UNIVERSITÄT LEIPZIG

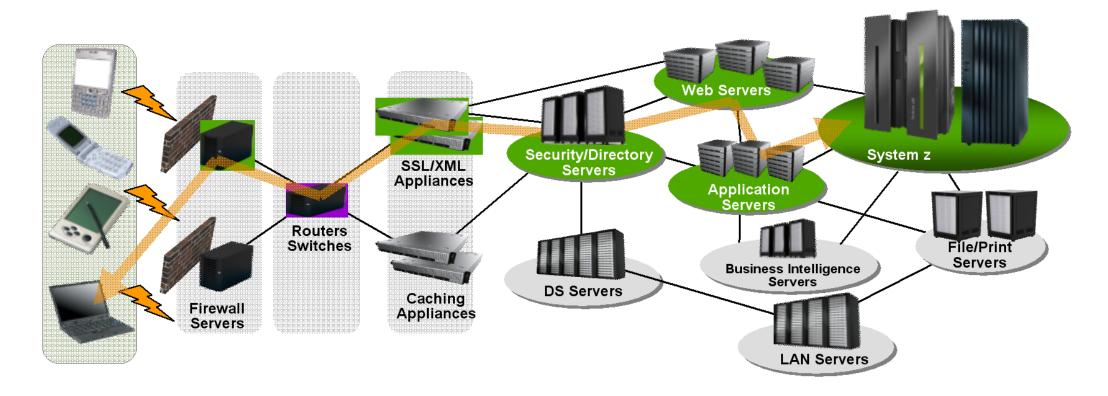
Mainframe Internet Integration

Prof. Dr. Martin Bogdan Prof. Dr.-Ing. Wilhelm G. Spruth

SS2012

zBX Teil 3

zBX Einführung

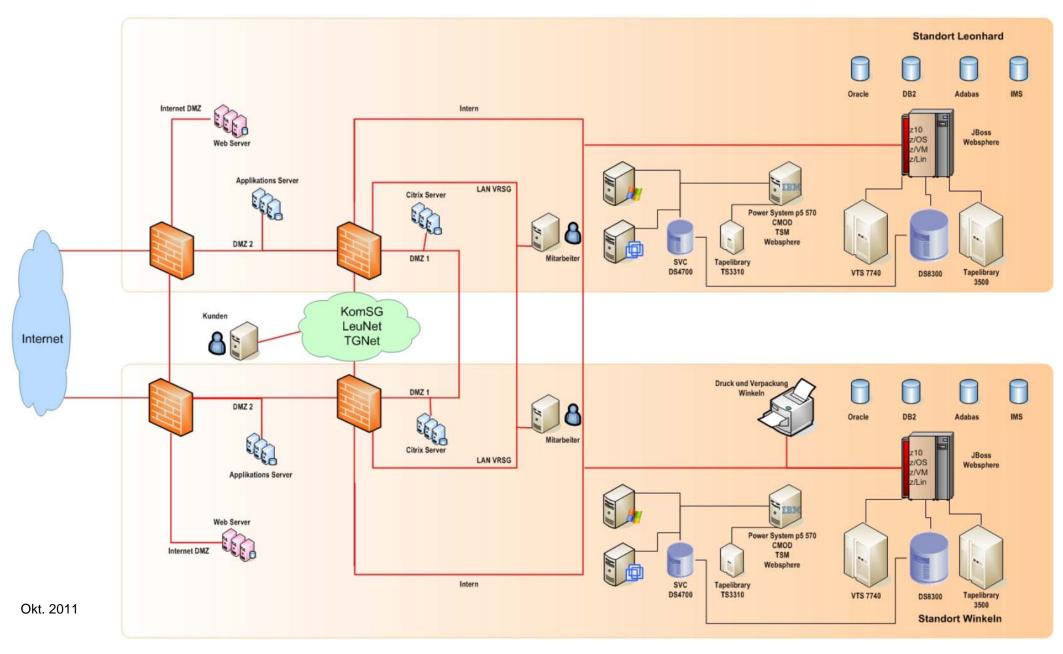


IT Infrastruktur in einem Großunternehmen

Die IT Infrastruktur eines Großunternehmens besteht neben den Mainframes typischerweise aus einer großen Anzahl weiterer nicht-Mainframe Server. Dies können häufig viele tausend Server sein, die spezielle Aufgaben wie Kryptografie (Secure Socket Layer), LDAP Directory Services, Routing, Netzwerk Management, Business Intelligence, Präsentation Logik, SAP Server, Web Application Server und vieles anderes.

