February 2010

Towards Integrated Data Center Energy Management

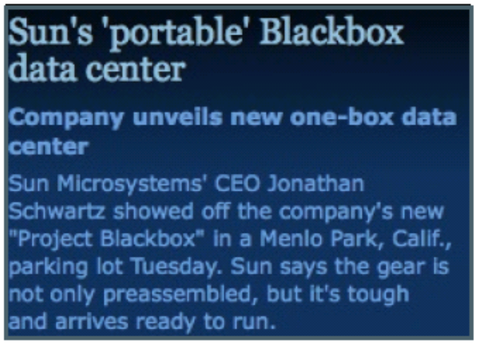
An IBM Research Strategic Initiative

Jody Glider (gliderj@almaden.ibm.com)

IBM Data Center Energy Management Strategic Initiative

Lots of focus!

- Prior to 2001
  - $(IT equipment cost) > $(Power + Cooling)
- After 2001
  - $(Power + Cooling) > $(IT equipment cost)



Sun's 'portable' Blackbox data center

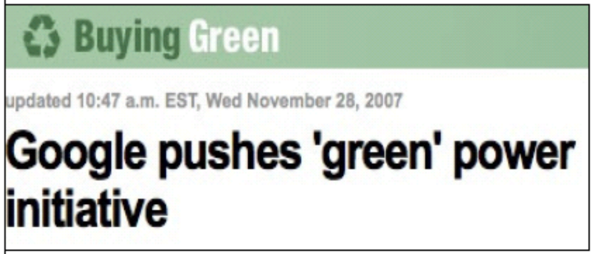
Company unveils new one-box data center

Sun Microsystems' CEO Jonathan Schwartz showed off the company's new "Project Blackbox" in a Menlo Park, Calif., parking lot Tuesday. Sun says the gear is not only preassembled, but it's tough and arrives ready to run.

HP's Green Business Technology Initiative

Innovative Dynamic Smart Cooling

Cut cooling costs in the data center as much as 40%.



Buying Green

updated 10:47 a.m. EST, Wed November 28, 2007

Google pushes 'green' power initiative

Intel Becomes Largest Purchaser of Green Power in the U.S.

Company Tops EPA Green Power Partner List, Vows to Drive for Greater Efficiency While Spurring Growth in Renewable Market

How Microsoft is going green

Biodiesel trucks, solar-powered data centers are just a couple environmentally friendly track

By John Fontana , Network World , 01/09/2008

IBM Project Big Green

Big Green Banner

Project Big Green is a $1 billion investment to dramatically increase the efficiency of IBM products. New IBM products and services,

IBM’s Efforts to Improve Data Center Energy Efficiency

- IBM announced Project Big Green in May 2007
  - Large set of data center energy monitoring and management efforts across IBM’s Software, Systems and Services divisions to create a family of energy-saving solutions that span the entire stack
    - Hardware platform
    - Software stack (OS to middleware)
    - Facilities (cooling and more)
  - $1B reallocated to Project Big Green per year
    - Guarantee R&D funding for IT energy efficiency technology.
    - Convert IBM facilities to Green Data Centers using energy-efficient technologies & services, building on IBM energy conservation measures started in 1990
    - Use virtualization as technology accelerator for IBM’s Green Data Centers
- IBM Research is a major source of innovation (and liaison) for the Software, Systems and Services divisions of IBM
- In 2009, Research launched Data Center Energy Management strategic initiative (DCEM SI)
  - Identify and exploit synergies across the worldwide Research laboratories
  - Project Big Green in microcosm

DCEM SI Mission

- Deliver energy monitoring and management products and solutions to the market by working with IBM product and service partners
  - innovatively augment, complement and integrate point products and solutions into comprehensive data center solutions that address customers' most important needs

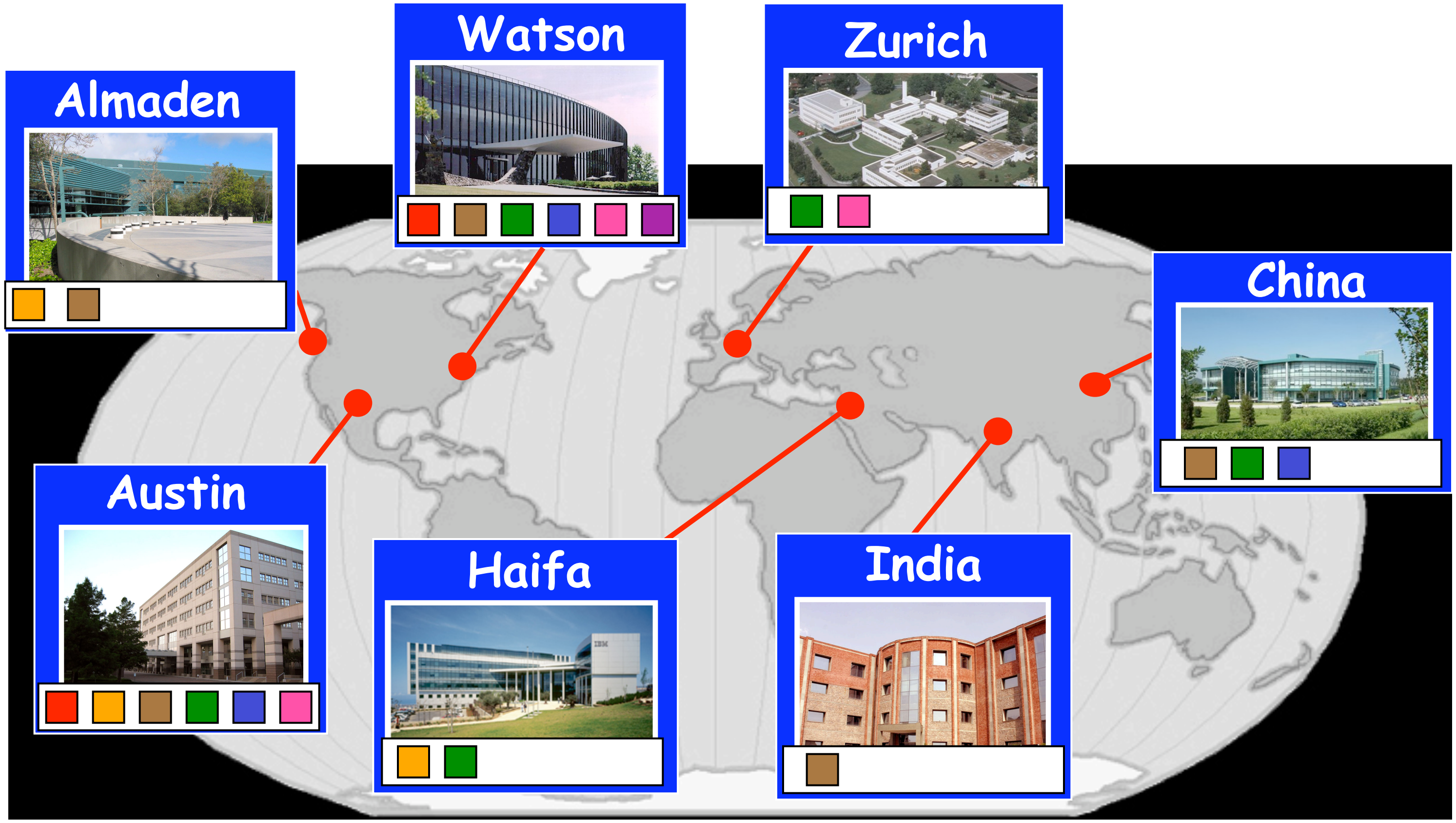
- Develop and showcase these solutions in a living data center lab

