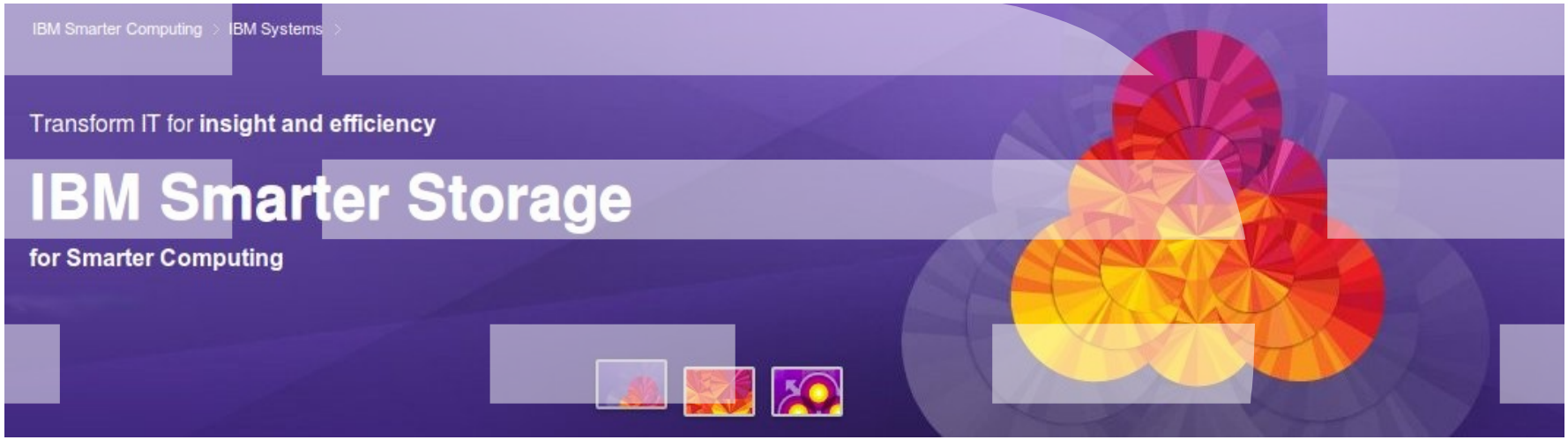
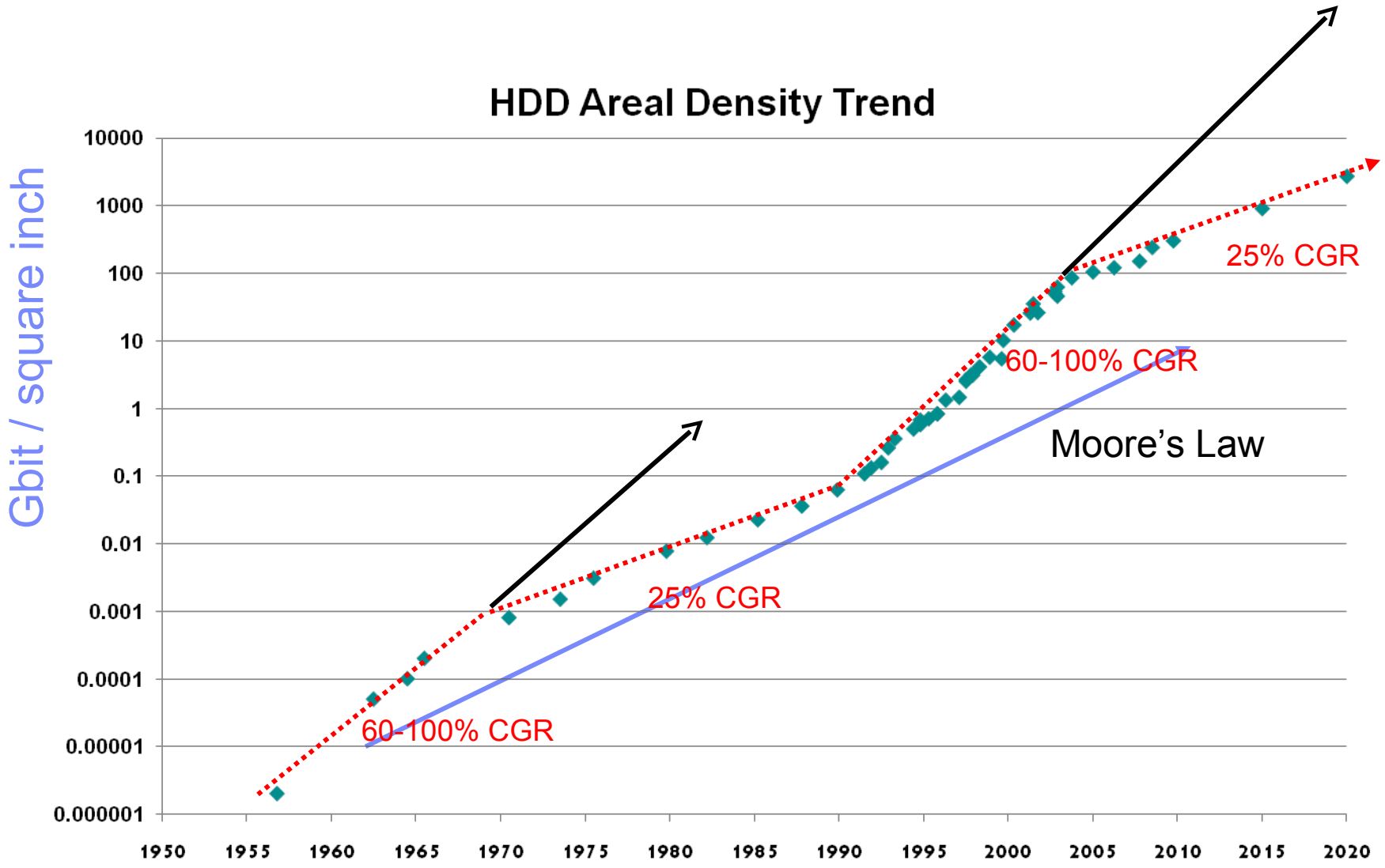
IBM SAN Volume Controller (SVC)
IBM Storwize V7000

GPFS User Group Meeting – Sept 2012

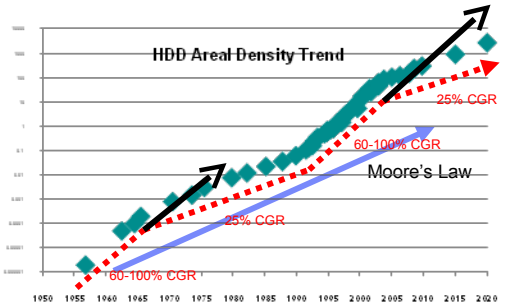


Can we change the laws of Physics ?



