

14. Hybrid Computing

14.1 Installationsbeispiele

14.1.1 Geographically Dispersed Parallel Sysplex

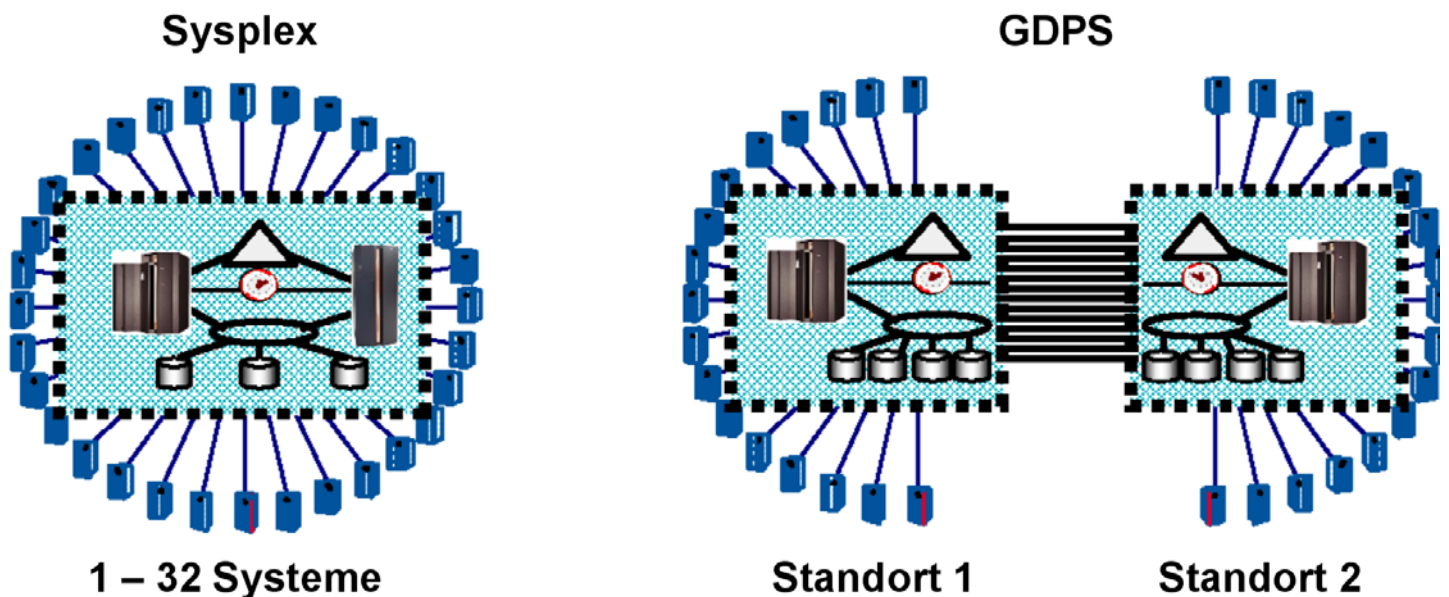


Abb. 14.1.1

Aufteilung des Rechenzentrums auf zwei Standorte

Ein (Parallel) Sysplex ist eine Konfiguration, bei der mehrere Mainframe Rechner zu einem Cluster (closely coupled) zusammengeschlossen sind. Dabei ist angenommen, dass die Rechner relativ nahe (Dutzende von Metern) zueinander untergebracht sind.

Ein „Geographically Dispersed Parallel Sysplex“ (GDPS) ist eine Konfiguration bei der die Rechner des Sysplex in zwei Rechenzentren untergebracht sind, die typischerweise 5 – 10 km voneinander entfernt sind.

14.1.2 Mainframe Installation

Abb. 14.1.2 zeigt eine typische Großrechner Konfiguration.

Die Konfiguration besteht aus zwei Rechenzentren (RZ 1 und RZ 2), die je einen (oder mehrere) Mainframe Rechner enthalten, und in einem Abstand von wenigen km untergebracht sind. Wenn ein Rechenzentrum ausfällt (Beispiel: ein Flugzeug stürzt auf dem Gebäude ab und setzt es in Brand), kann der Betrieb des Unternehmens mit dem 2. Rechenzentrum aufrecht erhalten bleiben. **Alle größeren Unternehmen unterhalten aus Gründen der Ausfallsicherheit zwei Rechenzentren.**

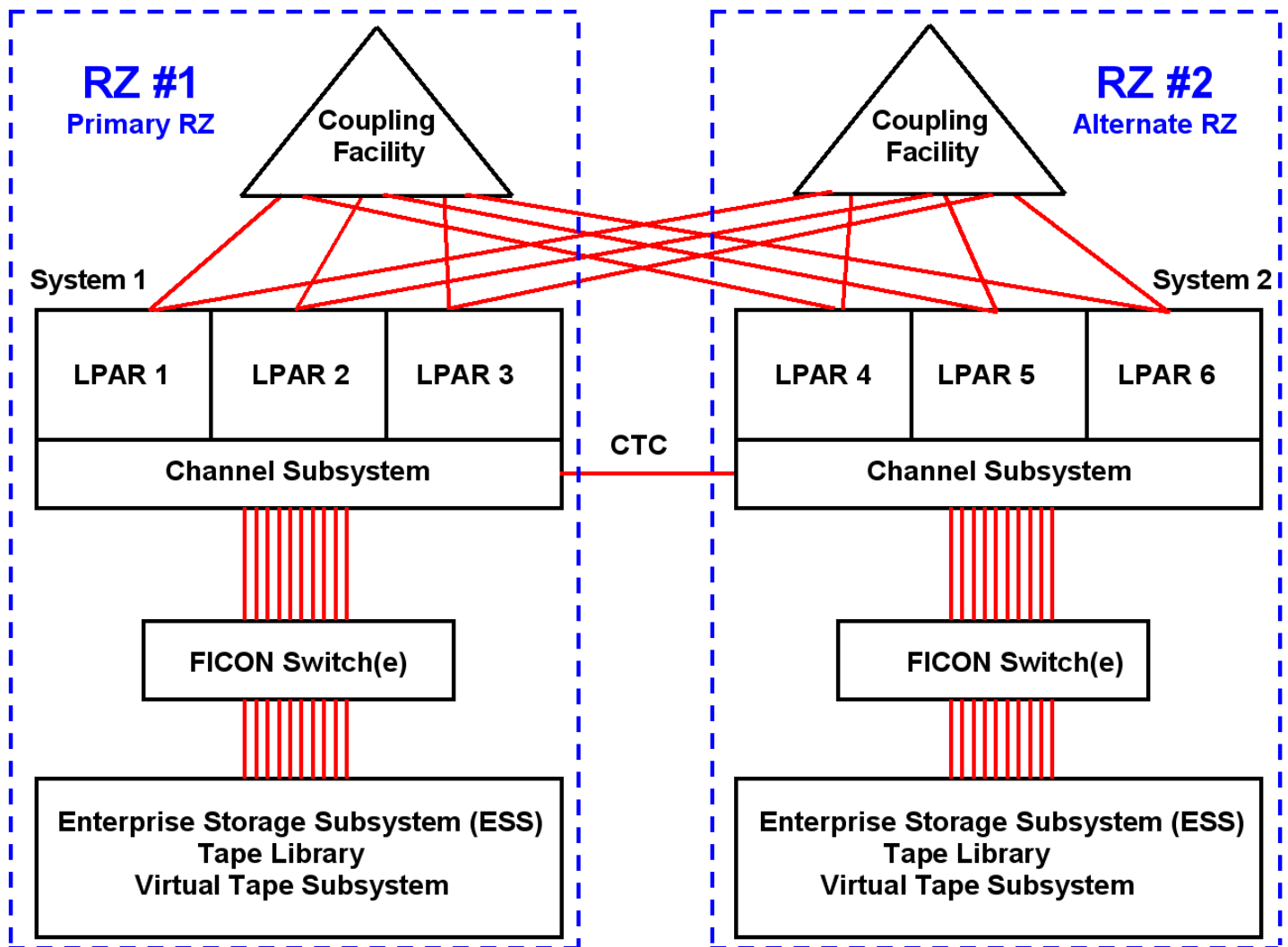


Abb. 14.1.2
Mehrfache LPARs in jedem Standort

Die Entfernung der beiden Rechenzentren untereinander beträgt typischerweise 5 – 10 km, in Ausnahmefällen 15 km. Man wünscht für den Katastrophenfall genügend Abstand. Auf der anderen Seite bewirkt die Lichtgeschwindigkeit (etwa 20 μ s für 5 km) eine unerwünschte Latenz für die Datenübertragung zwischen den beiden Rechenzentren. Abstände > 10 – 15 km erfordern eine andere Struktur als kleinere Abstände.

Ebenfalls aus Gründen der Ausfallsicherheit sind (mindestens) zwei Coupling Facilities vorhanden (je eine pro Rechenzentrum). Die Mainframes in den beiden Rechenzentren sind mit beiden Coupling Facilities verbunden.

Weiterhin sind die Mainframes in den beiden Rechenzentren über „Channel to Channel“ (CTC) Glasfaserkabel miteinander verbunden (Band 1, Abschnitt 5.3.6). In ganz großen Installationen befinden sich mehrere Mainframes in jedem der beiden Rechenzentren, z.B. 2 x 3 z196 Rechner bei Volkswagen und der UBS (Union Bank of Switzerland). In mittelgroßen Unternehmen existiert in jedem Rechenzentrum nur ein Mainframe Rechner mit weniger als maximaler Ausrüstung, z.B. mit nur 16 statt maximal 101 CPUs.

14.1.3 Datensicherheit

Alle kritischen Daten sind in beiden Rechenzentren auf getrennten Enterprise Storage Subsystemen und Tape Libraries dupliziert. Zahlreiche FICON Glasfaserverbindungen und mehrere FICON Switche (Director) stellen die Verbindung zu den Mainframes her. Modern ist der zusätzliche Einsatz eines „Virtual Tape Servers“ (VTS, Band 1 Abschnitt 5.4.6). Er enthält einen Plattenspeicher, der eine Art Cache für die Tape Library implementiert.

Typischerweise werden z.B. bei der Transaktionsverarbeitung die Datenbestände in beiden Rechenzentren **synchron** dupliziert. Eine Update Funktion erfolgt in beiden Rechenzentren (fast) gleichzeitig. Ein Commit erfolgt erst dann, wenn die Datenänderungen in beiden Rechenzentren erfolgreich abgeschlossen sind. Wegen der Latency der Signalübertragung benötigt dieses als „Metro Mirror“ bezeichnete Verfahren etwas mehr Verarbeitungszeit.

Eine als „Global Mirror“ bezeichnete **asynchrone** Alternative kann bei größeren Distanzen zwischen den beiden Rechenzentren eingesetzt werden, ist dafür aber nicht ganz so Katastrophensicher.

Eine Sysplex Konfiguration umfasst in jedem RZ mindestens eine LPAR, oft aber mehr. Es ist möglich und üblich, dass die LPARs in den beiden RZ als mehrere unabhängige und parallel laufende Sysplexe konfiguriert werden.

14.1.4 Mainframe Installation der Union Bank of Switzerland

Die Union Bank of Switzerland (UBS) ist die größte Bank in der Schweiz mit Hauptsitz in Zürich.

Die beiden Rechenzentren der UBS sind in 10 km Entfernung voneinander untergebracht. In jedem Rechenzentrum befinden sich (Stand 2008) 5 Mainframes plus zusätzlich drei weitere Mainframes, die als Coupling Facilities eingesetzt werden. Jedes Rechenzentrum enthält 12 FICON Switsche (Director) für den Anschluss der Plattenspeicher (Enterprise Storage Server) und Magnetband-Libraries.

Es ist durchaus üblich, dass Enterprise Storage Server und Magnetband-Libraries von anderen Firmen als IBM stammen. Bei der UBS stammen die Enterprise Storage Server teilweise von der Firma Hitachi. Die Magnetband Library stammt von der Firma Storage Tek, und kann insgesamt 26 400 Magnetbandkassetten aufnehmen. Zusätzlich werden Virtual Tape Server eingesetzt.

Anmerkung: In 2011 wurden die vorhandenen 2 x 5 älteren Mainframes durch 2 x 3 z196 Rechner mit insgesamt wesentlich höherer Leistung ersetzt. Vier weitere Mainframe Rechner dienen als Coupling Facilities.

Mainframe CPU Leistung wird traditionell in MIPS (Million Instructions per Second) gemessen. Hierzu dient ein IBM proprietäres Benchmark, welches in Assembler geschrieben ist und deshalb nicht auf andere Plattformen portiert werden kann. Mainframe Performance Vergleiche mit anderen Architekturen sind fast unmöglich. Als Faustformel kann dienen, dass ein moderner zEC12 Mikroprozessor marginal schneller als ein Hochleistungs- x86 oder PowerPC Mikroprozessor ist.

Site 1

Site 2

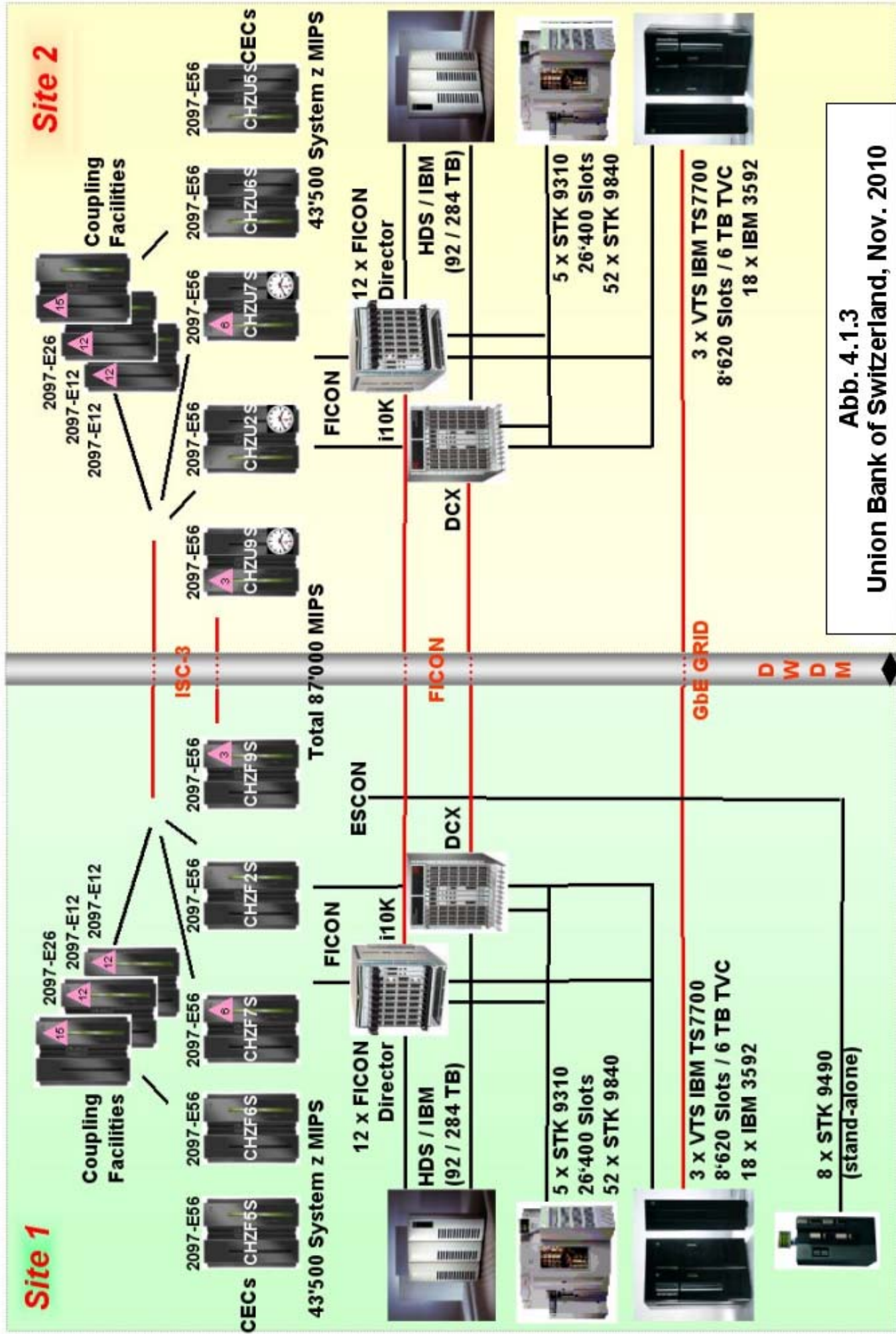


Abb. 4.1.3
Union Bank of Switzerland, Nov. 2010

STK 9310, 9490 StorageTek Cartridge Subsystem, HDS Hitachi Data Systems | 10 km

14.1.5 Multiple Sysplexes

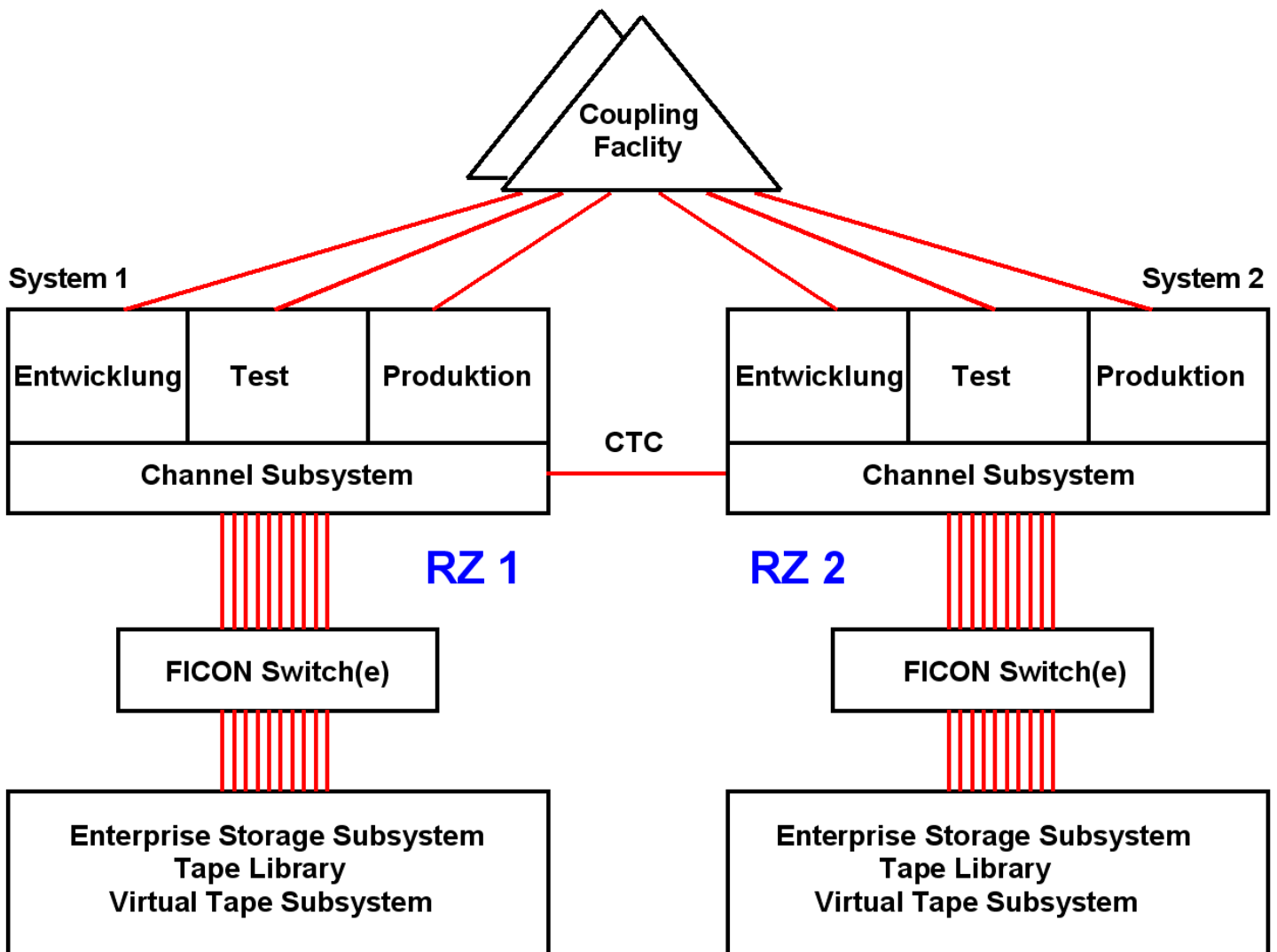


Abb. 14.1.4
Multiple Sysplex Configuration

Der eigentliche Haupteinsatz eines z/OS Rechners für Aufgaben wie Transaktionsverarbeitung, Stapelverarbeitung und Datenhaltung wird als "Produktion" bezeichnet.

Neben einem Sysplex für die Produktion existiert typischerweise ein getrennter Sysplex für die Entwicklung neuer Anwendungen. Ein weiterer Sysplex wird für das Austesten neu entwickelter Anwendungen benötigt, ehe sie für die Produktion frei gegeben werden.

Häufig existieren weitere Sysplexe für Sonderaufgaben.

Es ist ohne weiteres möglich (und üblich), mehrere logische Sysplexe auf eine physische GDPS Infrastruktur abzubilden. Dabei verfügt in jedem Rechenzentrum ein Sysplex über mindestens eine LPAR, wobei auch mehrere LPARs denkbar sind.

14.1.6 CreditSuisse

Die *Credit Suisse* Group ist die zweitgrößte Bank in der Schweiz mit Hauptsitz in Zürich.

Die beiden Rechenzentren in Zürich und Horgen sind etwa 15 km voneinander entfernt.

In beiden Rechenzentren stehen jeweils zwei physische Mainframe Rechner (PC10 und PC20 in Horgen) sowie PC60 und PC70 (in Zürich). Auf jedem der 4 Rechner befinden sich mehrere LPARs, z.B. die LPARs mit den Namen S11, GA1, S21, GB1, R21 usw. auf dem Rechner PC10 am Standort Horgen.

Unterhalten werden 3 logische Sysplexe für Entwicklung (RZ1), Test (RR2) und Produktion (RZ2). Hierbei erhält der Produktions-Sysplex den größten Teil der verfügbaren CPU Kapazität (40 00 MIPS) im Vergleich zu jeweils 5 000 MIPS für die beiden anderen Sysplexe, siehe Abbildung auf der folgenden Seite. Diese drei Sysplexe verfügen über LPARs auf allen 4 Rechnern.

Daneben existieren 6 weitere logische Sysplexe (RQ2, RZ4, RZ0, RZ8, RZZ und ZVM), die für Spezialaufgaben eingesetzt werden. Sie verfügen über weit weniger CPU Leistung (100 – 700 MIPS), und sind teilweise auf einen der beiden Standorte begrenzt. Auf z/VM laufen eine Reihe von zLinux virtuellen Maschinen für Entwicklungs- und Testzwecke.