- Deploy these solutions within data centers operated by IBM, and quantify the resulting financial and energy savings

- Deliver ideas and code as appropriate to IBM development and services partners

DCEM SI projects: a global research effort

- **Processor microarchitecture, memory design**
- **Energy-proportional storage**
- **Energy-aware dynamic server consolidation**
- **Monitoring and visualization**
- **What-if analysis and decision support**
- **Novel approaches to DC cooling**
- **Risk analysis and optimization for energy-efficient DC design**



- Over 15 Research teams collaborate on projects that
  - develop energy-efficiency technologies at all levels of the stack, across the widest span of spatial and temporal scales from microprocessor (microseconds) to data center (minutes to months)
  - explore algorithms and architectures for coordinating multiple energy-efficiency technologies

Outline for the rest of the talk

- Showcase of a few data center projects
  - Mobile Measurement Technology
  - Raleigh Leadership Data Center
- Future directions for DCEM SI
- Storage Energy Management Research
  - Motivation
  - Modeling and monitoring
  - Optimization techniques
    - Tiering of HDD/SDD
    - SSD caching
    - Consolidation and Migration between HDD/SDD

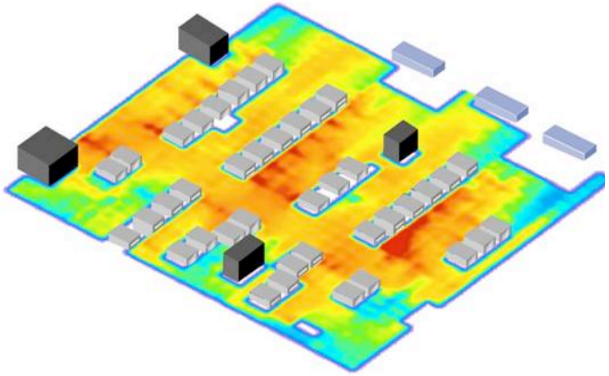
IBM Mobile Measurement Technology (MMT)

Solution Approach – *Three Steps*

1 Measure

Capture *high resolution temperature data*, *air flow data* *infrastructure & layout data*

2 Model



To identify improvement opportunities *model the data center* and use optimization algorithms ("*best practices rules*")

3 Manage "*Best Practices*"

- Realize air transport energy savings
- Realize thermodynamic energy savings

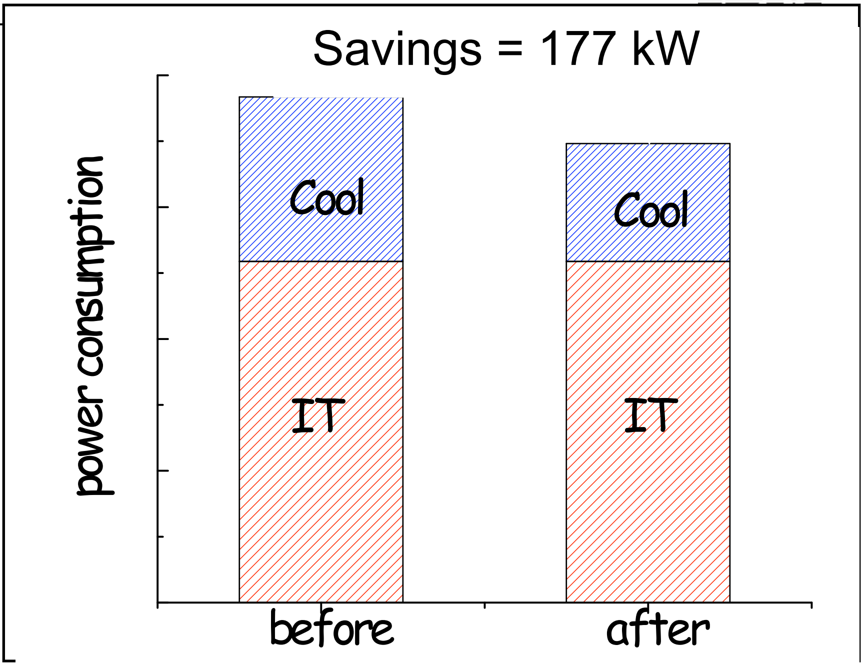
→ *Achieve reduced energy consumption*

→ *Potential for deferring new investments*

Contact: hendrikh@us.ibm.com

Typical Energy Savings

- Saved 177 kW with measurement / metrics driven best practices implementation
- Developed 6-tier metric to drive best practices implementation with minimal investments
- ROI: 1-2 Months
- Improved DC COP 2.39 to 3.44
  - COP_{thermo} from 4.5 to 5.1
  - COP_{trans} from 5.3 to 9.8



Finding / Metrics	Key Action / Solution
Horizontal hotspots (HH)	change tile layout & deploy high throughput tiles
Vertical hotspots (VH)	snorkels / fillers
Non-targeted air flow	close leaks / cable cutouts
Plenum temperatures	service ACUs supply side / increase ACU utilization
ACU utilization	turn under-utilized ACUs off
ACU flow	remove blockage

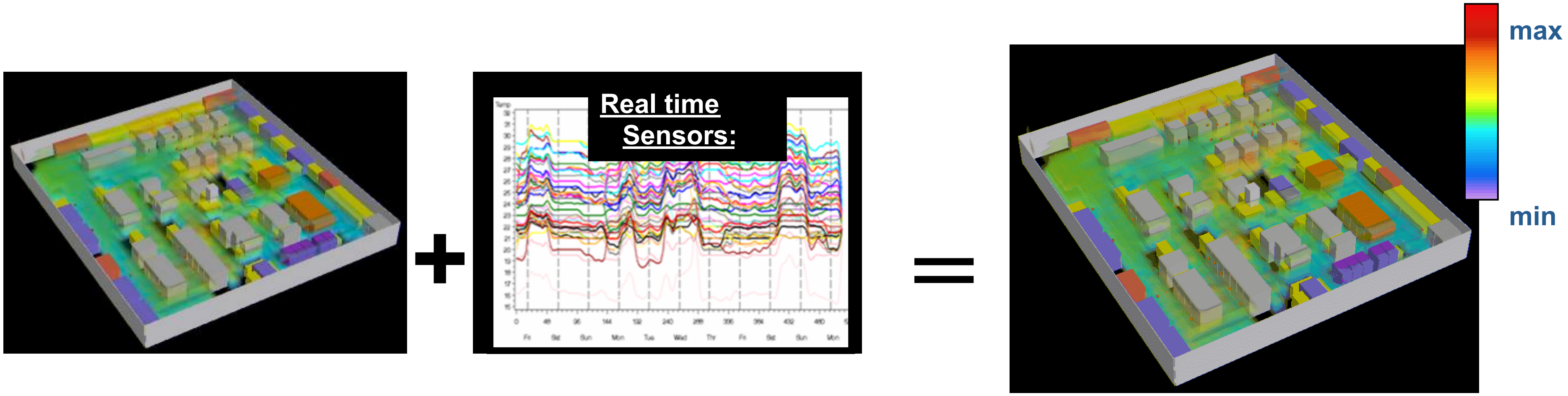
thermo

transport

Case Study: *DC Area = 20k sqf; Temp. Meas. = 200,000; Airflow Meas. = 1,200; Power density ~ 75 W / sqf*