State of the Storage Industry ~2000

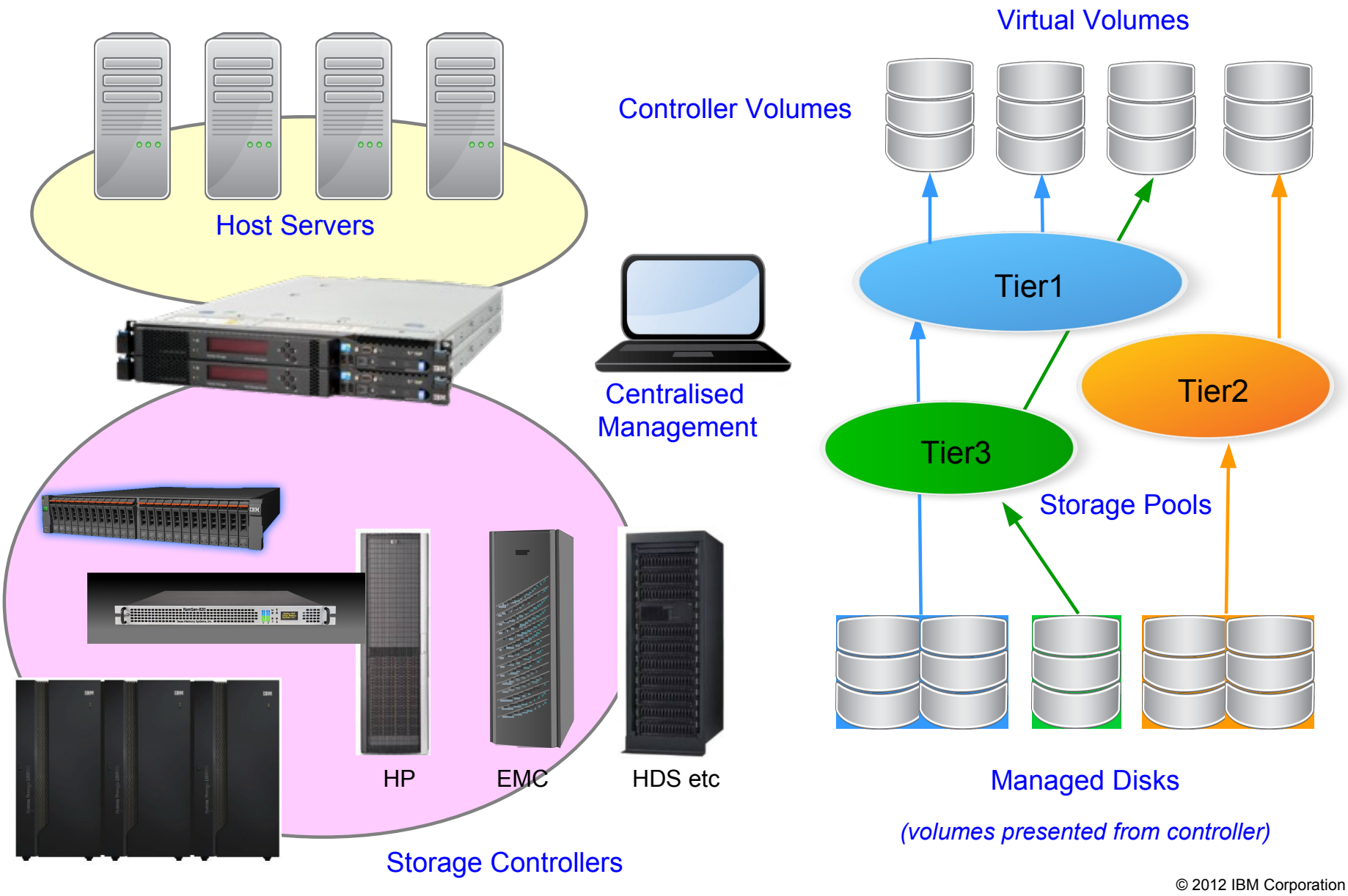
- **Back in 2000 we could see the second slow down was coming**



- **SAN complexity had grown "out the box"**
- **Storage itself was becoming a commodity**
- **Storage islands – wrong place – wasted capacity?**
- **At the same time :**
  - Custom storage function silicon development was too slow and costly – 3yrs etc
  - Commodity (x86) silicon was proving to be enterprise ready

- **Q : How can we tame heterogenous SANs?**
- **Q : How can we consolidate the advanced functions?**
- **Q : How can we make better use of what customers already have?**
- **Q : How can we ride the x86 technology wave?**

Block Storage Virtualization with SVC



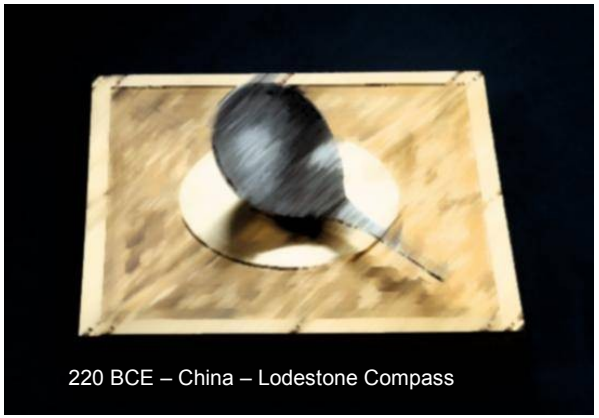
SVC Deployment

- **Nodes deployed in pairs – IO Group**
- **Volumes accessed through a single IO Group**
  - 2 way partitioned write back cache (pLRU)
  - Non-distruptive move capability between IO Groups (load balancing etc)

- **Up to 8 nodes in a single cluster**
- **Controller LUNs available to all nodes**
  - **Storage pools common across cluster**

Brief History of SVC Time

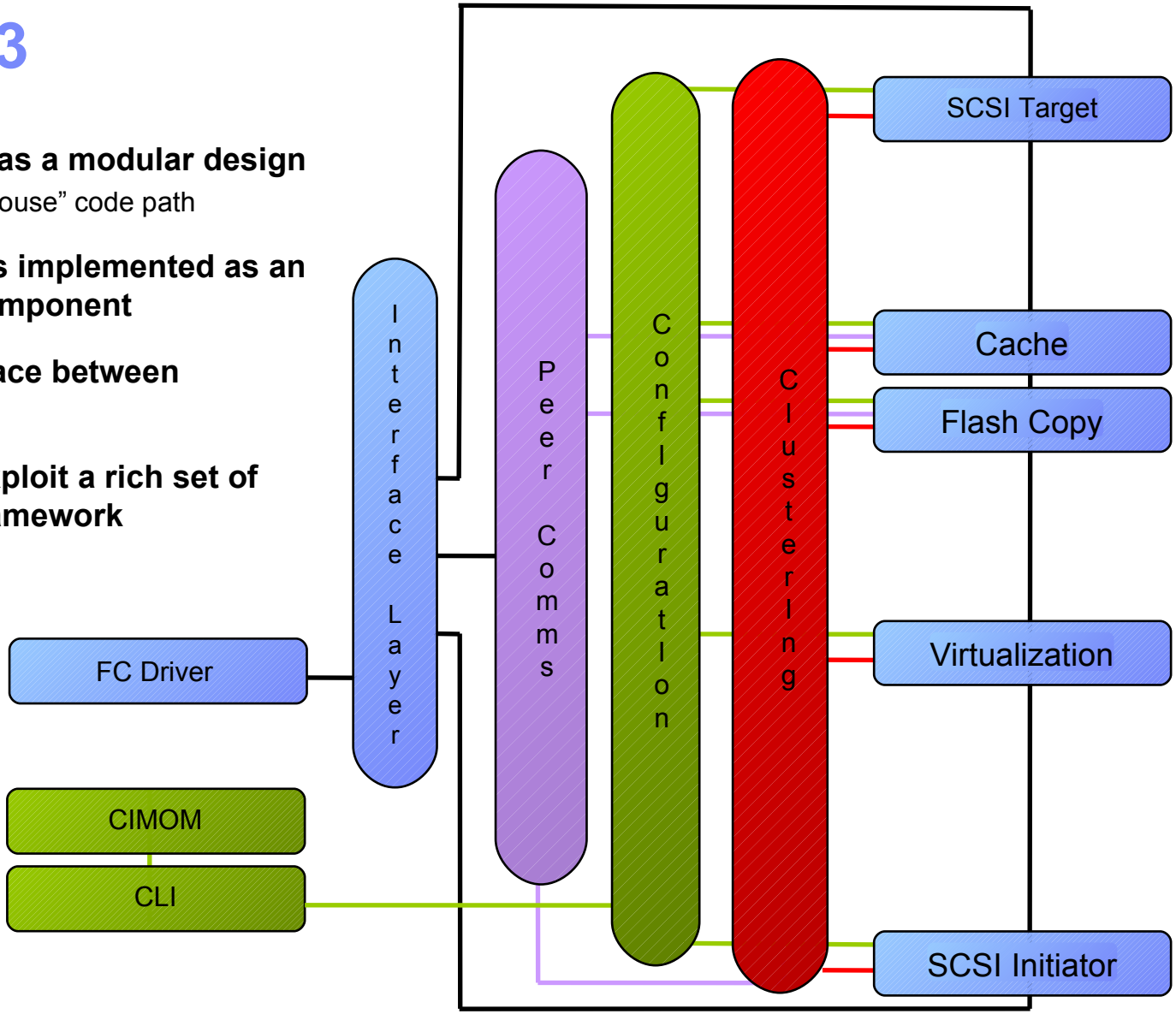
- **1999 –** Almaden Research group publish ComPaSS clustering
- **2000 –** SVC ‘lodestone’ development begins using ComPaSS



