Workload Thinking for zEnterprise
Fit for Purpose!

Mike Kott
IBM
SWG WW z Client Architect Community Leader
Los Angeles, CA
mkott@us.ibm.com

August 10, 2011
Session Number: **08937**

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

AIX*	FICON*	IBM*	POWER*	VSE/ESA	z/OS*
BladeCenter*	GDPS*	IBM eServer	POWER7*	WebSphere*	zSeries*
CICS*	HiperSockets	IBM (logo)*	PR/SM	XIV*	z/VM*
DataPower*	IBM*	IMS	System x*	z9*	z/VSE*
DB2*	IBM eServer	Parallel Sysplex*	System z*	z10 BC	
DFSMS	IBM (logo)*	POWER*	System z9*	z10 EC	
DS8000*	IMS	POWER7*	System z10*	zEnterprise	
ESCON*	Parallel Sysplex*		System z10 Business Class		

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
InfiniBand is a trademark and service mark of the InfiniBand Trade Association.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

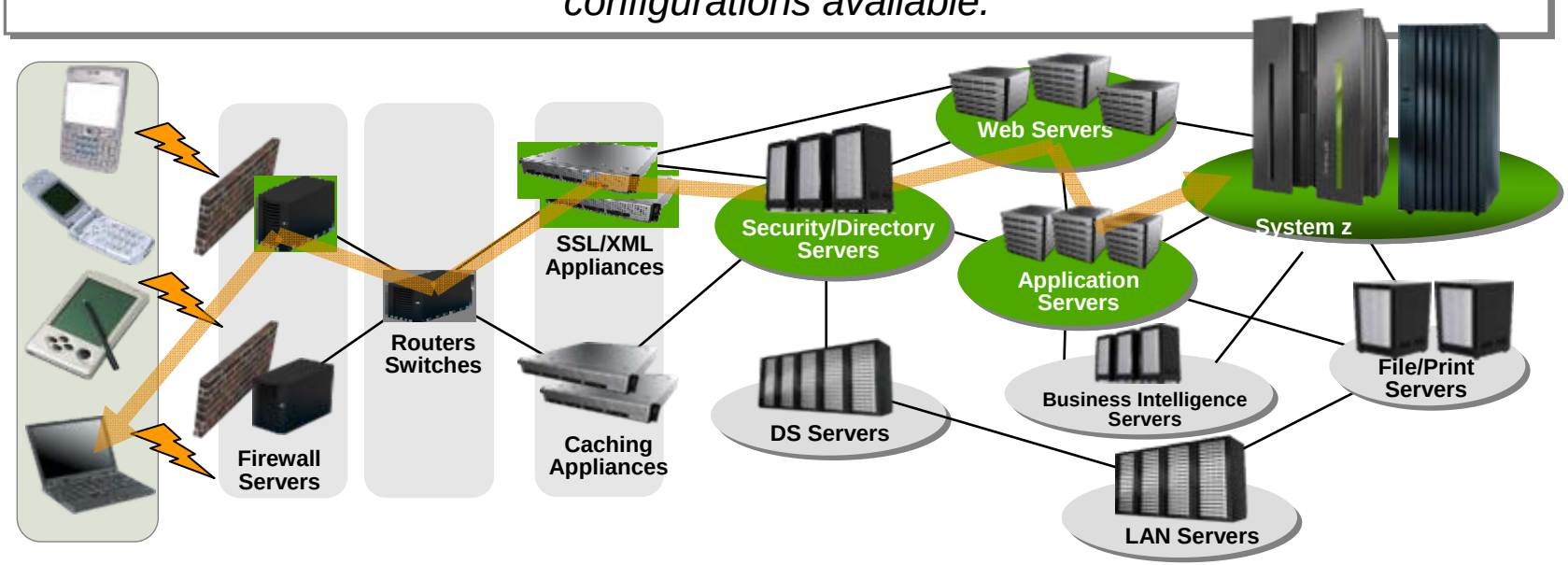
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

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: Limitations

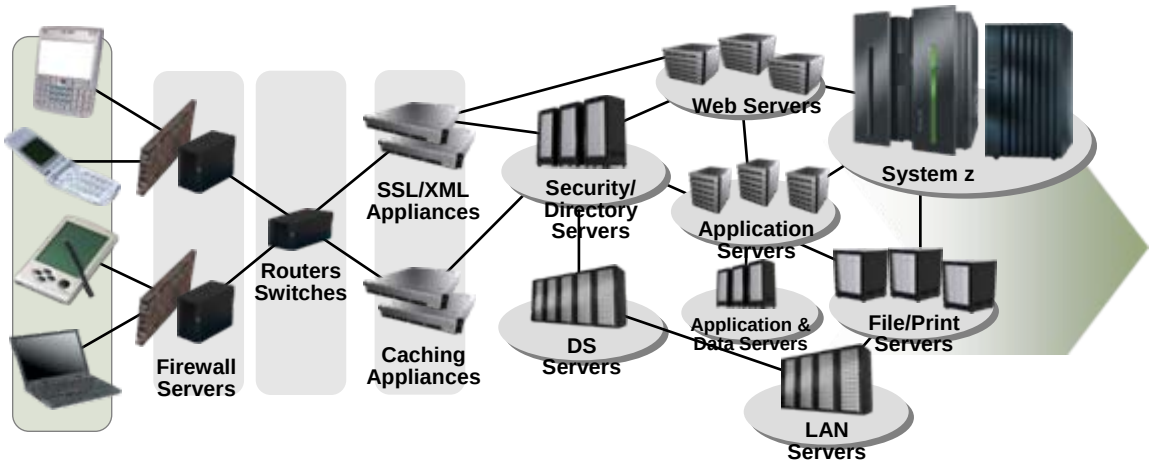
Information technology today is limited by the technology and architecture configurations available.



- Business processes and the applications that support them are becoming more service oriented, modular in their construction, and integrated.
- The components of these services are implemented on a variety of architectures and hosted on heterogeneous IT infrastructures.
- Approaches to managing these infrastructures along the lines of platform architecture boundaries cannot optimize: alignment of IT with business objectives; responsiveness to change; resource utilization; business resiliency; or overall cost of ownership.
- **Customers need better approach: The ability to manage the IT infrastructure and Business Application as an integrated whole.**

