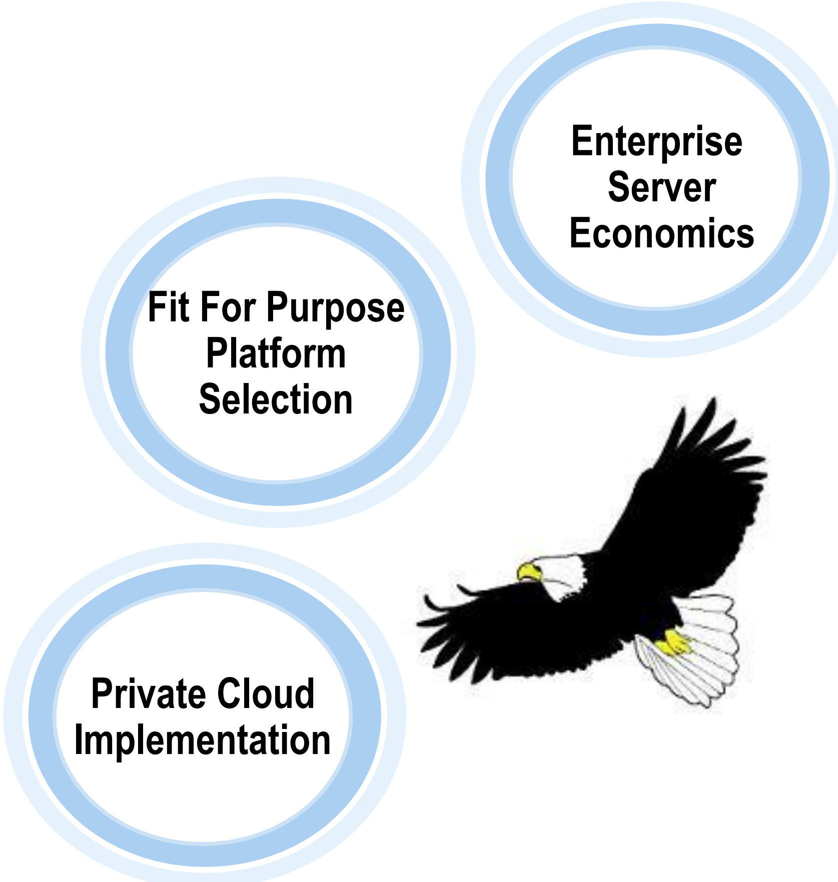
The New zEnterprise – A Cost-Busting Platform

TCO Lessons Learned, Part 1 – Establishing Equivalence

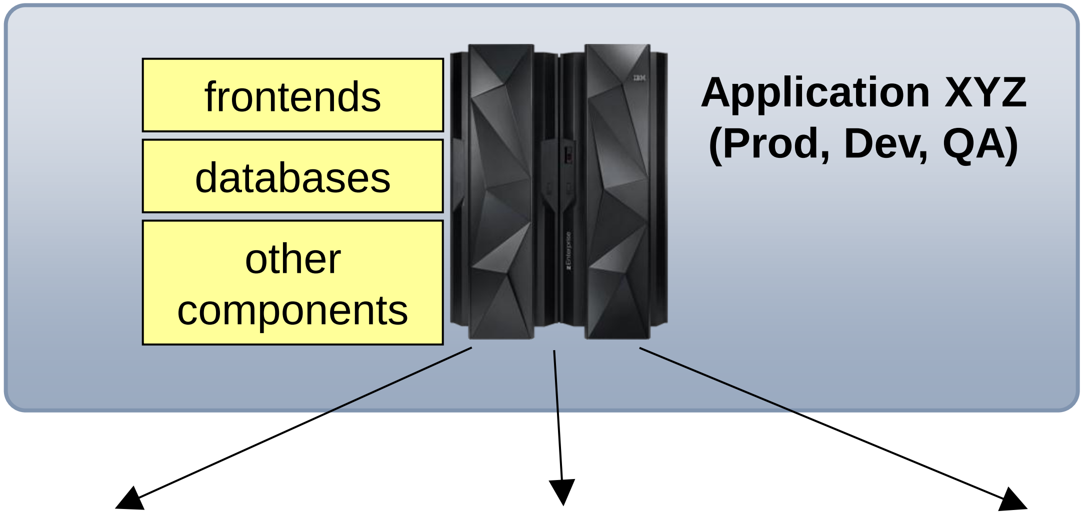
The IBM Eagle team helps customers understand mainframe costs and value

- **Worldwide** team of senior technical IT staff
- **Free of Charge** Total Cost of Ownership (TCO) studies
  - Help customers evaluate the lowest cost option among alternative approaches
  - Includes a one day on-site visit and is **specifically tailored to a customer's enterprise**
- Studies cover POWER, PureSystems and Storage accounts in addition to System z
  - For both IBM customer and Business Partner customer accounts
- Over 300 customer studies since formation in 2007
- Contact: eagletco@us.ibm.com



What happens in a TCO study?

Workload identified for analysis



Deployment Choices

- **Do nothing**
- **Optimize current environment**
- **Deploy on other platforms**

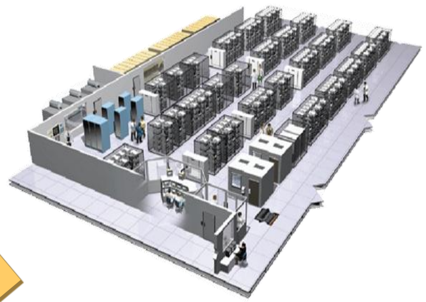
Key steps in analysis

1. Establish equivalent configurations
- Needed to deliver workload

2. Compare Total Cost of Ownership
- TCO looks at different dimensions of cost

How can we determine equivalent configurations?

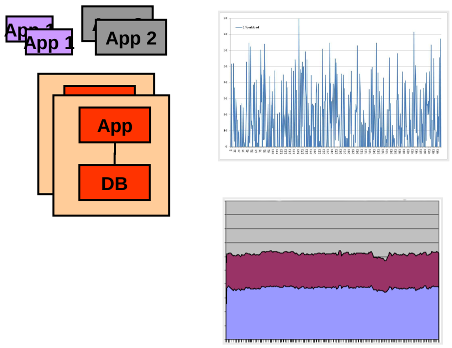
Real world aspects determine accurate equivalence



Top Down approach

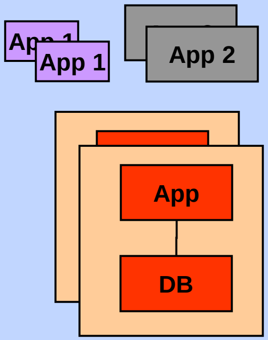
What we see in customer environments

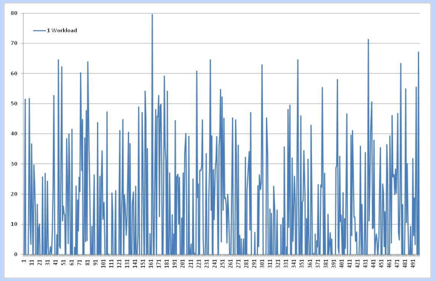
Bottoms up approach

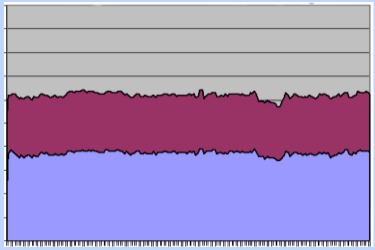


What we know about platforms and measure in atomic benchmarks

Platform differences and atomic benchmarks set a baseline for establishing equivalence







Platform factors

GHz, cache, I/O, co-location

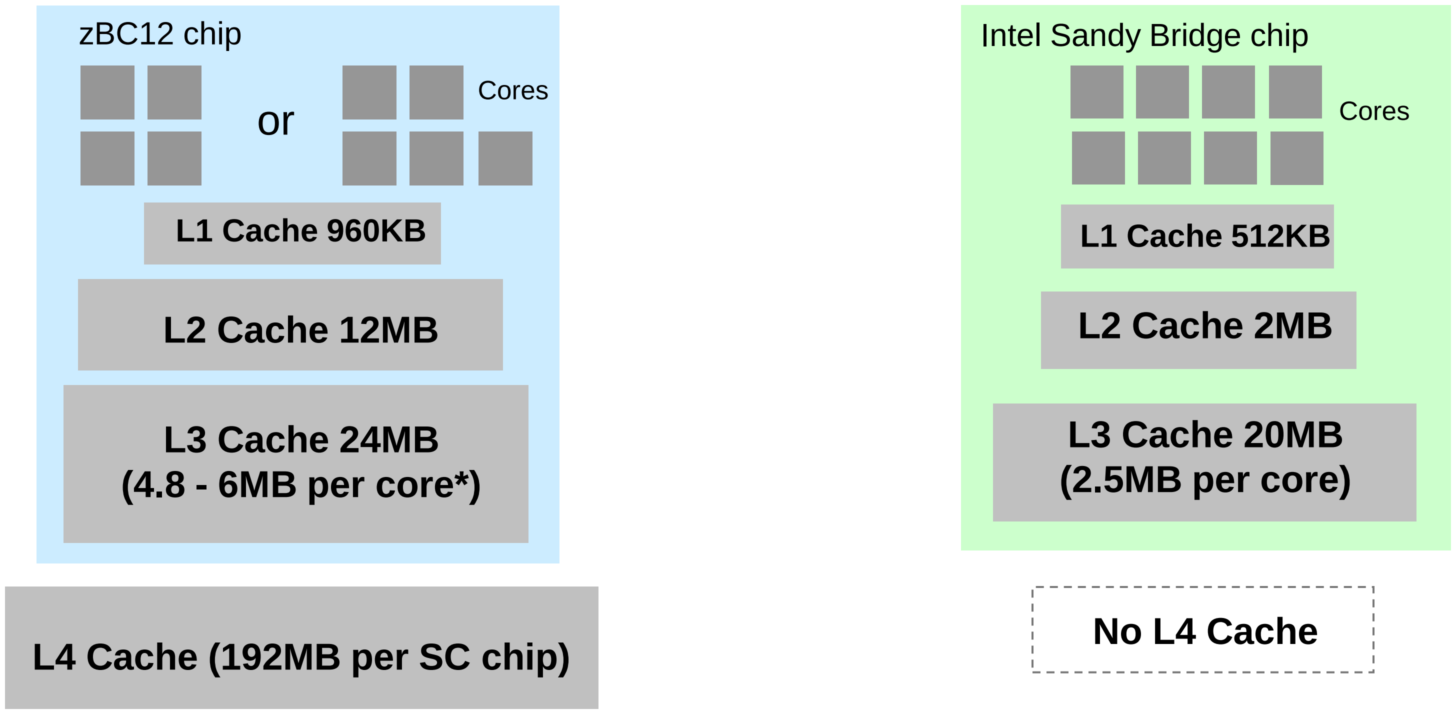
Variability in demand

Different size servers

Workload Management

Mix workloads with different priorities

Like zEC12, new zBC12 has larger cache structures to support more concurrent workloads



