Performance Analysis in the Real World of Online Services

Dileep Bhandarkar, Ph. D.
Distinguished Engineer

2009 IEEE International Symposium on Performance Analysis of Systems and Software

Microsoft®
Global Foundation Services

Microsoft

My Background:

Crossing the bridge from Architect to User!

- 1973: PhD in Electrical Engineering
  - Carnegie Mellon University
  - Thesis: Performance Evaluation of Multiprocessor Systems
- 4 years - Texas Instruments
  - Research
- 17 years - Digital Equipment Corporation
  - Processor Architecture and Performance
- 12 years - Intel
  - Performance, Architecture, Strategic Planning
- 2 years - Microsoft
  - Data Center Hardware Engineering

Global Foundation Services

Delivering across the company, all over the world, around the clock

Microsoft FlexGo
Windows Live
Microsoft Dynamics CRM Online
Microsoft Office Live WorkSpaces
Live Search
Microsoft Office Live Meeting
Microsoft Forefront
Windows Azure
Microsoft Office Online
Microsoft Hosted Messaging & Collaboration
Windows Live ID
Windows Live Hotmail
Windows Live SkyDrive
Microsoft Silverlight
XBOX LIVE
Microsoft Office Groove 2007
Windows Live Spaces
Microsoft HealthVault
Windows Live Search
Microsoft Online Services
msn
Live Search Maps
Connected Services SANDBOX
Enabling Managed Network Mashups
Microsoft Office SharePoint Online
msn adCenter
Microsoft Exchange Online
Microsoft Office Communications Online
Microsoft game studios
Windows Live Messenger

Plus over 150 more sites and services

Performance

- per·form·ance (pər fôr′məns)
- noun

1. the manner in which or the efficiency with which something reacts or fulfills its intended purpose.
2. operation or functioning, usually with regard to effectiveness, as of a machine
3. the ability to perform: efficiency b: the manner in which a mechanism performs <engine *performance*>

What Matters

- Throughput, Response Time
- Cost Effectiveness
  - Performance / $
  - Performance / Watt
  - Performance / $ / Watt
- Energy Consumption / Efficiency
- Reliability
- Maintainability

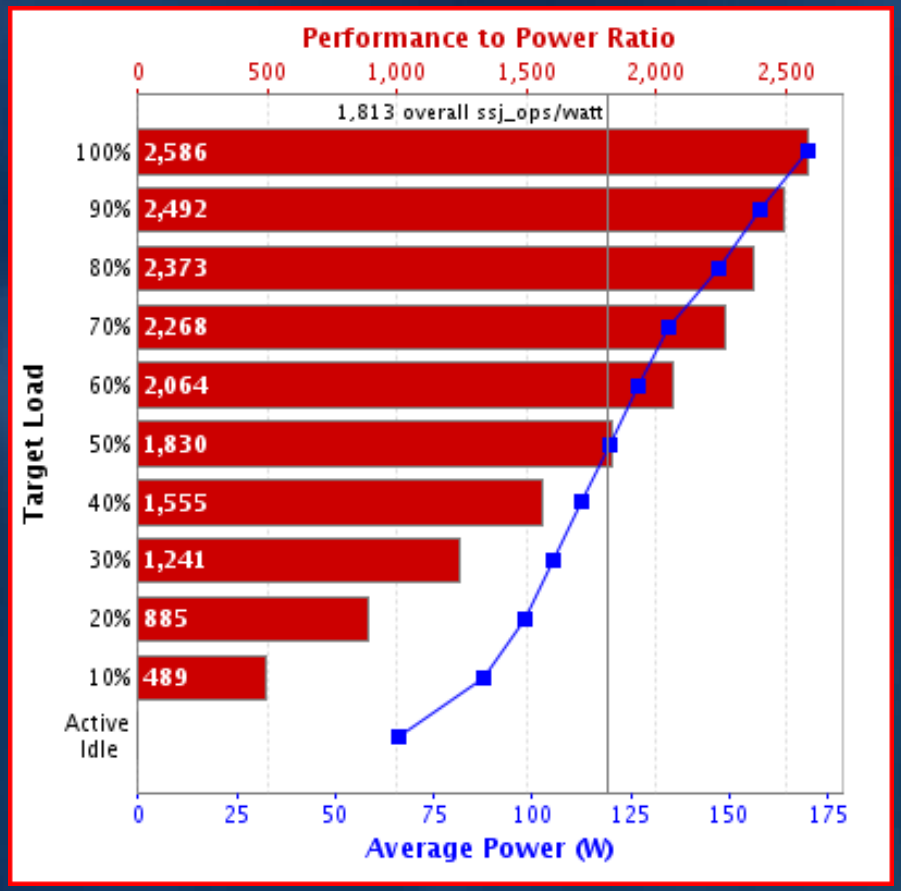
Common Performance Metrics

- SPEC CPU Benchmarks
  - Highly tuned using compiler versions you may not have access to today
  - Geometric Mean of several individual benchmarks
  - Single and multistream
  - Integer and Floating Point
- Transaction Processing Council
  - TPC-C, TPC-E, TPC-H
- SPECpower
- Lots of lesser known benchmarks – SPECweb, IOmeter, FSCT, WCAT etc

End User Issues

- Performance (speed) on old binaries
- My code compiled with my favorite compiler using my optimization level running on my operating system
- Industry standard benchmarks are a good first indicator, but not sufficient
- Configuration dependencies

Read the Fine Print!



- SPECpower Example:
- Impressive Result!
- 2 processors
- 8 GB memory
- 1 SSD!

