QPS,KW-hr,MTBF,ΔT,PUE,IOPS,DB/RH: *A Day in a Life of a Datacenter Architect*

Kushagra Vaid

Principal Architect, Datacenter Infrastructure

Microsoft Online Services Division

kvaid@microsoft.com

First… lets explain the lingo in the talk title…

Metric	Definition	Usage
QPS	Queries per second	Performance of Index-serving engine
KW-hr	Kilowatt-hr: Energy consumption metric	Effectiveness of power conservation schemes
MTBF	Mean Time Between Failures	Understanding component reliability
ΔT	Temperature delta between front and rear of server	Server chassis design and airflow CFD analysis
PUE	Power Usage Effectiveness	Measure of datacenter power efficiency
IOPS	Disk Input/Output Operations per second	Storage subsystem performance analysis
DB/RH	Dry Bulb: Air temperature Wet Bulb: Temperature indicating amount of moisture in the air	Determining server operating range and acceptable environmental specs

Microsoft Datacenters: Providing services 24x7

Windows Live ID

More than 1 billion authentications per day

bing

More than 2 billion queries per month

Microsoft Exchange Hosted Services

Processes 2–4 billion e-mails per day

Windows Live Messenger

320 million active accounts

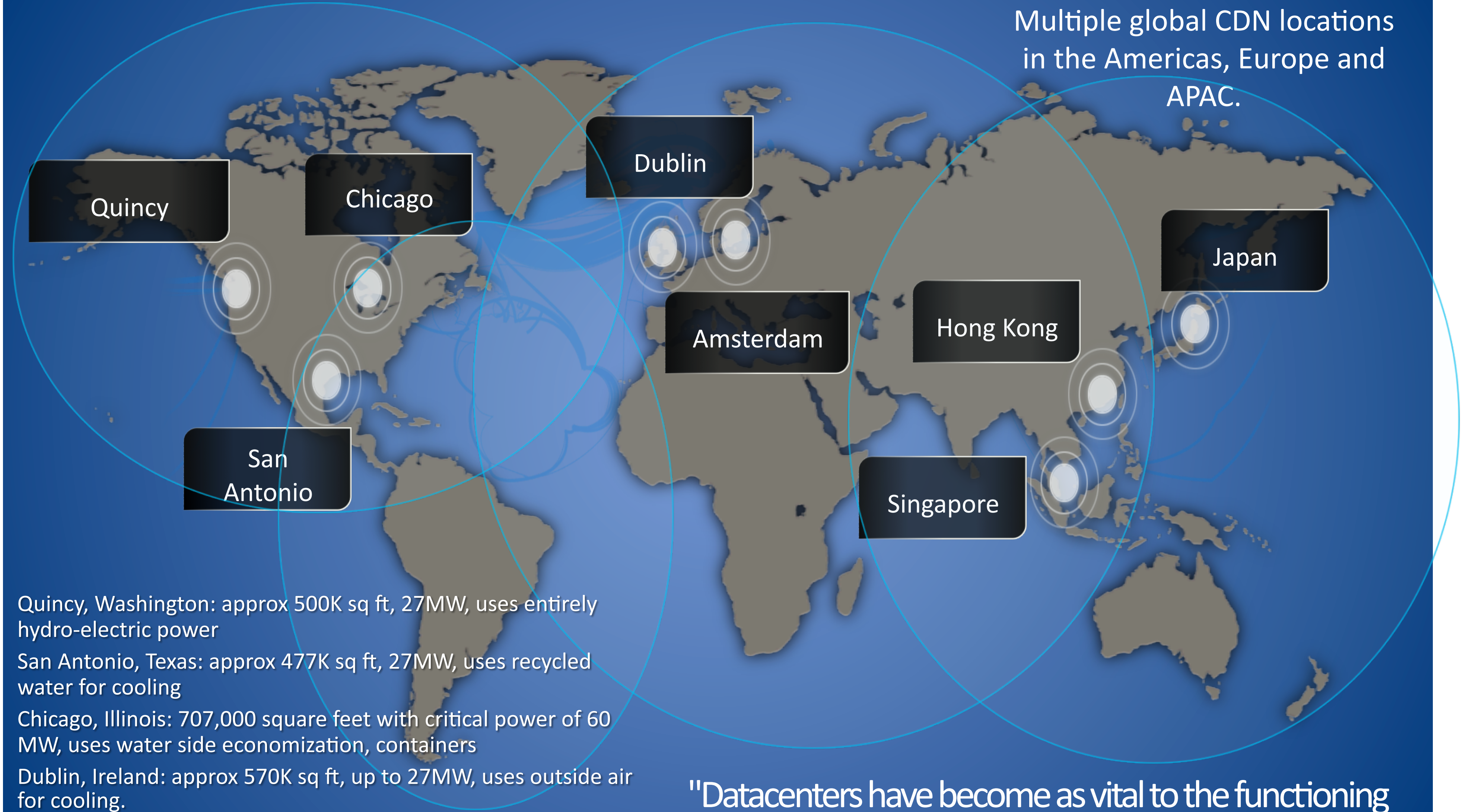
msn

550 million unique visitors monthly

Windows Live Hotmail

400 million active accounts

Microsoft Datacenters – Global scaling!



Multiple global CDN locations in the Americas, Europe and APAC.

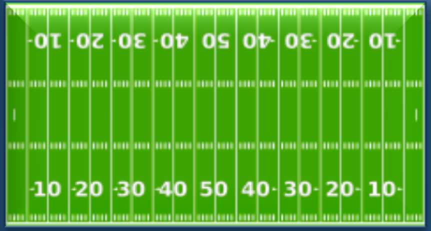
Quincy, Washington: approx 500K sq ft, 27MW, uses entirely hydro-electric power

San Antonio, Texas: approx 477K sq ft, 27MW, uses recycled water for cooling

Chicago, Illinois: 707,000 square feet with critical power of 60 MW, uses water side economization, containers

Dublin, Ireland: approx 570K sq ft, up to 27MW, uses outside air for cooling.

"Datacenters have become as vital to the functioning of society as power stations." – *The Economist*



A large mega-datacenter is

11 times

the size of a football field

Design factors for Datacenter Infrastructure

Datacenter and Server architecture

- Power distribution efficiency
- Cost efficiency
- Thermal design analysis

Platform architecture

- Application performance analysis
- New platform architecture exploration

Reliability analysis

- Environmental operating ranges and impact on MTBF

Bringing it all together – Holistic Systems Design

Design factors for Datacenter Infrastructure

Datacenter architecture

- **Power distribution efficiency**
- Cost efficiency
- Thermal design analysis

Platform architecture

- Application performance analysis
- New platform architecture exploration

Reliability analysis

- Environmental operating ranges and impact on MTBF

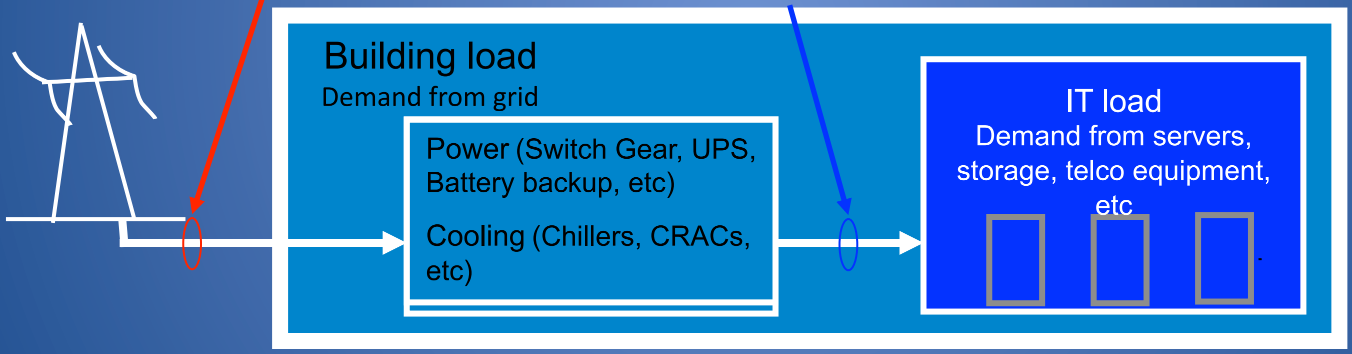
Bringing it all together – Holistic Systems Design

Cost and efficiency facts

For a typical large mega-datacenter…

Building costs are between $10M to $15M per MegaWatt

PUE = Total Facility Power / IT Equipment Power



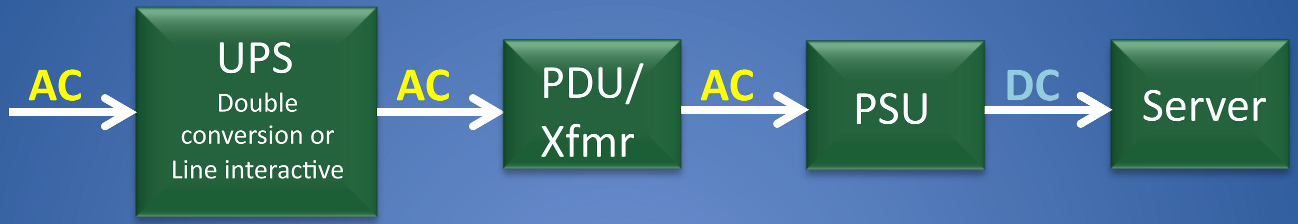
More about PUE: http://thegreengrid.org/en/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE

Typical industry PUE ranges from 1.5-2.0

Energy Consumption: US power rate 10.27 cents per Kilowatt hour) according to DOE/eia (http://www.eia.doe.gov/cneaf/electricity/epm/table5_3.html)

