



System Architectures for Cloud Computing

Richard Kaufmann

Grizzled Hack, Scalable Computing & Infrastructure Group

ISC, June 2009

© 2008 Hewlett-Packard Development Company, L.P.
The information contained herein is subject to change without notice



Disclaimer

HP makes no warranties regarding the accuracy of this information. HP does not warrant or represent that it will introduce any product to which the information relates. It is presented for sheer entertainment value only. Trademarks remain the property of their original owners. So there.



"chocolate ice cream","vanilla ice cream"

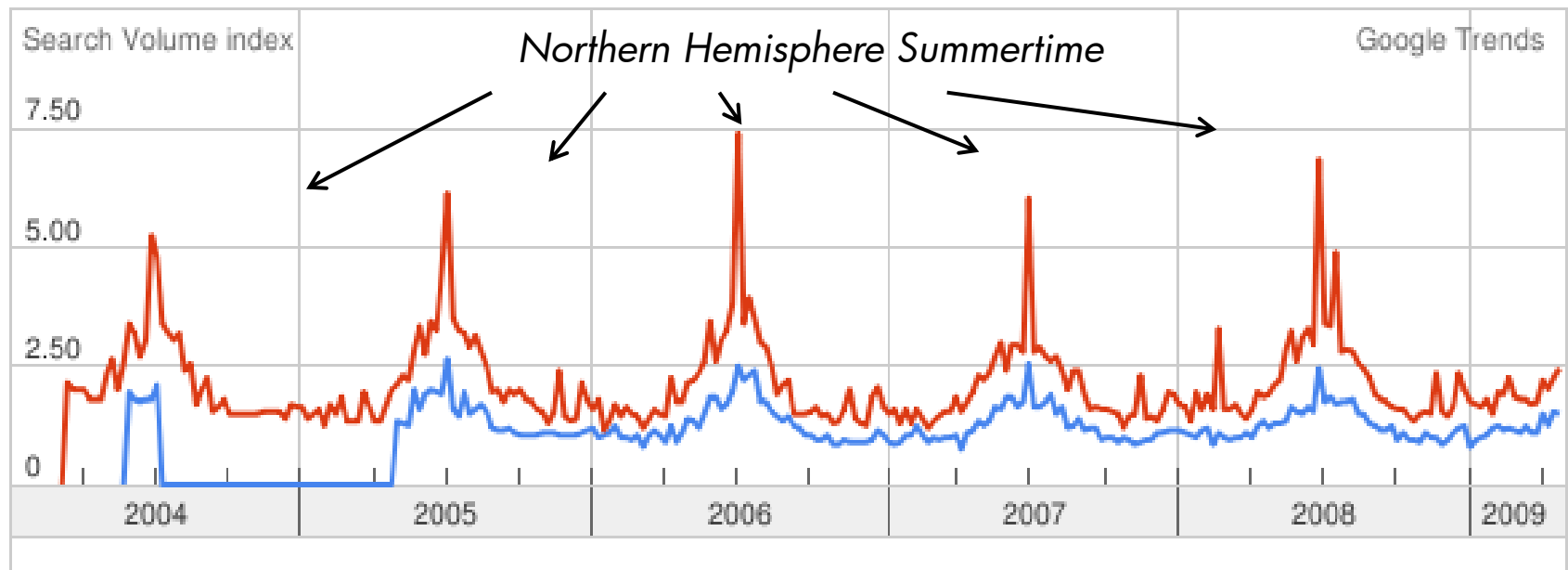
Search Trends

Tip: Use commas to compare multiple search terms.

Searches Websites

Scale is based on the average worldwide traffic of "chocolate ice cream... in all years. [Learn more](#)