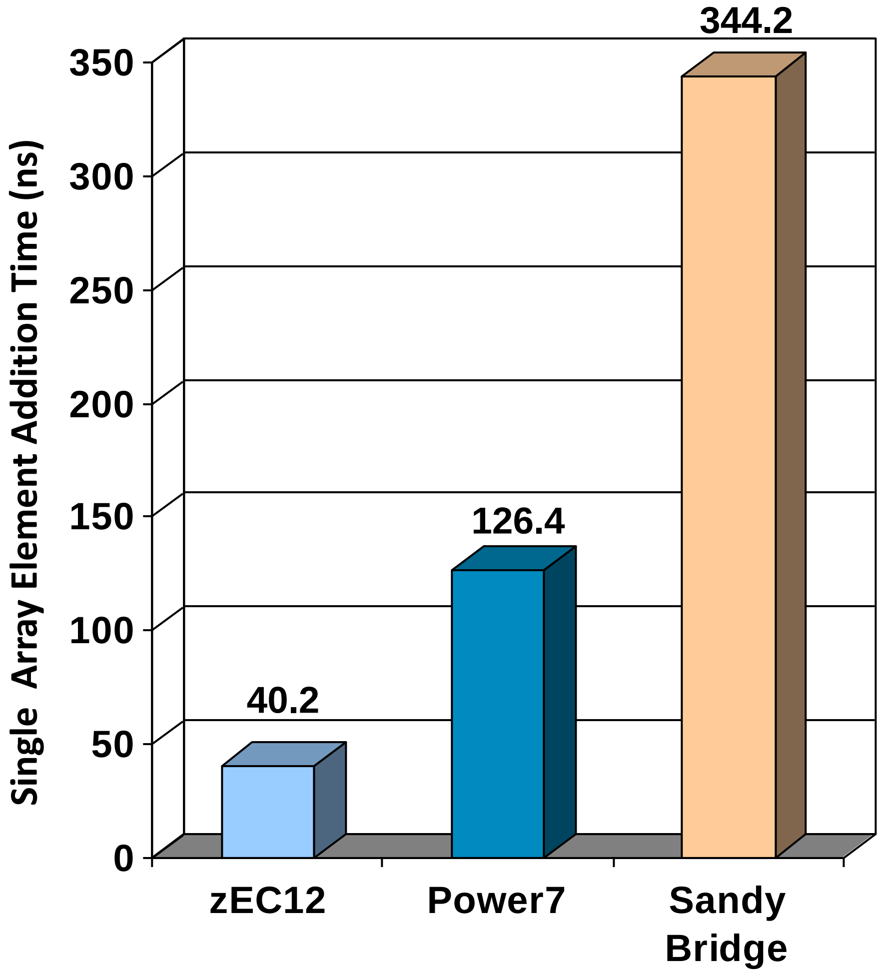
Advantages of large cache:

- Fewer cache misses help maintain thread processing speed
- Improves database performance by holding larger working sets
- Improves consolidated workload performance by supporting more working sets

* Six core PU chips using 4 and 5 active core per PU chip. 4.8 MB L3 cache if 5 active core per chip. 6MB L3 cache if 4 active core per chip.

Intel servers slow down under cache intensive workloads

- Multiple concurrent processes introducescache contention
  - Example: 5 processes each with 70MB working set size
- Intel workloads significantly slowed due to cache contention
- System z with z/OS showed results 8x faster than Intel system



Larger cache is beneficial for SAP workloads – as well as CICS, VSAM and Batch workloads

Cost advantage for smaller scale SAP database:

SQL Server on Intel

SAP Applications

4 x HP DL980 2.13GHz 4ch/32co

128 DB cores

SAP Applications

DB2 on z/OS

zEC12 with 3 GP + 2 zIIPs

5 cores

***29%** lower unit cost*

Database Unit Cost
$86/User

# of Users	23,000
Hardware	$0.34M
Software	$1.64M
Total (3 yr. TCA)	$1.98M

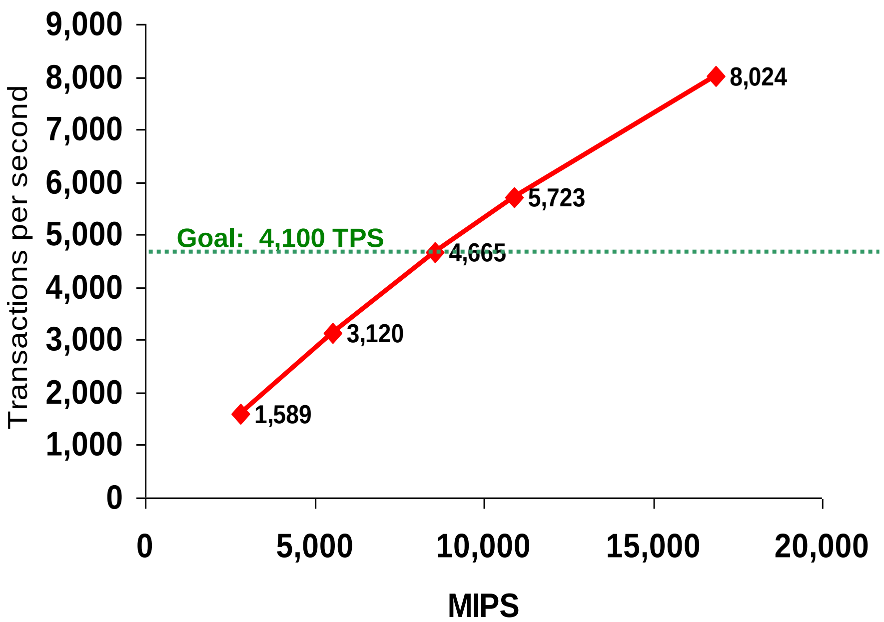
Database Unit Cost
$61/User

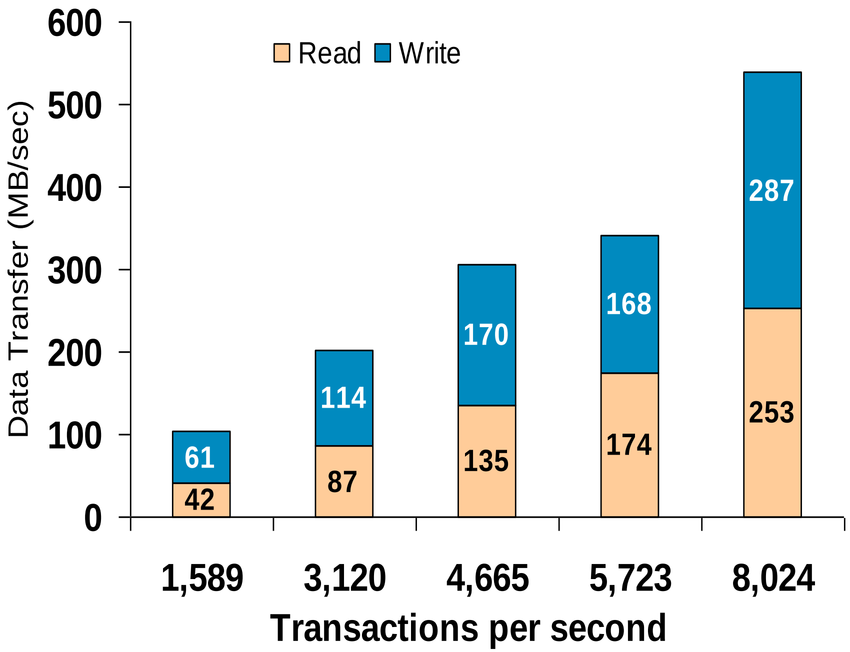
# of Users	23,000
DB2 Solution Edition(HW+SW)	$1.40M
Total (3 yr. TCA)	$1.40M

Note: Workload Equivalence established from a large US Retailer SAP DB offload incorporating estimated CPU Savings from DB2 for z/OS upgrade (107 Performance Units per MIPS). Upgrading from DB2 V8 to V10 reduces average CPU usage by 28%. DB2 V10 for z/OS on zEC12 and SQL Server 2008 on Intel

Dedicated I/O subsystem means System z is ideal for high bandwidth workloads

Capacity benchmark for Bank of China:



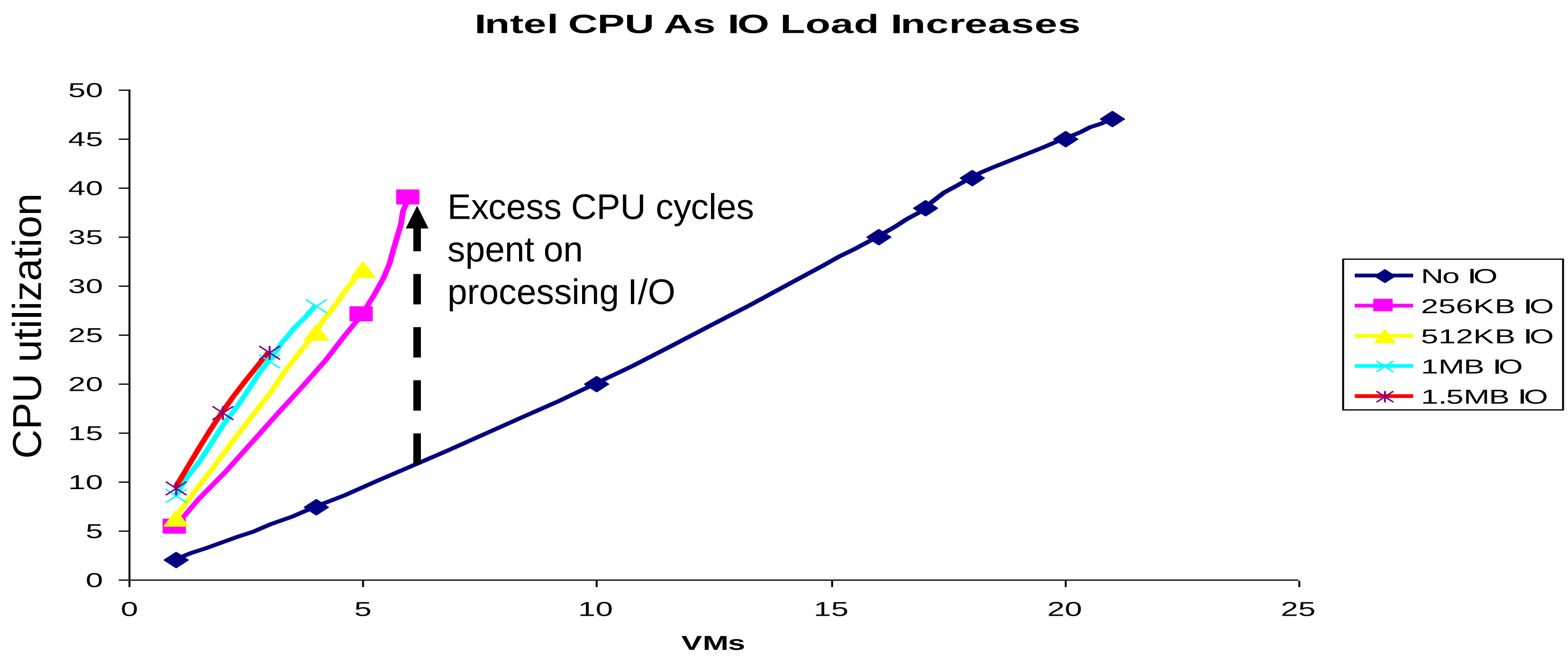


System z easily surpassed benchmark goal, and demonstrates near linear scalability

Reads and writes are well-balanced and scale linearly, demonstrating no constraints on I/O constraint

Tests show Intel's performance degrades as I/O demand increases

- Test case scenario: Run multiple virtual machines on x86 server
  - Each virtual machine has an average I/O rate
  - x86 processor utilization is consumed as I/O rate increases
- With no dedicated I/O subsystem, Intel's performance degrades



Multi-tenant database testing also demonstrates System z's superior ability to handle I/O load

Which platform can achieve the lowest cost per workload?