Häufig werden von Spezialfirmen Anwendungen entwickelt, die nur unter einem spezifischen Betriebssystem (Windows, Linux, Solaris, MacOS, andere) und/oder einer spezifischen Hardware (x86, Sparc, PowerPC) laufen.

Im Gegensatz zum Mainframe wird dieser Teil der IT Landschaft als "distributed" bezeichnet.



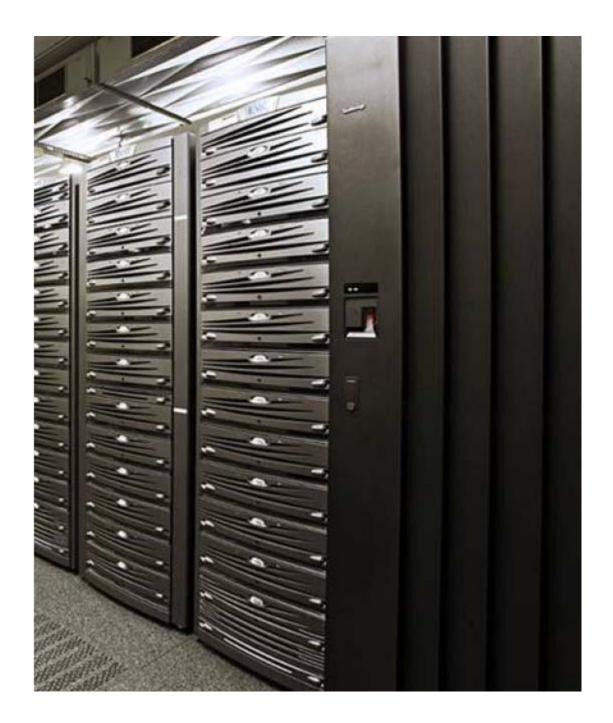
Das mittellgroße VRSG, Verwaltungsrechenzentrum AG St.Gallen, Dienstleistungsunternehmen für Städte und Gemeinden in der Ost-Schweiz, zeigt die Komplexität der distributed IT Infrastruktur (DMZ = Demilitarized Zone).

IT Infrastruktur

Der distributed Teil der IT Infrastruktur stellt fast immer einen sehr unorganischen Wildwuchs dar. In kleinen Unternehmen ist eine einheitliche IT Strategie relativ leicht zu etablieren und einzuhalten, z.B. "wir benutzen ausschließlich Microsoft Server Software (Directory Server, Transaction Server, SQL Server). Beim Focus Verlag in München werden seit vielen Jahren nahezu ausschließlich Mac OS Server eingesetzt. In großen Unternehmen fehlt oft eine einheitliche Strategie, oder wird nicht konsequent befolgt, ändert sich bei einem Wechsel in der Unternehmensführung oder scheitert an Sachzwängen.

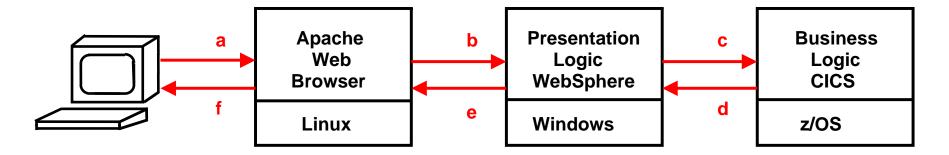
Die distributed Infrastruktur hat eine hohe Komplexität und einen großen und aufwendigen Administrationsaufwand zur Folge. Spezifisch haben die zahlreichen Plattformen unterschiedliche und inkompatible Management Interfaces.

In vielen Unternehmen sind die distribted Server in den einzelnen Fachbereichen installiert und über das Unternehmen verstreut. Häufig wird versucht, die Server im Rechenzentrum zu konzentrieren um einen (kleineren) Teil der Kosten und des Administrationsaufwandes einzuspen. So sind z.B. bei der Fiducia AG, Karlsruhe neben den 5 Mainframes mit 60 000 MIPS weitere 6 611 Server vorhanden, davon heute 3.885 zentralisiert im Rechenzentrum.



Fiducia Rechenzentrum

Die Fiducia IT AG Karlsruhe hat einen Teil der dezentral in den Fachabteilungen stehenden Server zentralisiert und einheitlich auf Sun Solaris umgestellt. Im Rechenzentrum stehen neben den 5 Mainframes weitere 3.885 Unix Solaris Server. die als Blades implementiert sind (linke Abbildung). Dies verringert deutlich den Administrations-Aufwand, der aber dennoch wesentlich höher als der Mainframe Administrations-Aufwand bleibt.



Die Komplexität sei an Hand eines Beispiels demonstriert.

Gezeigt ist eine typische Konfiguration, bei der die Business Logic unter z/OS und CICS läuft, und die Presentation Logik unter WebSpere auf einem Windows Rechner.

