What System z Can Do That Intel Can't

The New zEnterprise – A Cost-Busting Platform

IBM CPO System z Customer Briefing 2013

What System z Can Do That Intel Can't

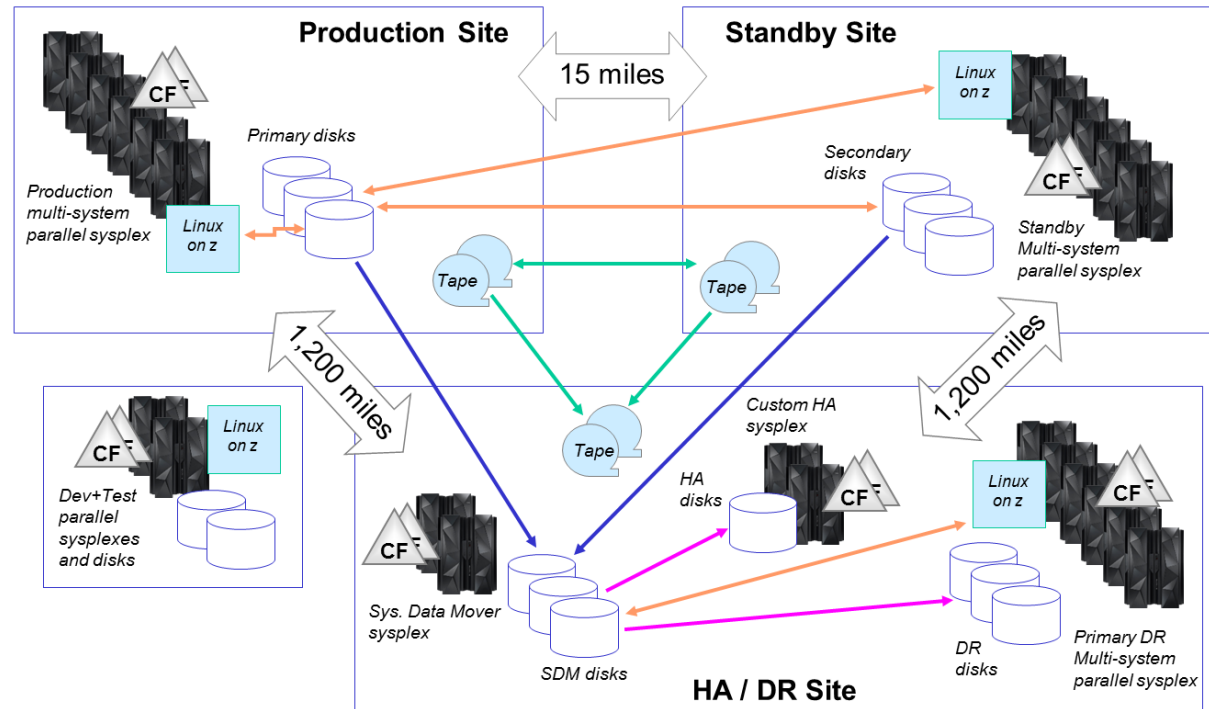1. Run Bigger and More Workloads

Intel Sandy Bridge

System z

Intel Does Not Have The Physical Capacity For State-of-the-Art Systems Of This Magnitude

- 1B CICS trans/day
- 4,000 IMS trans/sec
- 14M ACH transactions in 2.5 hours
  - 6-way sysplex
  - 30ms response
  - 216 CPU's at primary site
  - 200K MIPS

- Zero outages, zero customer impact
- Linux is Active-Active in the two data centers, with zero downtime
  - 15% Linux, growing at 30%
- *"Crazy about security overall, and the z system has a fortress around it"*

System z Delivers More Raw Processing Capacity Than Intel

*World's **fastest** clock speed*	**5.5 GHz**
Total cores	**120**
Configurable cores	**101**
General processor core performance	**1,514 MIPS**
Specialty processor core performance	**1,514 MIPS**
Total Capacity	**78,426 MIPS**

zEC12

Maximum x86 clock speed = **3.4 GHz**

Maximum x86 cores = **32**

Intel Sandy Bridge

Parallel Sysplex Enables System z To Scale To Capacities Far Beyond What Intel Can

Parallel sysplex clustering delivers **highest availability**

Single System Sysplex

Cross Connected Servers with internal Coupling Facilities

External Coupling Facility (Can be different class server)

Potentially 2.5 million MIPS per 32-way cluster

Supports rolling software updates via automatic sysplex failover

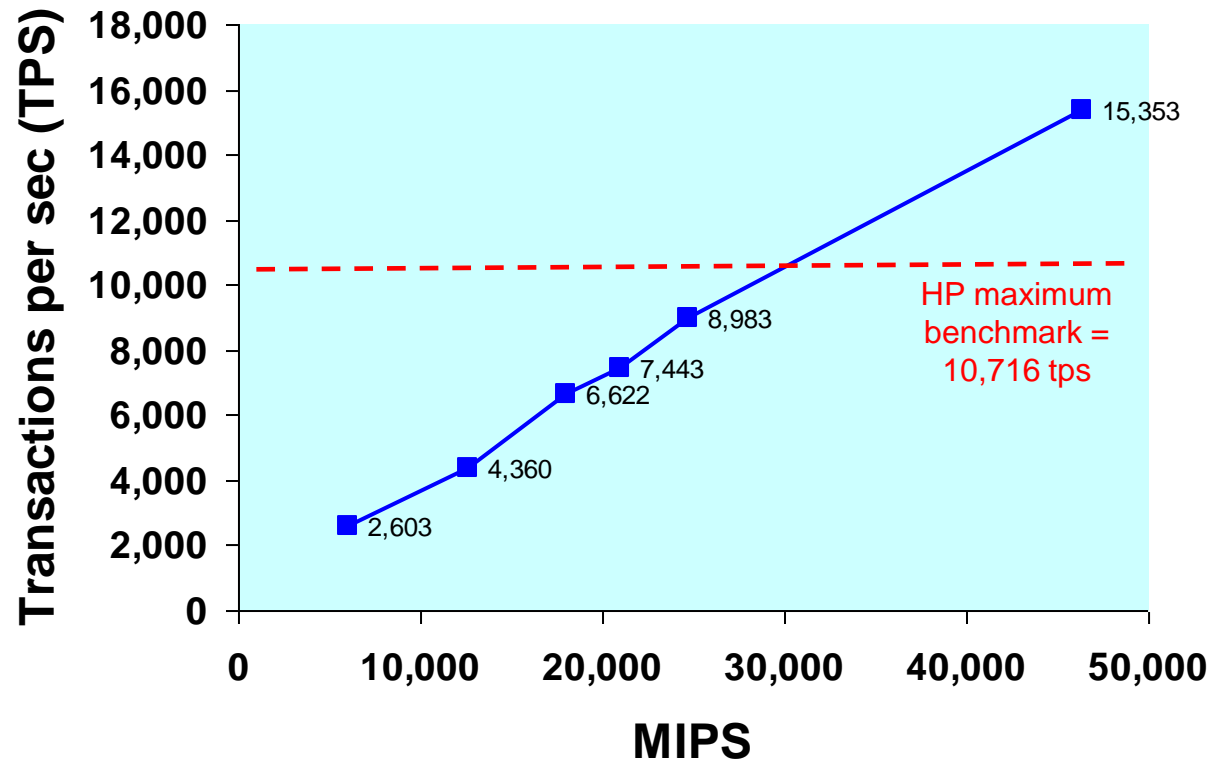# Real-World Benchmarks Show System z Runs Bigger Workloads Than Intel

Kookmin Bank
- **IBM System z and DB2**
- TCS BaNCS
- **15,353** Transactions/second
- **50** Million Accounts
- IBM benchmark for customer
- DB2 V9, CICS 3.1, z/OS V1.8

State Bank of India[1]
- **HP Superdome**
- **TCS BaNCS**
- **10,716** Transactions/second
- **500** Million Accounts
- Largest banking benchmark performance claimed by HP

System z and BaNCS Online Banking Benchmarks

Transactions per sec (TPS) vs MIPS

- 2,603
- 4,360
- 6,622
- 7,443
- 8,983
- 15,353

HP maximum benchmark = 10,716 tps

[1] Source: http://www.enterprisenetworksandservers.com/monthly/art.php?2976 and *InfoSizing FNS BANCS Scalability on IBM System z – Report Date: September 20, 2006;* Clement Report; http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA1-4027ENW.pdf Feb 2010

System z Has More Cache Than Intel To Support Cache Intensive Workloads

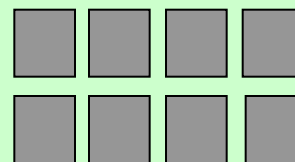**zEC12 chip**
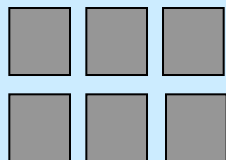
6 cores per chip

L1 Cache 960KB

L2 Cache 12MB

L3 Cache 48MB
(**8MB per core**)

L4 Cache 1,536MB across 4 books

Memory 3TB

Sandy Bridge chip

8 cores per chip

L1 Cache 512KB

L2 Cache 2MB

L3 Cache 20MB
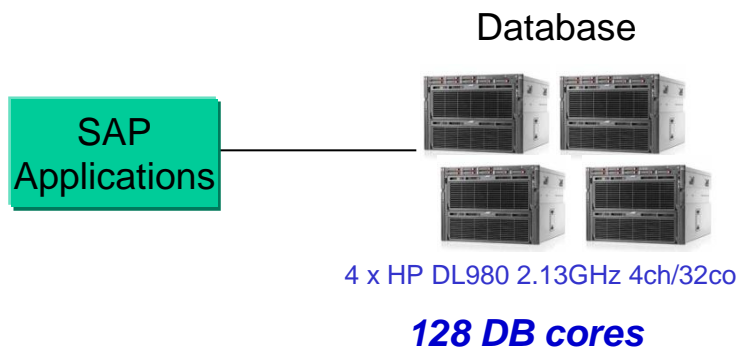(**2.5MB per core**)

No L4 Cache
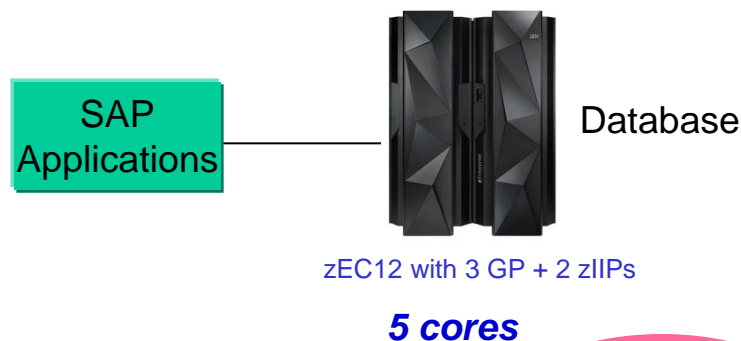Processor reported busy during a memory fetch, but no useful work is getting done