I/O Intensive Database Workload

Brokerage high volume trading workload, each driving a minimum* of **243** transactions per second on 200GB database

1 workload on 16-core quarter unit

Pre-integrated DB Competitor V2 Multi-Tenant Private Cloud

$2.27M/workload

5 multi-tenant workloads

on zEC12

2 GPs + 2 zIIPs

DB2 10 for z/OS on zEC12

$1.73M/workload

25%
lower cost!

* Maximum TPS was measured at 270 based on 70 ms injection interval for customer threads. SLA requires no more than 10% degradation in throughput, yielding a Minimum TPS of 243

z/OS database workloads benefit from higher I/O bandwidth

Competitor DB on Intel

SAP Applications

8x 3850 x5 with 32 cores (dual active clusters)

128 DB cores

SAP Applications

DB2 on z/OS

zEC12 2-way data sharing Sysplex

44 DB cores

***41%** more postings at ½ cost!*

Database Unit Cost
$0.30/Postings per hour

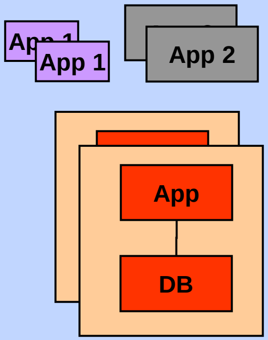
Postings per Hour	42.0M
# of Accounts	90M
Hardware	$0.63M
Software	$12.0M
Total (5 yr. TCA)	$12.6M

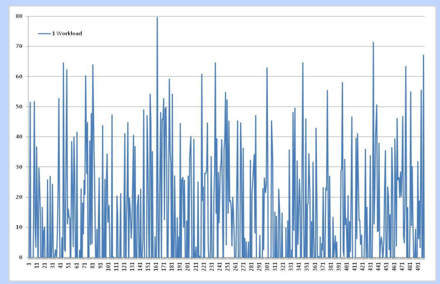
Database Unit Cost
$0.15/Postings per hour

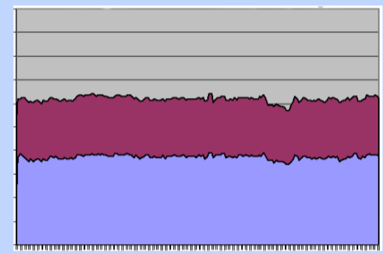
Postings per Hour	59.1M
# of Accounts	150M
DB2 Solution Edition (HW+SW)	$7.49M
Capacity Backup (CBU)	$1.24M
Total (5 yr. TCA)	$8.73M

Cost of platform infrastructure for comparative transaction production.
Cost of packaged application software not included. List prices used.

Platform differences and atomic benchmarks set a baseline for establishing equivalence







Platform factors

GHz, cache, I/O, co-location

Variability in demand

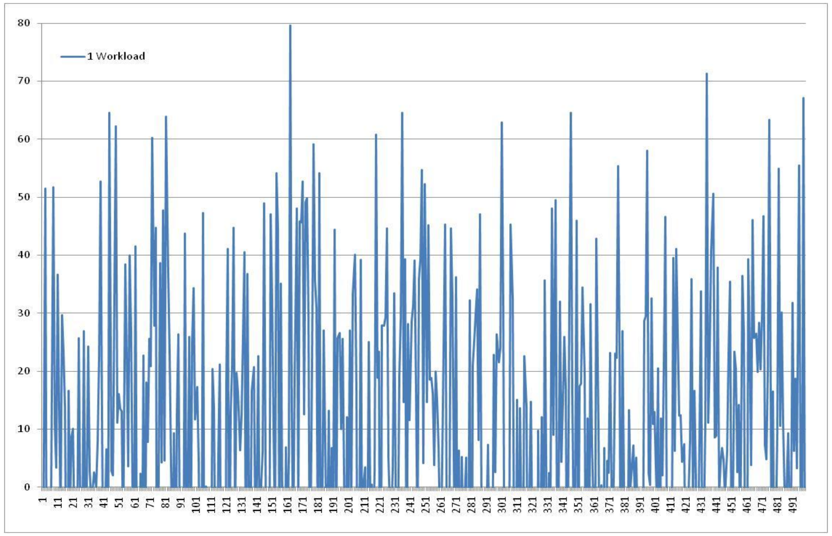
Different size servers

Workload Management

Mix workloads with different priorities

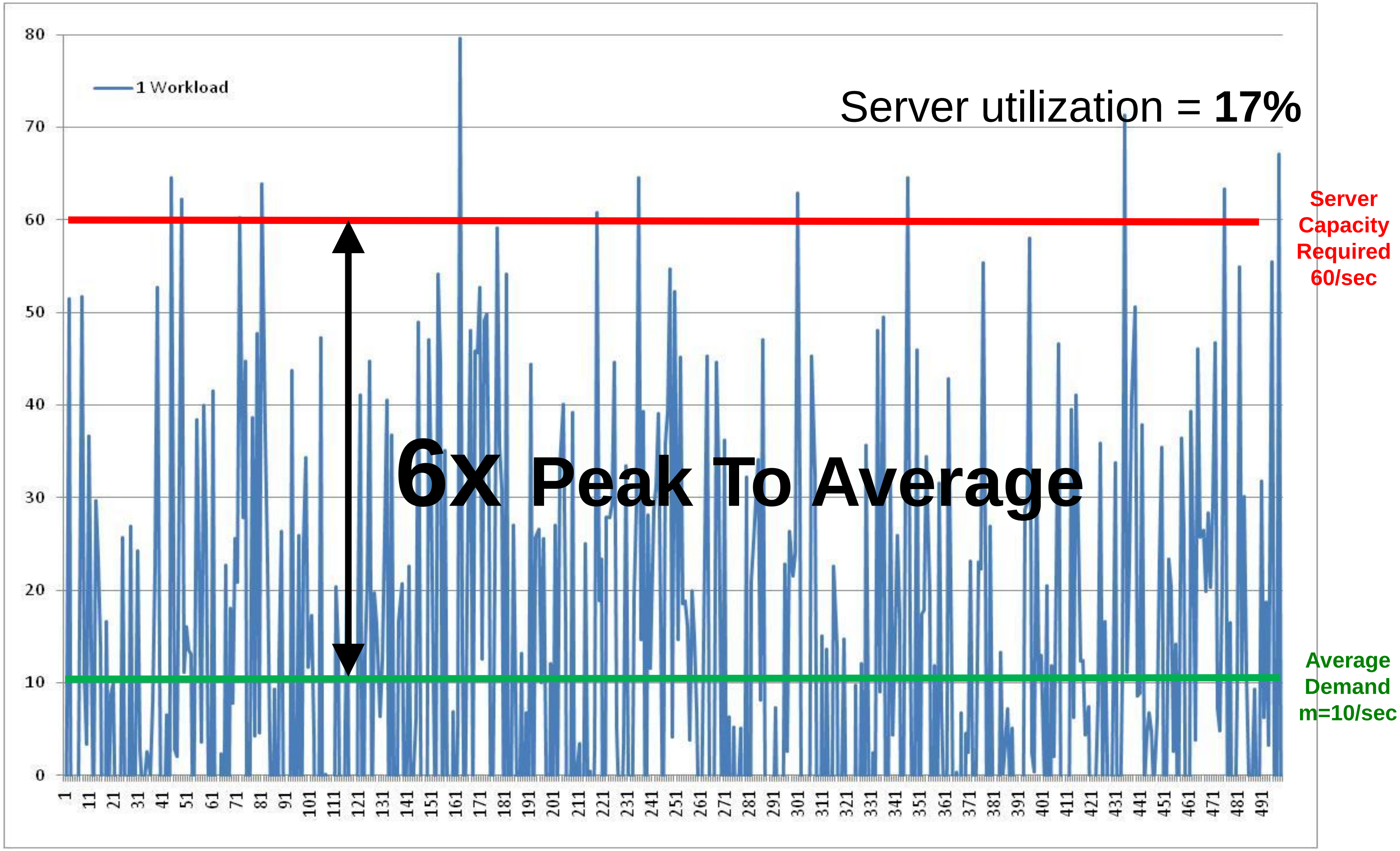
Larger servers with more resources make more effective consolidation platforms