"chocolate ice cream..." 1.00 "vanilla ice cream" 2.05





xeon, opteron, itanium, sparc

Search Trends

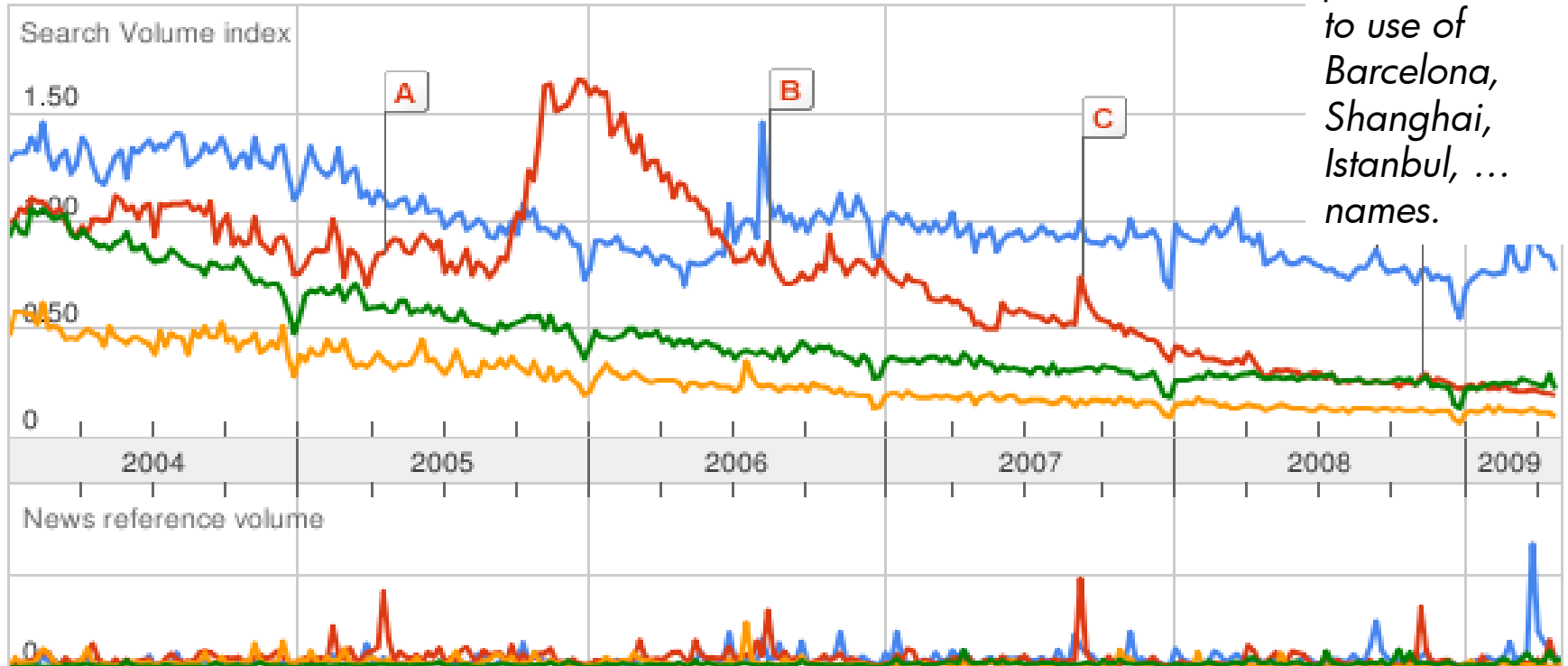
Tip: Use commas to compare multiple search terms.

Searches [Websites](#)

Scale is based on the average worldwide traffic of **xeon** in all years. [Learn more](#)

Opteron scores fall off, probable due to use of Barcelona, Shanghai, Istanbul, ... names.

xeon 1.00 **opteron** 0.74 **itanium** 0.28 **sparc** 0.00





"cloud computing", "grid computing", "distributed computing"

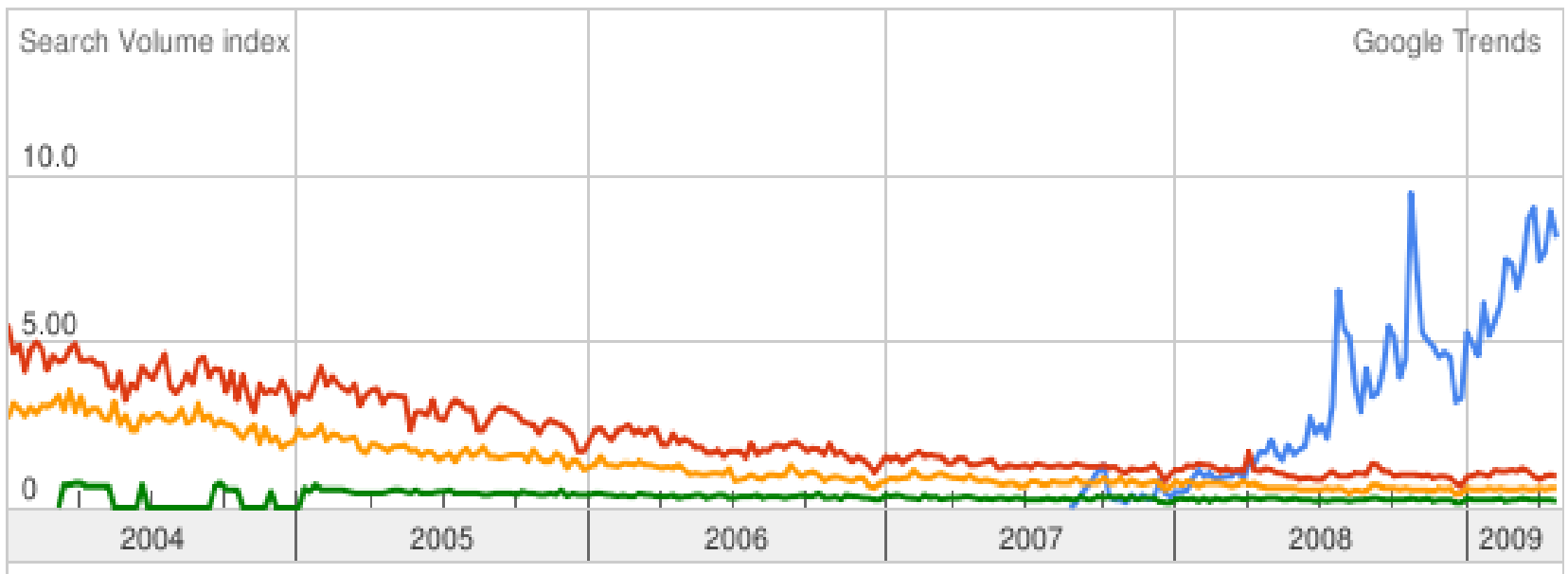
Search Trends

Tip: Use commas to compare multiple search terms.