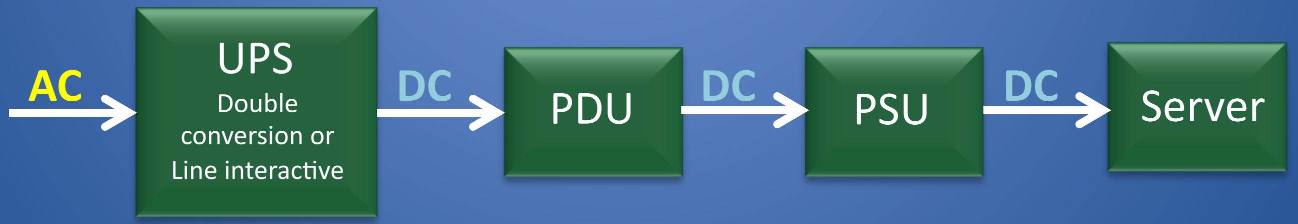
Facility level distribution options

Consider the following common topologies …



AC → UPS Double conversion or Line interactive → AC → PDU/ Xfmr → AC → PSU → DC → Server

480Vac	480Vac	208Vac	12Vdc
480/277Vac	415/240Vac	240Vac	12Vdc
480/277Vac	480/277Vac	277Vac	12Vdc
600Vac	600Vac	208Vac	12Vdc

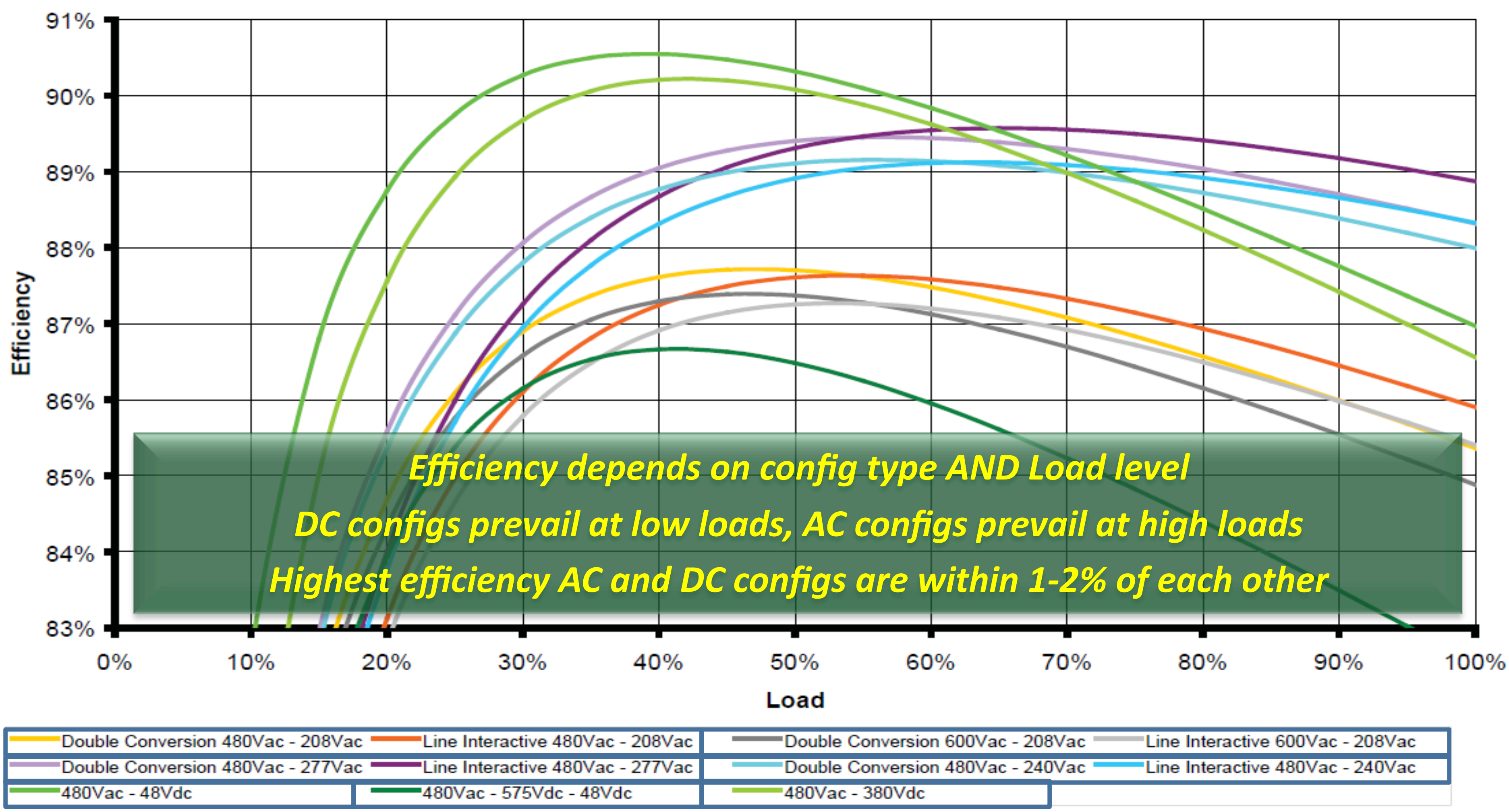


AC → UPS Double conversion or Line interactive → DC → PDU → DC → PSU → DC → Server

480Vac	575Vdc	48Vdc	12Vdc
480Vac	380Vdc	380Vdc	12Vdc
480/277Vac	48Vdc	48Vdc	12Vdc

Analyzing Facility level distribution efficiency

Which topology is the most efficient (lowest PUE)?



Efficiency depends on config type AND Load level

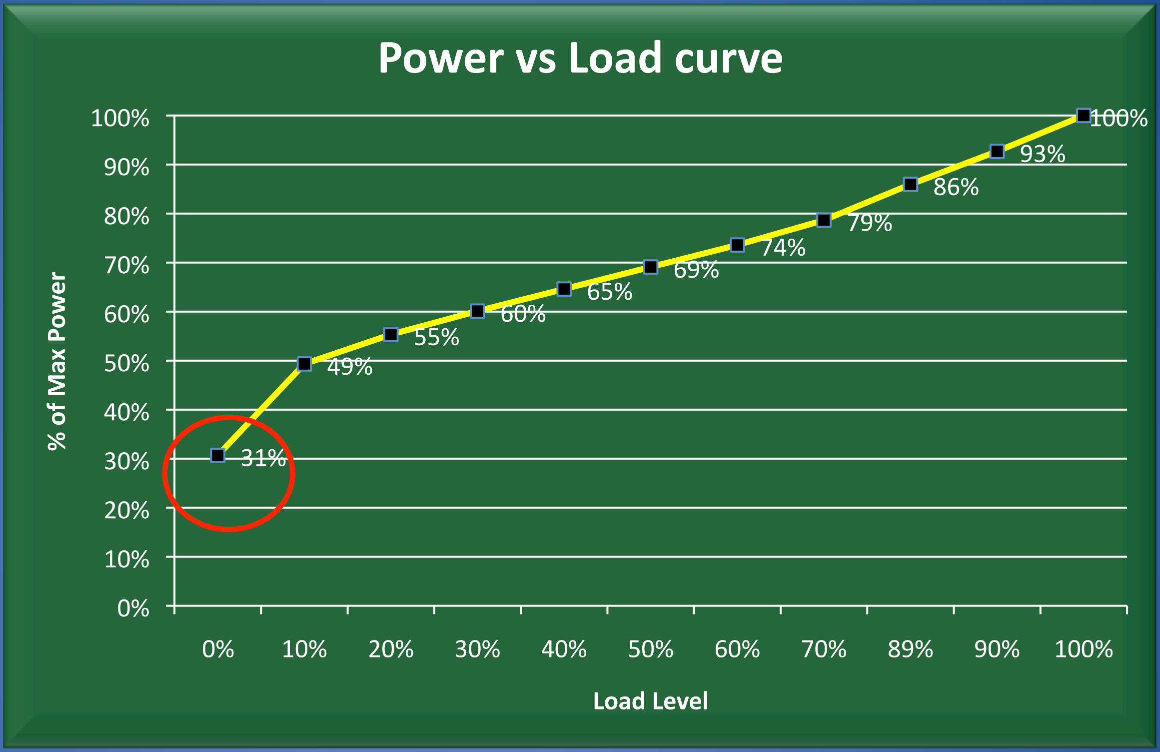
DC configs prevail at low loads, AC configs prevail at high loads

Highest efficiency AC and DC configs are within 1-2% of each other

Source: Green Grid (http://www.thegreengrid.org/~/media/WhitePapers/White_Paper_16_-_Quantitative_Efficiency_Analysis_30DEC08.ashx)

