Power Management in Storage Systems

Kaladhar Voruganti
Technical Director
CTO Office, Sunnyvale

June 12, 2009
Outline

- Power Consumption Background in Data Centers and Storage Systems
- Power Management Strategies for Storage Systems
- Power Management Metrics
- Impact of Server Power Management Strategies on Storage Systems
Storage Consumption Constantly Increasing

- More types of information are being digitized and stored persistently (emergence of newer types of applications)

- Data is being stored persistently for longer periods of time (for legal and sentimental reasons)

- More people are persistently storing their information (computer usage globally is increasing)
What is a Storage Controller?

Traditional Storage Controller

Cluster of thousands of Google Servers

Source: IBM Redbook DS8000

Source: CNET News April/2009
Where Does the Power Go in a Data Center?

50

35

9

4

Server/Storage

Cooling

UPS

Lighting

Percentages

Source: IDC Report 2008
Where Does the Power Go in a Storage Controller?

Google Box Power Consumption

<table>
<thead>
<tr>
<th>Component</th>
<th>Watts</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU (2)</td>
<td>10</td>
</tr>
<tr>
<td>Memory (4)</td>
<td>25</td>
</tr>
<tr>
<td>Disk (1)</td>
<td>36</td>
</tr>
<tr>
<td>PCI Slots (2)</td>
<td>50</td>
</tr>
<tr>
<td>Motherboard (1)</td>
<td>25</td>
</tr>
<tr>
<td>Fan (1)</td>
<td>80</td>
</tr>
</tbody>
</table>

Watts Per Disk Total Power Cost

<table>
<thead>
<tr>
<th>Component</th>
<th>Watts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>275</td>
</tr>
<tr>
<td>Disks + Shelf Tax (12)</td>
<td>288</td>
</tr>
</tbody>
</table>

Drive: 12 Watts
Shelf (Fans, Power Supply): 12 Watts
Cooling: 24 Watts

Source: Google Paper, ISCA 2007

Source: NetApp Internal Study
Storage Power Management Strategies

- **Hardware**
  - Can select the appropriate storage architecture
  - Can select the appropriate storage system
  - Can select the appropriate hardware features

- **Software**
  - Storage Efficiency
  - Migration and Spinning/Shutting Disks Down
Power Management Strategies (Hardware)

- Hardware Techniques
  - Architectural Level
    - DAS versus Storage Controllers
    - Single Node Battery versus UPS Technology
  - Storage Box Level
    - Disks versus SSDs
    - Efficient Power Sources
    - Use of higher capacity disks
    - Use of lower RPM disks
DAS versus Storage Controllers

- **DAS Storage**
  - Application and storage are co-located
  - Power consumption is efficient if box is used to run applications (e.g. map-reduce applications)
  - Power consumption is not efficient if the cluster nodes are only used to serve storage
  - Low powered shared-nothing nodes are being proposed for archival storage (Pergamum work from UC Santa Cruz)
Single Node Battery Versus Centralized UPS

- Large UPSs can reach 92 to 95 percent efficiency at full load
  - Operating at lower load results in inefficiency which results in the generation of heat
  - Need cooling to remove the heat from the data center
- By having 12 volt battery at each of the storage nodes, Google is able to get 99.9 percent efficiency

Source: CNET.com Article on Google, April, 2009
Power Management Strategies (Hardware)

- Hardware Techniques
  - Architectural Level
    - DAS versus Storage Controllers
    - Single Node Battery versus UPS Technology
  - Storage Box Level
    - Disks versus SSDs
    - Efficient Power Sources
    - Use of higher capacity disks
    - Use of lower RPM disks
Efficient Power Supply

- Want Power Supplies that are Efficient for a wider range of load
  - These cost more
  - But offer savings in power consumption due to less heat generation (less cooling required)

- If there are multiple Power Supplies usually each one is run at lesser load, and thus, has higher inefficiency

<table>
<thead>
<tr>
<th>Load</th>
<th>Less Efficient Power Supply (Efficiency)</th>
<th>More Efficient Power Supply (Efficiency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 % Load</td>
<td>50 % Eff</td>
<td>80 % Eff</td>
</tr>
<tr>
<td>50 % Load</td>
<td>75% Eff</td>
<td>80% Eff</td>
</tr>
<tr>
<td>100 % Load</td>
<td>75% Eff</td>
<td>90% Eff</td>
</tr>
</tbody>
</table>

Source: NetApp Internal Study
# Higher Capacity Disks

Source: NetApp White Paper: WP 7010-0207

<table>
<thead>
<tr>
<th></th>
<th>Old Systems</th>
<th>New Systems</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td># of systems</td>
<td>11 Old Systems:</td>
<td>1 FAS 3020 with 3 disk shelves</td>
<td>11:1</td>
</tr>
<tr>
<td></td>
<td>4 F880</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 F810</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 F820</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 F825</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 F840</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power* (kW HRs)</td>
<td>113,651</td>
<td>20,915</td>
<td>81% Decrease</td>
</tr>
<tr>
<td>* Does not include power for cooling.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space (Cubic Feet)</td>
<td>63.0</td>
<td>4.3</td>
<td>93% Decrease</td>
</tr>
<tr>
<td>Capacity (GBs)</td>
<td>9,776</td>
<td>14,000</td>
<td>16% Increase</td>
</tr>
</tbody>
</table>
Power Management Strategies (Software)