- Most workloads experience variance in demand



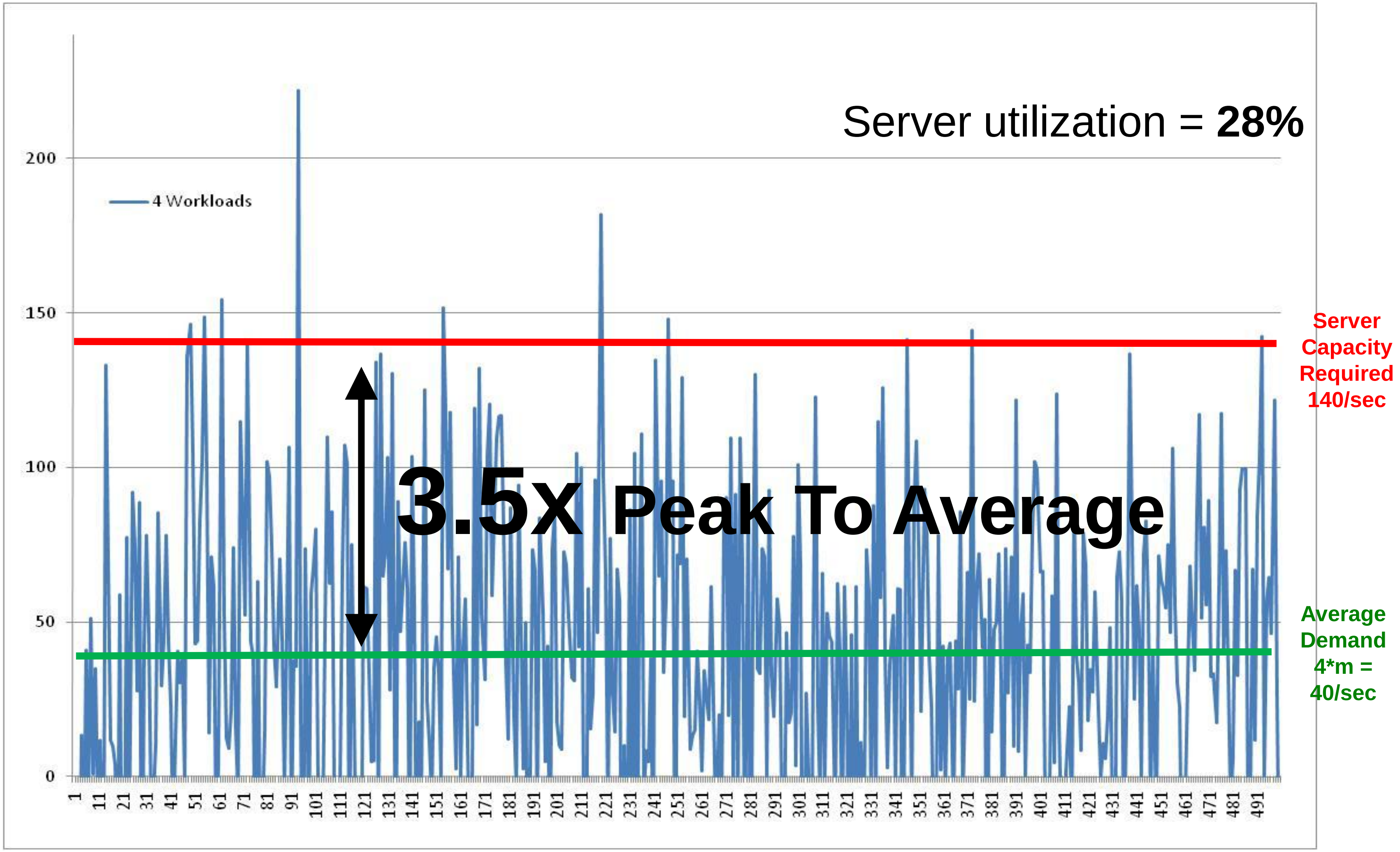
- When you consolidate workloads with variance on a virtualized server, the variance of the sum is less (statistical multiplexing)
- The more workloads you can consolidate, the smaller is the variance of the sum
- Consequently, bigger servers with capacity to run more workloads can be driven to higher average utilization levels without violating service level agreements, thereby reducing the cost per workload

A single workload requires a machine capacity of 6x the average demand



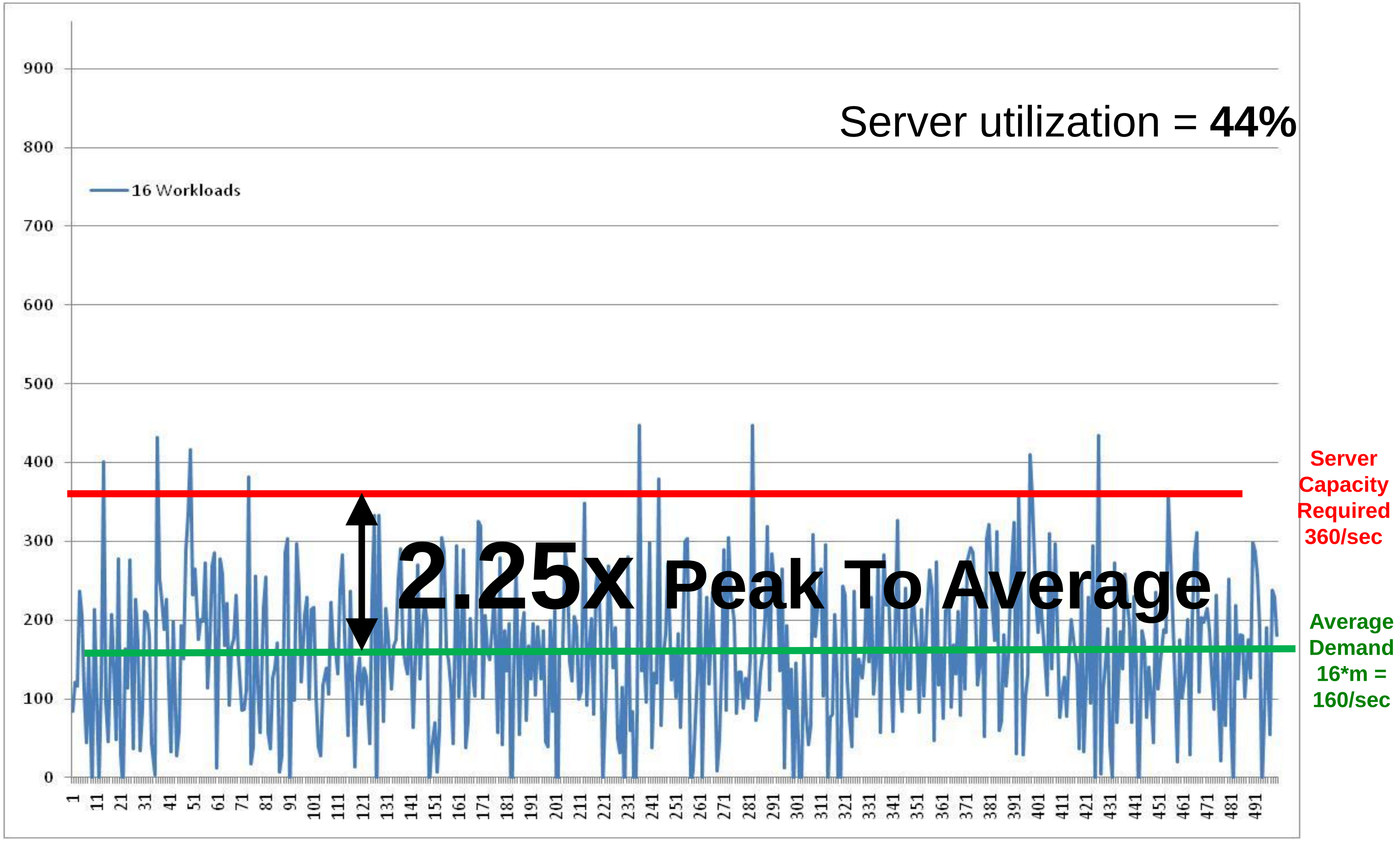
Assumes coefficient of variation = 2.5, required to meet 97.7% SLA

Consolidation of 4 workloads requires server capacity of 3.5x average demand



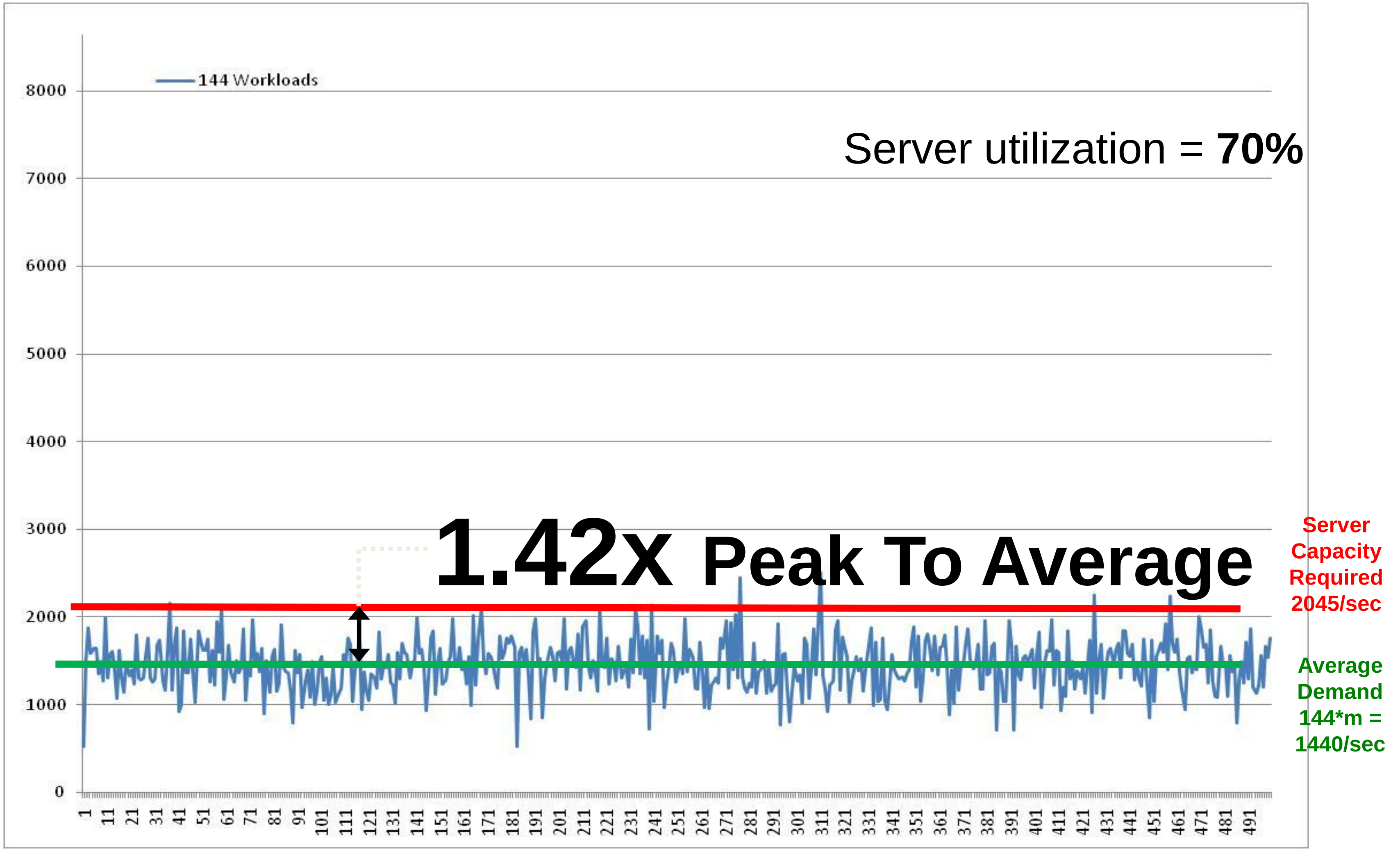
Assumes coefficient of variation = 2.5, required to meet 97.7% SLA