Searches Websites

Scale is based on the average worldwide traffic of "cloud computing" in all years. [Learn more](#)

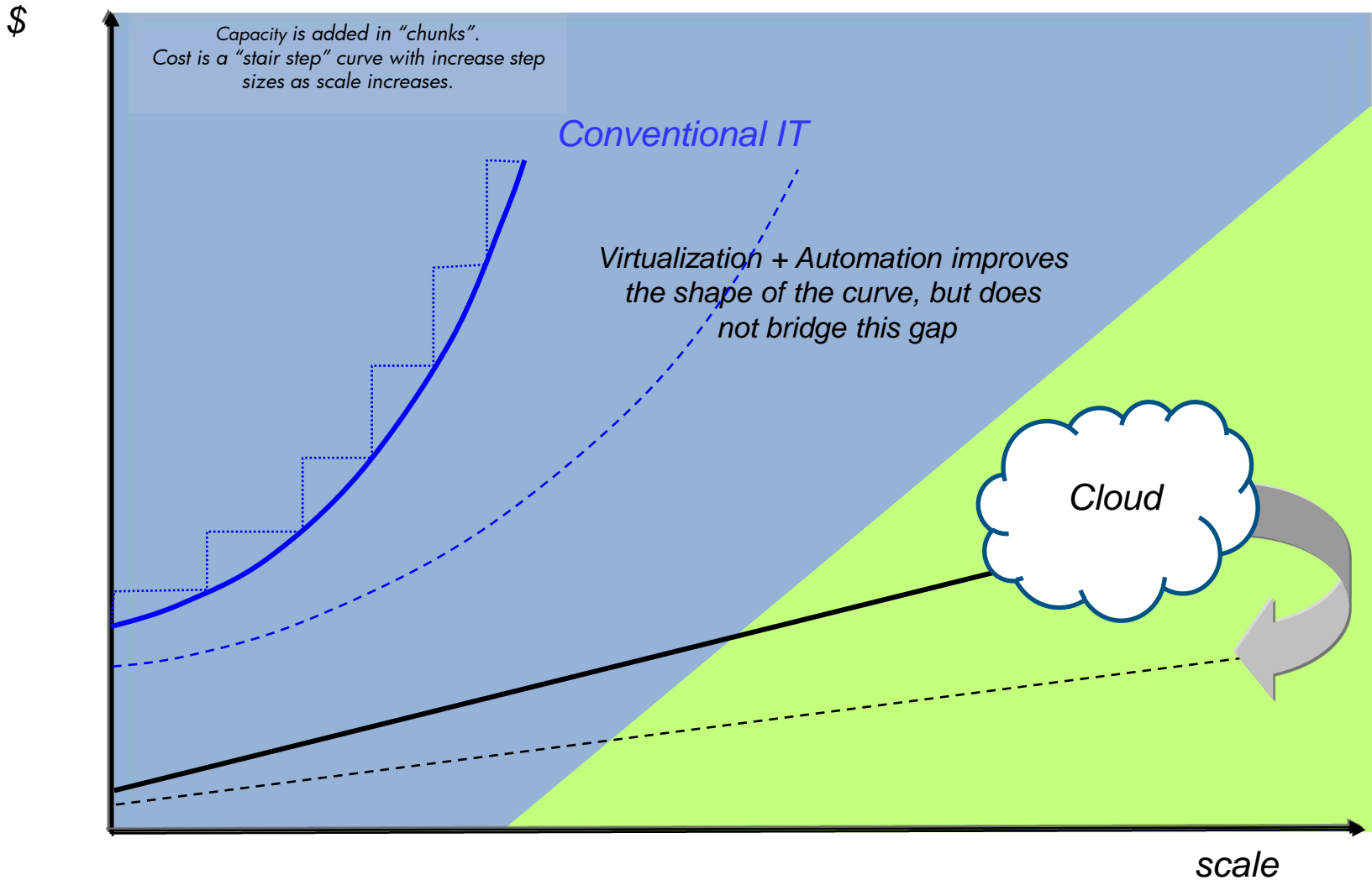
"cloud computing" — 1.00 "grid computing" — 2.20
"distributed computi..." — 1.35 "high performance co..." — 0.30



Cloud Workloads

- Designed for massive scale
- Tolerate Infrastructure Failures
 - Racks may die, servers may even lie?!, but your web property is always alive
- Designed for Co-tenancy
 - The next machine over (physical or virtual!) may be occupied by your competitor