- Software Techniques
  - Storage Efficiency
    - Reduce overhead per amount of usable storage
    - Efficient Copies
    - Data De-duplication/Compression
    - Thin Provisioning
    - Number of copies
    - Protection Mechanism
  - Migration and disk shutdown/spin-down
# Storage Efficiency Techniques

*Source: Oliver Wyman Article, Dec 2007 “Making Green IT a Reality”*

- Reducing Storage Overhead
- Thin Provisioning
- Efficient Protection Mechanisms
- Consolidation of Protocols
- Efficient Copies
- De-duplication/Compression

<table>
<thead>
<tr>
<th>Capability</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snapshot</td>
<td>20% storage overhead compared to 100% for full copies or BCV</td>
</tr>
<tr>
<td>Thin Provisioning (FlexVol)</td>
<td>20% - 33% savings by growing and shrinking volume sizes on demand</td>
</tr>
<tr>
<td>RAID 6 Implementation (RAID-DP™)</td>
<td>14% - 17% overhead compared to 100% for RAID 1</td>
</tr>
<tr>
<td>Multi-protocol (Unified Storage) &amp; FC/SATA Drives</td>
<td>2-3x savings if running multiple protocols with low util rates; up to 50% savings with SATA instead of FC</td>
</tr>
<tr>
<td>Multiple Writeable Snapshot Copies (FlexClone®)</td>
<td>Up to 66% savings if creating five copies of original data compared to BCV or full copies</td>
</tr>
<tr>
<td>Deduplication</td>
<td>10% - 80% space savings depending on data set</td>
</tr>
</tbody>
</table>
Disk Spin-down/Shut-down Techniques

- Migrate less accessed data to lower tiers of storage and shut-down disks
- Archival data can be stored on disks that are shut down because of write-once and read-maybe properties
- Difficult to shut-down disks for those applications that have strict latency requirements and have long-tailed distribution access patterns
- Spinning things down to lower RPM and then spinning them up is difficult because constant spinning up disks can actually result in higher power consumption
- COPAN has shown roughly 5x times power savings compared to normal storage controllers in cases where things can be shut down.
  - Very dense packaging than traditional storage controllers
  - Keeps application meta-data in cache
  - Spins disks down but keeps the electronics up
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Online</th>
<th>Near online</th>
<th>Removable Media</th>
<th>Virtual Media Library</th>
<th>Appliance</th>
<th>Interconnect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Pattern</td>
<td>Random</td>
<td>Random</td>
<td>Sequential write</td>
<td>Sequential write</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaxTTD (t)</td>
<td>t &lt; 80 ms</td>
<td>t &gt; 80 ms</td>
<td>t &gt; 80 ms</td>
<td>t &lt; 80 ms</td>
<td>t &lt; 80 ms</td>
<td>t &lt; 80 ms</td>
</tr>
<tr>
<td>User accessible data</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Prohibited</td>
<td>Prohibited</td>
</tr>
</tbody>
</table>
Configuration & Workload Example

Storage Subsystem/Server Under Test

Workload Generator
IOmeter

AC Voltage 115V

Watt Meter
True RMS

Temperature sensor
Ambient

Interface Cable
FC/SAS/SATA/SCSI

Temperature °C data logging

High Performance
Server
Windows

Watt data logging Cable
SNIA Green Power Profile Example

- **Green Power Profile**
  - **Phase 0**
    - Preconditioning phase no power or IO measurements are necessary
    - 5 minutes maximum OLTP workload
    - 5 minutes no workload (idle)
    - 5 minutes maximum OLTP workload
  - **Phase 1**
    - Idle measurement phase. No user initiated commands allowed
    - 60 minutes measurement period
  - **Phase 2**
    - OLTP workload
    - 20 minutes measurement period
    - 5 minutes rest no OLTP workload
  - **Phase 3**
    - Sequential Throughput 50% Read 50% Write 1MB transfer
    - 20 Minutes measurement period
SNIA Green Power Profile Example

Green Power Profile

High End Storage Subsystem Example
End User Capacity = 1.6TB

- Phase 0
  - 272 Watts @ Idle

- Phase 1
  - 356 Watts @ 6680 IOs/s

- Phase 2
  - 308 Watts @ 809 MB/s

- Phase 3

AC Watts rms

Time 5 mins/div

© 2008 NetApp. All rights reserved.
# Power Management Metrics

<table>
<thead>
<tr>
<th>I/O Performance (OLTP Type Workloads)</th>
<th>IOPs/Watt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>GB/Watt</td>
</tr>
<tr>
<td>Usable Capacity</td>
<td>GB/Watt</td>
</tr>
<tr>
<td>Sequential I/O Throughput</td>
<td>MBps/Watt</td>
</tr>
<tr>
<td>Availability</td>
<td>RTO/Watt 9s/Watt</td>
</tr>
</tbody>
</table>
Impact of Server Virtualization on Storage

- Server Virtualization is being used to consolidate physical servers to obtain cost, space and power efficiencies

- Impact of Server Virtualization on Storage
  - Need for shared storage
  - Need for efficient storage with de-duplication
  - I/O interference due to Multi-tenancy
  - Mis-match of Hypervisor and Storage constructs
  - Block Mis-alignment due to layers of storage software
Conclusion

- Key takeaways
  - Storage Efficiency is the primary mechanism for saving power by having fewer number of disks
  - Flash is emerging as a power efficient alternative to disks (price is still an issue)
  - Shutting-down disks is only attractive for archival storage

- Need for:
  - Power Management Metrics Standards
  - Power aware storage management tools