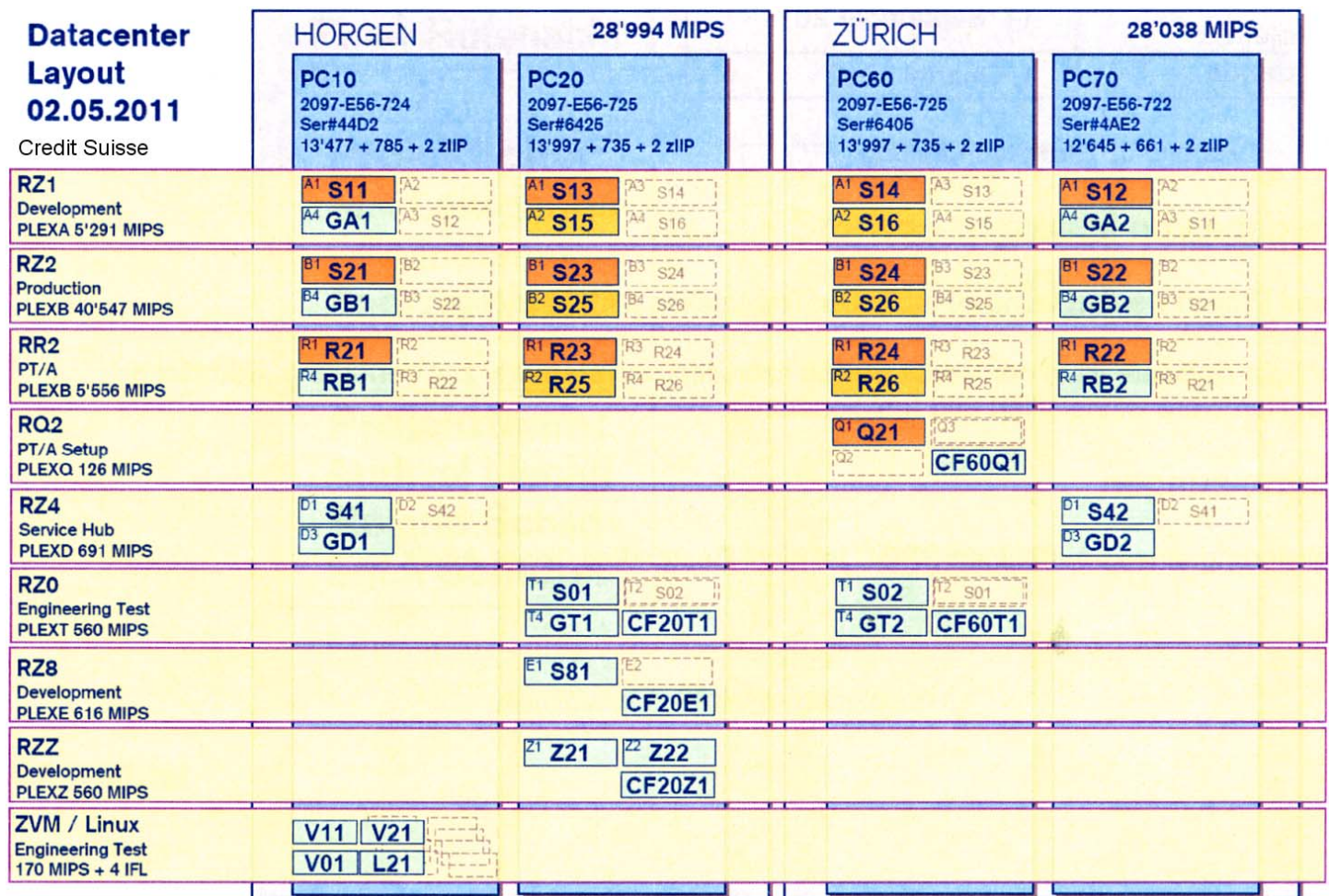


Abb. 14.1.5
Data Center Layout der Firma Credit Suisse, 2011

Kenndaten des Credit Suisse Data Centers:

Server

	6 751
• Windows Application Servers	2 726
• Unix Servers	3 131
• File- & Print-Server	894
• Storage Capacity	4,600 PByte

Mainframes

• Mainframe Systems (plus weitere in anderen Standorten)	4
• Installed MIPS	70 000
• Storage Capacity (DASD)	1 008 PByte
• Anzahl Datenbanken	54 500
• Anzahl IMS Transactions	27 Million pro Tag
• Anzahl Batch Jobs	86 000 pro Tag
• Tape Library Storage (plus weitere 13 PByte Shelf storage)	4.6 PByte

Kenndaten des Production-Sysplex:

- 100 Million MQ-Gets und 0.5 TByte Daten verarbeitet pro Tag
- 12 Million Corba Calls für IMS oder CICS pro Tag
- 20 Million CICS Transactions pro Tag (peak)
- 37.6 Million IMS Transactions pro Tag (peak)
- 220 IMS Databases
- 76 200 DB2 Tables
- 4.9 Billion SQL Calls pro Tag (peak)

50 000 unterstützte Benutzer in 550 Locationen

97 000 Workstations / Laptops

75 000 email Accounts

339 Millionen Emails/Jahr

100 Million MQ-Gets und 0.5 TB data processed per day besagt, dass WebSphere MQ in großem Umfang eingesetzt wird.

12 Million Corba Calls for IMS oder CICS per day besagt, dass RMI/IIOP in großem Umfang eingesetzt wird. RMI/IIOP wird später in Kapitel 6 behandelt.

Credit Suisse setzt sowohl DB2 wie die nicht relationalen IMS Datenbanken ein. Dies ist nicht untypisch. Weltweit sind auf allen in der Wirtschaft und in der Verwaltung eingesetzten Mainframe Rechnern etwa 60 % des Datenvolumens in DB2 Datenbanken, 30 % in IMS Datenbanken und 5 % in Adabas Datenbanken gespeichert. (Das VSAM Datenvolumen ist hierbei nicht berücksichtigt). Dabei existieren große Abweichungen von Unternehmen zu Unternehmen. So existieren zahlreiche Fälle, in denen nur DB2 oder nur IMS eingesetzt wird.

Für IMS existiert ein eigener Transaktionsmonitor, der an Stelle von CICS eingesetzt wird. Im Falle der Credit Suisse übersteigt das IMS Transaktionsvolumen das CICS Transaktionsvolumen, was allerdings nicht sehr typisch ist.

Credit Suisse Storage Management:

Plex	Data-Class	Storage-Class	Management-Class	Storage-Group
RZ0	130	24	90	32 (davon 5 tape, 1 VIO)
RZ1	136	26	96	30 (davon 2 tape, 1 VIO)
RZ8*	133	18	83	23 (davon 2 tape, 1 VIO)
PLEXZ	106	26	93	25 (davon 2 tape, 1 VIO)
RR2	108	28	100	37 (davon 2 tape, 1 VIO)
Q21	108	28	100	37 (davon 2 tape, 1 VIO)
RZ4	103	24	93	24 (davon 1 tape, 1 VIO)
RZ2	108	28	100	37 (davon 2 tape, 1 VIO)
Restliche*	(n/a)	(n/a)	(n/a)	(n/a)

Mai 2011

Abb. 14.1.6
Von der Credit Suisse verwendete DFSMS Konstrukte

In Band 1, Abschnitt 3.3.9 wurde dargestellt, wie mit Hilfe der z/OS DFSMS Komponente Data Sets allocated und jeweils einer Data Class, einer Storage Class und einer Management Class zugeordnet werden müssen.

Plattenspeicher (Volumes) mit identischer Data Class, Storage Class und Management Class werden zu einer Storage Group zusammengefasst.

Abb. 14.1.6 stellt dar, wie viele Data Classes, Storage Classes und Management Classes sowie Storage Groups in den einzelnen Sysplexes bei der Credit Suisse vorhanden sind. Die große Anzahl verdeutlicht die Komplexität der Datenverwaltung in einem modernen Rechenzentrum.

VIO (Virtual I/O) ist eine andere Bezeichnung für Virtual-Tape-Server, Band 1, Abschnitt 5.4.6.

Im Vergleich dazu verfügt unser Rechner am Lehrstuhl Technische Informatik der Uni Leipzig über je 1 Data Class, Storage Class, Management Class und Storage Group.

14.1.7 FIDUCIA IT AG , Karlsruhe

Die Fiducia ist ein führender IT-Dienstleister. Sie betreibt IT-Outsourcing für 790 Volksbanken und Raiffeisenbanken, Privatbanken sowie Zentralinstitute. Installierte Endgeräte bei den angeschlossenen Banken und Instituten:

- 100.998 PC-Arbeitsplätze in den Banken
- 6.611 Server – 7 x 24 Stunden im Online-Betrieb, davon 3.885 Solaris Server, u.a. distributed WebSphere
- 10.971 Geldautomaten
- 10.757 Kontoauszugsdrucker

Installation:

- 2 Rechenzentren, 7 x 24 Stunden im Online-Betrieb
- 5 x z10 Rechner (59.005 MIPS), je 512 GByte, 5 LPAR/z10, bis zu 25 CPU/LPAR
- 612 TB Plattenspeicherkapazität HDS, EMC
- 7.492 TB Kassettenpeicher-Roboter
- 40 % IMS, 60 % DB2

Verarbeitet über 16,22 Milliarden Transaktionen pro Jahr.

Spitzenvolumen am 2.1.2009, IMS/DC

- 70 Millionen Host-Transaktionen/Tag
- 3.062 Host-Transaktionen/Sek.

30 Mil. LOC (Lines of Code) in Cobol; weitere 30 Mill. LOC in Java

14.1.8 Generali Versicherung, Aachen

Als weiteres Beispiel sei die Generali Information Services erwähnt. Installiert sind (Mai 2010)

- 15 000 MIPS mit 15 % MIPS Wachstum / Jahr
- ca. 100 TByte Disk, 30 000 aktive Benutzer
- Print Output: 270 Mill. Seiten / Jahr, 50 Mill. Briefe / Jahr.
- 1,2 Mill. € Energiekosten / Jahr

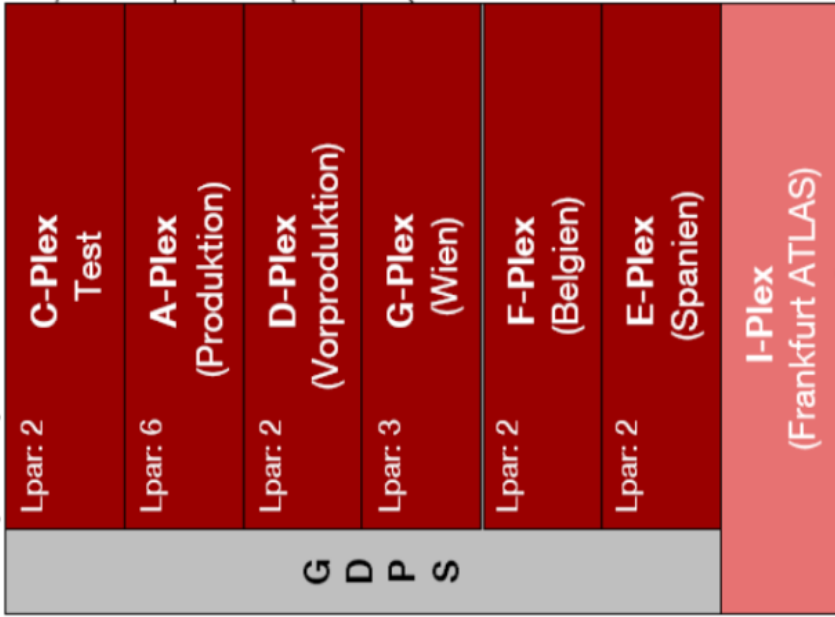
Generali Information Services beschäftigt 650 Mitarbeiter in der Anwendungsentwicklung

WebSphere wurde offloaded nach AIX auf System p; Domino auf System x Blade Server

Generali benutzt eine typische GSPS Konfiguration mit drei Sysplexes für Entwicklung, Produktion und Test, sowie weiteren Sysplexes für die Tochtergesellschaften in Belgien, Österreich und Spanien.

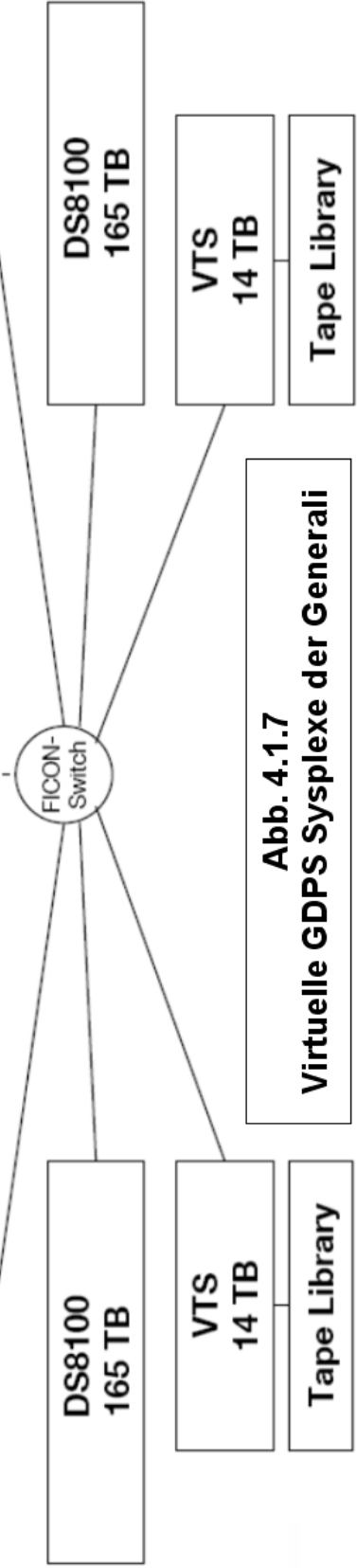
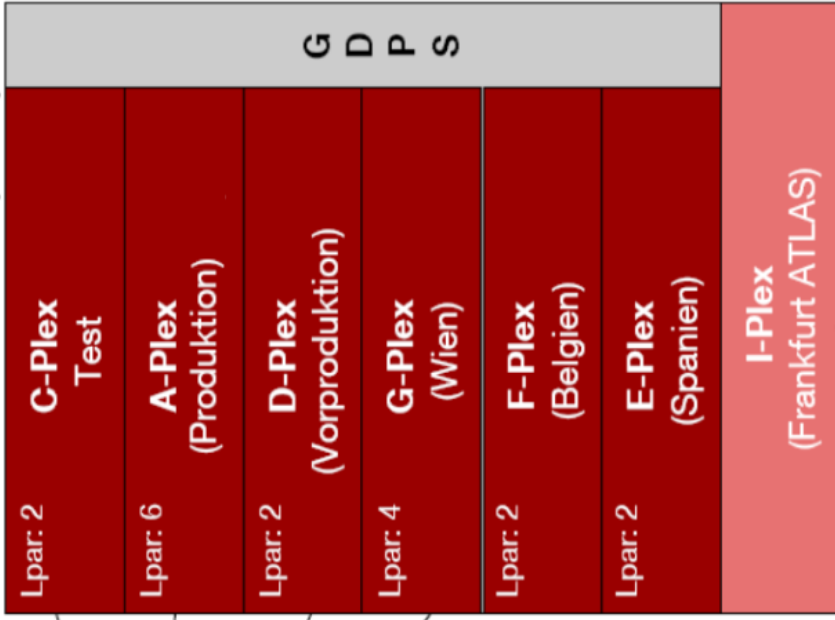
Ein Virtual Tape Server (VTS, anderer Name Virtual I/O, VIO) ist ein Speicher auf Basis eines Festplatten Arrays, der nach außen hin eine Tape-Library emuliert. Siehe Band 1, Abschnitt 5.4.6.

IVZ1 (CPU1) -725



**GDPS der Generali
Versicherung Aachen**

IVZ2 (CPU2) -725



14.1.9 Rechenzentrum der gkv Informatik AG in Wuppertal

Die gkv Informatik ist ein IT-Dienstleister für gesetzliche Krankenversicherungen im Verbund von AOKs und BARMER GEK. Die Unterhält die IT Infrastruktur für 37 000 Mitarbeiter.

Bewältigt den zentralen Beitragseinzug für 16 Millionen Versicherte. Stand Dezember 2010:

- zentrales Rechenzentrum mit 1.600 qm Nutzfläche,
- 20.000 Glasfaser Kabel mit einer Gesamtkabellänge > 100 km
- 4 Diesel- Notstromaggregate mit 4,4 MWatt Leistung,
- 4 IBM z10-Mainframes mit 40.000 MIPS,
- 400 TByte Festplatten Kapazität
- sechs Bandrobotersysteme mit 3.500 TByte Kapazität ,
- 400 UNIX-Server (AIX, Solaris, Linux)
- 3.000 Windows-Server, davon 1.500 in Citrix Server-Farmen.
- Serverfarm mit 1.000 Bladesystemen für simultanen-Zugriff von > 12.000 Mitarbeitern
- 120.000 Clients (PCs, Notebooks, Thin-Clients, Multifunktionsdruckers)
- 700 Mitarbeiter der gkv Informatik sichern die kontinuierliche Erbringung der IT-Dienstleistungen
- VoIP-Netz mit 44.000 Endgeräten an 1.000 Orten

14.1.10 Kommunales Rechenzentrum St. Gallen

Das Verwaltungsrechenzentrum AG (VRSG) in St. Gallen ist ein Dienstleistungsunternehmen für Städte und Gemeinden in der östlichen Schweiz. Es bedient 260 Kunden (Kantone/Gemeinden/öffentliche Verwaltungen), 7500 Benutzer, 50 Mio. Franken Umsatz (2010), 220 Mitarbeiterinnen und Mitarbeiter. VSRG betreibt 2 Rechenzentren in der St.Leonardstr. in St. Gallen sowie im St. Gallen Stadtteil Winkeln. Die Entfernung zwischen beiden Standorten beträgt etwa 7 km. Es handelt sich um eine weniger große Installation.

An beiden Standorten ist je ein (kleinerer) z10 Rechner installiert (Sept. 2011). Die beiden Standorte implementieren zwei logische „Geographically Dispersed Parallel Sysplex“ (GDPS). Jeder Rechner verfügt über 14 Prozessoren plus SAPs. (Ein z10 Rechner kann maximal 64 Prozessoren plus 12 SAPs enthalten). Bei den Prozessoren (PUs in IBM Terminology) handelt es sich um 6 normale CPUs (CPs in IBM Terminology), sowie 8 Spezialprozessoren, spezifisch

- 2 zIIP (System z Integrated Information Processor) für DB2 Verarbeitung
- 2 zAAP (System z Application Assist Processor) für Java und XML Verarbeitung
- 3 IFL (Integrated Facility for Linux) für zLinux Verarbeitung
- 1 ICF (Integrated Coupling Facility). Dies ist ein Prozessor für eine LPAR, welche die Aufgabe einer Coupling Facility übernimmt.

zIIPs, zAAPs, IFLs und ICFs benutzen die gleiche Mikroprozessor Hardware wie CPUs (und SAPs), werden jedoch bei der System Installation für ihre Spezialaufgaben „charakterisiert“, was Konsequenzen in Bezug auf Performance, Verkaufspreis und Software Lizenz Gebühren hat.

Standort Leonhard
IBM 2097-604 (277 MSU)

4CP (+ 2 CBU) / 2 ZIIP / 2 ZAAP / 3 IFL / 1ICF

CBU = Capacity Back Up

Standort Winkel
IBM 2097-602 (149 MSU)
2CP (+ 4 CBU) / 2 ZIIP / 2 ZAAP / 3 IFL / 1ICF

nicht aktiviert	P1	VRSGPLX3	VRSGPLX3
Adabas Prod	P2		
Adabas Prod Tessin	P3		
Adabas Prod Auswertung	P4		
Websphere Prod Cluster	P5		
nicht aktiviert (BRS X6)	P6		
WAS Deployment Manager Prod	Q1		
nicht aktiviert	Q2		
Inubit Produktion (Websphere)	Q3		
DB2 Cluster	Q4		
Websphere Test Cluster	Q5		
nicht aktiviert	Q6		
nicht aktiviert	T1		
z/VM mit z/Linux Produktion Oracle und ACE	V1PL	z/VM	
z/VM mit z/Linux Test Oracle und ACE	V2TL		

X1	IMS Produktion
X2	nicht aktiviert (BRS P2)
X3	nicht aktiviert (BRS P3)
X4	nicht aktiviert (BRS P4)
X5	Websphere Prod Cluster
X6	zOSMF, Stor Mgmt, TPC/R, WAS Compute Grid
Y1	Adabas Test
Y2	IMS Test
Y3	Inubit Test (Websphere)
Y4	DB2 Cluster
Y5	Websphere Test Cluster
Y6	nicht aktiviert
T2	Test LPAR Gruppe Systeme
V1PW	z/VM mit z/Linux Produktion Oracle und ACE
V2TW	z/VM mit z/Linux Test Oracle und ACE

Abb. 4.1.8
Kommunales Rechenzentrum St. Gallen, Sept. 2011

14.2 Integrated Coupling Facility

14.2.1 Alternative zu einer Freestanding Coupling Facility

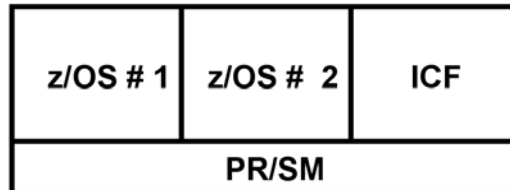


Abb. 14.2.1
Coupling Facility in einer LPAR eines Mainframe Rechners

Die Coupling Facility (CF) kann statt in einem getrennten Rechner auch in einer LPAR eines zSeries Knotens (System) untergebracht werden, wobei in anderen LPARs zwei oder mehr z/OS Images laufen. Diese Art der CF wird als Integrated Coupling Facility (ICF) bezeichnet. Die ICF läuft auf einer oder mehreren hierfür dedizierten CPUs eines SMP. ICFs haben in den letzten Jahren zunehmend an Popularität gewonnen. Das Kommunale Rechenzentrum St. Gallen benutzt Integrated Coupling Facilities.

Die ICF stellt CF Funktionalität ohne Coupling Links zu Verfügung. Letztere werden durch den PR/SM Hypervisor emuliert.

Aus Zuverlässigkeitsgründen sollten immer mindestens 2 physische oder integrierte CFs vorhanden sein.

Die entsprechende GDPS Konfiguration ist in Abb. 14.2.2 dargestellt.