MMT: Evolution from static to dynamic solution

- DC can change over time
  - IT power levels can change (e.g., 10-15 % during a day)
  - cooling conditions change etc..
  - new racks / new servers / re-arrangement of tiles etc..
- MMT 1.0 is “sparse” in time but “dense” in space
- Real-time sensor are “sparse” in space but dense in time
- MMT 1.5 provides high time & spatial resolution combining
  - MMT 1.0 for **base model** generation, sensor placement etc..
  - real-time sensors for creating **dynamic models**



Animation of 3D heat map over 24 hours

Summary of IBM Raleigh Leadership Data Center

IT Related

Design / Build

- 100,000 sq ft raised floor
- Economies of scale – 200,000 sq ft site
- High availability design point
- Support liquid cooled IT equipment
- Up to 15MW of IT power

Operate

- Cloud computing support for public and private clouds
- Energy efficient of DCiE 66% to 71%
- High virtualization utilization
- Best practices equipment layout with hot / cold aisle configuration
- Continuously self-regulating energy efficient data center with integrated IT, facilities and building automation systems
- Real-time CFD monitoring and control for IT equipment placement

Contact: kephart@us.ibm.com

Facilities Related

Cooling

- Free cooling for 44% of the calendar year with water-side economizer
- Variable speed drives on chilled water pumping, chillers and air handlers
- Variable speed motors on CRAC's
- Increased set point to 50ºF chilled water supply

Electrical

- Modular power density – up to 32 kilowatts per rack
- High efficiency components for UPS, PDU
- Support for high density zones

Other building systems

- Energy Efficient Lighting
- High "R" Value Insulation

Green

LEED Certification Objective

- Pursuing LEED Gold
- 95% original building shell reused
- 80% original materials recycled
- 20% new materials from recycled products
- Harvesting rainwater

Energy Management Programs

- Government Incentives of $ 750,000
- Renewable Energy Certificates

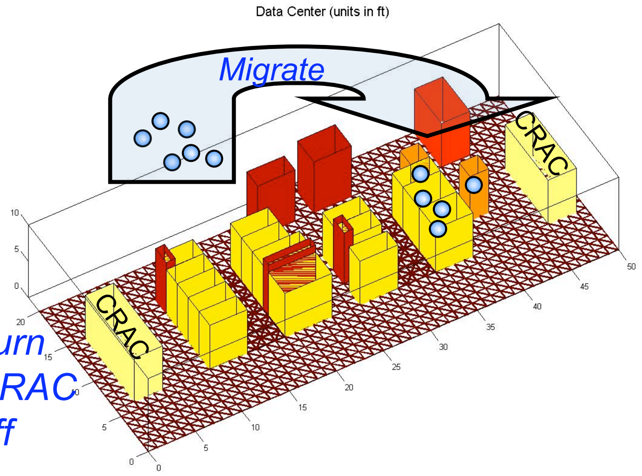
Data Center Energy Management Strategic Initiative 2010

Towards Greater Energy Proportionality

- Ideally, energy use in data centers should be *proportional* to useful computation
  - But today, underutilized resources waste lots of energy at all scales: chip to device to data center
- We are placing special focus on improved energy proportionality in two key realms
- **1. Storage systems.** Develop intelligent energy-aware hierarchical storage management technologies
- **2. Coupled cooling & workload migration.** Use workload migration to move computation to most efficient region, and adjust cooling intelligently to save energy
- **Other goals**: continuing stream of energy visualization, analytics, and management innovation into IBM products and services

	SATA	SAS/FC	SSD/Flash
Capacity	High	Low	Low
Performance	Low	High	High
Energy use	Low	High	Low

Intelligently manage storage to achieve acceptable capacity, performance and energy consumption.

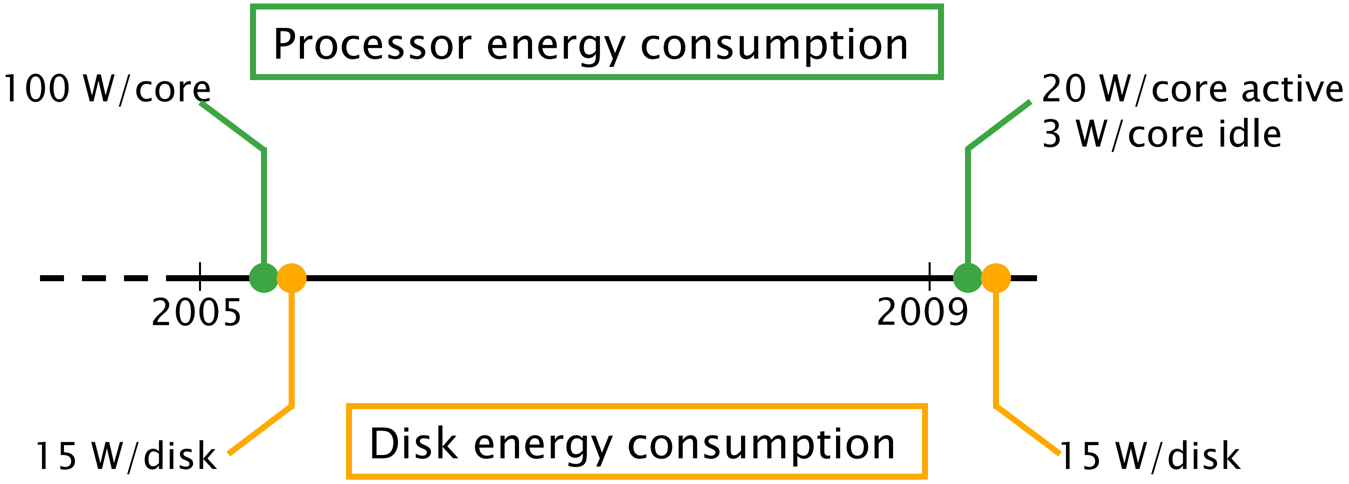


Increase CRAC fan speed: 60%→80%

Turn CRAC off

IBM Storage energy management research

Storage energy consumption is becoming significant



Processor energy consumption

100 W/core

20 W/core active
3 W/core idle

15 W/disk

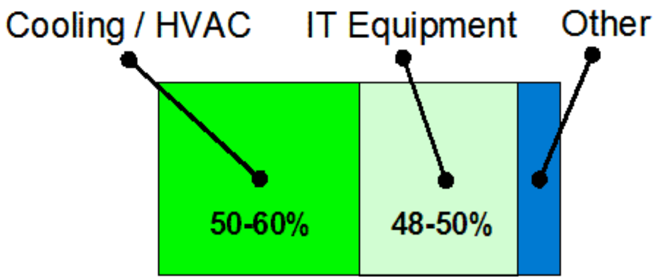
Disk energy consumption

15 W/disk

- Disk drive power consumption remains the same
- Processors become energy efficient