Memory 768GB

System z Is More Efficient For Data Processing Workloads

Cost advantage for smaller scale SAP database

SQL Server on Intel

Database

SAP Applications

4 x HP DL980 2.13GHz 4ch/32co

128 DB cores

DB2 on z/OS

SAP Applications

Database

zEC12 with 3 GP + 2 zIIPs

5 cores

96% less core
29% less 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

I/O Bandwidth Is Also Important For Critical Data Workloads
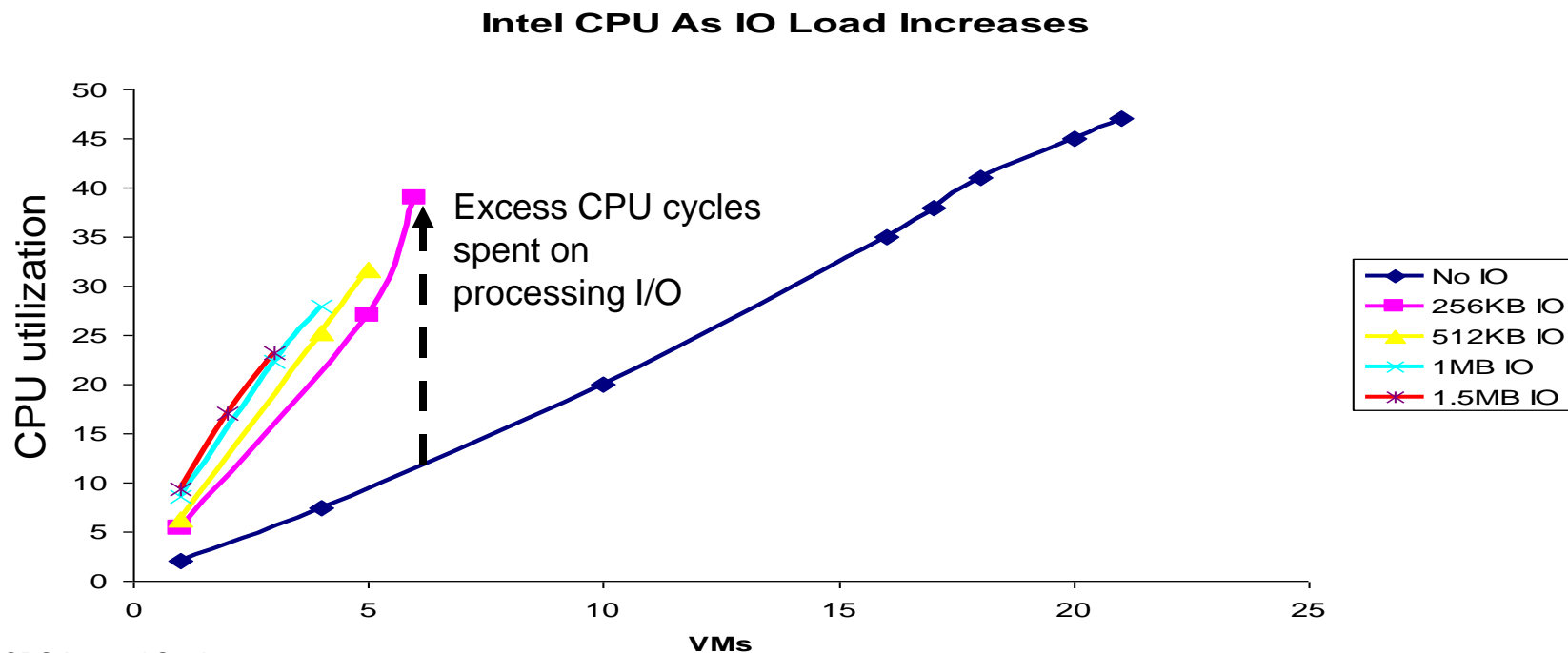
Bank of China System z Benchmark required huge I/O bandwidth capacity
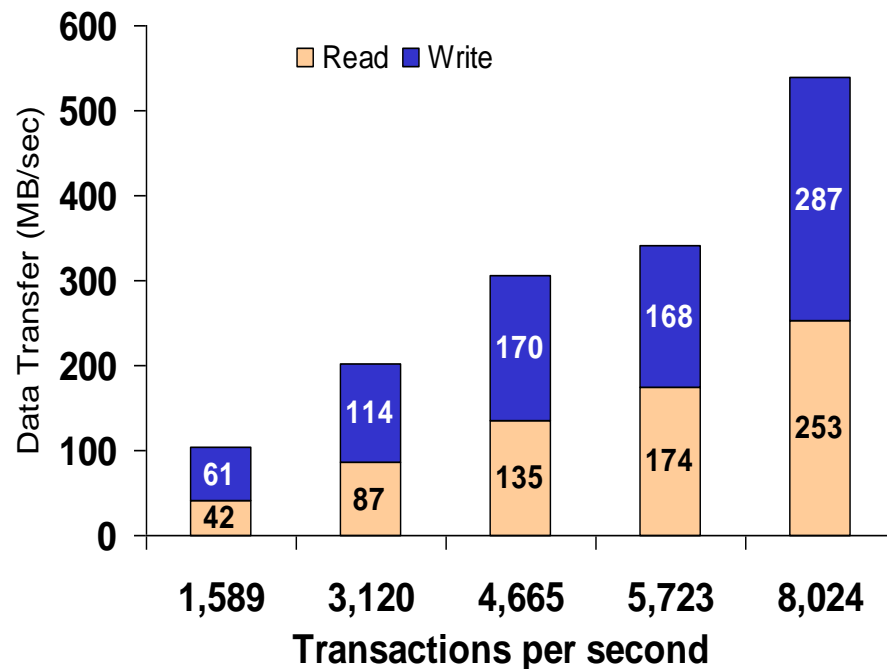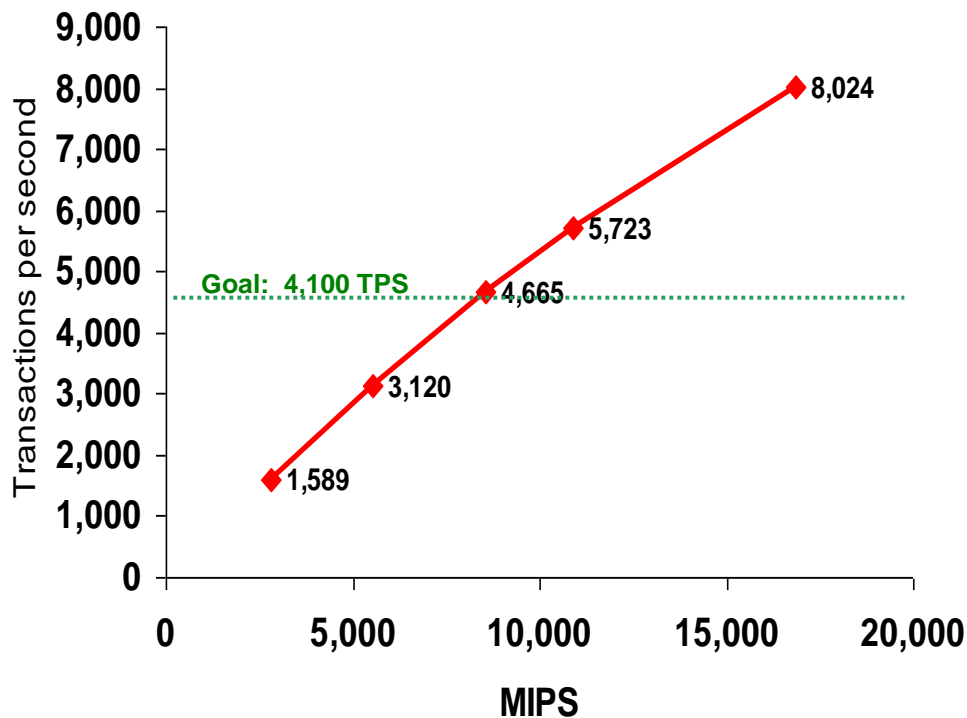
Bigger workloads require more I/O bandwidth capacity

System z's Dedicated I/O Subsystem Delivers More I/O Processing Capacity Than Intel

- Intel's performance degrades as I/O demand increases
  - No dedicated I/O subsystem
- 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

Intel CPU As IO Load Increases

Excess CPU cycles spent on processing I/O

Legend:
- No IO
- 256KB IO
- 512KB IO
- 1MB IO
- 1.5MB IO

Y-axis: CPU utilization (0 to 50)
X-axis: VMs (0 to 25)

Source: CPO Internal Study. 12-core Westmere EP with KVM. FB at 22 tps with varying IO per transaction.

z/OS Database Workloads Benefit From Higher I/O Bandwidth
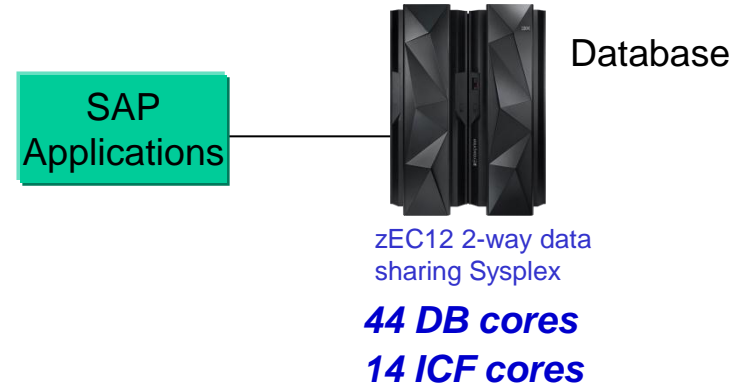
Competitor DB on Intel

Database

SAP Applications

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

128 DB cores

DB2 on z/OS

Database

SAP Applications

zEC12 2-way data
sharing Sysplex
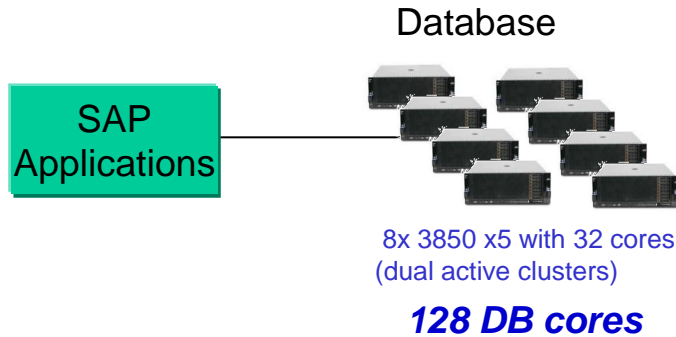
44 DB cores
14 ICF cores

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

*A world record at **half** the cost!*

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

Linux On System z Workloads Also Benefit From Higher I/O Bandwidth

Comparison of consolidation platforms
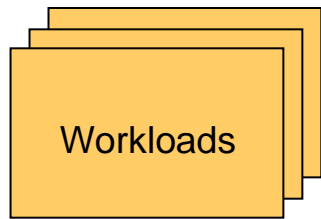
Workloads

Online banking workloads, each driving **22** transactions per second, with **1 MB I/O per transaction**