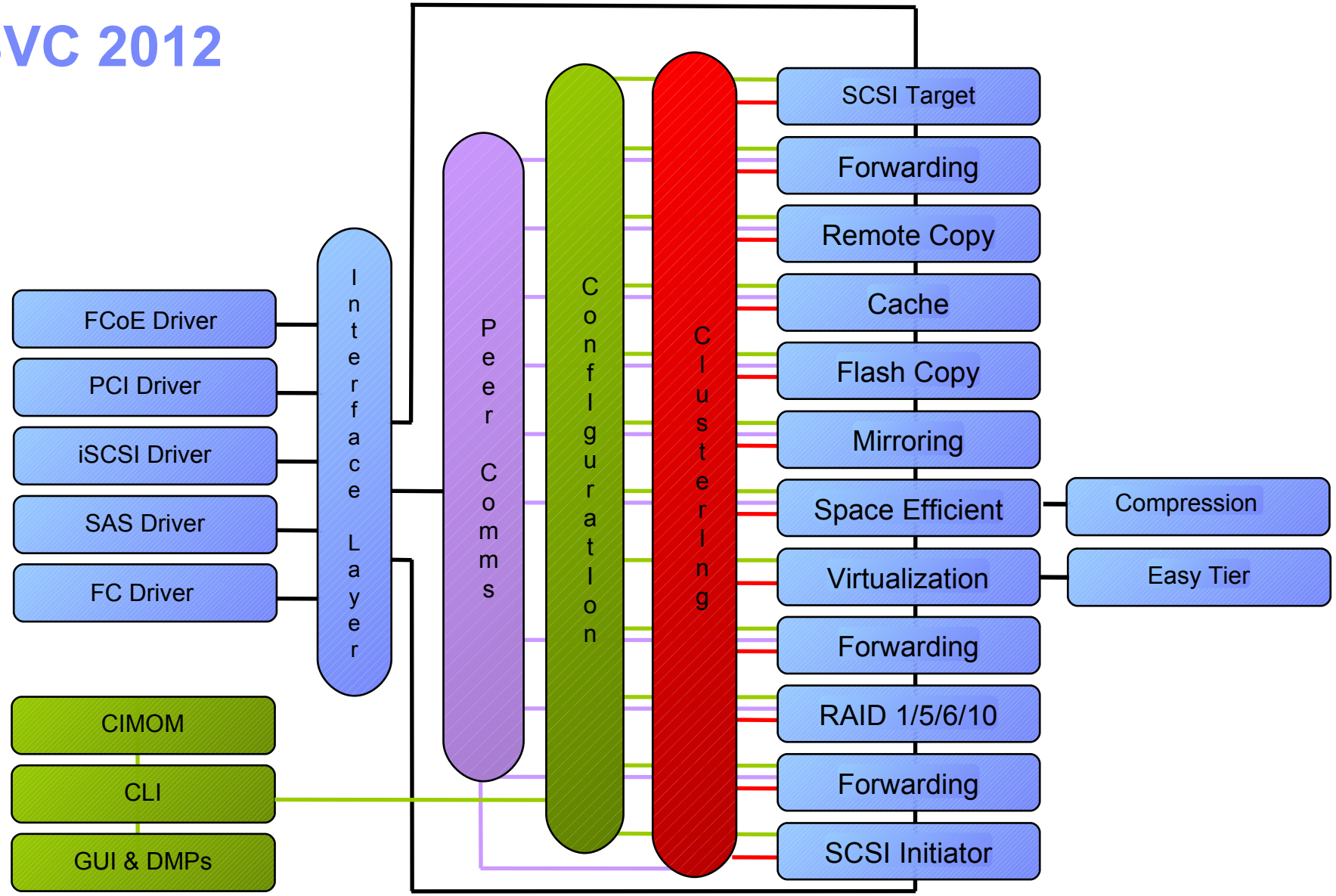
- **2003 –** SVC 1.1 – 4F2 Hardware 4 node
- **2004 –** SVC 1.2 – 8 node support
- **2004 –** SVC 2.1 – 8F2 Hardware
- **2005** – SVC 3.1 – 8F4 Hardware
- **2006** – SVC 4.1 – Global Mirror, MTFC
- **2007** – SVC 4.2 – 8G4 Hardware, FlashCopy enh
- **2008** – SVC 4.3 – Thin Provisioning, Vdisk Mirror 8A4 Hdw
- **2009** – SVC 5.1 – CF8 Hardware, SSD Support, 3 Site
- **2010** – SVC 6.1 – V7000 Hardware, RAID, Easy Tier
- **2011** – SVC 6.2/3 – CG8, V7000U, 10G iSCSI, xtD Split Cluster
- **2012** – SVC 6.4 – RtCompression, FCoE, NDVM, ++ hdw

SVC 2003

- **SVC software has a modular design**
  - 100% “In-house” code path
- **Each function is implemented as an independent component**
- **Standard interface between components**
- **Components exploit a rich set of libraries and framework**