Challenges with datacenter power allocation

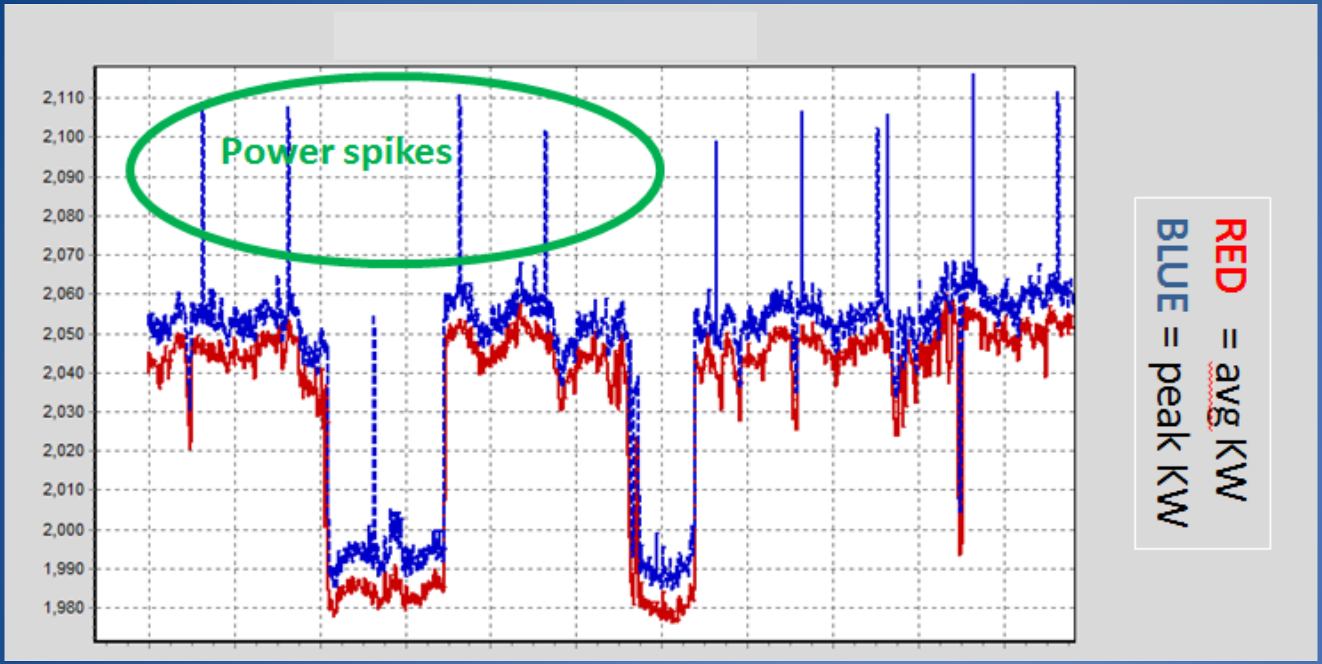
SPECpower2008 benchmark
2 x Intel L5640/6c/2.26Ghz
16GB DDR3
1x160GB SSD
Windows Server 2008 R2



http://www.spec.org/power_ssj2008/results/res2010q3/power_ssj2008-20100714-00275.html

- Problem statement: Given the above power load curve, how would you maximize server density for the datacenter power envelope
- Tradeoffs:
  - Provision based on peak load →stranded power from variable load traffic patterns
  - Provision based on average load → risk of tripping circuit breakers at higher load levels

Challenges with datacenter power allocation



Output of datacenter power meter

- Short duration power spikes from sudden load increases occur in reality
- Server heterogeneity and app power profile changes also need to be considered
- Power Capping can be used, but requires understanding tradeoff to performance SLAs and may require sophisticated policy management
- The eventual allocation is a calculated risk to maximize server density without stranding power, taking load patterns into consideration

Design factors for Datacenter Infrastructure

Datacenter architecture
- Power distribution efficiency
- **Cost efficiency**
- Thermal design analysis

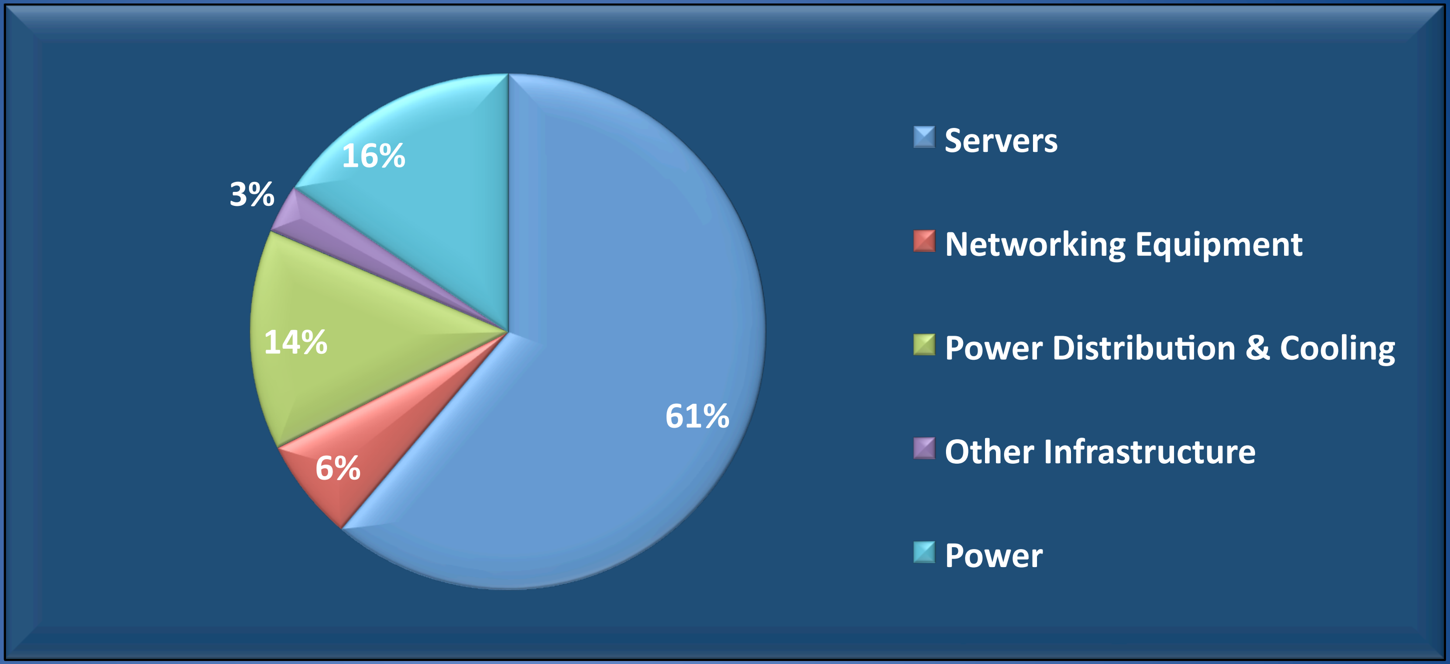
Platform architecture
- Application performance analysis
- New platform architecture exploration

Reliability analysis
- Environmental operating ranges and impact on MTBF

Bringing it all together – Holistic Systems Design

Datacenter TCO breakdown

Assumptions:
10MW facility
PUE 1.25
$10/W construction costs
$0.10c/KWhr power costs
Server: $2000, 200W
3yr server amortization
15yr datacenter amortization



Source: James Hamilton (http://mvdirona.com/jrh/TalksAndPapers/PerspectivesDataCenterCostAndPower.xls)

- Datacenter build costs are 17% of TCO
- Energy usage costs are 16% of TCO
- Server capex accounts for largest portion of TCO (61%)
  - Invest in mechanisms to improve work done per watt

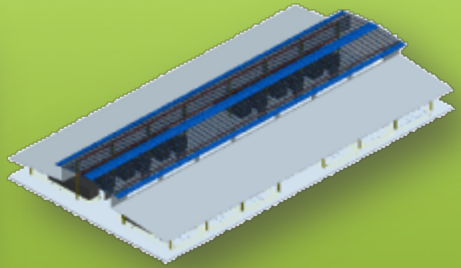
Microsoft's Datacenter Evolution

Datacenter Colocation
Generation 1

San Antonio & Quincy
Generation 2

Chicago & Dublin
Generation 3

Modular Datacenter
Generation 4



2005 2006 2007 2008 2009 2010+

DEPLOYMENT SCALE UNIT

EFFICIENT RESOURCE USAGE

Server

Capacity
~2 PUE

Rack

Density and
Deployment
1.4 – 1.6 PUE

Containers & Pods

Scalability and
Sustainability
1.2-1.5 PUE

ITPAC

Faster Time to Market
Reduced Carbon
1.05-1.15 PUE

Microsoft's Chicago Data Center

$500M+ investment

3000 construction related jobs

707,000 sq ft

60 MW Total Critical Power

1.5 million man-hours-of-labor

3400 tons of steel

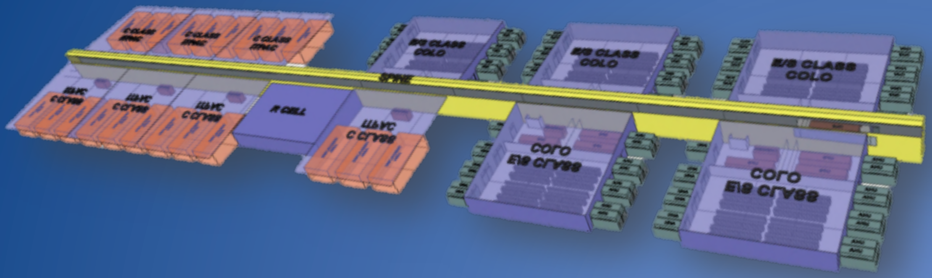
190 miles of conduit

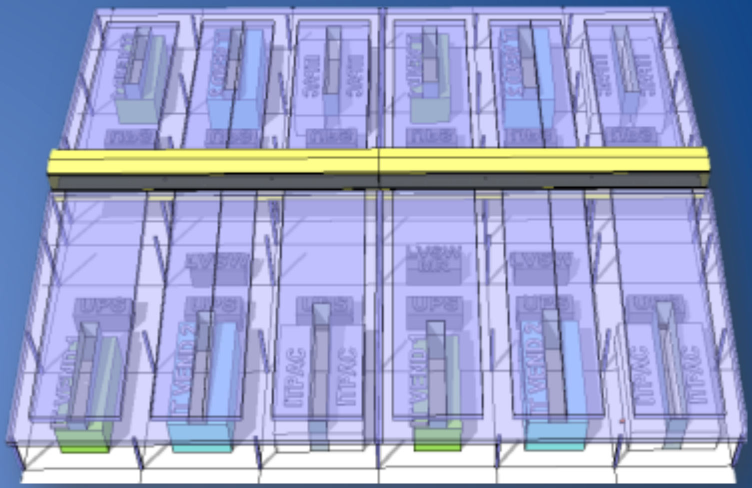
2400 tons of copper

7.5 miles of chilled water piping

26,000 cubic yards of concrete

Microsoft's Modular Datacenters





- ***Airside economization – outside air cooling!***
- Ultra-efficient water utilization
- Focus on renewable materials
- 30-50 % more cost effective
- 1.05 – 1.15 PUE
- ITPAC is the "datacenter-in-a-box"

