

La virtualizzazione dello storage nell'infrastruttura Cineca per il calcolo ad alte prestazioni

Speaker: Alessandro Tonello
Storage Networking Group
Cineca



- IBM pSeries: cluster 1600 with 64 nodes, 512PE, (p5 generation), 'SP5'
- IBM xSeries: cluster 1350 with 512 nodes, 1024PE, x345 and x346, 'CLX'
- IBM Blade cluster with 274 nodes, 1096PE, (AMD-64 Opteron), 'BCC'
- Cray XD1 Linux Cluster, 72 nodes, 144PE (AMD Opteron)
- SGI Altix 3700, 64 nodes
- IBM xSeries Linux cluster with 14 nodes, 28PE, x346



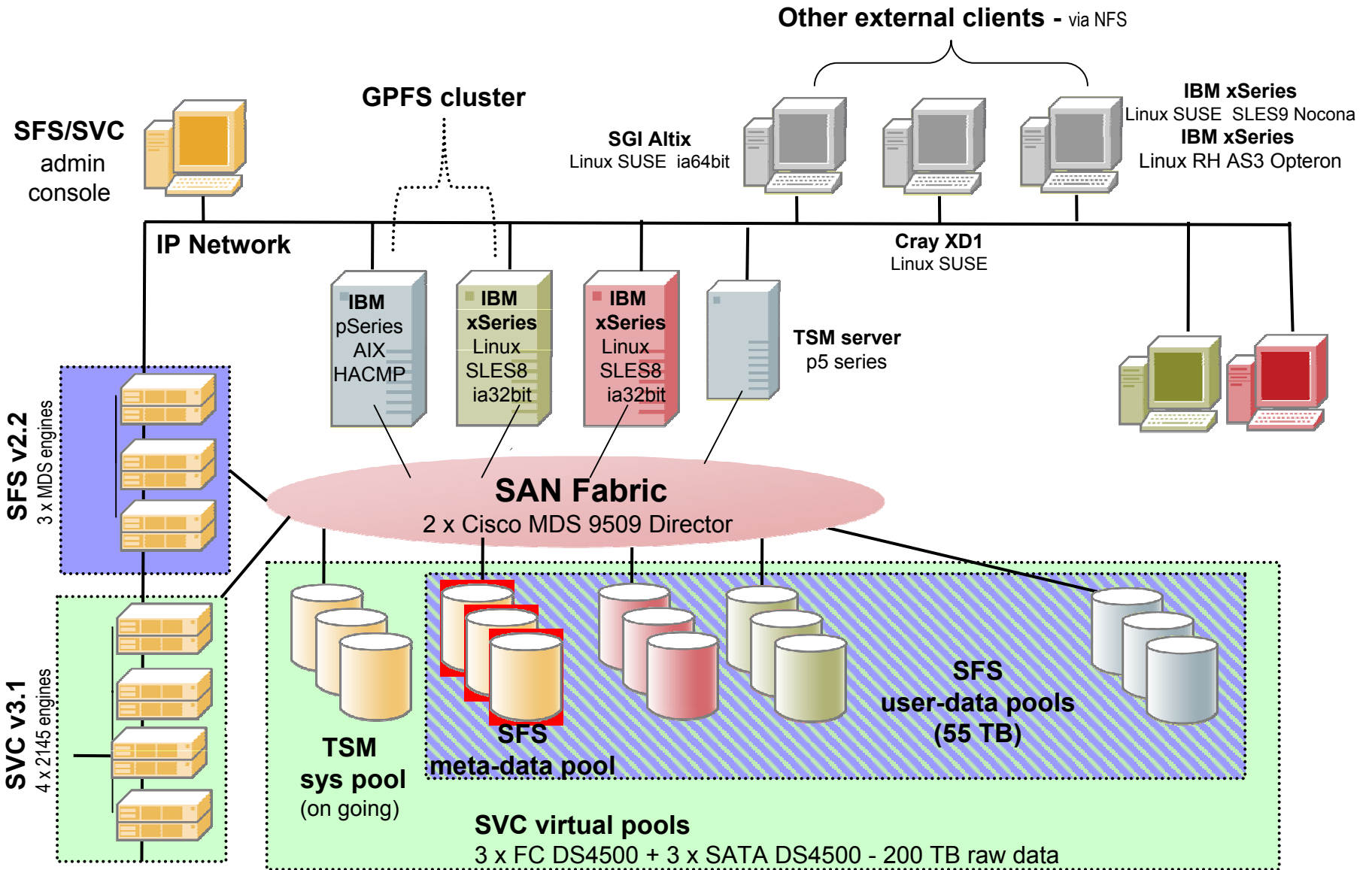
Specifiche iniziali del progetto di mass-storage

