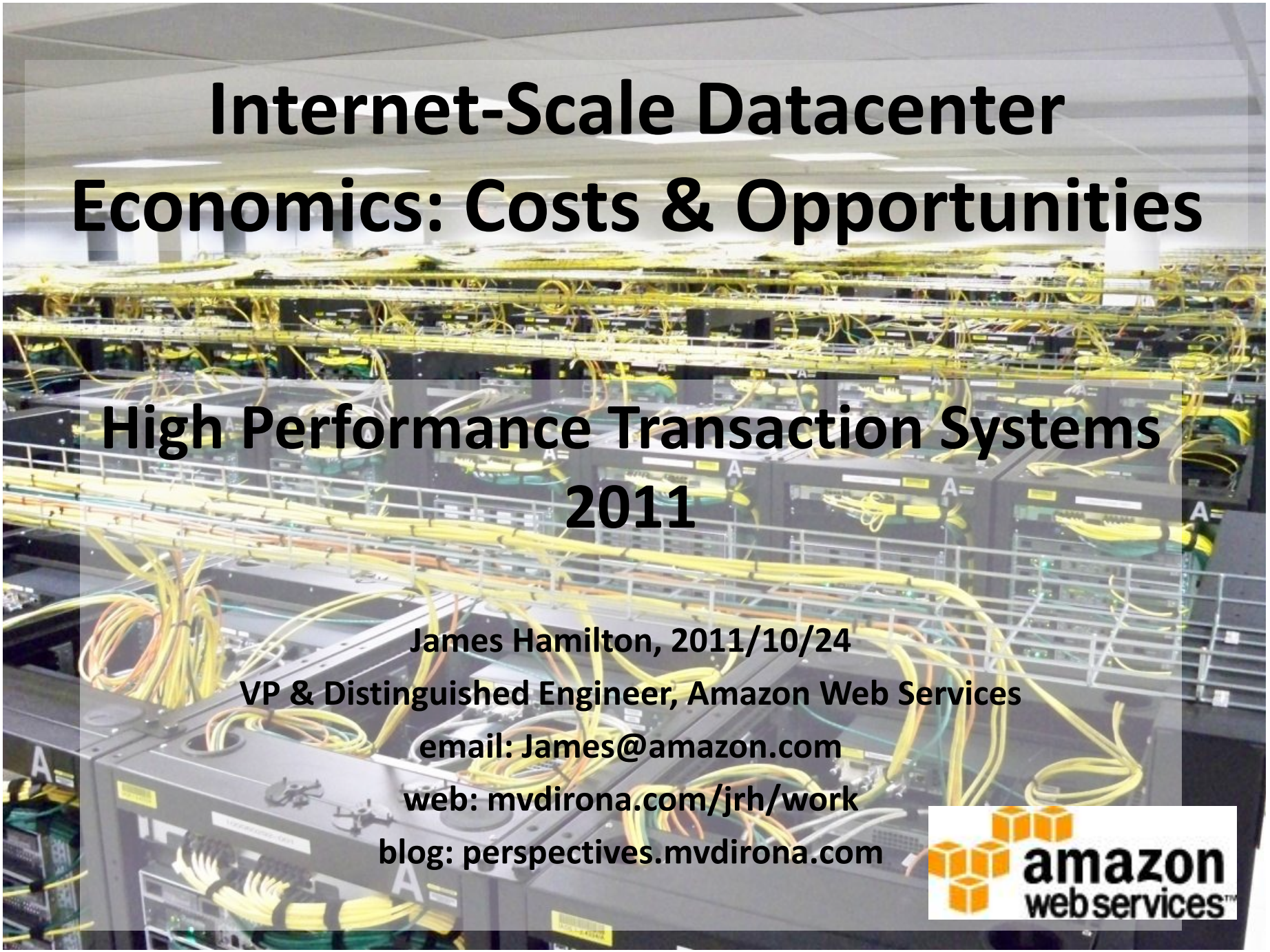


Internet-Scale Datacenter Economics: Costs & Opportunities

High Performance Transaction Systems 2011

James Hamilton, 2011/10/24

VP & Distinguished Engineer, Amazon Web Services

email: James@amazon.com

web: mvdirona.com/jrh/work

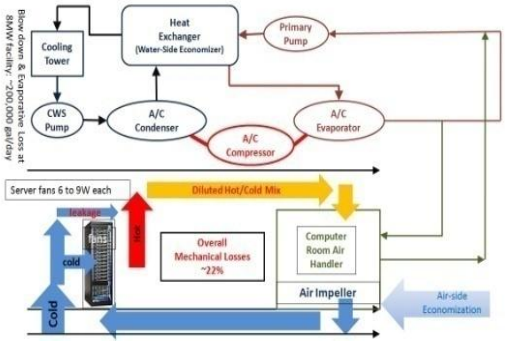
blog: perspectives.mvdirona.com

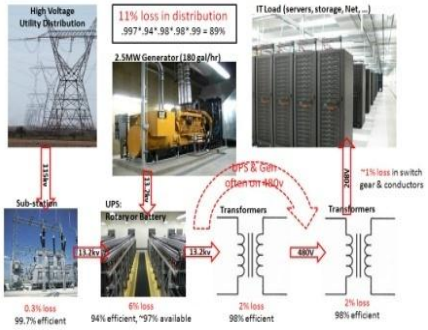
amazon web services™

Agenda

- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing

- Power Distribution
- Cooling & Building Designs
- Networking & Server Innovation