1 workload per 16-core x86 blade

Virtualized on x86 16 core HX5 Blade

48 workloads per 32-way z/VM

I/O bandwidth large scale pool

Virtualized on z/VM on zEC12 32 IFLs

24x more workload density than x86

What System z Can Do That Intel Can't

1. Run Bigger and More Workloads

2. Perfect Workload Management

Intel Sandy Bridge

System z

System z Has Perfect Workload Management

Intel can't do this

- z/OS workload management is perfect for processes
  - ▶ I/O subsystem extends prioritization to the storage disks
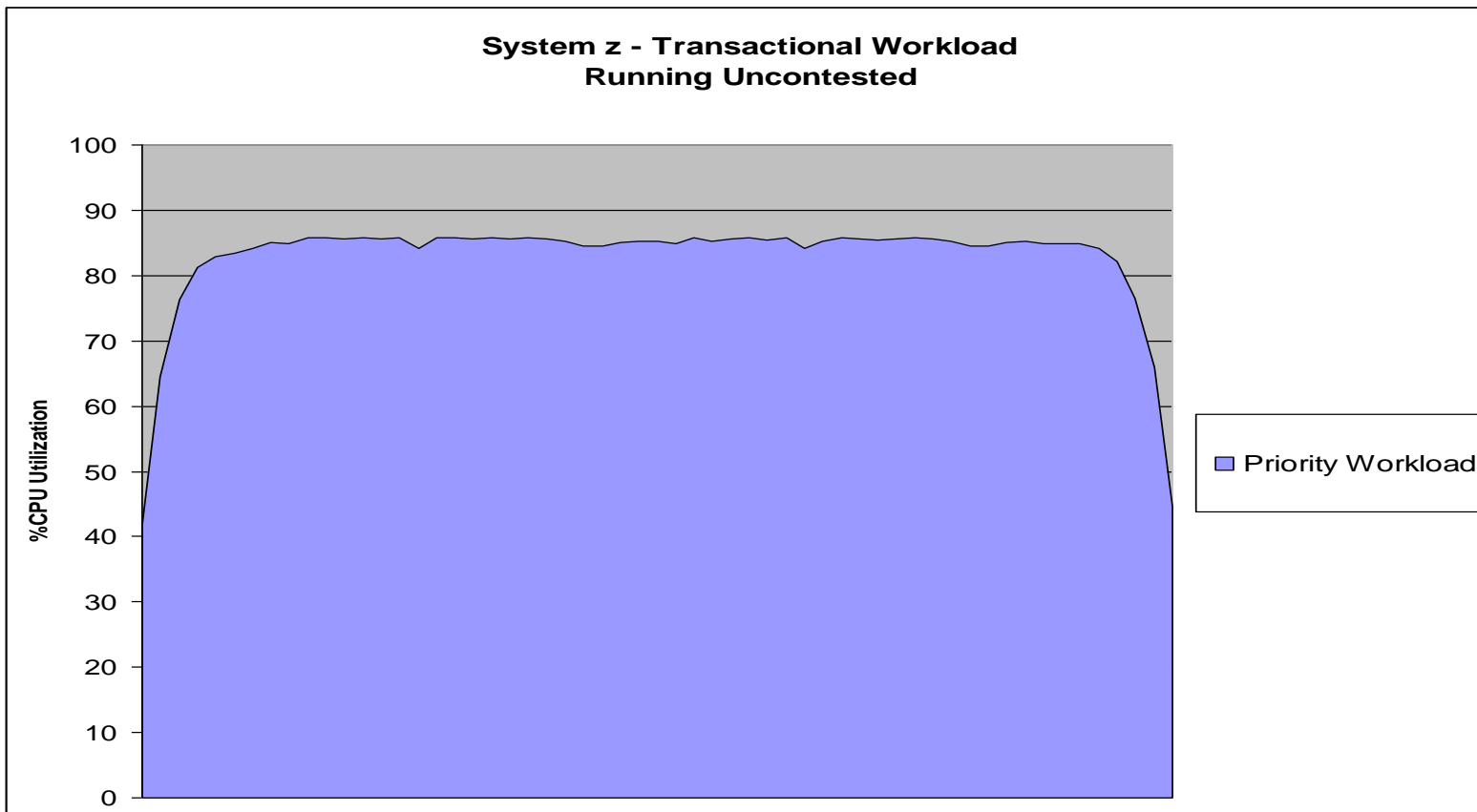- PR/SM workload management is perfect for LPARs

Intel Sandy Bridge

System z

Priority Transactional Workload With Constant Demand Running Standalone On System z PR/SM

**System z - Transactional Workload
Running Uncontested**

%CPU Utilization

(y-axis: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100)

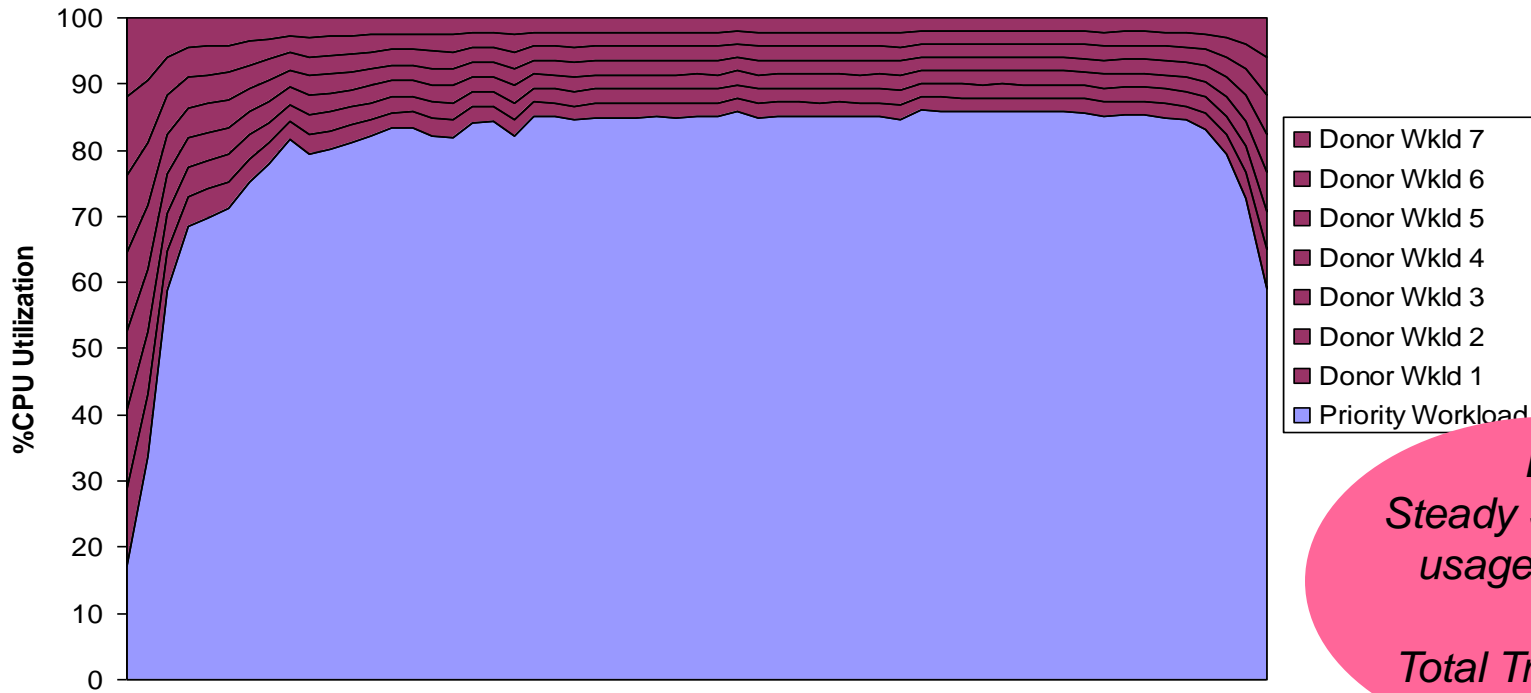■ Priority Workload

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

Priority Workload Metrics
Total Throughput: 418K
Maximum TPS 130

Priority Transactional Workload On System z Does Not Degrade When Low Priority Donor Workload Is Added

z/OS WLM - Priority Transactional Workload
Running With Other Workloads - 1 Hour Run

%CPU Utilization

- Donor Wkld 7
- Donor Wkld 6
- Donor Wkld 5
- Donor Wkld 4
- Donor Wkld 3
- Donor Wkld 2
- Donor Wkld 1
- Priority Workload

NO
Steady State CPU usage leakage
1%
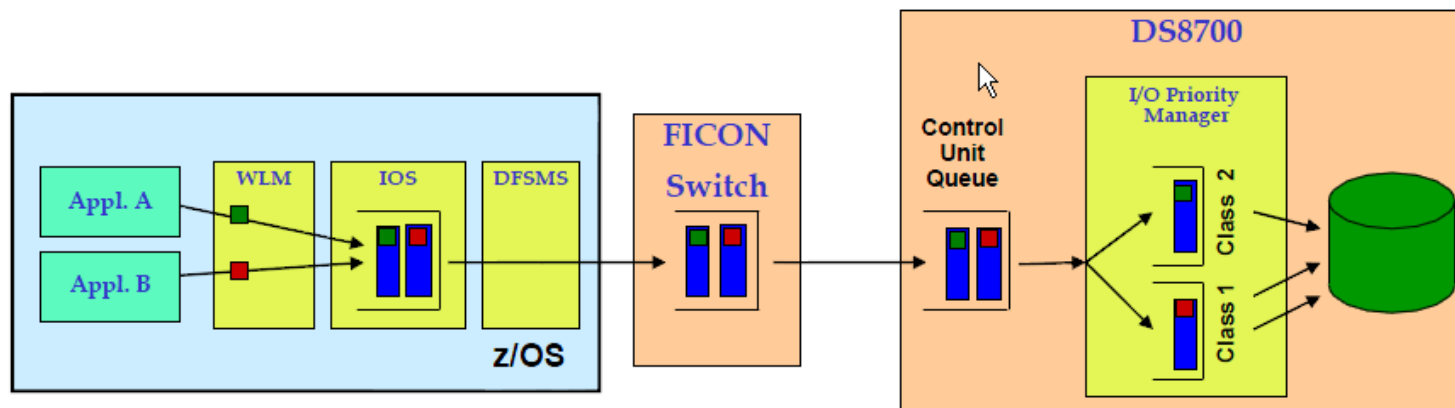*Total Transaction leakage*

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

Priority Workload Metrics
Total Throughput: 415K
Maximum TPS 128

z/OS Workload Management 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
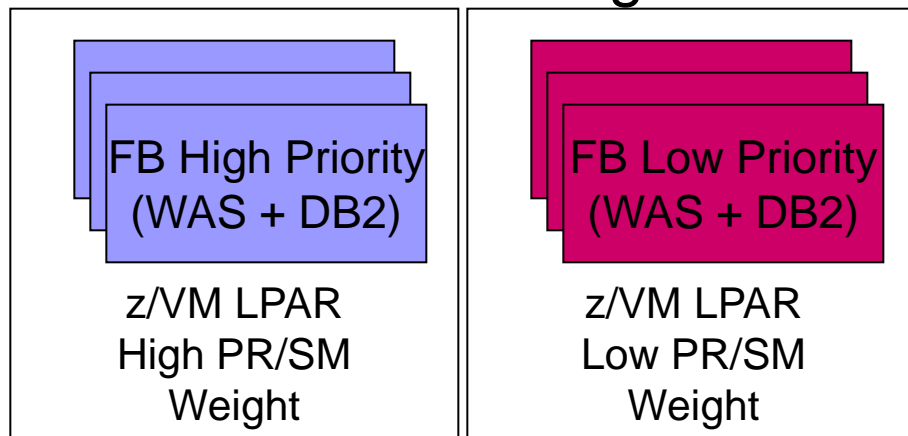

