

**SHARE**

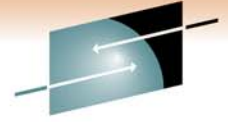
Technology • Connections • Results

# Workload Thinking for zEnterprise *Fit for Purpose!*

Mike Kott  
IBM  
SWG WW zITA Community Leader  
Los Angeles, CA  
[mkott@us.ibm.com](mailto:mkott@us.ibm.com)

March 1, 2011  
Session Number: 8937



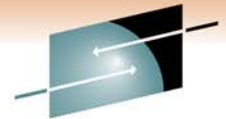


**SHARE**  
Technology • Connections • Results

# Agenda

- The Right Fit, Using the Right Tool
- TCA versus TCO
- Fit For Purpose – Many Factors to Consider
- Workloads & Platform Requirements
- zEnterprise Best Fit Workloads
- Recommended Next Steps

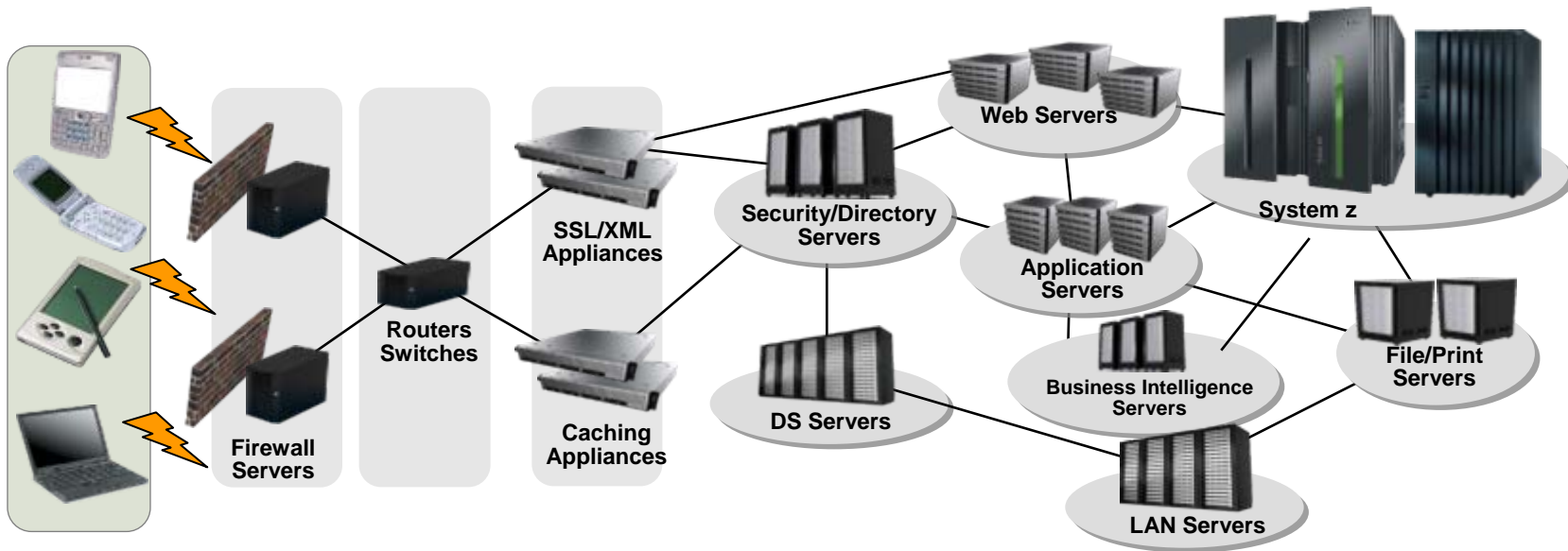
# Information Technology Today: Complex



**SHARE**

Results

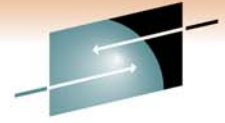
Information Technology today is limited by the technology and architecture configurations available



## Complexity and Limitations of Today's Environment

- **Many tiers/nodes of independent resources** connected over corporate network
- **System management** information typically **not end-to-end view**
- **Automation policies** are limited to tier/node boundaries
- **Redundancy** is pervasive for Operational staff, HW, Software and policies across architectures
- **Managing this complexity now consumes the majority of IT budgets**

# Platform Selection Is All about Using the Right Tool for the Right Job (Fit for Purpose)



**SHARE**  
Technology • Connections • Results



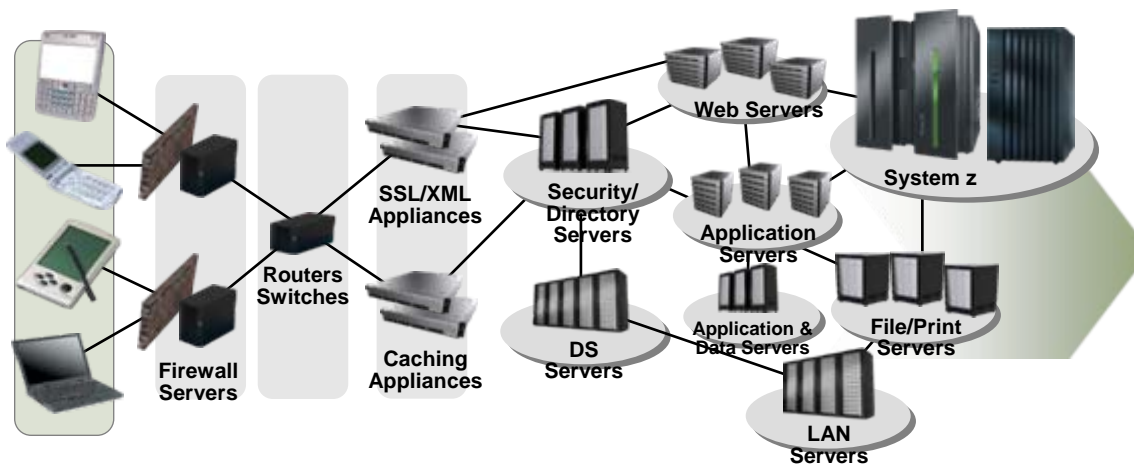
System x

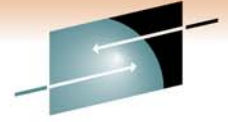


POWER7



System z





**SHARE**  
Technology • Connections • Results

# The Right Fit: Using the Right Tool for the Right Job

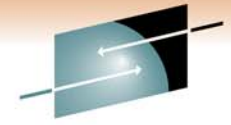


All of these “tools” can very quickly move a person from one place to another.

But, which one is the right tool to move one person?

One hundred people? Four hundred people?