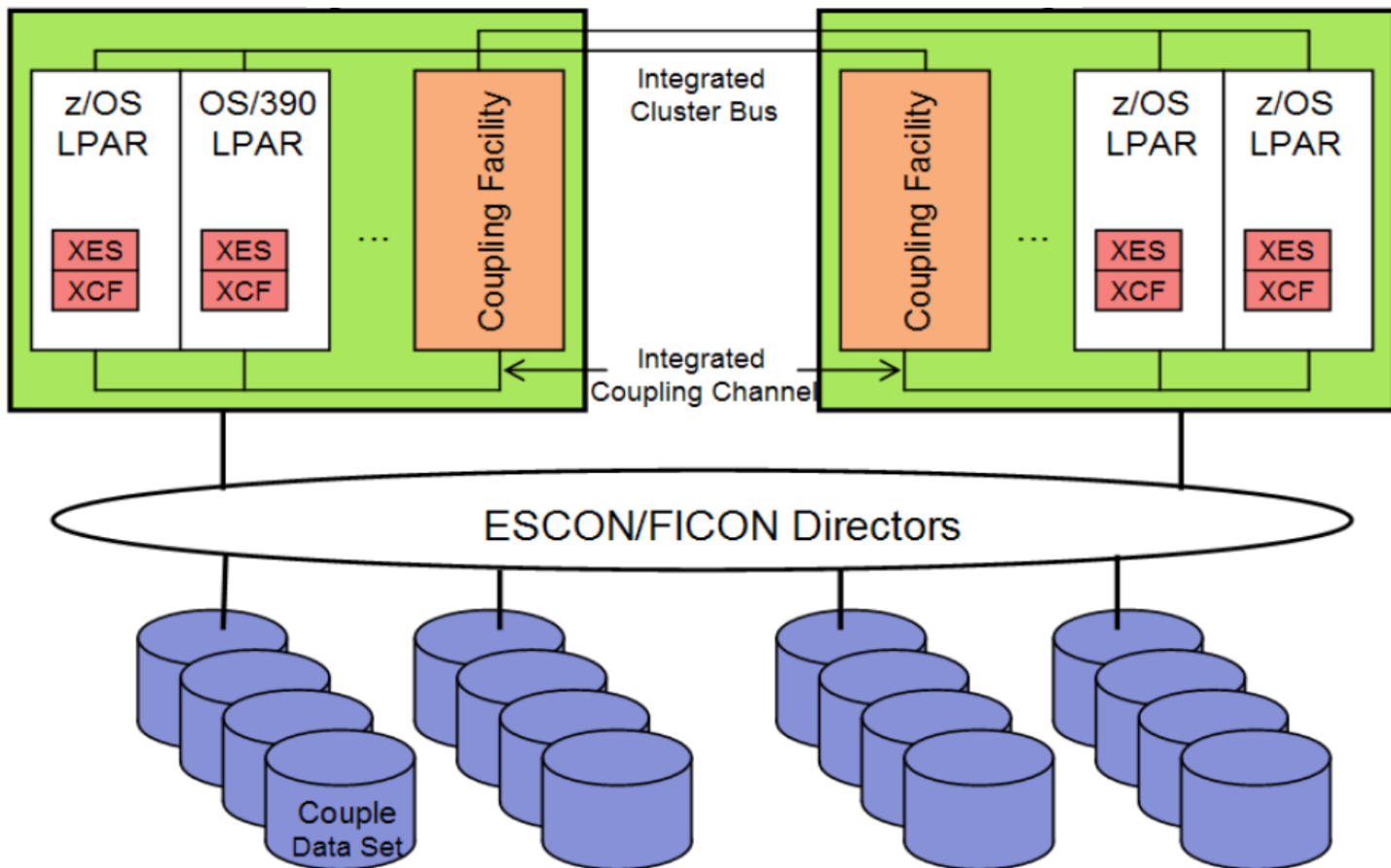


Abb. 14.2.2
Zwei Systeme mit integrierter Coupling Facility

Der Integrated Coupling Channel ist eine logische Struktur, welche die Aufgabe eines Coupling Links übernimmt, siehe Abschnitt 11.2.2.

14.2.2 Firma Endress+Hauser, Weil am Rhein

Ein faszinierendes Beispiel für eine ICF Installation ist die Firma Endress + Hauser. Die Endress+Hauser Gruppe ist einer der international führenden Anbieter von Messgeräten, Dienstleistungen und Lösungen für die industrielle Verfahrenstechnik. Das Unternehmen liefert Sensoren, Geräte, Systeme und Dienstleistungen für Füllstand-, Durchfluss-, Druck- und Temperaturmessung sowie Flüssigkeitsanalyse und Messwertregistrierung.

Zwei z10 EC (Modell E64) Systeme im Parallel-Sysplex-Betrieb. Pro System:

- 17 CPUs für z/OS Datenbankanwendungen
- 2 CPUs für ICF
- 32 CPUs für zLinux,
- 1,5 TB Hauptspeicher
- Taktfrequenz: 4,2 GHz
- Stellfläche: je 2 m² (!) pro System.

Unter zLinux laufen vor allem SAP Anwendungen, unter z/OS die dazugehörigen Datenbankanwendungen. Die HiperSockets-Technologie (Abschnitt 12.4.7) bietet extrem schnelle und sichere Konnektivität zwischen den SAP Anwendungs- und Datenbankservern.

19 SAP R/3 Produktionssysteme und 20 SAP R/3 Test-Systeme. verteilt auf 14 logische Partitionen (LPARs), laufen unter zLinux. SAP-R/3 Datenbanken auf der Basis von z/OS DB2 sind auf 6 logische Partitionen verteilt.

Unterstützt werden 3.700 registrierte Benutzer (drunter bis zu 2.500 gleichzeitige Benutzer). Diese Benutzer arbeiten in 35 Tochtergesellschaften an 71 Standorten weltweit und erhalten eine durchschnittliche Reaktionszeit von 0,5 Sekunden für SAP Anwendungen.

Endress + Hauser unterhält die vermutlich größte zLinux Produktionsinstallation in Deutschland.

Die folgende Abbildung 14.2.3 ist leider nur schlecht lesbar. Mit maximaler Vergrößerung werden jedoch viele Einzelheiten sichtbar.

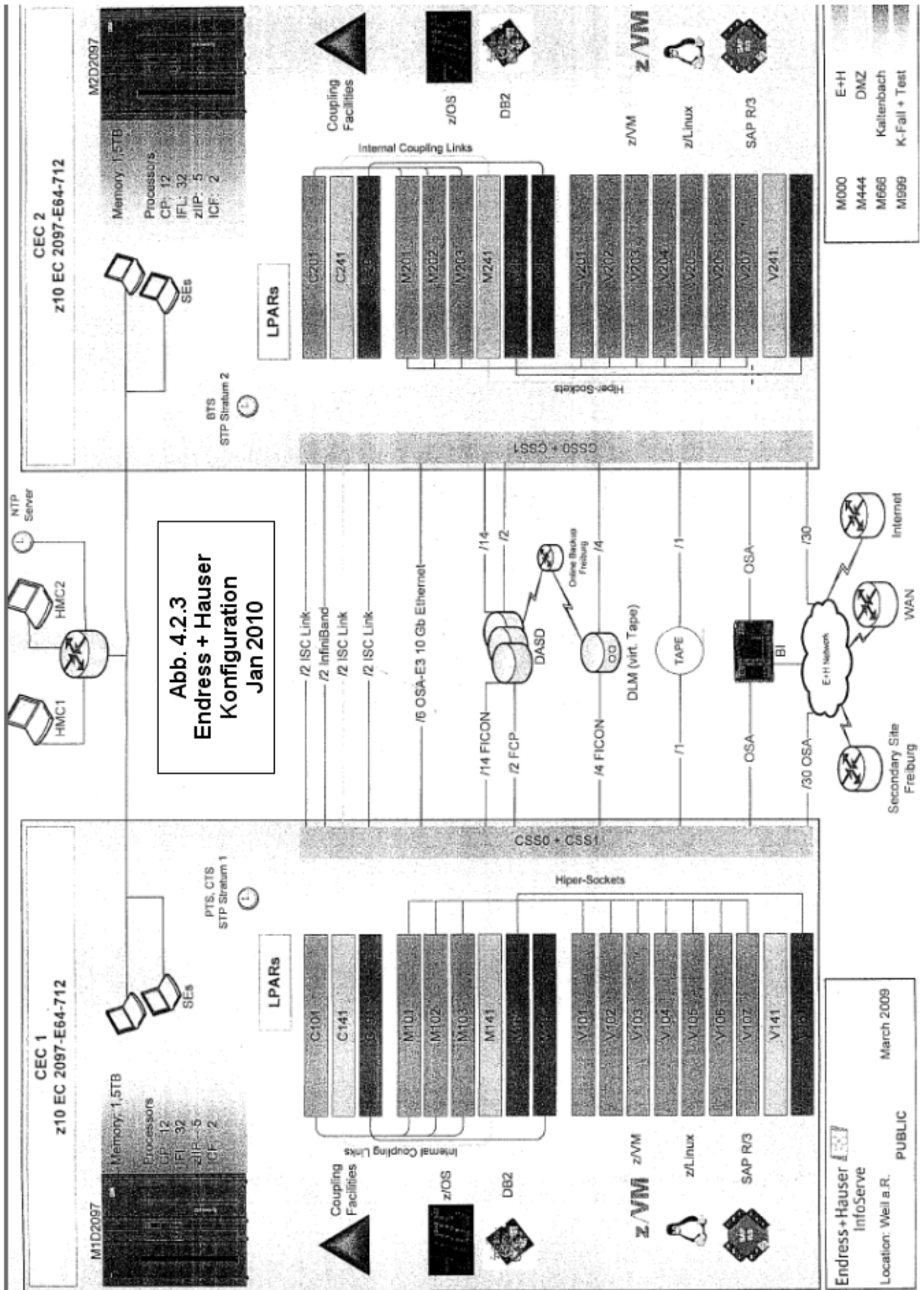


Abb. 4.2.3
Endress + Hauser
Konfiguration
Jan 2010

Endress + Hauser
InfoServe
Location: Weil a.R.
PUBLIC
March 2009

14.2.3 Hardware Management Console

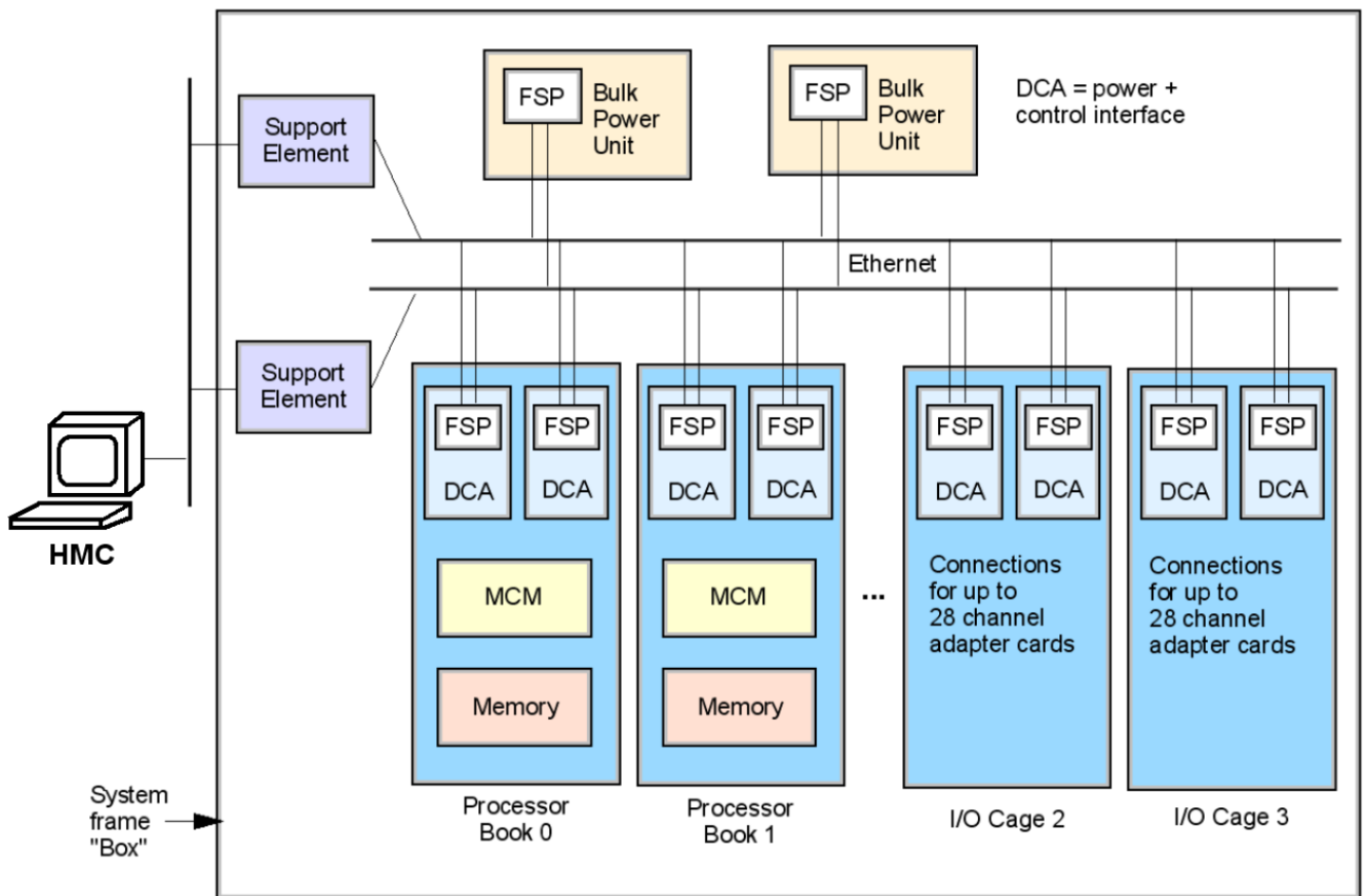


Abb. 14.2.4
Anschlüsse für das Support Element

Abbildung 14.2.4 zeigt das Innere eines System z Rechners. Für die Administration ist das Support Element (eingebauter Laptop, andere Bezeichnung Service Element) wichtig.

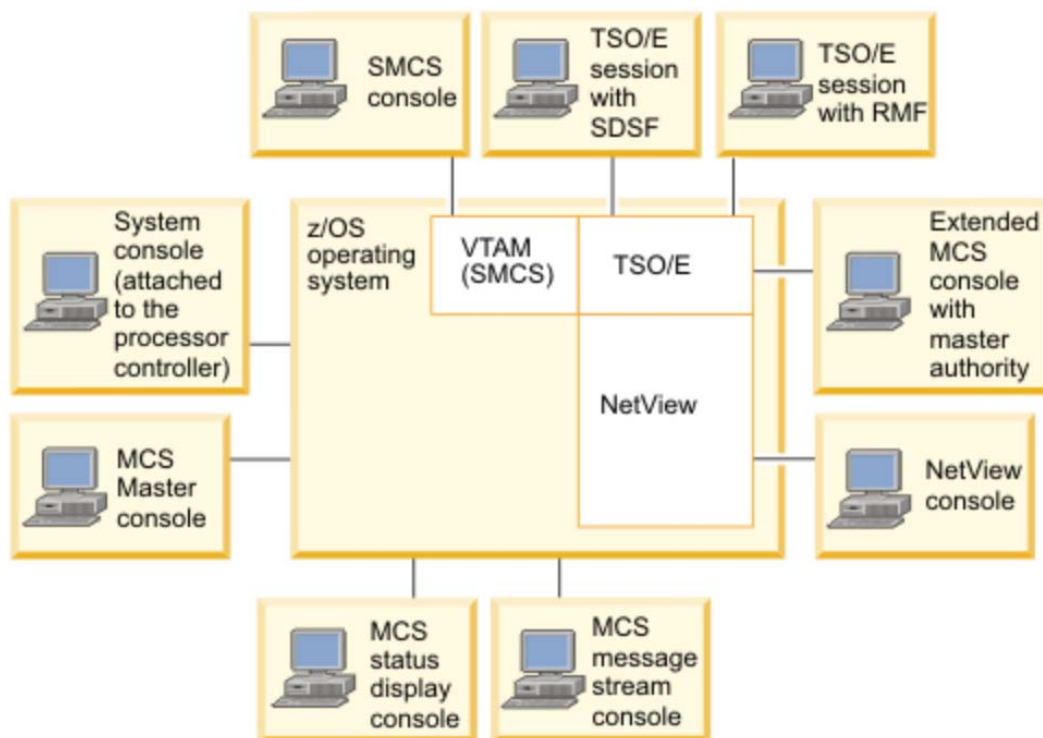
Das Support Element ist über eine interne Ethernet Verbindung mit einer ganzen Reihe von „Flexible Service Prozessoren“ (FSB) verbunden. FSBs sind kleine Baugruppen, die in die Books, I/O Cages und Power Supply Module integriert sind, siehe Band 1, Abschnitt 4.3.3. Sie erstellen kontinuierlich diagnostische Daten, die an das Support Element weitergegeben werden. Über das Support Element kann der Rechner auch gesteuert werden, z.B. konfiguriert oder hochgefahren werden.

Aus Gründen der Zuverlässigkeit sind das Support Element, die dazugehörigen FSBs und Ethernet Verbindungen doppelt vorhanden. Wenn ein Support Element ausfällt, kann der Betrieb mit dem zweiten Support Element fortgesetzt werden.

Die beiden Support Elemente sind in den Mainframe Rahmen fest eingebaut, und können nicht entfernt werden. Wenn ein Administrator das Support Element benutzen will, muss er die Tür des Mainframe Rahmens öffnen, um sich Zutritt zu verschaffen. Da dies lästig ist, existiert eine zusätzliche „Hardware Management Console“ (HMC), siehe Band 1, Abschnitt 1.3.11. Dies ist ein externer PC, der über ein unabhängiges Ethernetkabel mit den beiden Support Elementen verbunden ist.

Die HMC ermöglicht alle Funktionen, die auch mit dem Support Element durchführbar sind. Sie steht normalerweise in einem geschlossenen Raum mit kontrollierten Zutrittsrechten. Die HMC ist eine geschlossene Box, die bei einer gewaltsamen Öffnung selfdestructs. Die Kommunikation zwischen HMC und Support Element erfolgt kryptografisch verschlüsselt.

In einer Mainframe Installation existieren neben der HMC noch weitere Konsolen (Bildschirme) für die Administration des z/OS Betriebssystems, und für die Administration von Subsystemen wie z.B. DB2 oder WebSphere. Außerdem existieren in der Regel weitere zahlreiche Windows, Unix und Linux Rechner, die ebenfalls administriert werden müssen. Diese zahlreichen Bildschirm-Konsolen eines Unternehmens werden üblicherweise im gleichen Raum untergebracht und zu einem „Leitstand“ zusammengefasst.



Beispiel einer Console Konfiguration für eine Mainframe Installation

Die System Console ist eine Teilfunktion der Hardware Management Console.

14.2.4 Leitstand der Fiducia IT AG in Karlsruhe

Die Abbildungen 14.2.5 – 14.2.8 zeigen den Leitstand der Fiducia IT AG in Karlsruhe. Kennzeichnend ist neben den zahlreichen Computer-Bildschirmen eine zentrale Bildschirmwand, auf der selektiv die Inhalte einzelner Administrator-Arbeitsplätze wiedergegeben werden.

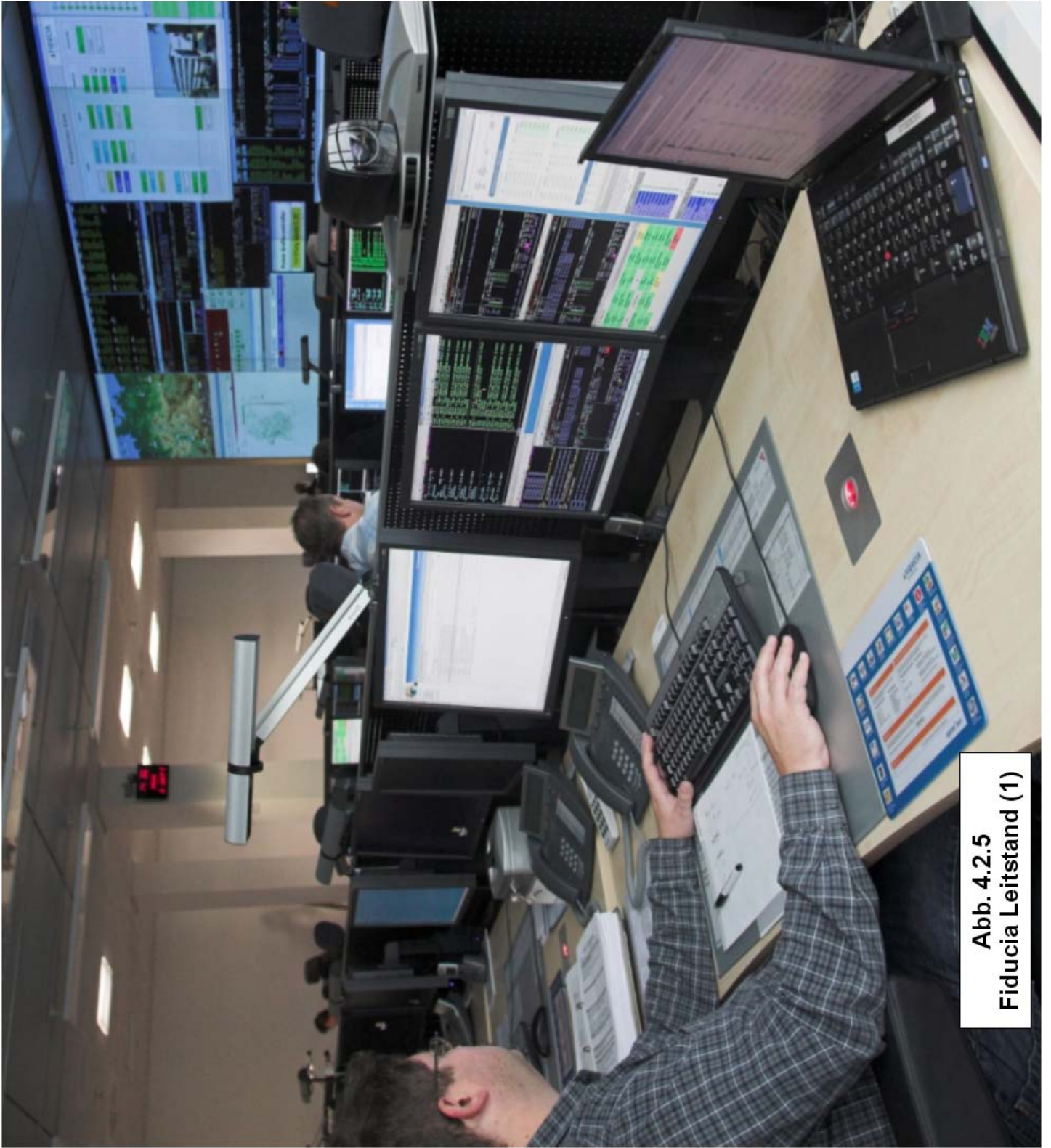
Der Leitstand steuert den Normalbetrieb und in einer Krise den Wiederanlauf aller Services.

Der Leitstand ist an 24 Stunden/Tag und 7 Tagen pro Woche besetzt. Die Systemadministratoren haben neben dem Platz im Leitstand noch einen normalen Büroarbeitsplatz für Ihre laufenden Tätigkeiten. Der normale Betrieb der IT Infrastruktur ist hochgradig automatisiert. Viele Funktionen sind nur für einen Störfall vorhanden, der hoffentlich nie auftritt.

Die Kosten, die bei einer Störung des IT Betriebes auftreten sind jedoch so astronomisch, dass alle Voraussetzungen geschaffen werden, um solche Störfälle möglichst rasch beenden zu können.

Bei (sehr seltenen) Störfällen versammeln sich der Krisenstab und die technischen Spezialisten in einem Konferenzraum, der durch eine Glaswand einen unmittelbaren Blick auf die Bildschirme des Leitstandes ermöglicht.

Leitstände, wie sie in der IT Industrie üblich sind, werden auch in anderen Industriezweigen eingesetzt, z.B. Kraftwerke, Raffinerien, Chemische Industrieanlagen, Raumfahrt, andere ...