Source: www.spec.org

TPC-C Example

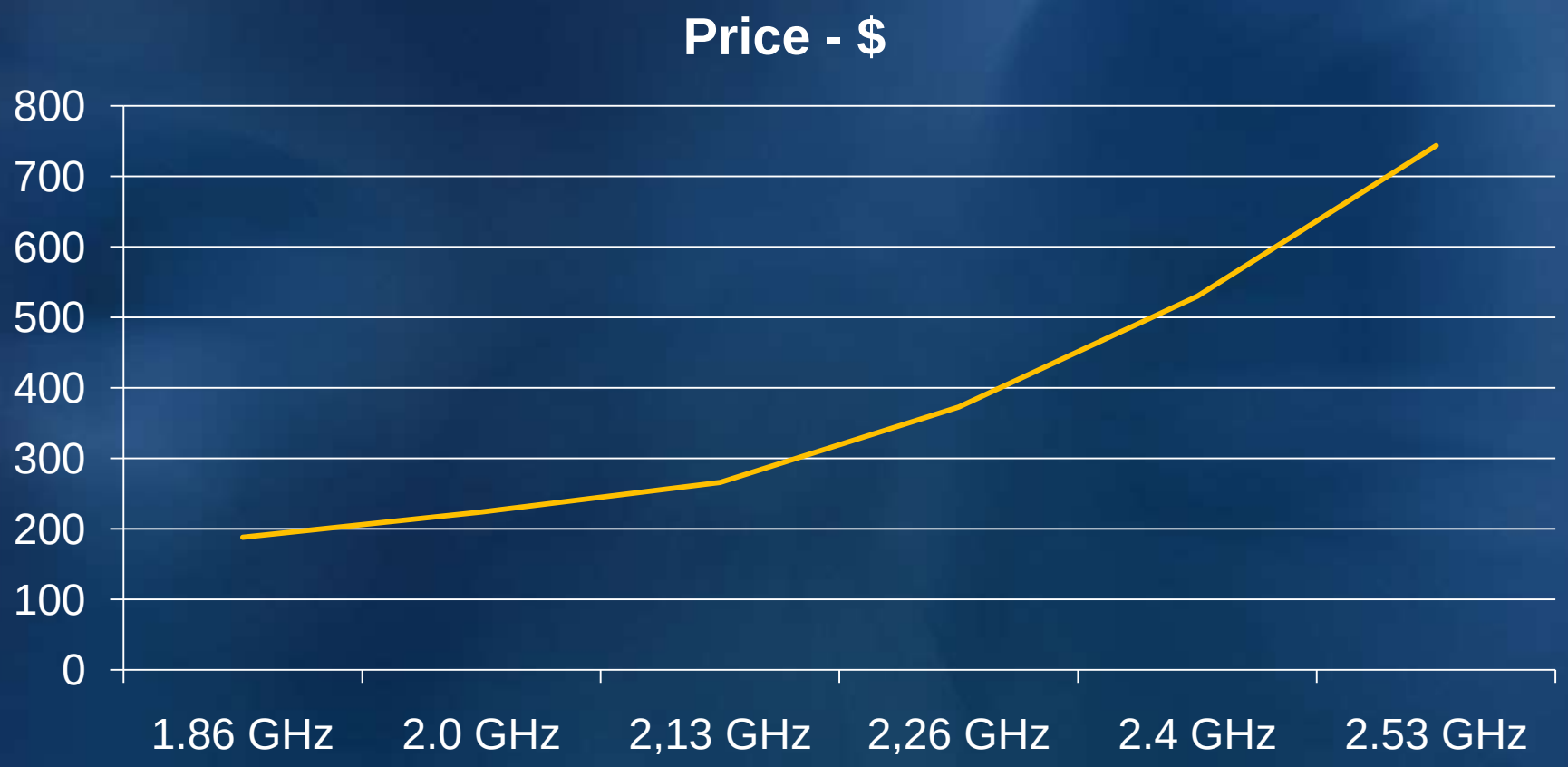
- Dual Processor Platform with price starting at $ 3,315
- TPC-C Throughput: 631,766 tpmC
- Price/Performance: $1.08 USD per tpmC
- Total System Cost: $678,231
- 144 GB memory (18 x 8 GB)
- 26 10K RPM disk drives
- 1184 15K RPM disk drives
- Typical User Configuration: 16 GB memory, mid bin processor, ~8 disk drives for ~$10K

Source: www.tpc.org

Balanced Performance

- Typical User Configuration
  - Smaller memory generates more disk I/O
  - Disk I/O limited
  - Low Processor Utilization
- User Options
  - Add more memory and drives – higher cost
  - Deploy just one processor
  - Use lower frequency processor to save $ and power

Non-Linear Price Performance



Source: www.intc.com

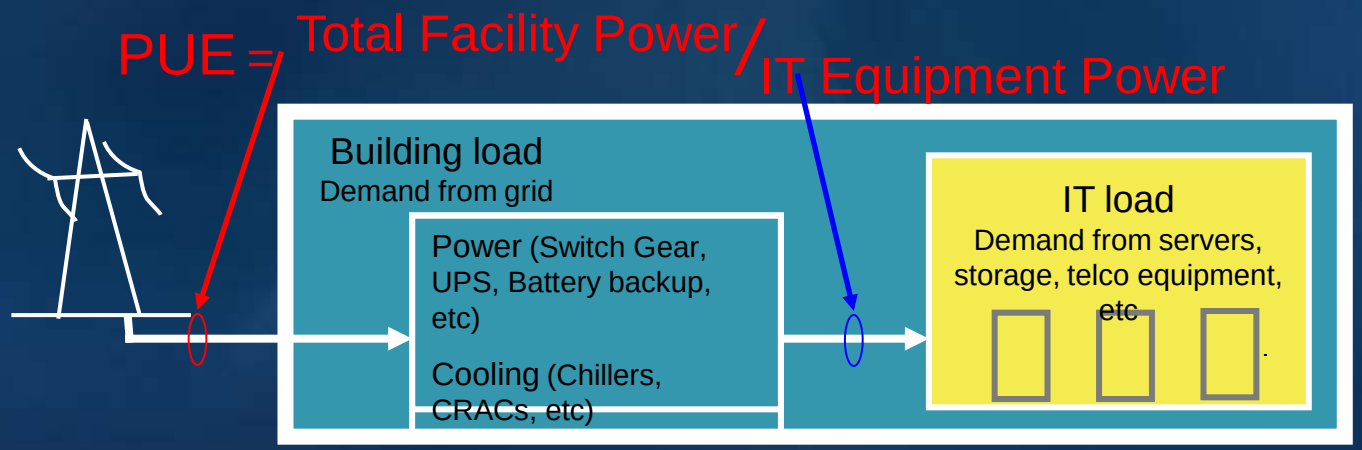
E5540 (8M L2 cache 2.53 GHz (80W) 5.86 GT/sec Intel® QPI 45nm)	-	$744
E5530 (8M L2 cache 2.40 GHz (80W) 5.86 GT/sec Intel® QPI 45nm)	-	$530
E5520 (8M L2 cache 2.26 GHz (80W) 5.86 GT/sec Intel® QPI 45nm)	-	$373
E5506 (4M L2 cache 2.13 GHz (80W) 4.80 GT/sec Intel® QPI 45nm)	-	$266
E5504 (4M L2 cache 2.00 GHz (80W) 4.80 GT/sec Intel® QPI 45nm)	-	$224
E5502 (4M L2 cache 1.86 GHz (80W) 4.80 GT/sec Intel® QPI 45nm)	-	$188