Der z/OS Work Load Manager kann die Response Time zwischen den Punkten c und d messen, nicht aber zwischen den Punkten a und f, also dem, was der Benutzer an seinem Bildschirm sieht. (Wäre der Browser und WebSphere auch unter z/OS installiert, wäre das kein Problem). Wenn die Response Time ungenügend ist, ist es schwierig herauszufinden, ob das Problem an dem Linux oder dem Windows Server liegt. Der Administrator tappt hier im Dunkeln. Dies gilt besonders, wenn auf Grund der hohen Work Load nicht ein sondern 10 oder 20 Apache und WebSphere Server installiert sind.

Was passiert: Der Administrator rät, dass das Problem bei dem WebSphere Server liegt, und installiert zwei zusätzliche Server. Wenn er Pech hat, wird die Response Time dadurch noch schlechter.

CPU Utilization

Die CPU Utilization (andere Bezeichnung System Utilization) in distributed Servern ist häufig nur 20 %, während Mainframes routinemäßig eine CPU Utilization von 90 % oder besser erreichen.

Hierzu Scott McNealy, Chairman, President, und CEO der Firma Sun Microsystems:

- A recent study shows that about 10% of IT costs are hardware, 10% software, and the rest administration and training.
- System utilization is around 15%; it should be 80%.
- Today, a system administrator can manage between 15 and 30 systems; it should be 500.

http://www.informationweek.com/story/IWK20030123S0020

Das war im Januar 2003. Seitdem hat sich wenig geändert. Ebenfalls von Jonathan Schwartz, Executive Vice President, Sun Software:

- IT infrastructure utilization rates are at an all-time low. The rapid spread of new services across a multiplicity of distributed systems has resulted in a gross underutilization of hardware.
- Management costs and complexity are at an all-time high. The proliferation of individual, low-level software and hardware components that deliver just one service requires too much management.

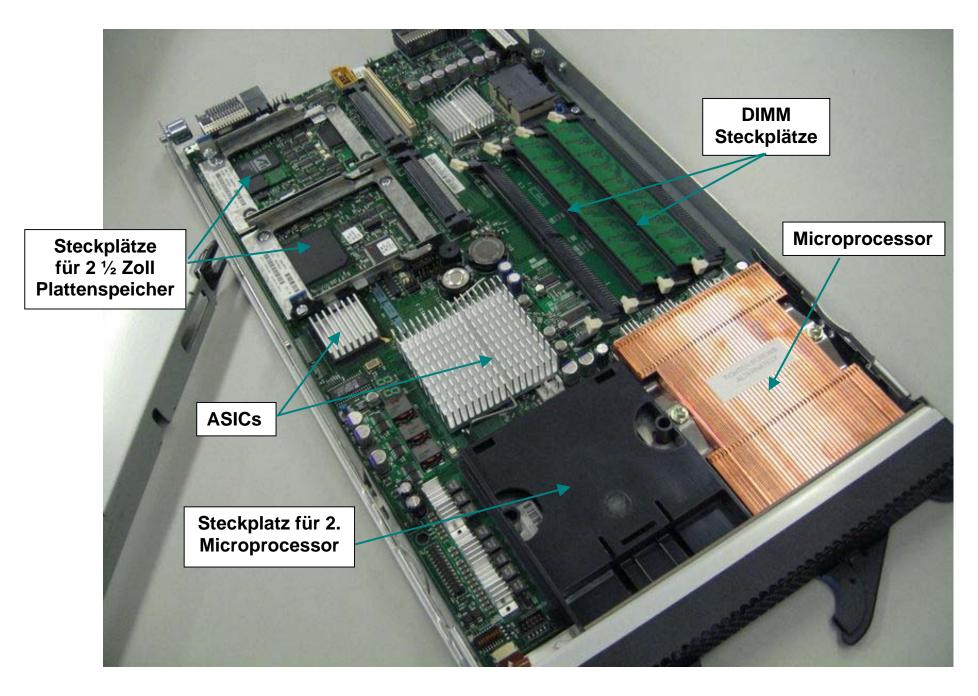
http://sysdoc.doors.ch/SUN/N1booklet.pdf

Blades

Distributed Server bestehen in der Mehrzahl der Fälle aus Baugruppen, die als "Blades" bezeichnet werden. Eine Blade ist ein Printed Circuit Board (ähnlich einem PC Mainboard), auf dem sich typischerweise zwei Microprozessor Chips befinden, normalerweise mit 4 oder 8 CPU Cores/Chip. Weiterhin enthält eine Blade Steckplätze für Hauptspeicher DIMMs (Dual Inline Memory Module), ASIC Chips für die Integration, für Ethernet und Serial SCSI Anschlüsse und eventuell Steckplätze für ein bis zwei 2 ½ Zoll Festplatten, auf denen das Betriebssystem untergebracht werden kann.