Intel can't do this

DEMO: z/OS Workload Management
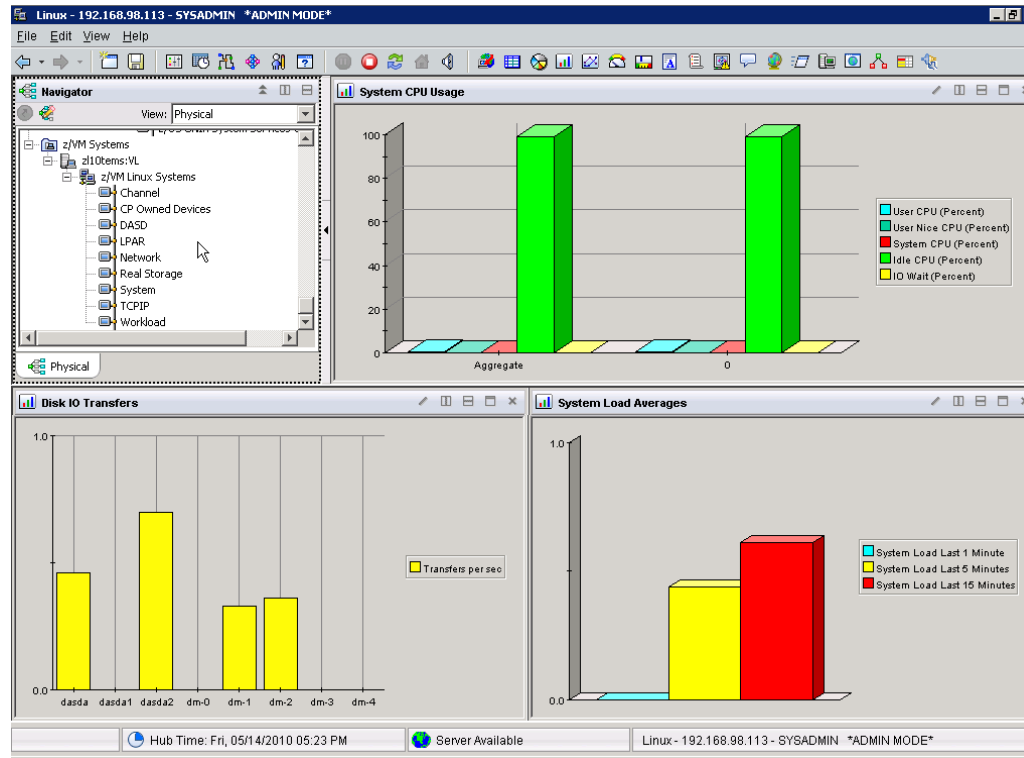
Comparison of System z PR/SM To Common Hypervisor Virtualization Environments

- 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

FB High Priority (WAS + DB2)	FB Low Priority (WAS + DB2)
z/VM LPAR High PR/SM Weight	z/VM LPAR Low PR/SM Weight

PR/SM Partitions
zEC12
32 Shared cores

FB High Priority (WAS + DB2) High share

FB Low Priority (WAS + DB2) Low share

Common Intel Hypervisor
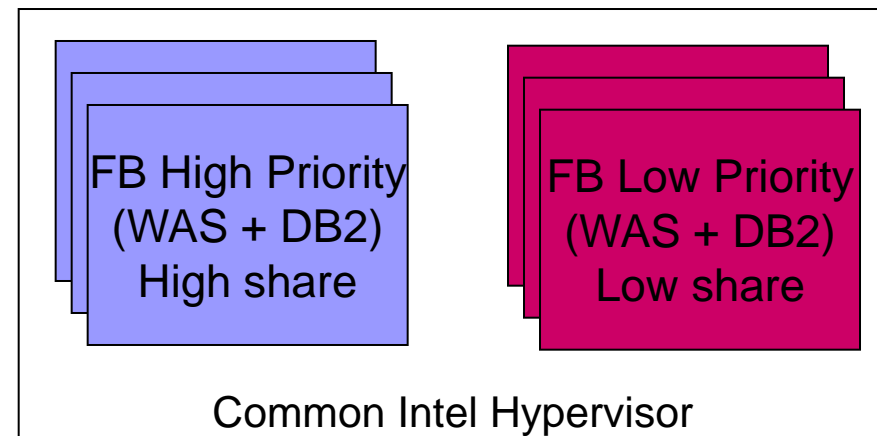
Intel Westmere EX
40 cores

Priority Workload With Varying Demand Running Standalone On System z PR/SM
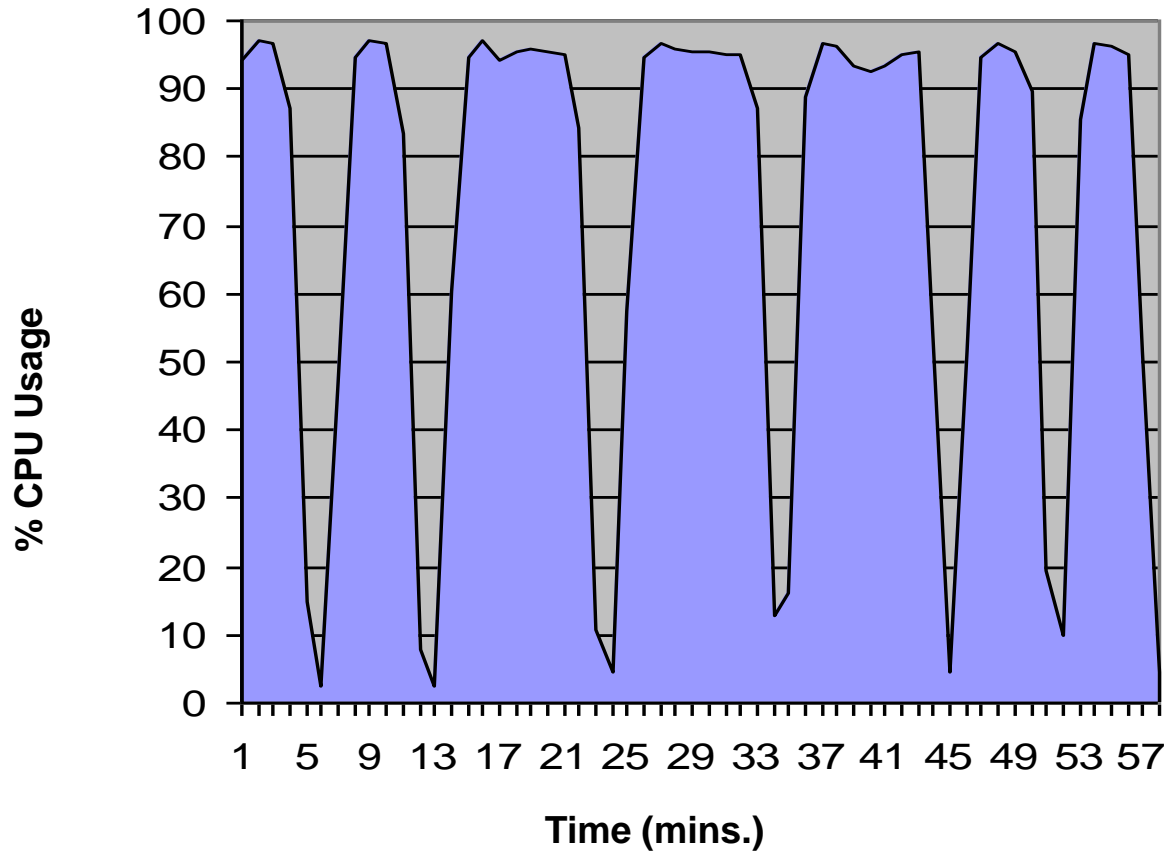
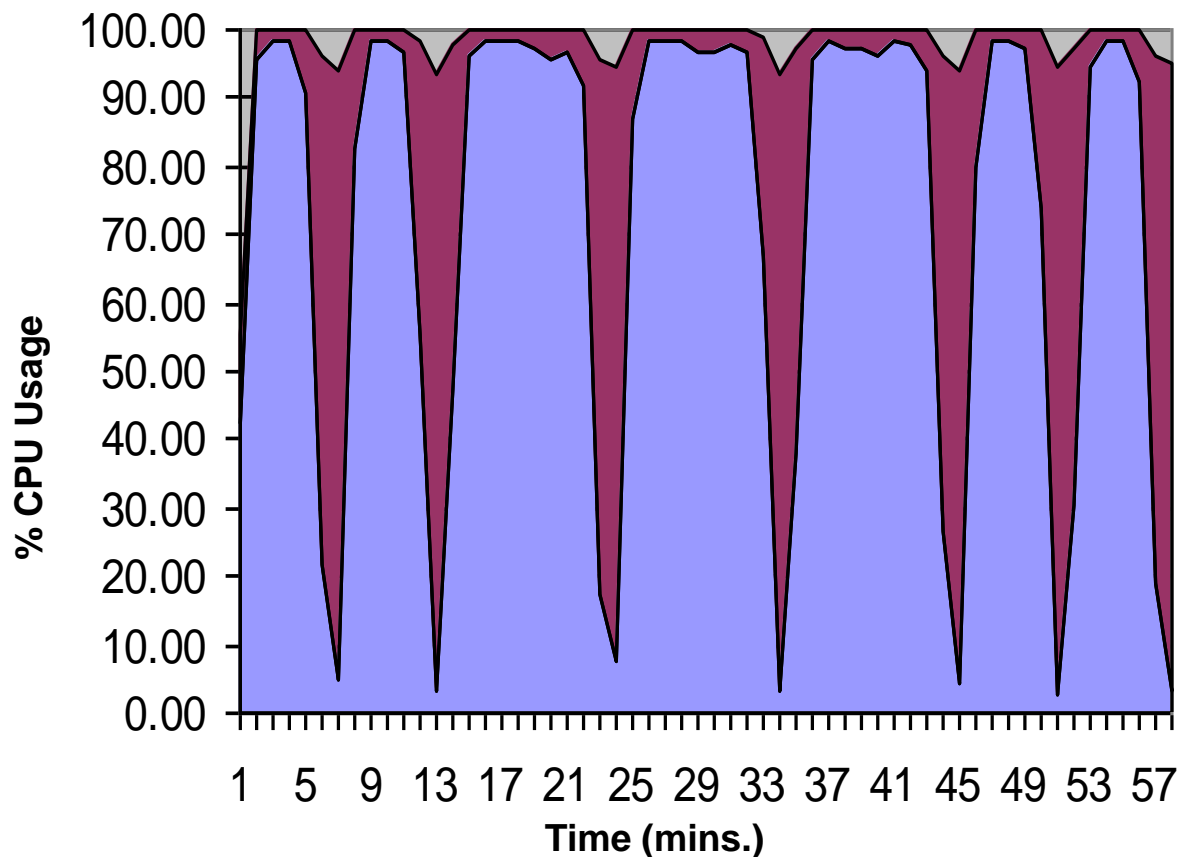High Priority Workload Demand Curve

Priority Workload

Capacity Used
High Priority – 72.2% CPU Minutes
Unused (wasted) – 27.8% CPU Minutes

Priority Workload Metrics
Total Throughput: 9.13M
Avg Response Time: 140ms

Priority Workload On System z Does Not Degrade When Low Priority Donor Workload Is Added

Run High Priority And Low Priority Workloads Together

- ■ Donor Workload
- □ Priority Workload
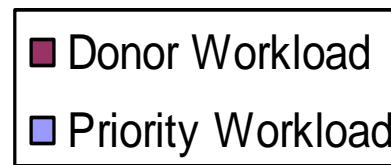
NO throughput leakage NO response time increase

Capacity Used
High Priority – 74.2% CPU Minutes
Low Priority – 23.9% CPU Minutes
Wasted – 1.9% CPU Minutes