**Abb. 4.2.5
Fiducia Leitstand (1)**

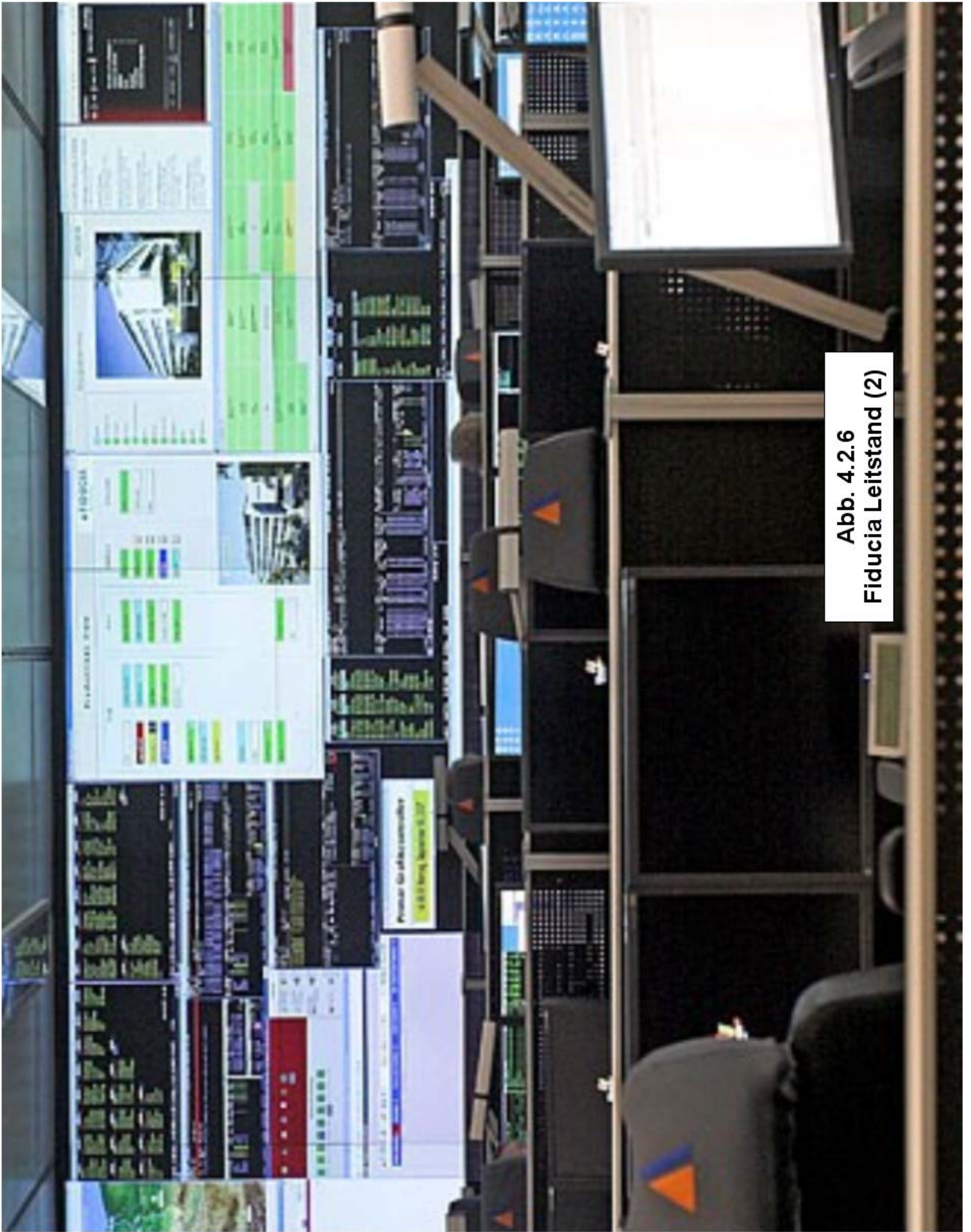


Abb. 4.2.6
Fiducia Leitstand (2)



Abb. 4.2.7
Fiducia Leitstand (3)

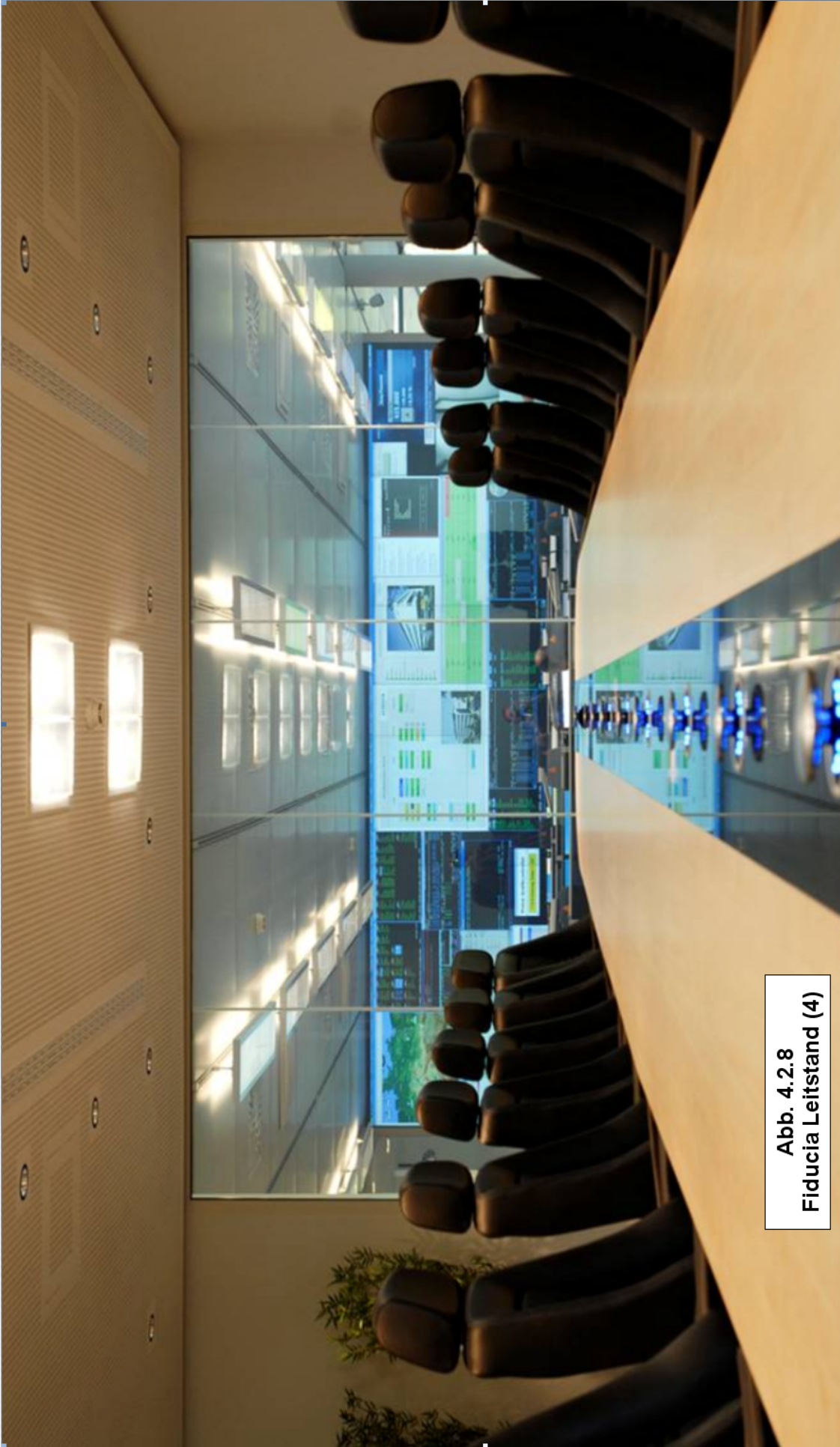


Abb. 4.2.8
Fiducia Leitstand (4)

14.2.9 Firma Jungmann Systemtechnik

Die Firma Jungmann Systemtechnik in Buxtehude, Landkreis Stade in Niedersachsen, ist der führende Hersteller von Leitständen in Deutschland.

Die folgenden Abbildungen 14.2.9 und 14.2.10 zeigen zwei Demonstrationsleitstände der Firma Jungmann.

Das Unternehmen liefert nicht nur die Computerindustrie, sondern auch andere Industriezweige wie z.B. Kraftwerke, Raffinerien, Chemische Industrieanlagen, Raumfahrt und andere.

Jungmann liefert auch ergonomisch besonders bequeme Sessel, die Ermüdungserscheinungen bei den Systemadministratoren verringern sollen.

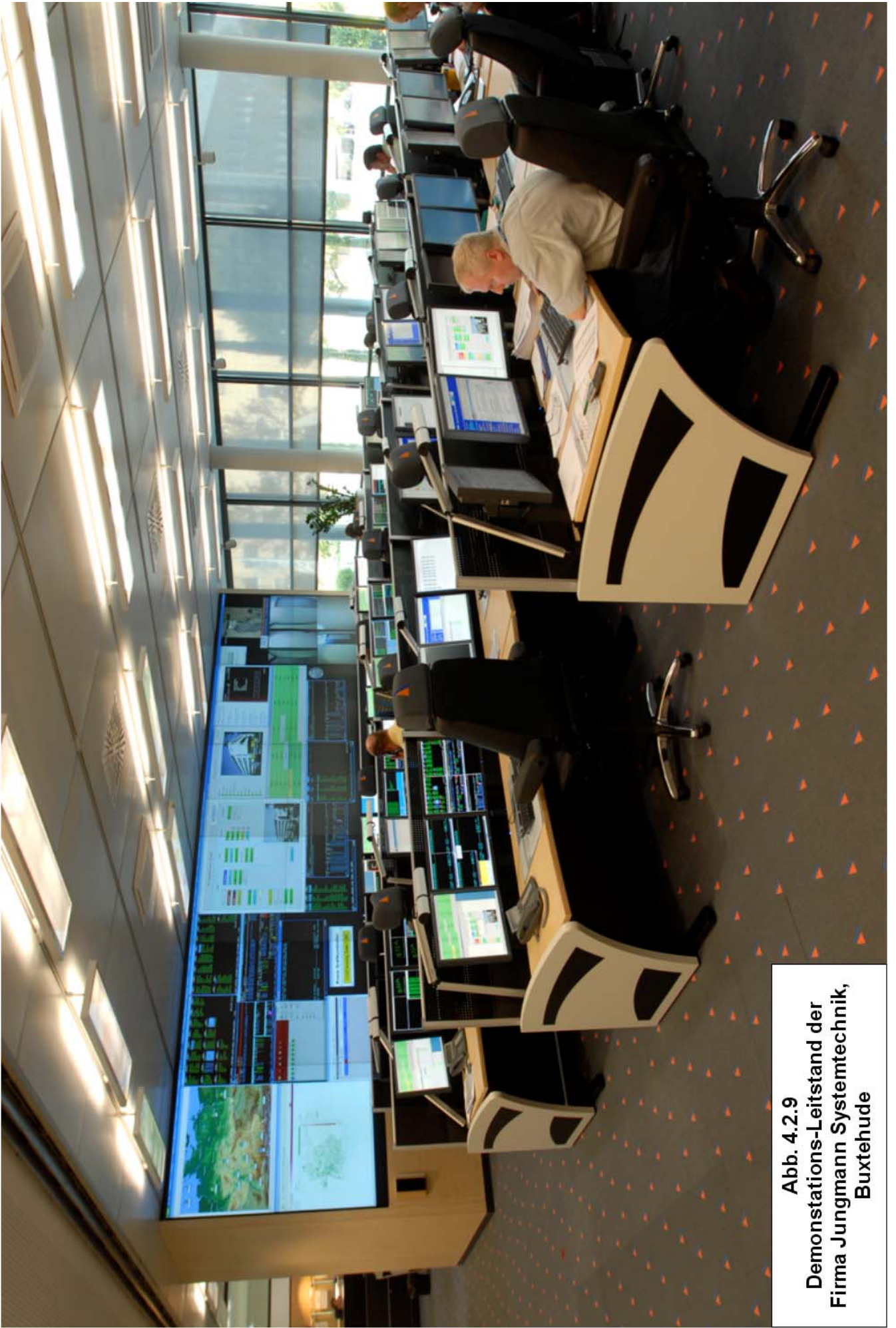


Abb. 4.2.9
Demonstrations-Leitstand der
Firma Jungmann Systemtechnik,
Buxtehude

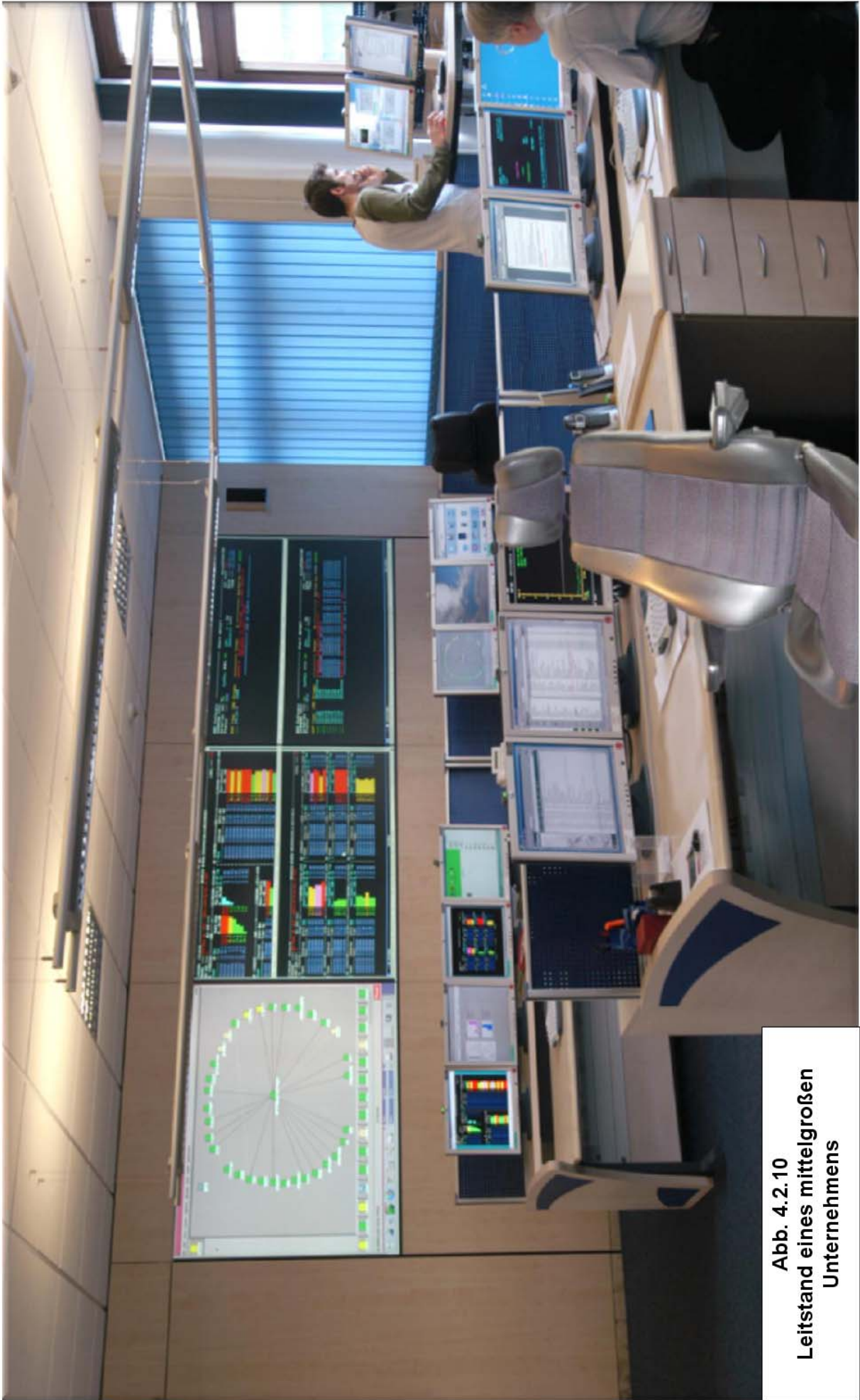


Abb. 4.2.10
Leitstand eines mittelgroßen
Unternehmens

14.3 zBX und der Unified Resource Manager

14.3.1 IT Infrastruktur in einem Großunternehmen

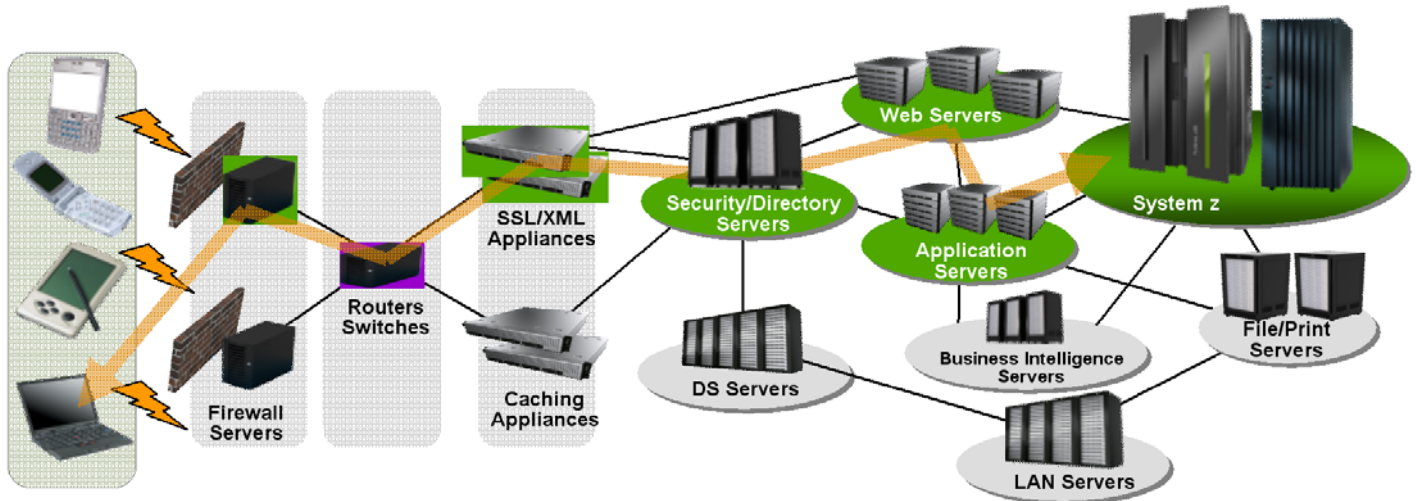


Abb. 14.3.1

Große Unternehmen benutzen Hunderte oder Tausende distributed Server

Die IT Infrastruktur eines Großunternehmens besteht neben den Mainframes typischerweise aus einer großen Anzahl weiterer nicht-Mainframe Server. Dies können viele tausend Server sein, die spezielle Aufgaben wahrnehmen, wie Kryptografie, Secure Socket Layer Verarbeitung, 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.

Siehe auch Abb. 14.3.2 , Das mittelgroße VRSG, Verwaltungsrechenzentrum AG St.Gallen, Dienstleistungsunternehmen für Städte und Gemeinden in der Ost-Schweiz, zeigt die Komplexität der distributed IT Infrastruktur (DMZ = Demilitarized Zone).

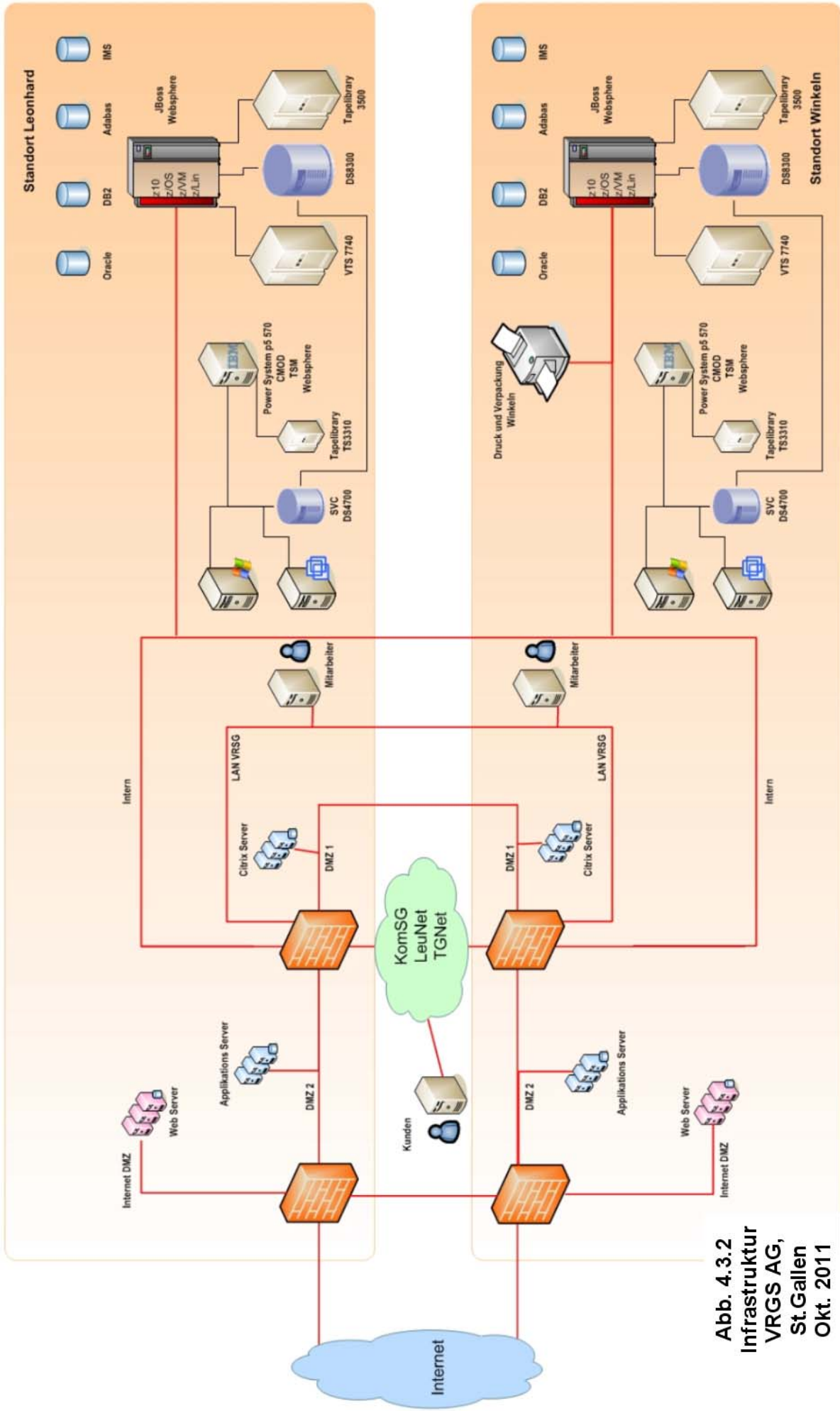


Abb. 4.3.2
 Infrastruktur
 VRGS AG,
 St.Gallen
 Okt. 2011

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“ (Microsoft Server, MS Directory Server, MS Transaction Server, MS SQL Server). Beim Focus Verlag in München werden seit Jahrzehnten 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.