Storage is consuming increasing percentage of data center energy

Breakdown of data center power



Source: StorageIO, Greg Sculz www.storageio.com

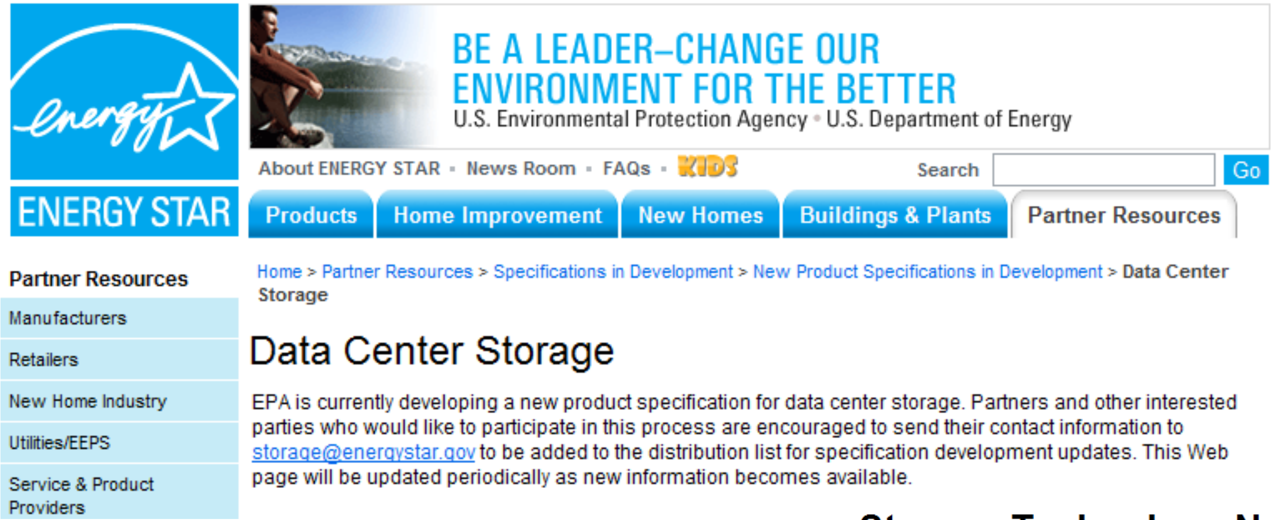
Storage accounts for up to 37-40% of the energy consumption of all IT components

Storage energy usage will get worse…

D. Reinsel. The Real Costs to Power and Cool All the World’s External Storage. IDC Report, Doc 212714, 2008

- Storage unit sales will outpace server unit sales in the next 5 years
- Storage capacity increases per drive are slowing
  - More physical drives to keep up with capacity demands
- Storage performance has not kept up with the increasing capacity
  - More physical drives to keep up with performance demands
- Data centers are moving towards 2.5” drives
  - 2.5” drives typically consume more energy/GByte than 3.5” drives

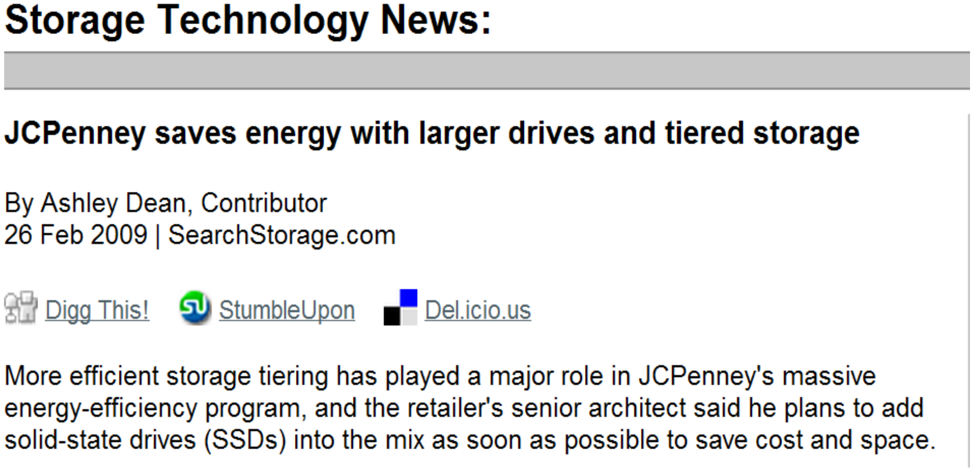
Industry is taking note!



Home > Partner Resources > Specifications in Development > New Product Specifications in Development > Data Center Storage

Data Center Storage

EPA is currently developing a new product specification for data center storage. Partners and other interested parties who would like to participate in this process are encouraged to send their contact information to storage@energystar.gov to be added to the distribution list for specification development updates. This Web page will be updated periodically as new information becomes available.

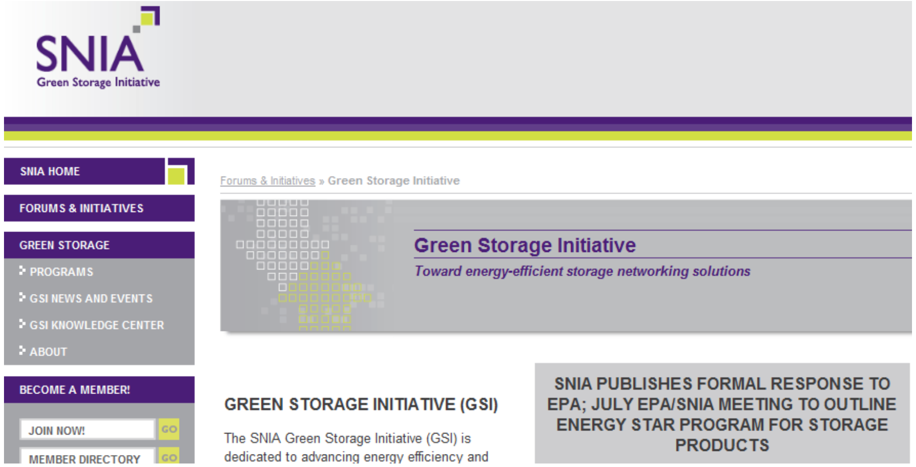


Storage Technology News:

JCPenney saves energy with larger drives and tiered storage

By Ashley Dean, Contributor
26 Feb 2009 | SearchStorage.com

More efficient storage tiering has played a major role in JCPenney's massive energy-efficiency program, and the retailer's senior architect said he plans to add solid-state drives (SSDs) into the mix as soon as possible to save cost and space.



Forums & Initiatives » Green Storage Initiative

Green Storage Initiative

Toward energy-efficient storage networking solutions

GREEN STORAGE INITIATIVE (GSI)

The SNIA Green Storage Initiative (GSI) is dedicated to advancing energy efficiency and

SNIA PUBLISHES FORMAL RESPONSE TO EPA; JULY EPA/SNIA MEETING TO OUTLINE ENERGY STAR PROGRAM FOR STORAGE PRODUCTS

Strategic directions for Storage Energy Efficiency

Monitoring and Modeling

- Storage systems MUST get better at monitoring energy…one energy number per petabyte machine ain’t enough!
- Better monitoring will enable better control and optimization
- Generating effective models enables planning and optimization

Dr. Phil says: “You can’t fix what you don’t acknowledge”!