Talk does not necessarily represent positions of current or past employers

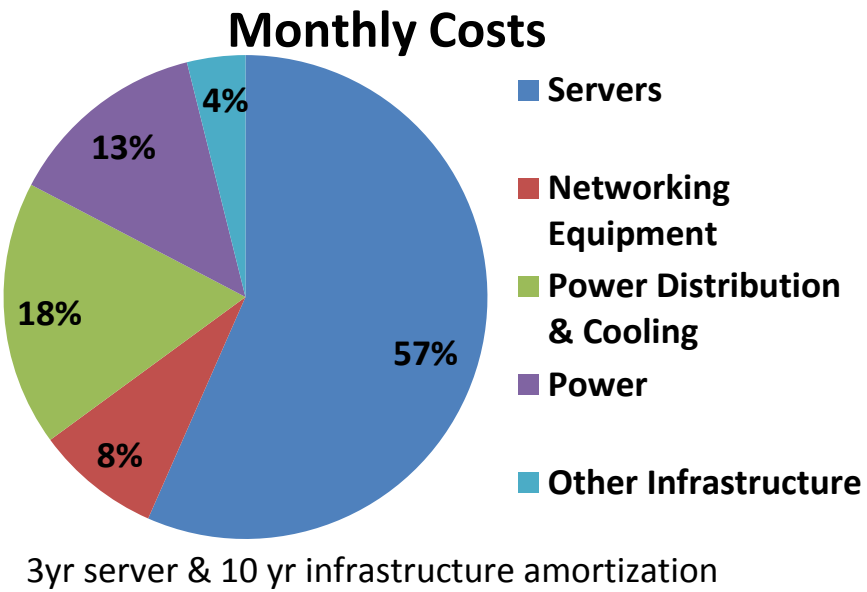
Pace of Innovation

- Datacenter pace of innovation increasing
  - More innovation in last 5 years than previous 15
  - Driven by cloud services & extraordinary-scale internet applications like search
  - Cost of infrastructure dominates service cost
  - Not just a cost center
- High focus on infrastructure innovation
  - Driving down cost
  - Increasing aggregate reliability
  - Reducing resource consumption footprint

Perspective on Scaling

Where Does the Money Go?

- **Assumptions:**
  - Facility: ~$88M for 8MW critical power
  - Servers: 46,000 @ $1.45k each
  - Commercial Power: ~$0.07/kWhr
  - Power Usage Effectiveness: 1.45



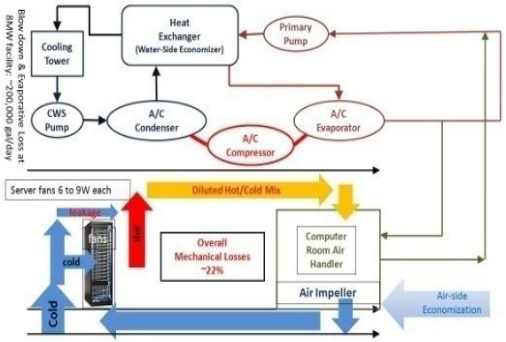
3yr server & 10 yr infrastructure amortization

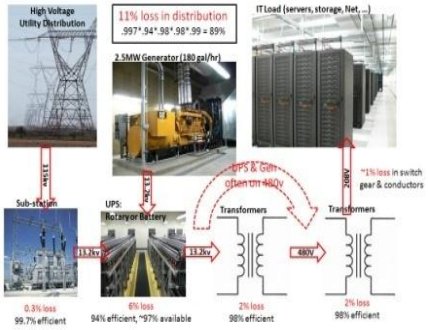
- **Observations:**
  - 31% costs functionally related to power (trending up while server costs down)
  - Networking high at 8% of overall costs & 19% of total server cost (often more)

From: http://perspectives.mvdirona.com/2010/09/18/OverallDataCenterCosts.aspx

Agenda

- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing
- Power Distribution
- Cooling & Building Designs
- Networking & Server Innovation





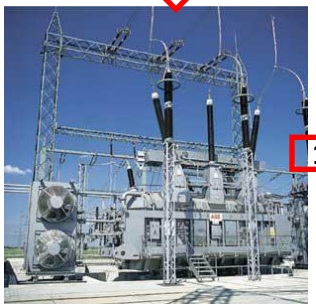
amazon web services™

Power Distribution

High Voltage Utility Distribution

115kv

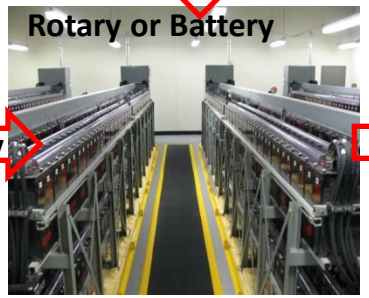
Sub-station



0.3% loss
99.7% efficient

13.2kv

UPS:
Rotary or Battery



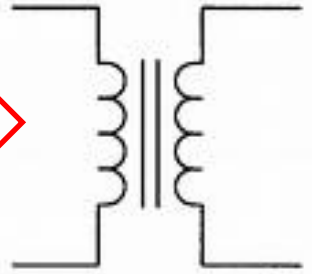
6% loss
94% efficient, ~97% available

Generators

13.2kv

13.2kv

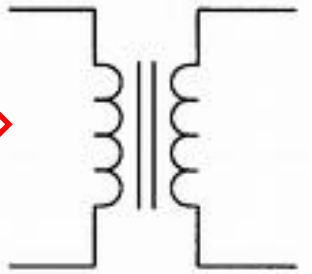
Transformers



2% loss
98% efficient

480V

Transformers



2% loss
98% efficient

208V

IT Load (servers, storage, Net, …)

UPS & Gen
often on 480V

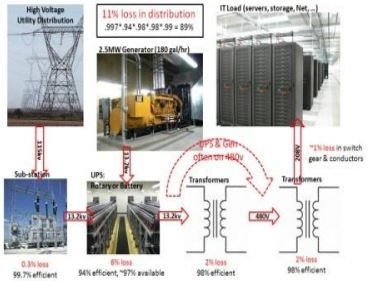
~1% loss in switch gear & conductors

~11% lost in distribution
.997*.94*.98*.98*.99 = 89%

Note: Two more levels of power conversion at server

Removing Power Conversions

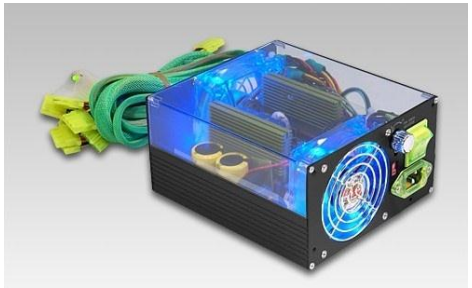
- Remove final conversion prior to server
  - 480VAC line-to-neutral yields 277VAC
  - 400VAC line-to-neutral yields 230VAC
  - Both good but later supports standard PSUs
- Another option is HVDC distribution
  - 400DC an interesting option
  - Improved efficiency but higher capital cost