SVC 2012



SVC Software Delivers Advanced Storage Functions

Volume Management

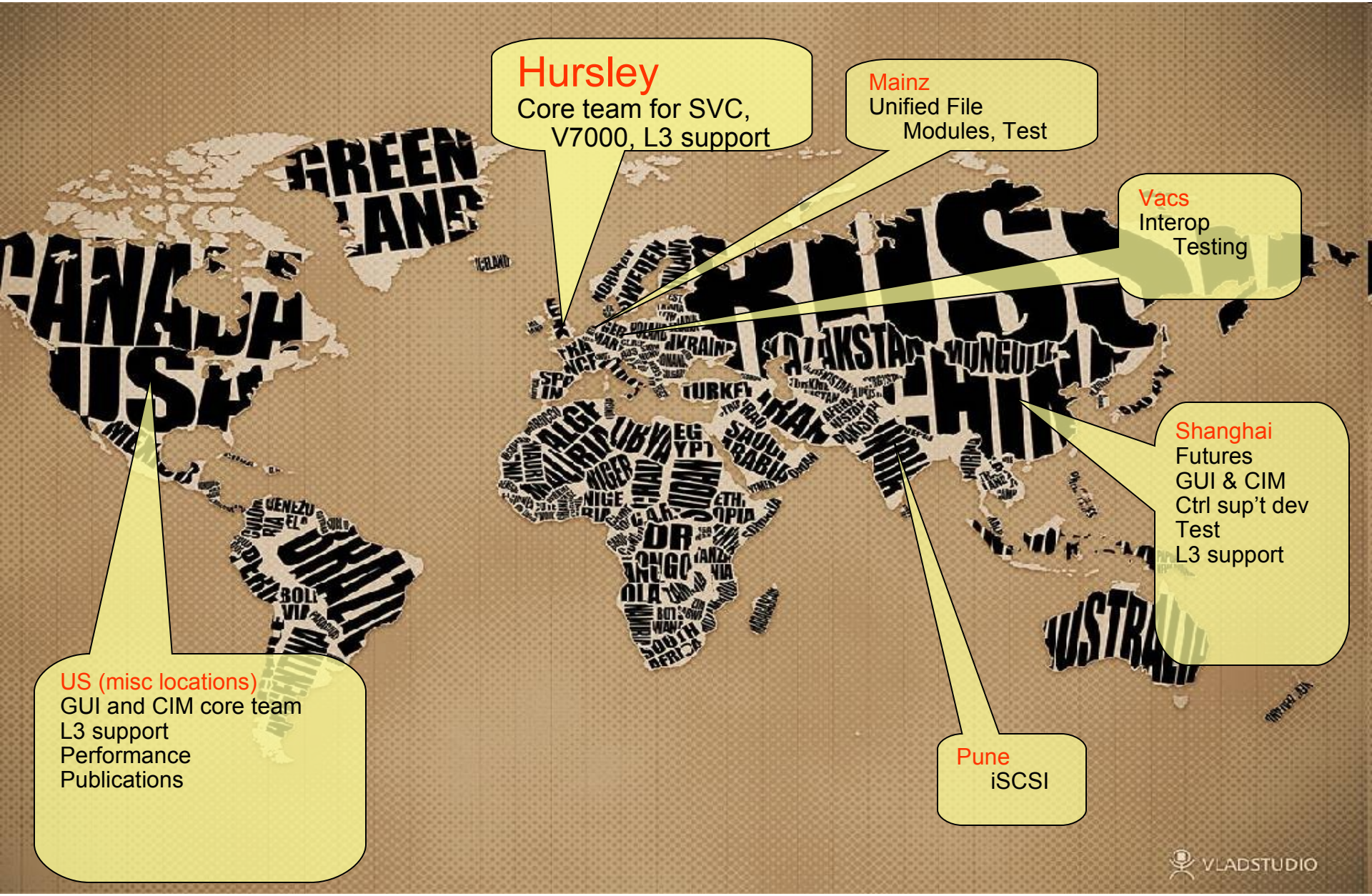
- RAID (0,1,5,6,10)
- Thin Provisioning with Zero Detect
- Non-disruptive virtual and controller volume migration (New!)
- Volume mirroring (inc thick to thin)
- Easy Tier automatic tiering
- Real-time Compression (New!)

DR and HA

- "Metropolitan" distance sync remote mirroring up to 300km between sites for business continuity
- Async remote mirroring up to 8000km between sites
- Multiple cluster partnerships (4)
- Extended Distance Split Cluster – HA – 100km for live data mobility, 300km for HA failover (New!)

Point In Time Copy

- Point in time copy of volumes for backup, business continuity and application test
- Up to 256 copies of an individual volume
- Incremental copies
- Cascaded copies
- Thin provisioned copies
- Simple "Reverse" operation



SVC 2145-CG8 Node

- **New SVC engine based on IBM System x3550 M3 server (1U)**
  - Intel® Xeon® 5600 (Westmere) 2.53 GHz hex-core processor
  - 24GB of cache
  - Four 8Gbps FC ports

- **Optional features, for each node you can choose:**
  - Solid State Devices (SSDs): up to four internal 146GB SSDs - OR -
  - 10 Gb Ethernet support for 10 Gb iSCSI and/or FcoE support
    - Two 10Gb ports per storage engine
    - Storage engine supports 10Gb iSCSI or internal SSDs *but not both*
- **Cluster non-disruptive upgrade capability to replace older nodes with new CG8 nodes**
- **Performance**
  - 125K OLTP IOPs per Node (500K per cluster)
  - 2.5GB/s sequential disk processing per Node (10GB/s per cluster)
- **Interoperability**
  - **Over 250 different models of heterogeneous storage controllers**
  - Support for mainstream offerings from all major vendors and smaller vendors (pillar, Nexsan etc)
  - Support for SSD specialist vendors (TMS/IBM) Violin
  - Every major OS / HBA / SAN vendor supported

10Gbps iSCSI/ FCOE

8Gbps FC

1Gbps iSCSI

IBM Storwize V7000 Combines IBM Technologies

Building a Platform to Deliver Superior Storage Efficiency

Storwize V7000 – Modularity and Advanced Software Functions

Modular Hardware Building Blocks in 2U

- Enclosures contain up to twelve 3.5” or twenty-four 2.5” drives in just 2U
- Control enclosure: dual active-active controllers and drives; Expansion enclosure: drives only
- Up to nine expansion enclosures attach to one control enclosure
- Mix drive sizes and HDD/SSD in enclosure
- Eight 8Gbps FC ports plus four 1Gbps iSCSI ports per controller pair; 16GB cache per controller pair

Software inherited from prior offerings plus enhancements

New **advanced software functions**

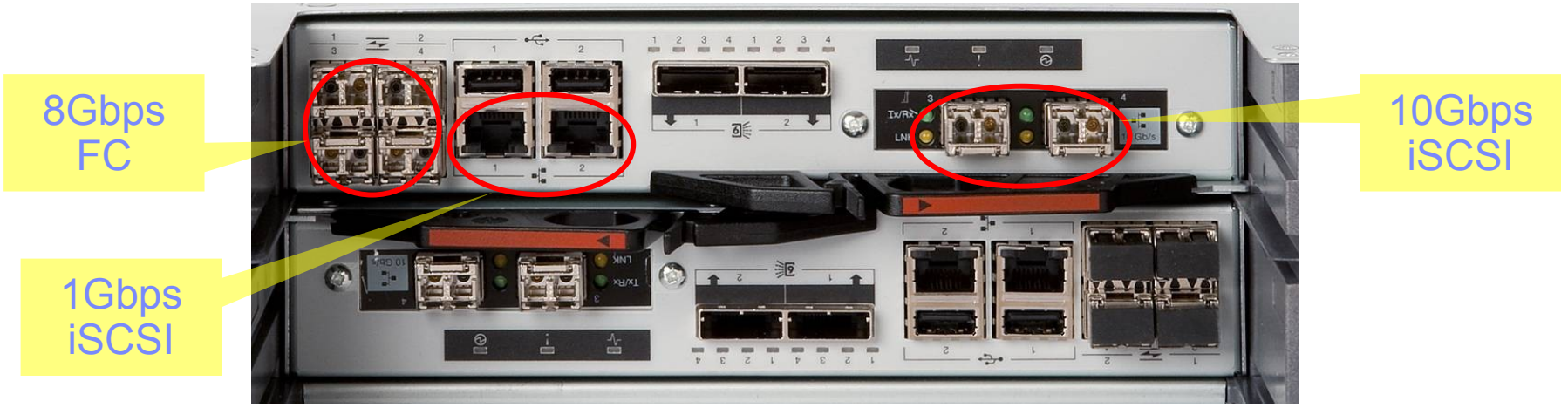
- New GUI *(easy-to-use, web based)*
- RAS services and diagnostics
- Additional host, controller and ISV interoperability
- Integration with IBM Systems Director
- Enhancements to TPC, FCM and TSM support