- Consolidamento dell'infrastruttura di storage
- Maggiore scalabilità
- Maggiori prestazioni (I/O)
- Maggiore disponibilità e condivisione dei dati in un ambiente eterogeneo
- Maggiore affidabilità -> aumentata ridondanza in SAN
- Aumento della capacità complessiva a breve: 200 TB
- Volume di dati: almeno 10TB/giorno sui vari sistemi di calcolo

Situazione precedente: storage dedicato per tutti i sistemi di calcolo, condivisione via NFS

Situazione desiderata:

Un'area dati condivisa, modulare e adattabile, a fronte sistemi di calcolo eterogenei (Unix e Linux) e con necessità di storage in rapida evoluzione.



- SAN
 - 2 x IBM 2062-D07 (Cisco MDS 9509 : (64 + 80) x 2 GBit/sec ports)
- Disk layer (6 x IBM DS4500)
 - FC 66 TB with 146GB/10000RPM HDD
 - SATA 134 TB with 250GB/7200RPM HDD
- SAN Volume Controller (V. 3.1) with 4 x 2145 engine nodes
- SAN File System (V. 2.2.2) with 3 x MDS nodes
- Master Console serving both SAN VC and SAN FS
- TSM server : 2 x p550 and TSM release 5.3

I cluster per il supercalcolo serviti da SAN File System:

- IBM pSeries : cluster 1600 with 64 nodes, 512PE, (p5 generation), 'SP5'
- IBM xSeries : cluster 1350 with 512 nodes, 1024PE, x345 and x346, 'CLX'
- IBM Blade cluster with 274 nodes, 1096PE, (AMD-64 Opteron), 'BCC'
- Via NFS server:
 - Cray XD1 Linux Cluster, 72 nodes, 144PE (AMD Opteron)
 - SGI Altix
 - IBM xSeries Linux cluster with 14 nodes, 28PE, x346

- 2 Switch Cisco 9509
- 3 FC modules per switch (2 x 64 ports 2Gbps)



DS4500 SATA

- 3 storage servers
- 250GB SATA disks, 2Gbps
- 13 disks enclosures x 182 disks
- RAID 5
- 133TB raw



