IBM Storage Virtualization for GRID environment

Stephen Chu, IT Infrastructure Solution, STG, GCG, IBM
Agenda

Why choose IBM for Storage Virtualization

How IBM Virtualize Storage
Why IBM?

The right doctor for your IT

IBM- Established in 1896
IBM’s Experience

Simpler is Better
Less is more
Reliability
Availability
Serviceability

Choose IBM for Storage Virtualization
Choose IBM for Storage Virtualization

Why IBM choose to build quality product

βWhy smart buyer buys high quality product?
<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Country/Year</th>
<th>Computer / Processors Manufacturer</th>
<th>Computer / Processors Manufacturer</th>
<th>%peak</th>
<th>%peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DOE/NA/LLNL</td>
<td>United States/2005</td>
<td>BlueGene/L eServer Blue Gene Solution</td>
<td>67036</td>
<td>156000</td>
<td>165000</td>
</tr>
<tr>
<td>2</td>
<td>IBM Thomas J. Watson Research Center</td>
<td>United States/2005</td>
<td>BGW eServer Blue Gene Solution</td>
<td>40860</td>
<td>91290</td>
<td>114885</td>
</tr>
<tr>
<td>3</td>
<td>NASA/Ames Research Center/NAS</td>
<td>United States/2004</td>
<td>Columbus SGI Altix 1.5 GHz, Voltaire /10162</td>
<td>SGI</td>
<td>51870</td>
<td>60950</td>
</tr>
<tr>
<td>4</td>
<td>The Earth Simulator Center</td>
<td>Japan/2002</td>
<td>Earth-Simulator / 5120</td>
<td>NEC</td>
<td>35820</td>
<td>40980</td>
</tr>
<tr>
<td>5</td>
<td>Barcelona Supercomputer Center</td>
<td>Spain/2005</td>
<td>MareNostrum 2860 cluster, PPC 970, 2.2 GHz</td>
<td>Myrinet / 4500</td>
<td>27910</td>
<td>42144</td>
</tr>
<tr>
<td>6</td>
<td>ASTRON University</td>
<td>Netherlands/2005</td>
<td>eServer Blue Gene Solution</td>
<td>12226</td>
<td>27430</td>
<td>34406.6</td>
</tr>
<tr>
<td>7</td>
<td>Lawrence Livermore National Laboratory</td>
<td>United States/2004</td>
<td>Thunder Intel Itanium2 T1020 1.4GHz - Quadrics / 4096</td>
<td>California Digital Corporation</td>
<td>19940</td>
<td>22938</td>
</tr>
<tr>
<td>8</td>
<td>Computational Biology Research Center</td>
<td>United States/2005</td>
<td>Blue Protein eServer Blue Gene Solution</td>
<td>8192</td>
<td>18200</td>
<td>22937.6</td>
</tr>
<tr>
<td>9</td>
<td>Ecole Polytechnique de Lausanne</td>
<td>Switzerland/2005</td>
<td>eServer Blue Gene Solution / 8192</td>
<td>IBM</td>
<td>18200</td>
<td>22937.6</td>
</tr>
<tr>
<td>10</td>
<td>Sania National Laboratories</td>
<td>United States/2005</td>
<td>Red Storm, Cray XT3, 2.0 GHz / 2002</td>
<td>Cray Inc.</td>
<td>102500</td>
<td>20000</td>
</tr>
<tr>
<td>11</td>
<td>Oak Ridge National Laboratory</td>
<td>United States/2005</td>
<td>Cray XT3, 2.4 GHz / 3748</td>
<td>Cray Inc.</td>
<td>14170</td>
<td>17990</td>
</tr>
<tr>
<td>12</td>
<td>Los Alamos National Laboratory</td>
<td>United States/2002</td>
<td>ASCI Q eServer pSeries 950, 1.25 GHz</td>
<td>8192</td>
<td>13880</td>
<td>20460</td>
</tr>
<tr>
<td>13</td>
<td>Lawrence Livermore National Laboratory</td>
<td>United States/2005</td>
<td>eServer pSeries 555 1.9 GHz / 2004</td>
<td>IBM</td>
<td>13000</td>
<td>15364.8</td>
</tr>
</tbody>
</table>

**IBM Blue Gene – 5**

**IBM Power Cluster - 2**
Blue Gene SoC is implemented in 0.13 μm technology integrating L1, L2, L3 (4MB) cache; quad-word memory controller; and five networks (torus, tree, interrupt, JTAG, Ethernet)
Blue Gene system modularity