Power vs Performance Tradeoff

- 2.4 GHz vs 2.26 GHz Frequency
  - ~5% frequency difference results in 2-3% performance difference on most workloads
- What about Power?
  - 80 W vs 60W TDP
  - 10-15 Watts savings dependent on workload
- What about Price?
  - Same!
  - Total Cost of Ownership lower for 60W CPU

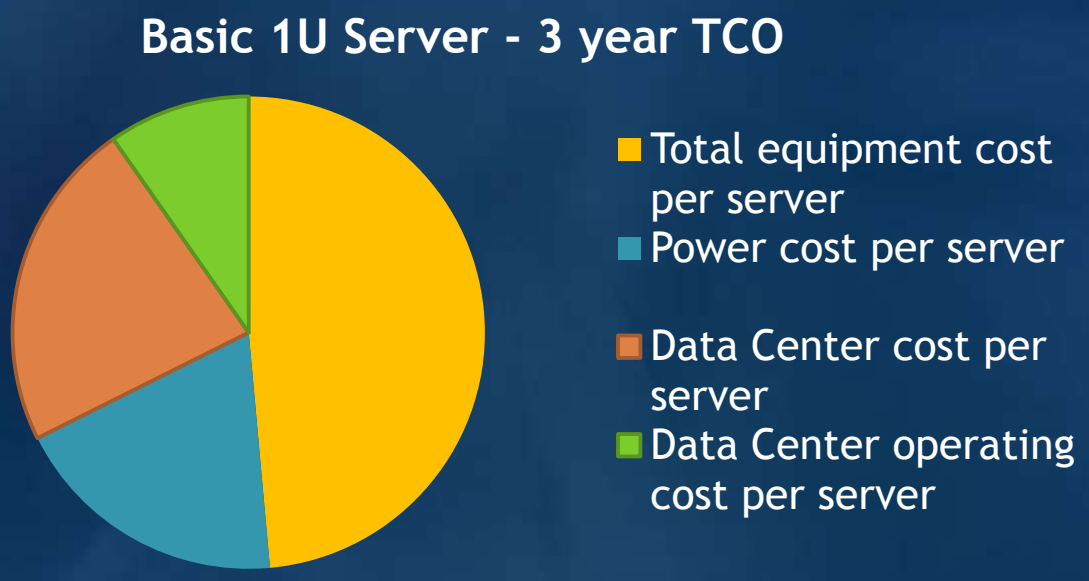
Why Power is Important?

- Energy Consumption: US power rate 10.27 cents per Kilowatt hour) in 2008 according to DOE/eia (http://www.eia.doe.gov/cneaf/electricity/epm/table5_3.html)
- In a typical data center for every watt in server power there can be another 0.5 to 1 watt consumed for power distribution losses and cooling.



But There is More!

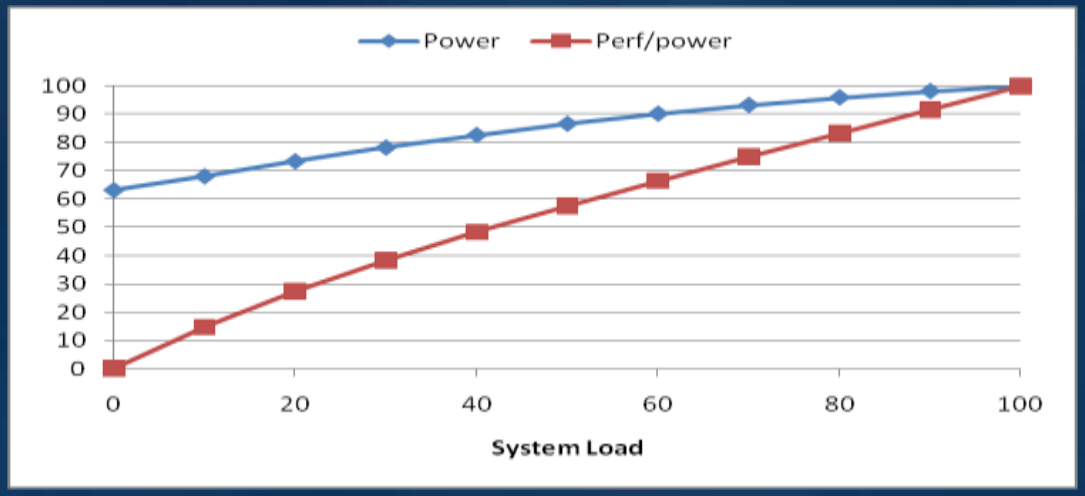
- Data centers can cost between $10M and $20M per Megawatt
- Capital Depreciation Costs are Important



Your Mileage May Vary!