Proven **IBM software functionalities**

- Easy Tier *(dynamic HDD/SSD management)*
- RAID 0, 1, 5, 6, 10
- Storage virtualization *(internal and external disks)*
- Non-disruptive data migration
- Global & Metro Mirror
- FlashCopy up to 256 copies of each volume
- Thin provisioning

Storwize V7000 – Connectivity

- Delivered in two new control enclosure models
  - 2076-312 (twelve LFF drives)
  - 2076-324 (twenty-four SFF drives)
- Each node canister (controller) has :
  - 2x 1Gbps iSCSI ports (one port on one node shared with management traffic)
  - 4x 8Gbps Fibre Channel ports
  - Optional :
    - 2x 10Gbps ports
    - Can be configured for either iSCSI or FcoE
- Single control enclosure capable of over 125K OLTP IO/s – assuming sufficient disk capability
- Single control enclosure capable of 4GB/s sequential read workload



Storwize V7000 Enclosures

Storwize V7000 Clustered Systems

Scale UP and/or OUT as business requirements grow

Control enclosure 1 Control enclosure 2 ..... Control enclosure 4

- ***SCALE OUT to add performance***
- **Clustered systems provide smarter scalability**
- **Buy only as much capability as you need, when you need it**
- **Increase performance, connectivity and cache**
- **Clustered system operates and is managed as a *single system***

Non-disruptive scalability vs disruptive controller swaps

- ***SCALE UP to add capacity***

IBM Storwize V7000: A New Era in Midrange Storage Efficiency

- Modular, midrange disk storage that grows as your needs grow
  - Scales to (960) HDDs (Clustered System)
- Enterprise class capabilities in a midrange offering
  - Essentially an SVC system with its own disks
- High Performance in a Midrange Disk System
  - Up to 500,000 IOPS for “database-like” 70/30 R/W workloads
  - Up to 16GB/s sequential disk read
- Easy to set up and manage
  - Innovative, intuitive GUI eliminates complexity
  - Nondisruptive data migration included, unique in a product of this class
- Ushering in a new era in storage efficiency
  - Increase disk utilization
- Investment protection
  - Virtualize external storage arrays and extend asset life

