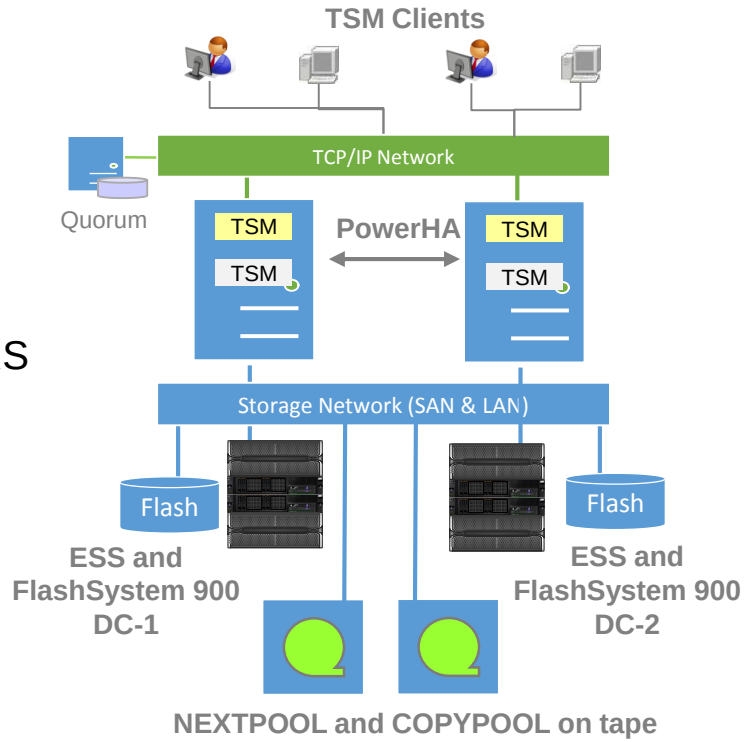
Real World Example: Spectrum Scale native

- Existing environment
  - 8 instance running on AIX LPARS with XIV
  - LVM mirror integrated with PowerHA in two DC
- Requirements
  - Scalable capacity and performance
  - High availability across two DC
- Solution
  - Spectrum Scale cluster with 3 nodes in AIX LPARS
  - IBM FlashSystem 900 and V7000
  - Spectrum Scale synchronous mirror
- Advantages
  - Seamless migration from XIV
  - Scalable performance
  - 50 % reduction of failover times
  - Easy administration



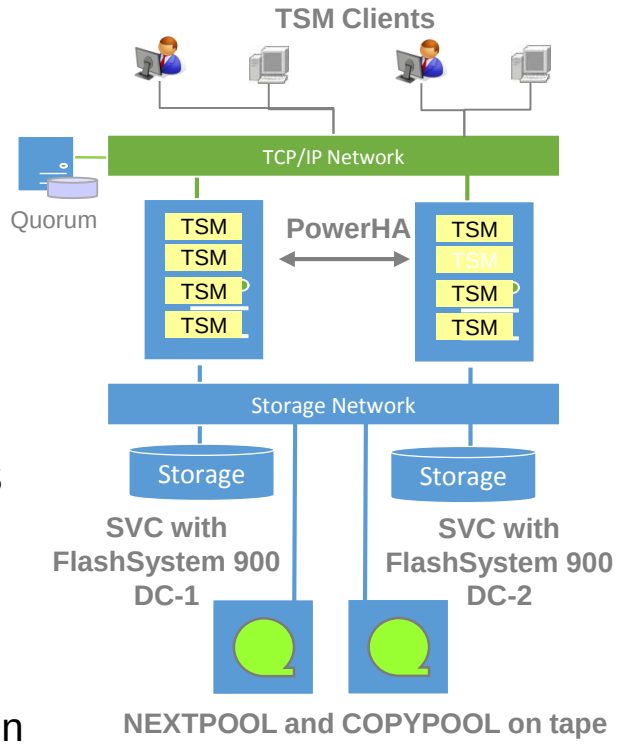
Real World Example: with ESS

- Existing environment
  - 4 instance running on AIX LPARS with SVC
  - LVM mirror integrated with PowerHA in two DCs
- Requirements
  - Scalable capacity and performance
  - Improve operational efficiency
- Solution
  - Spectrum Scale cluster with 5 nodes on AIX LPARS
  - IBM FlashSystem 900 for DB, ESS for STG
  - Spectrum Scale synchronous mirror
- Advantages
  - Scalable performance with Spectrum Scale
  - Better operational efficiency by storage consolidation
  - Easy administration based on solid standards



Real World Example: All Flash and Tape

- Existing environment
  - 2 x 8 instance running on AIX LPARs with DS8000 in four DCs
  - LVM mirror integrated with PowerHA always between two DCs
- Requirements
  - Scalable capacity and performance
  - Improve operational efficiency
  - Standardization of backup environment
- Solution
  - Spectrum Scale cluster with 9 nodes on AIX LPARS
  - IBM FlashSystem 900 managed by SVC
  - Spectrum Scale synchronous mirror
- Advantages
  - Scalable performance with IBM FlashSystem
  - Better operational efficiency by storage consolidation
  - Easy administration based on solid standards