Design factors for Datacenter Infrastructure

Datacenter architecture

- Cost efficiency
- Power distribution efficiency
- **Thermal design analysis (CFD modeling)**

Platform architecture

- Application performance analysis
- New platform architecture exploration

Reliability analysis

- Environmental operating ranges and impact on MTBF

Bringing it all together – Holistic Systems Design

ITPAC design overview (showing thermal modeling and construction)

video

Microsoft®

IT Pre-Assembled-Components (ITPAC)

Design factors for Datacenter Infrastructure

Datacenter architecture

- Cost efficiency
- Power distribution efficiency
- Thermal design analysis

Platform architecture

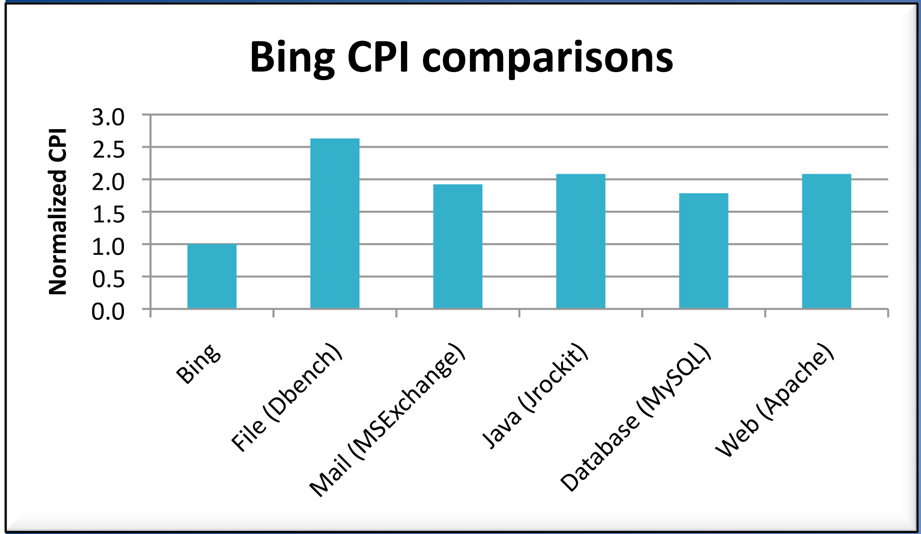
- **Application performance analysis**
- New platform architecture exploration

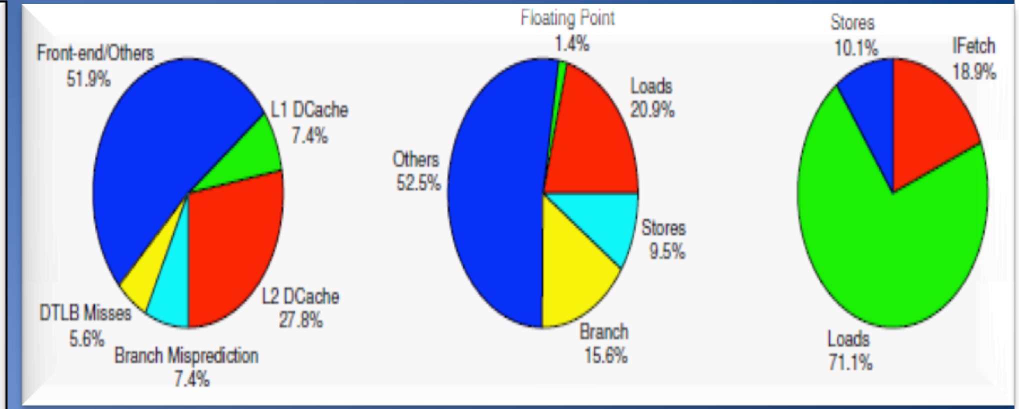
Reliability analysis

- Environmental operating ranges and impact on MTBF

Bringing it all together – Holistic Systems Design

CPU pipeline analysis



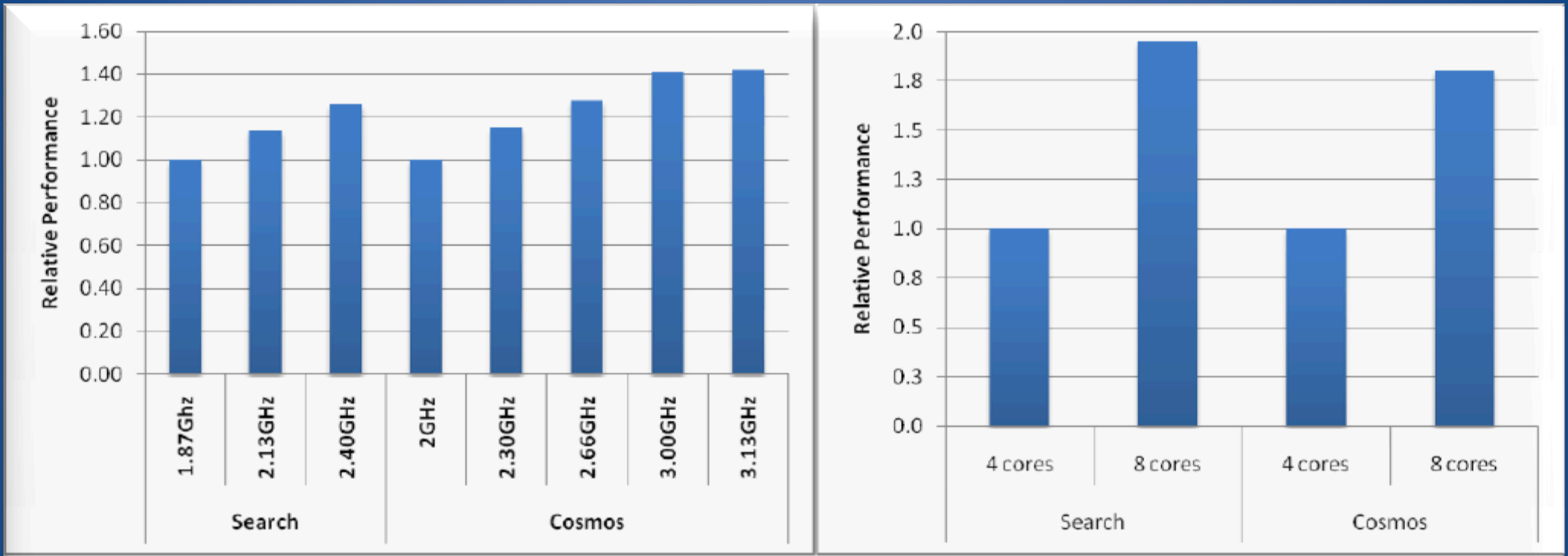


Stall cycle breakdown   Instruction Mix   Sources of cache access

Source: Reddi et al, "Web Search using Mobile Cores: Quantifying the Price of Efficiency, ISCA 2010

- Bing is computationally intensive and exhibits good instruction level parallelism
- Most memory operations are loads (as expected) for index reads
- Pipeline stalls due to cache misses are small (~7%) relative to other stall events
- Analysis helps understand how to optimize code for maximizing **QPS (Queries per Second)**

Workload scalability analysis

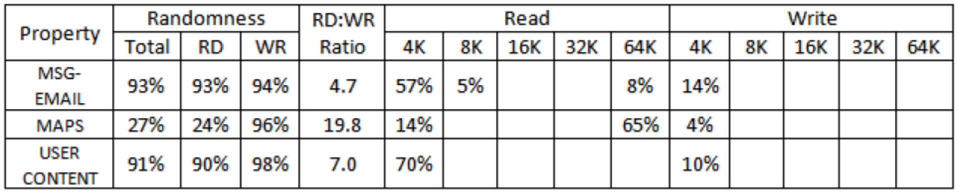


Performance scaling as a function of processor frequency and number of cores for Bing and Cosmos.

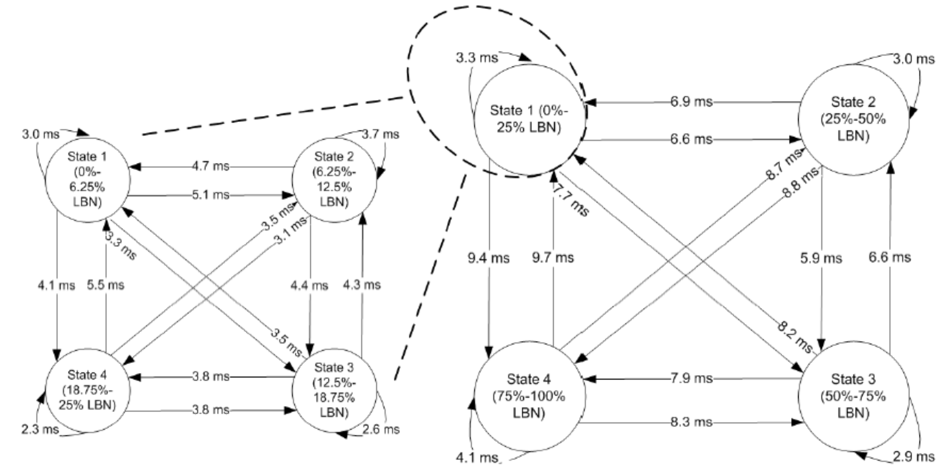
Source: Kozyrakis et al, "Server Engineering Insights for Online Services, IEEE Micro, July 2010

- How does the multicore trend affect online workloads?
- Identify workload scalability bottlenecks if any
- What is the tradeoff for different frequency offerings in commodity CPUs?