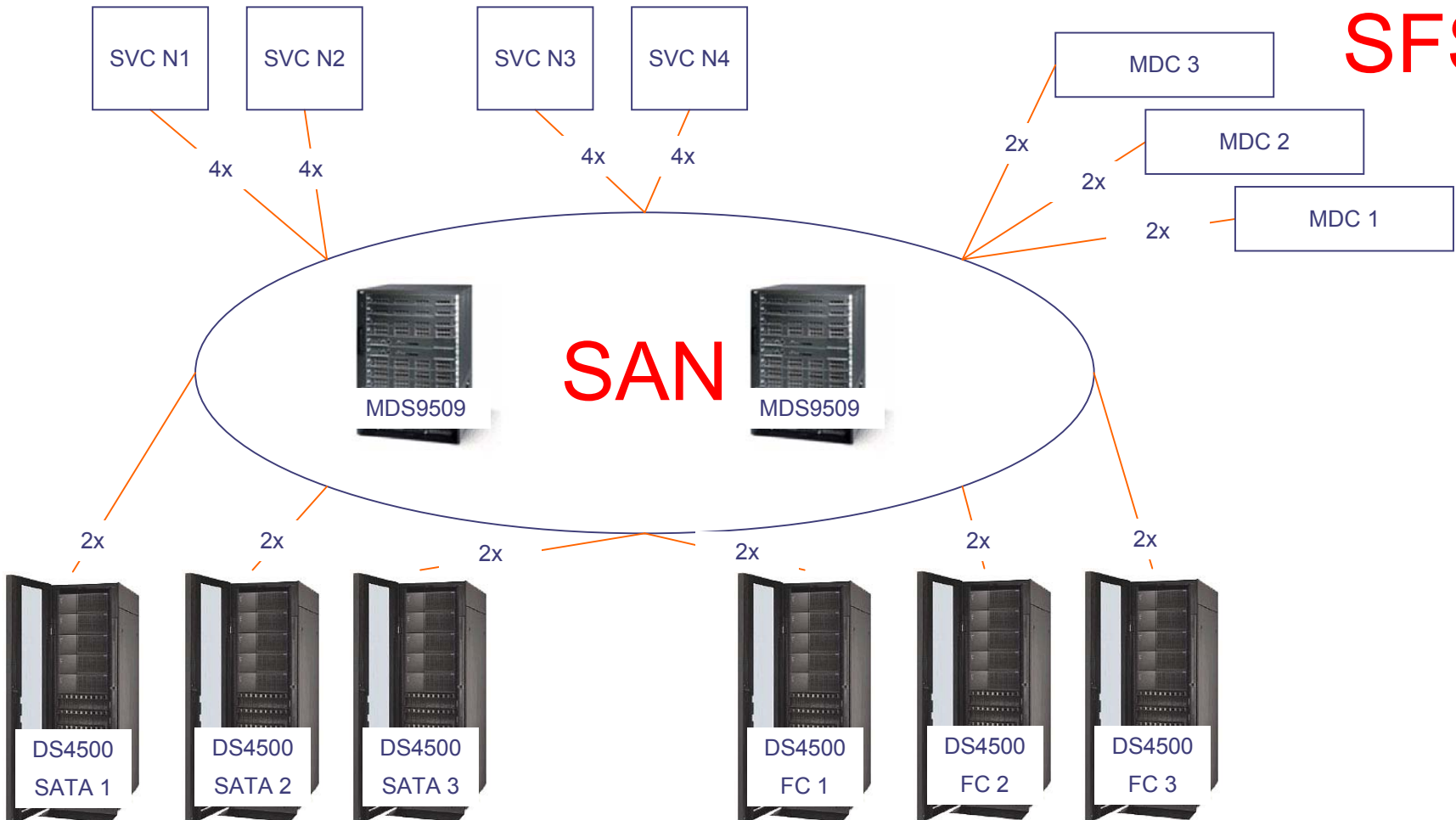
DS4500 FC

- 3 storage servers
- 146GB FC disks, 2Gbps
- 11 disks enclosures x 154 disks
- RAID 5
- 67TB raw



SVC

SFS

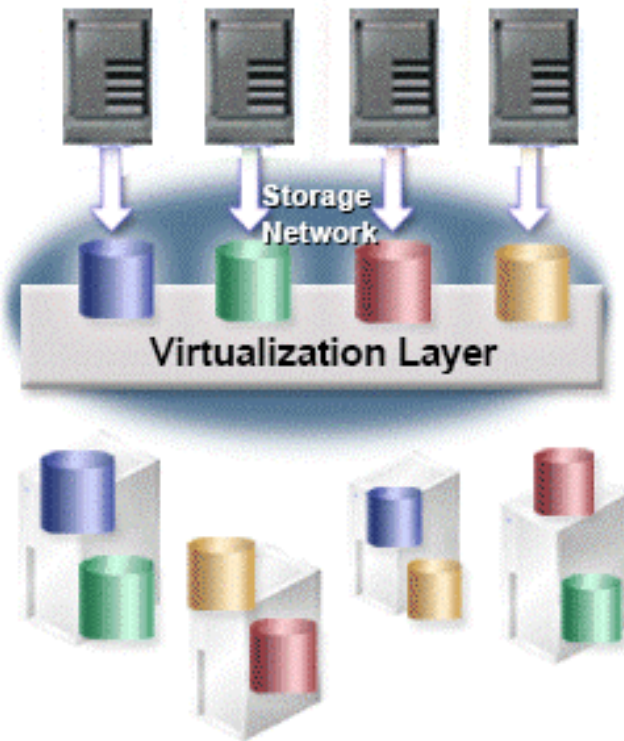
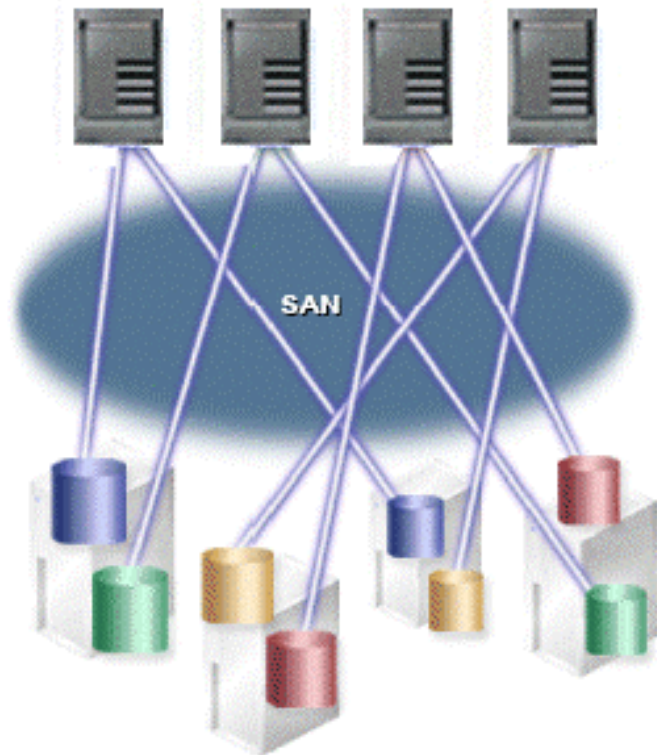