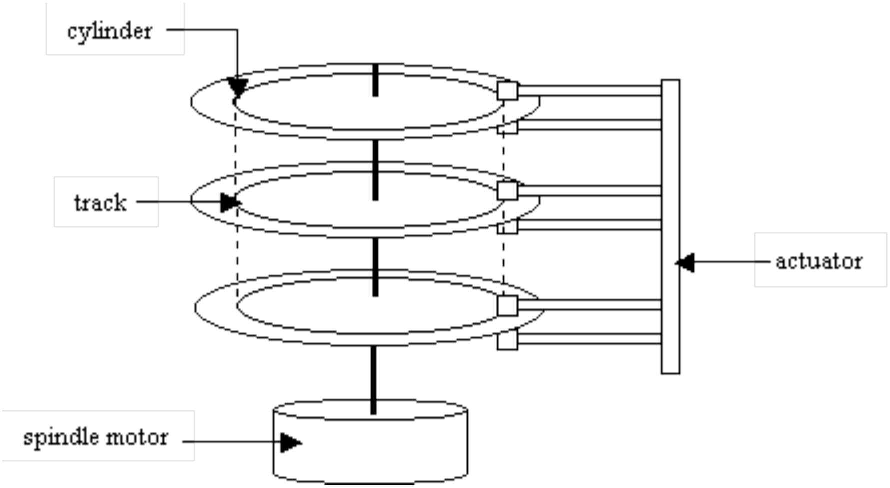
Control and Optimization

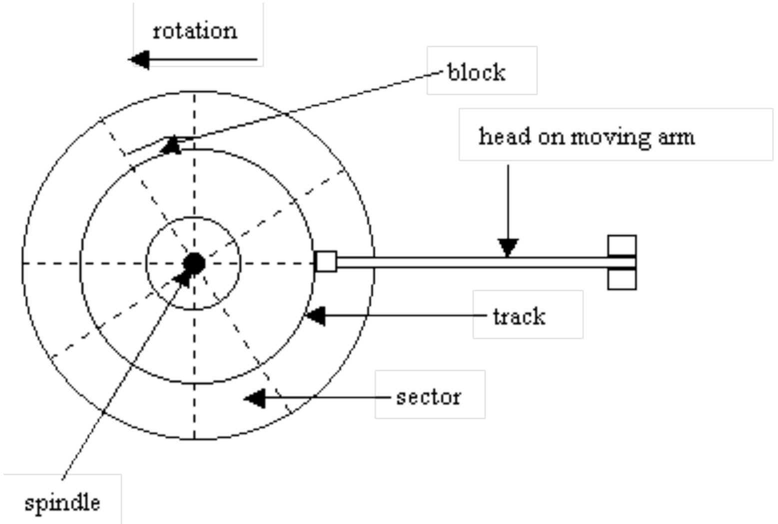
- Our vision is ‘energy proportionality’, which boils down to:
  - **When performance requirements matter most**: Energy should vary with workload. Storage systems should regulate energy consumption to match workload.
  - **When energy constraints matter most**: Performance provided by the storage should be vary to match energy constraints. Storage should regulate performance to lower the energy.

Reporting, control and optimization need to be tied into the bigger data center energy management environment.

Power modeling

Introduction to Power Modeling for Storage





Fixed: Spindle power, electronics

Dynamic: Seek activity, data transfer

12V

5V

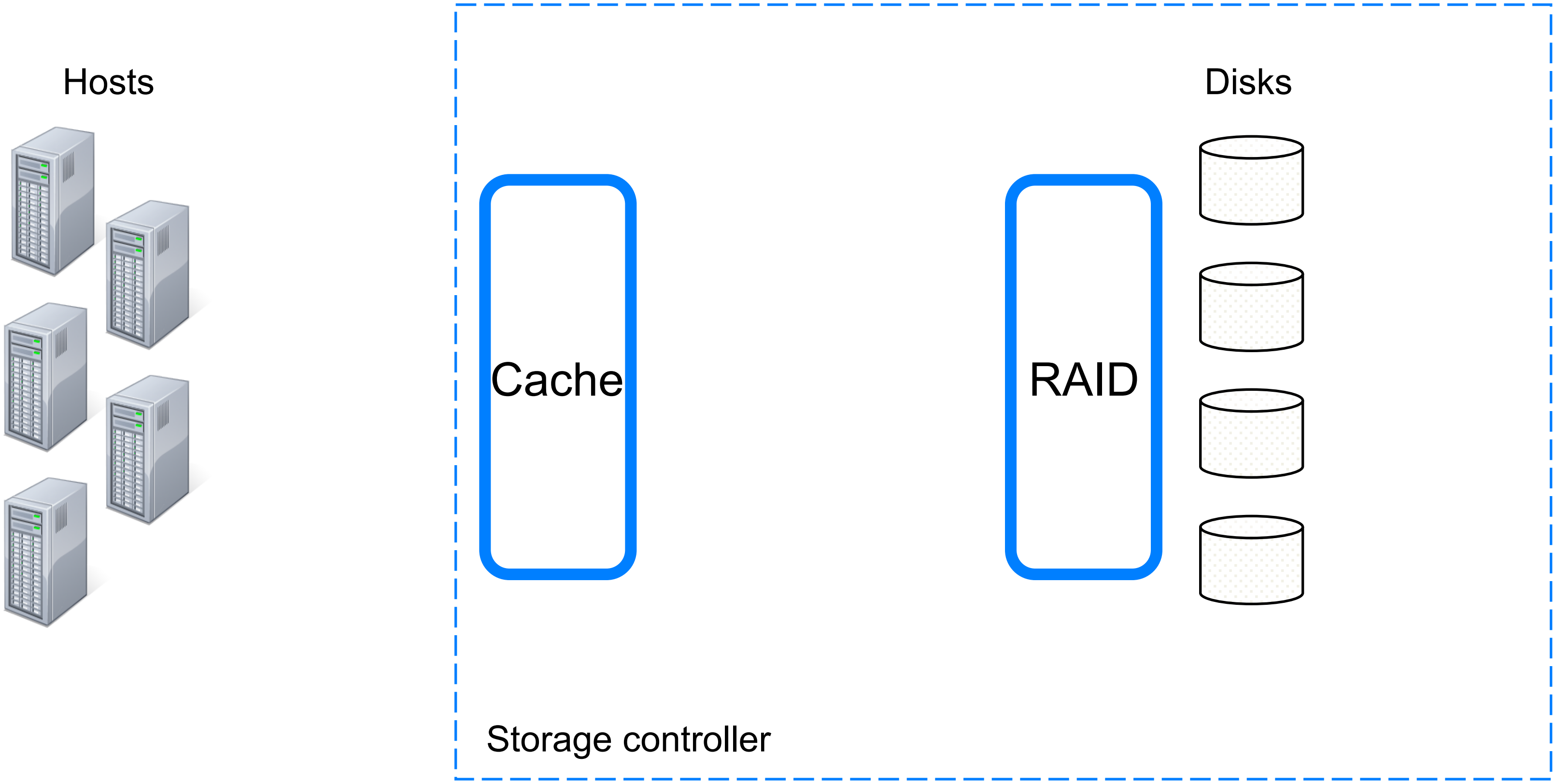
$$\mathrm{Power(disk)} = \mathrm{Fixed}_{\mathrm{power}} + \mathrm{Dynamic}_{\mathrm{power}}$$

$$\mathrm{Fixed}_{\mathrm{power}} = \mathrm{Power(spindle)} + \mathrm{Power(electronics)}$$

$$Dynamic_{power} = \sum_{\mathrm{I/O}} [Power(\mathrm{seek}) + Power(\mathrm{datasize})]$$