Achieving scale at a the right cost... the critical success factor for cloud service providers



Where Cloud Servers Fit In The Ecosystem

- Enterprise App & Database Servers
- Cloud Storage Nodes
- Cache Nodes
 - Higher memory + network b/w, either backed by disks or (essentially) diskless
- HPC Compute Nodes
- Cloud Compute Nodes, e.g. Search & Web Tiers
- Low-rent Hosters
 - e.g. Very modest computes, idle \rightarrow 0, flexible ram+hdd, anything to lower acquisition costs

HPC Nodes

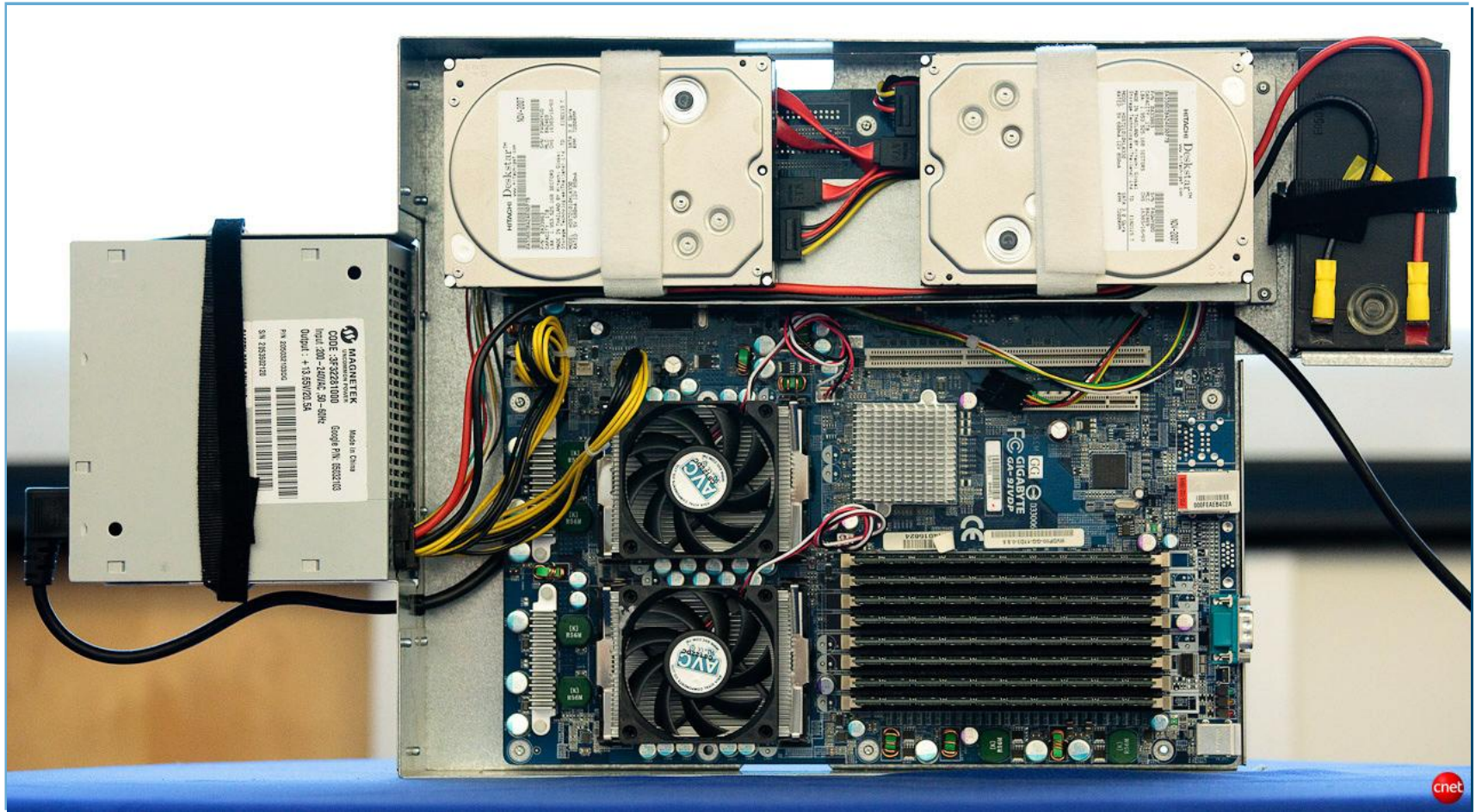
- ~50% IB-Connected
- Top500 dominated by blades
 - C-Class: integrated IB switch greatly simplifies cluster layout... and can reduce TCO
- Run hotter
 - Linpack == Power Virus
 - Rule of thumb: 250+ watts per 2s server



Cloud Nodes

- 100% 1GigE connected
 - Tin cans and a string would be fine for many users!
- Also dominated by 2s servers
 - 1U → Aggregated
- Run cool
 - Rule of thumb: 160 watts or less per 2s server