Virtualizzazione a blocchi: SVC

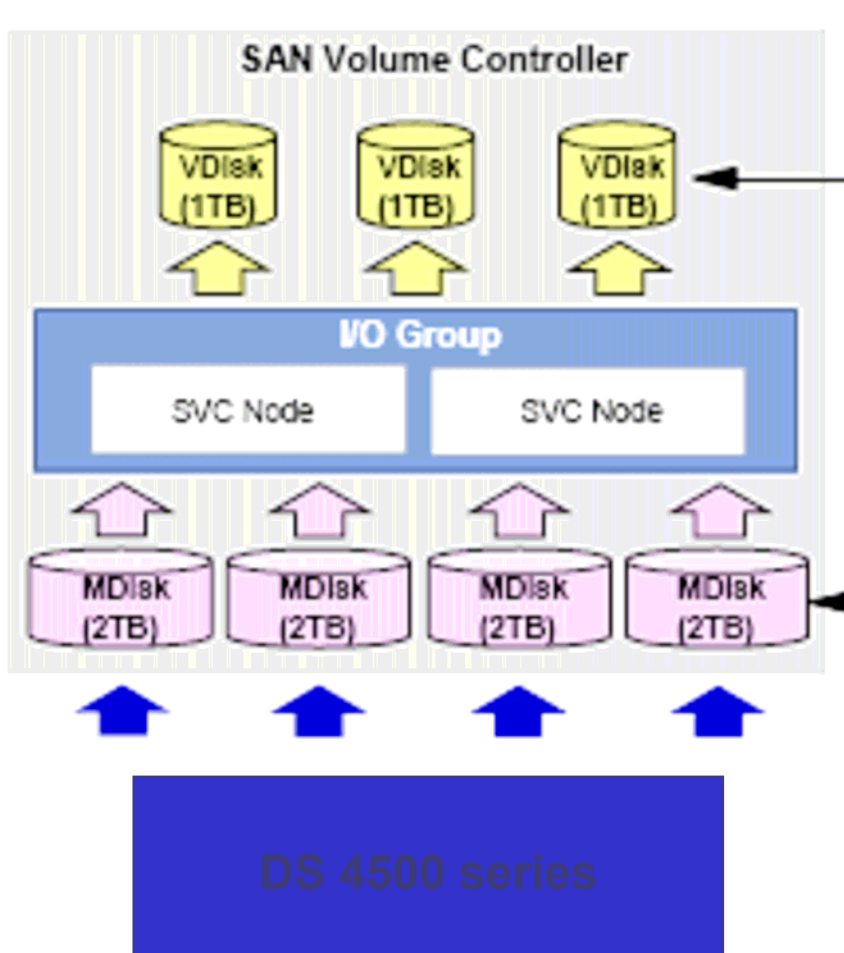
- Consolidamento
- Scalabilità
- Ridondanza