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

System x **POWER7** **zEnterprise**



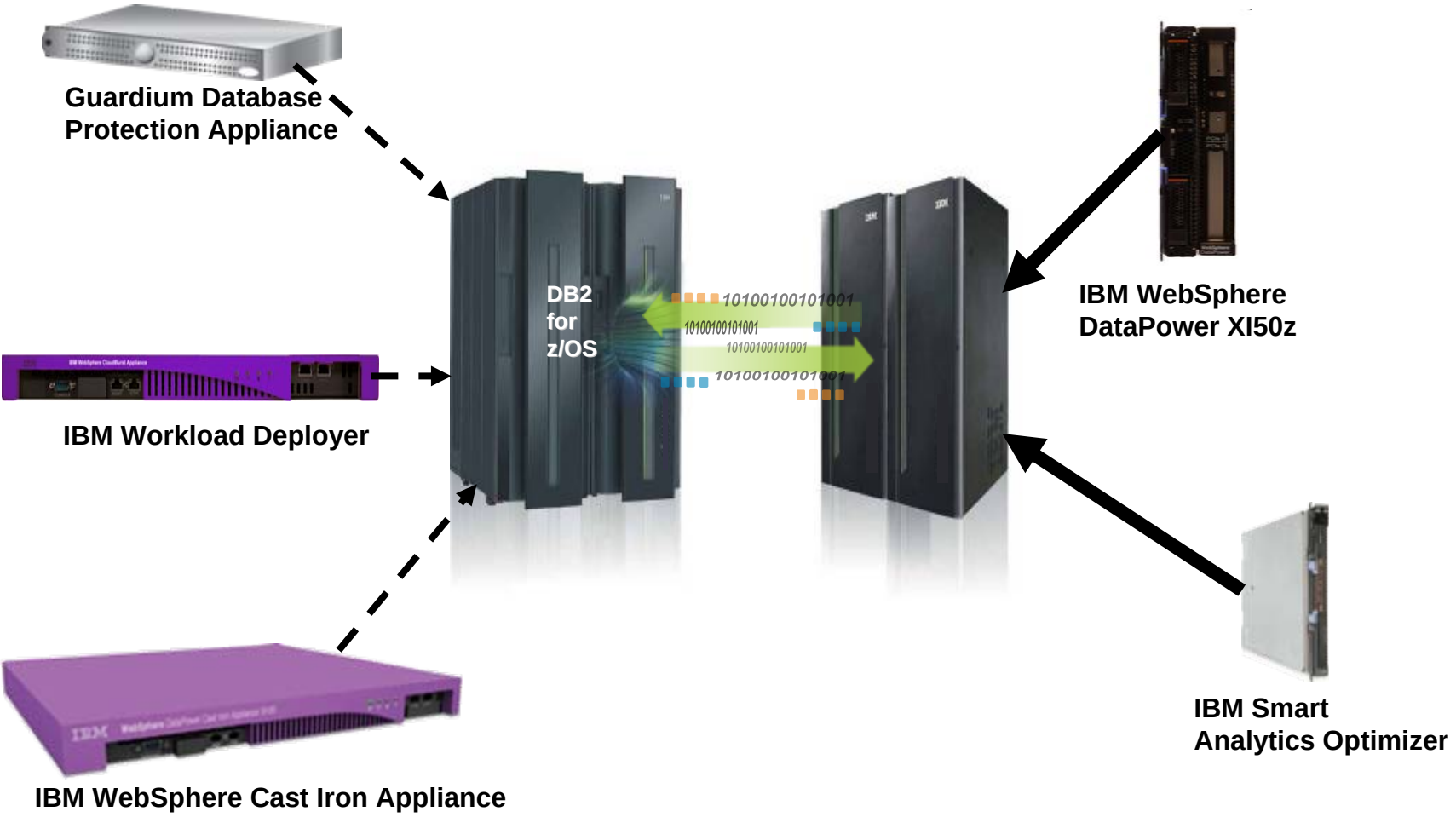
The Right Fit: Using the Right Tool for the Right Job

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 Appliances – Integrating with zEnterprise

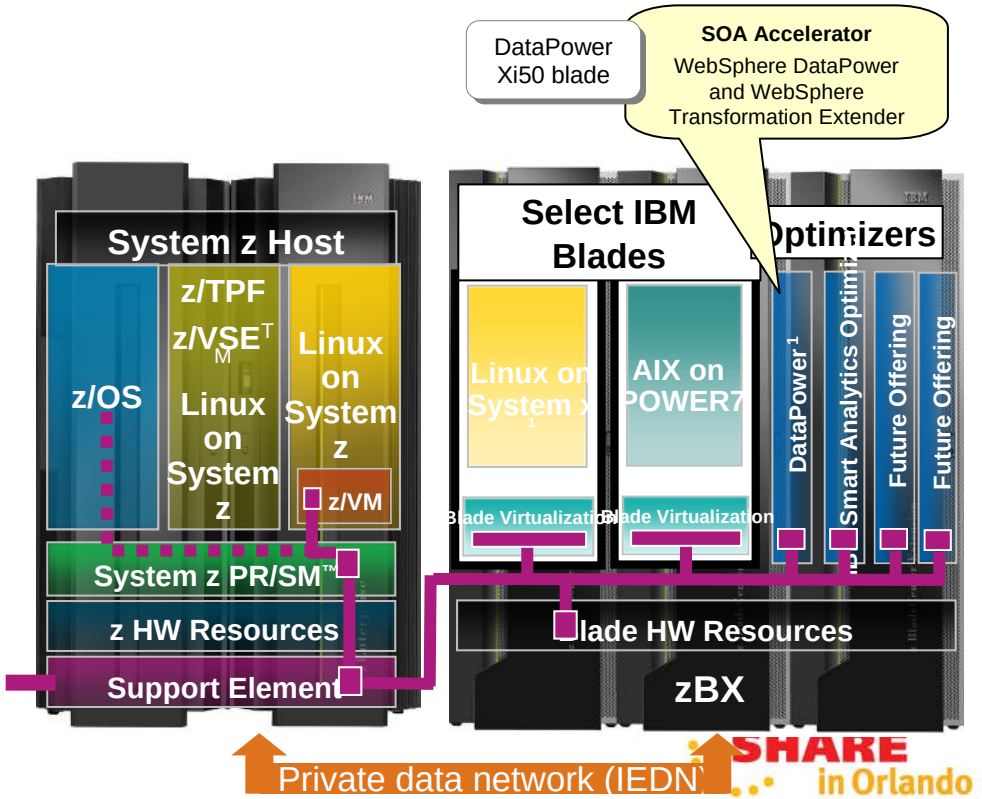


Purpose Built & Hybrid Computing

- ***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?"