**Blue Gene Rack(s)**
- Up to 1024 Compute Nodes / Rack
- Up to 128 IO Nodes / Rack

**Host System**
- Service and Front End Nodes (P5/SLES9), Storage System, Ethernet Switch, Cabling, SuSE SLES9, DB2, XLF/C Compilers
Three Walls in HPC

Memory Wall
- Latency induced performance and bandwidth limitations

Frequency Wall
- Diminishing returns from deeper and/or wider pipelines
  (can be negative if power is taken into account)

Power Wall
- Must improve efficiency and performance equally
IBM Cell Strategy

- Multi-Core Non-Homogeneous Architecture
  - Attacks Power Wall

- 3-level Model of Memory
  - Main Memory, Local Store, Registers
  - Attacks Memory Wall

- Large Shared Register File & SW Controlled Branching
  - Allows deeper pipelines
  - Attacks Frequency Wall
Highlight of CELL

- Supercomputer on a chip
- Multi-core microprocessor (9 cores)
- 3 to 4 GHz clock frequency
- 10x performance for many applications
- Digital home to distributed computing
Dynamic Power Management

No power management

Dynamic power management

Single thread

Simultaneous multi-threading

Photos taken with thermal sensitive camera while prototype POWER5 chip was undergoing tests
How IBM optimize Storage

Simplify

Virtualize

Automate
Simplify
Data Management Challenge — Data Growth

Source: International Technology Group, Sept 2003
* hardware, software, storage networking, personnel, backup operations, recovery, security
Data, is like your personal effect

- Gain More
- Deletion
- Deletion/Sharing
- Periodic Deletion
- Periodic Archiving
- Strengthen data access, achieving & protection

Source: International Technology Group, Sept 2003
* hardware, software, storage networking, personnel, backup operations, recovery, security
Cleaning your house every morning.
Mr Zhu’s proverb

<table>
<thead>
<tr>
<th>Free Storage Space</th>
<th>Gain More</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non Critical Data</td>
<td>Deletion</td>
</tr>
<tr>
<td>Replicated Data</td>
<td>Delete/Sharing</td>
</tr>
<tr>
<td>log files, dump files</td>
<td>Periodic Deletion</td>
</tr>
<tr>
<td>Older, Isolated Data</td>
<td>Periodic Archiving</td>
</tr>
<tr>
<td>Effective Data</td>
<td>Strengthen Data Access, Archiving &amp; Protection</td>
</tr>
</tbody>
</table>

Shelf space makes your home more homy.
My wife’s proverb
Simplification

Summary

SAN
Virtualize
Virtualize Storage Device
Ecosystem for Information Life Cycle

Data Management

- Enterprise class
- Mid-range
- Low-cost

Compress?

Encrypt?

Long Term Archive

Replication Management

WORM

Automated

Media Management

Manual

Offline Archive Management

WORM

WORM
SAN Volume Controller (SVC)

Change in Storage will not impact application

Centralize Storage Management

Data replication between Storage Pool

Multiple Storage RAID to form a single storage pool

Storage Pool

DS8000

HDS

DS4000

EMC

HP

SAN Volume Controller

Advanced Copy Services

SAN

Virtual Disk

Virtual Disk

Virtual Disk

Virtual Disk

SAN Volume Controller (SVC)
Traditional RAID

Every Logical Volume map into a physical disk

Physical storage configuration change will impact application & server availability
The SAN Volume Controller “Isolate” server from changes in storage.
Managed disks was divided into disk group in order to manage different type of storage device.
SVC Terminology

Virtual Disks:
Max 4096 Vdisks total (16MB – 2TB each)
Each virtual disk assigned to:
• Specific Node-pair
• Specific Managed Disk Group

Cluster:
• Max 4 Node-pairs (8 Nodes total)

Managed Disks:
• Select LUNs from up to 64 physical disk subsystems
• Max 128 Managed Disk Groups
• Max 128 LUNs per group
3 SVC virtual disk mode

(Image Mode)
Virtual Disk = Physical LUN

(Sequential Mode)
Virtual Disk mapped sequentially to a portion of a managed disk

(Striped Mode)
Virtual Disk striped across multiple managed disks
SVC supported configuration