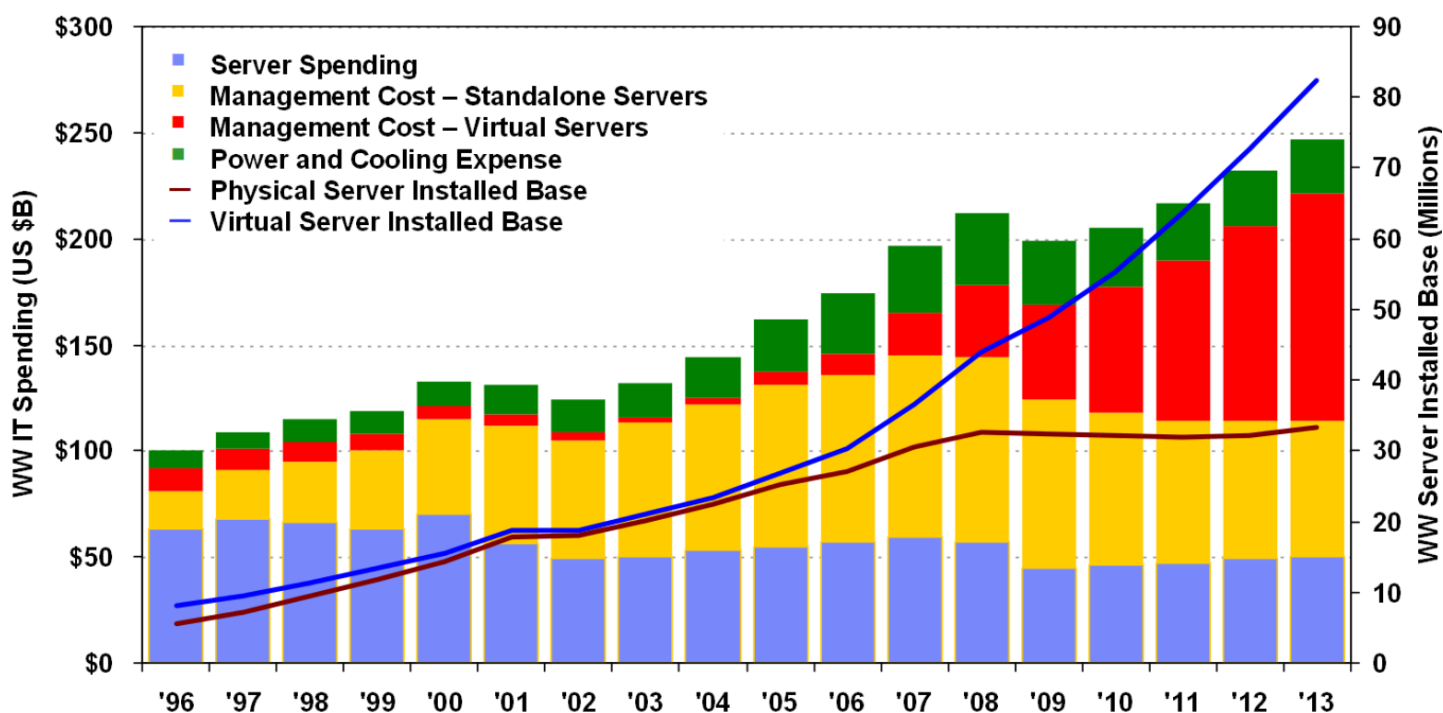


Abb. 14.3.3
Management Kosten übertreffen bei Weitem die Kosten für Hardware und Software

Source: IDC – “Three Data Centers – One Vision?”, March 2010

www.cedix.de/VorlesMirror/Band2/RayJones02.pdf

In vielen Unternehmen sind die distributed Server in den einzelnen Fachbereichen installiert und über das Unternehmen verstreut. Häufig wird versucht, die Server im Rechenzentrum zu konzentrieren um einen Teil der Kosten und des Administrationsaufwandes einzusparen. So sind z.B. bei der Fiducia AG, Karlsruhe neben den 5 Mainframes mit 60 000 MIPS weitere 6 611 Server vorhanden, davon heute (2011) 3.885 zentralisiert im Rechenzentrum.



Abb. 14.3.4
Server Farm im 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. Dies verringert deutlich den Administrations-Aufwand, der aber dennoch wesentlich höher als der Mainframe Administrations-Aufwand bleibt.

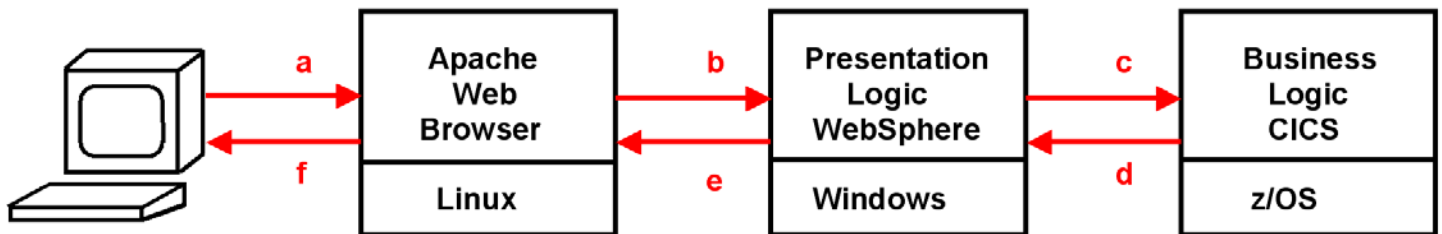


Abb. 14.3.5
Reichweite des Work Load Managers

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 einem „distributed“ WebSphere 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.

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.

www.cedix.de/VorlesMirror/Band2/DistrCost1.pdf

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.

www.cedix.de/VorlesMirror/Band2/DistrCost2.pdf

14.3.2 Blades

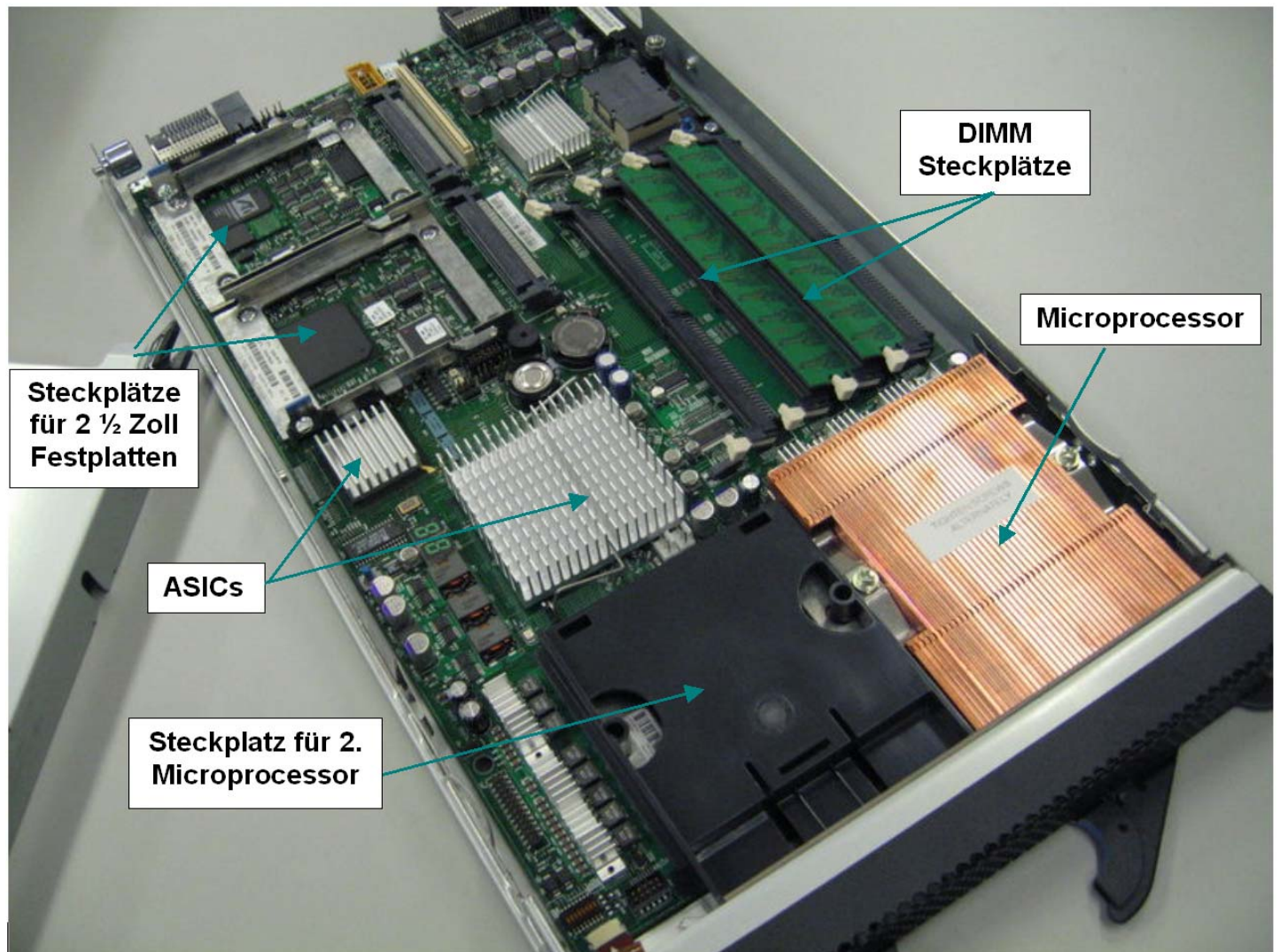


Abb. 14.3.6
IBM Blade für x86 und PowerPC Mikroprozessoren

Distributed Server bestehen in vielen Fällen aus Baugruppen, die als “Blades” bezeichnet werden. Eine Blade ist ein Printed Circuit Board (ähnlich einem PC Mainboard), auf dem sich typischerweise zwei Mikroprozessor 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 oder Fibre Channel SCSI Anschlüsse und eventuell Steckplätze für ein bis zwei 2 ½ Zoll Festplatten, auf denen das Betriebssystem untergebracht werden kann.

Blades werden in 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, oder alternativ mit proprietären Chips der Hersteller (HP Itanium, IBM PowerPC sowie Sun Sparc). Auf den Blades laufen Betriebssysteme wie Windows, Linux, Solaris, HP-UX, IBM AIX und Sun Solaris.

Abb. 14.3.6 zeigt eine IBM Blade.

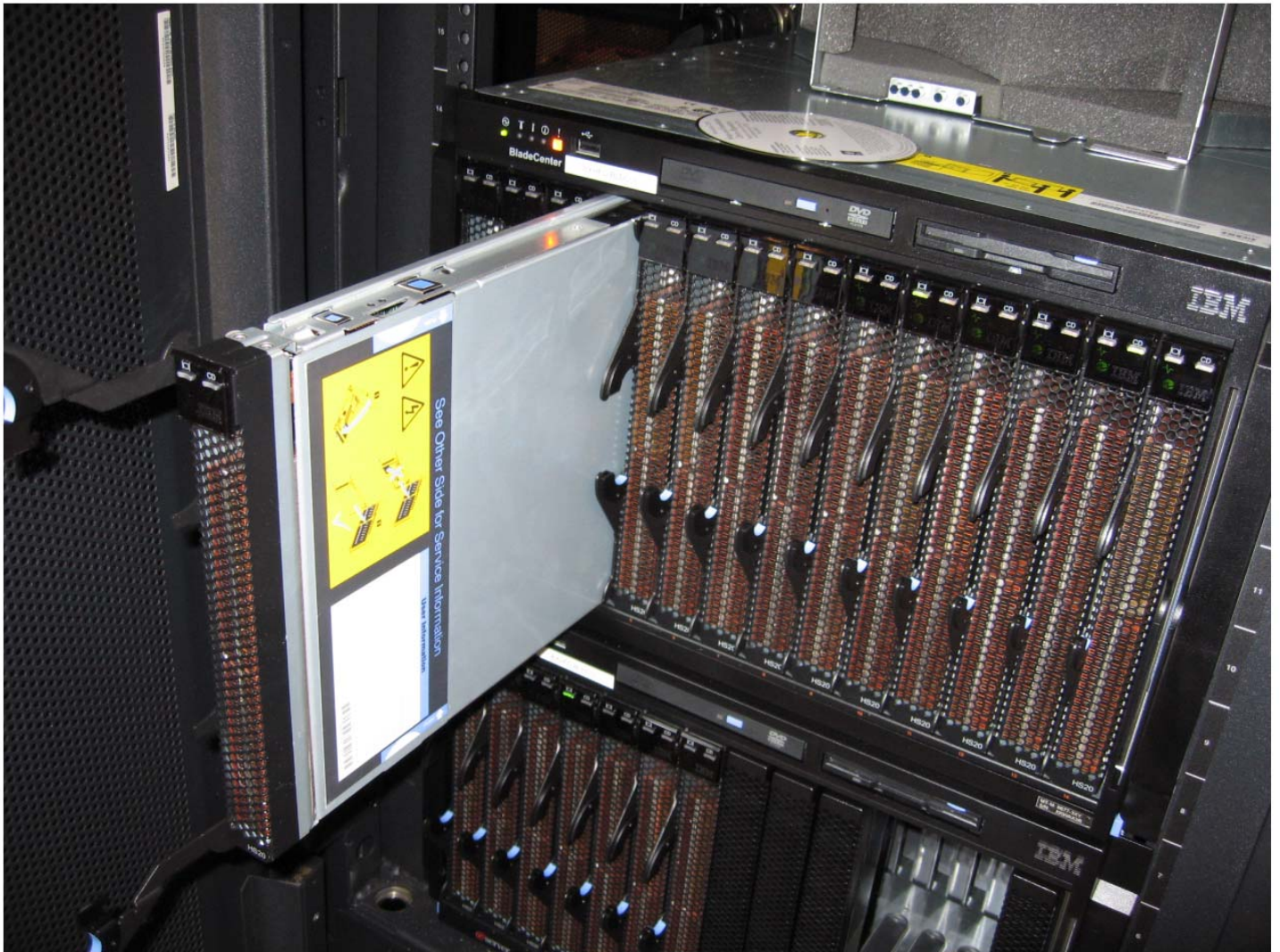


Abb. 14.3.7
IBM Blade Center mit 14 Blades

Blades werden in einem Blade Gehäuse (Blade Enclosure, IBM Bezeichnung „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, siehe Band 1, Abschnitt 5.2.3) und Management Funktionen zur Verfügung.

Abbildung 14.3.7 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, und/oder zu einem Storage Area Netzwerk (SAN) ermöglicht. Mehrere Blade Center können in einem 19 Zoll Rack untergebracht werden.

Blades können einen sehr unterschiedlichen Funktionsumfang haben.

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 Chips, zwei Quad-Core-Xeon Chips oder zwei Opteron Chips 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.

Blade Center der Firma Sun



Abb. 14.3.8
Sun Blade Center

In großen Unix Konfigurationen wird eine ähnliche Technologie eingesetzt. Die System Boards eines Sun 25K (Abschnitt 11.1.4) oder M9000 Rechners sind eine evolutionäre Erweiterung der Sun Blade 6000 Blades. Ähnliches gilt für die „Cell Boards“ des HP Superdomes. Sie unterscheiden sich durch zusätzliche Funktionen, z.B.:

- Direkte Verbindung über einen zentralen Switch (ohne Ethernet Protokoll)
- NUMA Fähigkeit
- Partitionsmöglichkeiten (Harte oder virtuelle Partitionierung)
- Zusätzliche unterstützende Hardware für die HP-UX oder Solaris Betriebssysteme
- Einrichtungen für eine zentrale Administration
- Verbesserte I/O Einrichtungen, höhere I/O Kapazität
- Zusätzliche Verbesserungen für Ausfallsicherheit, Zuverlässigkeit und Verfügbarkeit
-

14.3.3 System z Blade Center Extension

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 (IEDN), mit einem zEnterprise-Rechner - entweder zEC12, z196 oder z114 – verbunden sind. Eine zBX kann speziell angepasste PowerPC oder x86 Blades aufnehmen, auf denen die AIX, Linux oder Windows Betriebssysteme und deren Anwendungen laufen können. Weiterhin sind Blades mit leistungsfähigen Spezialprozessoren für bestimmte Workloads (z.B. XML oder Krypto Verarbeitung) verfügbar. Sie werden als „Accelerator“ bezeichnet.

Die zBX ist eine neue Infrastruktur für die Erweiterung von System z Servicequalität- und Verwaltungsfunktionen auf integrierte POWER7- und x86 Komponenten. Die Kombination eines oder mehrerer zEC12, 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 Rechner.

Eine zBX besteht aus einem Gehäuse, in dem 1 oder 2 Blade Center , Stromversorgung, Verkabelung in der Form von mehreren Ethernet Switches und weiteren Komponenten untergebracht sind. Die zBX wird in der Regel unmittelbar neben dem zEC12, z196 oder z114 Rechner aufgestellt. Die Datenverbindungen zwischen den beiden Einheiten sind von außen nicht zugreifbar.

Eine zBX kann zunächst Blades mit Standard Microprozessoren aufnehmen, spezifisch x86 und PowerPC. Auf ihnen sind alle unterstützten Betriebssysteme lauffähig, besonders Windows und Linux, aber in Zukunft auch andere Betriebssysteme wie z.B. Solaris, Mac OS, 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 Appliances wie

- WebSphere DataPower Integration Appliance XI50
- WebSphere DataPower XML Accelerator Appliance XA35
- WebSphere DataPower Message-Routing Appliance XM70
- WebSphere DataPower XML Security Gateway Appliance XS40
- WebSphere DataPower (B2B Business to Business) Gateway Appliance XB60

<http://www.cedix.de/VorlesMirror/Band2/SpecBlade.pdf>

Weitere Special Blades werden in Zukunft erwartet. Ein Beispiel ist, besonders Performance-kritische Algorithmen nicht durch einen Mikroprozessor, sondern durch einen ASIC oder ein FPGA verarbeiten zu lassen. Gameframe ist ein weiteres Beispiel (<http://www.ti.uni-tuebingen.de/index.php?id=381&L=1>).



Abb. 14.3.9
zBX Konfiguration

Eine zBX wird über eine interne, von außen nicht zugängliche Verkabelung direkt an ein Modell zEC12, z196 oder Modell 114 Mainframe angeschlossen.

Es ist eine zentrale Switch Steuerung vorhanden, in mancher Beziehung vergleichbar mit den zentralen Switches der Sun M9000 oder HP Superdome Rechner.

IBM bezeichnet Konfigurationen mit einer zBX als „zEnterprise“.

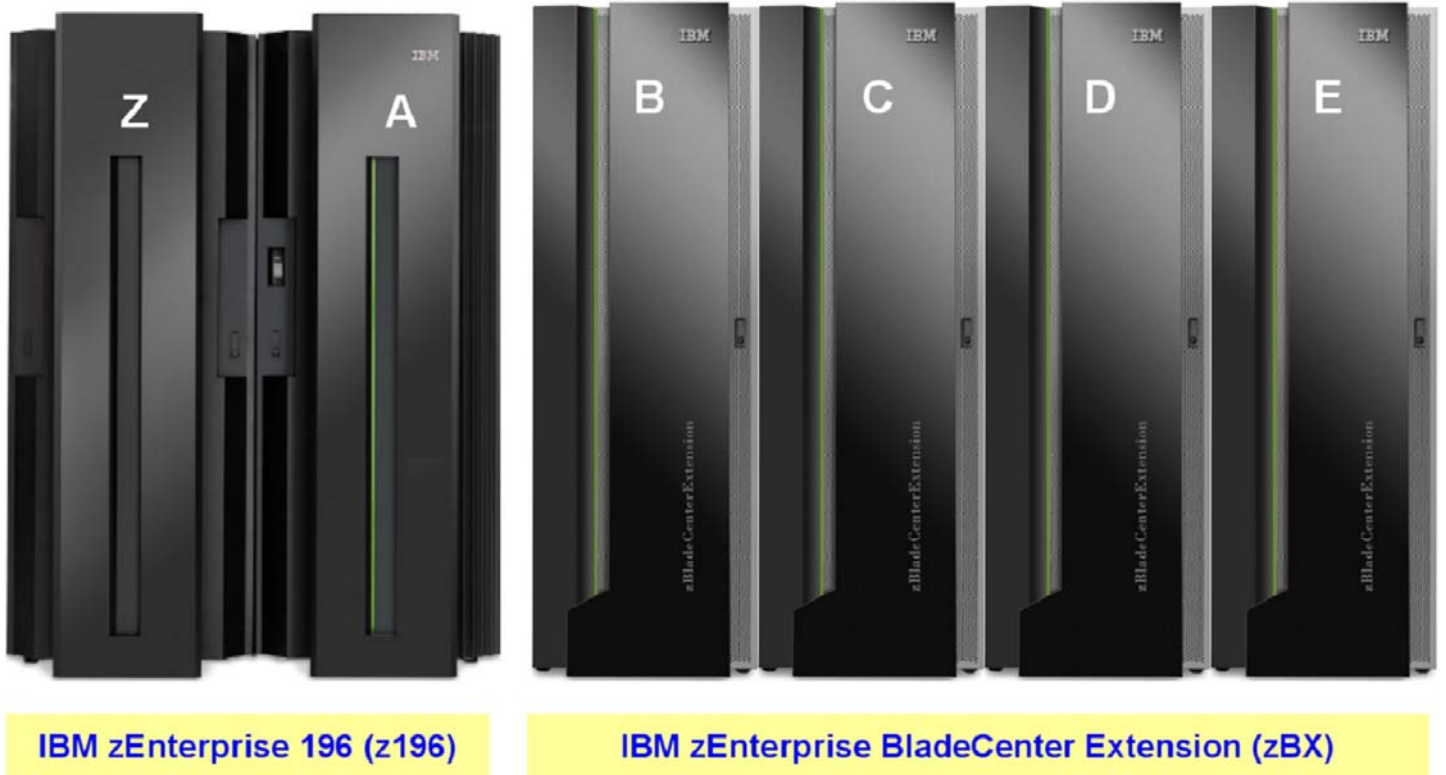


Abb. 14.3.10
Konfiguration mit der maximal zulässigen Anzahl von 4 zBX Einheiten

Es ist möglich, an einen zEnterprise 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.