Blades werden sehr ähnlicher Ausführung von zahlreichen Herstellern produziert, darunter Dell, HP, IBM und Sun, normalerweise mit einer ganzen Reihe von unterschiedlichen Ausstattungsvarianten. Die Blades werden mit x86 Prozessoren der Firmen Intel (Xeon) und AMD (Opteron) bestückt, dazu mit proprietären Chips der Hersteller (HP Itanium, IBM PowerPC und Cell sowie Sun Sparc. Auf den Blades laufen Betriebssysteme wie Windows, Linux, Solaris, HP-UX, IBM AIX und Sun Solaris,

Auf der folgenden Folie ist eine IBM Blade zu sehen.

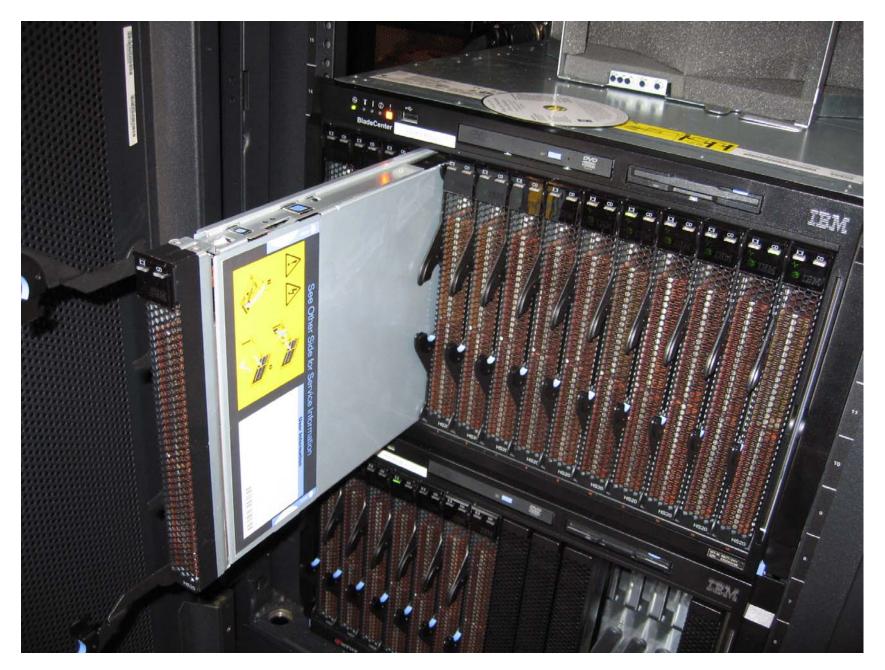


IBM Blade für Intel-, PowerPC- und Cell Microprozessoren

Blade Center

Blades werden in einem Blade Gehäuse (Blade Enclosure, IBM Bezeichnug "Blade Center") untergebracht. Eine Blade Enclosure enthält eine Anzahl von Blades, wobei auf jeder Blade typischerweise ein eigenes Betriebssystem läuft. Weiterhin stellt die Blade Enclosure zusätzliche Funktionen wie Stromversorgung, Kühlung, Ethernet Netzwerkfunktionen, Plattenspeicheranschlüsse (häufig über Fibre Channel SCSI) und Management Funktionen zur Verfügung.

Die folgende Abbildung zeigt ein IBM Blade Center, welches 14 IBM Blades aufnehmen kann. Es existieren unterschiedliche Blade Center Konfigurationen, die sich z.B. bezüglich der Netzwerk Konfigurationen unterscheiden. Es kann ein recht komplexer Ethernet Switch eingebaut sein, welcher eine direkte interne Verbindung der Blades untereinander ermöglicht. Mehrere Blade Center können in einem 19 Zoll Rack untergebracht werden.



IBM BladeCenter

Blade Enclosures werden von zahlreichen Herstellern vertrieben und häufig für distributed Server eingesetzt. Sun Microsystems bietet die "Sun Blade 6000" Blade-Server-Plattform an, die sowohl mit Sparc- als auch x86-Blades bestückt werden kann. Auf ihnen laufen die Solaris, Linux und Windows Betriebssysteme. Die Blades können mit jeweils zwei UltraSPARC T1 Cips, zwei Quad-Core-Xeon Chipss oder zwei Opteron Chipss bestückt werden.

Die Basis stellt ein Blade-Gehäuse dar, das bis zu zehn Blades aufnimmt, bei bis zu vier Gehäusen pro Rack.