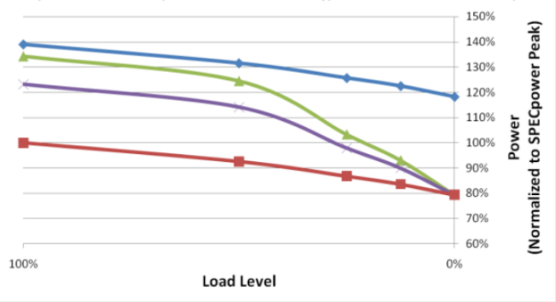
Storage rightsizing for power provisioning efficiency



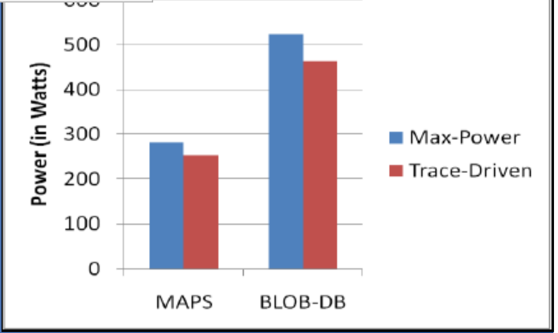
Property	Randomness			RD:WR	Read					Write				
	Total	RD	WR	Ratio	4K	8K	16K	32K	64K	4K	8K	16K	32K	64K
MSG-EMAIL	93%	93%	94%	4.7	57%	5%			8%	14%				
MAPS	27%	24%	96%	19.8	14%				65%	4%				
USER CONTENT	91%	90%	98%	7.0	70%					10%				



(a) Hierarchical sublevel for MSG-EMAIL (b) MSG-EMAIL with interarrival time



Provisioning through Trace-Driven Approach



Sankar et al, "Storage Characterization for Unstructured Data in Online Services Applications", IISWC 2009
Sankar et al, "Addressing the Stranded Power Problem in Datacenters Using Storage Workload Characterization", WOSP-SIPEW 2010

- E.g. above shows how in-depth storage trace analysis can be used to understand I/O workload patterns and **IOPS** rates
- HDD power models can then be used to determine optimal power provisioning values – minimizing stranded power and allowing higher density

Design factors for Datacenter Infrastructure

Datacenter architecture

- Cost efficiency
- Power distribution efficiency
- Thermal design analysis

Platform architecture

- Application performance analysis
- **New platform architecture exploration**

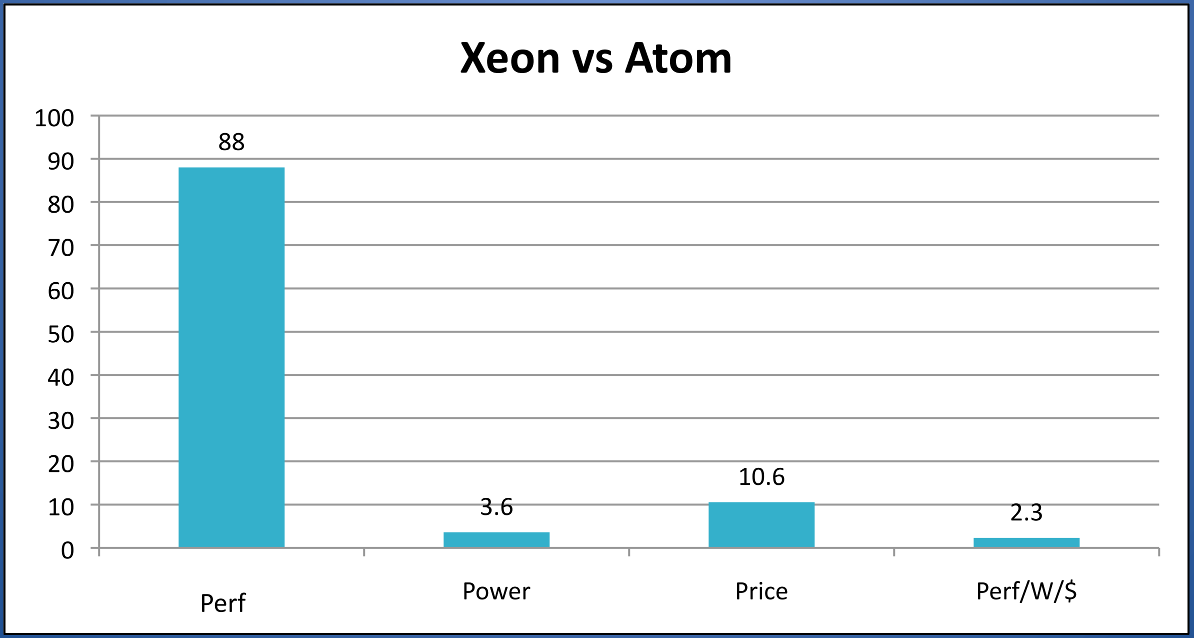
Reliability analysis

- Environmental operating ranges and impact on MTBF

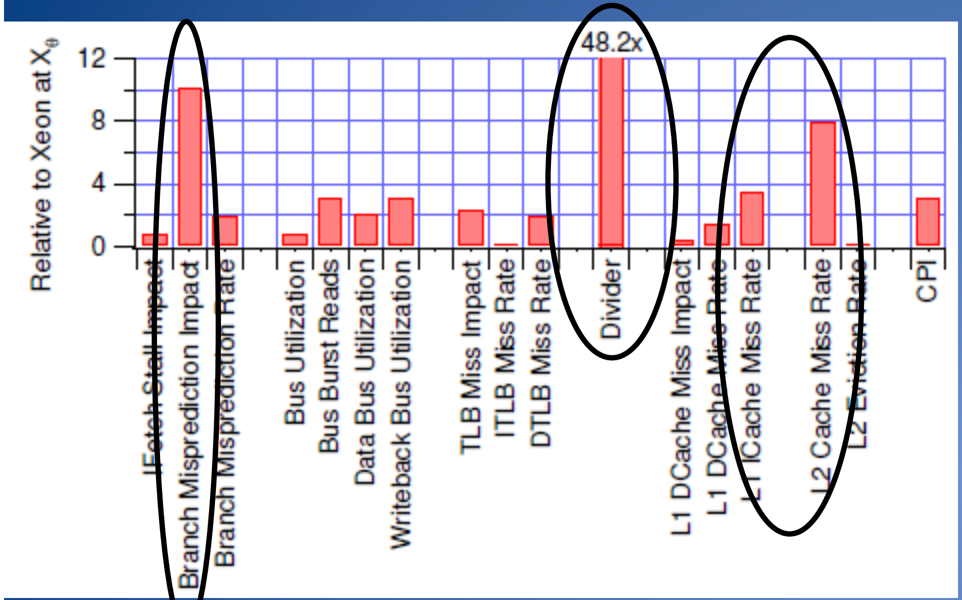
Bringing it all together – Holistic Systems Design

Evaluating mobile CPUs (Atom) for Bing

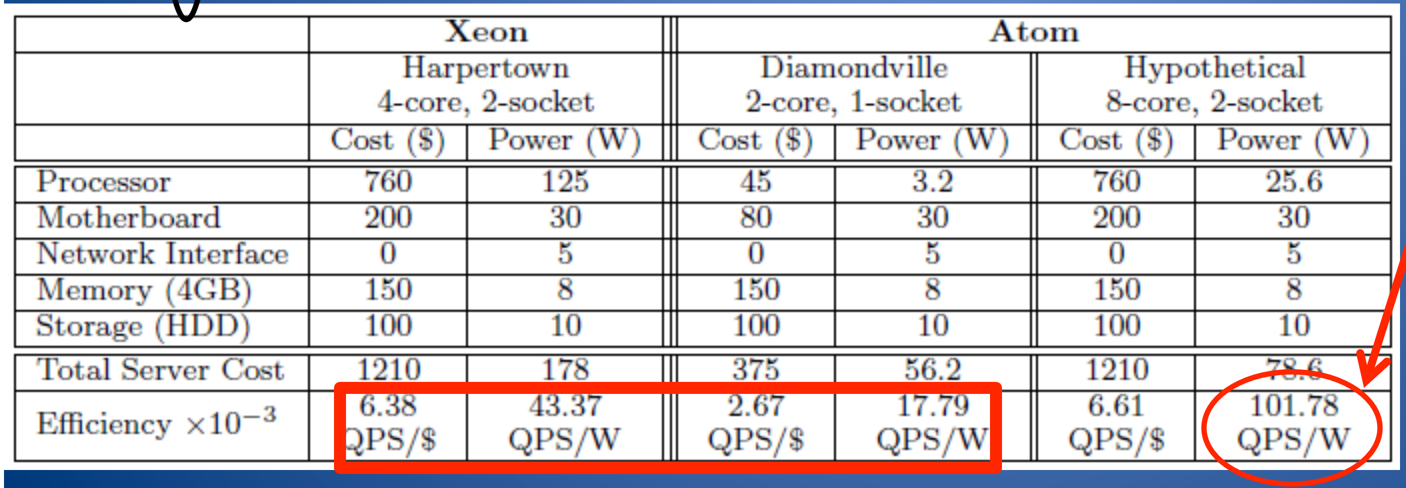
- Concept: Scale down the Bing platform to use ultra-low power mobile CPUs
  - Advantages: Low cost and power
  - Disadvantages: Performance, Significant tuning
- Results below show how the current systems compare
  - Overall, Xeon systems are currently 2.3x better than Atom on a Perf/W/$ basis
  - Atom system power is high, even though CPU power is low (few watts)
  - However, significant room for improvements as Atom-based servers are optimized in the future