Power vs System Load

- Typical Web Applications are not CPU or disk intensive
- Platform imbalance can keep processor utilization low
- Idle power is typically over 50% of peak



Our Approach

- Invest in understanding your workloads
- Measure, Model, Validate, Predict, Measure
- Focus on Entire System
- Balance the Platform
- Focus on Total Cost of Ownership
  - Acquisition Cost
  - Energy Consumption Cost
  - Data Center Capital Cost

2 Socket Catches up to 4 Socket

Source: www.tpc.org, www.intc.com

	HP ProLiant DL580 G5 2.67 GHz 16MB L2 C/S with 8 ProLiant DL360G5		TPC-C Version 5.9 Report Date January 16, 2009
Total System Cost	TPC-C Throughput	Price/Performance	4 x Intel x7460 2.67GHZ/16M 256GB
$615,914 USD	639,253 tpmC	$0.97 USD/ tpmC	

Intel® Xeon® processor MP² 7400 series, 6 and 4 core versions Server/Workstation (uFCPGA)	Mar '09 (03/15) Price
X7460 (16M L3 cache 2.66 GHz (130W) 1066 MHz FSB 45nm)	$2,729
E7450 (12M L3 cache 2.40 GHz (90W) 1066 MHz FSB 45nm)	$2,301

	HP ProLiant DL370 G6 2.93 GHz 8 MB L2 C/S with 8 ProLiant DL360G5		TPC-C Version 5.10 Report Date March 30, 2009
Total System Cost	HP ProLiant DL370 G6 w/ 2xIntel X5570, 144GB		Availability Date
$678,231 USD	631,766 tpmC	$1.08 USD/ tpmC	March 30, 2009

X5570 (8M L2 cache 2.93 GHz (95W) 6.40 GT/sec Intel® QPI 45nm)	-	$1,386
X5560 (8M L2 cache 2.80 GHz (95W) 6.40 GT/sec Intel® QPI 45nm)	-	$1,172
X5550 (8M L2 cache 2.66 GHz (95W) 6.40 GT/sec Intel® QPI 45nm)	-	$958
E5540 (8M L2 cache 2.53 GHz (80W) 5.86 GT/sec Intel® QPI 45nm)	-	$744
E5530 (8M L2 cache 2.40 GHz (80W) 5.86 GT/sec Intel® QPI 45nm)	-	$530

Best Price/Performance

DELL	PowerEdge 2900 Server with Oracle Database 11g Standard Edition One			TPC-C Rev 5.10 Original Report Date **February 20, 2009**
Total System Cost	TPC-C Throughput	Price/Performance		Availability Date
$62,567	**104,492 tpmC**	**$.60 / tpmC**		February 20, 2009
Processors	Database Manager	OS	Other Software	Number of Users
1/4/4 Quad Core Intel® Xeon® 5440,2X6MB Cache, 2.83GHZ 1333MHZ FSB	Oracle Database 11g Standard Edition One	Microsoft Windows Server 2003 Standard x64 Edition SP1	Windows Server 2003 Standard Edition w/ COM+ Internet Information Server 6.0 Microsoft Visual C++	82,600

Single Processor, 32 GB Memory, 102 Disk Drives

Typical Configuration: 1 Quad Core Processor, 16 GB Memory, 8 Disk Drives (10K RPM, 146 GB), List Price: ~$7K

Source: www.tpc.org

TPC-E Example

FUJITSU	PRIMERGY RX300 S5	TPC-E 1.7.0 TPC Pricing 1.3.0 Report Date March 30, 2009	
TPC-E Throughput 800.00 tpsE	Price/Performance $ 343.91 USD per tpsE	Availability Date April 1, 2009	Total System Cost $ 275,131
Database Server Configuration			
Operating System Microsoft Windows Server 2008 Enterprise x64 Edition	Database Manager Microsoft SQL Server 2008 Enterprise x64 Edition	Processors/Cores/Threads 2/8/16	Memory 96 GB

2x Intel Xeon X5570 2.93 GHz
96 GB Memory
2x 73 GB 15K SAS Drives
4x 300 GB 15K SAS Drives
Onboard SAS RAID Controller
5x SAS RAID Controller
Onboard 2x 1 Gb/s

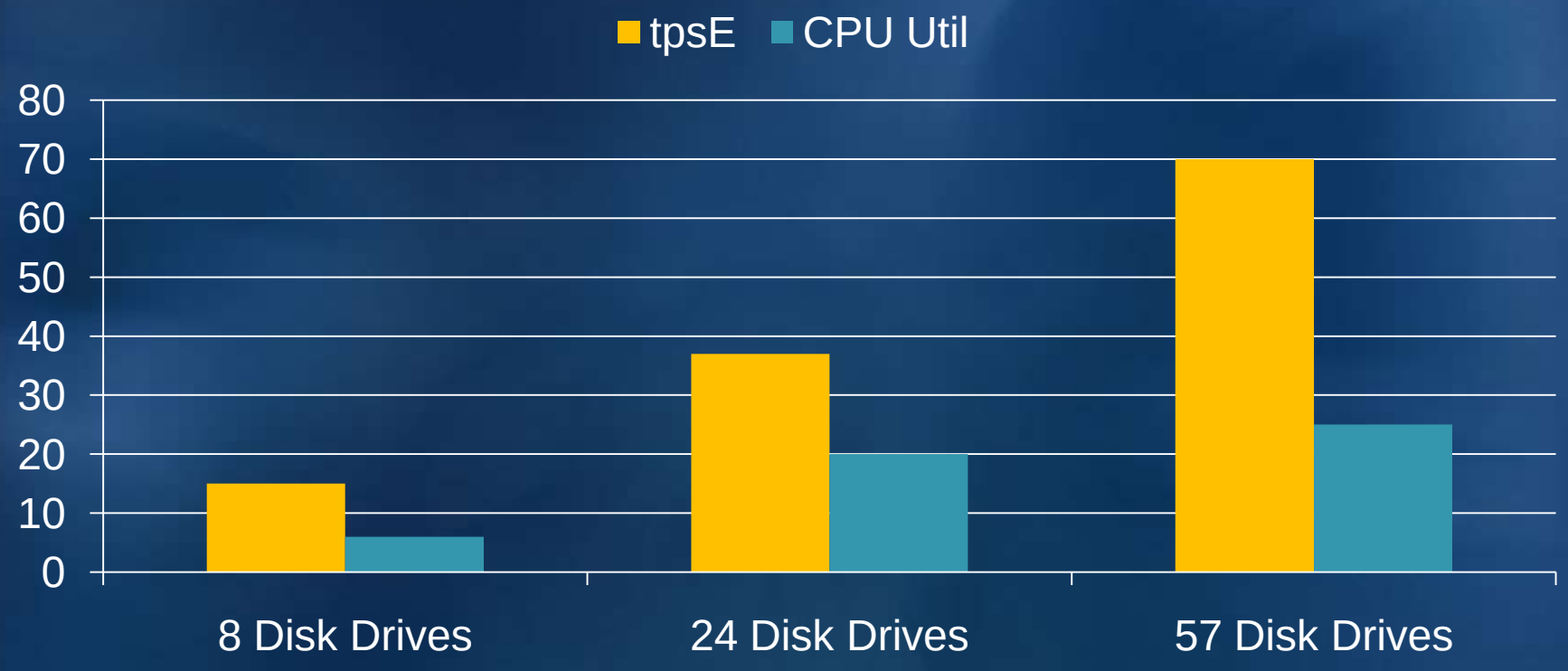
Storage
192 x 73 GB 15K
168 x 146GB 15K
4 x 300GB 15K