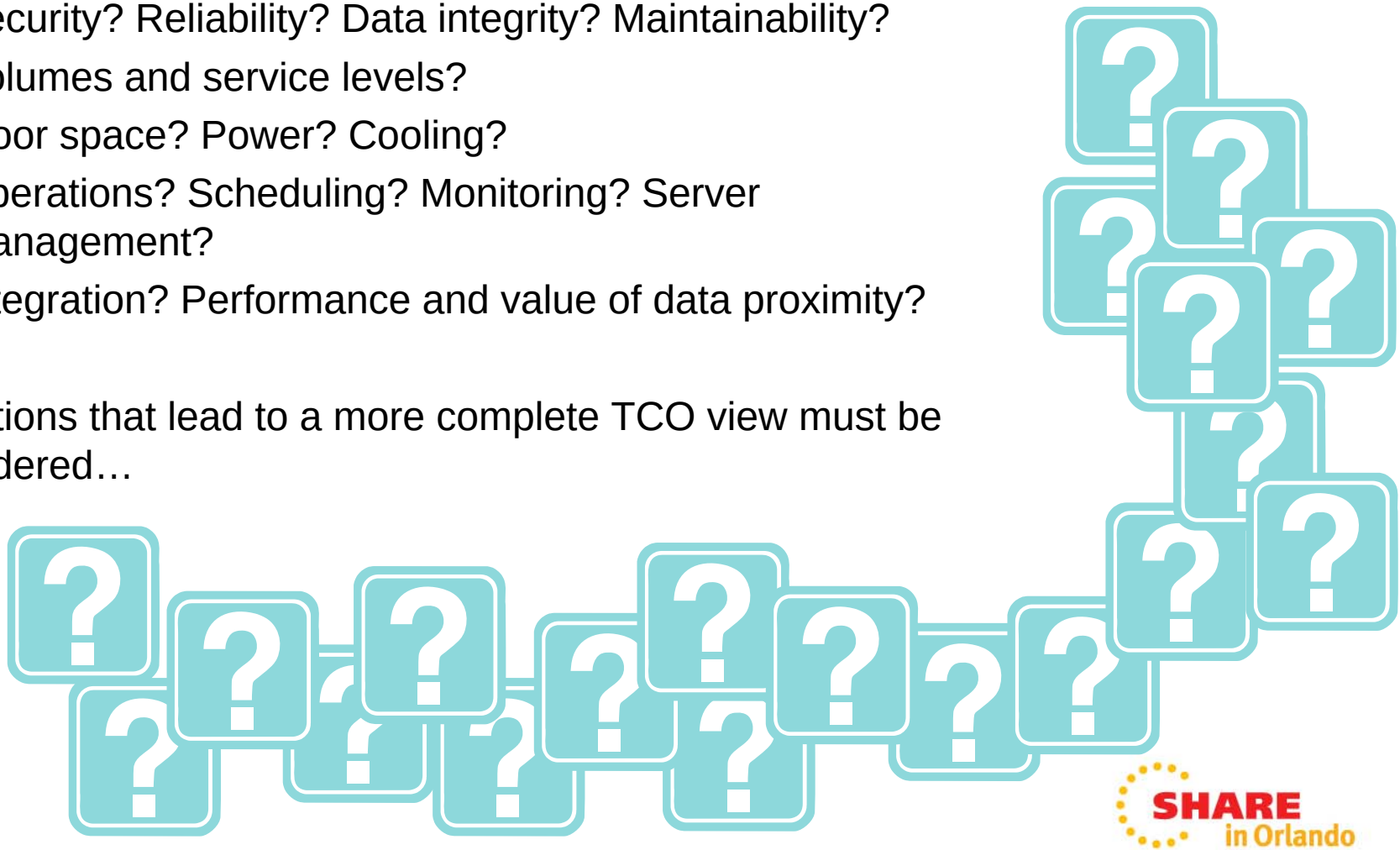
- They're done!

- 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…



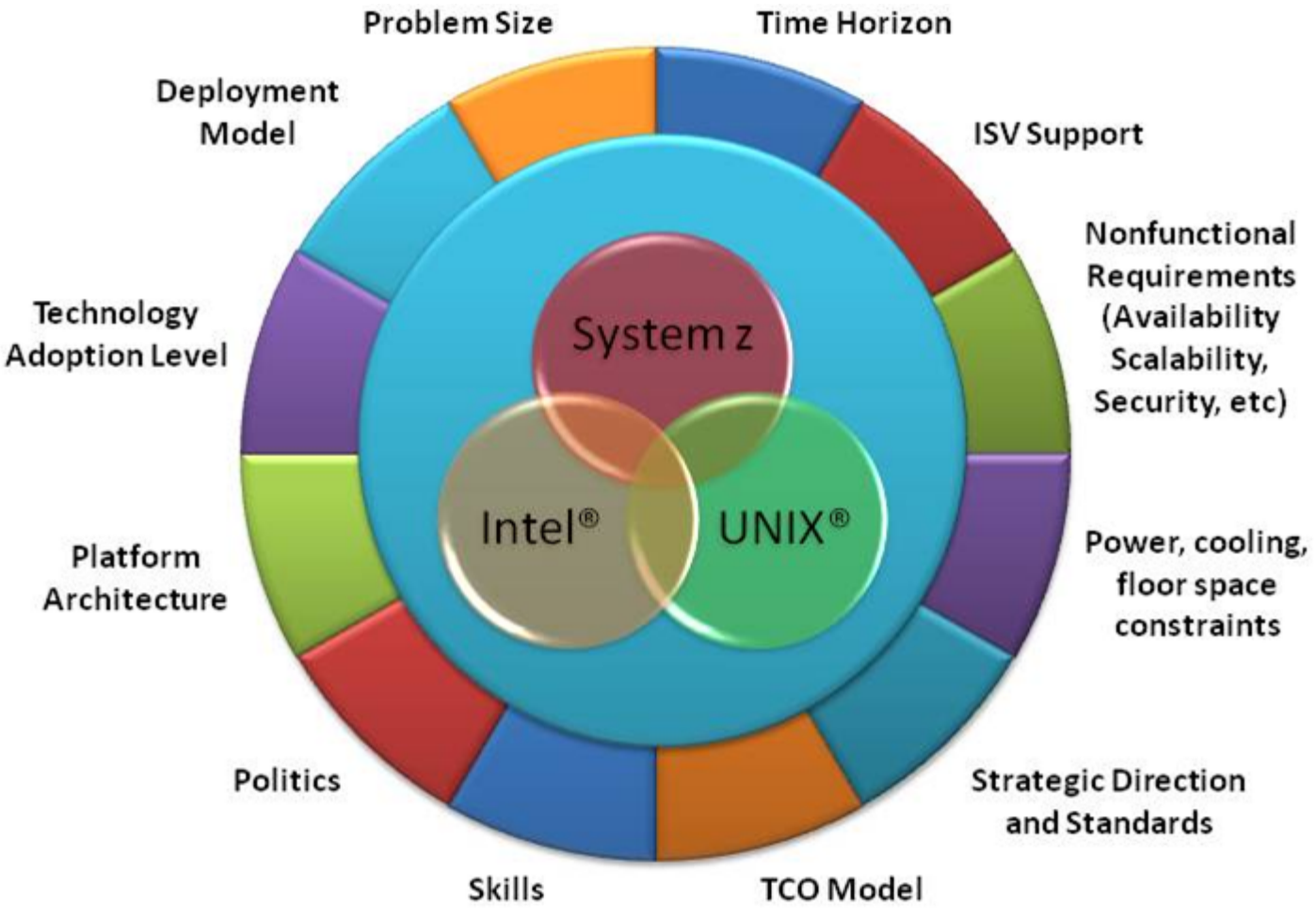
TCA versus TCO

- 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.