Power Distribution Efficiency Summary

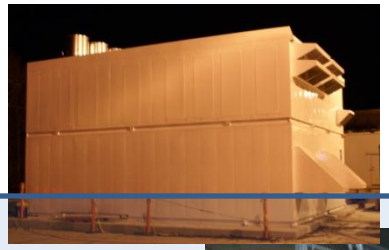
- 2 more power conversions at servers
  5. Power Supply: often under 80% at typical load
  6. On board voltage regulators (VRMs or VRDs)

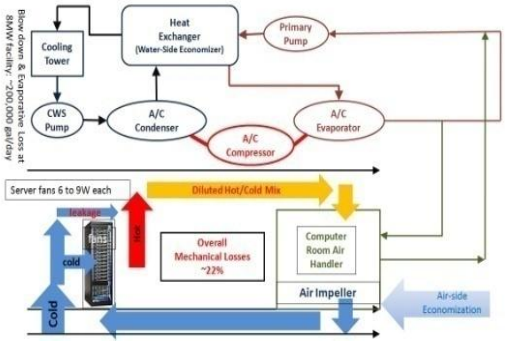
- Rules to minimize power distribution losses:
  - Oversell power (more load than provisioned power)
  - Avoid conversions (fewer & better)
  - Increase efficiency of conversions
  - High voltage as close to load as possible
  - Size voltage regulators to load & use efficient parts
  - High voltage direct current a small potential gain

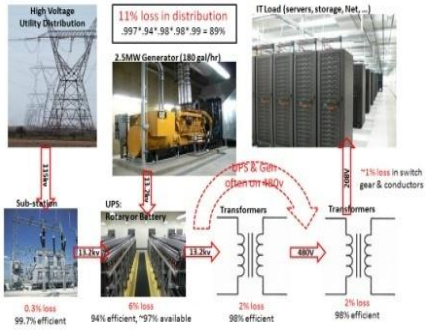


Agenda

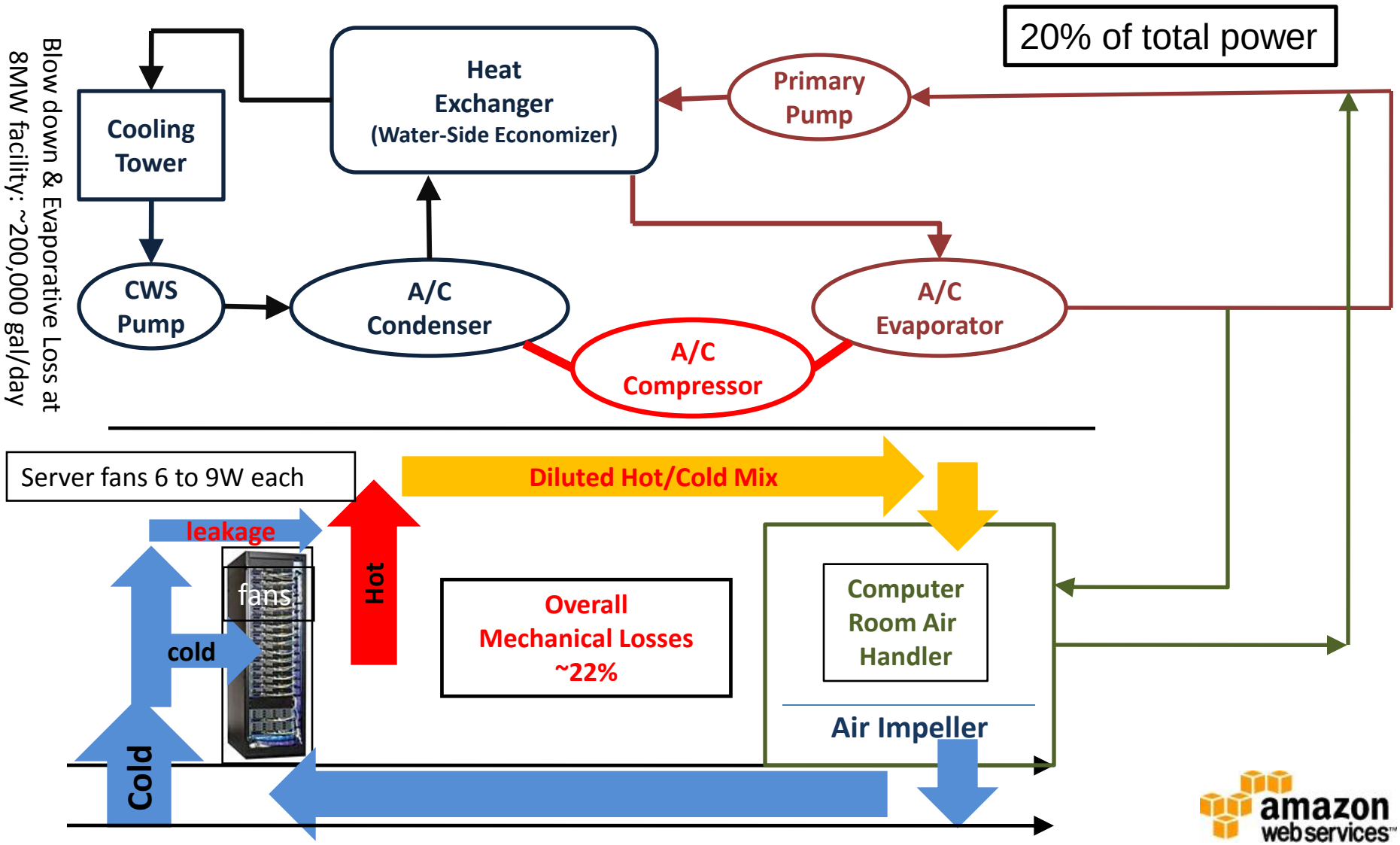
- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing
- Power Distribution
- Cooling & Building Designs
- Networking & Server Innovation