Contact: meth@il.ibm.com

Introduction to Power Modeling for Storage



Hosts

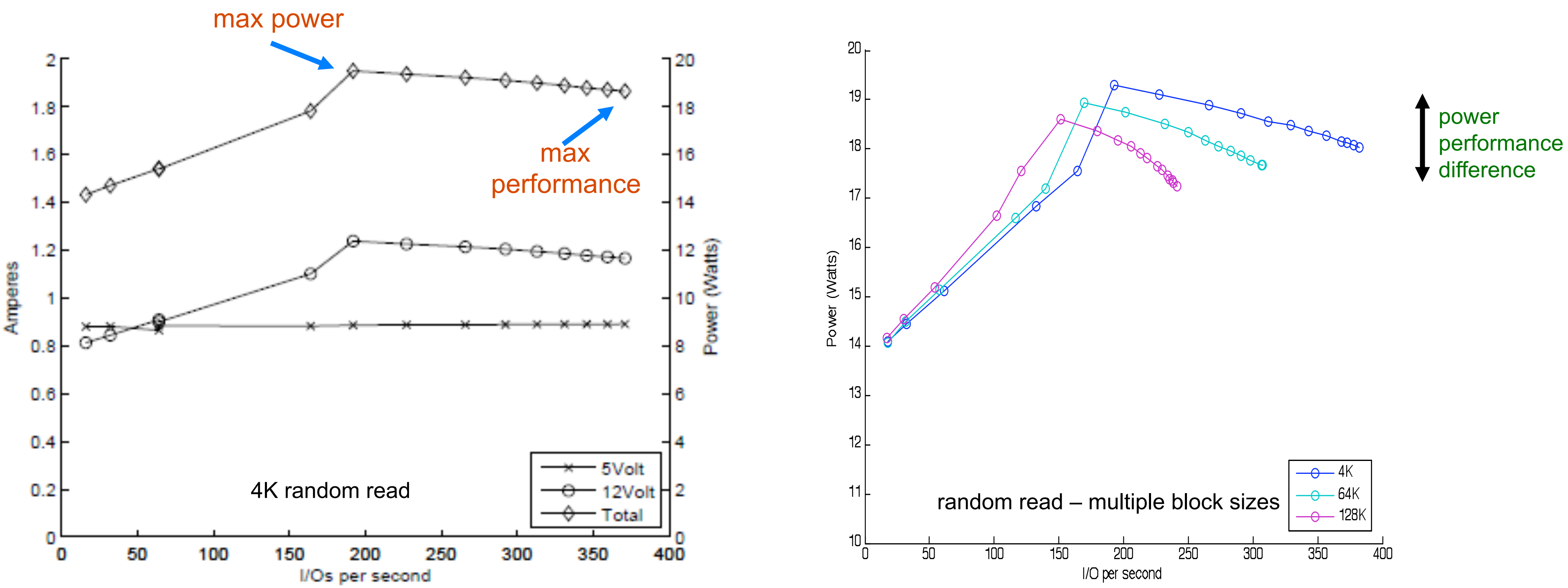
Disks

Cache

RAID

Storage controller

Initial Measurements Results



max power

max performance

power performance difference

300GB 15K Enterprise FC disk drive

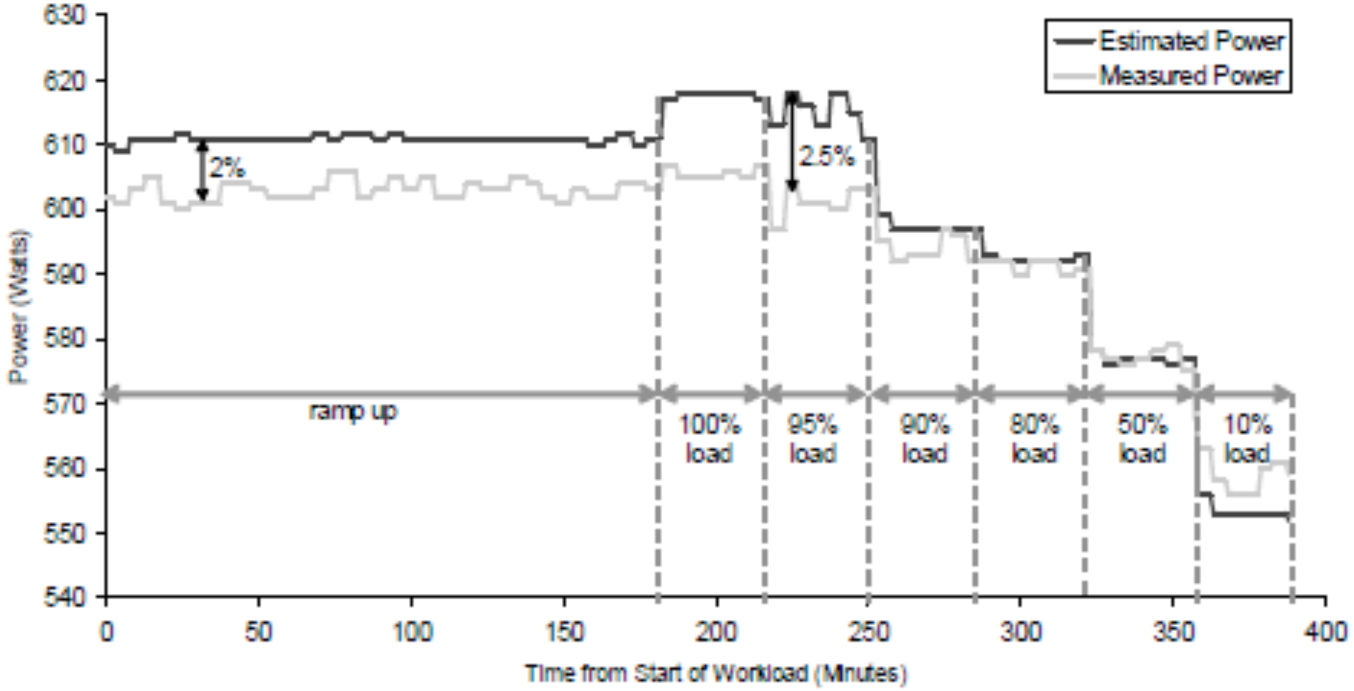
Power consumption is determined by primitive activities:

- Seek activity power (12V)
- Data transfer activity power (5V)

Power Estimation Methodology (cont)

- Power tables
  - Span the performance range of the disk
  - Seek activity <#seek, power>
  - Data transfer <MBPS, power>
- Interpolation
  - Use the data point in the power tables for each activity and use interpolation to estimate disk power
- Assumptions
  - Uniform distribution over the array disks
  - Uniform distribution over the disk blocks

Power Model Validation using SPC-1 Benchmark



SPC-1 workload

- SPC-1 Industry standard benchmark - www.storageperformance.org
- Estimation error of up to 2.5%
- Power consumption drops with reduction of workload

Energy Optimization

Optimization techniques comparison

Technique	Time to adapt	Run-time delays	Potential energy savings (rough guesstimate)
Consolidation	Hours	none	50%
Tiering	Mins/hours	none	75%
Opportunistic spindown	secs	secs	25%
MAID	secs	secs	75%
Write Off-loading	msecs	msecs	25%
Adaptive seek speeds	msecs	msecs	25%
Workload shaping	msecs	msecs	25%
Thin provisioning	n/a	none	50%
Space-efficient snapshot	n/a	none	50%
Deduplication/compression	n/a	msecs	50% (active/archive), 90% (backup)