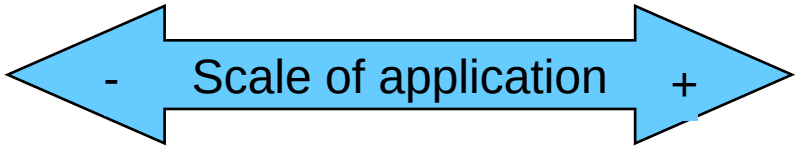
What Quality of Service Do Your Applications Require?

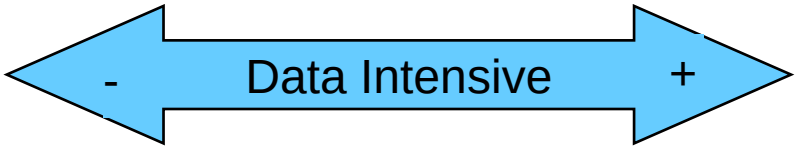
- What service level agreements are in place?
- What level security does the specific application implementation require?
- What response times are required?
- How much network latency can be tolerated?
- Is 24 x 7 x 365 availability required, or do systems just need to be up during business hours?
- 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

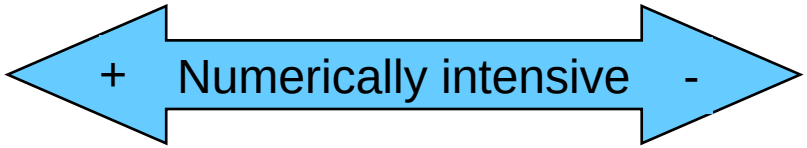
Hardware Considerations

Intel **UNIX** **System z**





UNIX/RISC Intel System z

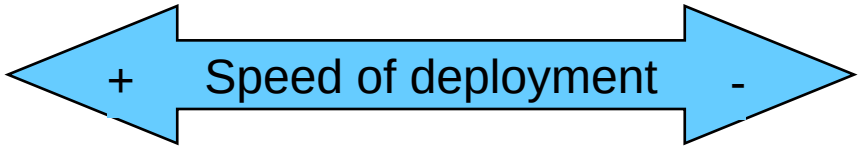


Software/OS Considerations

Linux on System z **z/OS**

AIX in zBX

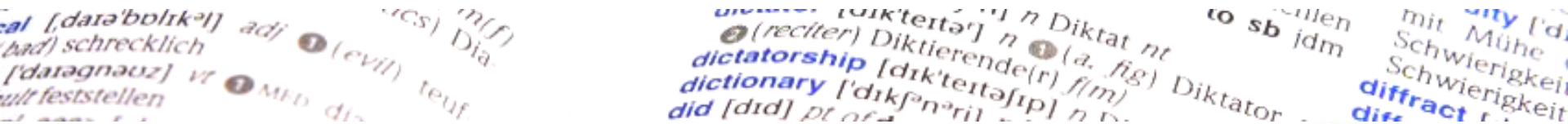
Linux on zBX



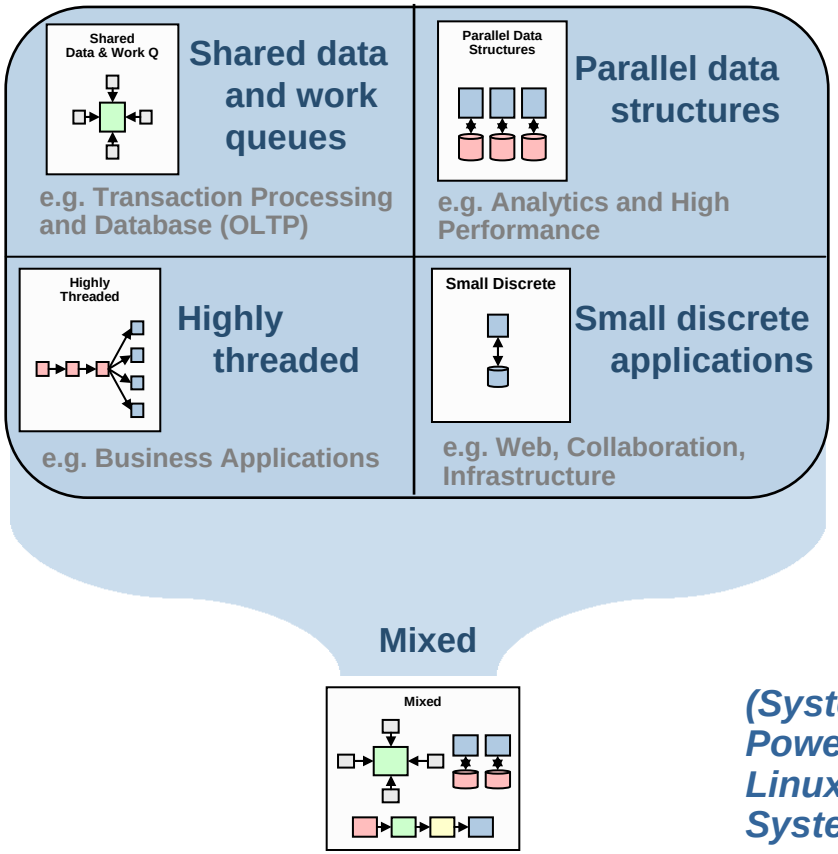
High Level Workload Definition

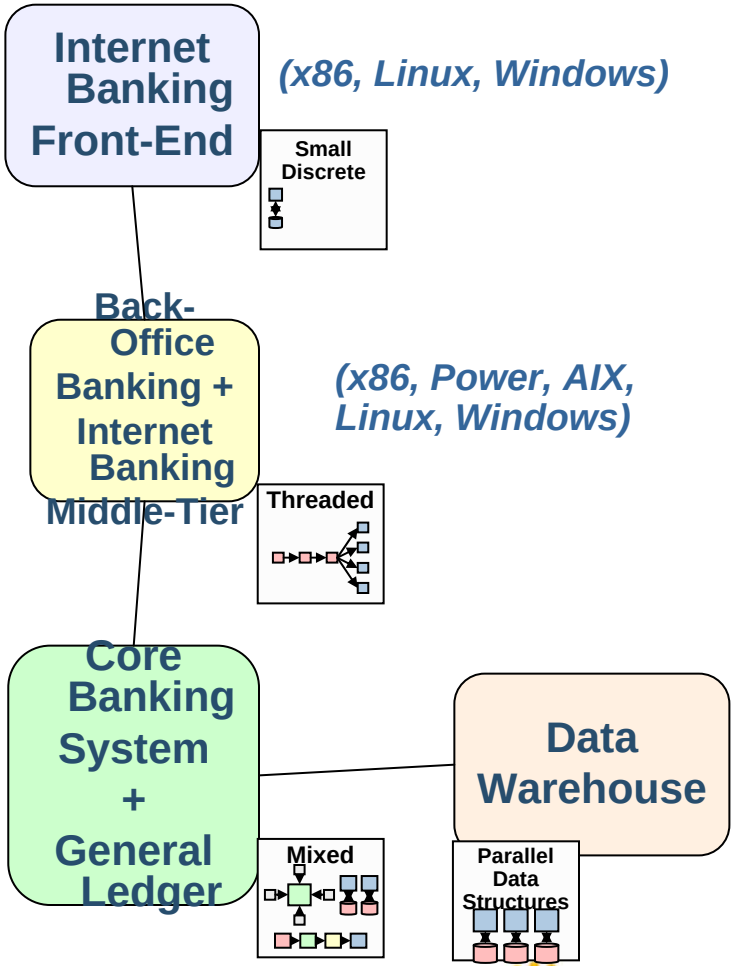
- **Workloads** are a combination of:
  - **Application function**: What it does and how it does it
  - **Data structure**: Data residency, topology, access model
  - **Usage pattern**: Utilization profile over time, mix of use cases
  - **Service level**: Non-functional requirements
  - **Integration**: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