Mechanical Systems



Hot Aisle Containment

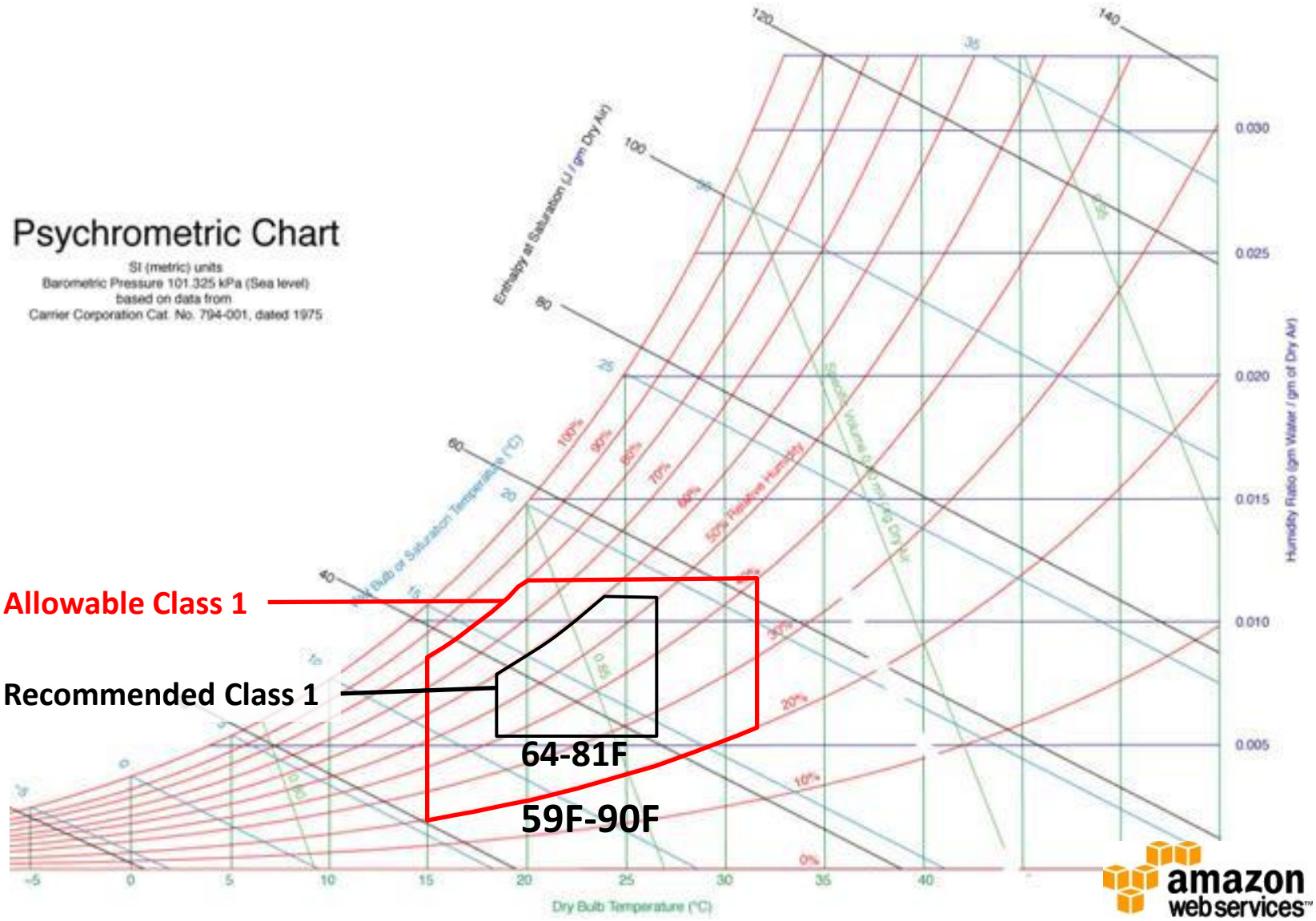
Facebook Open Compute

Intel

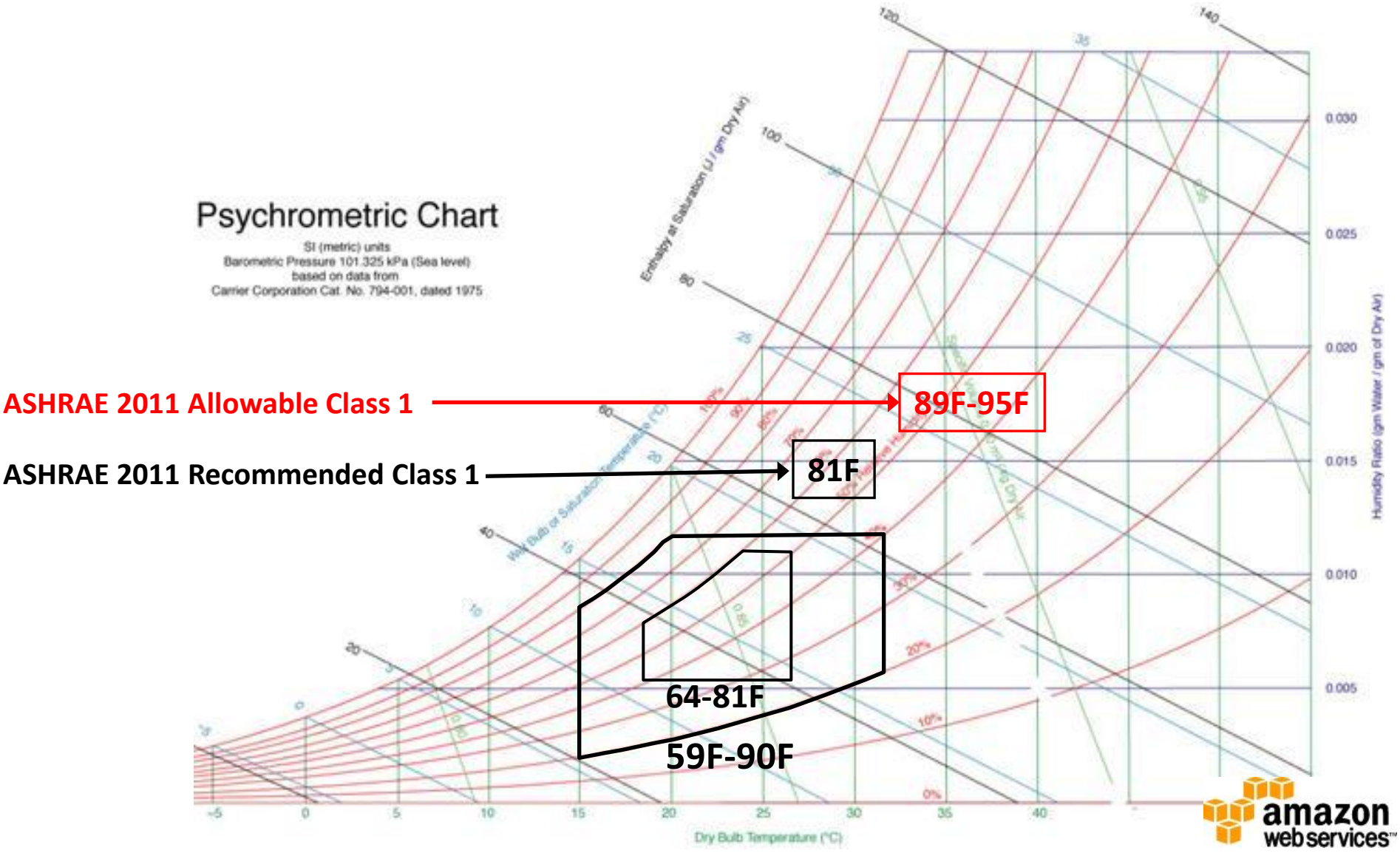
Intel

WriteLine

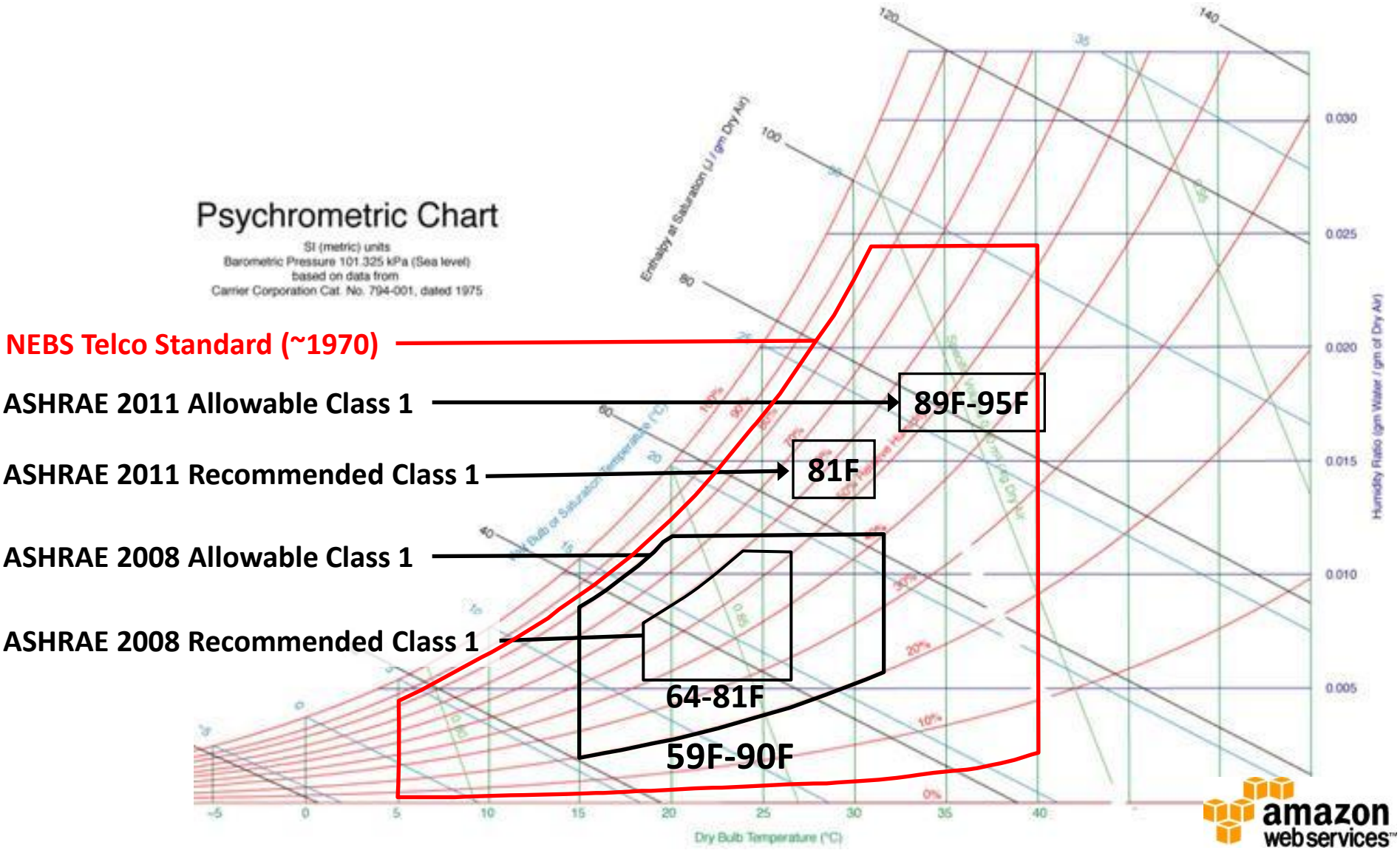
ASHRAE 2008 Recommendations



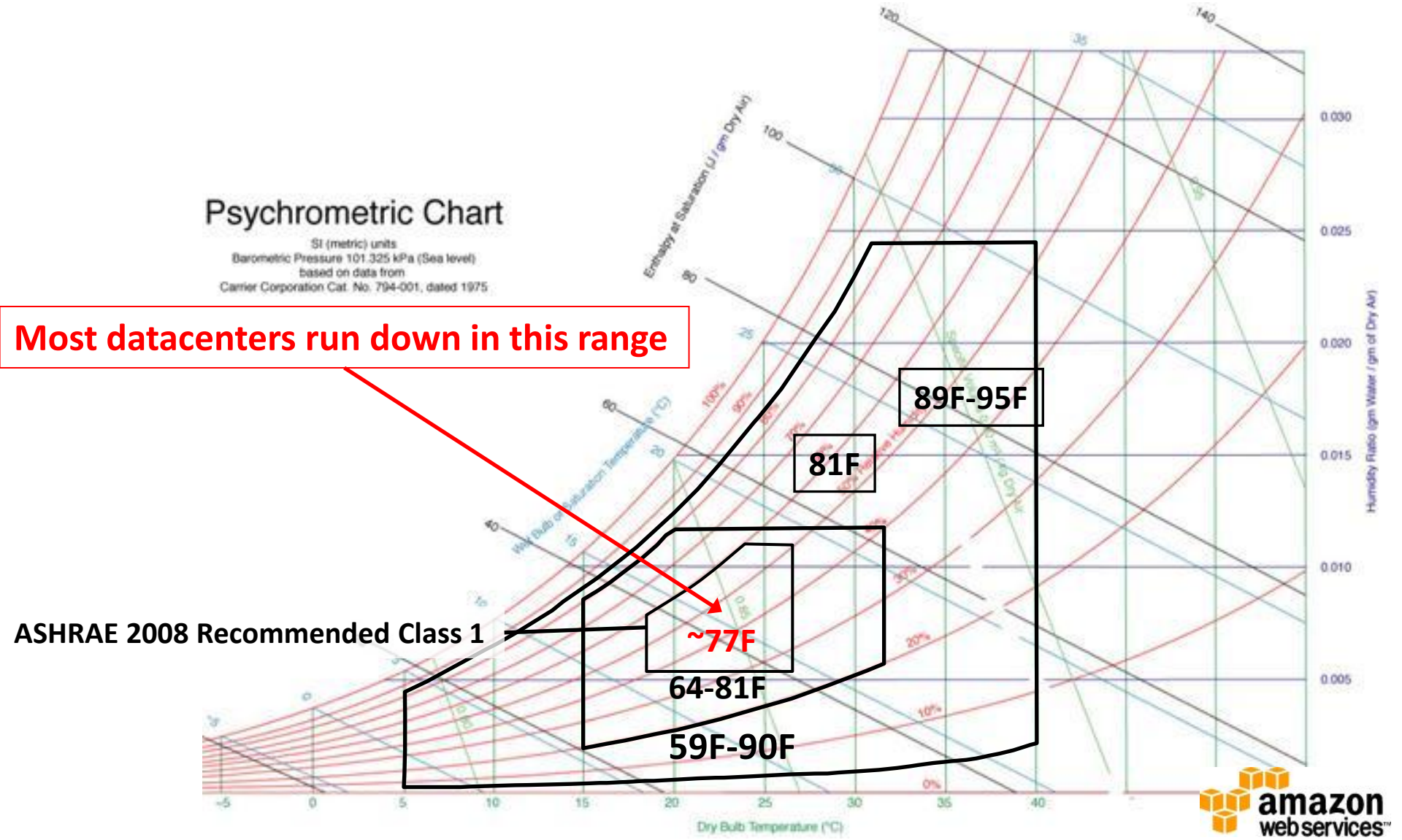
ASHRAE 2011 Recommendations



Network Equipment Building Standard

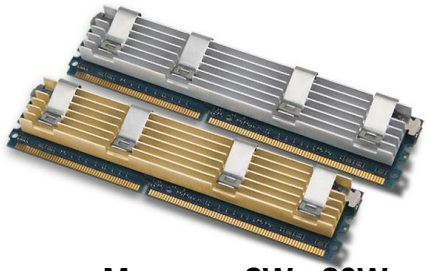


Most Datacenters Still Run Cold



Air Cooling

- Component temps specs higher than historically hottest place on earth
  - Al Aziziyah, Libya: 136F/58C (1922)
- Just a mechanical engineering problem
  - More air or better mechanical designs
- Tradeoff: semi-conductor leakage & power to move more air vs cooling savings
- Currently available equipment temp limits:
  - 40C/104F: CloudRack C2 & most net gear
  - 35C/95F: Most of the server industry