Typical Configuration:
2 processors
32 GB Memory
24 x 146 GB 10K RPM Drives
List Price: ~$15K

Testing with TPC-E

- Published results at www.tpc.org
  - Use very small portion of available disk capacity
  - Data on outer tracks of disks
    - seek distance per disk: minimal for random access pattern
    - spread out across numerous disks to get IOPS
  - Not representative for “real world usage”
- Our Methodology
  - Fill disk capacity for any server from 20-100% in increments of 20% (Simulate partial capacity utilization)
  - Vary Active customer load from 20-100% (Simulate partial working set)
  - Weighted Harmonic Mean to give a single “tpsE” score for the server
  - Representative of our usage scenario
    - Different customers can utilize capacity to different extent
    - Working sets are not usually the entire customer base

Typical System Performance



Disk I/O Performance is a Dominant Factor

Web Capacity Analysis Tool (WCAT)

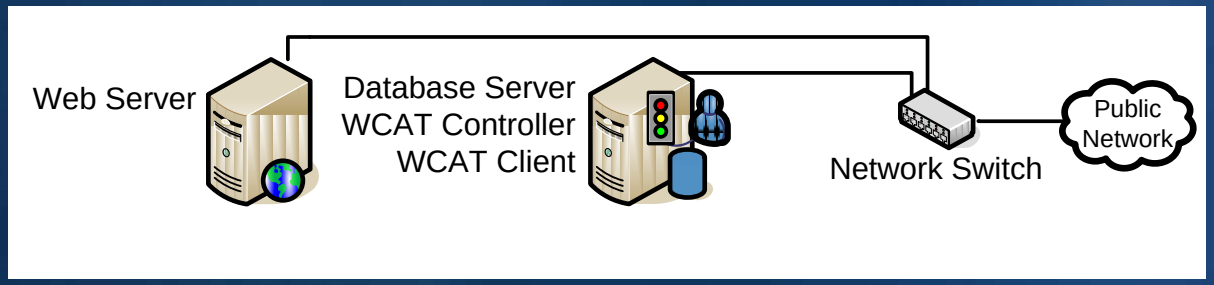
- Lightweight HTTP load generation tool
- Internal Microsoft tool created by Windows Server Performance Team
- Available for public download as IIS productivity tool (from www.iis.net)
- Clear structured tool with good scalability, allows specific setups for different usage scenarios

- Basic scenarios we use in our web testing:
  - Dynamic .ASPX content (CPU intensive)
  - Static “cold” content (Disk intensive)
  - Static “hot” cached content (network intensive)
  - Windows Live mix content

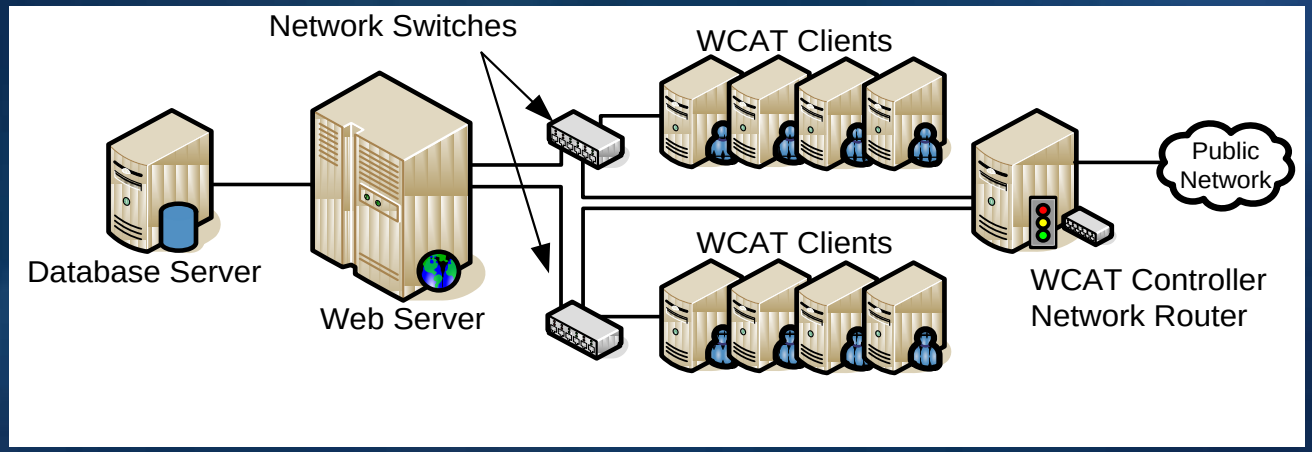
Web Capacity Analysis Tool Test Environment