- *The workload requirements will create varying demands when determining server alternatives*

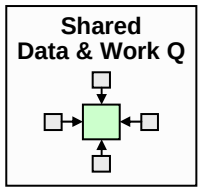
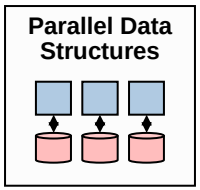
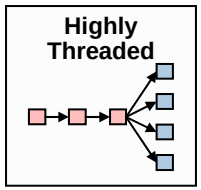
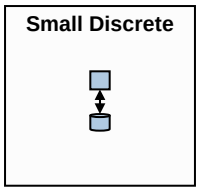
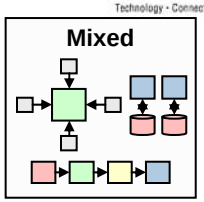


Workload Architectures





Workload Characteristics and Platform Requirements

	Shared Data & Work Q	Parallel Data Structures	Highly Threaded	Small Discrete	Mixed
					
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

Workloads we see every day that match these characteristics

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.

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

Fit for Purpose Virtualization

IBM zEnterprise System Provisioning with IBM Workload Deployer…

IBM Workkoad Deployer Appliance dispenses WebSphere Application Server Hypervisor Edition Servers into a set of other machines

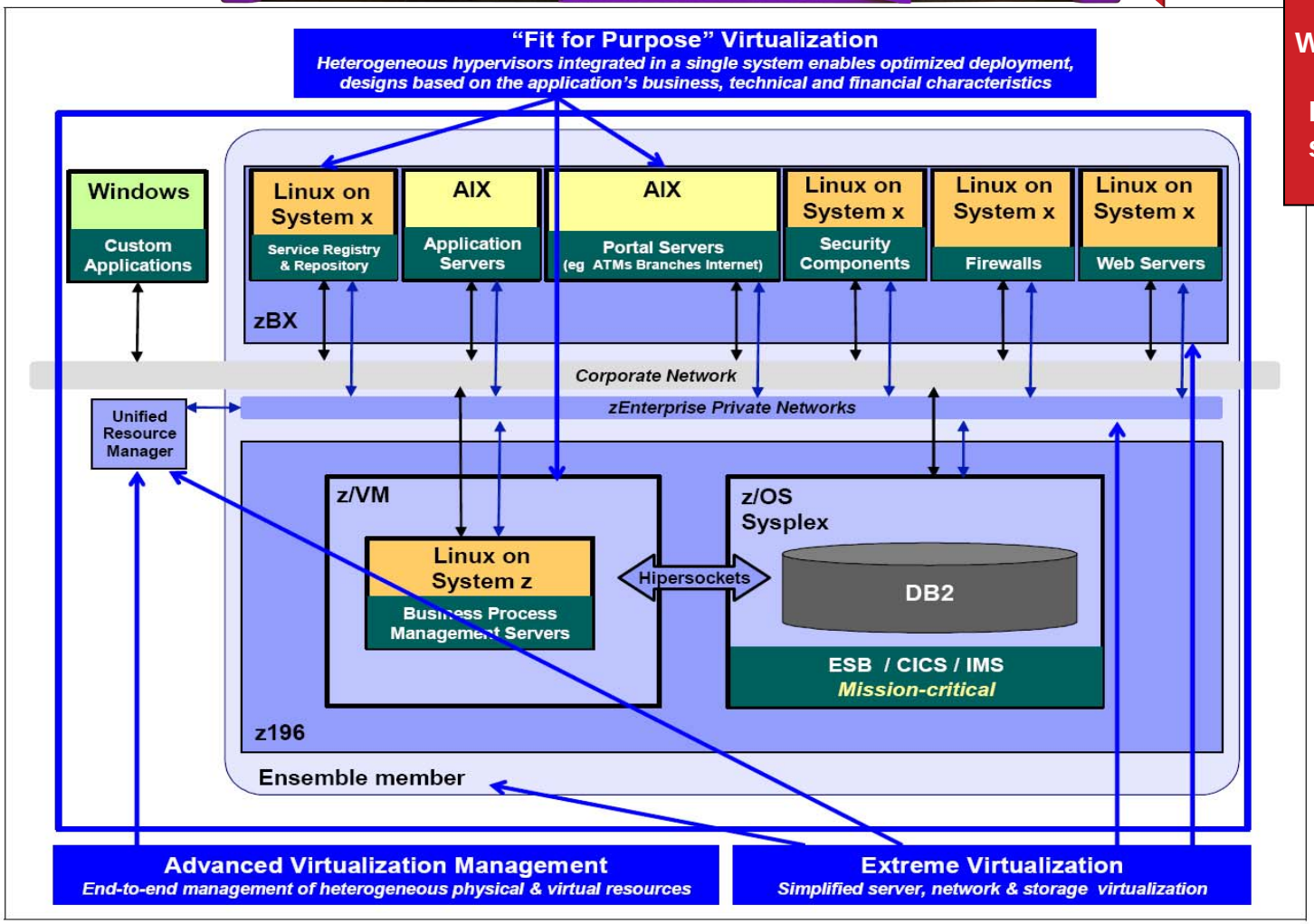


Figure 9 IBM zEnterprise virtualization: real business value for today's heterogeneous applications

Watson answers a grand challenge - Jeopardy

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

In July 2010, the IBM zEnterprise system introduced the first hybrid computing technology enabling clients to:

- Optimize the deployment of workloads by utilizing the best fit technology and operating environment
- Deploy enterprise private clouds that are ready for mission critical applications
- Establish a common management infrastructure for both mainframe and distributed-systems
- Take actionable insight based upon real time analytics

Central Processing Complex	IBM zEnterprise™ Unified Resource Manager	IBM zEnterprise BladeCenter® Extension (zBX)

IBM zEnterprise System – Best-in-class systems and software technologies

A "System of Systems" that unifies IT for predictable service delivery

IBM zEnterprise 196 (z196)

- 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

zEnterprise Unified Resource Manager

- Unifies management of resources, extending IBM System z® qualities of service end-to-end across workloads
- Provides platform, hardware and workload management

zEnterprise BladeCenter Extension (zBX)

- Selected IBM POWER7® blades and IBM System x® blades for deploying applications in a multi-tier architecture
- High-performance optimizers and appliances to accelerate time to insight and reduce cost
- Dedicated high-performance private network

What's New: Delivering innovation and value at all levels

Central Processing Complex

- *New I/O subsystem for improved system connectivity*
- *Security enhancements*
- *Clustering improvements*
- *New IBM zEnterprise 114 (z114) for mid-sized businesses*

zEnterprise Unified Resource Manager

- *Delivering APIs to enable management of Unified Resource Manager from external tools[1]*

zEnterprise BladeCenter Extension (zBX)

- *Introduction of select System x blades into zBX*
- *Support for Linux & in the futureWindows[1] to broaden application support and integration.*

1 All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represents goals and objectives only.

Introducing the IBM zEnterprise 114
Bringing hybrid computing to a broader set of businesses

New!

IBM zEnterprise 114 (z114)	zEnterprise Unified Resource Manager	zEnterprise BladeCenter Extension (zBX)
The next generation midrange mainframe delivering extensive growth options, flexibility, efficiency and improved price performance.	*Centralized management of heterogeneous resources for simplification and resiliency*	*Integrated IBM POWER7 blades, IBM System x blades, and High-performance optimizers and appliances*

zEnterprise technology designed for small and mid-sized businesses

The Value Begins At the Heart with the z114

zEnterprise 114 (z114)

2 Models: M05 & M10

- **New technology in a new package**
  - Modular 2 drawer design for lower cost of entry
  - Granularity for right-sizing your system
  - Additional Scale for consolidation and growth
  - Improved data center efficiency
  - Same Qualities of Service as the z196
  - Hybrid enabled to drive workload expansion and integration
- **Improved Platform Economics**
  - New Software Curve
  - Lower Hardware Maintenance
  - Lower specialty engine and memory prices
  - Upgradeability for investment protection

Up to 18% Improvement for traditional z/OS workloads [1]

Up to an ADDITIONAL 25% Improvement in CPU intensive workloads via compiler enhancements[2]

Up to 12% Total capacity improvement [1]

Scales From **26 - 3100** MIPS

Up to 130 available capacity settings

From 1-10 configurable cores for client use includes CPs, IFL, zIIP, zAAP, and ICFs

From 0-2 IBM provided spare cores

Up to 256 GB RAIM fault tolerant memory

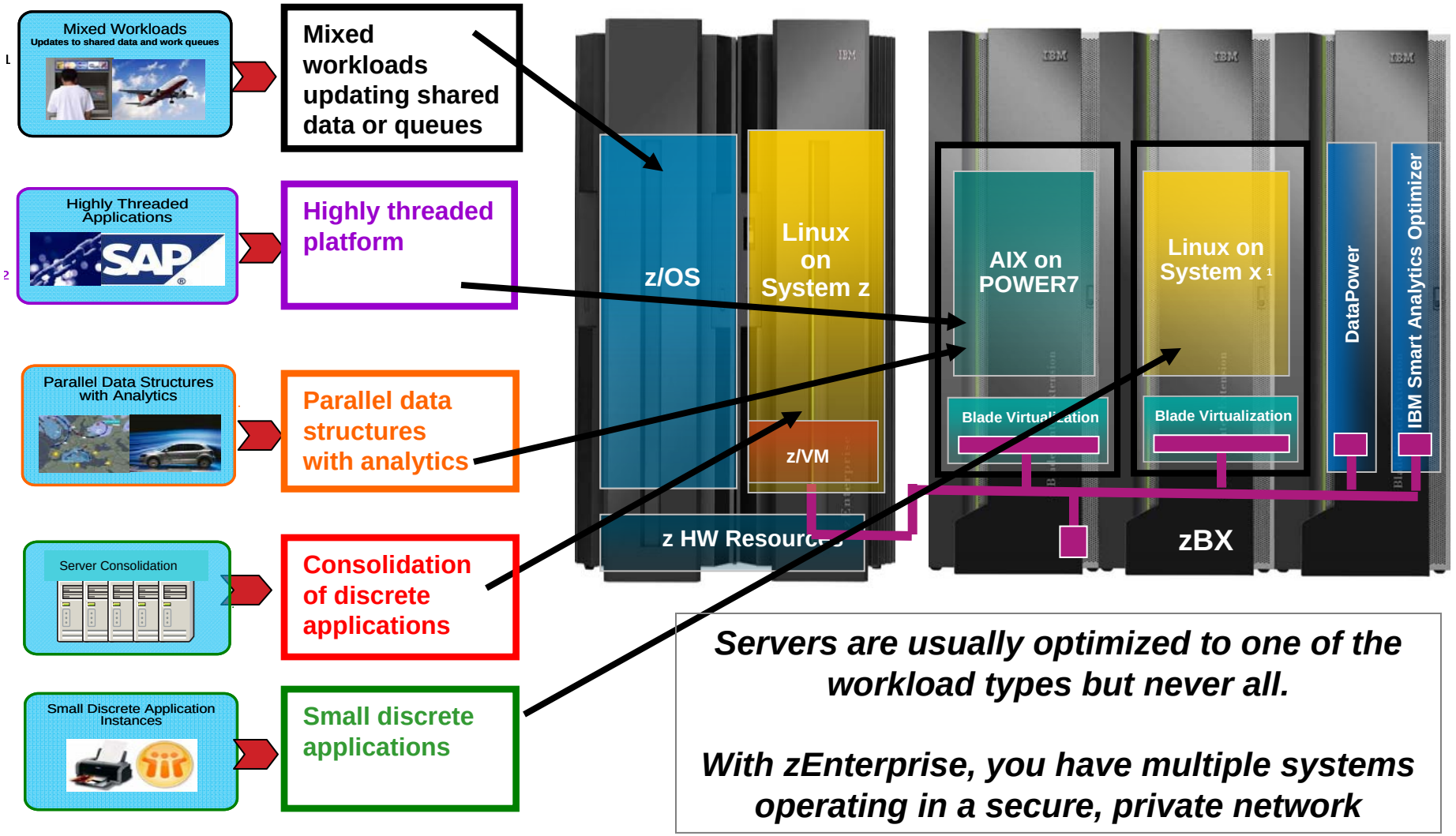
Fully Upgradeable from the IBM System z10 Business Class™ (z10 BC) & IBM System z9® Business Class (z9 BC); and to the z196 M15