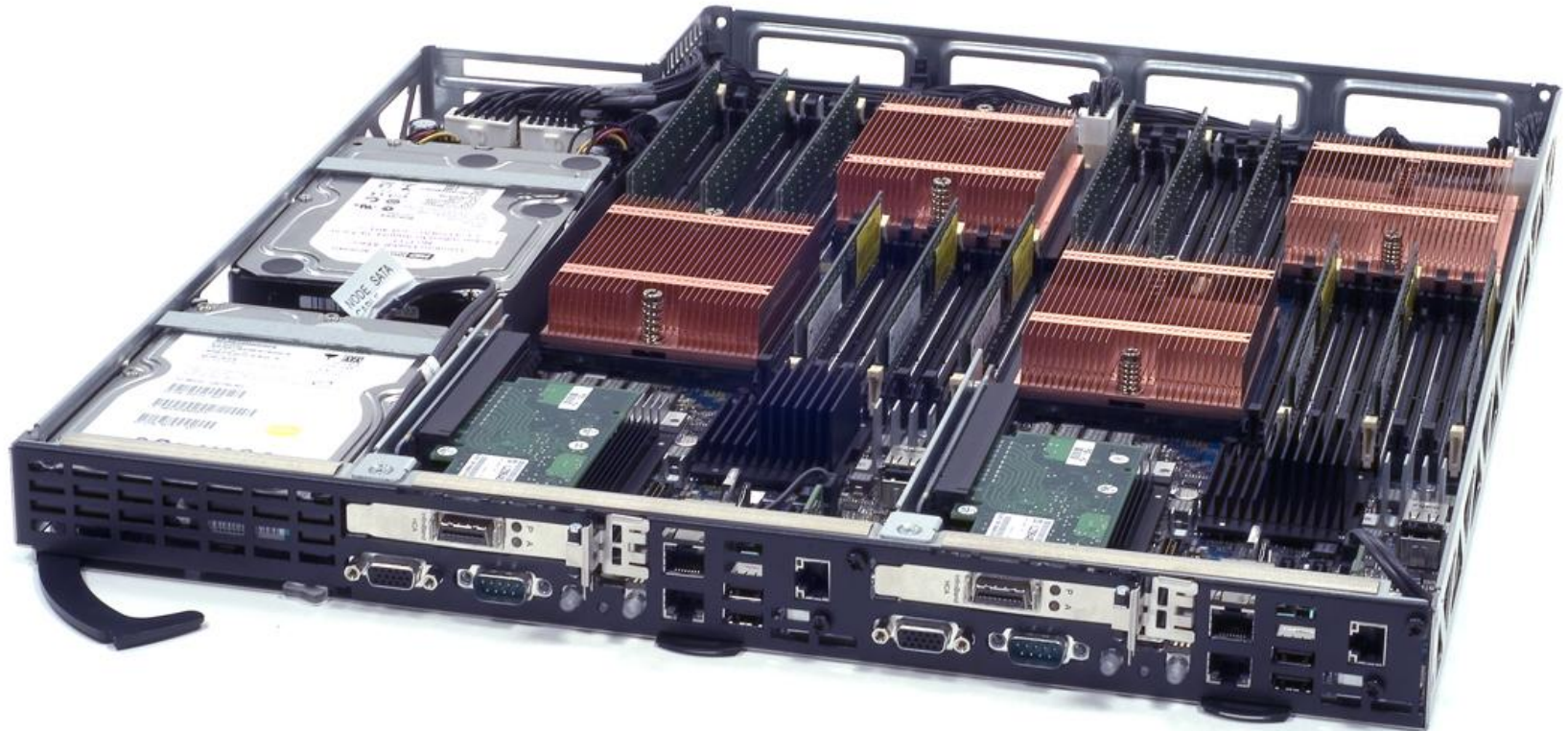
A Server Only A Mother Could Love...



http://news.cnet.com/8301-1001_3-10209580-92.html

2U, a bit out of date. Note: 12V Lead Acid battery & Velcro™!

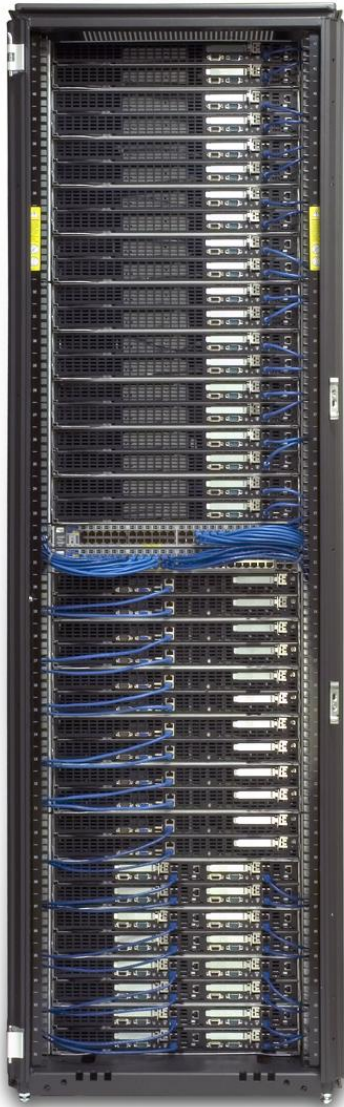
Two 2S servers (one LFF drive each)



Single 2s Server + 6 LFF Drives



Designed at a rack level.