Expand Easily

Start Small

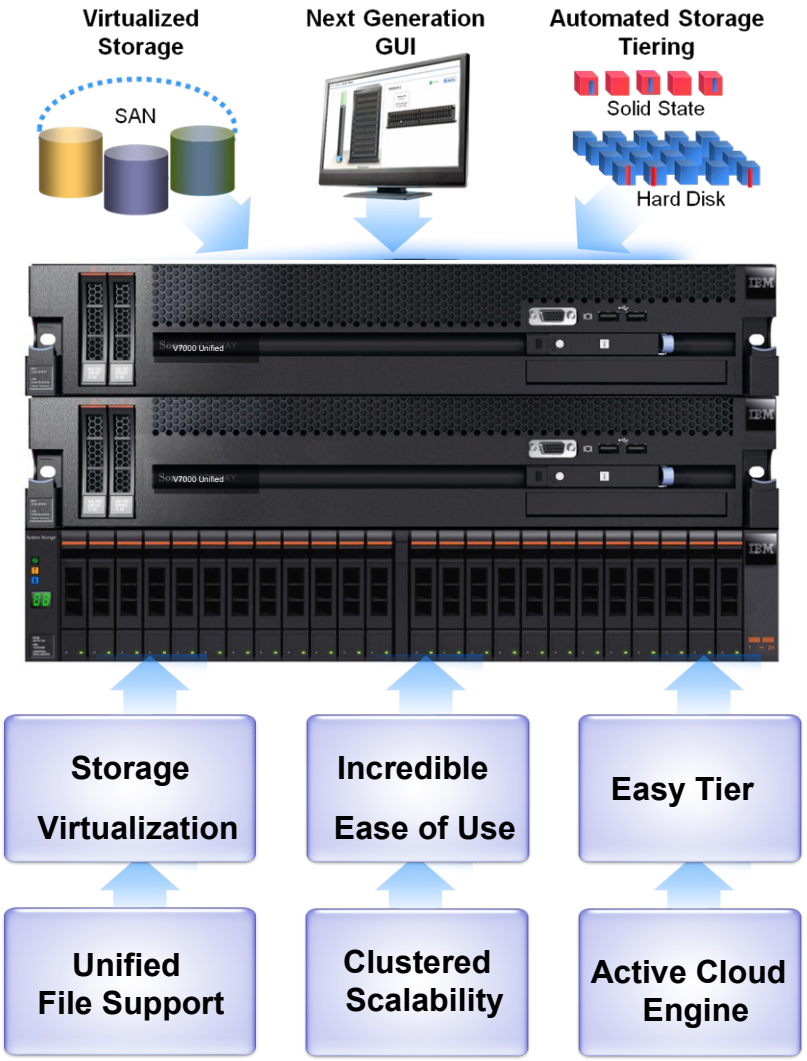
Storwize V7000 Unified

What this is

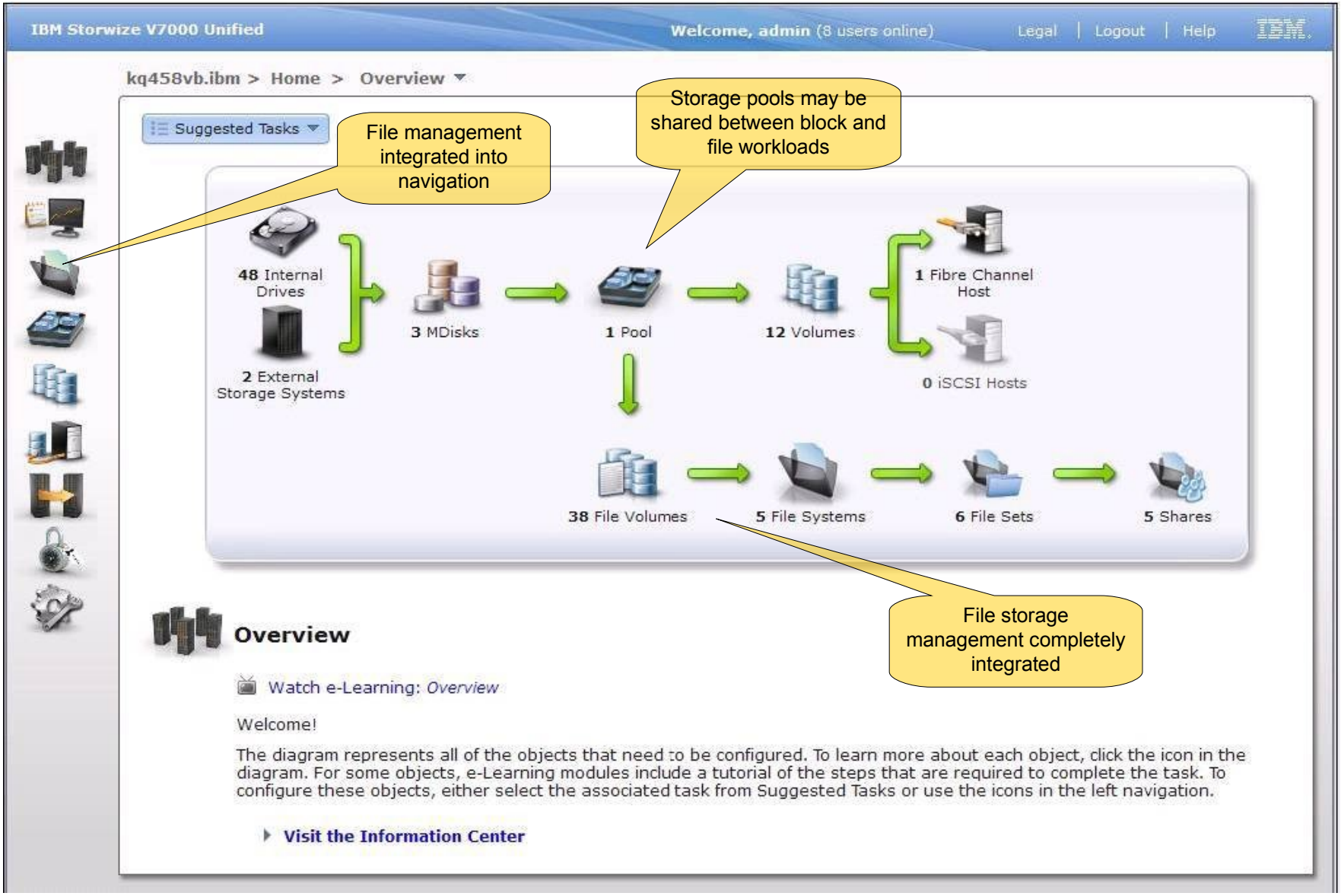
- Unified block and file storage system with a tightly integrated management console
- Support for NFS/CIFS/FTP/HTTS/SCP file protocols in addition to existing block functions
- File replication and file level snapshots for business continuity and disaster recovery

Continuing the evolution with GPFS at its heart

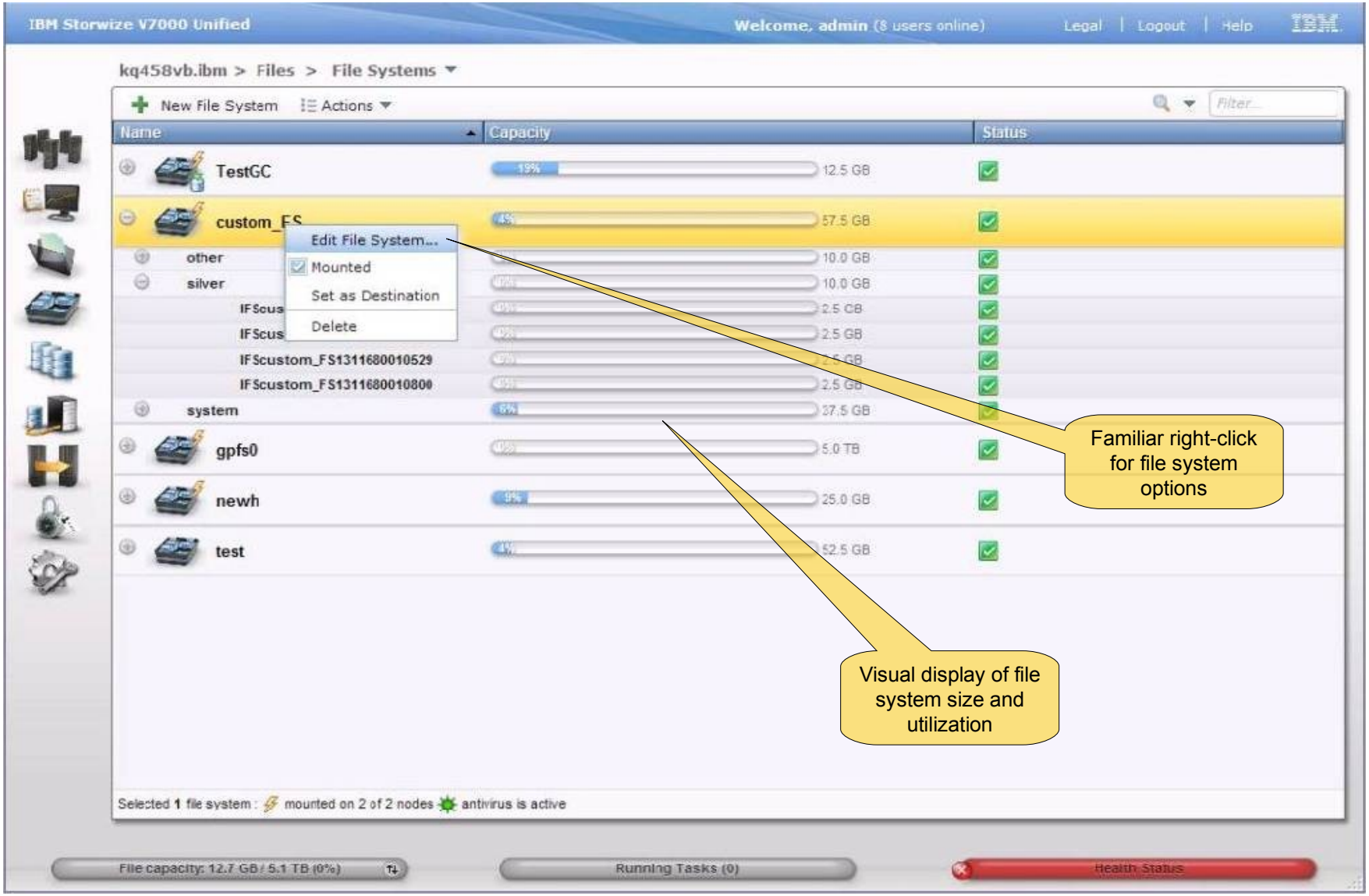
- Network filesystem built ontop of IBM's GPFS massively scale out filesystem
- Fully integrated file and block object management
- ILM and HSM with IBM Active Cloud Engine – policy driven data placement and movement