14.3.4 Private Netzwerke

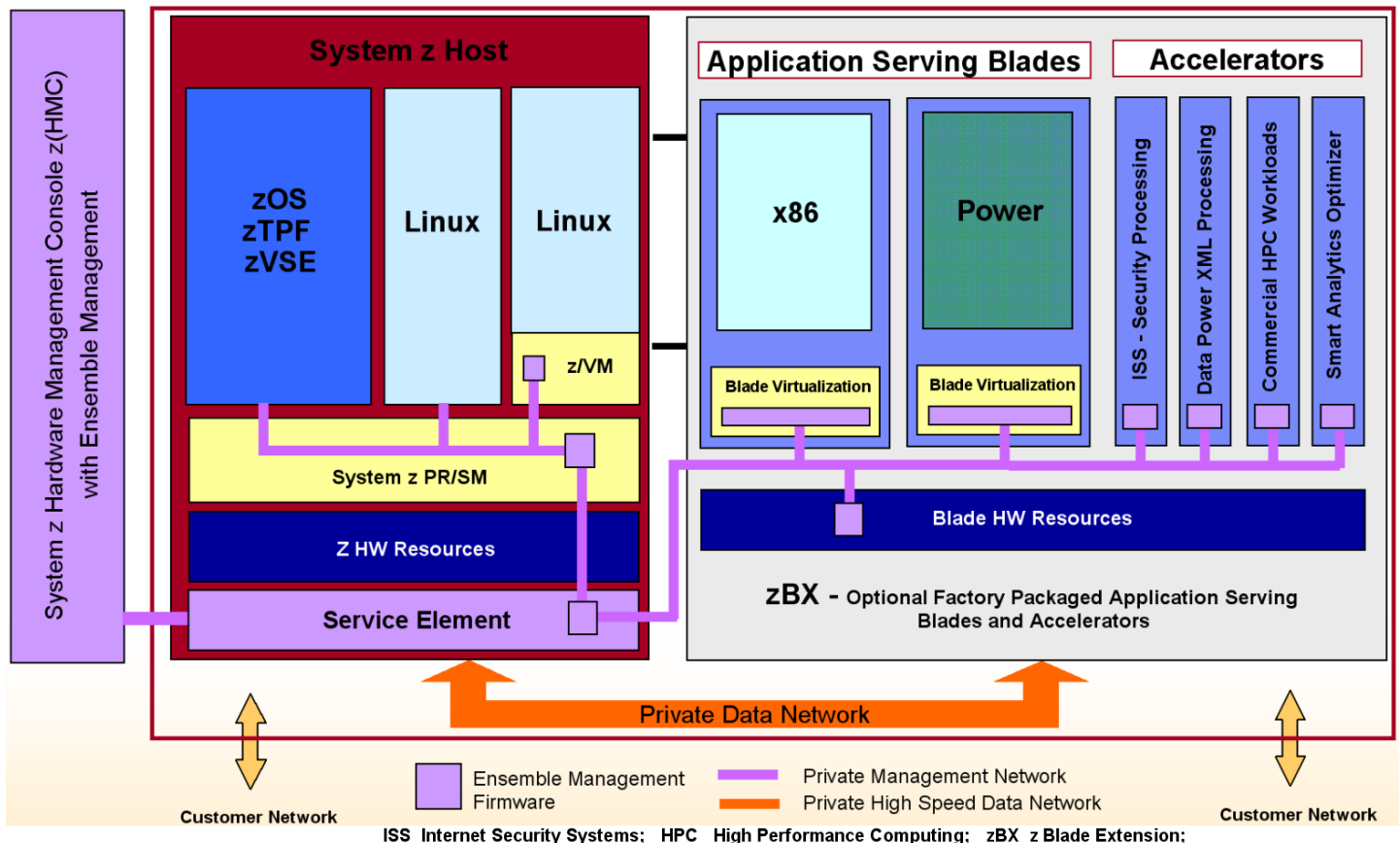


Abb. 14.3.11
Netzwerk Anbindung einer zBX

Die zBX ist sehr eng an den zEnterprise Rechner gekoppelt. Hierzu dienen zwei „Private Networks“. Private bedeutet, dass die Netzwerke nur zwischen der zBX und dem zEnterprise 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) **Private Management Network**.

Das Private Data Network wird als „Inter Ensemble Data Network“ (**IEDN**) bezeichnet. Es 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 Switches und der dazugehörige Support.

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

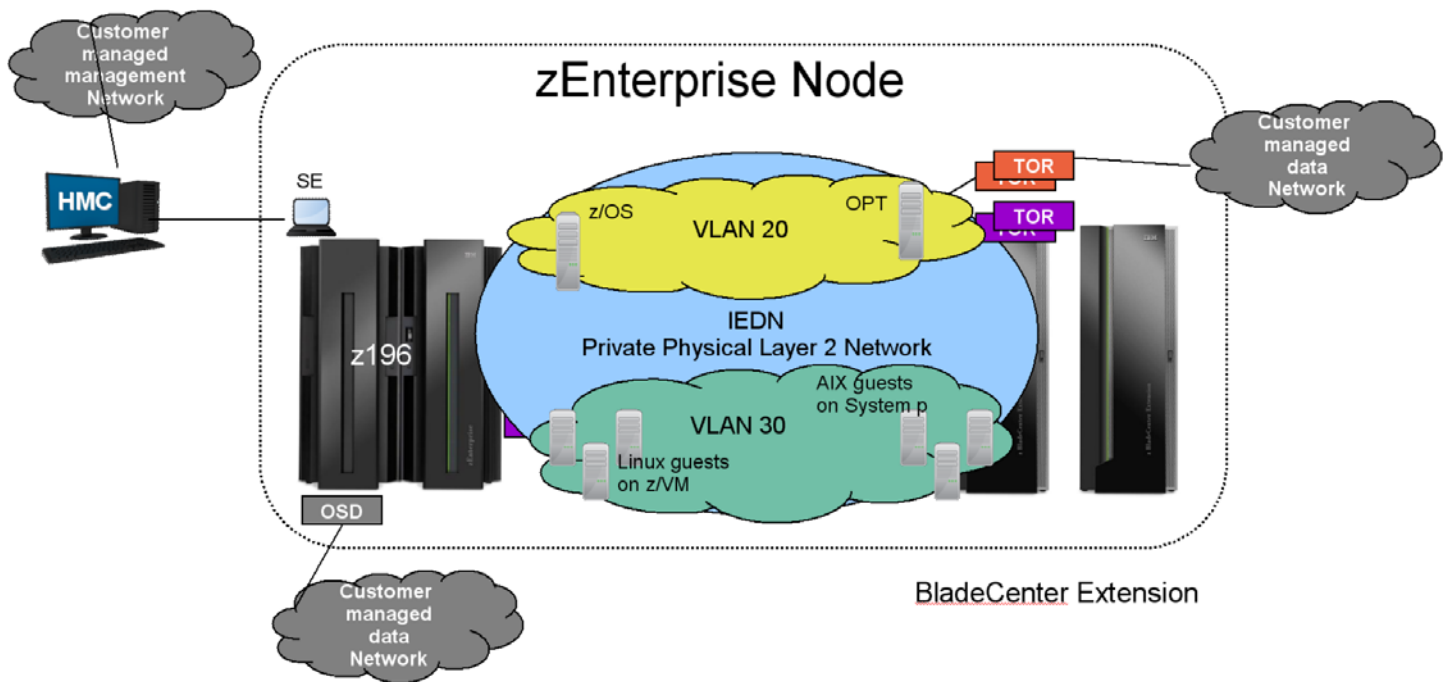


Abb. 14.3.12
zBX VLAN Anbindung

Ein Mainframe-Rechner mit angeschlossener zBX wird als zEnterprise Node bezeichnet. Innerhalb einer zEnterprise Node können mehrere virtuelle Netzwerke (VLAN) konfiguriert werden, welche die Verbindung zwischen den Blades und dem z/OS Rechner herstellen.

Sockets sind ein weit verbreitetes Protokoll in der Schicht 5 des OSI Modells für die Kommunikation zwischen zwei Rechnern. Sockets benutzen TCP/UDP in der Schicht 4 und IP in der Schicht 3.

Die in Abschnitt 12.4.7 erwähnten Hipersockets sind ein mit Sockets kompatibles z/OS Protokoll, welches an Stelle von OSI Schicht 4 und Schicht 3 eine Queue im Hauptspeicher für die Kommunikation zwischen 2 LPARs benutzt. Der Performance Gewinn auf Grund der Vermeidung des Schicht 4 und Schicht 3 Overhead ist signifikant.

HiperSockets Integration mit IEDN bewirkt, dass ein Mainframe über das physische IEDN mit einer zBX unter Vermeidung des Schicht 4 und 3 Overhead kommunizieren kann. Die Verbindung erscheint als ein Single Layer 2 Netzwerk. Über das interne zBX Private Data Network ist eine Hipersocket Kommunikation zwischen Blades und z/OS LPARs möglich. Dies verlängert die Reichweite des Hipersockets Netzwerks außerhalb des Mainframes hin auf die zBX.

OSX ist eine spezielle Version des OSA Ethernet Adapters für eine IEDN Kommunikation zwischen zwei zEnterprise Rechnern.

Die HiperSockets Virtual Switch Bridge kann mit dem Intraensemble Datennetz (IEDN) und OSX Adapter eine Hipersocket Verbindung zu einem anderen Mainframe Rechner durch dessen IEDN OSX-Adapter erstellen, wiederum unter Eliminierung des Schicht 4 und Schicht 3 Overhead.

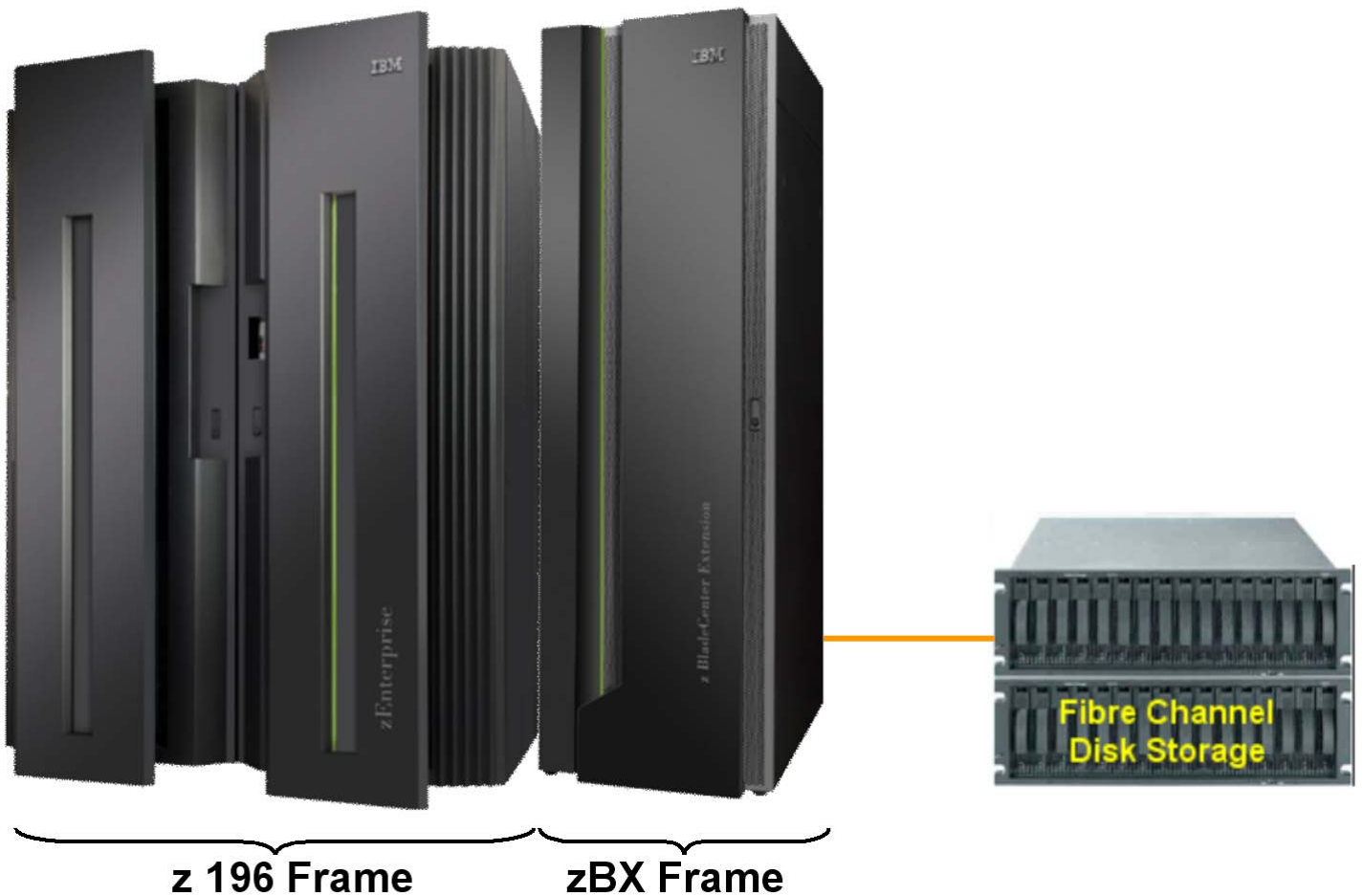


Abb. 14.3.13
Festplattenanschluss an die zBX

Die Blades benötigen eigene Plattenspeicher und entsprechende Anschlüsse. Hierfür ist innerhalb der zBX ein weiteres Fibre Channel Netzwerk und –Switches vorgesehen, über das jede Blade mit externen Fibre Channel SCSI Platten verbunden werden kann (Siehe Band 1 Abschnitt 5.2.3).

14.3.5 Unified Resource Manager

Ein Mainframe Rechner mit angeschlossener zBX wird als zEnterprise Node bezeichnet. Mehrere zEnterprise Nodes bilden ein Enterprise Ensemble. Innerhalb einer zEnterprise Node können mehrere virtuelle Netzwerke konfiguriert werden, welche die Verbindung zwischen den Blades und dem z/OS Rechner herstellen.

Die Hardware Management Konsole (HMC, siehe Abschnitt 14.2.3) spielt eine Schlüsselrolle bei der Administration eines Mainframe Rechners. 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 vom 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 zusätzlich 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**
- **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.

Six Functional Management Areas Within Unified Resource Manager

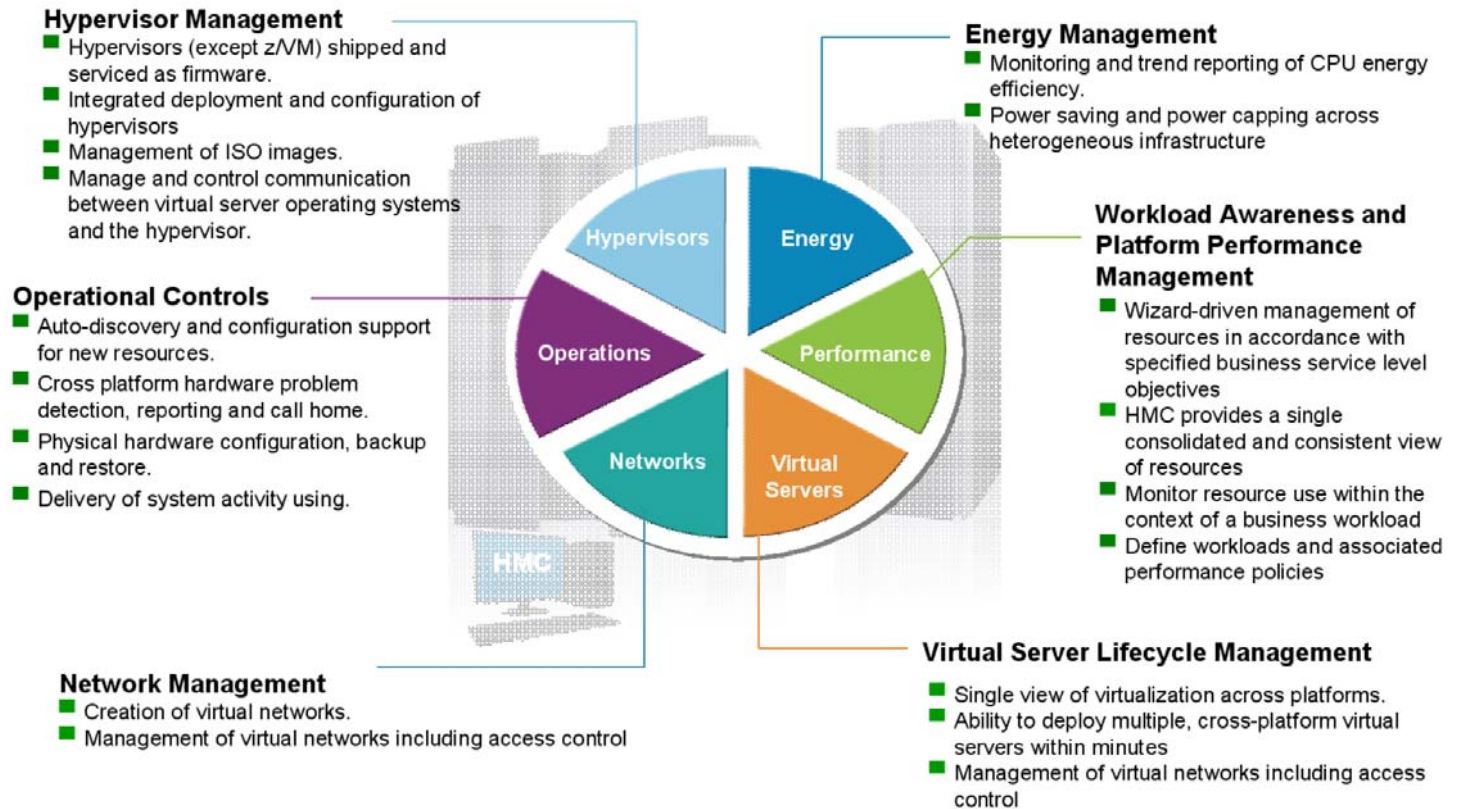


Abb. 14.3.14
Funktionen des Unified Resource Manager

Die zEnterprise Unified Resource Manager läuft auf der Hardware Maintenance Console (HMC). Er kooperiert mit verschiedenen Hypervisor Technologien. Diese sind:

- PR/SM unterstützt die Definition von logischen Partitionen oder LPARs in den zEnterprise Rechnern, so dass viele logische Partitionen die gleichen physischen Ressourcen gemeinsam nutzen können.
- z/VM bietet die Möglichkeit, Betriebssysteme wie Linux, z/OS und andere als Gäste von z/VM auf System z laufen zu lassen, so dass viele virtuelle Maschinen die gleichen physischen Ressourcen gemeinsam nutzen können.
- PowerVM Enterprise Edition bietet Virtualisierungsfunktionen für AIX. Es ermöglicht die Erstellung von logischen Partitionen auf den Blades und ermöglicht die gemeinsame Nutzung von Ressourcen zwischen mehreren Betriebssysteme.
- Ein in die x86 (System X) Blades integrierter Hypervisor „Kernel based Virtual Machines“ (KVM) bietet eine Virtualisierungslösung für Linux und Windows auf x86-Hardware.

Die Links in die verschiedenen Hypervisor und indirekte Verbindungen zu z / OS, falls erforderlich, erzeugen eine sehr heterogenen Umgebung. 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 ist es die Aufgabe des Unified Resource Managers, diese unterschiedlichen Virtualisierungslösungen zu konsolidieren. Hierfür existiert eine Unterstützung für „Federated Hypervisors“: PR/SM, z/VM, PowerVM und KVM , mit einem einheitlichem Satz von System Management Policies. Dies ist z.B. ein wichtiger Schritt für den Aufbau einer Cloud-Infrastruktur.

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 neue Architektur Plattformen.
- Ausdehnung des z/OS WLM auf zBX Blades.
- zBX Coupling Facility.
- zBX GDPS.
- Shared Memory zwischen Mainframe und zBX, z.B. für Hipersockets.

14.3.6 Ensemble



Abb. 14.3.15
Ensemble mit drei Nodes

Für Administrationszwecke ist es möglich, mehrere z196 oder zEC12 Server zu einem als „Ensemble“ bezeichneten Verbund zusammenzuschließen. Ein einziger zEnterprise Server mit angeschlossener zBX (1 – 4 Frames) wird hierbei als „Node“ bezeichnet. Ein Ensemble kann aus bis zu 8 Nodes bestehen.

In Abb. 14.3.15 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 zEnterprise Server einschließlich ihrer lokal angeschlossenen zBXs zu administrieren. Aus Gründen der Zuverlässigkeit ist wie immer eine zweite HMC als Backup vorhanden (Primary und Alternate HMC).

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

14.3.7 zBX Vorteile

Zuverlässigkeit, Verfügbarkeit und Wartungsfreundlichkeit (RAS) reduziert Ausfallzeiten. zBX ist ein Blade-Center, das für Mainframe-Konnektivität und RAS Leistung ausgelegt ist.

Alle Server laufen virtualisiert unter Kontrolle des Unified Resource Manager.

Die Konsolidierung verteilter Server auf entweder zBX oder zLinux virtuelle Server eliminiert mehr als 90 Prozent der Netzwerk-Infrastruktur-Komponenten (Kabel, Switches und Router). Die Verkabelung zwischen dem zBX und dem Mainframe kann auf wenige Netzwerk Komponenten reduziert werden. zBX konsolidiert System p, System x und Spezial Geräte in einem einzigen integrierten Chassis mit gemeinsamen Zugang zur Stromversorgung und Netzwerken. Dies vereinfacht das Management von Stromversorgung, Kühlung, Netzwerken und Stellfläche, bei deutlicher Senkung der Betriebskosten.

Verbesserte Sicherheit eliminiert Firewall Kosten. Die Konsolidierung verteilter Server in eine zBX verlagert Server-zu-Server-Datenverkehr in VLANs auf dem privaten IEDN und weg von Attacken ausgesetzten externen Netzwerken. Dies kann die Notwendigkeit für Firewalls zwischen zBX Servern sowie zwischen zBX Servern und dem Mainframe eliminieren.

Automation ersetzt viele manuelle Prozesse in fünf kritischen Management-Bereichen: Assets, Deployment, Capacity/Performance, Security sowie Change Management für z/OS, AIX, Linux und Windows Server.

In Zukunft werden wir weitere Verbesserungen in Richtung einer einheitlichen Management Struktur sehen, um all dies zu steuern.

14.4 Netezza

14.4.1 Data Warehouse

Seit vielen Jahren beobachten wir ein ständiges Wachstum der in unseren Computern gespeicherten Daten. Für die Führungskräfte eines Unternehmens ist die Nutzung dieser Daten für ihre Entscheidungsprozesse sehr wichtig.

Business Intelligence Systeme stellen die Informationen für die Entscheidungsprozesse bereit.

Aus technischen Gründen erweist es sich als sinnvoll, ein Business Intelligence System von den datenliefernden operativen Systemen (z.B. einer CICS – DB2 Kombination) zu entkoppeln und auf einer separaten Plattform, als „Data Warehouse“ bezeichnet, zu betreiben. Eine Entkopplung führt einerseits zu einer Entlastung der operativen Systeme und eröffnet andererseits die Option, das analyseorientierte Business Intelligence System auf die Belange von Auswertungen und Berichten hin zu optimieren.



Abb. 14.4.1
Extraktion, Transformation, Laden Prozess

Ein spezieller „Extraktion, Transformation, Laden“ (ETL) Prozess kopiert ständig die operativen Daten in die Data Warehouse Datenbank.

Im Idealfall soll eine derartige Datenbasis unternehmensweit ausgerichtet sein und das Informationsbedürfnis verschiedenster Anwendergruppen abdecken.

14.4.2 Data Mining

Unter Data-Mining – deutsch etwa: „aus einem Datenberg etwas Wertvolles extrahieren“ – versteht man die systematische Anwendung statistischer Methoden auf einen Datenbestand mit dem Ziel, bisher unbekannte Zusammenhänge zu erkennen. Hierbei geht es auch um die Verarbeitung sehr großer Datenbestände (die nicht mehr manuell verarbeitet werden könnten), wofür effiziente Methoden benötigt werden, deren Zeitkomplexität sie für solche Datenmengen geeignet macht.

Data-Mining benutzt Datenanalyse- und Entdeckungsalgorithmen, die unter akzeptablen Effizienzbegrenzungen eine spezielle Auflistung von Mustern (oder Modellen) der Daten liefern. Das Ziel ist die Extraktion von Wissen, das

- gültig (im statistischen Sinne),
- bisher unbekannt und
- potentiell nützlich ist,

„zur Bestimmung bestimmter Regelmäßigkeiten, Gesetzmäßigkeiten und verborgener Zusammenhänge“.