# Purpose Built & Hybrid Computing

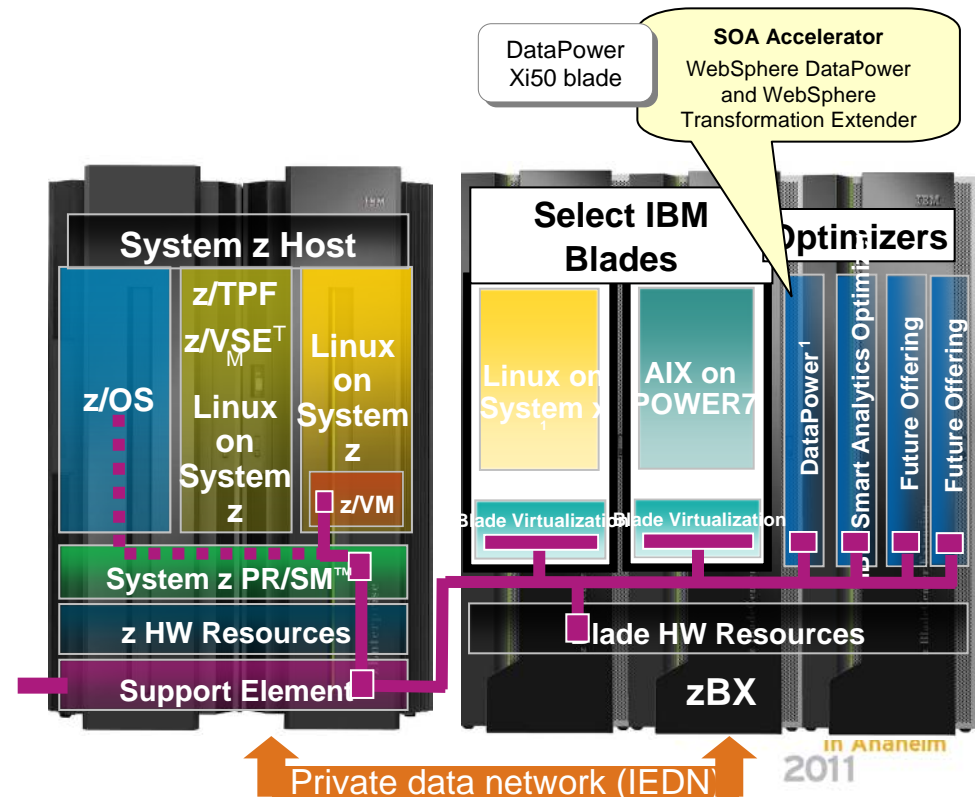


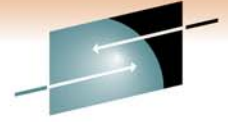
**SHARE**  
Technology • Connections • Results



- **Purpose-Built Computing** = certain kinds of workloads and applications are more appropriate to run on certain kinds of hardware

- **Now Hybrid Computing** = certain kinds of systems can work very well together for enterprise solutions





# How Do Companies Typically Select a Platform for Their Applications?



- Their first question is:

- “Will it run there?”



- Their second question is:

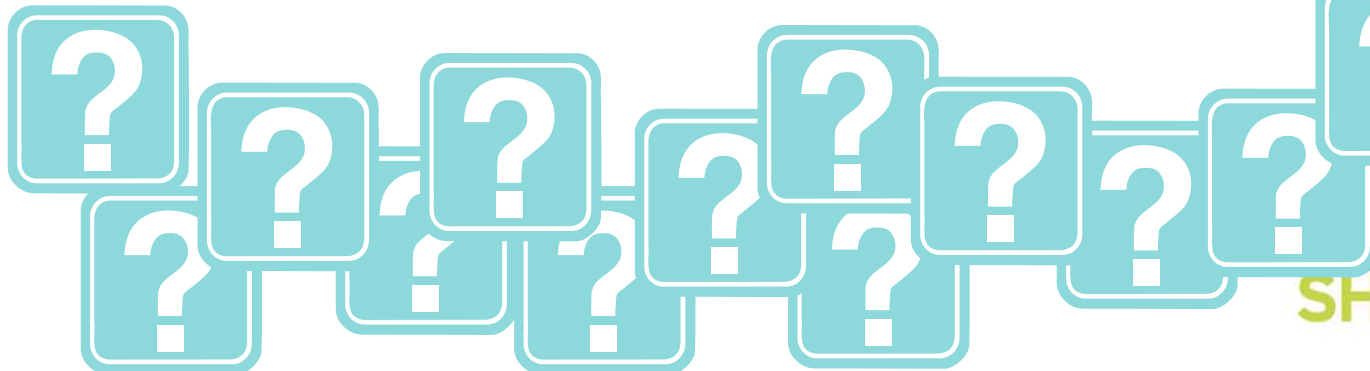
- “How much does the hardware cost?”

- They’re done!

- But this is just a TCA view...Is that all they should be thinking about?

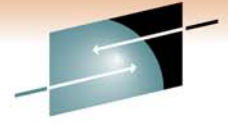
# What Did We Miss? Nonfunctional Requirements

- Shouldn't they have asked questions about:
  - Scalability? Availability? Backup? Site disaster recovery?
  - Security? Reliability? Data integrity? Maintainability?
  - Volumes and service levels?
  - Floor space? Power? Cooling?
  - Operations? Scheduling? Monitoring? Server management?
  - Integration? Performance and value of data proximity?
- Questions that lead to a more complete TCO view must be considered...





# TCA versus TCO



**SHARE**  
Technology • Connections • Results

- TCA is basically the purchase price
- TCO is the cost associated with the item purchased, the use of it, and the maintenance of it
- TCO deals with costs over time whereas TCA deals with upfront costs only