- Web Server under the test: contains workload content: ~ 2.5 GB, ~ 4 mln files (aspx, gif, html)
- WCAT Clients: provide actual load to the Web Server according to the test scenario
- WCAT Controller: configures client machines, runs test scenarios, creates log/report files with performance counters and WCAT runtime statistics

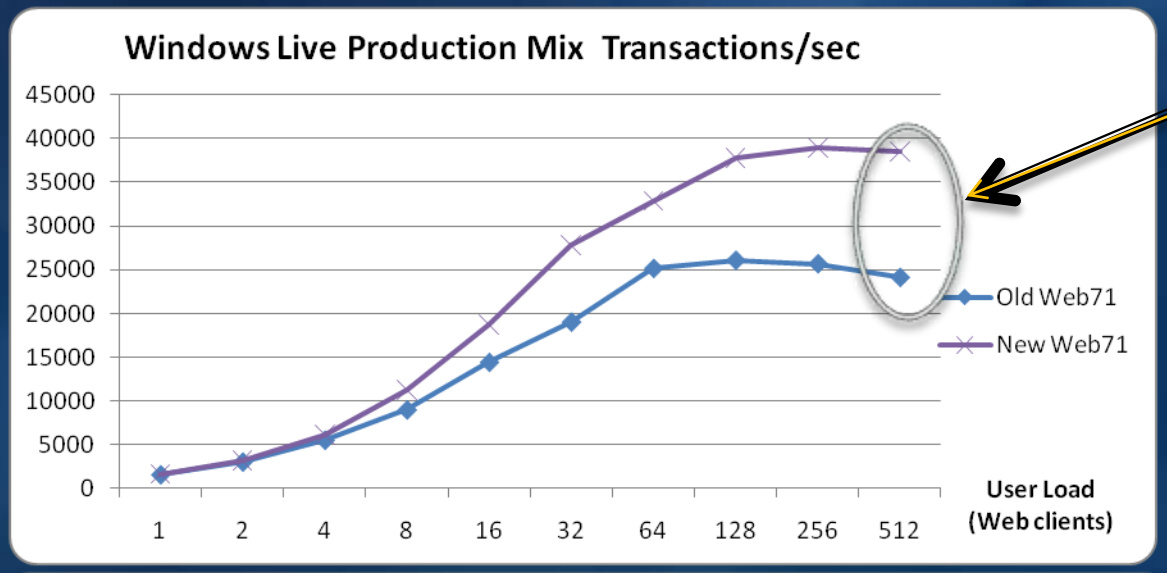
Dual Machine Environment



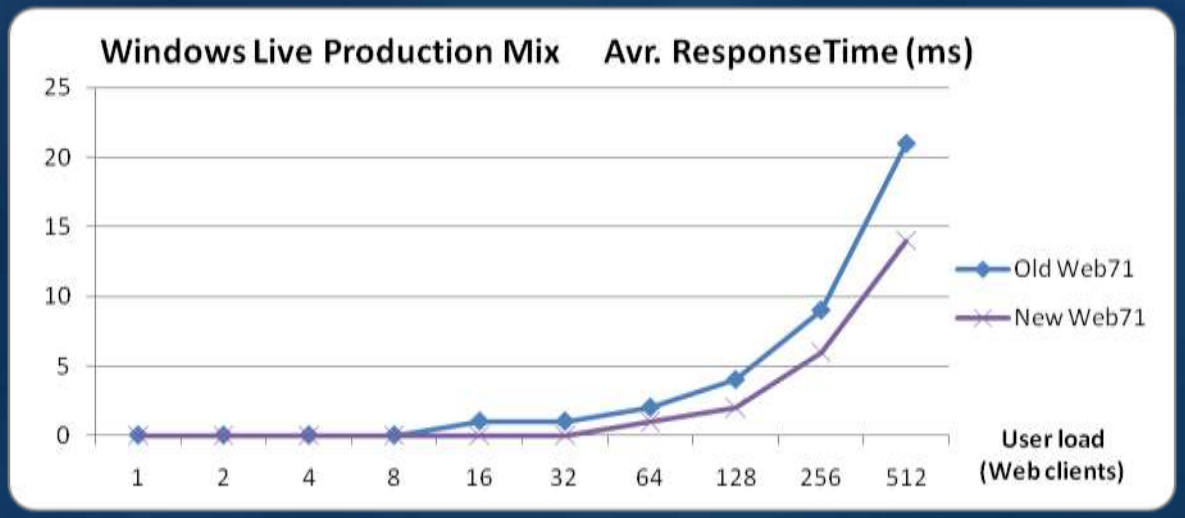
Multiple Machines Isolated Environment



WCAT Performance

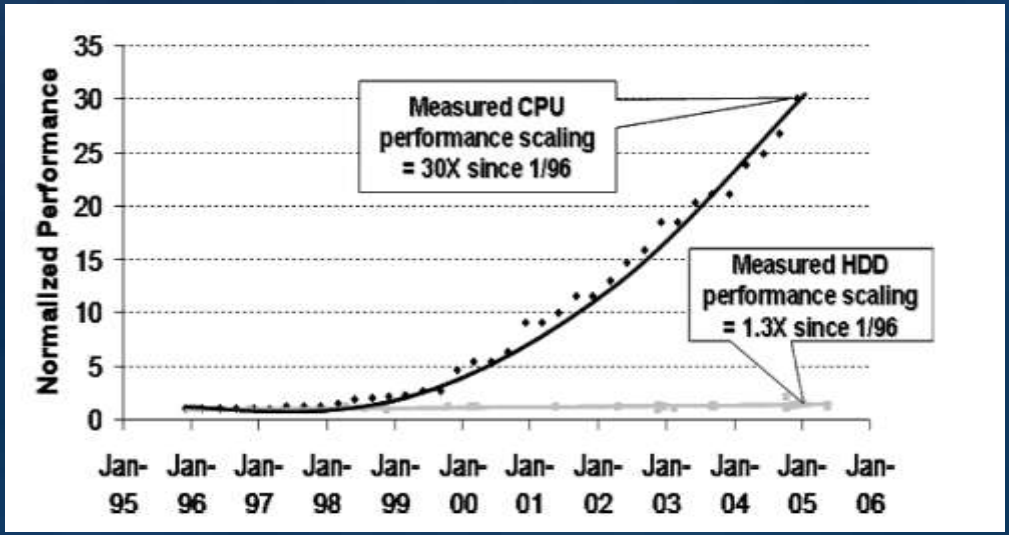


Faster CPU Increases Throughput



CPU vs Storage Performance

- Huge increase in CPU Performance
- Storage Performance not increasing proportionately
  - Need to understand storage requirements in enterprise datacenters



Source: Intel

Benchmark Development

- Most benchmarks are developed by system or component vendors