Ähnliche Produkte sind von den Firmen Cisco, Dell und HP erhältlich.

System z Blade Center Extension zBX

Die "System z Blade Center Extension" (zBX) ist eine Gruppierung von einem oder mehreren speziell hierfür optimierten IBM Blade Centern, die über ein sicheres privates Hochleistungsnetzwerk mit einem zEnterprise-Zentralprozessorkomplex (Central Processing Complex, CPC) – entweder der z196 oder der z114 – verbunden sind. Eine zBX kann speziell angepasste PowerPC oder x86 Blades aufnehmen, oder als "Accelerator" bezeichnete leistungsfähige Spezialprozessoren für bestimmte Workloads (z.B. XML oder Krypto Processing).

Die zBX ist eine neue Infrastruktur für die Erweiterung von System z Servicequalität- und Verwaltungsfunktionen auf integrierte POWER7- und x86 IT-Komponenten. Die Kombination eines oder mehrerer z196 oder z114 Rechner mit einer oder mehrerer zBXs wird als zEnterprise-System bezeichnet. Die Administration eines zEnterprise-Systems erfolgt über eine neue Komponente, dem zEnterprise Unified Resource Manager mit System z Wartungsstandards.

Für eine höhere Verfügbarkeit wurde in der zBX auf verschiedenen Ebenen Hardwareredundanz vorgesehen. Dies umfasst die Power-Infrastruktur, rack-einbaufähige Netzwerkswitches, Netzteile und Switcheinheiten im BladeCenter-Gehäuse, die redundante Verkabelung zu Supportzwecken sowie Datenverbindungen zum zEnterprise-CPC (z196 oder z114).

Eine zBX besteht aus einem Gehäuse, in dem 1 oder 2 Blade Center, Stromversorgung, Verkabelung in der Form von mehrern Ethernet Switche und weiteren Komponenten untergebracht sind. Die zBX wird in der Regel unmittelbar neben dem z196 oder z114 Rechner aufgestellt. Die Datenverbindungen zwischen den beiden Einheiten sind von außen nicht zugreifbar.

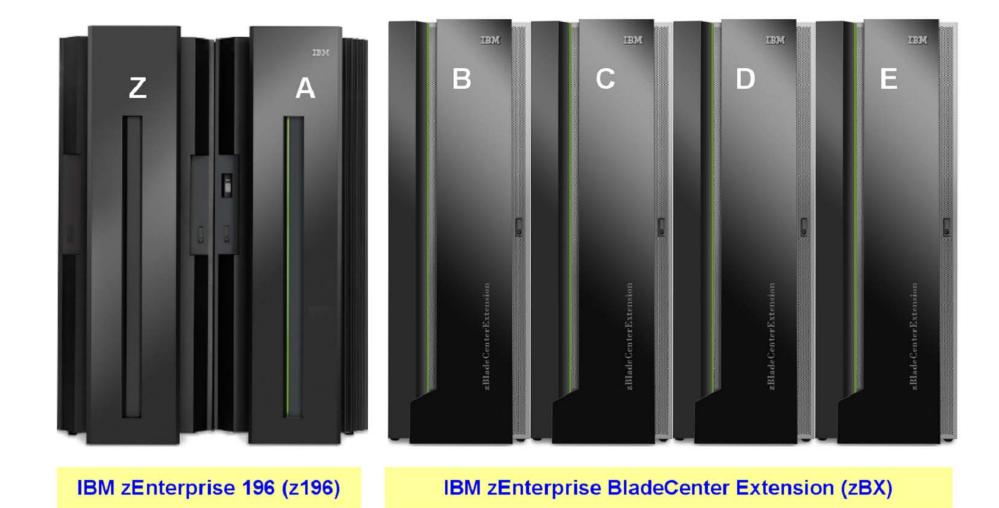


zEnterprise Blade Center Extension

Eine zBX wird über eine interne, von außen nicht zugängliche Verkabelung direkt an ein Modell z196 oder Modell 114 Mainframe angeschlossen.

z 196 Frame

zBX Frame



Es ist möglich, an einen z196 Rechner mehrere (bis zu 4) zBX Einheiten anzuschließen.

Ein Blade Center nimmt bis zu 14 Blades auf und ein zBX Frame entweder 1 oder 2 Blade Center, also insgesamt max. 28 Blades. Eine Maximalkonfiguration mit 4 angeschlossenen zBX kann somit 112 Blades enthalten-

Mit 2 CPU Chips pro Blade und 8 CPU Cores pro CPU Chip sind maximal 1 792 Blade CPU Cores möglich.