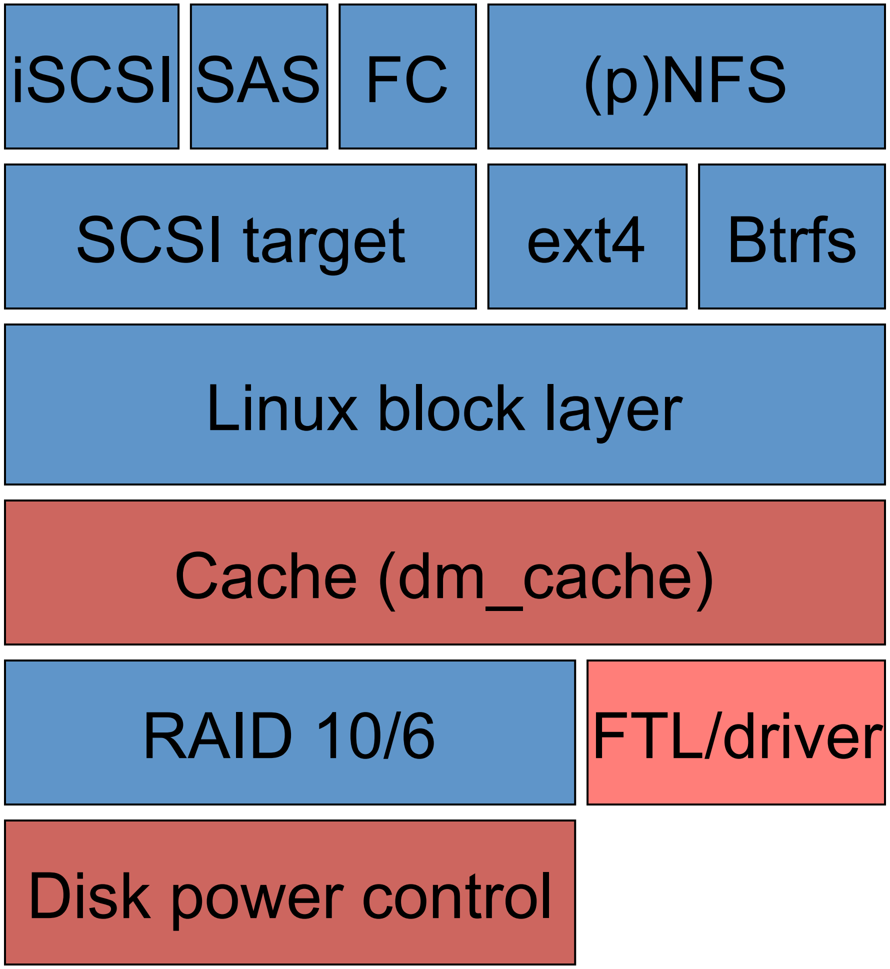
Optimization techniques need to be matched to environment

Tiering: Cost/Energy/Performance for different HDD/SSD combinations

Config	Typical Energy (Watts)	Total cost over two years (k$)	Total best case IOPs available in configuration (kIOPS)
All 7k RPM SATA	890.00	11.18	10.00
All SSD	1875.00	2004.60	18750.00
All 15k RPM SAS	3777.78	120.38	77.78
SATA + 1% SSD	908.75	31.23	197.50
SATA + 5% SSD	983.75	111.41	947.50
SATA + 10% SSD	1077.50	211.64	1885.00
50%-populated 15k SAS	7555.56	240.75	233.33
Full-pop 15k SAS + 1% SSD	3796.53	140.42	265.28
Full-pop 15k SAS + 5% SSD	3871.53	220.61	1015.28
Full-pop 15k SAS + 10% SSD	3965.28	320.84	1952.78

Tiering via FLASH cache

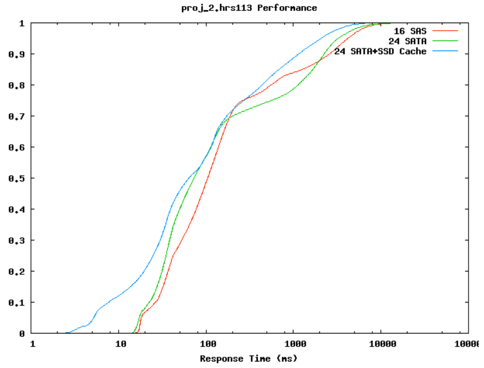
- Block-layer shim above RAID layer, below filesystem
- Based on Linux device mapper framework
- Caches all disk accesses onto SSD
  - Reads can bypass SSD when overloaded (if queue depth > 10)
- 32KB block size
- LRU replacement
- Tracks last access time for each underlying RAID rank (8 disks)
- Spin down a RAID rank that has been idle for N seconds
- Cache can be partitioned by RAID rank:
  - Baseline: no cache partitioning
  - Simple: each rank gets equal cache space
  - Power-aware: adjust partition size based on disk power state



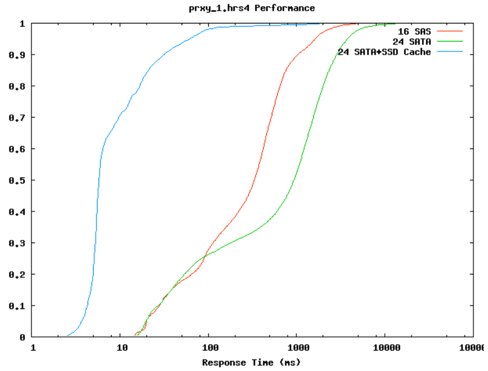
Contact: retrac@us.ibm.com

Some results simulating MSR traces (24 SATA vs. 16 SSD)

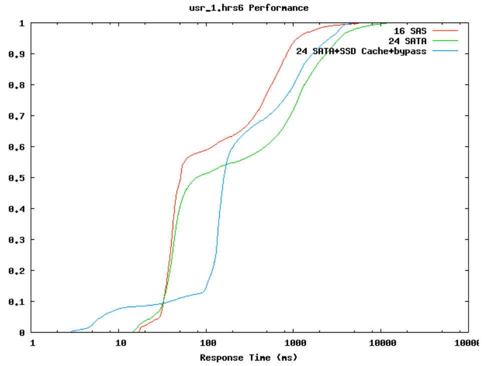
- Project file server trace: 7% less energy than SAS