Dies sind einige typische Data-Mining Anwendungen:

Im Finanzsektor: Rechnungsprüfung zur Betrugserkennung

Im Marketing:

- Marktsegmentierung, beispielsweise Kunden in Bezug auf ähnliches Kaufverhalten bzw. Interessen für gezielte Werbemaßnahmen identifizieren
- Warenkorbanalyse zur Preisoptimierung und Produktplatzierung im Supermarkt
- Zielgruppen-Auswahl für Werbekampagnen
- Kundenprofil-Erstellung zum Management von Kundenbeziehungen in Customer-Relationship-Management Systemen

Im Internet:

- Angriffserkennung (Security Attack)
- Empfehlungsdienste für Produkte wie beispielsweise Filme und Musik
- Netzwerkanalyse in sozialen Netzwerken
- Web-Usage-Mining um das Nutzerverhalten zu analysieren
- Textmining zur Analyse von großen Textbeständen

Im Gesundheitswesen:

- Pharmakovigilanz (Arzneimittelüberwachung nach Marktzulassung im Hinblick auf unbekannte unerwünschte Ereignisse oder Nebenwirkungen)
- Das BKK InfoNet des Bundesverbandes der Betriebskrankenkassen (BKK) bietet Kennzahlen und Analysen für Vertragspolitik und Versorgungsmanagement. Die Leistungs-, DMP- und Statistikdaten stehen Krankenkassen und Landesverbänden im Rahmen ihrer Zugriffsrechte zur freien Auswertung zur Verfügung.

Dies sind zwei Beispiele:

Beispiel Wal-Mart

Wal-Mart (www.wal-mart.com) ist der Marktführer im amerikanischen Einzelhandel, unterhält ein unternehmensweites Data Warehouse, Größe: ca. 300 TB (2003), mit täglich bis zu 20.000 Data Warehouse Anfragen. Jede Eingabe in einer Registrierkasse wird als Transaktion gespeichert. Es erfolgt eine tägliche Auswertung von

- Artikelumsätzen, Lagerbestand,:
- Warenkorbanalyse, Kundenverhalten, ...
- Überprüfung des Warensortiments zur Erkennung von Ladenhütern oder Verkaufsschlagern
- Standortanalyse zur Einschätzung der Rentabilität von Niederlassungen
- Untersuchung der Wirksamkeit von Marketing-Aktionen
- Auswertung von Kundenbefragungen, Reklamationen bzgl. bestimmter Produkte etc.
- Analyse des Lagerbestandes
- Warenkorbanalyse mit Hilfe der Kassensbons

Beispiel einer Anfrage

Welche Umsatzsteigerungen sind vom 1. – 10. Oktober in den Abteilungen Kosmetik, Elektro und Haushaltswaren in Birmingham, Alabama, angefallen, nachdem wir in Huntsville, Alabama, DVDs um 20 % billiger angeboten haben. ?

Beispiel Otto Versand

Der Versandhändler Otto verbessert mithilfe einer Spezialsoftware seine Bedarfsplanung für das gesamte Sortiment. Das Unternehmen füttert seine Software pro Woche mit 300 Millionen Datensätzen – und erstellt übers Jahr eine Milliarde Prognosen, wie sich der Absatz einzelner Artikel in den folgenden Tagen und Wochen entwickeln wird. Nach Konzernangaben ordert Otto durchschnittlich 30 Prozent weniger Ware als zuvor und hat dadurch deutlich weniger überschüssige Ware auf Lager. Von Einsparungen in zweistelliger Millionenhöhe ist die Rede.

Die wichtigsten Data Mining Prozesse sind ETL, Business Intelligence sowie Advanced Analytics.

ETL

Die Abkürzung ETL steht für „Extraktion, Transformation, Laden“. Dies ist ein Prozess, der Daten aus mehreren Datenquellen in einer Data Warehouse Zieldatenbank vereinigt.

Business Intelligence (BI)

Der englische Ausdruck „Intelligence“ (von lat. intellegere „verstehen“ und legere „lesen, wählen“) bezeichnet die aus dem Sammeln und dem Aufbereiten erworbener Informationen gewonnenen Erkenntnisse.

Der Begriff Business Intelligence (BI) bezeichnet Verfahren und Prozesse zur systematischen Analyse (Sammlung, Auswertung und Darstellung) von Daten in elektronischer Form. Ziel ist die Gewinnung von Erkenntnissen, die in Hinsicht auf die Unternehmensziele bessere operative oder strategische Entscheidungen ermöglichen. Dies geschieht mit Hilfe analytischer Konzepte sowie entsprechender Software bzw. IT-Systeme, die Daten über das eigene Unternehmen, die Mitbewerber oder die Marktentwicklung im Hinblick auf den gewünschten Erkenntnisgewinn auswerten. Mit den gewonnenen Erkenntnissen können Unternehmen ihre Geschäftsabläufe, sowie Kunden- und Lieferantenbeziehungen profitabler machen, Kosten senken, Risiken minimieren und die Wertschöpfung vergrößern.

Advanced Analytics

Der Begriff "Advanced Analytics" ist eine Erweiterung der Business Intelligence. Das Ziel ist jedoch weniger, durch die Analysen der Ist-Situation bzw. Vergangenheit aktuelle Probleme aufzuzeigen, sondern stärker auf Prognosen und Vorhersagen der zukünftigen Entwicklung zu setzen. Unterbereiche sind Predictive Analytics, Simulation und Optimierung

Predictive Analytics

- Was wird passieren, wenn unsere Kunden genauso einkaufen wie in der Vergangenheit?
- Was passiert mit unseren Vertriebsergebnissen, wenn sich die derzeitigen Trends fortsetzen?

Simulation

- Wenn wir ein neues Produkt einführen, wie werden unsere Konkurrenten wahrscheinlich reagieren?
- Wenn wir unsere Preispolitik ändern, was ist der Einfluss auf unsere Kundenbindung und Marktdurchdringung?

Optimierung

- Wie erreichen wir die besten Beladungspläne für unsere LKWs?
- In welche Drogen Forschungsprojekte sollen wir investieren, um unsere Gewinne in Anbetracht unserer auslaufenden Drogen Patente zu maximieren?

14.4.3 Data Sharing Alternativen

Data Mining kann mittels SQL Anfragen an die gleichen relationale Datenbanken erfolgen, die auch für die Transaktionsverarbeitung eingesetzt werden. Bei großen Datenbeständen können aber selbst einfache Anfragen Stunden an Verarbeitungszeit in Anspruch nehmen. In der Praxis dupliziert man die Datenbestände in einer getrennten Data Warehouse Datenbank. Das Data Warehouse verfügt über eigene CPUs, welche Data Warehouse Queries bearbeiten können. Eine Data Warehouse Datenbank ist typischerweise eine relationale Datenbank, und Abfragen erfolgen als SQL Queries. Data Mining Erweiterungen, englisch Data Mining Extensions (DMX), sind proprietäre Erweiterungen des SQL-Standards um die Fähigkeit zu verbessern, mit Data-Mining Modellen zu arbeiten.

Ein Transaktionsserver verarbeitet in der Regel mehrere Transaktionen parallel auf mehreren CPUs. Ebenso sind zahlreiche Plattenspeicher vorhanden. Diese sind nach dem „Shared Disk“ Verfahren an die CPUs angeschlossen, weil eine Transaktion in der Regel auf den ganzen Datenbestand zugreifen möchte. Bei einem Data Warehouse ist es möglich, die Datenbank auf einzelnen Festplatten zu partitionieren, jeder Festplatte eine getrennte CPU zuzuordnen, und eine Abfrage auf zahlreichen CPUs parallel gegen die partitionierte Datenbank durchzuführen. Diese Architektur wird als „Shared Nothing“ bezeichnet.

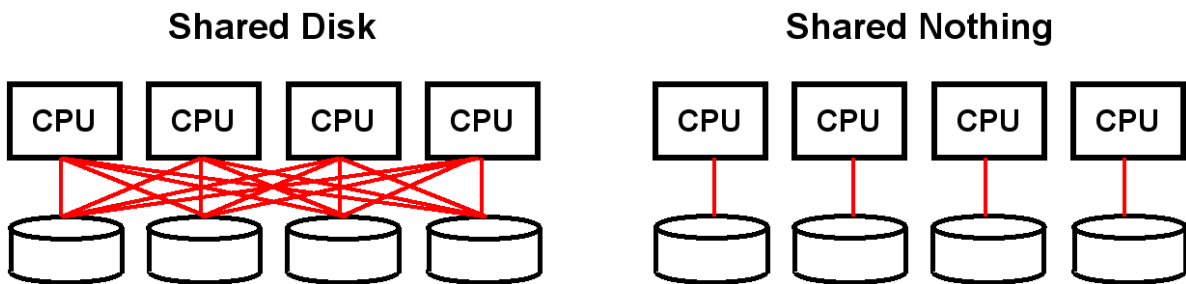


Abb. 14.4.2

Shared Nothing erlaubt nur Zugriffe auf zugeordnete Festplatten

Eine Shared-Nothing System gilt als besser als ein Shared Disk System für Anwendungen, die eine Reihe von Datenbanken parallel durchsuchen sollen. Dazu gehören Data Warehouse Anwendungen.

Die Shared-Nothing Architektur beschreibt eine Distributed Computing Architektur, bei der jeder Knoten unabhängig und eigenständig seine Aufgaben mit seinem eigenen Prozessor und den zugeordneten Speicherkomponenten wie Festplatte und Hauptspeicher erfüllen kann und kein bestimmter, einzelner Knoten für die Verbindung zu einer Datenbank notwendig ist. Die Knoten sind über ein LAN- oder WAN-Netzwerk miteinander verbunden. Jeder Knoten verfügt darüber hinaus über eine Kopie des Datenbank-Management-Systems. Ein Knoten kann Aufgaben an einen anderen nicht ausgelasteten Knoten weitergeben.

Shared Nothing ist auf Grund seiner Skalierbarkeit beliebt für Webanwendungen oder parallele Datenbanksysteme. Wie bei Google gezeigt werden konnte (Abschnitt 17.3.2), ist ein Shared Nothing System nahezu unbegrenzt durch Ergänzung zusätzlicher Knoten in Form preiswerter Computer ausbaufähig, weil kein einzelnes Netzwerkelement existiert, dessen begrenzte Leistung die Geschwindigkeit des gesamten Systems vermindert.

Die Firma Teradata ist ein bedeutender Data Warehouse Hersteller, der seit 1979 ein Shared nothing Relationales Datenbankmanagementsystem vertreibt. Die wichtigsten Mitbewerber im Data Warehouse Markt sind

- **IBM mit dem „PureData System for Analytics“ (Netezza) Produkt,**
- **Oracle mit dem „EXAdata“ Produkt und**
- **SAP mit dem „HANA“ In-Memory Datenbank Produkt.**

Apache Hadoop ist ein freies, in Java geschriebenes Framework für skalierbare, verteilt arbeitende Software. Es ermöglicht es, intensive Rechenprozesse mit großen Datenmengen (Petabyte Bereich) auf Computerclustern durchzuführen. Das IBM-Produkt InfoSphere BigInsights basiert auf Hadoop.

Problem: Während der Shared Nothing Ansatz eine deutliche Leistungssteigerung bringt, wachsen die Anforderungen so schnell, dass die Verarbeitungsdauer von Data Warehouse Anfragen ein echtes Problem darstellt.

14.4.4 PureData System for Analytics (Netezza)

Das IBM **PureData System for Analytics** (alte Bezeichnung: Netezza, englische Aussprache: „netiesä“) ist eine Data Warehouse Appliance, d.h. ein kombiniertes System aus Hardware und speziell darauf abgestimmter Software, welches sich von alternativen Produkten wie z. B. Teradata oder Oracle Exadata durch zwei Merkmale unterscheidet:

- Ein PureData System for Analytics ist zwar unabhängig von zBX. Der „IBM DB2 Analytics Accelerator for z/OS“ (IDAA) ermöglicht aber eine ähnlich enge Integration in z/OS (spezifisch DB2) wie bei den zBX Appliances.
- Mittels massiver Parallelisierung von Berechnungen und dem Einsatz von FPGAs (Field Programmable Gate Arrays) ist eine deutlich höhere Verarbeitungsgeschwindigkeit möglich.

Ein Field Programmable Gate Array (FPGA) ist ein Integrierter Schaltkreis (IC) der Digitaltechnik, in den eine logische Schaltung programmiert werden kann. Anders als bei der Programmierung von Computern oder Steuerungen bezieht sich hier der Begriff Programm nur in zweiter Linie auf die Vorgabe zeitlicher Abläufe im Baustein, sondern vor allem auf die Definition von dessen Funktionsstruktur. Durch die Programmierung von Strukturvorschriften wird zunächst die grundlegende Funktionsweise einzelner universeller Blöcke im FPGA und deren Verschaltung untereinander festgelegt. Durch die spezifische Konfiguration interner Strukturen können in einem FPGA verschiedene Schaltungen realisiert werden. Diese reichen von Schaltungen geringer Komplexität, wie z. B. einfacher Synchronzähler, bis zu hochkomplexen Schaltungen, wie Mikroprozessoren. FPGAs werden in allen Bereichen der Digitaltechnik eingesetzt, vor allem aber dort, wo es auf schnelle Signalverarbeitung und flexible Änderung der Schaltung ankommt, um beispielsweise nachträgliche Verbesserungen an den implementierten Funktionen vornehmen zu können, ohne dabei direkt die physische Hardware ändern zu müssen.

Mit der Einführung der FPGAs wurden kompakte, anwenderspezifische Schaltungen in geringen Stückzahlen ermöglicht. Heutzutage gestatten sie z. B. die preiswerte und flexible Fertigung komplexer Systeme wie Mobilfunk-Basisstationen als Alternative zur teureren Auftragsfertigung durch Halbleiterhersteller.

Die modernste (2013) PureData System for Analytics (PDSA) Implementierung wird als Modell N2001 bezeichnet.

http://de.wikipedia.org/wiki/Field_Programmable_Gate_Array enthält eine FPGA-Übersicht.

14.4.5 S-Blades und FPGA Konfiguration

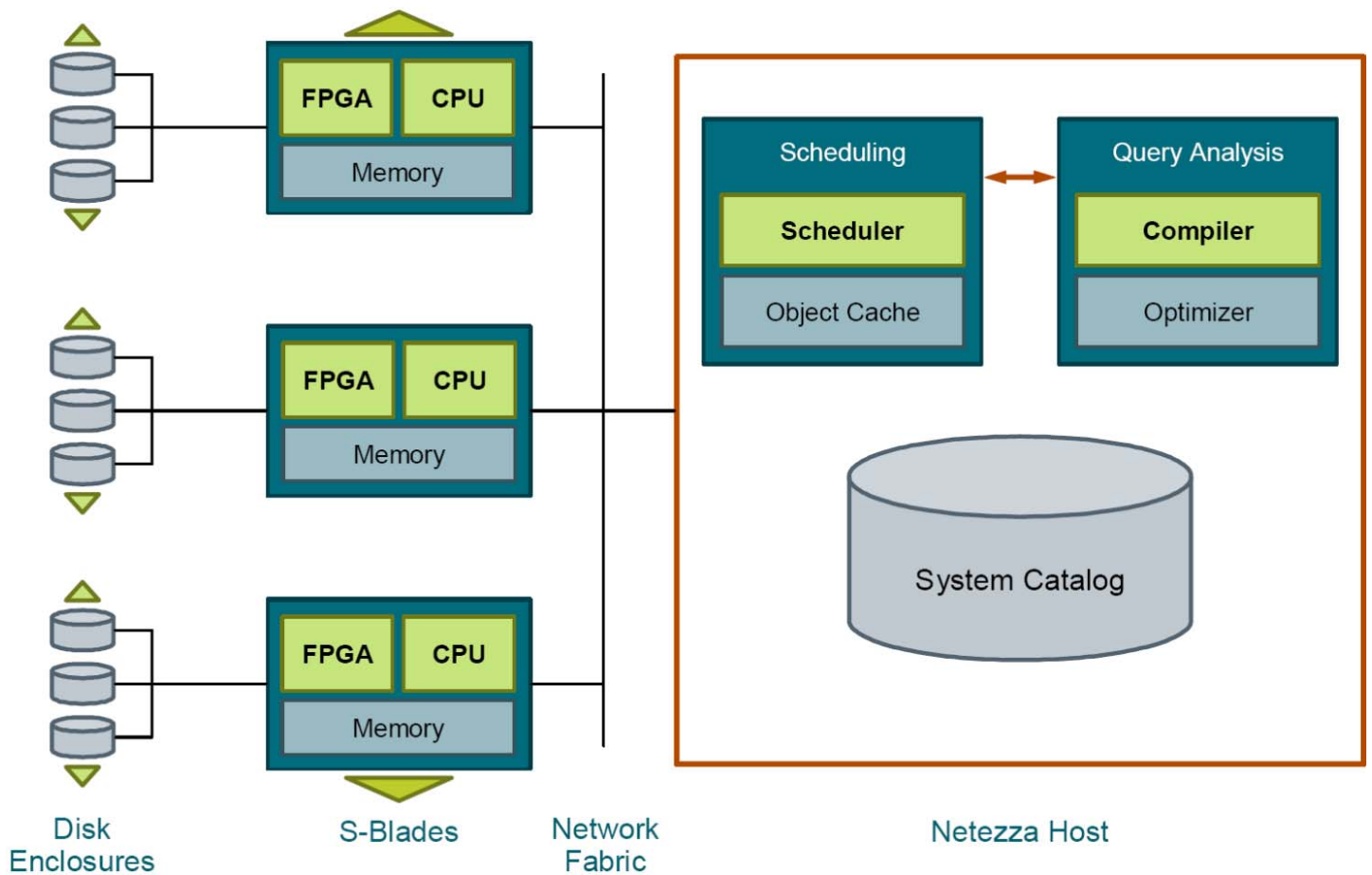


Abb. 14.4.3
S-Blades und der Netezza Host

Die Basiskonfiguration eines PureData System for Analytics (PDSA) besteht aus 288 SAS2 600-GB Festplatten (bei vierfacher Komprimierung stehen dem Benutzer 173 TByte zur Verfügung), 4 sogenannten S-Blades mit zusammen 64 CPU-Cores, 64 FPGA-Kernen, einem internen Netzwerk sowie einem internen Steuerrechner, der als Host bezeichnet wird.

Der Host ist ein selbständiger Hochleistungs-Linux-Server, der aus Verfügbarkeitsgründen doppelt vorhanden ist. Er kompiliert SQL-Abfragen von Advanced Analytics, Business Intelligence oder ETL-Anwendungen in ausführbare Code-Segmente und verteilt diese zur Ausführung auf die S-Blades.

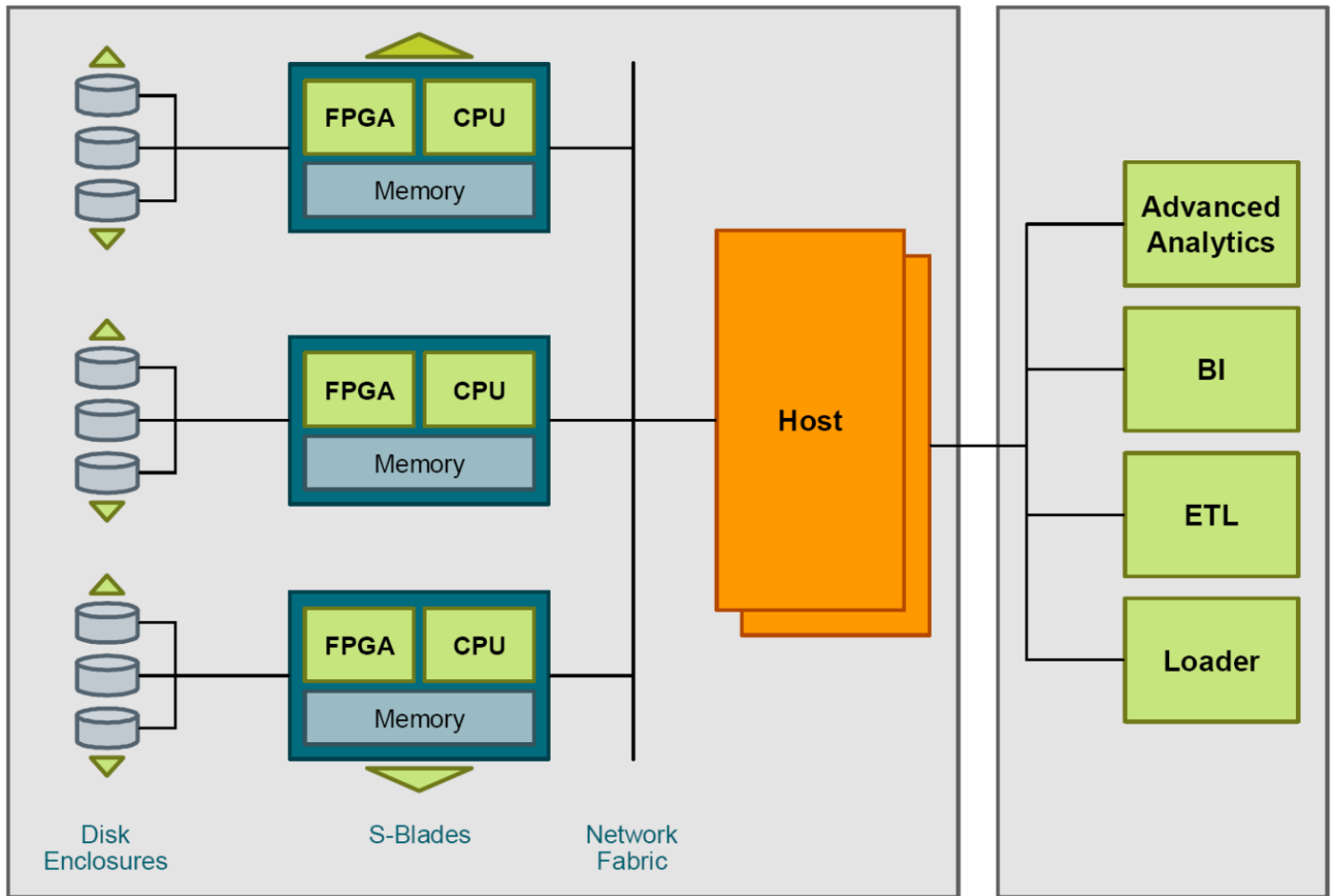


Abb. 14.4.4

Die Advanced Analytics, BI und ETL Anwendungen laufen auf einem getrennten Rechner

Einmal verteilt, führen die S-Blades die Verarbeitung durch. Wenn die Verarbeitung abgeschlossen ist, aggregiert der Host die Ergebnisse. Eine S-Blade ist ein kompletter Computer mit einem Prozessor, Hauptspeicher, Festplattenspeicher und einem Kernstück, dem „FPGA (Field Programmable Gate Array)“. Ein S-Blade enthält 16 CPU-Kerne und 16 FPGA-Kerne, wobei jedem CPU-Kern ein FPGA-Kern zugeordnet ist. Jedem FPGA-Kern sind darüber hinaus drei Festplatten zugeordnet.

Der „Loader“ bewirkt das Kopieren der Unternehmensdaten auf die Festplatten.