- **VMWare**: Windows / Linux / NetWare guests
- **Microsoft Windows**: MSCS
- **IBM AIX**: HACMP
- **Sun Solaris**: HP/UX
- **Linux (Intel)**: IBM BladeCenter
- **VMWare**: Windows/Linux
- **VMWare**: Windows / Linux / NetWare guests
- **Continuous Copy**: Synchronous
- **Asynchronous** (CNT & Other 3rd party solutions)
- **Point-in-time Copy**: Full volume Copy on write
- **SAN Volume Controller**: IBM ESS
  - 750 800
  - DS6000 & DS8000 series
  - DS4000
  - Lightning
  - Hitachi Thunder
  - 99xxV
  - EVA
  - HP EVA
  - MA
  - 8000
  - 12000
  - 16000
  - Symmetrix
  - 8000 Series
  - 3000
  - 5000
  - CLARiiON
  - All CX Models
  - FC4700
  - DMX
  - 800/1000
  - 2000
  - 3000
  - StoreEdge
  - SUN
Virtualize File System
File System Type

Local File System
- integral part of OS
- NTFS, FAT, JFS

LAN File Systems
- use LAN for data
- NFS, AFS, DFS, CIFS

SAN File Systems
- use SAN for data
- SAN FS

SAN File Systems (SAN)
Storage Subsystem
- Metadata Server
- Leo/Iva/Lou files

Local File System
- Leo files
- File ServerA
- File ServerB
- Leo/Iva files
- Iva/Lou files

LAN File Systems
- Leo
- Iva
- Lou

SAN File Systems
- Leo
- Iva
- Lou
Current File System Environment

- AIX servers
- Solaris servers
- Linux servers
- Windows servers

Filesystem:
- AFilesystem
- SFilesystem
- LFilesystem
- WFilesystem
Single Name Space

ROOT

fileset 1

fileset 3

fileset 4

fileset 5

fileset 6
Current SAN

![Diagram showing a network of file systems connected to a SAN (Storage Area Network).]
SAN File System and SAN Volume Controller Synergy

SAN File System and SAN Volume Controller Synergy
Automation
Build a storage ecosystem so cost of storage device can match value of the data it stores.

Migrating data among differing storage devices will not impact application.
IBM TotalStorage Productivity Center (TPC)

Knowledge is Power – knowing your data’s behavior

Automate Data Management with centralized tool
- Manage Virtualized and non-virtualized storage device
- Manage IBM & non-IBM Storage Device
- Provide Single Point of Control to Storage Ecosystem
TotalStorage Productivity Center (TPC) – Simplify Capacity Management

SAN Monitor

SAN Volume Controller
- IBM DS8000
- Hitachi Thunder
- HP EVA
- IBM DS6000
- Hitachi Lightning
- HP EMA
- EMC DMX
- EMC CLARiiON
- IBM DS4000
- IBM ESS

Virtual Disk

SAN Monitor

www.pdffactory.com
Disk Configuration Monitor

- Grouping Disks
- Journaling
- Automate Logical Volume mirroring and/or replication
- Monitor state of mirroring and/or replication
TPC – optimize storage capacity

- Manage file system utilization rate
- Identify and manage non-critical data
  - Duplicates
  - Old
  - No owners
- Forecast data growth rate with AI.
TPC – optimize storage capacity

Monitor Storage
- Device Status
- Performance
- Alert
- Generate events based on error frequency
Recommend “Best Performing LUN” based on statistics such as:

- Average transfer size
- I/O rate (per sec, per GB)
- Random or sequential reads/writes
- Read cache hits
- Write destages
TPC – reduce data access failure

- Real time monitoring
- Automatic status report and alert
- Integrate with Tivoli System Management
- Integrate with 3rd party management system via SNMP.
- SAN error prediction
TPC – reduce human error caused data access failure

User defined script invoked by user defined event

Automatic file system expansion

SAN Volume Controller

IBM DS8000 IBM DS6000 Hitachi Lightning Hitachi Thunder HP EVA HP EMA EMC DMX EMC CLARiiON
TPC support virtualized storage
Full Spectrum of Storage Offering

Enterprise-class Disk

Mid-range Disk with FibreChannel

Mid-range Disk with S-ATA

Virtual Tape

Enterprise-class Tape

Mid-range Tape

Serial ATA disks

Enterprise Storage Server
DS8100/8300

FAS IT Storage Server
DS6800
DS4800

TotalStorage
Enterprise Storage Server
DS8100/8300

TotalStorage
Virtual Tape Server

TotalStorage
Enterprise and LTO Tape
Automation
Summary
TPC
IBM service help optimize
Your storage virtualization
We hike because the peak is there.
And it is not coming towards us.
IBM optimize Storage to facilitate your GRID project

Simplify

Virtualize

Automate
Thank You