  - Goal is to showcase their products in best light
  - May not relate well to real world applications
  - Common framework for comparison
- End Users need to be more active
- Internal Benchmarks are most appropriate

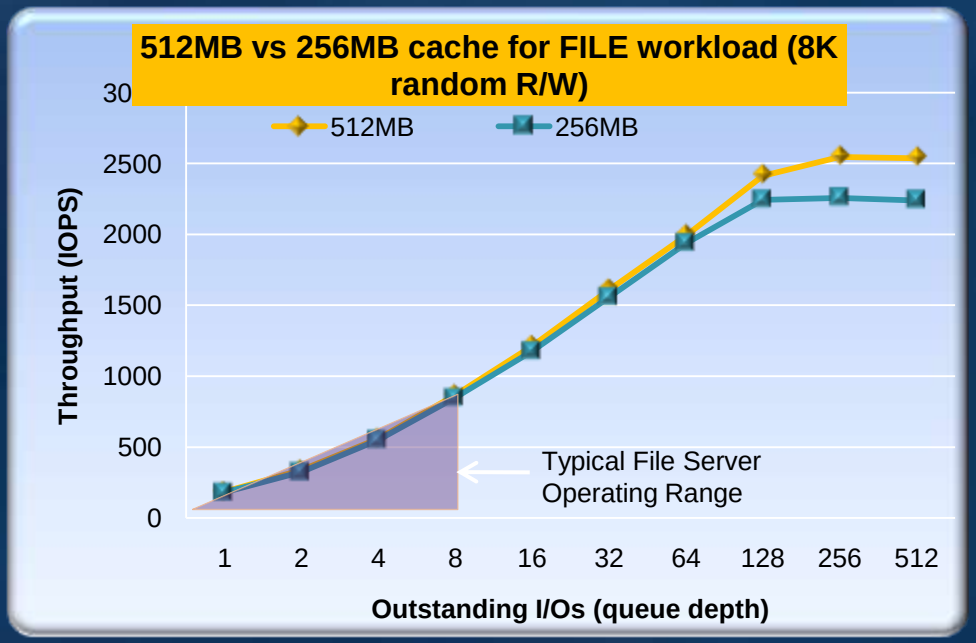
Your Mileage Will Vary!

Storage Workload Characterization

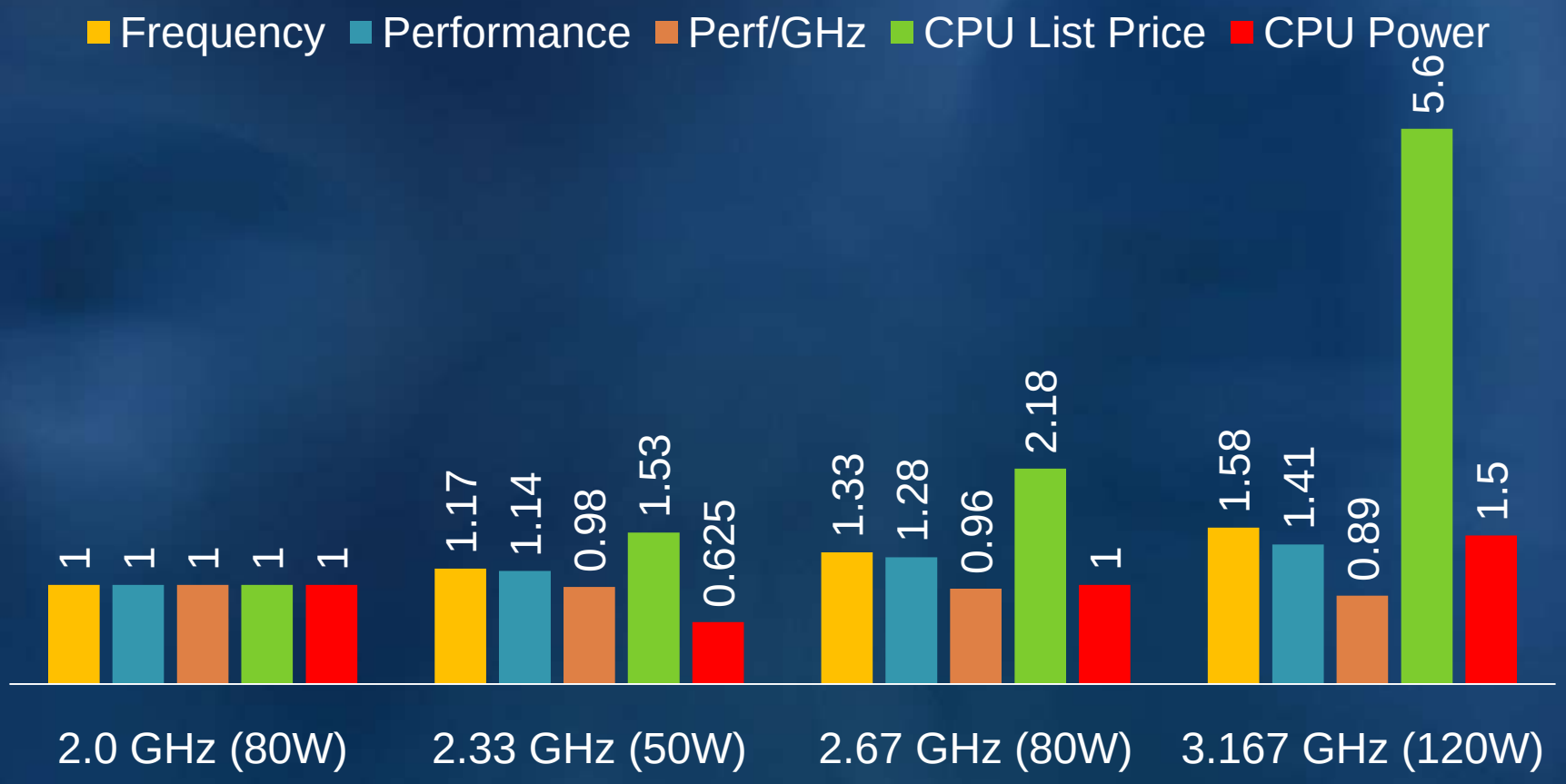
- How do we do it?
  - Event Tracing for Windows (ETW)
    - Collect disk event traces through Windows Instrumentation
    - Production traces taken for particular time periods to observe workload behavior on storage
  - Analysis
    - Summary characteristics
      - Block sizes
      - Queue depth
      - Randomness of workload
      - Read/Write patterns
      - IOPS, MBPS
    - Temporal analysis
      - Outstanding I/Os
      - Interarrival Time
      - Latency