Spezielle Blades

Eine zBX kann zunächst Blades mit Standard Microprozessoren aufnehmen, spezifisch x86 und PowerPC. Auf ihnen sind alle unterstützten Betriebssystem lauffähig, besonders Windows und Linux, aber in Zukunft auch Exoten wie z.B. Solaris, Xenix, Unixware, SCO Unix oder FreeBSD.

Alternativ zu Blades mit Standard Microprozessoren sind "Spezial Blades" verfügbar. Auf Ihnen laufen bestimmte Anwendungen besonders performant dank spezieller Hardware und Software. Ein Beispiel hierfür sind die WebSphere DataPower Integration Appliances für Anwendungen wie

- Data Mining,
- Security,
- XML Processing,
- Message routing,
- Business to Business (B2B).

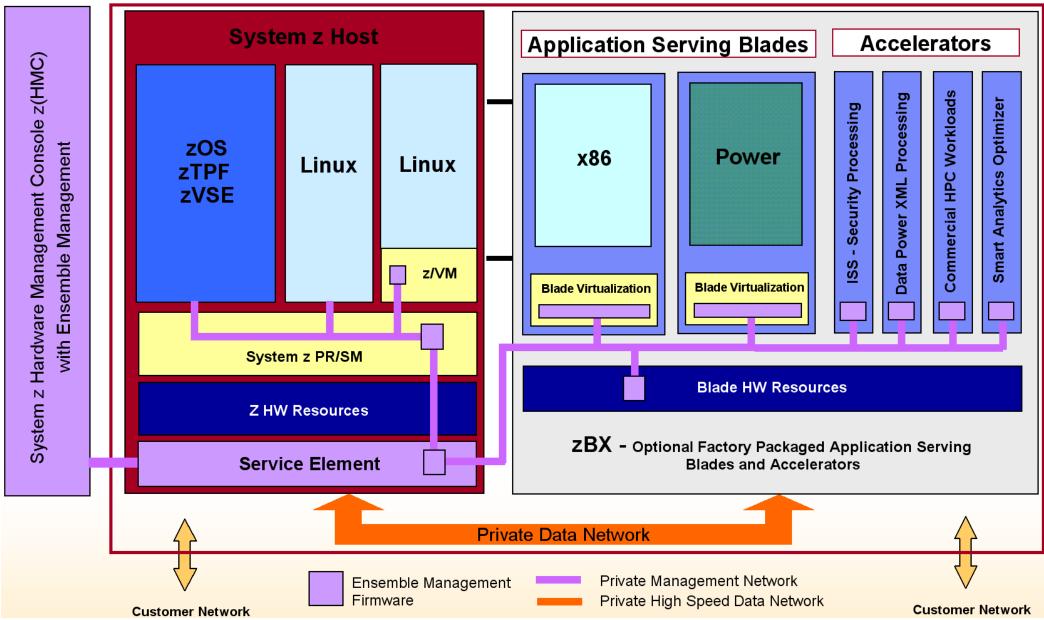
Weitere Special Blades werden in Zukunft erwartet. Ein Beispiel ist, besonders Performance-kritische Algorithmen nicht durch einen Microprozessor, sondern durch einen ASIC oder ein FPGA verarbeiten zu lassen. Gameframe ist ein weiteres Beispiel (http://forum.teamxbox.com/showthread.php?t=521848).

Private Netzwerke

Die ZBX ist sehr eng an den z196 (oder z114) Mainframe Rechner gekoppelt. Hierzu dienen zwei "Private Networks". Private bedeutet hierzu, dass die Netzwerke nur zwischen der zBX und dem Mainframe Rechner kommunizieren, und von außen nicht zugänglich sind. Ein Zugang von außen erfolgt nur über den Mainframe Rechner oder eine Blade in der zBX. Bei den beiden Netzwerken handelt es sich einmal um ein (10 Gbit Ethernet) Private Data Network sowie ein (1000BASE-T Ethernet) Management Network.

Das Data Network wird benutzt, um Daten zwischen den Mainframe Prozessoren und den Blade Prozessoren (bzw. deren Anwendungen) auszutauschen. Hierfür existieren innerhalb des zBX Frames mehrere sehr leistungsfähige Ethernet Switche und der dazugehörige Support. Da es sich hierbei um ein internes Netzwerk handelt, ist es möglich, an Stelle des sehr verarbeitungsaufwendigen TCP/IP Stacks den im Thema Partitionierung. Teil 4 erwähnten Hipersocket Mechnanismus einzusetzen. Mittels Hipersockets können zwei LPARs über eine Queue im Hauptspeicher unter Umgehung eines TCP/IP Netzes direkt miteinander kommunizieren. Über das interne zBX Private Data Network ist eine ähnliche Hipersocket Kommunikation zwischen blades und z/OS LPARs möglich.