Priority Workload Metrics
Total Throughput: 9.13M
Avg Response Time: 140ms

Priority Workload With Varying Demand Running Standalone On Common Hypervisor

High Priority Guest
CPU Demand
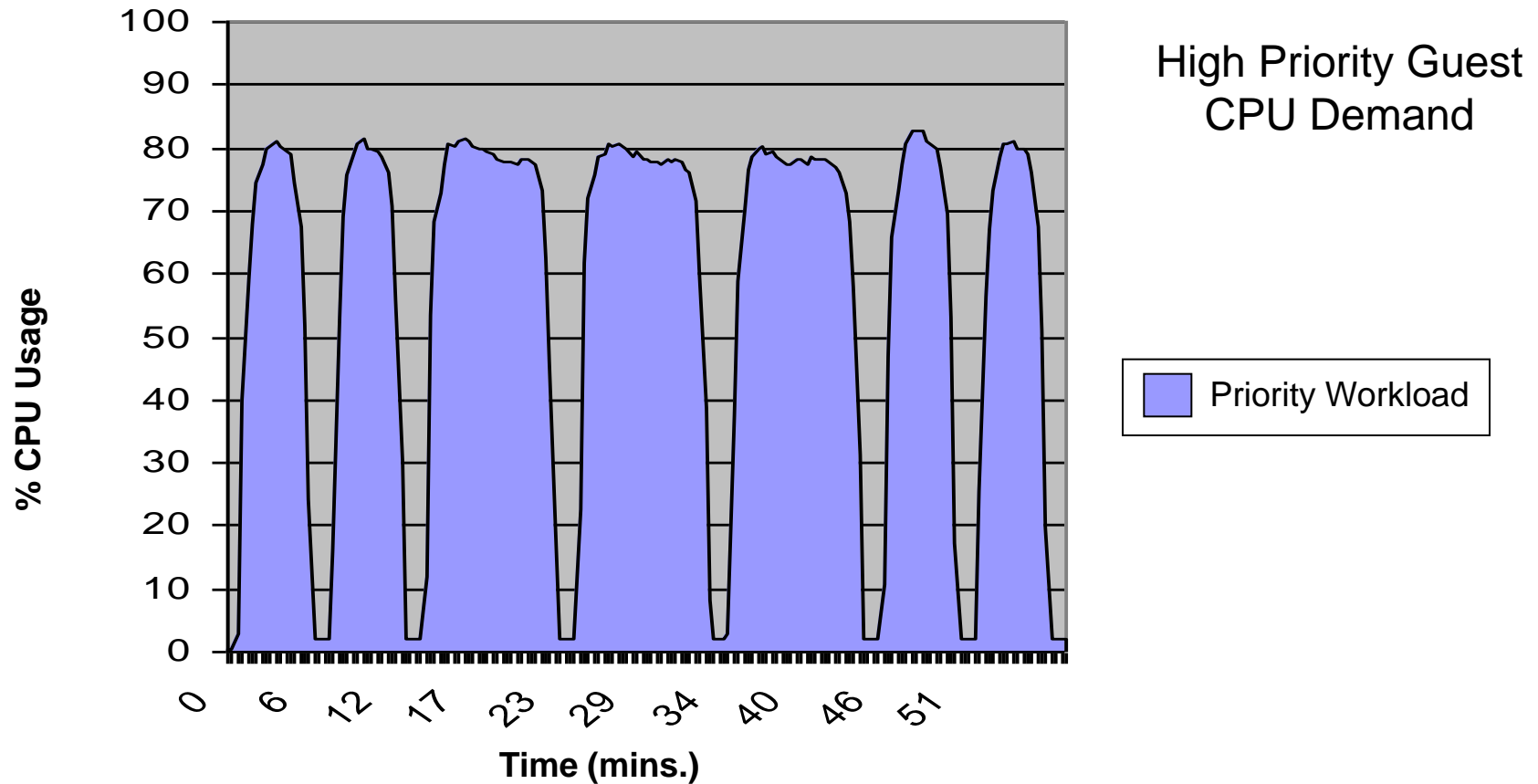
Priority Workload

Capacity Used
High Priority – 57.5% CPU Minutes
Unused (wasted) – 42.5% CPU Minutes

Priority Workload Metrics
Total Throughput: 6.47M
Avg Response Time: 153ms

Priority Workload On Common Hypervisor Degrades Severely When Low Priority Workload Is Added

Run High Priority And Low Priority Workloads Together

Legend:
- Donor Workload
- Priority Workload
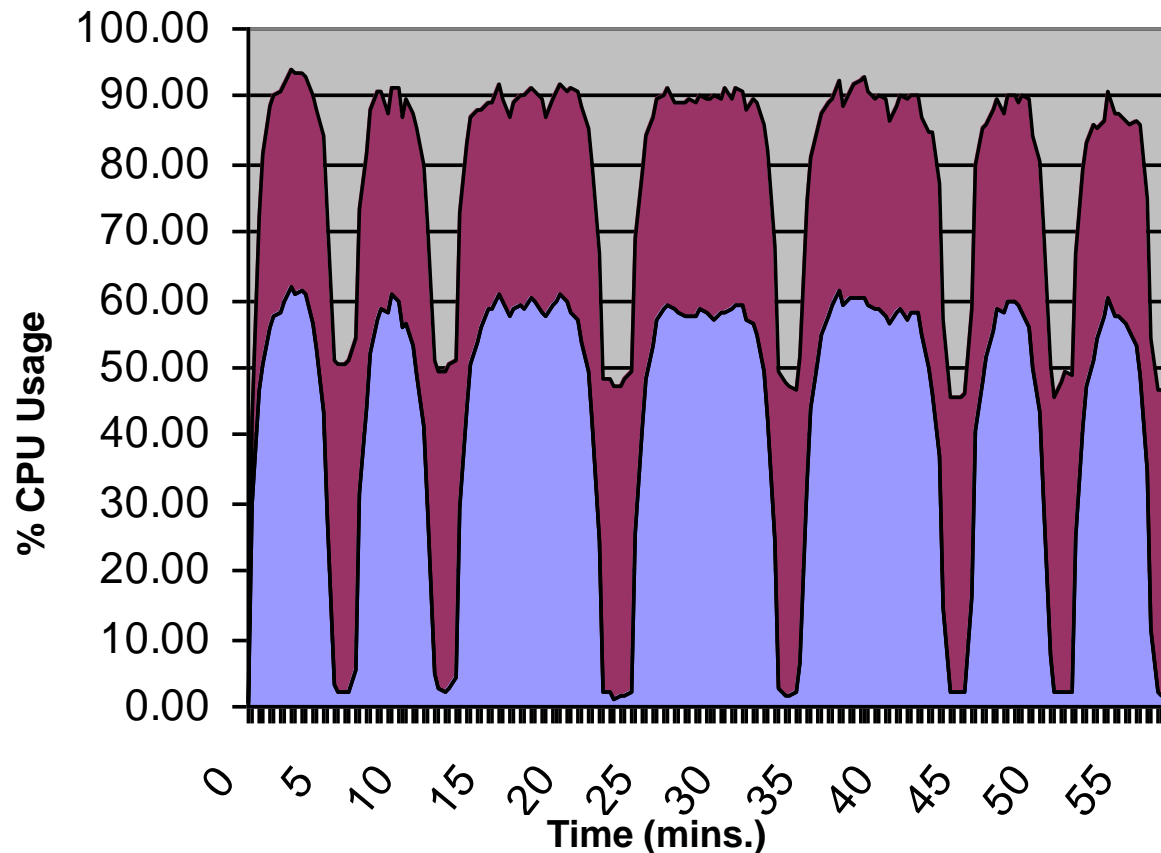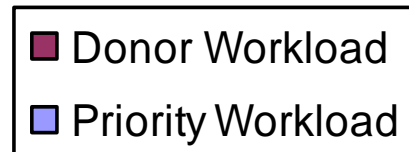
30.7%
throughput leakage
45.1%
response time increase
21.9%
wasted CPU minutes

Capacity Used
High Priority – 42.3% CPU Minutes
Low Priority – 35.8% CPU Minutes
Wasted – 21.9% CPU Minutes

Priority Workload Metrics
Total Throughput: 4.48M
Avg Response Time: 220ms

System z Virtualization Enables Mixing Of High And Low Priority Workloads Without Penalty

System z

Leading Intel Hypervisor

Too much resource given to Low Priority workload

High Priority workload gets less resource than needed
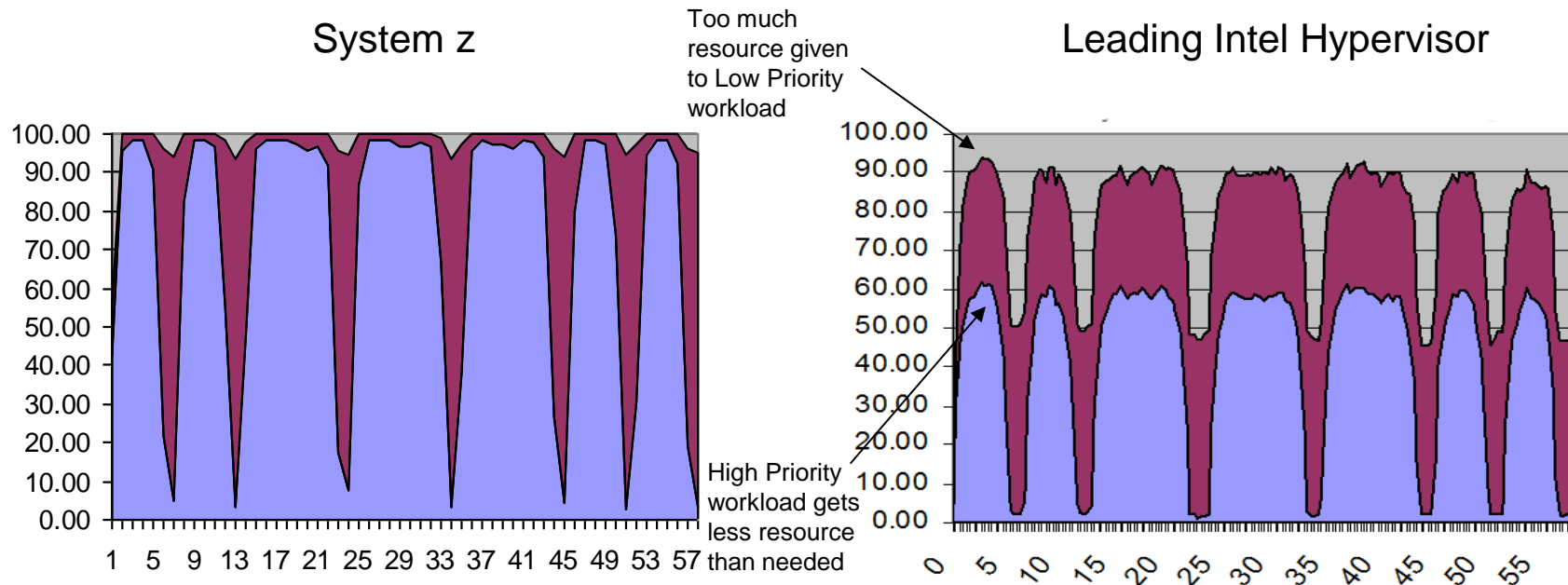
- **Priority Workload**
  - No throughput reduction
  - No response time increase
- **Low Priority Workload**
  - Soaks up remaining CPU minutes
- Unused CPU minutes 1.9%

- **Priority Workload**
  - 31% throughput reduction
  - 45% response time increase
- **Low Priority Workload**
  - Soaks up more CPU minutes
- Unused CPU minutes 21.9%

System z Virtualization Enables Mixing Of High And Low Priority Workloads Without Penalty