Memory: 3W - 20W
Temp Spec: 85C-105C

Hard Drives: 7W- 25W
Temp Spec: 50C-60C

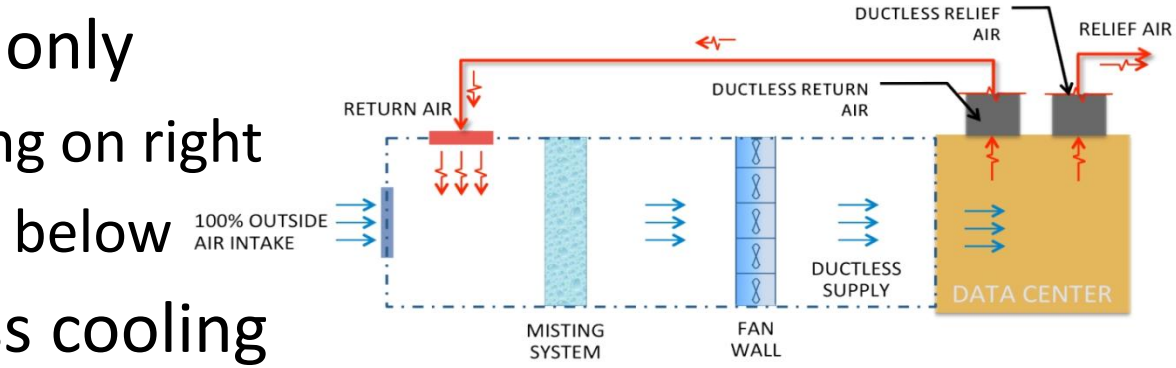
I/O: 5W - 25W
Temp Spec: 50C-60C

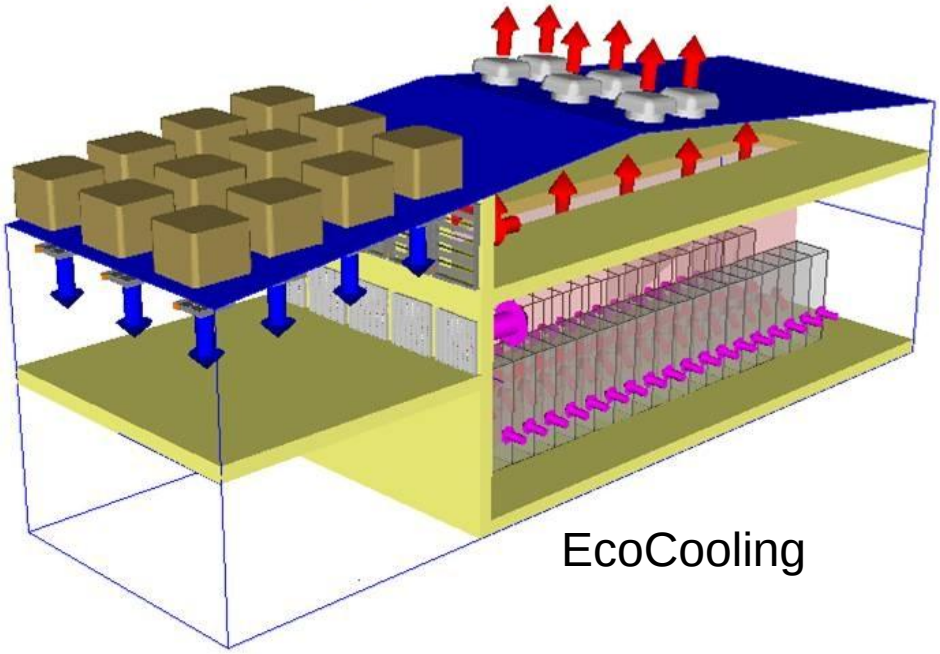
Processors/Chipset: 40W - 200W
Temp Spec: 60C-70C

Thanks to Ty Schmitt& Giovanni Coglitore

Innovative Shell Designs

- Evaporative cooling only
  - High pressure misting on right
  - Damp media design below
- Full building ductless cooling





EcoCooling

Facebook Prineville above & below

amazon web services™

Modular and Pre-fab DC Designs

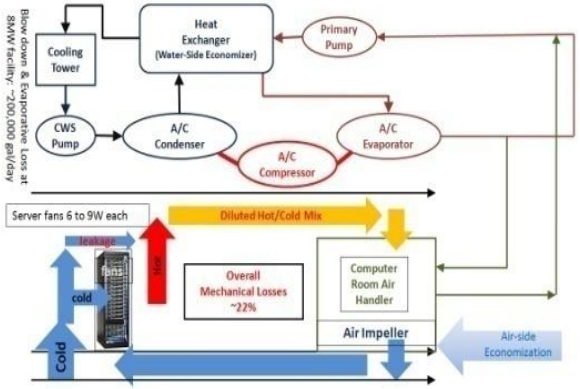
Microsoft ITPAC

Amazon Perdix

- Fast & economic deployments
- Sub-1.2 PUE designs
- Air-side economized
  - In some cases no mechanical cooling
- ISO standard shipping containers offered by Dell, HP, SGI, IBM, …

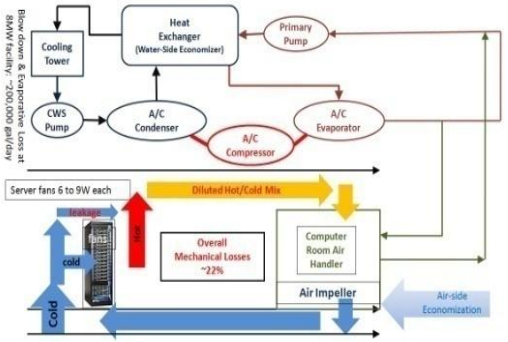
Cooling Looking Forward

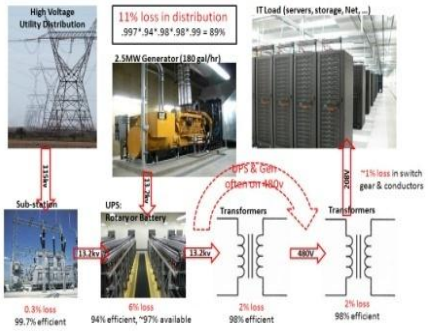
- River water or lake water cooling
  - Google Belgium & Finland, Deepgreen Switzerland
  - Not new: Toronto metro area cooling
- Water direct to the rack
  - IBM iDataPlex
- Water direct to components
  - Done before 80's IBM S/370 308x & 3090 series
  - Again when heat densities climb back to that level
- Direct on component spray cooling



Agenda

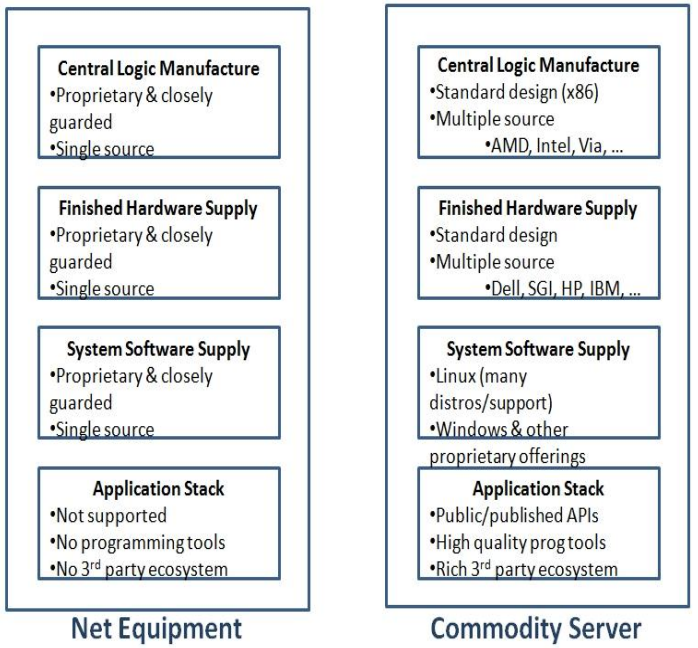
- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing
- Power Distribution
- Cooling & Building Designs
- Networking & Server Innovation