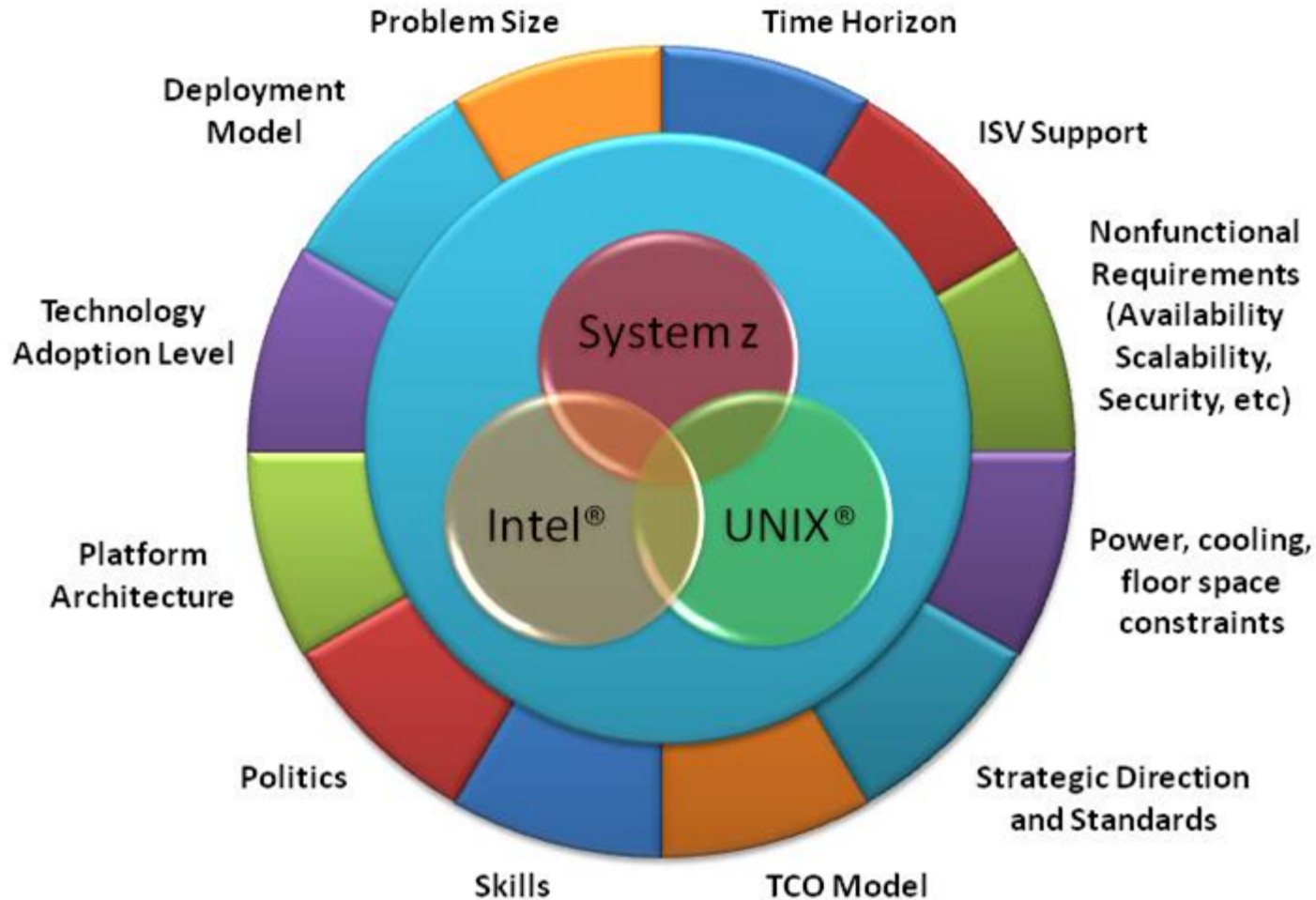


- Cost categories
  - Hardware, Software, People, Network, Storage, Facilities
- Environment
  - Production, Development, QA, Test, Disaster Recovery
- Time
  - Growth, Refresh (or upgrades), Events
- Quality of Services
  - Availability, Resiliency, Security, Manageability, Scalability, Usability, Maintainability, Extensibility



# Platform Choice: Fit for Purpose

Many factors influence platform selection, making it difficult to present a simple selection matrix.



Some factors are specific to each business; others are common to all and can be generalized.



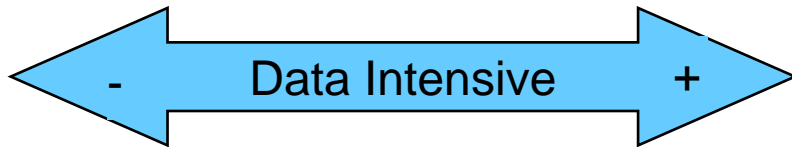
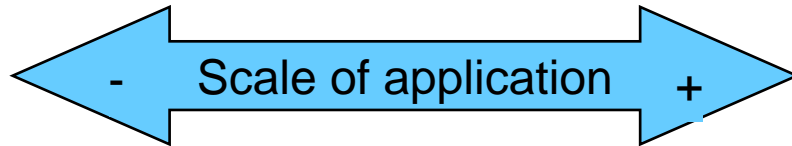
# What Quality of Service Do Your Applications Require?

- **What service level agreements are in place?**
- **What level security does the specific application implementation require?**
- **What response times are required?**
- **How much network latency can be tolerated?**
- **Is 24 x 7 x 365 availability required, or do systems just need to be up during business hours?**
- **What are business hours?**
- **Is the operation local or global?**
- **How much planned and unplanned down time can your organization tolerate?**
- **How much “head room” is needed for scaling, and how does related traffic or demand vary?**
- **Does the solution require encryption at all levels?**

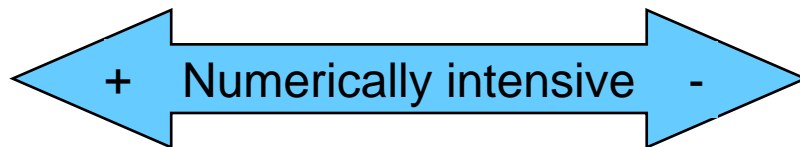
# Choosing the Right System and O/S

## Hardware Considerations

Intel      UNIX      System z



UNIX/RISC      Intel      System z



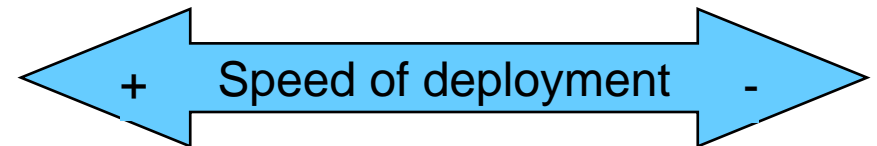
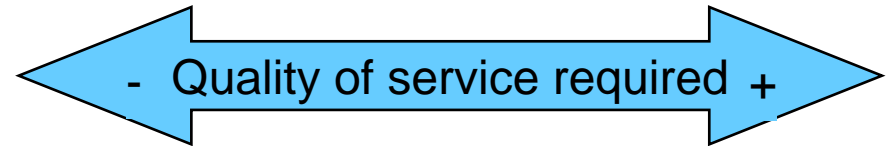
## Software/OS Considerations

Linux on System z

z/OS

AIX in zBX

Linux on zBX

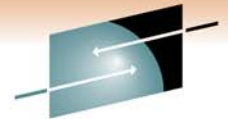


# High Level Workload Definition

- **Workloads** are a combination of:
  - **Application function:** What it does and how it does it
  - **Data structure:** Data residency, topology, access model
  - **Usage pattern:** Utilization profile over time, mix of use cases
  - **Service level:** Non-functional requirements
  - **Integration:** Interaction between application & data components
- *The workload requirements will create varying demands when determining server alternatives*

cal [ˈdɑːbɒlɪkəl] adj ① (evil) teuflisch  
[ˈdɑːrəɡnəʊz] vr ① mit d. ...  
[ˈdɑːrəɡnəʊz] vr ① mit d. ...

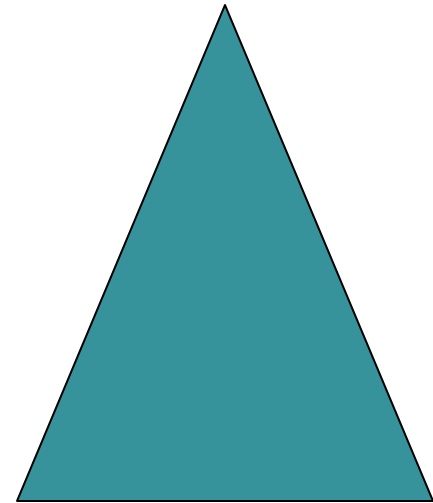
dictatorship [dɪk'tetərɪʃɪp] n ① (a. fig) Diktatur  
dictionary ['dɪkʃənərɪəri] n ① (a. fig) Wörterbuch  
diffraction [dɪfræk'tʃən] n ① (a. fig) Beugung



# What about workload optimization?

- There are strong tradeoffs that result from variability in usage patterns, parallelism in the application/data and the service level.
  - **You can't fully exploit parallelism and virtualization at the same time.**
- Machines have fitness for parallelism, fitness for data handling and fitness for serialization
  - **You cannot maximize all three in the same design**

**Service Level**  
(response time)  
Serial Fitness

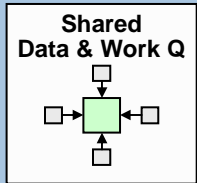
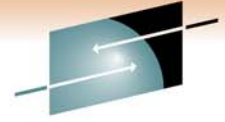


**Throughput**  
(concurrent threads)  
Parallel Fitness

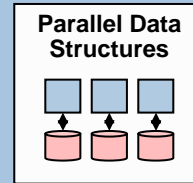
**Efficiency**  
(utilization)  
Data Fitness

If you **drive up resource sharing** to gain utilization you **reduce parallelism** and **increase response time**. If the load is highly variable, **adding parallelism will reduce the utilization** dramatically. Note that **integration increases sharing**, creating serializations and **impacting service levels**.