CPU pipeline analysis for Atom



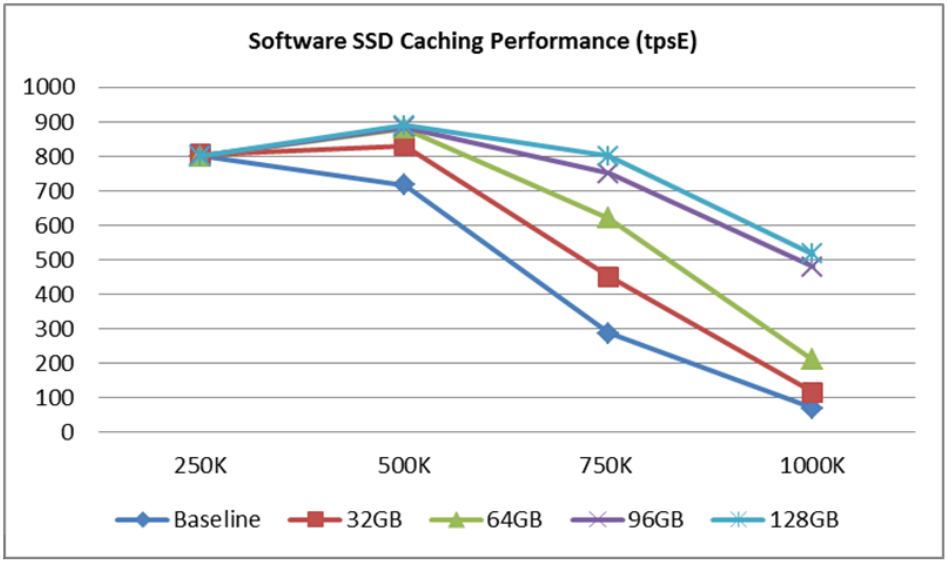
- Atom CPU weak on Branches, FP and caches → 3x worse CPI
- Overall TCO is 2.5x worse on both Perf/ $ and Perf/Watt
- Need future Atom CPUs to either provide improved perf or much lower power
- ***Source: "Web Search Using Mobile Cores", ISCA 2010***

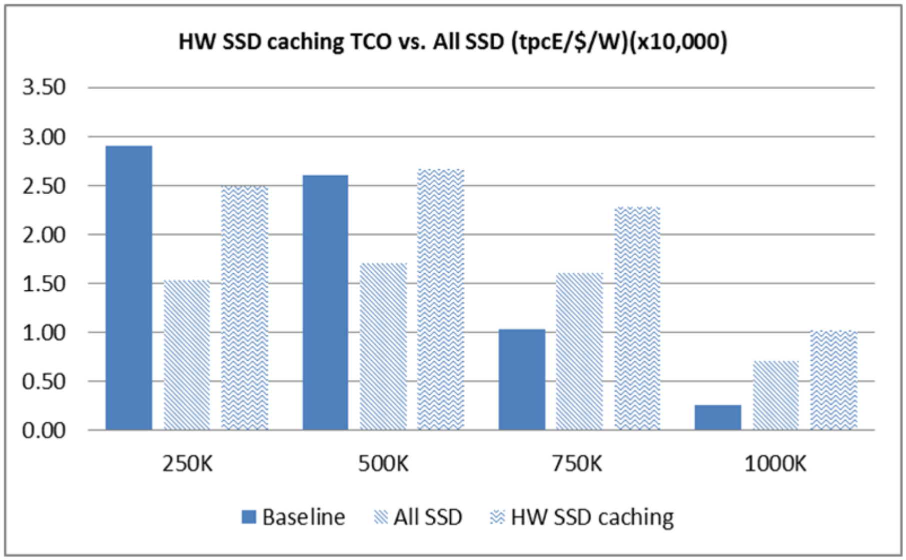


	Xeon		Atom			
	Harpertown 4-core, 2-socket		Diamondville 2-core, 1-socket		Hypothetical 8-core, 2-socket	
	Cost ($)	Power (W)	Cost ($)	Power (W)	Cost ($)	Power (W)
Processor	760	125	45	3.2	760	25.6
Motherboard	200	30	80	30	200	30
Network Interface	0	5	0	5	0	5
Memory (4GB)	150	8	150	8	150	8
Storage (HDD)	100	10	100	10	100	10
Total Server Cost	1210	178	375	56.2	1210	78.6
Efficiency $\times 10^{-3}$	6.38 QPS/$	43.37 QPS/W	2.67 QPS/$	17.79 QPS/W	6.61 QPS/$	101.78 QPS/W

Improving Perf/W for SQL workloads

- Seek architectural opportunities for ***dramatic*** Perf/W improvements
- E.g. shown below for SSD caching solutions implemented in HW RAID controllers and in SW buffer pool management algorithms
- Upto ~3x server consolidation possibility for database workloads


Software SSD Caching Performance (tpsE)


HW SSD caching TCO vs. All SSD (tpcE/$/W)(x10,000)

Khessib et al, "Using Solid State Drives as a Mid-Tier Cache in Enterprise OLTP applications", TPC Technology Conference on Performance Evaluation and Benchmarking (TPC-TC), Sept 2010

Design factors for Datacenter Infrastructure

Datacenter architecture

- Cost efficiency
- Power distribution efficiency
- Thermal design analysis

Platform architecture

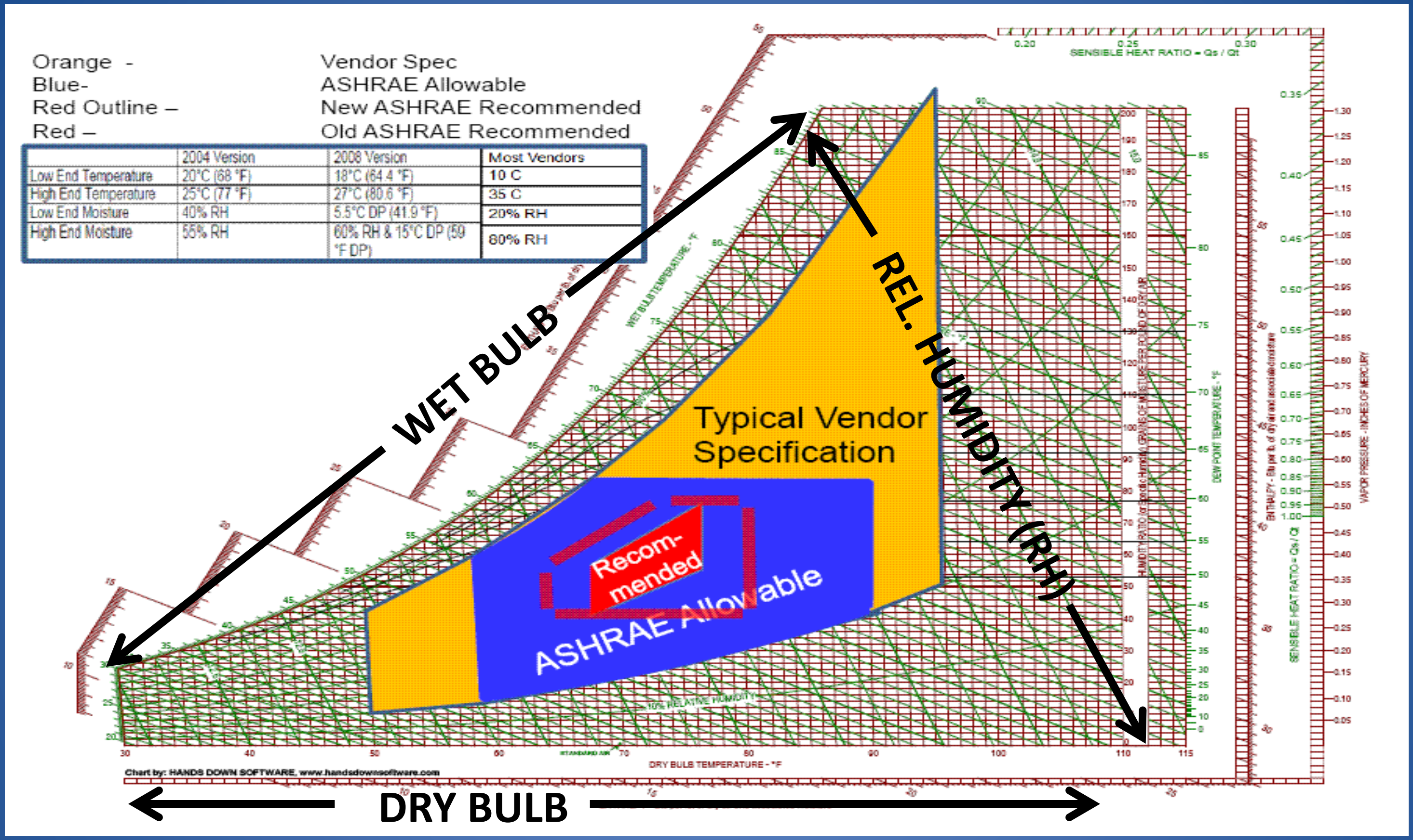
- Application performance analysis
- New platform architecture exploration