Consolidation of 16 workloads requires server capacity of 2.25x average demand



Assumes coefficient of variation = 2.5, required to meet 97.7% SLA

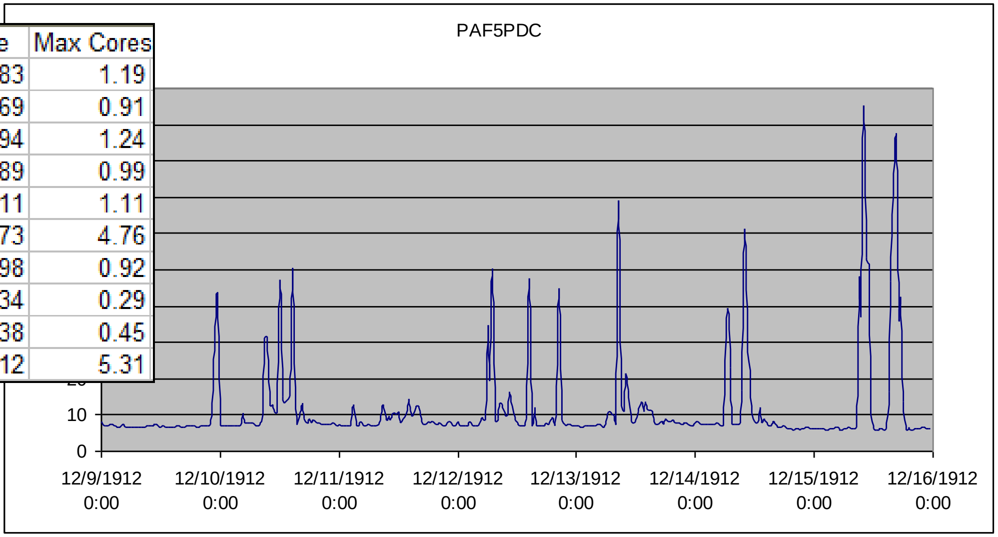
Consolidation of 144 workloads requires server capacity of 1.42x average demand



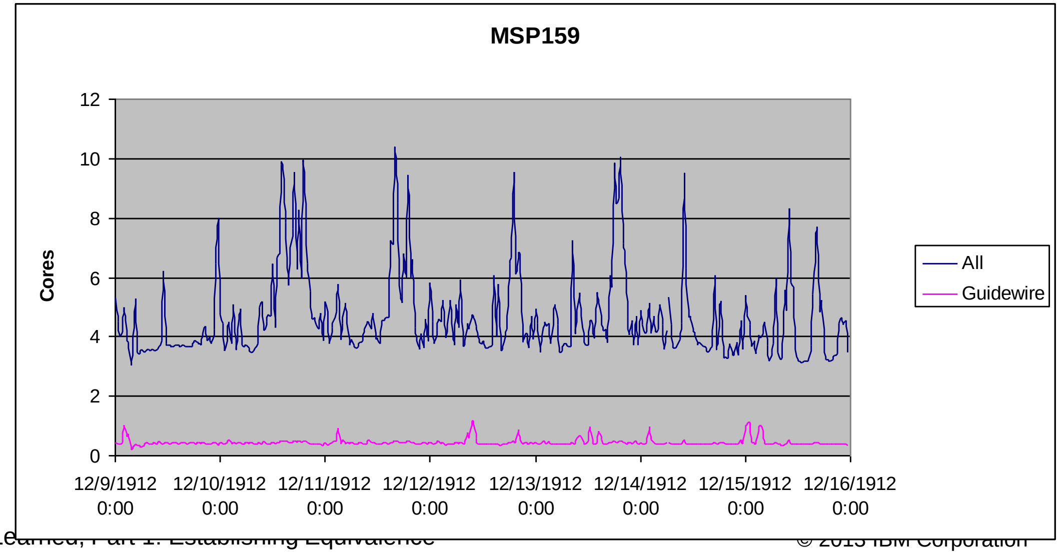
Assumes coefficient of variation = 2.5, required to meet 97.7% SLA

Actual data from a POWER customer demonstrates how statistical multiplexing applies to all large scale virtualization platforms

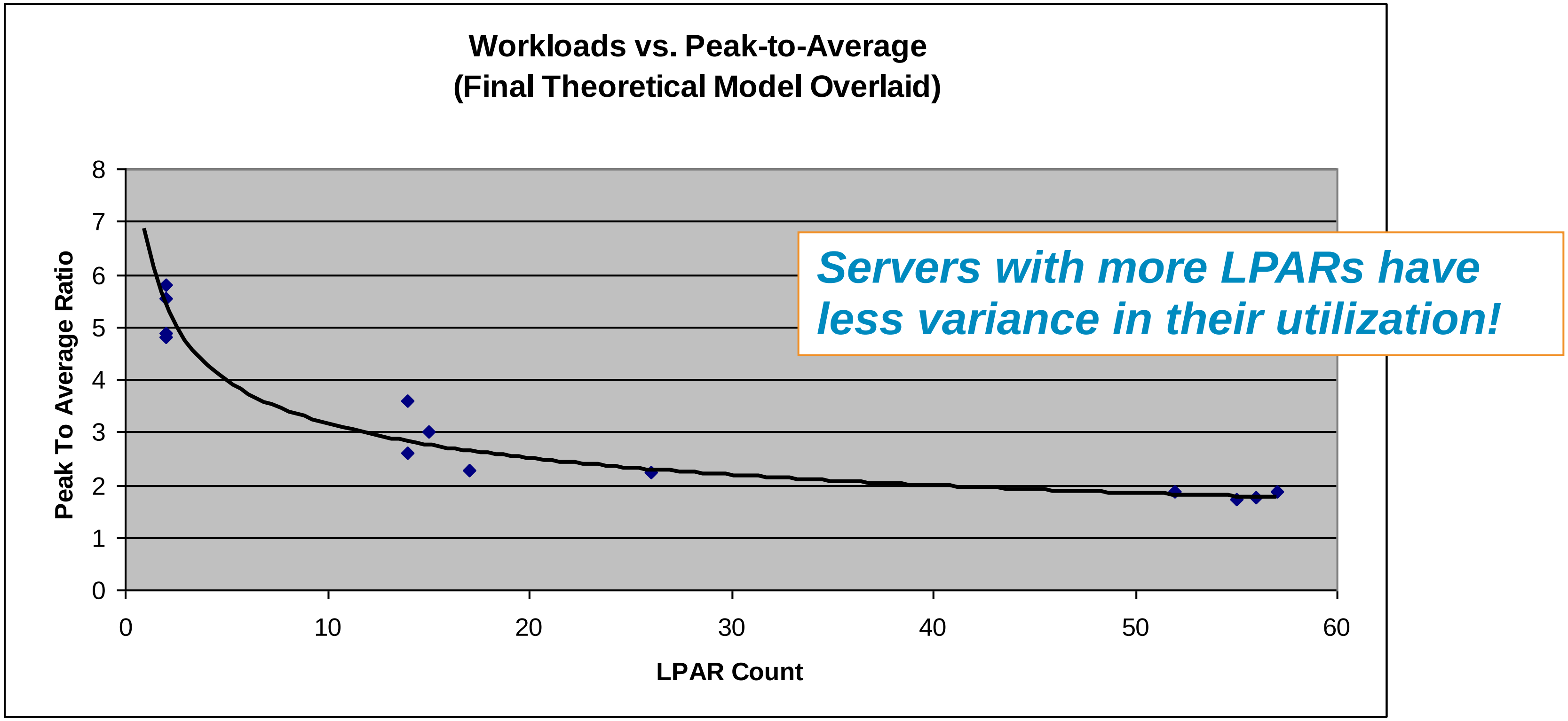
Frame	LPAR	Min	Max	Std. Dev.	Average	Variance	Max Cores
MSP159	PA3APDC	10.44	59.57	6.46	22.37	0.83	1.19
MSP159	PC2APDC	14.40	45.29	5.19	19.11	0.69	0.91
MSP159	PC18PDC	10.36	41.48	5.19	14.45	0.94	1.24
MSP159	PB5BPDC	9.49	32.92	3.23	11.83	0.89	0.99
MSP159	PB4EPDC	9.26	37.16	3.54	11.57	1.11	1.11
MSP159	PAF5PDC	6.00	95.27	11.78	11.25	3.73	4.76
MSP159	PFE2PDC	4.43	46.23	6.63	9.33	1.98	0.92
MSP159	PB3EPDC	7.83	14.31	0.60	8.53	0.34	0.29
MSP159	MSP159VIO2	4.33	14.95	1.86	8.51	0.38	0.45
MSP159	PCB1PDC	0.79	88.48	17.73	7.88	5.12	5.31



- Large US insurance company
- 13 production POWER7 frames
  - Some large servers, some small servers
- Detailed CPU utilization data
  - 30 minute intervals, one whole week
  - For each LPAR on the frame
  - For each frame in the data center
- Measure peak, average, variance

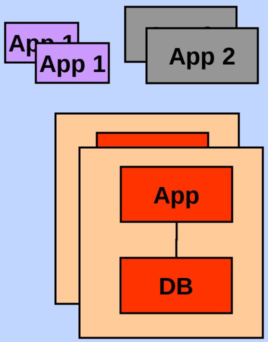


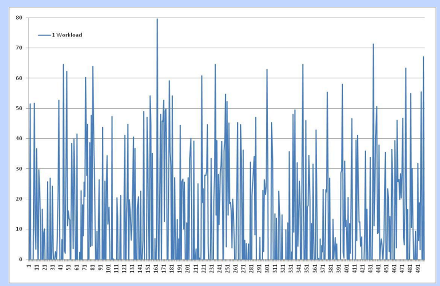
Customer data confirms statistical multiplexing theory

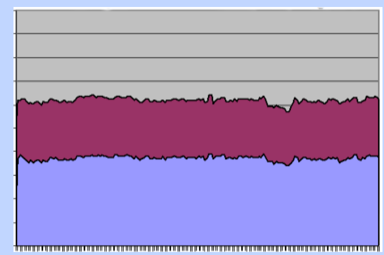


- The larger the shared processor pool, the greater the statistical benefit
- Large scale virtualization platforms are able to consolidate large numbers of virtual machines because of this
- Servers with capacity to run more workloads can be driven to higher average utilization levels without violating service level agreements

Platform differences and atomic benchmarks set a baseline for establishing equivalence