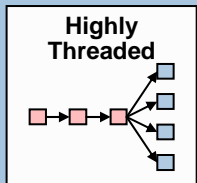
# Workload Architectures



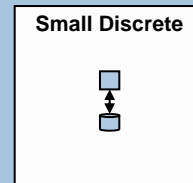
- Shared data and work queues



- Parallel data structures

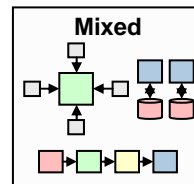


- Highly threaded



- Small discrete applications

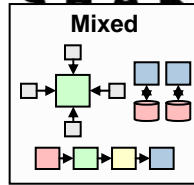
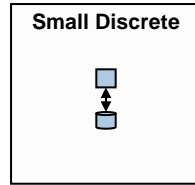
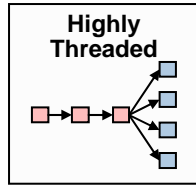
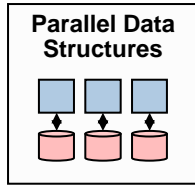
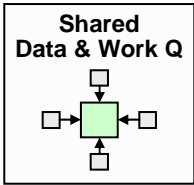
**Mixed**



# Workload Characteristics and Platform Requirements



**SHARE**  
sults



**Examples**

- OLTP databases
- N-Tier transaction processing

- Structured BI
- XML parsing
- HPC applications

Web app servers  
SAP app servers

- HTTP, FTP, DNS
- File and print
- Small end user apps

- z/OS and IBM i
- Hypervisors with virtual guests, WPAR

**Characteristics**

- Thread interaction raises contention & coherence delays
- Coherency traffic increases memory & cache bus utilization
- High context switch rates

- Low thread interaction
- High memory bandwidth
- Low context switch rates

Lots of software threads  
Modest thread interaction

- Does not pressure any resource
- Requires minimal memory footprint
- Inefficient on dedicated resources
- No shared data

- Different SLAs
- Varying sizes and number of threads
- May be N-Tier or independent
- Variable context switch rates

**Platform Considerations**

- Scale on robust SMP
- Cluster technology dependent
- Large shared caches and wide busses
- Fewer, bigger threads

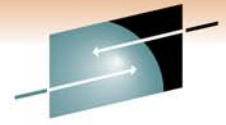
- Scale well on clusters
- Large private caches
- High thread count
- High memory and I/O bandwidth
- Often on dedicated machines

Scale on large SMP  
Can scale on clusters  
High thread count  
Large number of memory busses  
Large private caches

- Scale on almost any hardware
- Ripe for virtualization
- Can exist on low cost hardware

- Scale on robust SMP
- High internal bandwidth
- Thread speed and number is workload dependent
- Large, close caches
- High memory bandwidth





# zEnterprise Provides the Foundation for the “Smart” Infrastructure on Which We Can Build the Workloads of Today and Tomorrow

*They are workloads that.....*

- *Rely on data serving and application components on System z*
- *Solutions that need to leverage strengths of System z... Security, Reliability, Availability.*
- *Have application components on Power or x86 but require a higher level of integration capabilities and efficiency*



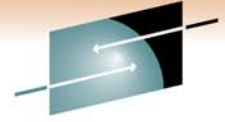
*....and / or.....*

- *Reside in low utilization / development environments*
- *Can be made more efficient through consolidation*
- *Can be optimized by using the newest virtualization technology*

*....but also may.....*

- *Reside in complex multi-platform IT environments*
- *Require flexible development and test infrastructure*
- *Require simplified, integrated policy and management*

# Workloads we see every day that match these characteristics



**SHARE**  
Technology • Connections • Results

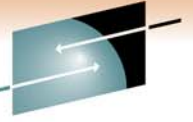
## What is a workload?

The relationship between a **group** applications and/or systems related across several business functions to satisfy one or more business processes.

				
<b>Banking</b>	<b>Insurance</b>	<b>Retail</b>	<b>Healthcare</b>	<b>Public Sector</b>
<i>Core Banking</i>	<i>Internet Rate Quotes</i>	<i>On-line Catalog</i>	<i>Patient Care Systems</i>	<i>Electronic IRS</i>
<i>Wholesale Banking – Payments</i>	<i>Policy Sales &amp; Management (e.g. Life, Annuity, Auto)</i>	<i>Supply Chain Management</i>	<i>On-line Claims Submission &amp; Payments</i>	<i>Web-based Social Security</i>
<i>Customer Care &amp; Insight</i>	<i>Claims Processing</i>	<i>Customer Analysis</i>		

**SHARE**  
in Anaheim  
2011

# These workloads have recognizable patterns



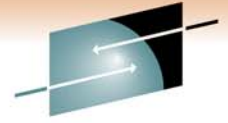
Core Applications	
<b>Database (z)</b> • DB2® for z/OS®, IMS™	<b>Database (z)</b> • DB2 for z/OS • Oracle on Linux for z
<b>Application (z)</b> • CICS® • COBOL • WebSphere®	<b>Application (z)</b> • WebSphere

SAP	
<b>Database (z)</b> • DB2 for z/OS	<b>Database (z)</b> • DB2 for z/OS
<b>Application (z)</b> • Linux® for z	<b>Application (x86)</b> • Linux for x86
<b>Database (z)</b> • DB2 for z/OS	
<b>Application (Power)</b> • AIX®	

Multi-Tier Web Serving	
<b>Database (z)</b> • DB2 for z/OS	<b>Database (z)</b> • DB2 for z/OS
<b>Application (z)</b> • WebSphere	<b>Application (Power / UNIX)</b> • WebSphere • JBoss
<b>Application (x86)</b> • WebSphere • Apache / Tomcat	
<b>Database (z)</b> • DB2 for z/OS, IMS	<b>Database (z)</b> • DB2 for z/OS or IMS
<b>Transaction Processing (z)</b> • CICS, MQ	<b>Application (Power / UNIX)</b> • WebSphere • JBoss
<b>Application (Power / UNIX)</b> • WebSphere • JBoss • WebLogic	<b>Presentation (x86)</b> • WebSphere • Apache / Tomcat
<b>Presentation (z)</b> • WebSphere • Windows	

Data Warehouse & Analytics	
<b>Master Data Management</b> <b>Database (z)</b> • DB2 for z/OS <b>Application (z)</b> • WebSphere MDM (AIX, Linux on z)	
<b>Analytics</b> • System z/OS • DB2 • Cognos® • SAS • Linux for System z • Cognos • SPSS • InfoSphere™ Warehouse	

# Workload Attributes and Market Segmentation



**SHARE**  
Technology • Connections • Results

## Transaction Processing and Database



- High Transaction Rates
- High Quality of Service
- Peak Workloads
- Resiliency and Security

## Analytics and High Performance



- Compute or I/O intensive
- High memory bandwidth
- Floating point
- Scale out capable

## Business Applications



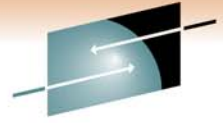
- Scale
- High Quality of Service
- Large memory footprint
- Responsive infrastructure

## Web, Collaboration and Infrastructure



- Highly threaded
- Throughput-oriented
- Scale out capable
- Lower Quality of Service

# Fit for Purpose Virtualization



**SHARE**  
Technology • Connections • Results

## IBM zEnterprise System Provisioning with WebSphere CloudBurst Appliance



WebSphere Cloudburst Appliance dispenses WebSphere Application Server Hypervisor Edition Servers into a set of other machines