Note front cabling.



Mgmt Network, Fans & Power from the back.



Horses for Courses

- Cloud applications want the cheapest node possible, and are willing to accept node failures *and even wrong answers* up to a point.
 - Trend: reliability implemented in software, with less need for reliable hardware
 - Example: cloud file services use replication instead of RAID
 - Map/Reduce (Hadoop), Bigtable (Cassandra), ...
 - Hard to do; Impractical for some applications; Only goes so far per unit of time...
- HPC Applications need higher node reliability, and wrong answers are just bad.
- Enterprise servers go where $O(\text{secs})$ of downtime mean $O(\text{US\$Ms})$ of lost transactions

Cloud vs. HPC Datacenters

- Massive scale
 - **Modest** max power/rack
 - Few Node Types
 - Power Utilization Efficiency (PUE) a major concern
 - **Modest** interconnect bandwidth/server
 - Willing to tradeoff node reliability for efficiency
 - Ugly, ugly, ugly
- Massive scale
 - **High** max power/rack
 - Few Node Types
 - Power Utilization Efficiency (PUE) a major concern
 - **Massive** interconnect bandwidth/server
 - Node reliability very important for capability computing
 - Typically Pretty

Pretty...



<http://www.sandia.gov/ASC/images/library/metropoliscenter.jpg>



http://www.sgi.com/global/de/images/lrz/sgi-hamTop3_scn.jpg

Not...



Microsoft's Vision of a Containerized Datacenter

Datacenter Futures

- Containers becoming mainstream
 - HPC will have trouble giving up pretty datacenters. It's hard to give a tour of a trailer park...
- Near-term Futures...
 - Scenario 1:
 - Cloud/Co-lo: 5-10KW rack → 10-14KW/rack
 - HPC: 50KW+ racks. Or maybe not!
 - Scenario 2a: Extremely dense containers. High power density, water-cooled. (PUE of 1.2ish).
 - Scenario 2b: Containers cooled with outside air: not very dense, even better PUE (approaching 1.1)

Interior view

Heat Exchangers

High efficiency, variable speed blowers

Separate Utility module segregates IT/UPS security access and environmentals

Standard 50U racks

Facilities management on exterior of cold aisle

Hot aisle with rear access through doors in the container

36" cold aisle



Trends: Things are getting a bit toasty...