System z

Leading Intel Hypervisor

Too much resource given to Low Priority workload

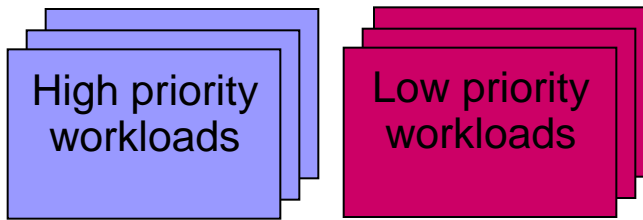High Priority workload gets less resource than needed
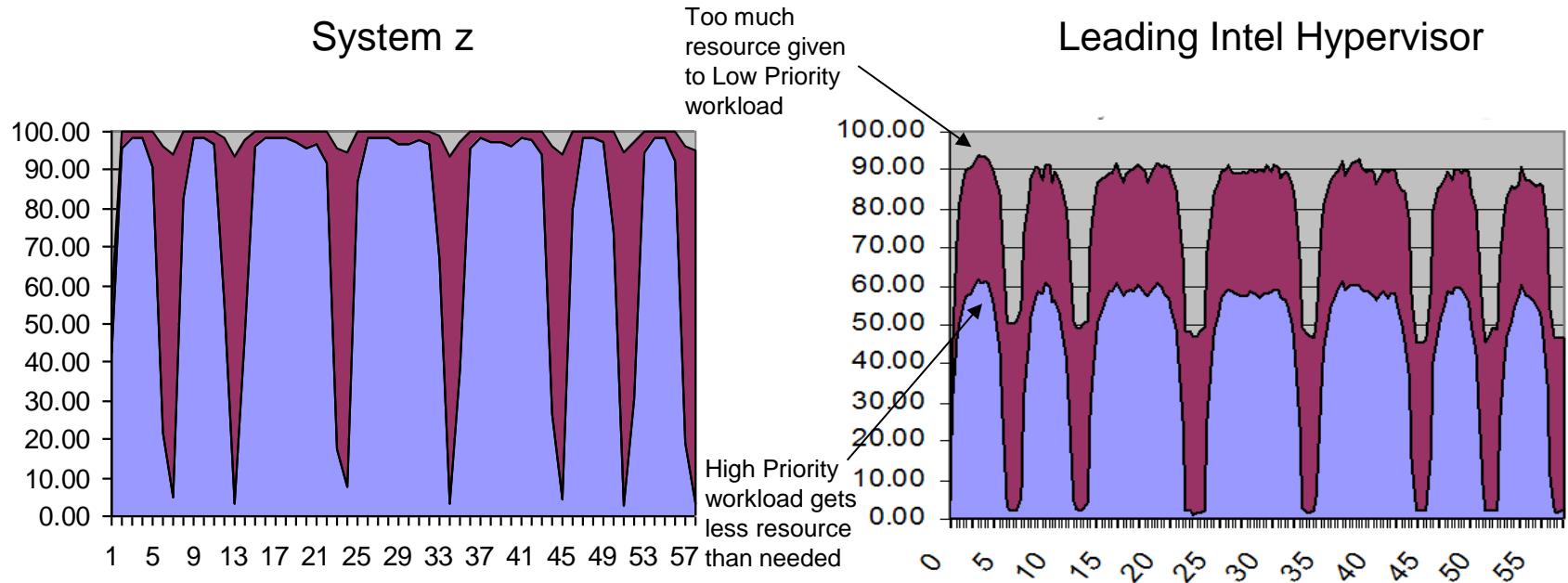
- Perfect workload management
- Consolidate workloads of different priorities on the same platform
- Full use of available processing resource (high utilization)

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

Deliver High And Low Priority Workloads Together While Maintaining Response Time SLA

Comparison to determine which platform provides the lowest TCA over 3 years

VMs

VMs

VMs

Virtualized on 3 Intel 40 core servers

$13.7M (3 yr. TCA)

High priority workloads

Low priority workloads

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

High priority online banking workloads driving a total of **11.9M** transactions per hour and low priority discretionary workloads

VMs

VMs

z/VM LPAR   **z/VM LPAR**
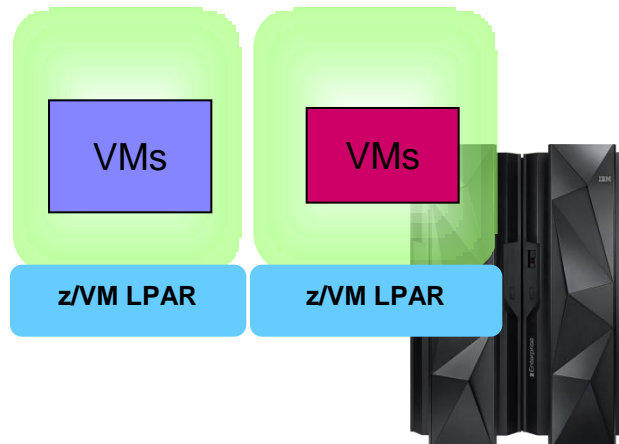
z/VM on zEC12

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.

What System z Can Do That Intel Can't

1. Run Bigger and More Workloads

2. Perfect Workload Management

3. Greater Core Density

Intel Sandy Bridge

System z

Why Core Proliferation Happens When Moving Workload From System z To Intel

- De-consolidation of applications to dedicated servers – decomposing highly tuned co-located components

- Processing expansion requirements for CICS/COBOL applications

- 3x expansion when converting hierarchical databases to relational

- Functional segregation into production, development and test

- 100% hardware coverage for Disaster Recovery costs double

Intel Sandy Bridge

System z

Core Proliferation For A Large Workload

16x 32-way HP Superdome
App. Production / Dev / Test

8x 48-way HP Superdome
DB Production / Dev /Test

zEC12 41-way Production / Dev / Test

41 GP processors
(38,270 MIPS)
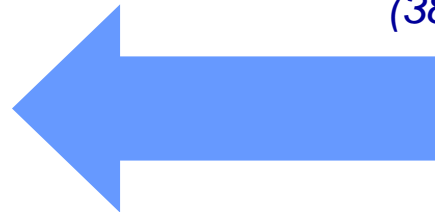
896 processors
(3,668,600 Perf Units)

22x more cores!

$180M (5 yr. TCA)

$111M (5 yr. TCA)

NOTE: To cover DEV/QA capacity, add 100% servers for distributed servers, add 25% MIPS (8,000) to System z

Core Proliferation For A Mid-sized Workload

6x 8-way Production / Dev
2x 64-way Production / Dev
    Application/MQ/DB2/Dev partitions

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

6 processors
(1,660 MIPS)

176 processors
(800,072 Performance units)

$25.4M (5 yr. TCO)

$17.9M (5 yr. TCO)

29x more cores!

482 Performance Units per MIPS

Core Proliferation For Oracle Workloads

TCO study for a Media and Entertainment Industry customer

107 HP servers

1440 cores total

30x more cores!

zEC12

48 IFLs

1 PS701

1 HX5

Hardware	$2.9M
Software	$24.2M
Labor	$7.9M
Space, Power and cooling	$1.2M
Disaster Recovery	$6.5M
Total (5 yr. TCO)	**$42.7M**

|	
Hardware	$4.9M
Software	$8.5M
Labor	$1.8M
Space, Power and cooling	$0.5M
Disaster recovery	$4.8M
Total (5 yr. TCO)	**$20.5M**

Intel: Oracle DB + App costs = $13.1M (LIC + maint over 5 yrs.).
IBM: Oracle DB + App costs = $1.92M (LIC + maint over 5 yrs.)

Migration Offloads Have Additional Costs

Typical Eagle TCO Study For A Financial Services Customer

x86 – 4 HP Proliant DL 980 G7 servers

Production   Development
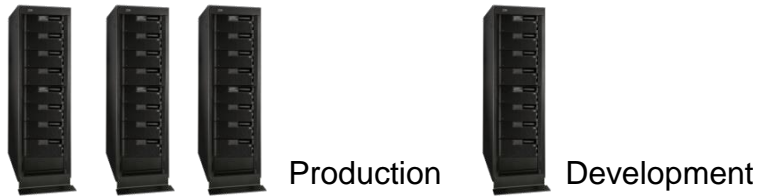
256 cores total

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 (5 yr. 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 (5 yr. TCO)	**$52M**

What System z Can Do That Intel Can't

1. Run Bigger and More Workloads

2. Perfect Workload Management

3. Greater Core Density

4. Spare Capacity for Growth

Intel Sandy Bridge

System z

System z's Integrated Capacity On Demand (CoD) Extends To Storage

- System z ships with spare processors installed
  - ▶ CoD can turn on spare processors without service interruption
  - ▶ **Intel can't do this**

System z

- CoD extends to DS8870
  - ▶ Up to six standby CoD disk drive sets (96 disk drives total) can be concurrently field-installed into the system*
  - ▶ Non-disruptive activation
  - ▶ Easy to logically configure the disk drives for use – no IBM intervention required

  Midrange storage typically can't do this

DS8870

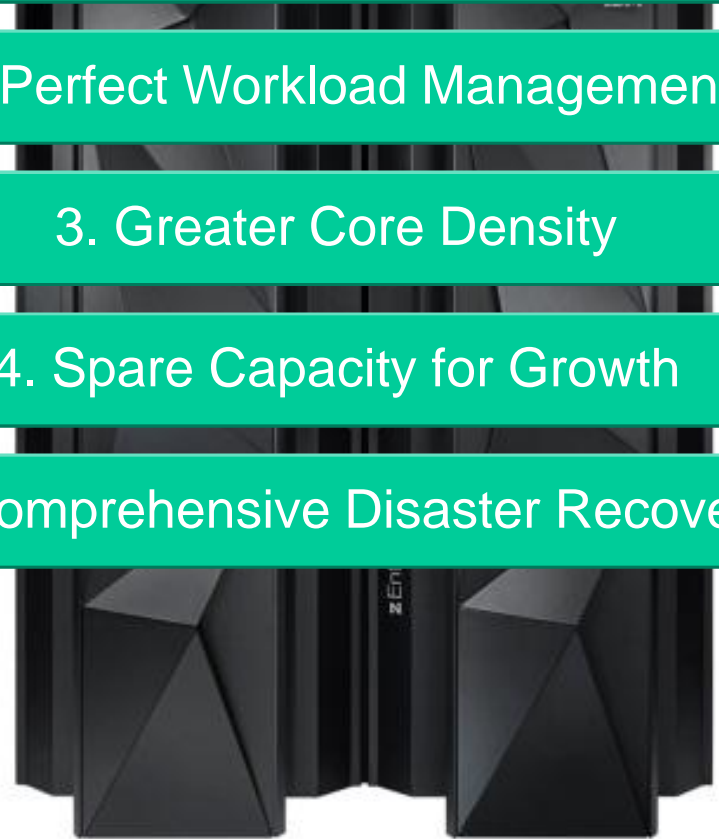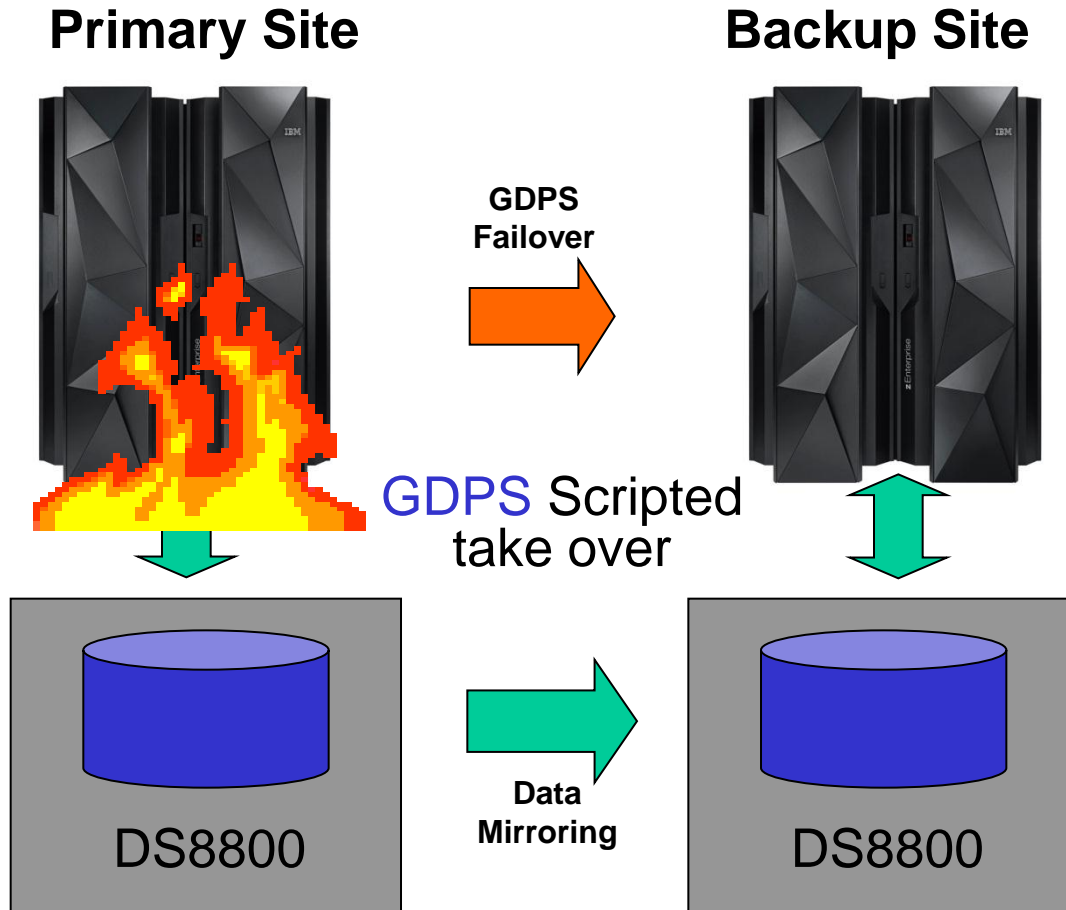*SSDs not available for CoD configurations