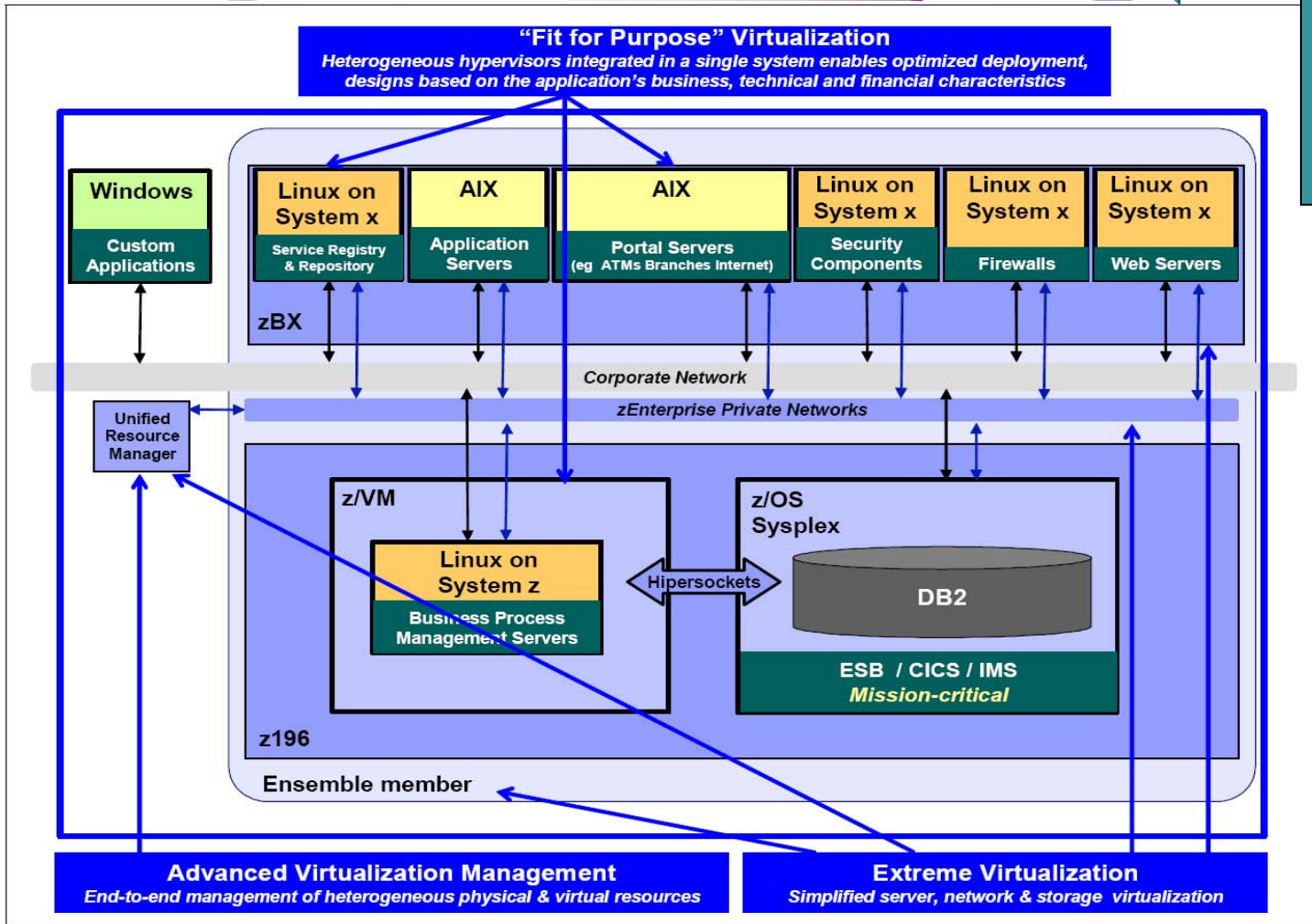
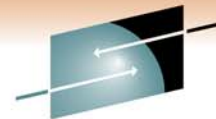


Figure 9 IBM zEnterprise virtualization: real business value for today’s heterogeneous applications

# Watson answers a grand challenge - Jeopardy



**SHARE**  
Technology • Connections • Results

Watson, named after IBM founder Thomas J. Watson, was developed by a team of IBM scientists who set out to accomplish a grand challenge – build a computing system that rivals a human’s ability to answer questions posed in natural language with speed, accuracy and confidence.

On February 14, 2011 Watson challenged *Jeopardy!* world champions Ken Jennings and Brad Rutter to a two-match contest aired over three consecutive nights- and **WON!**



**Fit for Purpose:** The software behind Watson can run on a variety of hardware platforms but in this instance has been deployed through an IBM POWER7 platform, which is especially **well-suited** to Watson. The server is optimized to handle the massive number of tasks across 2,880 high performance cores. The combination of massive parallelism, on-chip bandwidth, and memory capacity, combined with the ability to optimize application performance, **makes POWER7 ideal** for running complex analytics workloads.



## SW products:

- UIMA IBM Content Analytics
- InfoSphere Warehouse DB2, Informix, Netezza BA appliance
- Business Analytics
- Enterprise Content Mgmt solutions



# zEnterprise

*A system of systems that unifies IT for optimized service delivery*



## Unified management for a smarter system: **zEnterprise Unified Resource Manager**

Scale up with your business!  
The world's fastest and most  
scalable enterprise system:  
**IBM zEnterprise™ 196 (z196)**

- Unifies management of resources, extending IBM System z qualities of service end-to-end across workloads
- Part of the IBM System Director family, provides platform, hardware and workload management

Scale out to a trillion  
instructions per second:  
**zEnterprise BladeCenter  
Extension (zBX)**