Storwize V7000 Unified User Interface

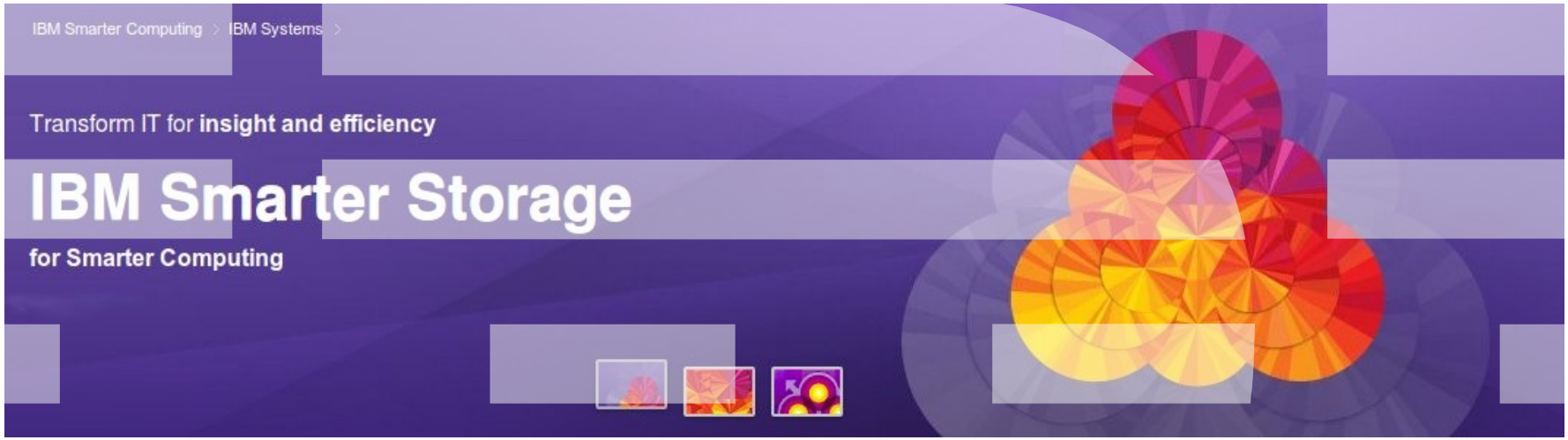


Storwize V7000 Unified User Interface

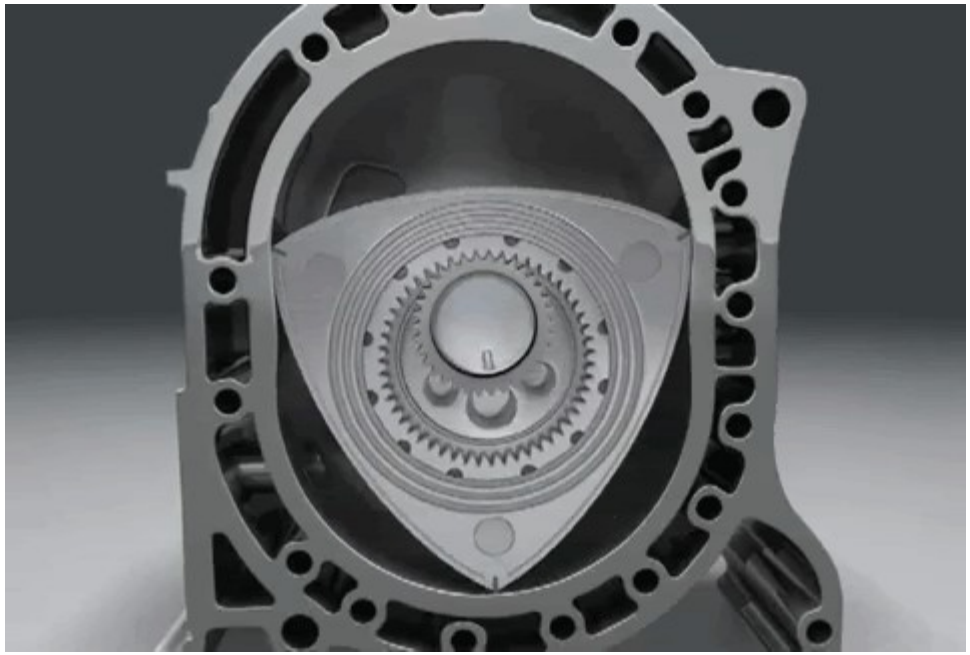


IBM Easy Tier and Real time Compression

GPFS User Group Meeting – Sept 2012



Wankel Rotary Engine – Renesis (Mazda RX-8)



Official figures – 21mpg urban, 27mpg extra-urban

Actual figures – 15mg urban, 21mpg extra-urban

Official Government MPG Figures

Urban cycle test explained

"It is carried out on a rolling road in a laboratory with an ambient temperature of 20 degrees Celsius (°C) to 30°C. … The maximum speed in the test is 31mph, average speed 12 mph and the distance covered is 2.5 miles"

Extra-urban cycle test explained

"The maximum speed in the test is 75 mph, average speed is 39 mph (63 km/h) and the distance covered is 4.3 miles"

Combined fuel consumption figure

"It is an average of the two tests, weighted by the distances covered in each one"

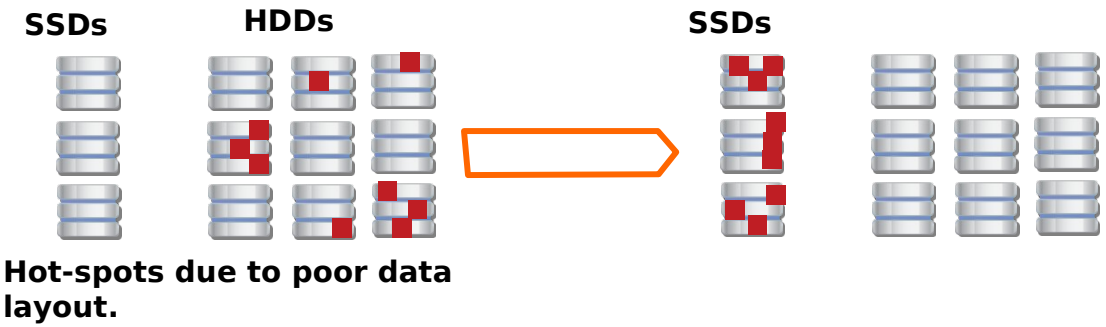
Taken from :

http://www.direct.gov.uk/en/Motoring/BuyingAndSellingAVehicle/AdviceOnBuyingAndSellingAVehicle/CalculatethefuelconsumptionCO2andtaxcosts/DG_195297

Your mileage may vary !

IBM Easy Tier

IBM Easy Tier SSD Management



- ✓ **"Easy Tier" pools identify the busiest data extents and automatically relocate them to highest performing Solid-state Disks**
- ✓ **Remaining data extents can take advantage of higher capacity, price optimized disks**

Benefits

Deploy SSDs in an optimal manner- improve performance and avoid unnecessary expenses

- Reduces administrative effort and costs
- Setup takes Minutes

Higher performance with 50% of the footprint and 40% energy consumption

Virtually no administrative effort: No need to bother with tiering policies that must be manually set to accommodate changing workload dynamics

SVC / V7000 Easy Tier Considerations

- Extent granularity movement :
  - Attribute of storage pool (16,32,64,128,256,512,1024,2048,4096,8192 MB (256MB default))
  - I/O > 64KB and sequential I/O ignored
  - Looking for small block random workloads

- “Greedy” Algorithm
  - Always looking to find hottest extents across “hybrid” pool.
  - Aims to improve latency of <64KB I/O accesses
  - Reads and write are equal
  - Will evict colder extents back to HDD
  - Learning over 24 hour period / evaluations / history up to 28 days
    - Paced migrations after providing recommendations back to SVC virtualization layer

- GPFS Considerations
  - Internal SpecSFS testing on V7000U shows more advantage (higher IOPs) via SSD meta-data volumes
  - However, if certain parts of data volumes are hot, then some benefit
  - GPFS fine grained striping tends to “smooth” hotness