File System condiviso: SAN FS

- Ambiente eterogeneo
- Prestazioni con accesso diretto ai dati



Struttura di virtualizzazione dello storage con SAN Volume Controller



- 60 TB tipo FC
- 62 TB tipo SATA
- 4 nodes, 2 I/O groups
- 4 HBAs per node

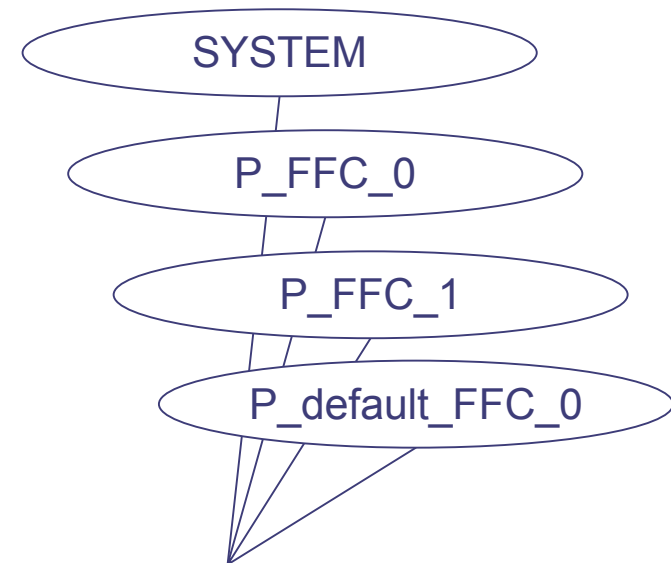
5 Storage Pools:

1 System Pool (600GB) w/ 3 volumes

1 Default Storage Pool (200GB) w/ 1 volume

1 SATA Storage Pool (37TB) w/ 18 volumes

2 FC Storage Pool (18TB) w/ 10 volumes



Benefit	SAN Volume Controller	SAN File System
Create a single pool of storage from multiple disparate storage devices	✓ Virtual Volumes from the storage pool	
File, Data sharing across heterogeneous Servers, OS		✓ Single SAN-wide File System, global namespace
Centralized Management	✓ Single interface for the storage pool	✓ Single view of file space across heterogeneous servers
Improved Capacity Utilization	✓ Pools volumes across disparate storage devices	✓ Reduces storage needs at File Level
Improved Application Availability	✓ No downtime to manage LUNs, migrate volumes, add storage	✓ Non-disruptive additions/changes to file space, less out-of-space conditions
Policy Based Automation		✓ Files, Data, Quality-of-Service based pooling

Necessità del Cineca	SAN Volume Controller	SAN File System
Gestione della crescita e ottimizzazione, consolidamento delle risorse.	<ul style="list-style-type: none"> • Migliore gestione su storage serves e tecnologie eterogenee • Espansione LUN scorrelata da dischi fisici • Storage devices suddivisi per Managed Disk Groups 	<ul style="list-style-type: none"> • Non necessarie copie di file per la condivisione • Gli Storage Pools permettono di differenziare la tecnologia di storage per qualità di servizio richiesto • Scalabilità e prestazioni di accesso
Gestione della complessità nell'infrastruttura di storage	<ul style="list-style-type: none"> • Centralizzazione e semplificazione della gestione delle risorse • Gestione di storage FC e SATA unificata 	<ul style="list-style-type: none"> • Gestione per poliche e regole delle risorse • Semplificazione nella migrazione dati • Name Space singolo globale per la condivisione dei dati
Gestione più efficiente di Backup e Disaster Recovery	<ul style="list-style-type: none"> • Abilita il backup a livello di dischi virtuali, mantenendo la disponibilità dei dati per la produzione 	<ul style="list-style-type: none"> • Consolidamento degli agenti di backup su un singolo backup server • Granularità a livello di file nella selezione dei dati oggetto di backup e restore
Application Availability	<ul style="list-style-type: none"> • Eliminazione dei fermi produzione per cambio di storage, gestione delle LUNs, migrazione dei volumi • Migrazione dei dati tra storage devices senza fermare la produzione 	<ul style="list-style-type: none"> • Ampliamento a caldo del file space • File, data administration senza dover pianificare fermi • Riduzione dei casi di fallimento applicazioni per esaurimento dello spazio di memorizzazione

- Cineca usa SVC anche senza SFS per aree di lavoro temporaneo (scratch GPFS)
- Migrazione di dati: il nostro uso delle features di migrazione di SFS. Effetto: consolidamento a caldo delle risorse storage, scambio di storage server per manutenzione, migrazione di volumi
- SFS da client Linux, problematiche rispetto ad aggiornamenti del kernel

- Apparati in rete privata
- Backup delle configurazioni degli apparati (corruzione, DR)
- Controllo di accesso fisico
- Politiche di sicurezza nella gestione dei supporti (es. nastri per archiviazione)
- Cifratura dei dati
- Permessi di sistema unix - linux



- Cineca con IBM ha allestito una SAN virtualizzata con benefici in termini di:
- consolidamento
- flessibilità
- disponibilità
- scalabilità
- prestazioni