Reliability analysis

- **Environmental operating ranges and impact on MTBF**

Bringing it all together – Holistic Systems Design

Server operating range specification



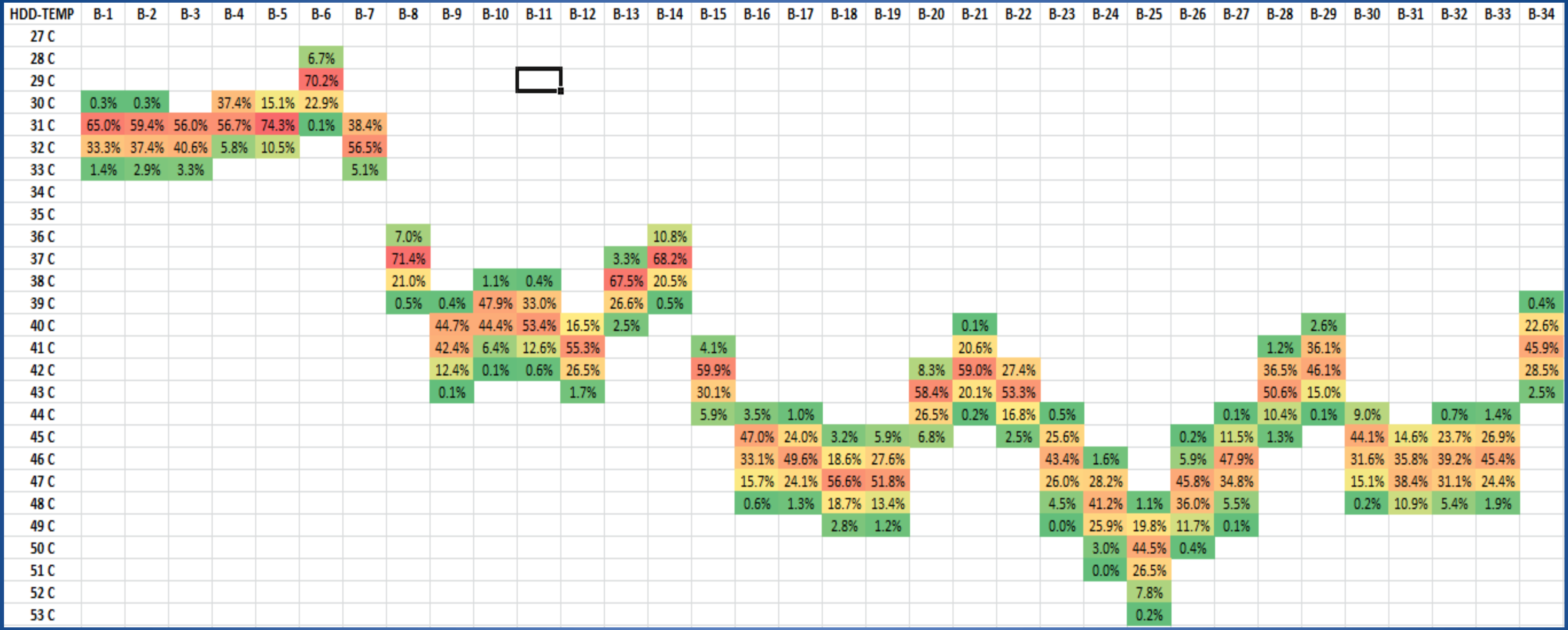
Orange - Vendor Spec
Blue- ASHRAE Allowable
Red Outline – New ASHRAE Recommended
Red – Old ASHRAE Recommended

	2004 Version	2008 Version	Most Vendors
Low End Temperature	20°C (68 °F)	18°C (64.4 °F)	10 C
High End Temperature	25°C (77 °F)	27°C (80.6 °F)	35 C
Low End Moisture	40% RH	5.5°C DP (41.9 °F)	20% RH
High End Moisture	55% RH	60% RH & 15°C DP (59 °F DP)	80% RH

Source: http://media.techtarget.com/digitalguide/images/Misc/temp_hr.gif

ASHRAE recommended range: 64F-81F Drybulb, max 60% RH
However, most server vendors specify 50F-95F, max 90%RH
Higher temperature operation → Lower fan speeds to cool servers → Improved PUE
However, need to take into account reliability aspects for server components

Temperature sensitivity analysis

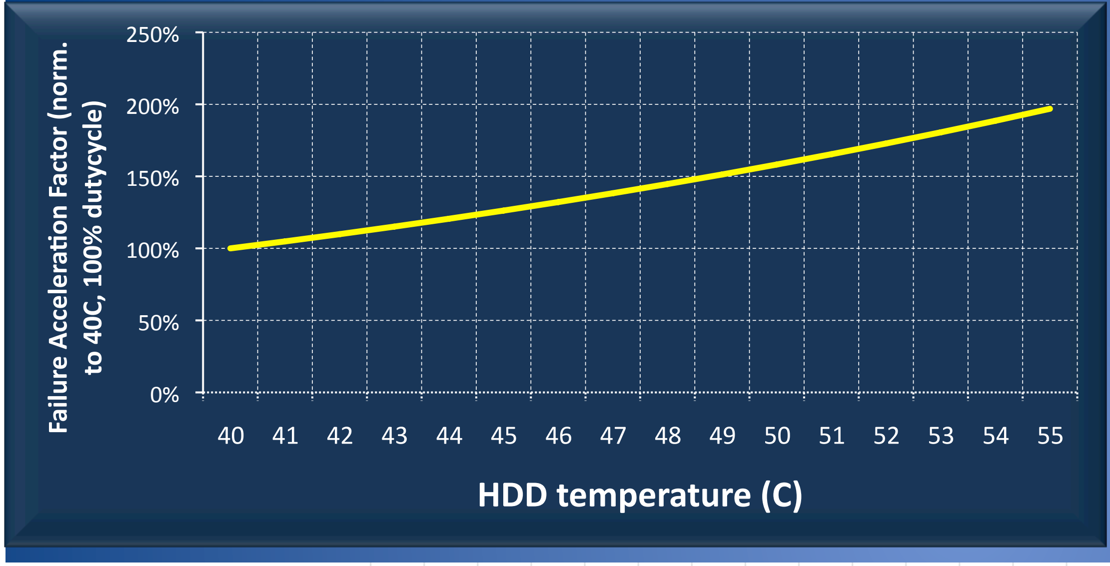


HDD-TEMP	B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8	B-9	B-10	B-11	B-12	B-13	B-14	B-15	B-16	B-17
27 C																	
28 C						6.7%											
29 C						70.2%											
30 C	0.3%	0.3%		37.4%	15.1%	22.9%											
31 C	65.0%	59.4%	56.0%	56.7%	74.3%	0.1%	38.4%										
32 C	33.3%	37.4%	40.6%	5.8%	10.5%		56.5%										
33 C	1.4%	2.9%	3.3%				5.1%										
34 C																	
35 C																	
36 C								7.0%						10.8%			
37 C								71.4%					3.3%	68.2%			
38 C								21.0%		1.1%	0.4%		67.5%	20.5%			
39 C								0.5%	0.4%	47.9%	33.0%		26.6%	0.5%			
40 C									44.7%	44.4%	53.4%	16.5%	2.5%				
41 C									42.4%	6.4%	12.6%	55.3%			4.1%		
42 C									12.4%	0.1%	0.6%	26.5%			59.9%		
43 C									0.1%			1.7%			30.1%		
44 C															5.9%	3.5%	1.0%
45 C																47.0%	24.0%
46 C																33.1%	49.6%
47 C																15.7%	24.1%
48 C																0.6%	1.3%
49 C																	
50 C																	
51 C																	
52 C																	
53 C																	

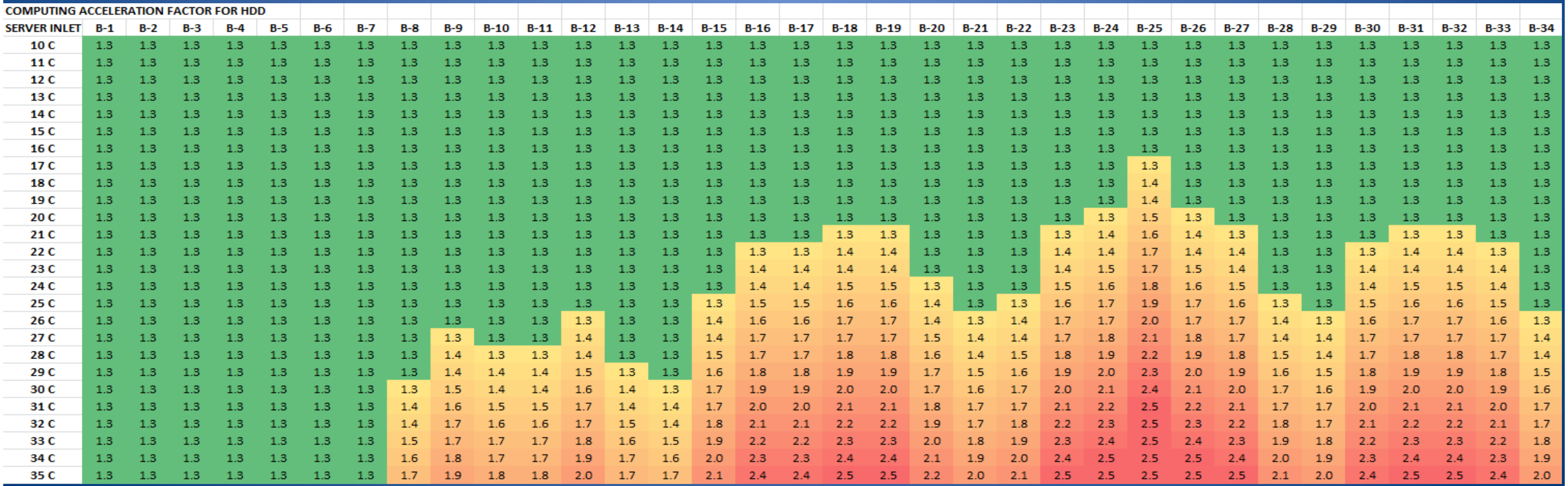
HDD-TEMP	B-18	B-19	B-20	B-21	B-22	B-23	B-24	B-25	B-26	B-27	B-28	B-29	B-30	B-31	B-32	B-33	B-34
27 C																	
28 C																	
29 C																	
30 C																	
31 C																	
32 C																	
33 C																	
34 C																	
35 C																	
36 C																	
37 C																	
38 C																	
39 C																	0.4%
40 C				0.1%								2.6%					22.6%
41 C				20.6%							1.2%	36.1%					45.9%
42 C			8.3%	59.0%	27.4%						36.5%	46.1%					28.5%
43 C			58.4%	20.1%	53.3%						50.6%	15.0%					2.5%
44 C			26.5%	0.2%	16.8%	0.5%				0.1%	10.4%	0.1%	9.0%		0.7%	1.4%	
45 C	3.2%	5.9%	6.8%		2.5%	25.6%			0.2%	11.5%	1.3%		44.1%	14.6%	23.7%	26.9%	
46 C	18.6%	27.6%				43.4%	1.6%		5.9%	47.9%			31.6%	35.8%	39.2%	45.4%	
47 C	56.6%	51.8%				26.0%	28.2%		45.8%	34.8%			15.1%	38.4%	31.1%	24.4%	
48 C	18.7%	13.4%				4.5%	41.2%	1.1%	36.0%	5.5%			0.2%	10.9%	5.4%	1.9%	
49 C	2.8%	1.2%				0.0%	25.9%	19.8%	11.7%	0.1%							
50 C							3.0%	44.5%	0.4%								
51 C							0.0%	26.5%									
52 C								7.8%									
53 C								0.2%									