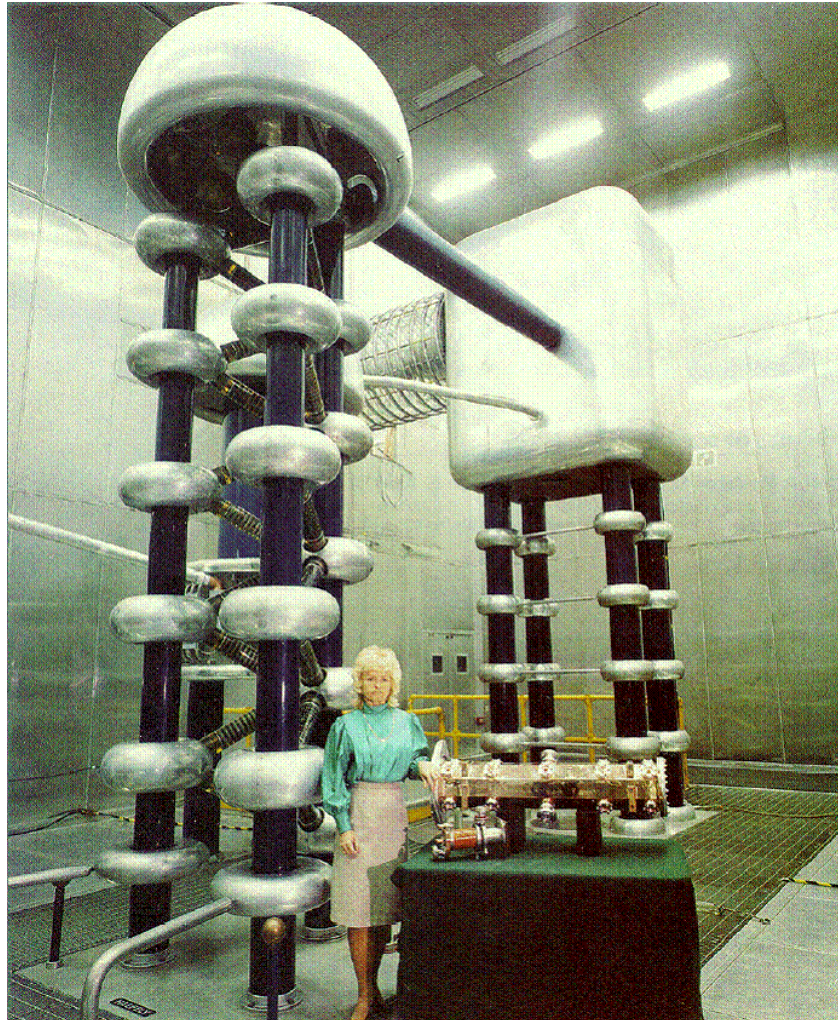


PC Cooling Strategy

Air

2006

2014 Desktop Heatsink?

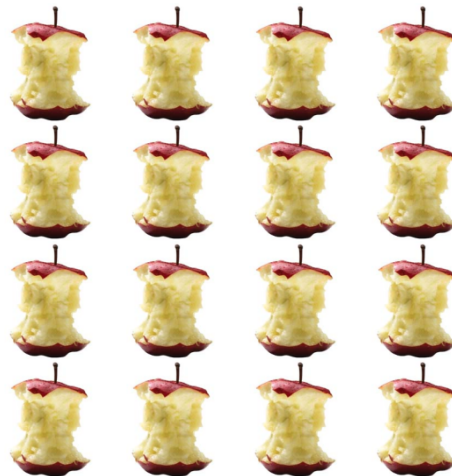


Actually a big Van der Graaf generator. Picture source: Unknown.

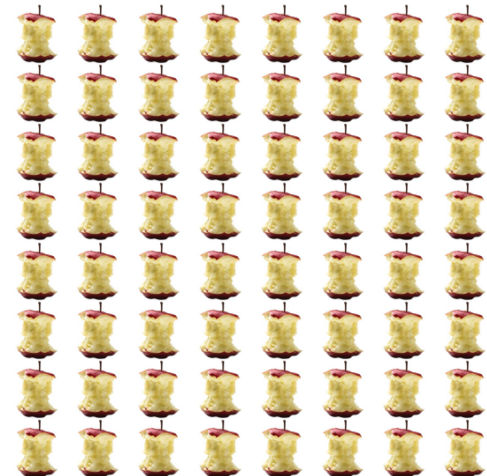
Dealing With The Problem, Part 1



2007



2011



2015

Dealing With The Problem, Part 2: Buy Better Stuff

- Efficient Power Supplies Cost More
 - >90% vs. ~65% eff.
 - For 400W load, saves \$516 (\$199 – \$715) over three years
 - (Details: \$0.10/kwh, \$10/W/10yr infrastructure)