[1]Relative capacity and performance compares at equal software levels as measured by IBM Large System Performance Reference (LSPR) workloads using z/OS® 1.11, Results may vary
[2]The z114 will exhibit up to 25% increase for CPU intensive workload as provided by multiple C/C++ compiler level improvements when going from z/OS 1.09 to z/OS 1.12

Providing investment protection while enabling growth

- Continuing to protect your investment with two generation upgrades
- Full upgradeability within each server family
- Temporary or permanent growth when you need it
- z114 (M10) is upgradeable to the z196 (M15 Air cooled only)

Workload harmony with zEnterprise – a "right-fit" decision



Servers are usually optimized to one of the workload types but never all.

With zEnterprise, you have multiple systems operating in a secure, private network

European Utility Company Using SAP

The Current: ***z10 + p595 AIX for SAP Central Instance and Application Servers, with DB2 for z/OS database, 60K bills per hour***

Client Requirement

- *Achieve 200K bills per hour*

Provide up-to-date technology
- z196+ p770

✓Results: achieved **250K** bills per hour

✓**70+%** improvement

Provide hybrid technology
- z196+ zBX

✓Results: achieved **430K** bills per hour

✓**600+%** improvement

Hybrid Computing Benefits:

✓Over **600%** improvement in current configuration

✓Hardware setup: implementation of zBX Power Blades in only **2 days**

✓Very good linear **scalability** either on scale-up for DB2 on z, or scale out on pBlades on zBX

✓**Low latency** due to the dedicated IEDN network

What Makes the Best Fit Workload for 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
  - 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

29

Best Fit Workloads for System z

- 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

30

What Makes a Good Fit Workload for System z

- **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

31

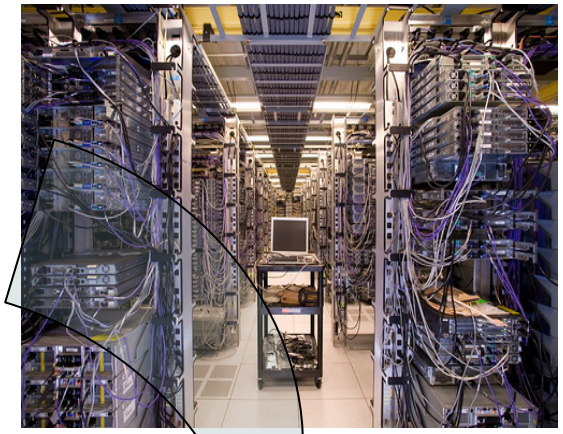
Good Fit Workloads for System z

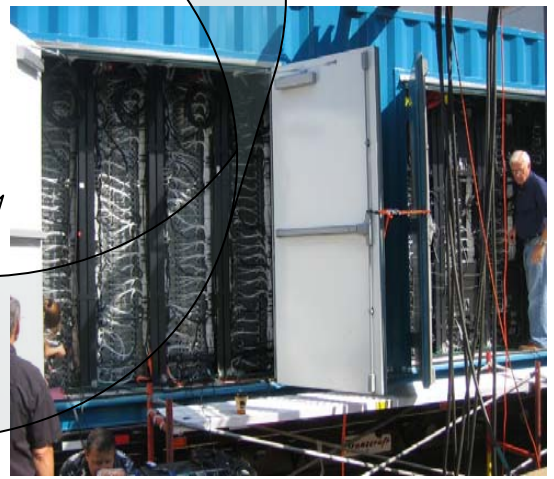
- Security
- Systems Management
- Networking
- Linux on z
  - UDB LUW
  - Informix, (IDS)
  - Apache web serving
  - SAMBA
  - TIM/TAM (LDAP Services)
  - TSM
  - Existing Linux Workloads on Distributed

32

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

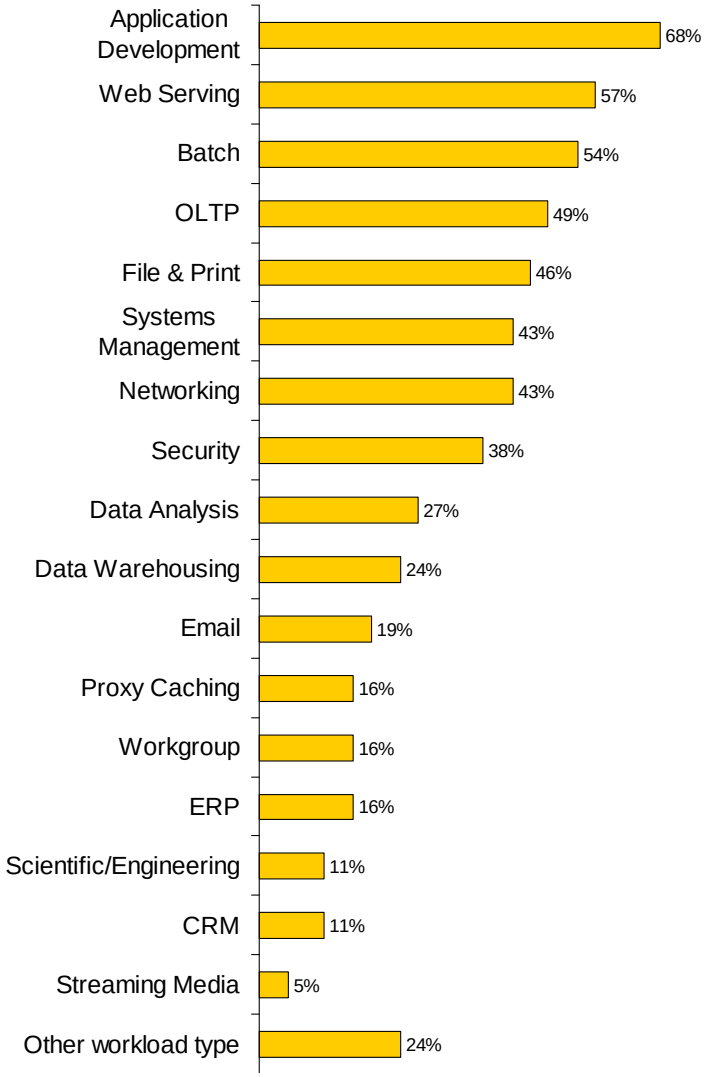
- **Leverage classic strengths of the System z**
  - High availability
  - High i/o bandwidth capabilities
  - Flexibility to run disparate workloads concurrently
  - Requirement for excellent disaster recovery capabilities
  - Security
- **Shortening end to end path length for applications**
  - **Co-location 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"