Das Management Network wird benutzt, um die für ein Maneframe bereits vorhandenen Management Funktionen auf die Blades der zBX auszudehnen. Die einzelnen Blades, die Switche und die Stromversorgungen einer zBX sind hierfür mit zusätzlichen Support Prozessoren und/oder Firmware ausgestattet, die eine zentrale Administration ermöglichen.



ISS Internet Security Systems; HPC High Performance Computing; zBX z Blade Extension;

IBM

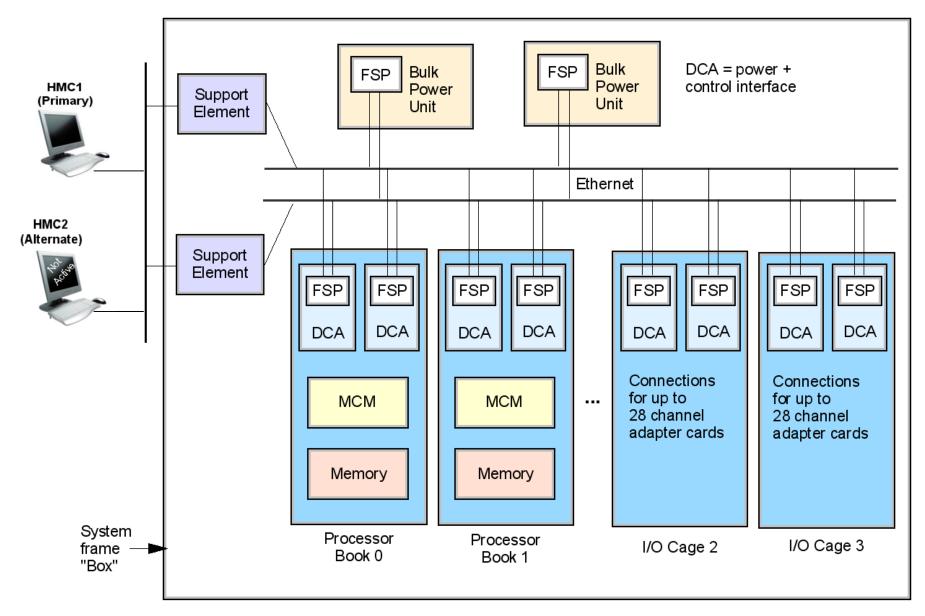
Plattenspeicher Anschluss

Die Blades benötigen eigene Plattenspeicher Und entsprechende Anschlüsse. Hierfür ist innerhalb der zBX ein weiteres Fibre Channel Netzwerk und –Switche vorgesehen, über das jede Blade mit externen Fibre Channel Platten verbunden werden kann.



z 196 Frame

zBX Frame



Hardware Management Console (HMC)

FSB = Flexible Service Processor, DCA = zusätzliche Distributed Converter Assembly Stromversorgungen

Unified Resource Manager

Die obige Abbildung ist eine Wiederholung aus dem letzten Kapitel. Wir erinnern uns, bei dem Code in den Flexible Support Prozessoren, den Support Elementen (Think Pad) sowie der HMC handelt es sich um Firmware. Firmware wird immer als Teil der Hardware betrachtet, weil er com Benutzer oder Administrator nicht einsehbar oder modifizierbar ist.

Beim Anschluss einer zBX wird die Rolle der HMC mittels einer zusätzlichen Funktion drastisch aufgewertet, dem "Unified Resource Manager". Der Unified Resource Manager übernimmt Management Funktionen der zBX Blades. Dabei werden teilweise bereits vorhandene Mainframe Werkzeuge in angepasster Form verwendet. Die zBX Blades werden auf die gleiche Art administriert und konfiguriert wie die bisher existierenden Komponenten eines Mainframe Servers.

Die Unified Resource Manager Funktionen sind:

- Operational controls (Operations)
- Virtual server lifecycle management (Virtual servers)
- Hypervisor management (Hypervisors)
- Energy management (Energy)
- Network management (Networks)
- Workload Awareness and platform performance management

Das Management einer heterogenen Umgebung bestehend aus PowerPC und x86 Prozessoren und AIX, Linux und Windows Betriebssystemen soll damit langfristig so vereinfacht werden wie das heute nur in und zwischen Mainframe Komponenten möglich ist. Dies ist allerdings ein langfristiges Ziel, von dem bisher nur Teile verwirklicht worden sind.

Ensemble

Für Administrationszwe ist es möglich, mehrere z196 Server zu einem als "Ensemble" bezeichneten Verbund zusammenzuschließen. Ein einziger z196 Server mit angeschlossener zBX (1 – 4 Frames) wird hierbei als "Node" bezeichnet. Ein Ensemble kann aus bis zu 8 Nodes bestehen.