What System z Can Do That Intel Can't

1. Run Bigger and More Workloads

2. Perfect Workload Management

3. Greater Core Density

4. Spare Capacity for Growth

5. Comprehensive Disaster Recovery

Intel Sandy Bridge

System z

System z Disaster Recovery Is Systematic And Comprehensive

Primary Site

Backup Site

GDPS Failover

GDPS Scripted take over

DS8800

Data Mirroring

DS8800

- Site Failover
  - ▶ Failover to secondary site in case of complete site failure

- Data Mirroring
  - ▶ Protect data in the event of a disk system failure

Supports systematic Disaster Recovery for virtualized Linux environments also

Complexity Of Intel Disaster Recovery Solutions Prohibits Wide Spread Use

- Workloads on standalone Intel servers require a disaster recovery solution for each server
  - ▶ Data mirroring
  - ▶ Failover and restart

- Embedded storage is difficult to mirror
- Comprehensive workload failover is not feasible for hundreds of servers
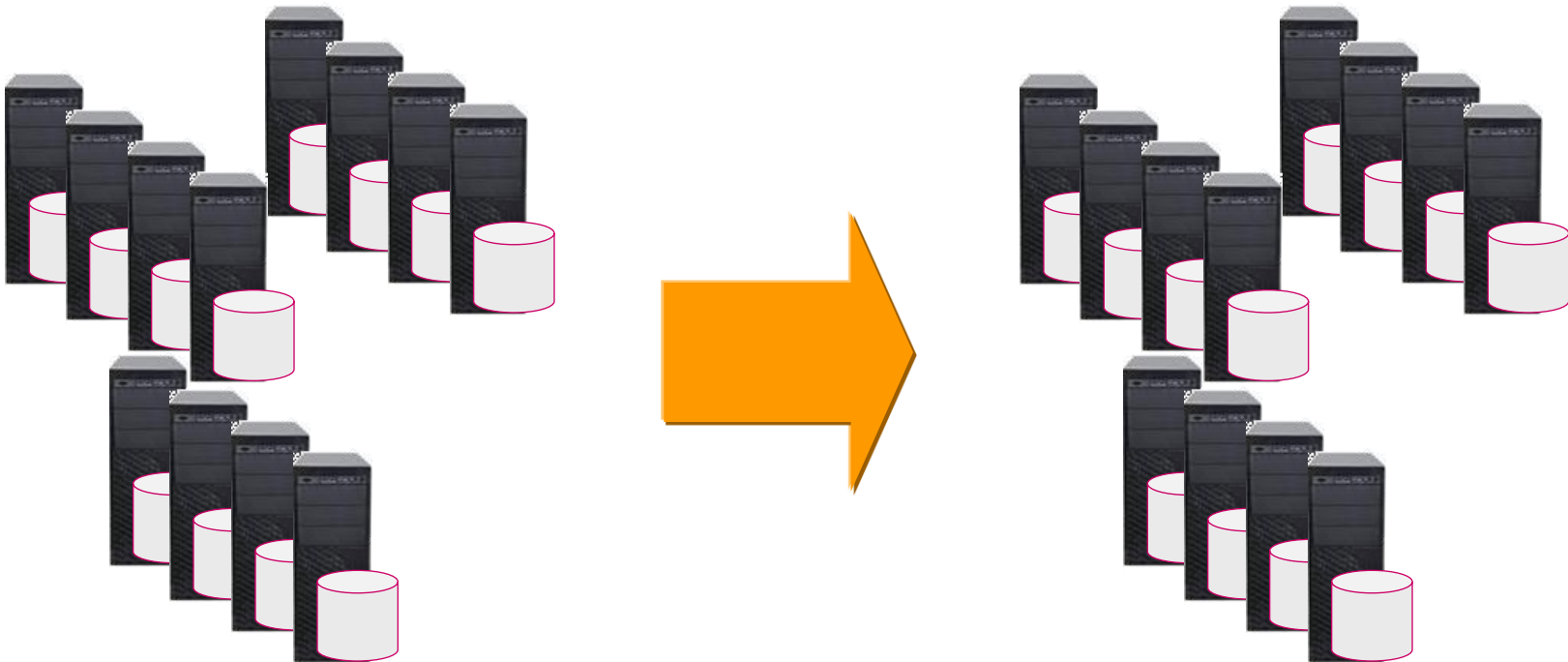
Consolidation Of Workloads On System z Simplifies Disaster Recovery

- Workloads are consolidated onto z/VM partitions as Linux guests

- Linux on System z can be failed over as part of GDPS
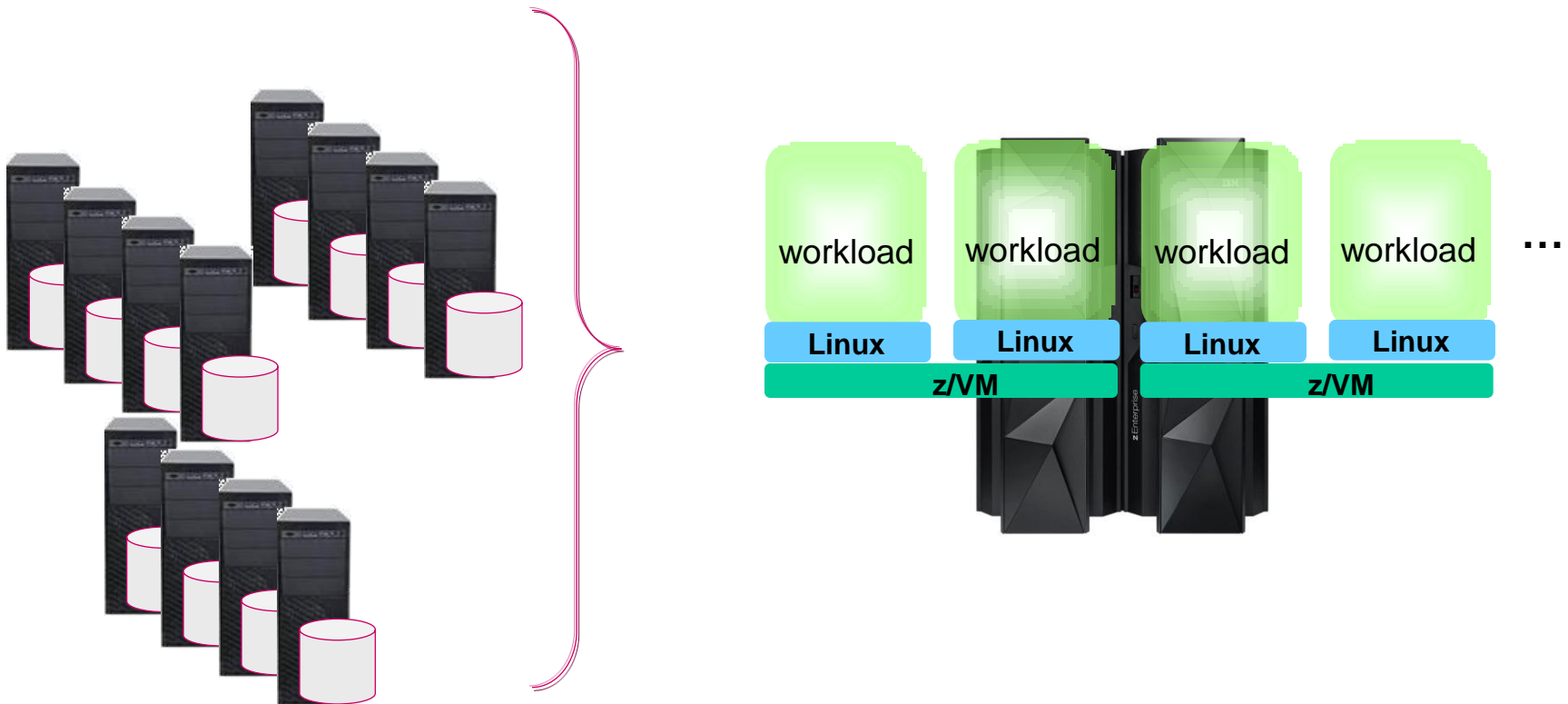
What System z Can Do That Intel Can't
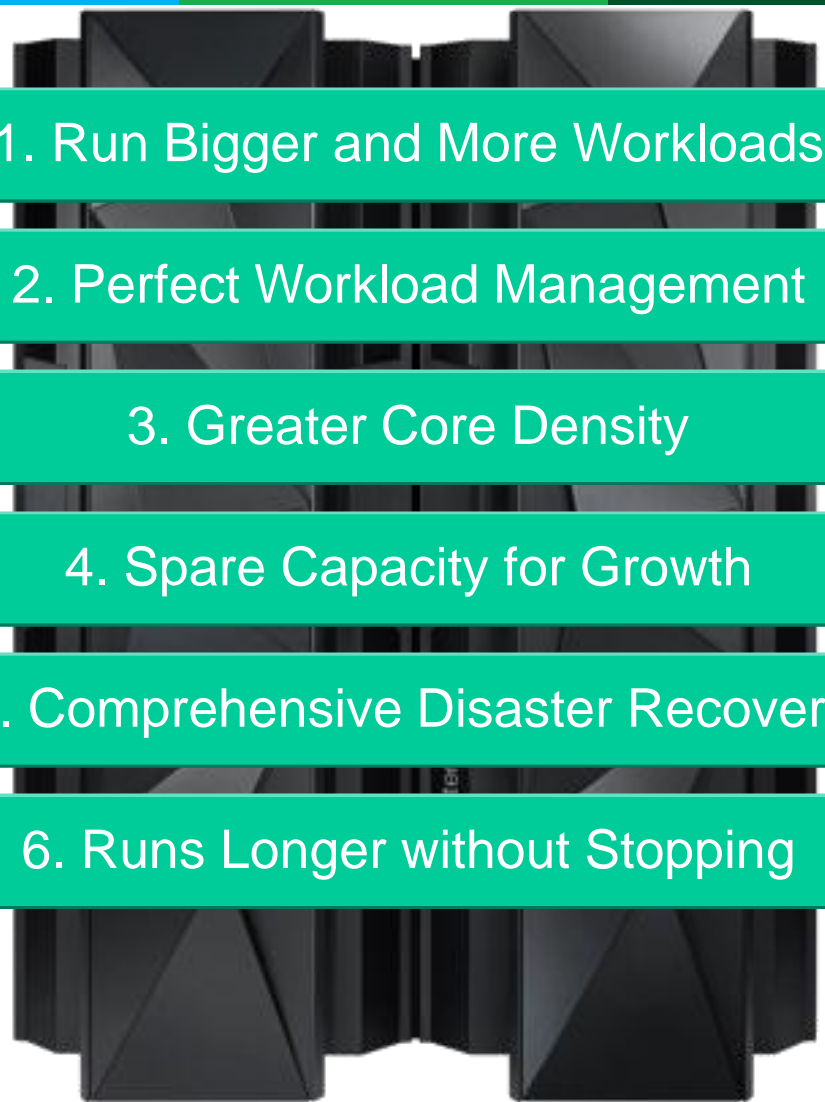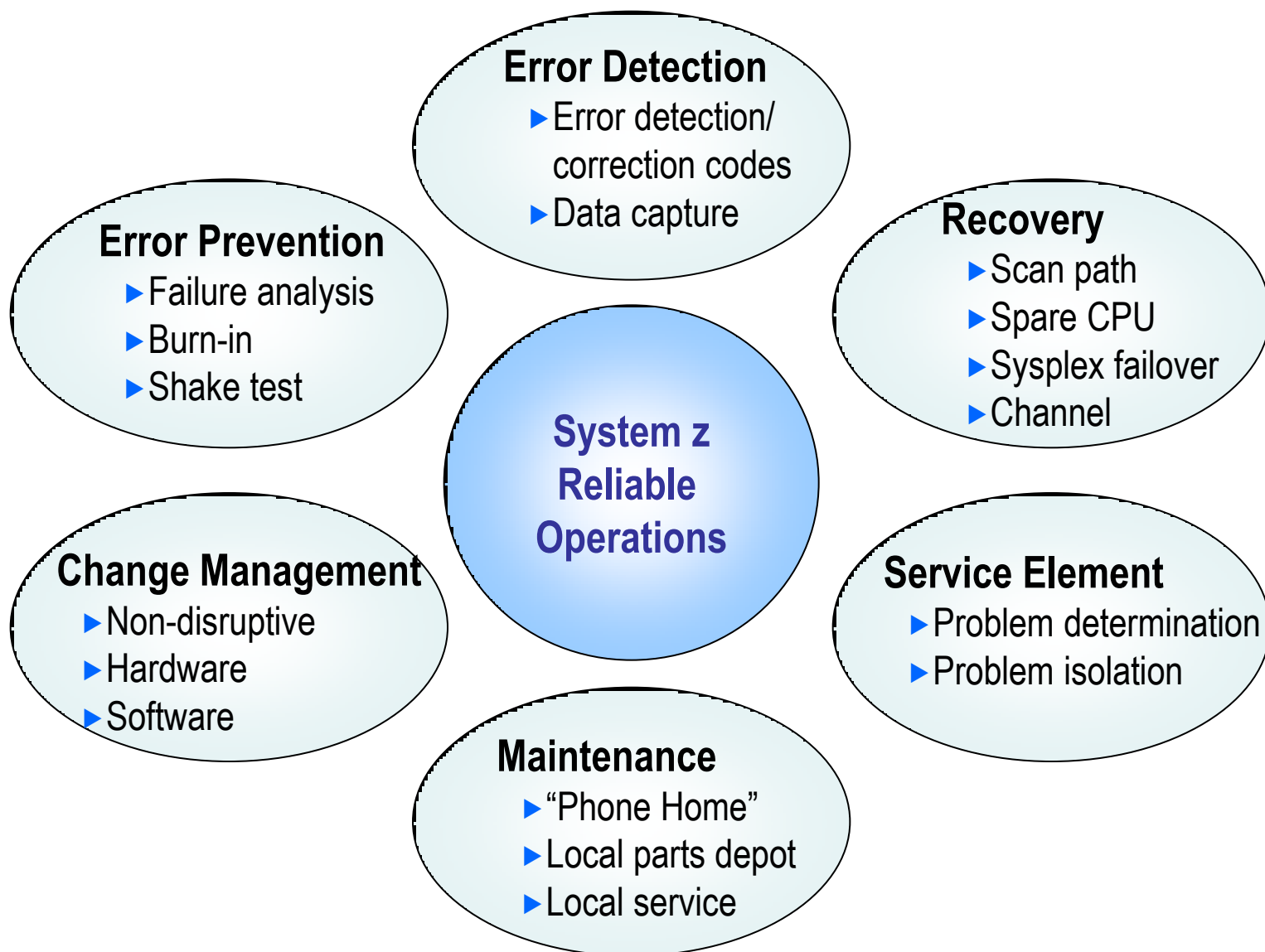
1. Run Bigger and More Workloads