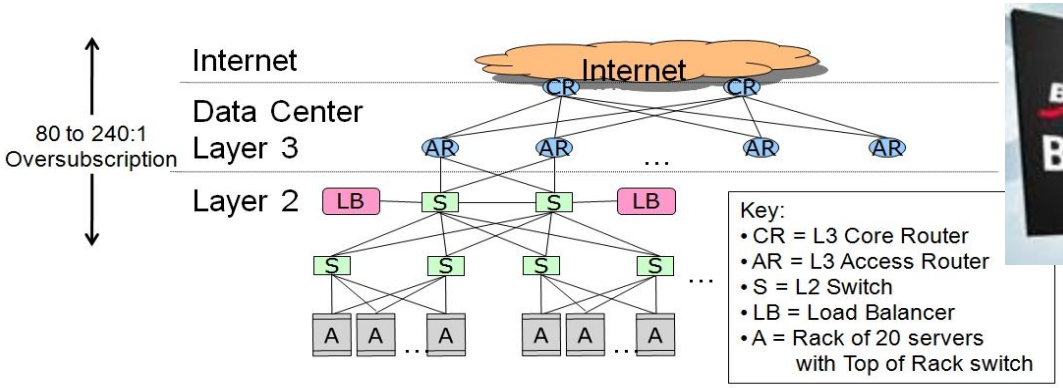




Sea Change in Networking

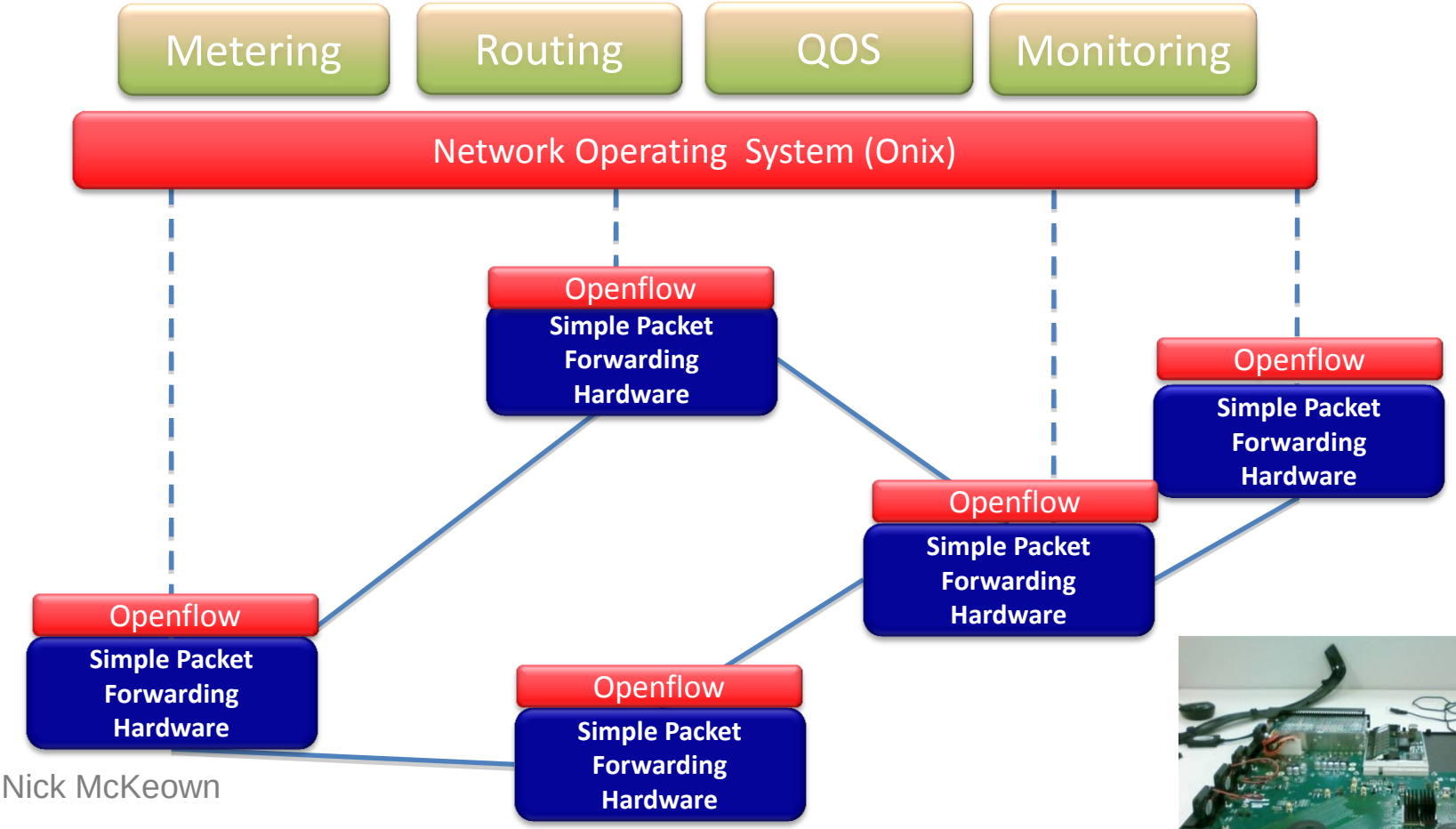
- Current networks over-subscribed
  - Forces workload placement restrictions
  - Goal: all points in datacenter equidistant
- Mainframe model goes commodity
  - Competition at each layer over vertical integ.
- Get onto networking on Moore's Law path
  - ASIC port count growth at near constant cost
  - Competition: Broadcom, Marvell, Fulcrum,…





amazon web services

Software Defined Networking



- Distributed control plane with central control
  - Research examples: VL2, PortLand, & others
  - Onix/OpenFlow gaining industry support & traction quickly

Server Innovation

- Removing all unnecessary cost & power
  - Omit lid, Depop board, efficient components, 12V-only PSUs
- Form factor: fractional RU & multi-server modules
- Shared power supplies
  - N supplies for M servers
  - Run supplies at most efficient load
- Shared large back-of-rack fans
- Cell phone technology predicts future server generations
- Super high-density storage platforms
  - Increasing server to disk ratio for cold storage
- Soon: ARM architecture, low power servers, multiple servers on board, proprietary net fabric