In dem unten gezeigten Beispiel sind drei Nodes (drei z196 Server, teilweise mit angeschlossenen zBXs,) zu einem Ensemble zusammengeschlossen. Eine einzige HMC mit installierter Unified Resource Manager Firmware, ist in der Lage, alle angeschlossenen z196 Server einschließlich ihrer lokal angeschlossenen zBXs zu administrieren. Aus Gründen der Zuverlässigkeit ist wie immer eine 2. HMC als Backup vorhanden (Primary uns Alternate HMC).

Bitte beachten: Ein Ensemble ist eine administrative Struktur. Unabhängig davon können die z196 Rechner des Ensembles einen oder mehrere Sysplexe oder GDPS implementieren. Dies wird wohl eher die Regel als die Ausnahme sein.



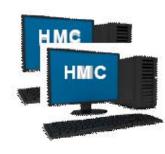


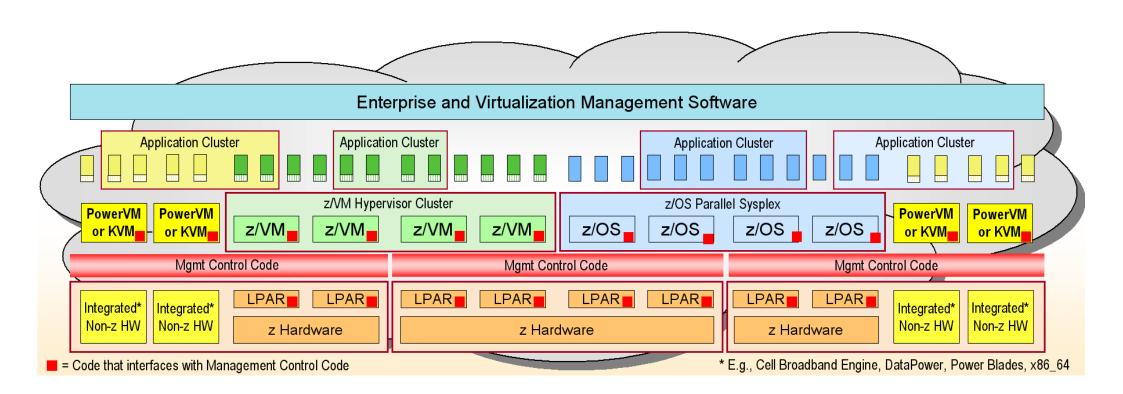
One ensemble

Three nodes









Virtualisierung

Die auf den Blades laufenden Betriebssysteme (AIX, Linux, Windows) werden häufig unter einem Hypervisor laufen: KVM für Windows und Linux, PowerVM für AIX. KVM ist ein Public Domain Hypervisor und unter anderem Bestandteil der Red Hat und OpenSuse Distributionen.

Auf dem Mainframe können mehrere LPARs mit z/OS unter PR/SM laufen. Eine oder mehrere LPARs können nauch z/VM enthalten, wobei unter z/VM wiederum zahlreiche virtuelle Maschinen, z.B. z/OS oder zLinux laufen können.

All dies erfordert eine komplexe Management Struktur.

Virtualisierungsmanagement

Diese Virtualisierungsumgebungen, zusätzlich zu VMWare Virtual Box, Xen und anderen sind heute Bestandteil einer IT Infrastruktur. zBX und Unified Resource Manager bieten als Verbesserung ein Integriertes Resource Monitoring, Workload Management, Image Management, Availability Management, Failure Management und Energie Management an.

Spezifisch existiert eine Unterstützung für "Federated Hypervisors": z/VM, PowerVM, and KVM, mit einem einheitlichem Satz von System Management Policies.

Trotzdem bleibt die Komplexität groß. Wir können an dieser Stelle weitere technologische Entwicklungen erwarten, z.B. :

- Integrationserweiterung der Blades für neue Architektur Plattformen, z.B GPU, FPGA, ARM, IPad, Android.
- Neue Special-Engines, z.B. mit Algorithmen die in FPGAs oder ASICs ausgeführt werden.
- Entwicklungswerkzeuge für für neue Architektur Plattformen.
- Ausdehnung des z/OS WLM auf zBX Blades
- zBX Coupling Facility.
- zBX GDPS.
- Shared Memory zwischen Mainframe und zBX, z.B. Hipersockets

Am Lehrstuhl Technische Informatik läuft derzeitig ein Forschungsprojekt, bei dem untersucht wird, die Funktionen des z/OS Workload Managers auf Blades in der zBX auszudehnen.