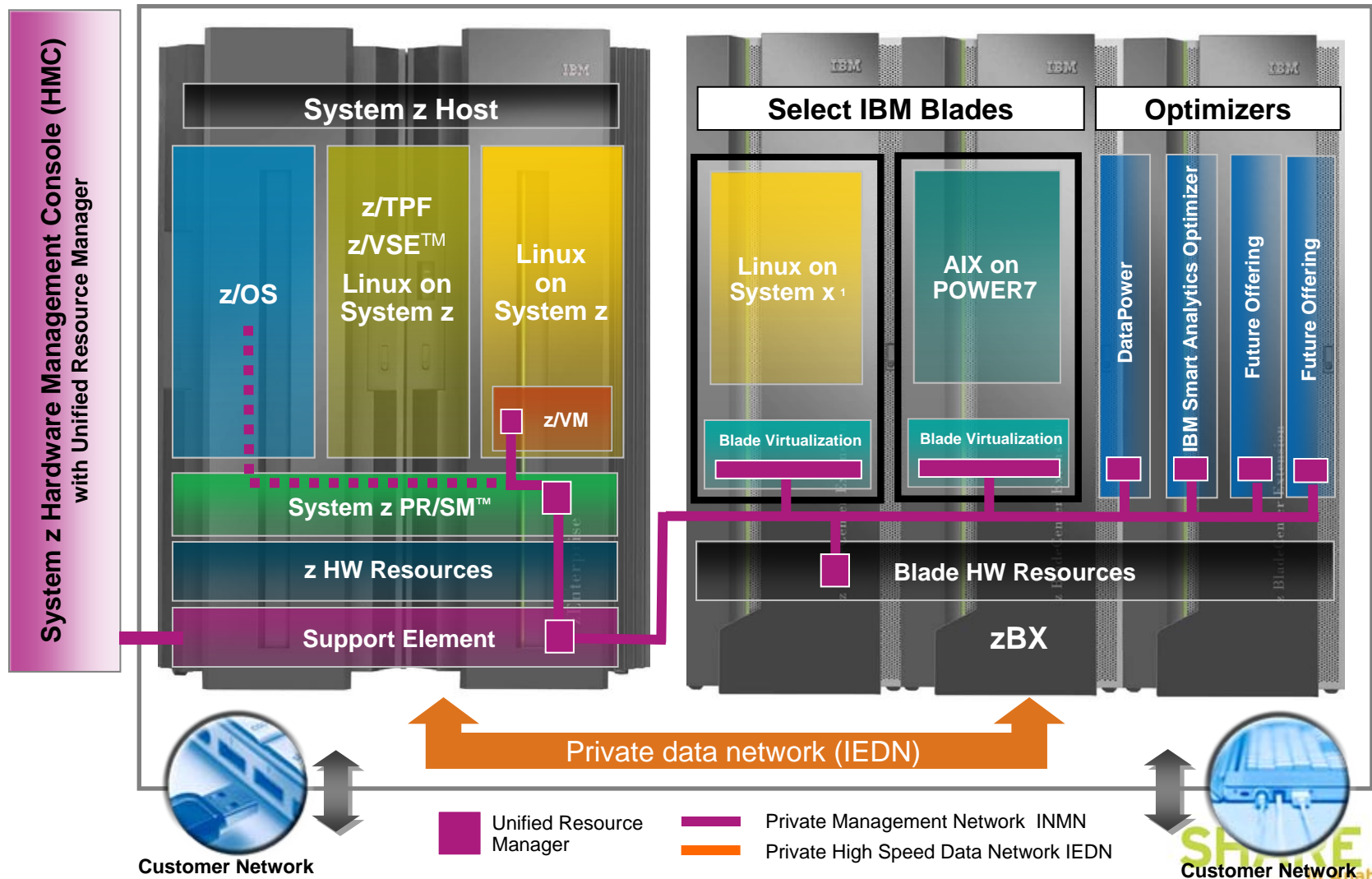
- Optimized to host large scale database, transaction, and mission critical applications
- The Most efficient platform for Large-scale Linux® consolidation
- Capable of massive scale up
- New easy to use z/OS® V1.12



- Selected IBM POWER7™ blades and IBM System x® Blades\* for tens of thousands of AIX® and Linux applications
- High performance optimizers and appliances to accelerate time to insight and reduce cost
- Dedicated high performance private network

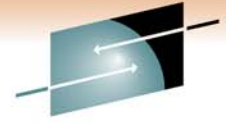
# A look inside the IBM zEnterprise System

## Enabling a new dimension in application architecture



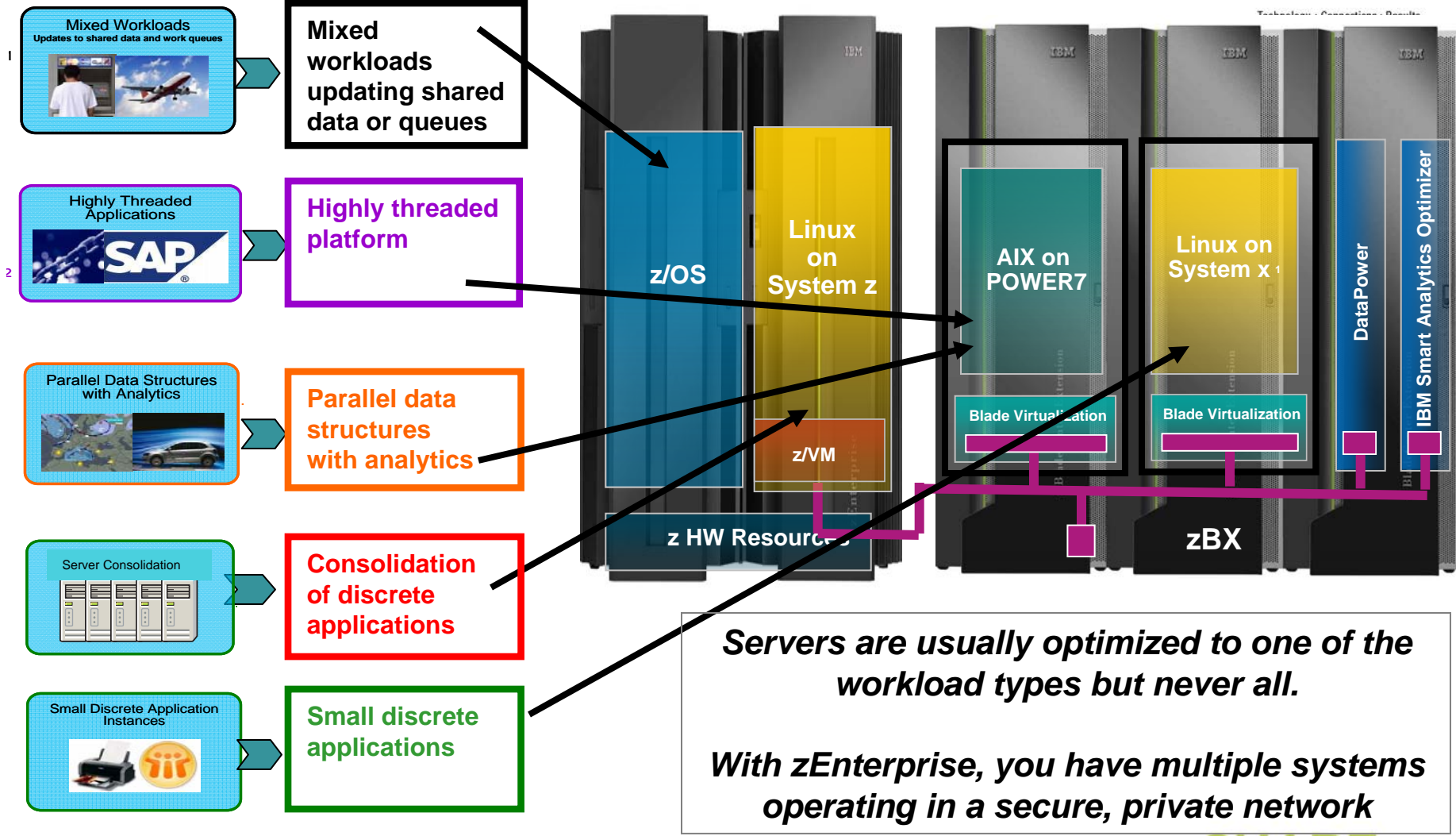


# Workload harmony with zEnterprise – a “right-fit” decision



**SHARE**

Technology • Cooperation • Results



# Cost Per Workload





## Metric To Compare True Value

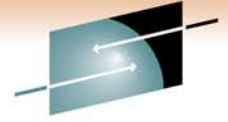
**Cost Of Hardware Alone  
Is Misleading**

**True Value:  
Cost Per Workload**



	Cost of Hardware (servers + infrastructure)	Cost of Software	Number of Workloads	Cost per Workload
 <p><b>56 Intel Blades</b> (8 cores per blade) <b>448 cores total</b></p>	<b>\$1.7M</b>	<b>\$9.8M</b>	500	<b>\$23.0K</b>
 <p><b>1 zEnterprise</b> <b>192 cores total</b> (z196 + zBX cores)</p>	<b>\$2.7M</b>	<b>\$4.7M</b>	500	<b>\$15.0K</b>

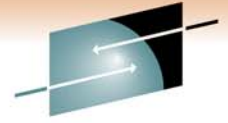
# What Makes the Best Fit Workload for System z