- HDD case temp distribution shown for a 35-bay JBOD array (over 3 months) for fixed server inlet temperature (25C)
- **Note increase in ΔT** for inner drive bays
- Airside economization scenarios may imply higher inlet temperatures - **What are implications to HDD MTBF?**

Modeling HDD AFR sensitivity to temperature



- Model the HDD Failure Acceleration Factor, adjusted for production failure data (left figure)
- Create mapping based on temperature distribution data (figure below)
- Calculate overall AFR for a given input server inlet temperature



COMPUTING ACCELERATION FACTOR FOR HDD

SERVER INLET	B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8	B-9	B-10	B-11	B-12	B-13	B-14	B-15	B-16	B-17	B-18	B-19	B-20	B-21	B-22	B-23	B-24	B-25	B-26	B-27	B-28	B-29	B-30	B-31	B-32	B-33	B-34
10 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
11 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
12 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
13 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
14 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
15 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
16 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
17 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
18 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
19 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
20 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.5	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
21 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.6	1.4	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
22 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.4	1.3	1.3	1.3	1.4	1.4	1.7	1.4	1.4	1.3	1.3	1.3	1.4	1.4	1.3	1.3
23 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.4	1.4	1.4	1.3	1.3	1.3	1.4	1.5	1.7	1.5	1.4	1.3	1.3	1.4	1.4	1.4	1.4	1.3
24 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.4	1.5	1.5	1.3	1.3	1.3	1.5	1.6	1.8	1.6	1.5	1.3	1.3	1.4	1.5	1.5	1.4	1.3
25 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.5	1.5	1.6	1.6	1.4	1.3	1.3	1.6	1.7	1.9	1.7	1.6	1.3	1.3	1.5	1.6	1.6	1.5	1.3
26 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.6	1.6	1.7	1.7	1.4	1.3	1.4	1.7	1.7	2.0	1.7	1.7	1.4	1.3	1.6	1.7	1.7	1.6	1.3
27 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.3	1.3	1.4	1.7	1.7	1.7	1.7	1.5	1.4	1.4	1.7	1.8	2.1	1.8	1.7	1.4	1.4	1.7	1.7	1.7	1.7	1.4
28 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.3	1.3	1.4	1.3	1.3	1.5	1.7	1.7	1.8	1.8	1.6	1.4	1.5	1.8	1.9	2.2	1.9	1.8	1.5	1.4	1.7	1.8	1.8	1.7	1.4
29 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.4	1.4	1.5	1.3	1.3	1.6	1.8	1.8	1.9	1.9	1.7	1.5	1.6	1.9	2.0	2.3	2.0	1.9	1.6	1.5	1.8	1.9	1.9	1.8	1.5
30 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.5	1.4	1.4	1.6	1.4	1.3	1.7	1.9	1.9	2.0	2.0	1.7	1.6	1.7	2.0	2.1	2.4	2.1	2.0	1.7	1.6	1.9	2.0	2.0	1.9	1.6
31 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.6	1.5	1.5	1.7	1.4	1.4	1.7	2.0	2.0	2.1	2.1	1.8	1.7	1.7	2.1	2.2	2.5	2.2	2.1	1.7	1.7	2.0	2.1	2.1	2.0	1.7
32 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.7	1.6	1.6	1.7	1.5	1.4	1.8	2.1	2.1	2.2	2.2	1.9	1.7	1.8	2.2	2.3	2.5	2.3	2.2	1.8	1.7	2.1	2.2	2.2	2.1	1.7
33 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.5	1.7	1.7	1.7	1.8	1.6	1.5	1.9	2.2	2.2	2.3	2.3	2.0	1.8	1.9	2.3	2.4	2.5	2.4	2.3	1.9	1.8	2.2	2.3	2.3	2.2	1.8
34 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.6	1.8	1.7	1.7	1.9	1.7	1.6	2.0	2.3	2.3	2.4	2.4	2.1	1.9	2.0	2.4	2.5	2.5	2.5	2.4	2.0	1.9	2.3	2.4	2.4	2.3	1.9
35 C	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.7	1.9	1.8	1.8	2.0	1.7	1.7	2.1	2.4	2.4	2.5	2.5	2.2	2.0	2.1	2.5	2.5	2.5	2.5	2.5	2.1	2.0	2.4	2.5	2.5	2.4	2.0

Design factors for Datacenter Infrastructure

Datacenter architecture

- Cost efficiency
- Power distribution efficiency
- Thermal design analysis

Platform architecture

- Application performance analysis
- New platform architecture exploration

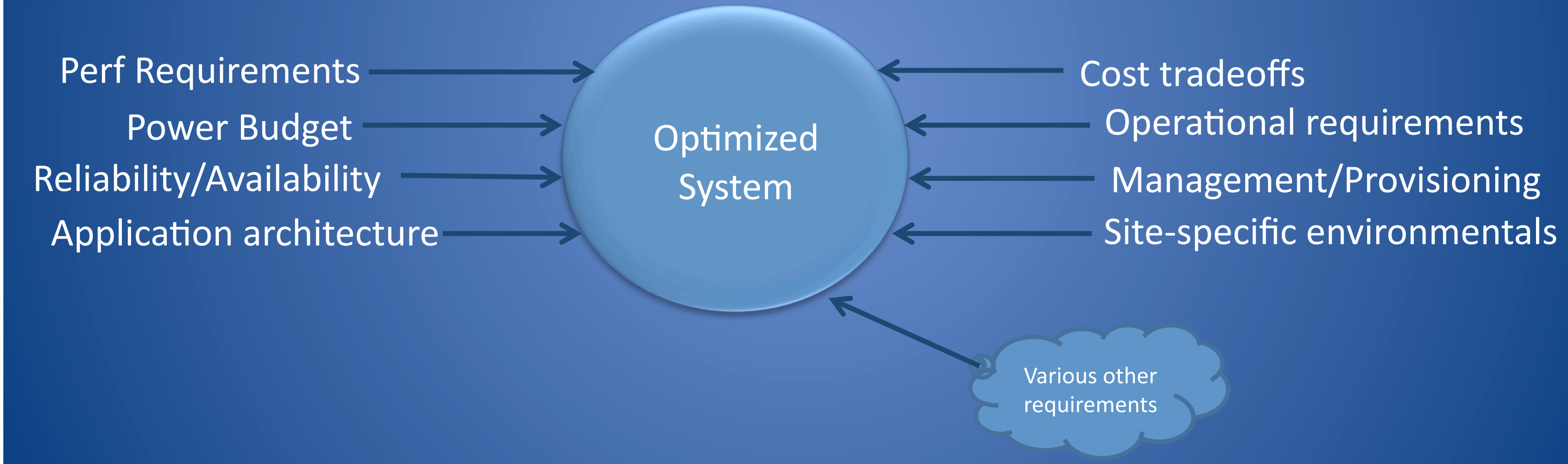
Reliability analysis

- Environmental operating ranges and impact on MTBF

Bringing it all together – Holistic Systems Design

Optimal systems design (Infrastructure perspective)

Multi-dimensional optimization problem



Perf Requirements

Power Budget

Reliability/Availability

Application architecture

Optimized System

Cost tradeoffs

Operational requirements

Management/Provisioning

Site-specific environmentals

Various other requirements

Research areas

- Optimal Power provisioning for high dynamic range workloads (idle at night, peak load during day)
- Power aware task scheduling on large clusters
- Energy proportionality via system architecture innovations
- Reliability model for entire datacenter, correlated for parameters such as temperature, humidity, load levels
- Failure prediction opportunities based on log analysis of continuous feeds from management consoles

- Several others…

Summary

- Datacenter design and optimization for various application scenarios involves several disciplines with complex interactions
- An optimal design is usually a delicate balance between various tradeoffs – no easy answers
- Extensive data analysis is the key to making effective design choices
- Several areas for improved design and reliability via data mining and predictive analysis

Q & A