2. Perfect Workload Management

3. Greater Core Density

4. Spare Capacity for Growth

5. Comprehensive Disaster Recovery

6. Runs Longer without Stopping

Intel Sandy Bridge

System z

System z Has More Comprehensive Protection To Ensure Better Availability Than Intel

Error Detection
- ▶ Error detection/ correction codes
- ▶ Data capture

Error Prevention
- ▶ Failure analysis
- ▶ Burn-in
- ▶ Shake test

Recovery
- ▶ Scan path
- ▶ Spare CPU
- ▶ Sysplex failover
- ▶ Channel

System z Reliable Operations

Change Management
- ▶ Non-disruptive
- ▶ Hardware
- ▶ Software

Service Element
- ▶ Problem determination
- ▶ Problem isolation

Maintenance
- ▶ "Phone Home"
- ▶ Local parts depot
- ▶ Local service

zEC12 Provides Transparent CPU Sparing

- Transparent sparing for all CPU types - CP/ICF/IFL/zAAP/zIIP

- zEC12 has 2 spare CPUs per server
  - Spares do not have to be local to the same book

- Processor Availability Facility (PAF) saves state & switches to spare CPU
  - System controllers can detect a failing processor chip
  - Status of the unit of workload running on the failing CPU can be saved
  - Failing CPU can be switched with the spare with NO interruption with the workload

OS workload
PR/SM partition

PU	CP4

CP Sparing Flow

1. CP4 fails
2. Error Detection
3. Save current state
4. Spare PU assigned as CP4
5. Error Recovery: restart application on new CP4

Dynamic SAP sparing/reassignment

•Upon failure, spare activated as new SAP

•No spare, active CP reassigned

Maintains I/O Performance

# System z Supports Concurrent Operations During Hardware Repair – *Intel Can't*

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

The Choice Is Clear!

**System z
is better than Intel
for Systems of Record**
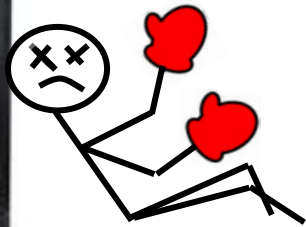
1. Run Bigger and More Workloads

2. Perfect Workload Management

3. Greater Core Density

4. Spare Capacity for Growth

5. Comprehensive Disaster Recovery

6. Runs Longer without Stopping
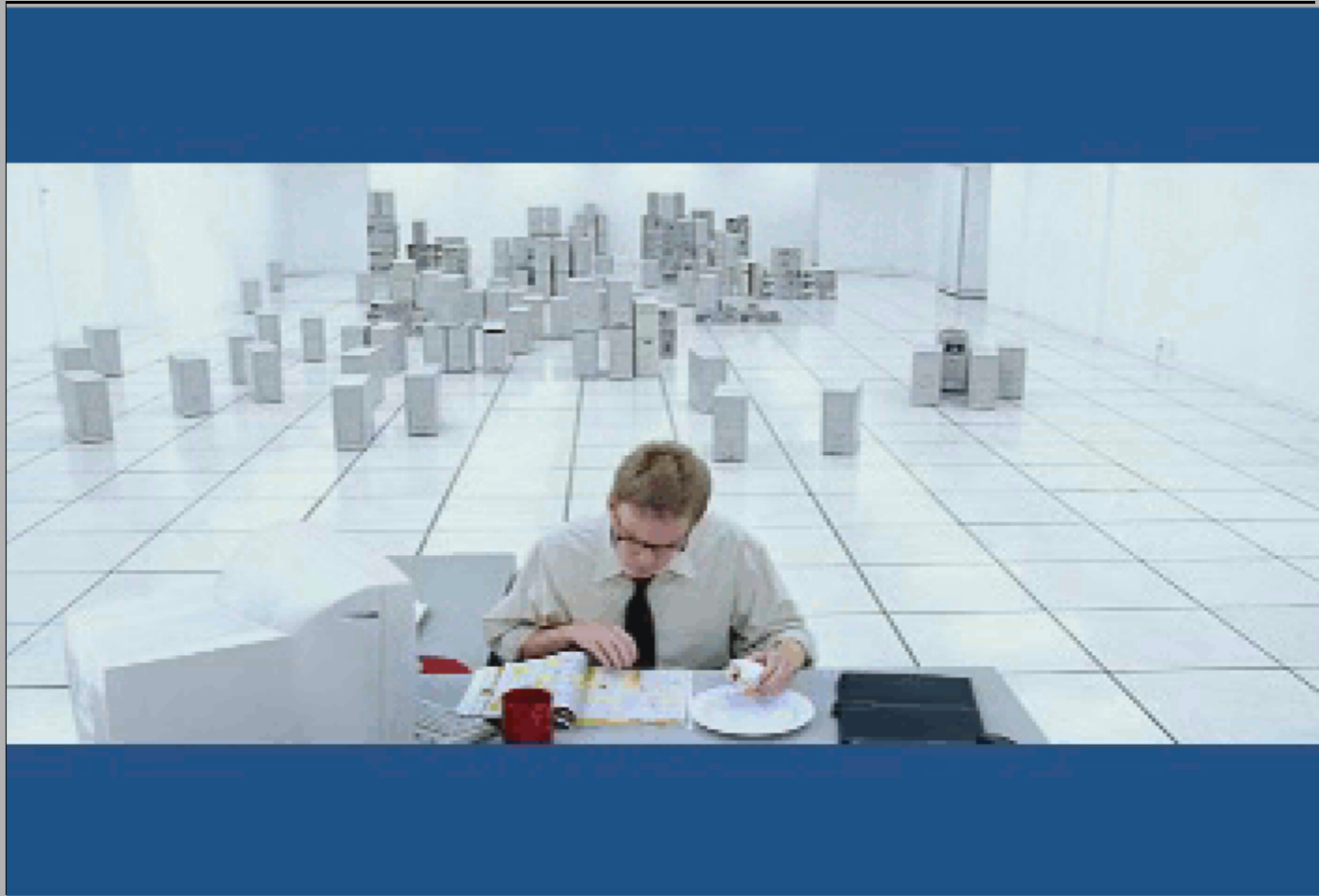
Notice Regarding Specialty Engines (e.g., zIIPs, zAAPs and IFLs):

Any information contained in this document regarding Specialty Engines ("SEs") and SE eligible workloads provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at
www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT").

No other workload processing is authorized for execution on an SE.

IBM offers SEs at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.