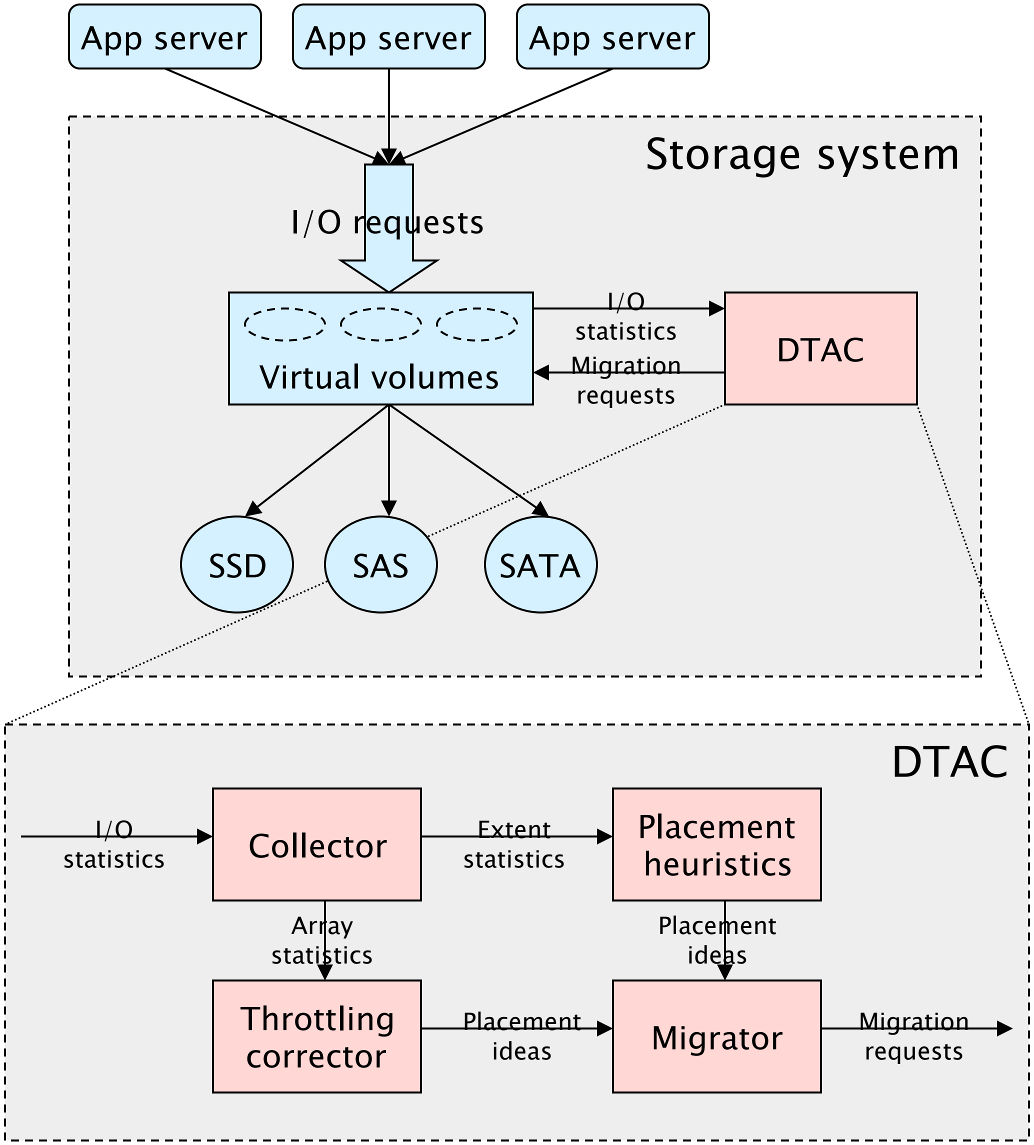
- Firewall/web proxy trace: 11% less energy than SAS



- User home directory trace: 11% less energy than SAS



Dynamic Tiering and Consolidation extent migration



- Oriented toward critical performance enterprise workloads
- Extent-based dynamic placement of extents into tiers of storage
- Match the performance requirements with the most appropriate tier
- Consolidate where possible and turn off drives that hold no data

Contact: hpucha@us.ibm.com

Some DTAC experimental results

Scenarios

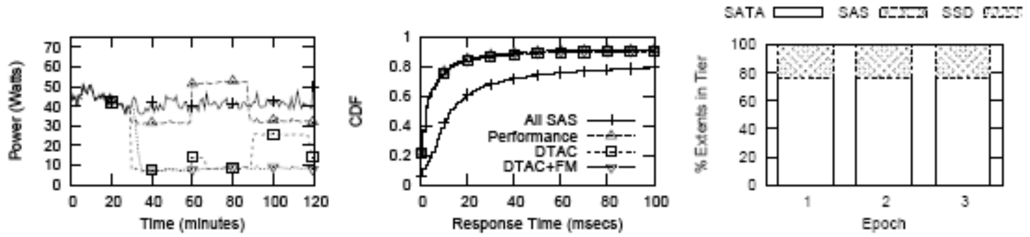
- First row: Space utilization high, IOPS low
- Second row: Space utilization low, IOPS utilization high
- Third row: Space and IOPS utilization medium
- Fourth row: Space and IOPS utilization high

Methodology

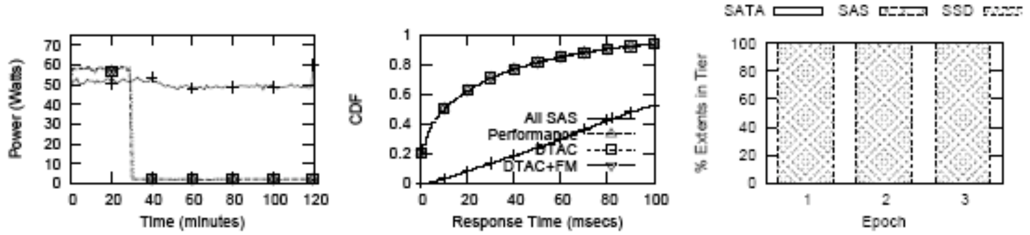
- Run selected MSR traces against prototype system for 2 hours, with epoch size of 30 minutes.
- One free migration after 30 min to ‘warm’ the placement

Summary of results

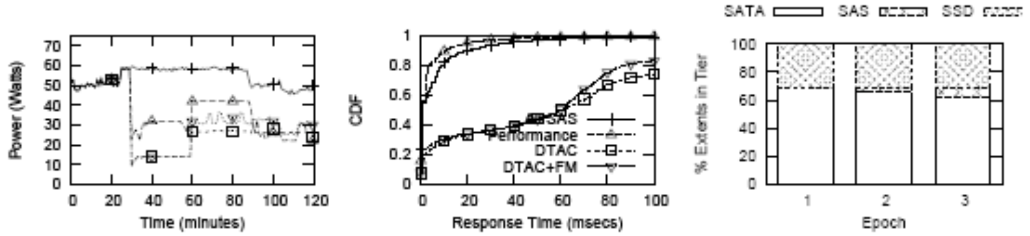
- Performance was generally better than all SAS configuration
- Energy consumption was always less than all SAS configuration



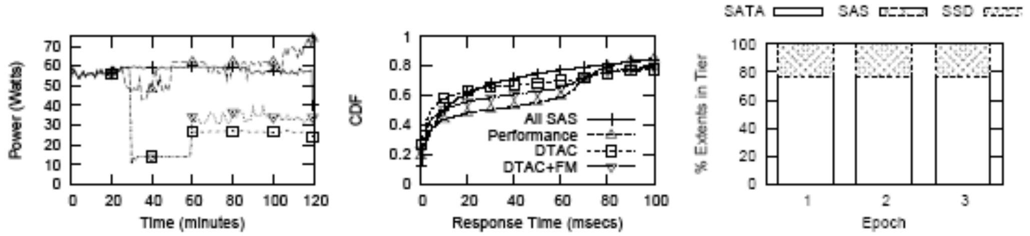
(a) Scenario 1



(b) Scenario 2



(c) Scenario 3



(d) Scenario 4

Future directions

Toward the vision: Possible areas for focus

- Push storage systems toward more variable energy cost components and less fixed cost components
  - Fixed cost components today can easily account for 40%-60% of storage systems energy consumption
- Demand response planning
  - The flipside of energy proportionality
  - Important in developing countries where electricity supply doesn't run 24/7
  - Important in developed countries where energy is increasingly costly and increasingly constrained
- Coordinated energy consumption optimization of all IT equipment and cooling
  - Ensure that optimizations in one area don't defeat optimizations elsewhere (e.g. storage system slowdown defeating servers hurrying to get work done before turning off)
  - Unified modeling/analysis to determine best operating point given performance/energy considerations of all equipment

Thank you!

Questions?