Summary

Multiple Spectrum Protect servers on Spectrum Scale can benefit from

- Global name space provided within high available cluster
- More flexibility with synchronous replication across sites
- Better cost efficiency
- Better performance
- Ease of use

Links

- IBM Spectrum Storage home:
  http://www-03.ibm.com/systems/uk/storage/spectrum/
- IBM Spectrum Scale Home Page
  http://www-03.ibm.com/systems/storage/spectrum/scale/overview.html
- Spectrum Scale Knowledge Center:
  http://www-01.ibm.com/support/knowledgecenter/STXKQY_4.2.0/ibmspectrumscale42_welcome.html?lang=en
- IBM Spectrum Scale Wiki
  http://www.ibm.com/developerworks/wikis/display/hpccentral/General+Parallel+File+System+(GPFS)
- Spectrum Scale Redbook:
  http://www.redbooks.ibm.com/abstracts/sg248254.html?Open
- ESS home page
  http://www-03.ibm.com/systems/uk/storage/spectrum/ess/
- ESS knowledge center including Native RAID
  http://www-01.ibm.com/support/knowledgecenter/SSYSP8_3.5.0/sts35_welcome.html?cp=SSYSP8_3.5.0%2F0&lang=en
- TSM Blueprints
  https://ibm.biz/TivoliStorageManagerBlueprints

Disclaimer

This information is provided on an "AS IS" basis without warranty of any kind, express or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Some jurisdictions do not allow disclaimers of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This information is provided for information purposes only as a high level overview of possible future products. PRODUCT SPECIFICATIONS, ANNOUNCE DATES, AND OTHER INOFORMATION CONTAINED HEREIN ARE SUBJECT TO CHANGE AND WITHDRAWAL WITHOUT NOTICE.

USE OF THIS DOCUMENT IS LIMITED TO SELECT IBM PERSONNEL THIS DOCUMENT SHOULD NOT BE GIVEN TO A CUSTOMER EITHER IN HARDCOPY OR ELECTRONIC FORMAT.

Important notes:

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

IBM makes no warranties, express or implied, regarding non-IBM products and services, including but not limited to Year 2000 readiness and any implied warranties of merchantability and fitness for a particular purpose. IBM makes no representations or warranties with respect to non-IBM products. Warranty, service and support for non-IBM products is provided directly to you by the third party, not IBM.

All part numbers referenced in this publication are product part numbers and not service part numbers. Other part numbers in addition to those listed in this document may be required to support a specific device or function.

MHz / GHz only measures microprocessor internal clock speed; many factors may affect application performance. When referring to storage capacity, GB stands for one billion bytes; accessible capacity may be less. Maximum internal hard disk drive capacities assume the replacement of any standard hard disk drives and the population of all hard disk drive bays with the largest currently supported drives available from IBM.

IBM Information and Trademarks

The following terms are trademarks or registered trademarks of the IBM Corporation in the United States and / or other countries: IBM, IBM Spectrum Storage, IBM Spectrum Protect, IBM Spectrum Scale, IBM Spectrum Accelerate, IBM Spectrum Virtualize, IBM Spectrum Control, Tivoli, IBM Elastic Storage

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Microsoft Windows is a trademark or registered trademark of Microsoft Corporation.

Isilon is a registered trade mark of EMC Corporation in the United States and other countries

Other company, product, and service names may be trademarks or service marks of others.

Disclaimer – Performance and Capacity

Performance is based on measurements and projections using standard benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including 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 results similar to those stated here.
There are many ways to set up a filesystem, the figures shown are for a filesystem implemented with parameters for "maximum performance" (typically 16MB blocksize). If more efficient space utilisation is required for small files (e.g. 512KB blocksize) then the achievable performance will be affected.

Throughput figures (MB/sec) are based on benchmarks using the IOR benchmark and the ESS set up with 8+2P data protection and a combined Data+Metadata storage pool. Where actual results for a model/storage combination are not available, figures are estimates which have been scaled from real benchmarks. IOR tests were performed using Infiniband interconnect and RDMA, figures for other network interconnects may vary significantly.

IOPS figures are "uncached, raw" IOPS tested using the gpfsperf benchmark.

The Separate capacity columns assume 7% to 10% of the filesystem capacity is set aside and dedicated for Metadata (MD), this is a typical figure but can vary depending on filesystem requirements (e.g. a very large number of files may require >7%). The figure shown is approximately 2%-3% usable MD with 3-way replication assumed. The combined Data+MD figure provides an indication of maximum filesystem capacity, a combined MD+Data is the default way in which most filesystems work. Separated MD and Data pools provide better performance, total capacity is seen as reduced because the whole of the MD space is assigned explicitly and is seen as separate- this is the recommended)way to set up a GPFS filesystem. The ESS default is to use separate MD and Data pools- 7% (GUI) or 10% (ESS scripts) is a typical MD capacity (approx 2% actual, 3-way or 4-way replicated).

These calculations take into account the equivalent of 2 spares worth of disk space per Declustered Array (DA) which is the default and can be changed.

These calculations also take into account GNR overheads (checksums) etc.
These calculations assume all vdisks are of the same RAID type, although mixing is allowed in practice.
This does not take into account any GPFS file system level replication which will reduce the net usable space appropriately.