Storage Workload Characterization

- Why do we do it?
  - Understand workload profiles in production
    - Mostly random, with high read:write ratios
  - Design next generation servers
    - Server rightsizing
      - Balancing CPU and Storage performance needs
    - Explore new technologies in the light of Microsoft workloads
      - SSD for storage – evaluate performance-power-cost advantage for enterprise profiles
    - Work with OEM partners to optimize server components for our profiles



Search Performance Scaling

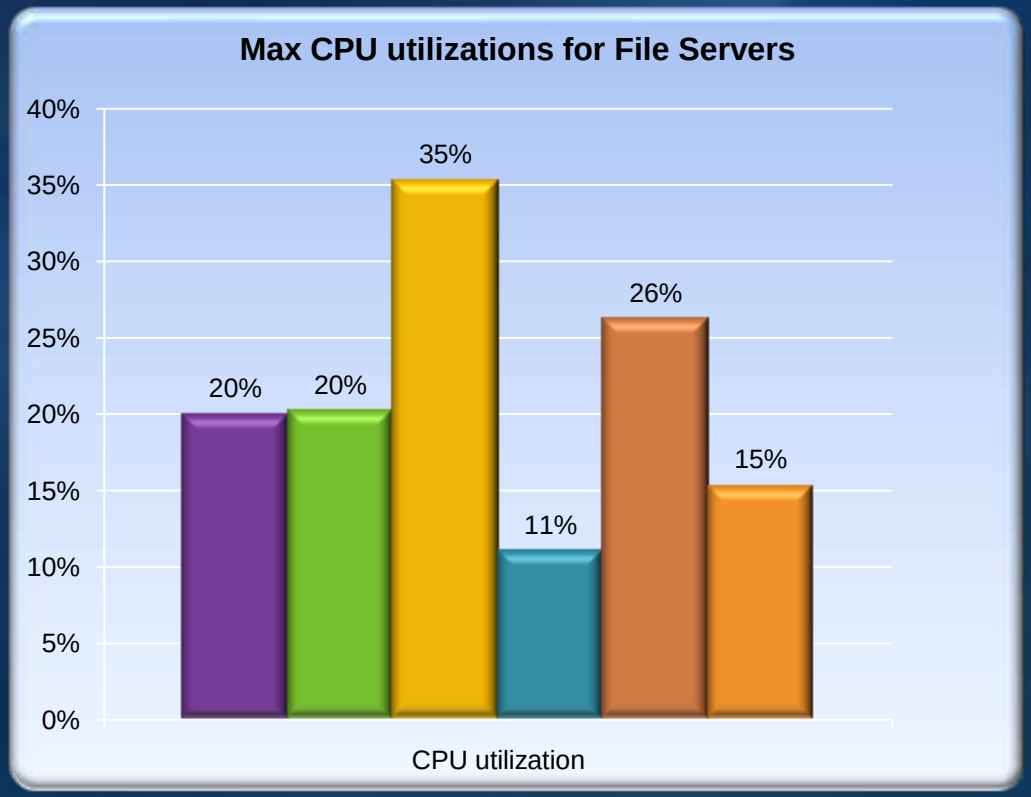


Sweet Spot is often at mid bin frequency,
especially when price and power are considered

Business As Usual

- Microprocessor Pricing History
  - Increased Performance at Constant Price
  - Multicore drove big increase going from single to dual to quad core
- Easy to ride the technology curve
  - Single -> Dual -> Quad
- Issue: Performance Scaling with Threads
- Result: Decreasing CPU Utilization
- Opportunity to Right-Size
  - Quad -> Dual -> Single

Right-Sizing Example
Dual Socket Platform



Removing the second processor would save power and cost!

Conclusions

- There is more to “performance” than speed.
- Processor performance has outpaced our ability to consume it in many cases.
- Difficult to exploit CPU performance increase across the board
- Platform imbalance is an opportunity to right-size to save power and cost.
- Power is an important “performance” metric.
- Industry Standard benchmarks may not reflect your environment.
- Do your own workload characterization.

Microsoft®

Your potential. Our passion.™