**SHARE**  
Technology • Connections • Results

- Leverage classic strengths of the System z
  - High availability
  - High i/o bandwidth capabilities
  - Flexibility to run disparate workloads concurrently
  - Requirement for excellent disaster recovery capabilities
  - Security
- Shortening end to end path length for applications
  - Collocation of applications
  - Consolidation of applications from distributed servers
  - Reduction in network traffic
  - Simplification of support model
- Consolidation Effect
  - Power requirements
  - Software costs
  - People Costs
  - Real Estate
  - Workloads requiring EXTREME Flexibility

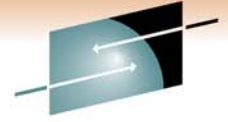
# Best Fit Workloads for System z



**SHARE**  
Technology • Connections • Results

- OLTP
- Database
- ERP
- Batch
- CRM
- Data Warehousing/Data Mart
- Applications requiring top end disaster recovery model
- Linux on z
  - WebSphere MQ Series
  - DB2 Connect
  - CICS Transaction Gateway , IMS Connect for Java
  - SAP
  - WebSphere Portal, Process Server and JAVA applications development and hosting
  - Lotus Notes, Sametime, Quickr, Connections
  - Network Infrastructure, FTP, NFS, DNS etc..,
  - Oracle Database
  - Virtualization and Security Services
  - BI Applications, Cognos, InfoSphere, SPSS, FileNet

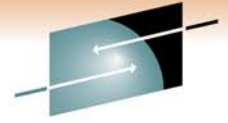
# What Makes a Good Fit Workload for System z



**SHARE**  
Technology • Connections • Results

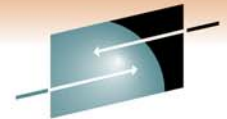
- Evaluate server choices
  - Correct application availability
  - Supporting applications
  - Total Cost of Ownership (TCO)
  - **Politics** within the organization.
  - Porting issues
- Shortening end to end path length for applications
  - Collocation of applications
  - Consolidation of applications from distributed servers
  - Reduction in network traffic
  - Simplification of support model
- Consolidation Effect
  - Power requirements
  - Software costs
  - People Costs
  - Real Estate
  - Workloads requiring EXTREME flexibility

# Good Fit Workloads for System z



**SHARE**  
Technology • Connections • Results

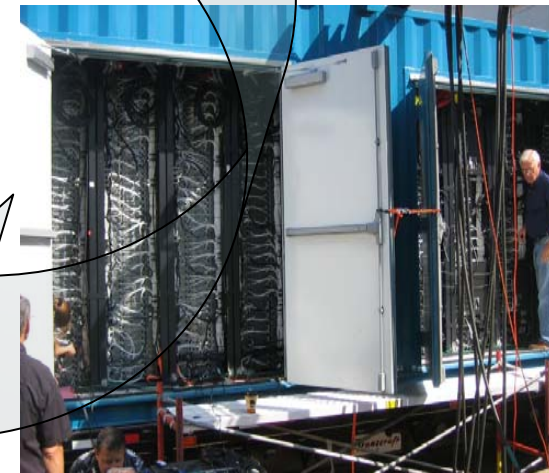
- Security
- Systems Management
- Networking
- Streaming Media
- Linux on z
  - UDB LUW
  - Informix, (IDS)
  - Apache web serving
  - SAMBA
  - TIM/TAM (LDAP Services)
  - TSM
  - Existing Linux Workloads on Distributed



**S H A R E**

# What Makes A Best Fit Workload for Linux on System z?

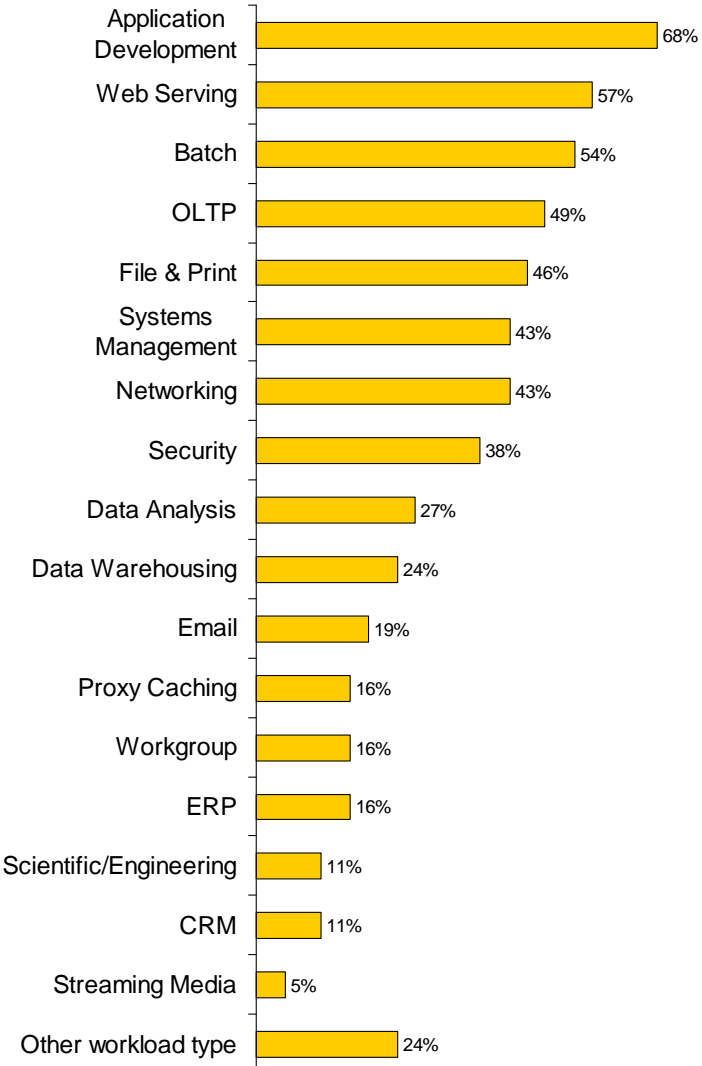
- Leverage classic strengths of the System z
  - High availability
  - High i/o bandwidth capabilities
  - Flexibility to run disparate workloads concurrently
  - Requirement for excellent disaster recovery capabilities
  - Security
- Shortening end to end path length for applications
  - Co-location of applications
  - Consolidation of applications from distributed servers
  - Reduction in network traffic
  - Simplification of support model
- Consolidation Effect
  - Power requirements
  - Software costs
  - People Costs
  - Real Estate
  - Workloads requiring **EXTREME Flexibility**



**“Do More with Less”**

**SHARE**  
in Anaheim  
2011

# What are Linux users running on System z?



## Surveys indicate customers use:

- Web Serving
- Data Services
- Web Application Serving
- Systems Development

## Best Fit Workloads for Linux on System z:

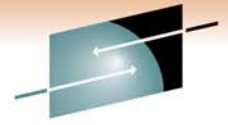
- **Web Application Servers:** WebSphere Application Server
- **Email and collaboration:** Domino, Web 2.0
- **Data services:** Cognos, DB2, Oracle, Informix, Information Server, Information Builders WebFOCUS
- **Business critical ISV applications:** e.g., SAP
- **Development of WebSphere and Java applications**
- **Virtualization and security services**
- **Business connectors:** WebSphere MQSeries, DB2 Connect, CICS Transaction Gateway, IMS Connect for Java
- **Network Infrastructure:** FTP, NFS, DNS, etc., and Communications Server and Communications Controller for Linux, CommuniGate Pro (VoIP)
- **Applications requiring top end disaster recovery model**