Platform factors

GHz, cache, I/O, co-location

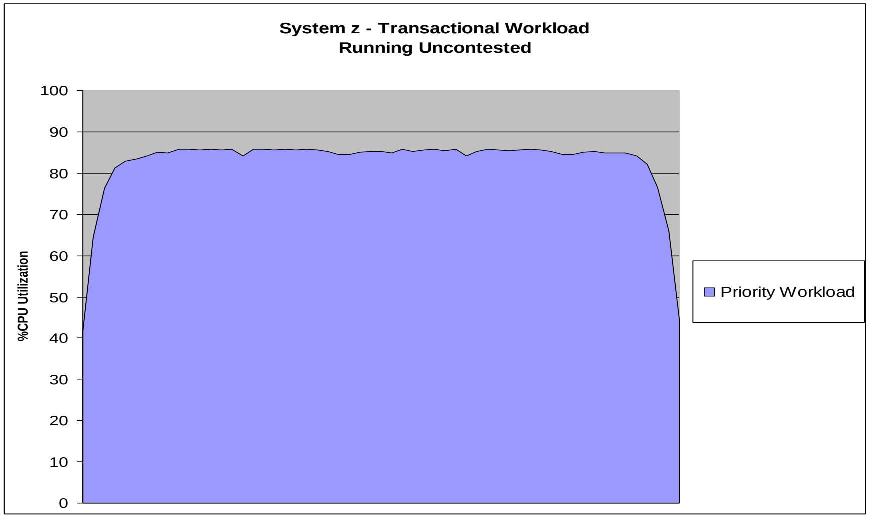
Variability in demand

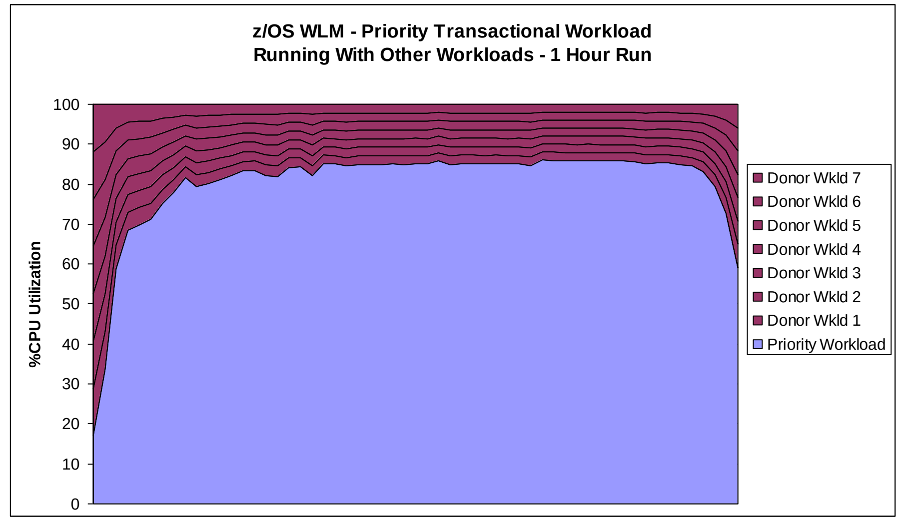
Different size servers

Workload Management

Mix workloads with different priorities

Priority transactional workload does not degrade when low priority workloads added





Capacity Used
High Priority Steady State - 85.2% CPU Minutes
Unused (wasted) - 14.8% CPU Minutes

Capacity Used
High Priority Steady State - 85.3% CPU Minutes
Unused (wasted) - 0% CPU Minutes

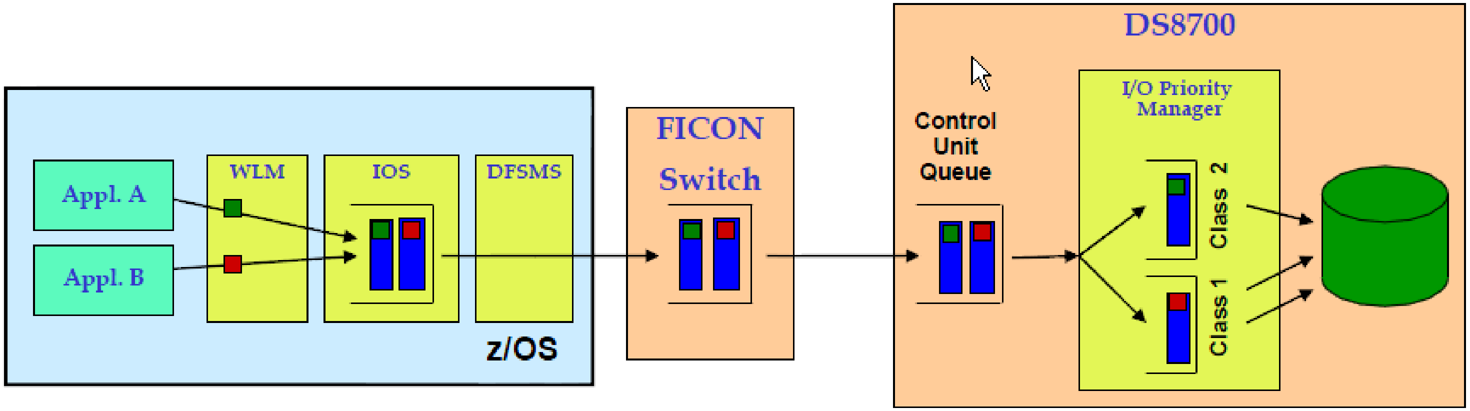
Priority Workload Metrics
Total Throughput: 417.8K
Maximum TPS 129.7

Priority Workload Metrics
Total Throughput: 414.7K
Maximum TPS 128.1

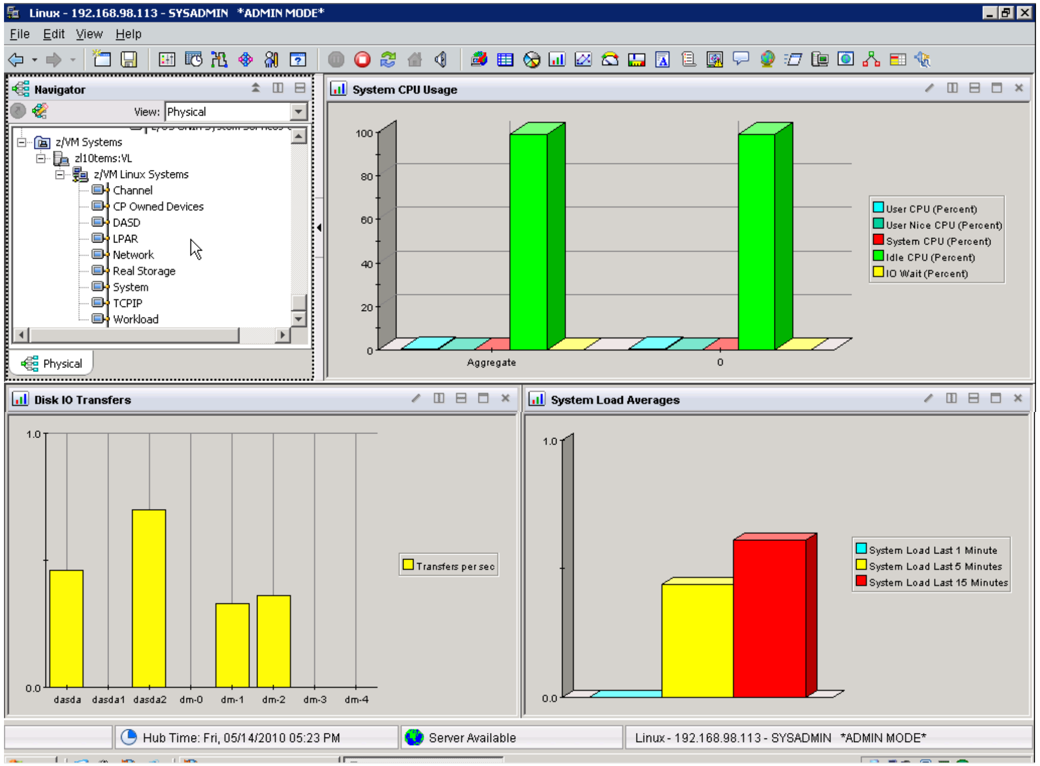
NO *steady state CPU usage leakage* ***1%*** *total transaction leakage*

z/OS Workload Manager (WLM) extends priority all the way down to storage

- FICON protocol supports advanced storage connectivity features not found in x86
- Priority Queuing:
  - Priority of the low-priority programs will be increased to prevent high-priority channel programs from dominating lower priority ones

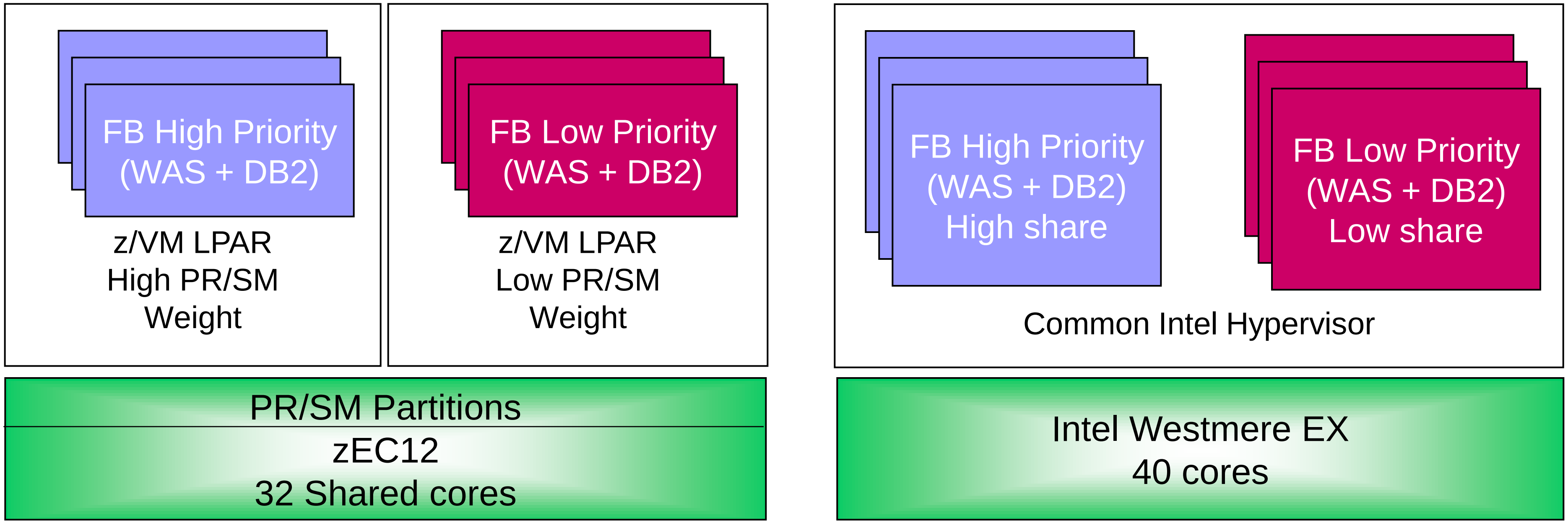


DEMO: z/OS Workload Manager

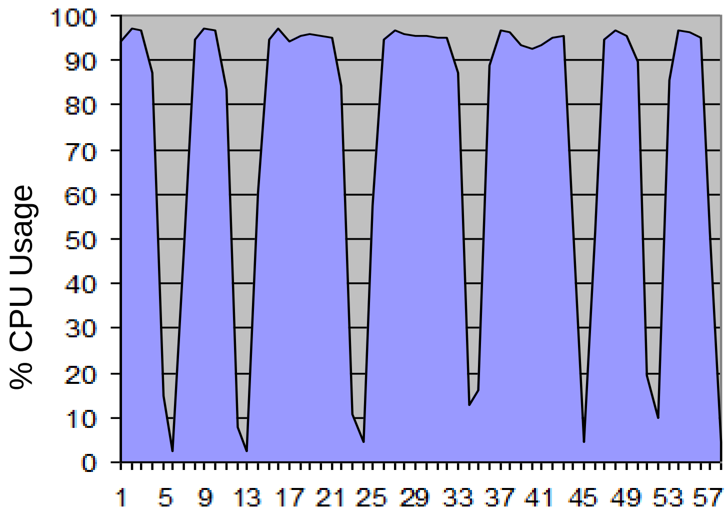


Tests demonstrate comparison of System z PR/SM virtualization to a common hypervisor