What **are Linux users running on System z?**



Source: 2009 IBM Market Intelligence

Surveys indicate customers use:

- Web Serving
- Web Application Serving
- Data Services
- 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

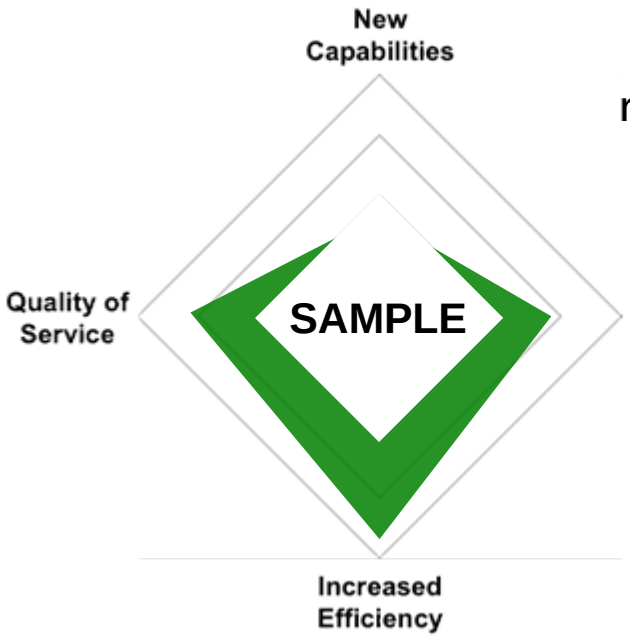
Numerous Fit for Purpose Workshops

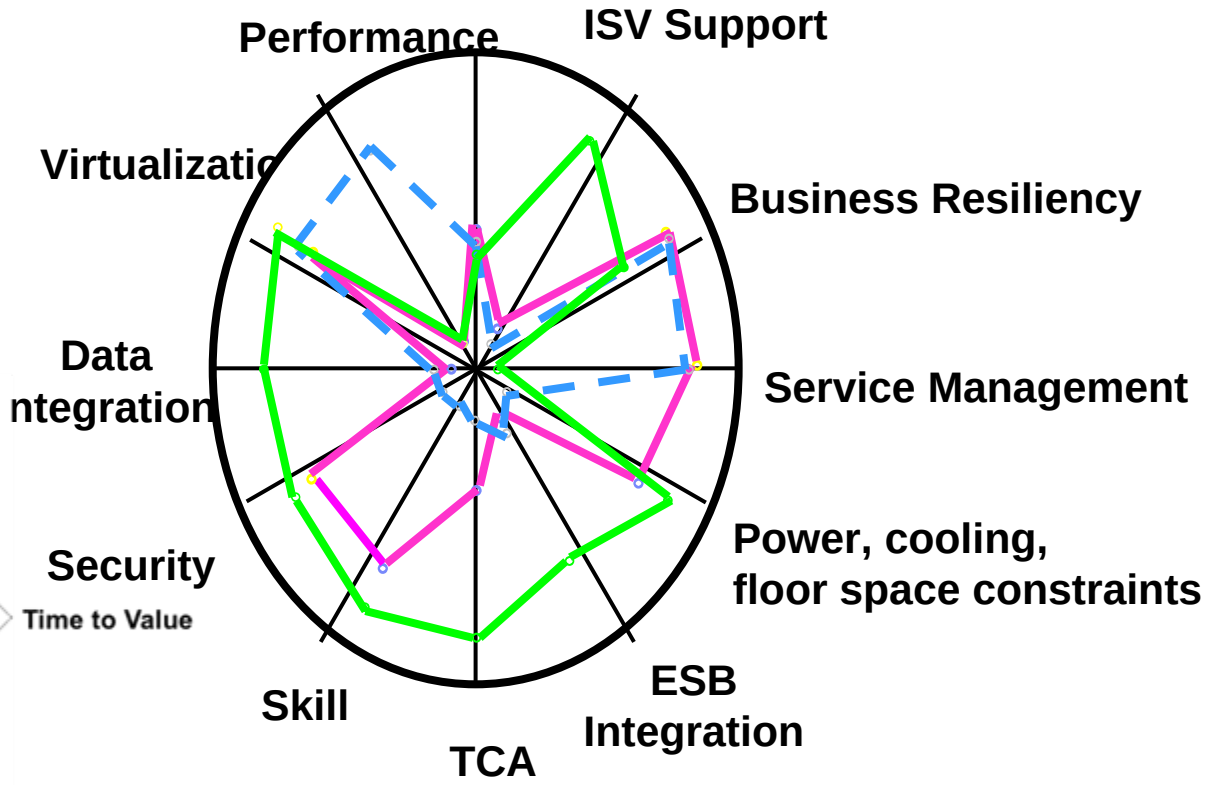
IBM Will Work with You to Determine the Most Appropriate for You

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

Numerous Tools & Methodologies to Measure Workload Requirements

Radar charts
Business Value Assessment Tool





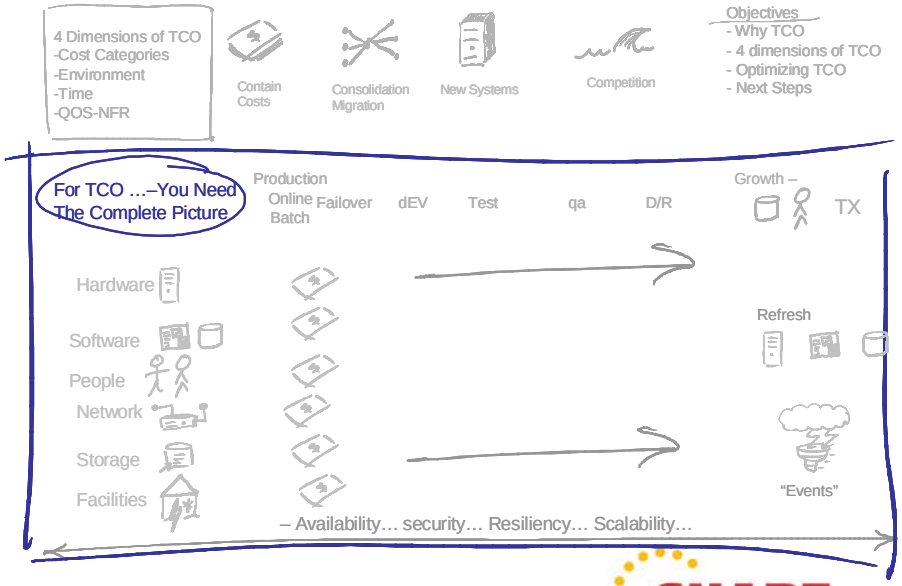
Spider charts

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 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-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