# Numerous Total Cost of Ownership Studies

*IBM Will Work with You to Determine the Most Appropriate for You*

- Eagle Study
- Scorpion Study
- RacezOS Study
- Rapid Workload Optimization Assessment
- IBM Migration Factory
- Server Optimization and Integration Services

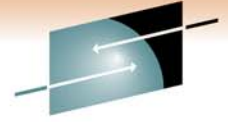


**S H A R E**

Results



2011



**SHARE**  
Technology • Connections • Results

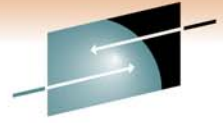
# Numerous Fit for Purpose Workshops

*IBM Will Work with You to Determine the Most Appropriate for You*

- 1 Day F4P Workshop Summary
- Multi-Day Detailed F4P Workshop
- F4P Strategy Services

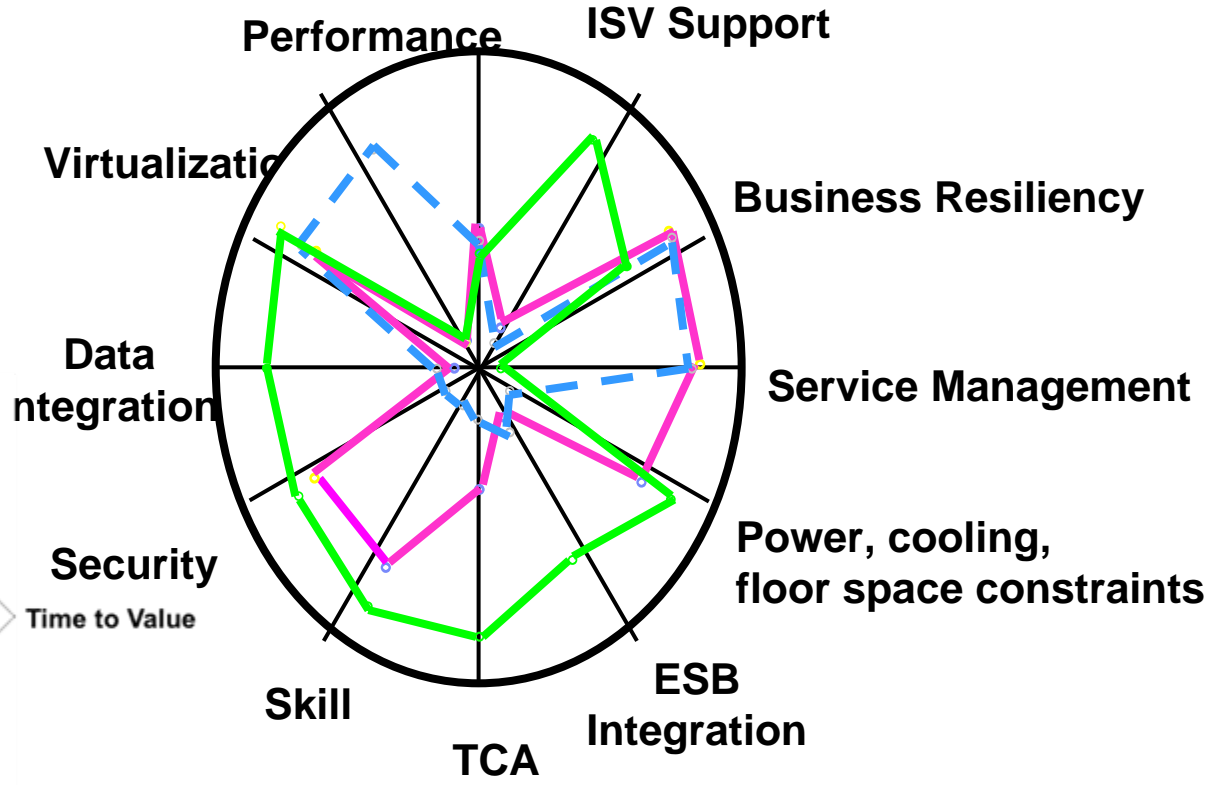


# Numerous Tools & Methodologies to Measure Workload Requirements



**SHARE**  
Technology • Connections • Results

**Radar charts**  
*Business Value Assessment Tool*



**Spider charts**

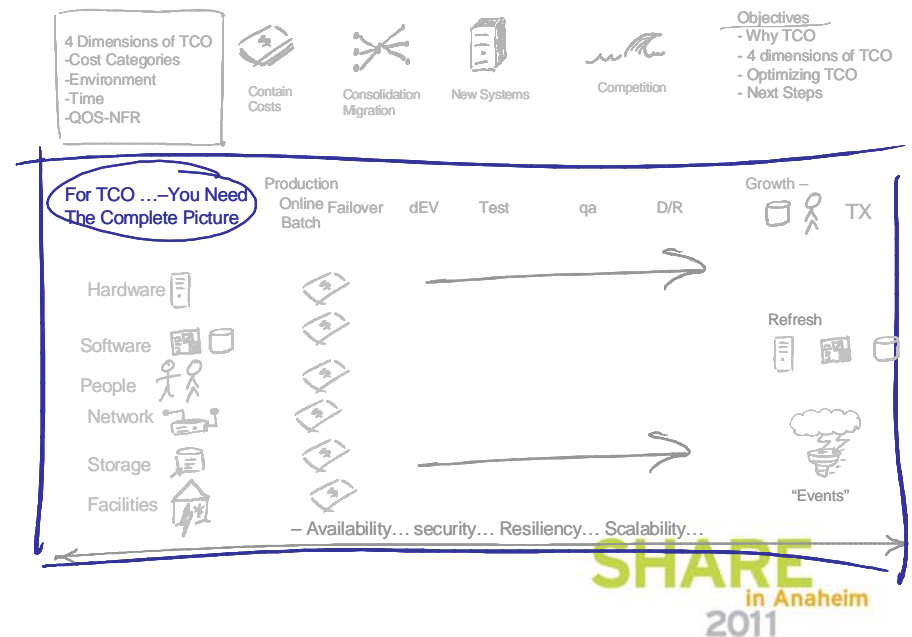
# Summary of Key Points

- Many factors influence platform selection – a simple matrix does not exist
- Local factors affect platform selection
- Infrastructure size matters
- Each deployment model has its place – virtualize or centralize where possible
- There is no single platform or middleware capacity metric
- Larger servers offer virtualization advantages
- Non-functional requirements are the significant element of platform selection
- Select platforms based upon workload requirements not middleware
- An enterprise wide view provides the best optimization opportunity
- The choice of cost and value elements, along with time horizon, can dictate which platform is considered the lowest cost
- Cost models have different purposes – use the right one for the job



## Recommended Next Steps

- Measure real costs of your workloads - know the difference between TCA and TCO
- Conduct a Fit for Purpose Workshop to evaluate your critical workloads
- Select specific workload and conduct a TCO Study to define true cost of ownership and Fit for Purpose platforms



## Additional Information

- Visit IBM *Booth # 213* at the SHARE Technology Expo to see the zEnterprise in action
- zEnterprise information on Fit for Purpose, TCO Studies, etc. on [www.ibm.com](http://www.ibm.com)

<http://www-03.ibm.com/systems/z/hardware/zenterprise/index.html>

- Contact your IBM Sales Rep, or contact me to get more information on Fit for Purpose, TCO, Workshops, Studies  
[mkott@us.ibm.com](mailto:mkott@us.ibm.com)