- High Priority web workload has defined demand over time
- SLA requires that response time does not degrade
- Low Priority web workload has unlimited demand
- It “soaks up” unused CPU minutes

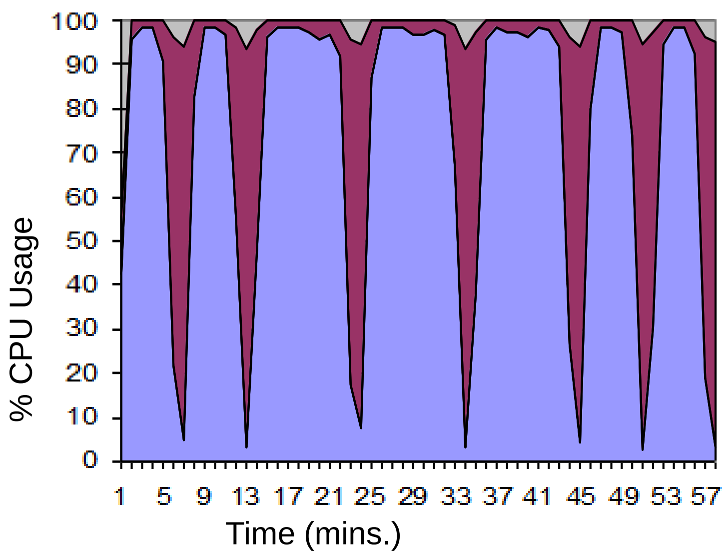


System z demonstrates perfect workload management…



Demand curve for 10 high priority workloads running in 1 z/VM LPAR (PR/SM weight = 99)

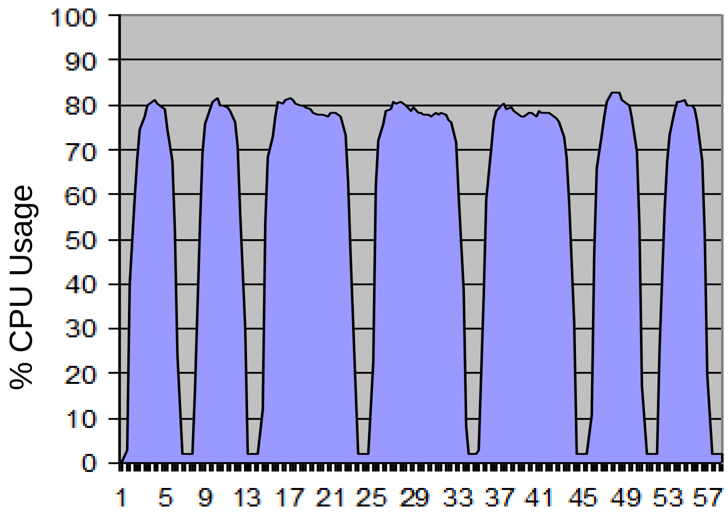
- **Workloads consume 72% of available CPU resources (28% unused)**
- **Total throughput: 9.13M**
- **Average response time: 140ms**



Demand curve when 14 low priority (PR/SM weight = 1) workloads are added in a second z/VM LPAR

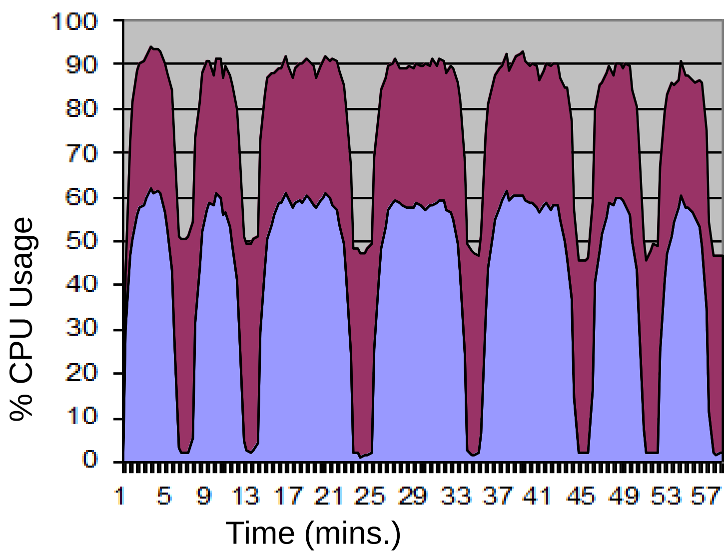
- **All but 2% of available CPU resources is used (high=74%, low=24%)**
- **High priority workload throughput is maintained (9.13M)**
- **No response time degradation (140ms)**

…Unlike this common Intel hypervisor which demonstrates imperfect workload management



Demand curve for 10 high priority workloads running on a common Intel hypervisor (high share)

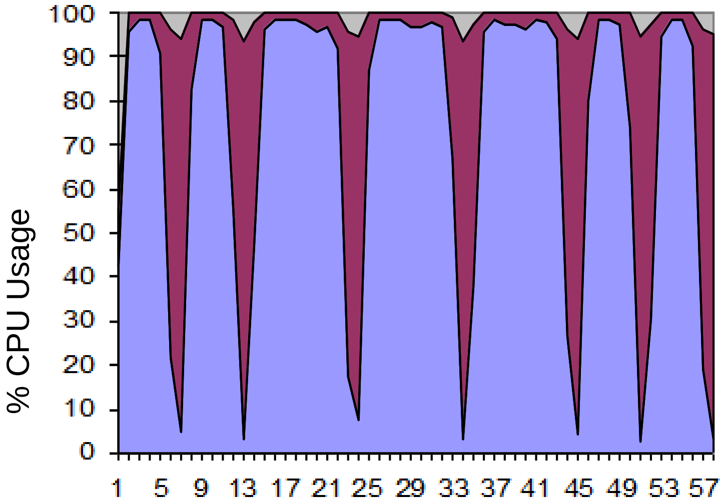
- **Workloads consume 58% of available CPU resources (42% unused)**
- **Total throughput: 6.47M**
- **Average response time: 153ms**



Demand curve when 14 low priority (low share) workloads are added

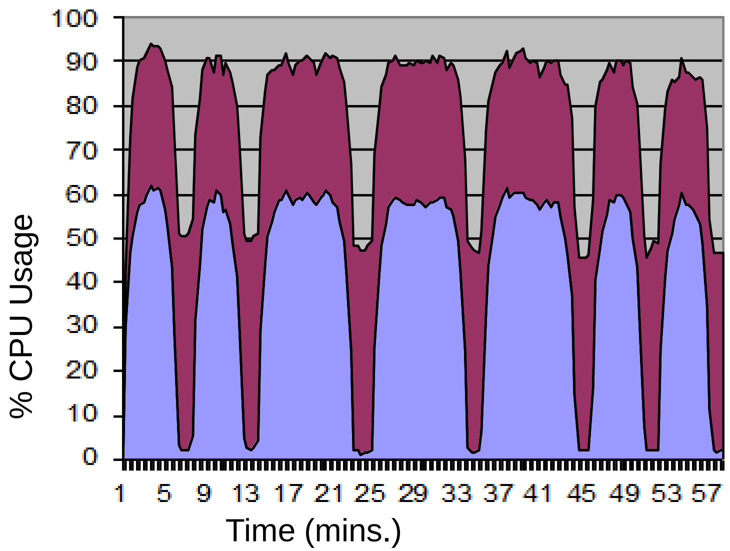
- **22% of available CPU resources is unused (high=42%, low=36%)**
- **High priority workload throughput drops 31% (4.48M)**
- **Response time degrades 45% (220ms)**

System z virtualization enables mixing of high and low priority workloads without penalty



System z

- Perfect workload management
- Consolidate workloads of different priorities on the same platform
- Full use of available processing resource (high utilization)

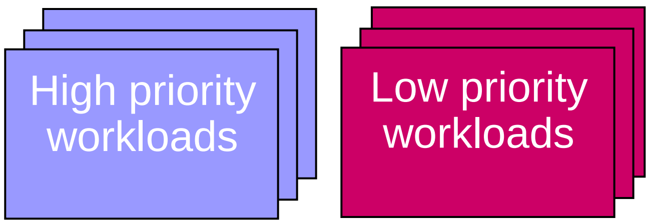


Common Intel hypervisor

- Imperfect workload management
- Forces workloads to be segregated on different servers
- More servers are required (low utilization)

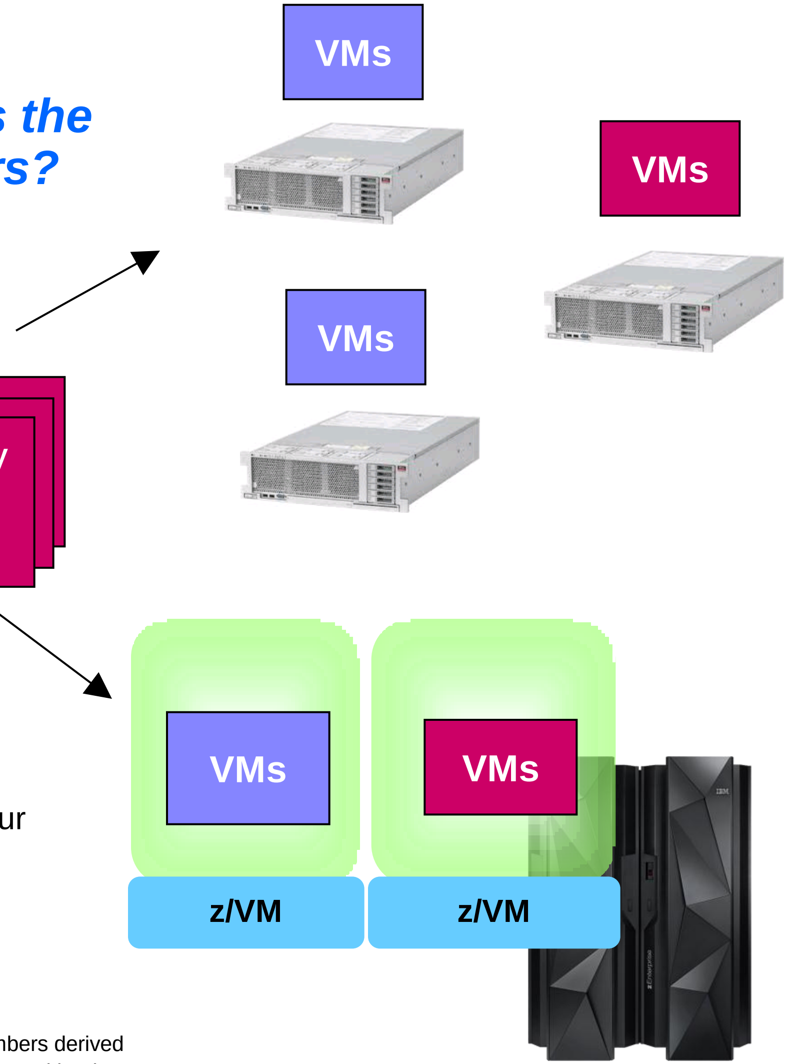
Imperfect workload management leads to core proliferation and higher costs

Which platform provides the lowest TCA over 3 years?