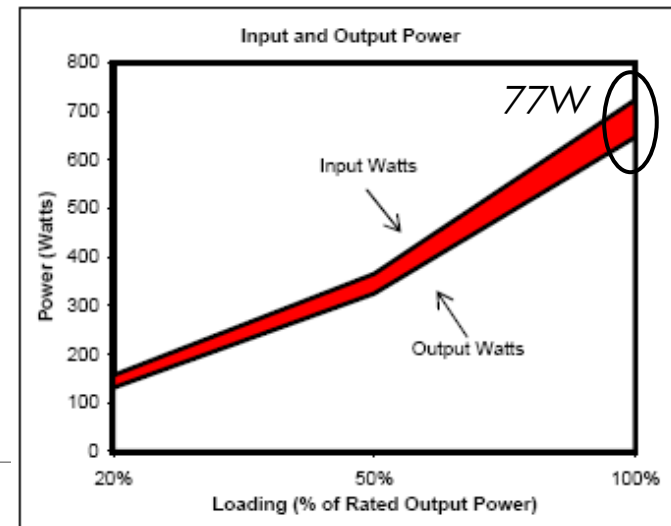
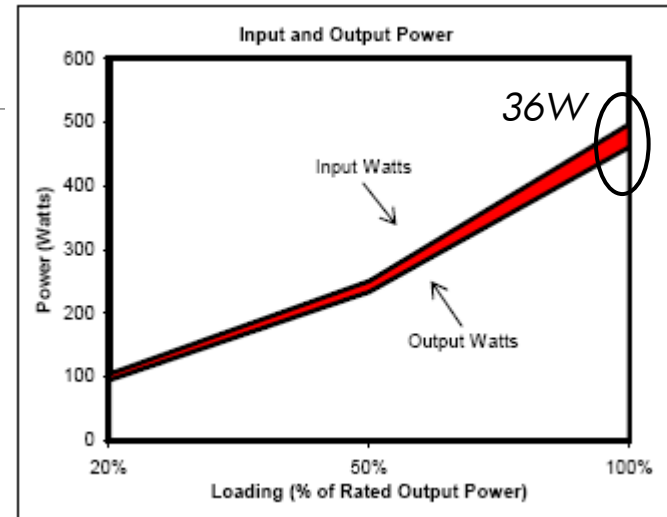
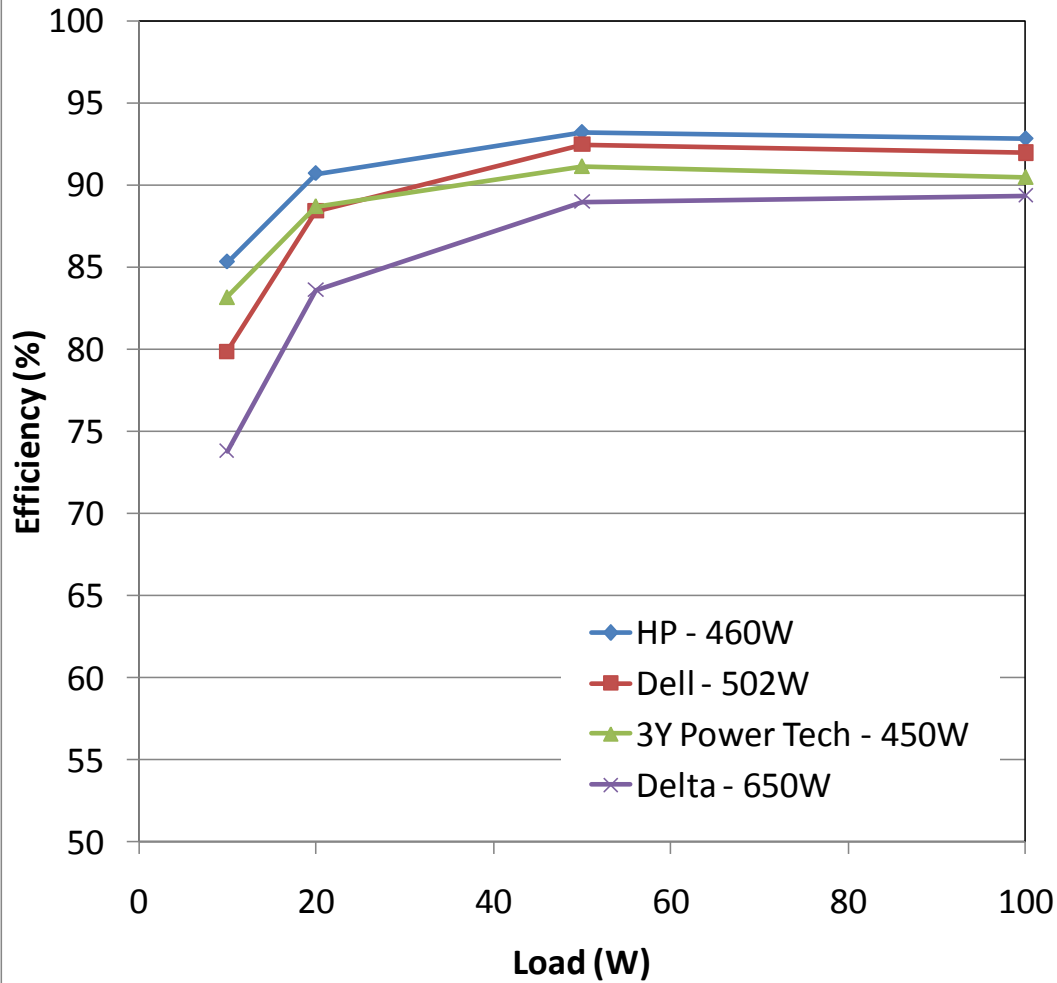
Power supply

650W Power Supply (Non-Hot Plug, Autoswitching) ▼

650W Power Supply (Non-Hot Plug, Autoswitching)

HP Power Supply 1200W with backplane [Add \$199.00]

Efficient Power Supplies



Dealing With The Problem, Part 3: Is it just money?

- When TCO doesn't justify lower power components...
 - What would your company do? (Heck, what would you do?)
- Personal conclusion: governmental policy will influence this calculation
 - Probably by higher energy/recycling taxes

Harder To Find Efficiencies

	5 yrs. ago	Today
PUE	2, 3, Higher	1.5 Good 1.2 Great
UPS Efficiency (Part of PUE)	94%	98%+
Power Supply Efficiency	75%	94%
Fan Power per 2s Node	60+ W	5-10 W ~1 W

Dealing With The Problem, Part 4: “Unobtainium”

- Optical links to memory, etc.
 - When will it not be “5 years away”?
- Stacked memory (outside the lab)
- NV memory replacing server storage, at all capacities, and even for write-heavy traffic

HPC ↔ Cloud?!

- HPC folks are becoming more cloud-like
 - ~~Just scratched out “grid computing” from their business cards and replaced it with “cloud computing”~~
 - Adopting new programming methodologies, e.g. map/reduce
 - More throughput jobs → less dependence on differentiated networks
- Cloud folks are getting ~~older~~ more HPC-like
 - (Some) looking to thicken up the interconnect bandwidth, both within racks and within a datacenter

Cloud Computing > Sliced Bread!

"cloud computing" — 1.00 "sliced bread" — 0.15