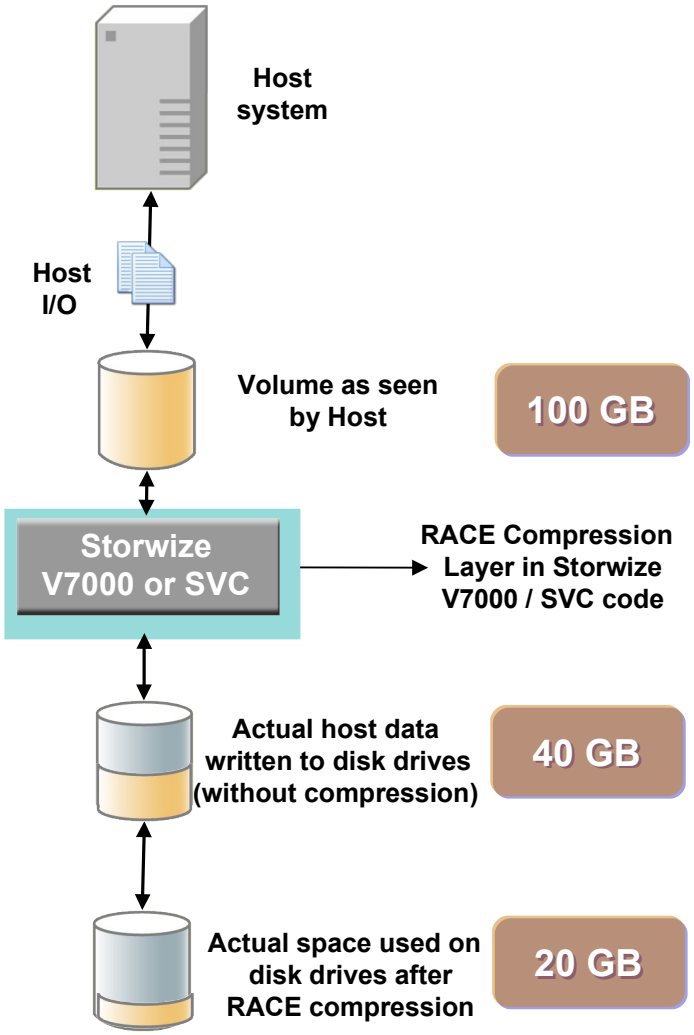
Real Time Compression

What this is

- IBM recently acquired the company known as Storwize (that name has now been applied to the new IBM Storwize V7000) which produced appliances that sit in front of NAS (network-attached storage) arrays and compress data being written to the array, using Lempel-Ziv algorithms in its Random Access Compression Engine (RACE).
- This RACE technology will be embedded into our code to provide the same compression capabilities for block storage

Why it matters

- Growing customer data demands more disk space and drives up the cost of storage in the data center
- The RACE engine, which can achieve upwards of 50 per cent data reduction, allows customers to reduce the amount of disk required for their data

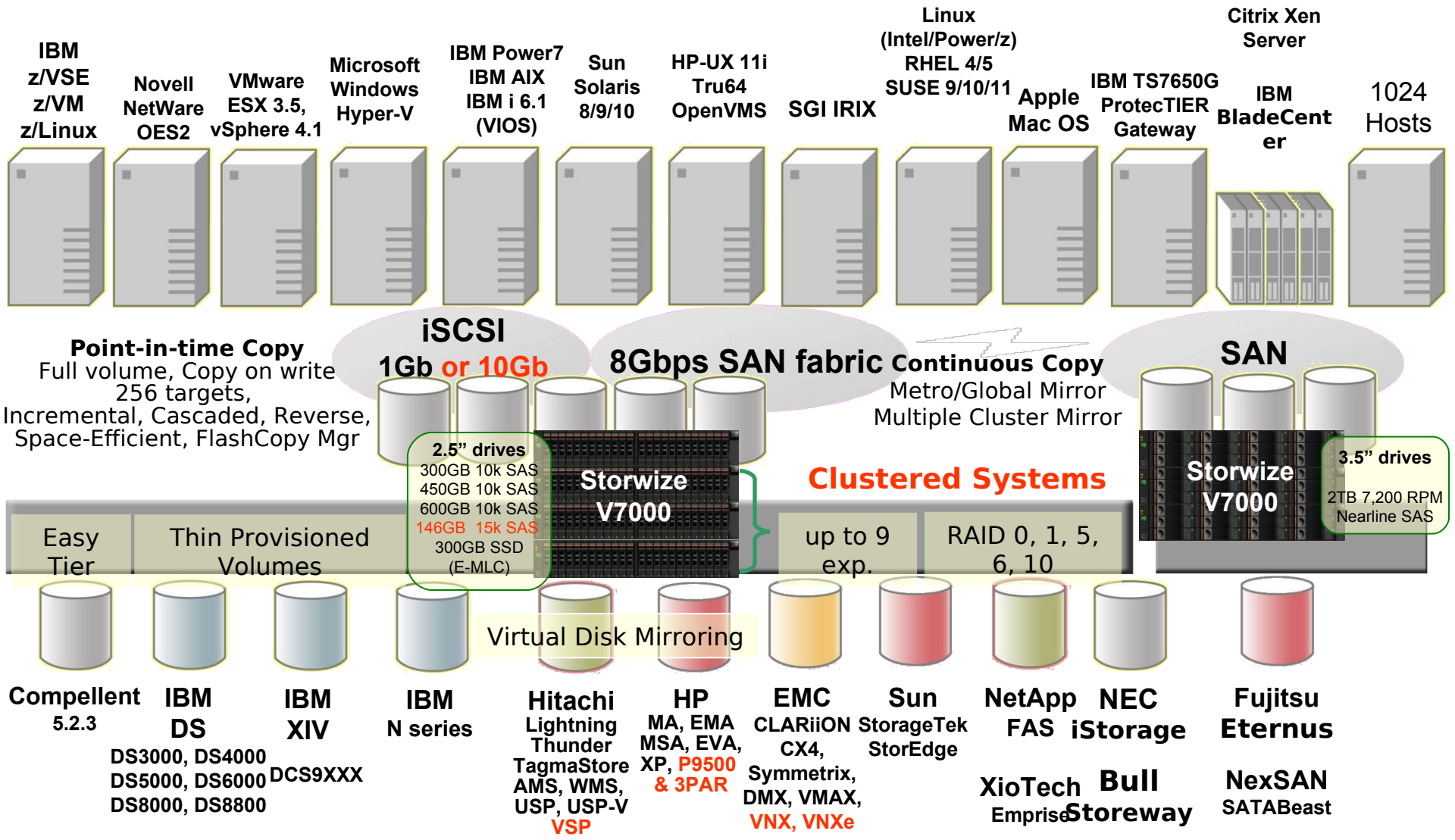


Back

RTC Considerations

- GPFS Considerations
  - Testing on V7000U shows NEVER compress meta-volumes
  - If using compression, use meta-only and data-only volumes
- Performance will vary based on temporal locality
- Do not use compressed volumes for high bandwidth workloads
  - SVC and V7000 CPU will become the bottleneck
- Current node hardware dedicates up to 75% of CPU for compression
  - Will reduce system capability even for non-compressed IOPs/MBs
- Use distinct IO Groups for compressed volumes

Supported Environments



Notice, Disclaimer, and Trademark Information