- **IBM WebSphere 8.5 ND**
- **IBM DB2 10 AESE**
- **Monitoring software**

High priority online banking workloads driving a total of **9.1M** transactions per hour and low priority discretionary workloads driving **2.8M** transactions per hour



Virtualized on 3 Intel 40 core servers

$13.7M (3 yr. TCA)

z/VM on zEC12

32 IFLs

$5.77M (3 yr. TCA)

58% lower cost!

Consolidation ratios derived from IBM internal studies.. zEC12 numbers derived from measurements on z196. Results may vary based on customer workload profiles/characteristics. Prices will vary by country.

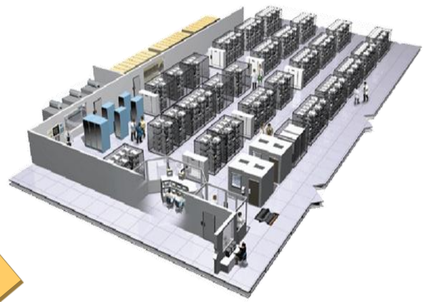
System z supports concurrent operations during hardware repair

Capability	zEC12	x86
ECC on Memory Control Circuitry	Transparent While Running	Can recognize/repair soft errors while running; limited ability with hard errors
Oscillator Failure	Transparent While Running	Must bring server down to replace
Core Sparing	Transparent While Running	Must bring server down to replace
Microcode Driver Updates	While Running	Some OS-level drivers can update while running, not firmware drivers; reboot often required
Book Additions, Replacement	While Running	Must bring server down to replace core, memory controllers, cache, etc.
Memory Replacement	While Running	Must bring server down to replace
Memory Bus Adaptor Replacement	While Running	Must bring server down to replace
I/O Upgrades	While Running	Must bring server down to replace (limited ability to replace I/O in some servers)
Concurrent Driver Maintenance	While Running	Limited – some drivers replaceable while running
Redundant Service Element	2 per System	“Support processors” can act as poor man’s SE, but no redundancy

Single book systems may not support concurrent memory upgrades

How can we determine equivalent configurations?

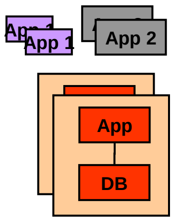
Real world aspects determine accurate equivalence

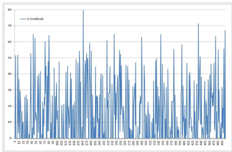


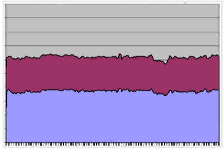
Top Down approach

What we see in customer environments

Bottoms up approach







What we know about platforms and measure in atomic benchmarks

Customer data often shows moving transaction processing off System z rarely reduces cost

Eagle TCO study for a financial services customer:

4 HP Proliant DL 980 G7 servers

Production

Development

256 cores total

Hardware	$1.6M
Software	$80.6M
Labor (additional)	$8.3M
Power and cooling	$0.04M
Space	$0.08M
Disaster Recovery	$4.2M
Migration Labor	$24M
Parallel Mainframe costs	$31.5M
Total (5yr TCO)	**$150M**

System z z/OS Sysplex

2,800 MIPS

Hardware	$1.4M
Software	$49.7M
Labor	Baseline
Power and cooling	$0.03M
Space	$0.08M
Disaster recovery	$1.3M
Total (5yr TCO)	**$52M**

65%
less cost!

Why are rehosting costs underestimated?

From HP's "Mainframe Alternative Sizing" guide, published in 2012…

MIPS Level	z196 Models	Actual MIPS	z10 EC Models	z10 Actual MIPS	z10 BC Models	z10 BC Actual MIPS	z114 Models	z114 Actual MIPS	HP Cores Estimate	Total HP equivalent MIPS
1,000	2817-701	1,202	2097-701	889	2098-Z02	1250	2818-Z01	782	2	866
2,000	2817-702	2,272	2097-702	1,667	2098-Z03	1784	2818-Z03	2026	5	1,860
3,000	2817-703	3,311	2097-704	3,114	2098-Z05	2760	2818-Z05	3139	8	3,021

Can a 2-chip, quad-core x86-based Blade server really replace 3,000+ MIPS?

- Simple core comparisons are inherently inaccurate…
- Real world use cases suggest this number is off by a factor of **10-20 times**

Eagle TCO study shows this mid-sized workload was *not* cheaper on the distributed platform

6x 8-way (x86) Production / Dev
2x 64-way (Unix) Production / Dev
Application/MQ/DB2/Dev partitions

2x z900 3-way Production / Dev / QA / Test

1,660 MIPS
(6 processors)

?

176 processors

$25.4M (5 yr. TCO)

$17.9M (5 yr. TCO)

29x
more cores!

482 Performance Units per MIPS

Eagle TCO Study shows a pure Intel offload was not cost-effective…

3x HP DL580 (2ch/20co)
Production / Dev / Test
(2011 x86 technology)

z800 Production /
Dev / Test
(2002 mainframe technology)

499 MIPS
(2.1 processors)

60 processors

Despite a 9-year technology gap,
the Intel platform still required
29x more processors

768 Performance Units per MIPS

"Performance units" used to define distributed server capacity

- Independent analyst measures and publishes capacity of all commercially-available distributed servers
- Provides relative comparison point across distributed servers
- Numerous Eagle TCO studies yield data on Performance Units per MIPS comparisons
  - Data feeds back into the Eagle model for predicting future case studies

Scenarios	zSW	MIPS	Dist. SW	Performance Units	Perf Units per MIPS ratio
Offloading Cases					
- Asian financial	CICS/DB2	6,700	OpenFrame/Oracle	816,002*	**122***
- Asian insurance	CICS/DB2	1,620	OpenFrame/Oracle	437,992	**270**
- NA financial services	CICS/DB2	1,660	UniKix/Oracle	800,072	**482**
- European financial	CICS/DB2	332	TXSeries/Oracle	222,292	**670**
- US County government	CICS/Datacom	88	Unikix/Oracle	43,884	**499**
Offload Studies					
- European agency	CICS/DB2/IMS	18,000	Tuxedo/Oracle	3,328,432 est	**185** est
- Restaurant chain	PeopleSoft/DB2	1,600	Oracle	186,224 est	**116** est
- Asian healthcare	CICS/DB2	671	Java	251,740 est	**375** est
- Asian bank	CICS/DB2	1,316	OpenFrame/Oracle	200,952 est	**153** est
- US utility	PeopleSoft/DB2	491	Oracle	163,744 est	**333** est
- US manufacturer	PeopleSoft/DB2	3,343	Oracle	774,120 est	**232** est

* Production workload only

Is there a cross over point? 1,000 MIPS? 500 MIPS?

A sampling of Eagle TCO data suggests there is no minimum MIPS value that automatically makes an offload financially beneficial…

			5-Year TCO		
Customer	z (MIPS)	distributed (PUs)	z	distributed	z/dist %
Average	**1,166**	**218,472**	**9,050,451**	**16,325,492**	
SA Government Agency	475	241,291	19,773,442	25,261,624	78.27%
German Financial	1,200	263,177	3,939,889	4,701,033	83.81%
NA Financial Services	2,526	308,144	3,456,611	5,939,476	58.20%
US utility company	456	163744	6,157,295	13,380,866	46.02%
European Insurance	904	171,062	13,019,980	15,877,484	82.00%
US Manufacturer	900	453,168	11,277,266	16,019,269	70.40%
Asian Bank	1,416	136,013	2,342,300	7,237,681	32.36%
US Retailer	1,700	215,124	3,543,154	8,951,851	39.58%
US County Government	88	43,884	4,717,394	8,108,668	58.18%
US Retailer	1,500	184,732	9,254,186	20,861,515	44.36%
AP bank	1,336	168,113	17,300,000	27,200,000	63.60%
AP bank	300	24,162	5,200,000	11,500,000	45.22%
US Manufacturer	1,917	261,040	4,758,313	7,350,216	64.74%
US Food Services	1,600	424,952	21,966,475	56,167,206	39.11%

The determining factor is really the *nature* of the workload…

Eagle TCO study shows this small workload was *not* cheaper on the distributed platform

2x 16-way (Unix) Production / Dev / Test / Education
App, DB, Security, Print and Monitoring
4x 1-way (Unix) Admin / Provisioning / Batch Scheduling

z890 2-way Production / Dev / Test / Education
App, DB, Security, Print, Admin & Monitoring

332 MIPS

(0.88 processors)

36 processors

$17.9M (4 yr. TCO)

41x
more cores!

$4.9M (4 yr. TCO)

670 Performance Units per MIPS

Eagle TCO study shows even this VERY small workload was not cheaper on the distributed platform

4x p550 (1ch/2co)
Application and DB

z890 Production / Test

88 MIPS
(0.24 processors)

8 processors

$8.1M (5 yr. TCO)

33x
more cores!

$4.7M (5 yr. TCO)

499 Performance Units per MIPS

What happens in a TCO study?

Workload identified for analysis

frontends

databases

other components

Application XYZ (Prod, Dev, QA)

Deployment Choices

Do nothing

Optimize current environment

Deploy on other platforms

Key steps in analysis

1. Establish equivalent configurations
- Needed to deliver workload

2. Compare Total Cost of Ownership
- TCO looks at different dimensions of cost