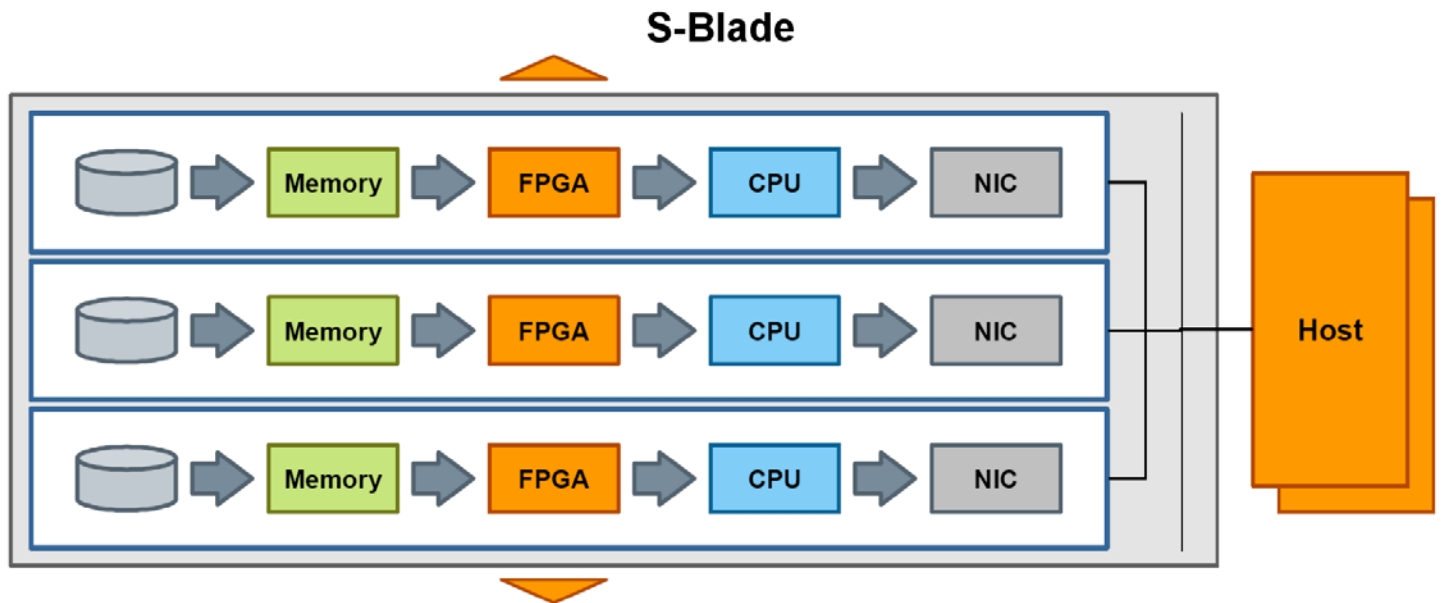


Abb. 14.4.5
Die Summe der S-Blades implementiert einen Shared Nothing Ansatz

Eine dediziertes High-Speed Interconnect Netzwerk (nicht gezeigt) liefert komprimierte Daten von den Festplatten in den Hauptspeicher der S-Blades mit maximaler Streaming-Geschwindigkeit. Die FPGAs entpacken und filtern 95-98% der Daten der gespeicherten Tabellen „on the fly“ heraus. Übrig bleiben nur die Daten, die benötigt werden, um die Anfrage zu beantworten. Diese werden von den CPUs der S-Blades weiterverarbeitet.

Anfragen werden im SQL Format gestellt (mit einigen Erweiterungen).

Der Network Interface Controller (NIC) stellt die Verbindung zum aktiven Host her.

<http://www.cedix.de/VorlesMirror/Band2/PDSA01.pdf>

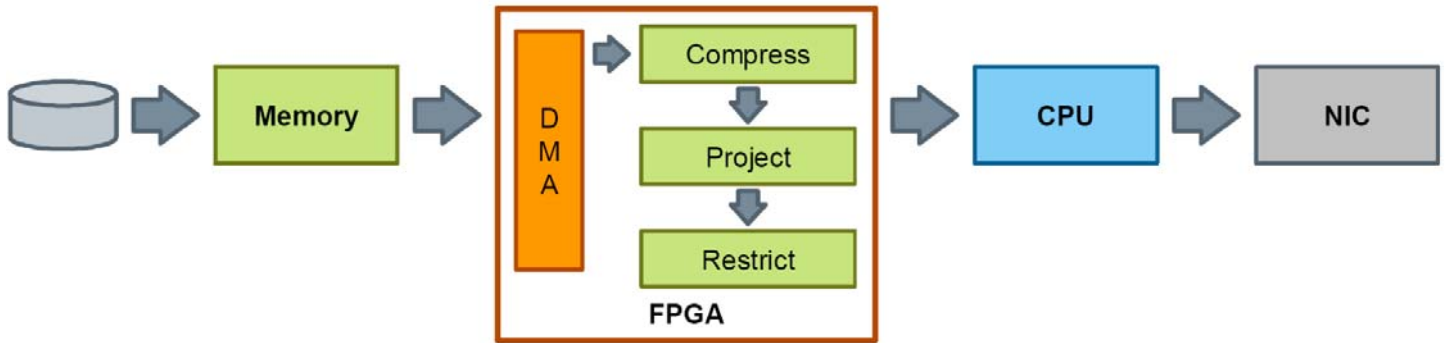


Abb. 14.4.6
Funktionen der FPGAs

Die FPGAs enthalten drei Engines:

- Die Compress Engine dekomprimiert Daten “on the fly”. Das Ergebnis ist eine signifikante Beschleunigung der langsamsten Komponente in jedem Data Warehouse, der Festplatte.
- Die Project and Restrict Engines bewirken eine weitere Leistungssteigerung durch Herausfiltern von Spalten und Zeilen, bezogen auf Parameter in den SELECT- und WHERE-Klauseln einer SQL-Abfrage.
- Die Visibility Engine spielt eine entscheidende Rolle bei der Aufrechterhaltung der ACID-Bedingungen. Sie filtert Zeilen aus, die bei einer Abfrage unsichtbar sein sollten, z. B. Zeilen, die zu einer Transaktion gehören, welche noch nicht abgeschlossen (COMMIT) wurde.

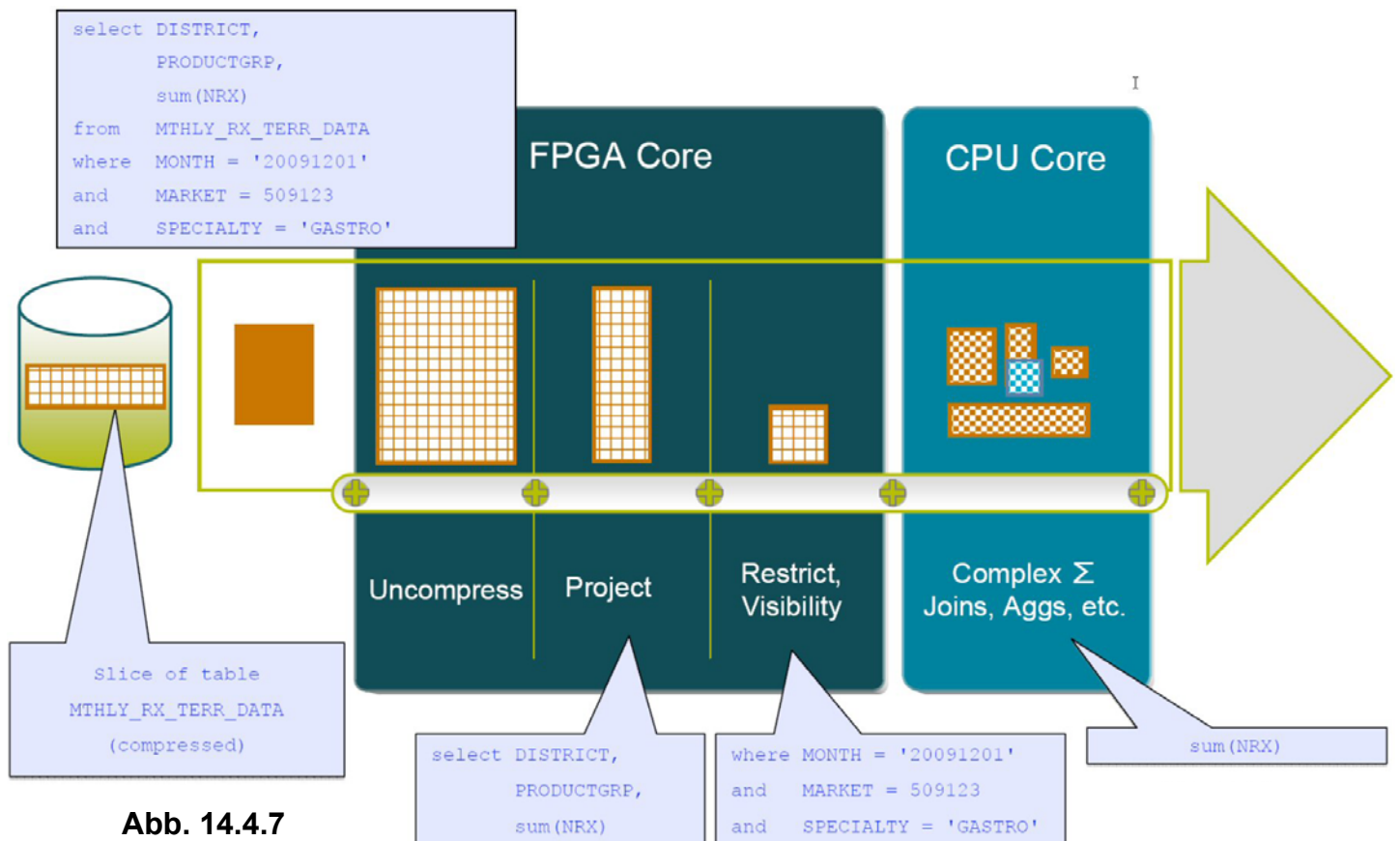


Abb. 14.4.7
FPGA Funktionen

14.4.6 PureData System for Analytics N2001-010 Hardware



Abb. 4.4.8
PureData System for Analytics N2001 Hardware

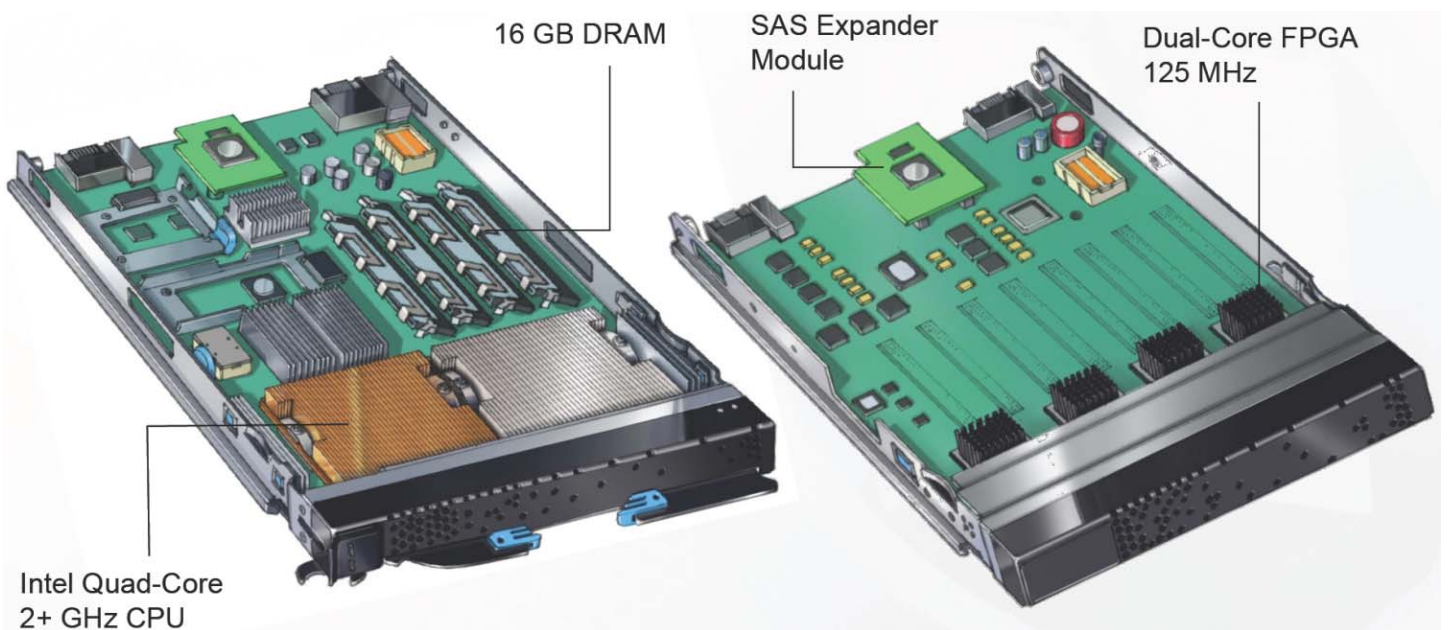
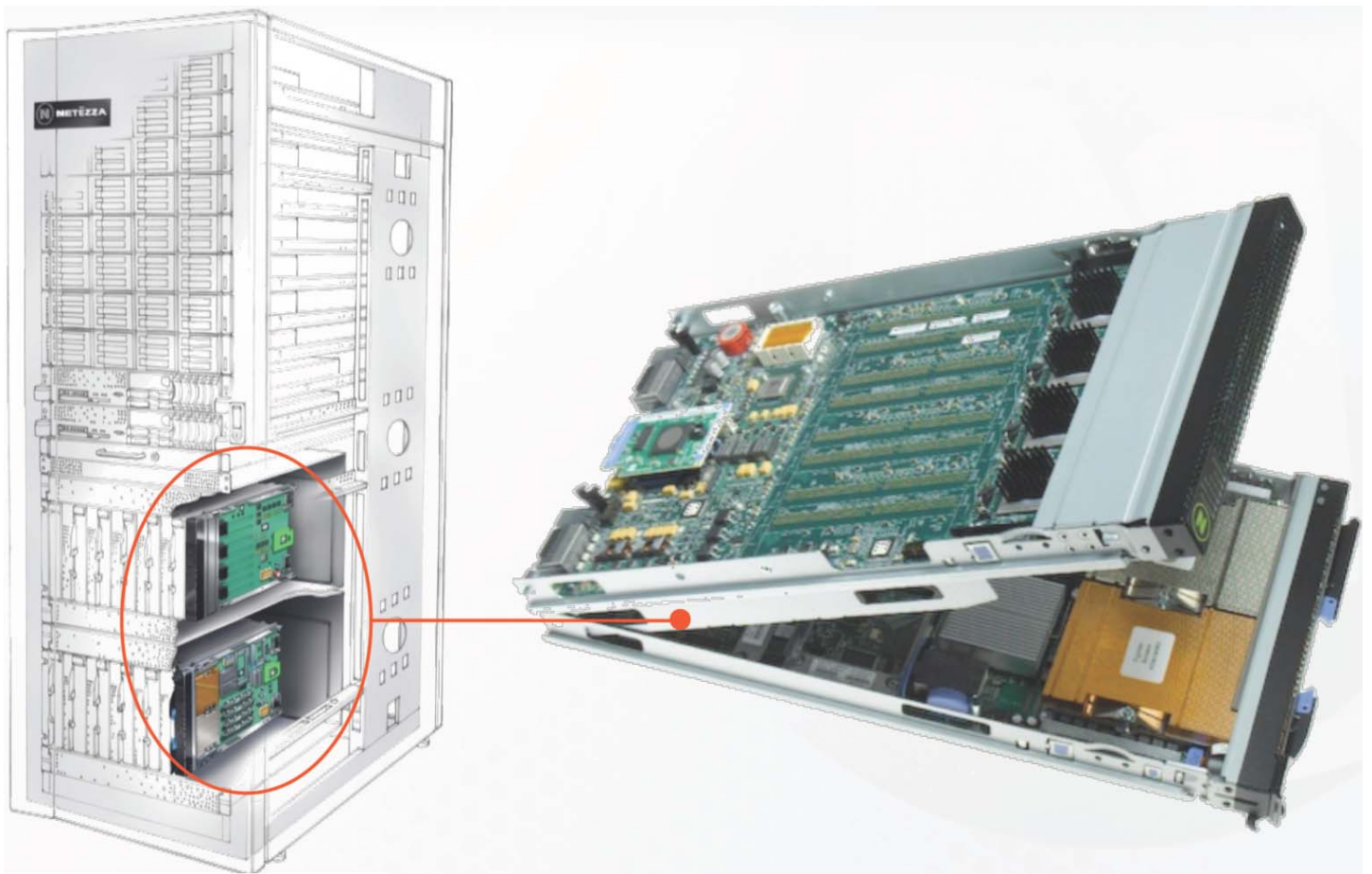


Abb. 14.4.9
Aufbau der S-Blades

Jedes PureData for Analytics S-Blade besteht aus 2 Platinen. Die linke Platine ist ein Standard-x86-Blade, wie es von IBM auch in anderen Blade-Center-Produkten eingesetzt wird. Die rechte Platine ist proprietär. Sie enthält die FPGA-Bausteine, die für den großen Performance-Gewinn verantwortlich sind.

14.4.7 Benutzer definierte Funktionen

FPGAs sind Halbleiterchips, die programmiert werden können. Sie werden seit vielen Jahren für eine Vielzahl von Anwendungen eingesetzt. Netezza ist das erste Produkt, welches FPGAs benutzt, um Streaming-Daten in einem Data Warehouse Appliance verarbeiten. In allen anderen Data Warehouse Systemen werden alle Daten für eine Abfrage eingelesen. Danach wird die SQL WHERE Clause verarbeitet. Mit Netezza, anstatt eine riesige Menge von Daten zu bewegen, verarbeitet ein FPGA die WHERE Clause als Teil des Streamings der Daten von der Festplatte. Nur die Daten, die für die Verarbeitung des nächsten Schrittes erforderlich sind, werden weiter gereicht.

PureData System for Analytics (Netezza) Benutzer können UDFs (User Defined Functions) und UDAs (User Defined Aggregates) schreiben (beide als UDxs bezeichnet). UDAs geben einen Wert für mehrere Zeilen zurück, während UDFs einen Wert pro Zeile zurückgeben. Auf UDAs- und UDFs kann mit einer einfache SELECT-Anweisung zugegriffen werden.

Mit UDxs erlaubt Netezza das Schreiben von eigener Verarbeitungslogik, wie Data Mining-Algorithmen, für die S-Blades. Durch die Umsetzung dieser und anderer Technologien erreicht Netezza Geschwindigkeiten, die 10-100 Mal höher als bei andere Lösungen sind.

Es ist möglich, eigene UDxs erstellen. Unabhängige Softwareanbieter wie Fuzzy Logix (<http://www.fuzzyl.com/>) bieten Software-Pakete an, die Hunderte von vorgefertigten Mathematik, Statistic, Data Mining Algorithmen, sowie finanziellen Modellen enthalten.

14.4.8 IBM DB2 Analytics Accelerator for z/OS

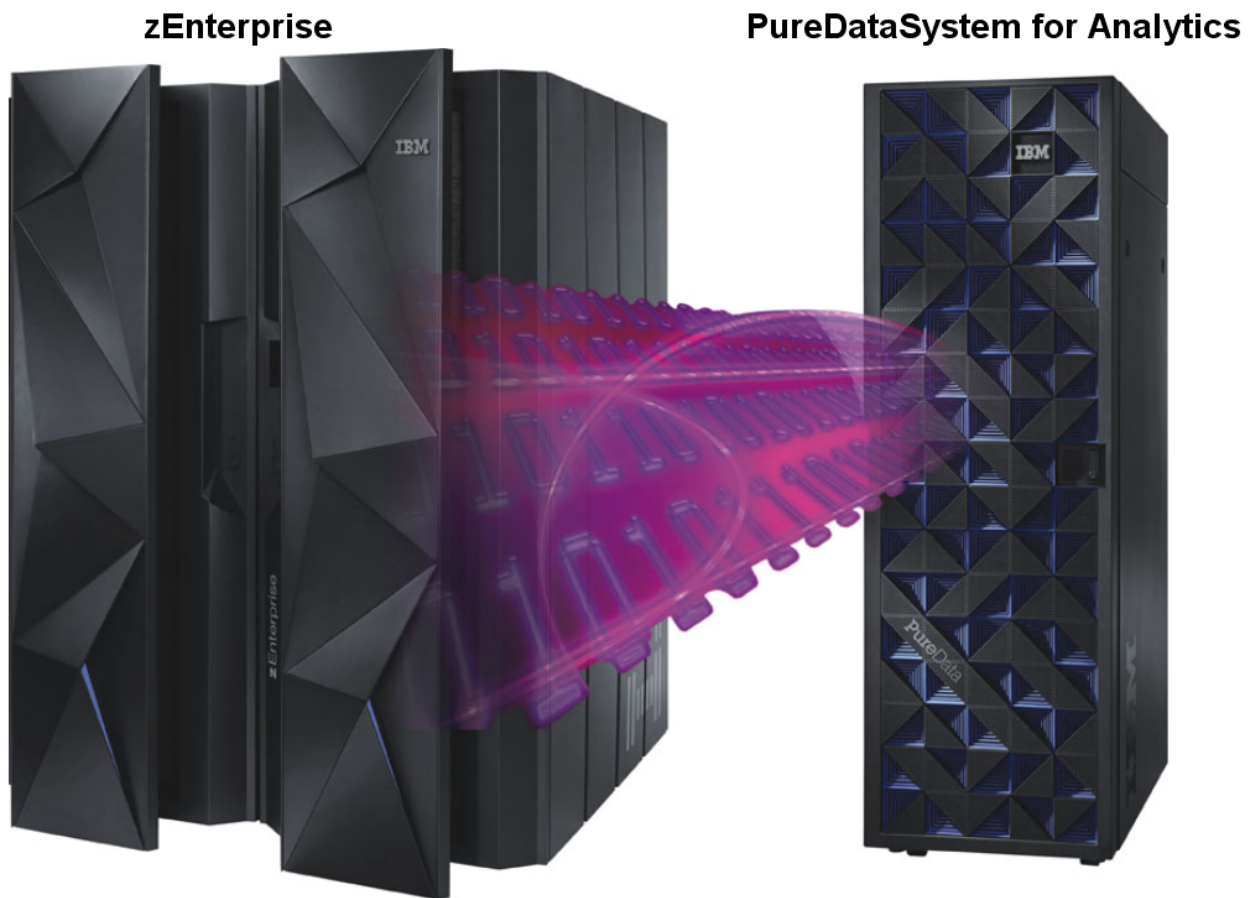


Abb. 14.4.10
14) Netezza Anschluss an einen zEC12 Rechner

In einem mittelgroßen Unternehmen kann ein Unix-, Linux- oder Windows-Rechner als Primärsystem dienen, auf dem die unternehmenskritischen Anwendungen laufen. Ein PureData System for Analytics kann an beliebige Primärsysteme angeschlossen werden. Von besonderem Interesse ist jedoch die Verbindung mit einem z/OS-Rechner, weil hier zahlreiche zusätzliche Funktionen möglich sind.

Der **“IBM DB2 Analytics Accelerator” for z/OS** (IDAA, englische Aussprache „Eye-Da“) ist eine Data Warehouse Appliance für IBM System z und basiert auf dem PureData System for Analytics (Netezza). IDAA kombiniert Eigenschaften von DB2 für z/OS mit den Stärken eines PureData System for Analytics.

IDAA ist in DB2 für z/OS integriert. Es wird eine vollständige Transparenz der Anwendungen erreicht, die Abfragen an DB2 für z/OS übergeben. Benutzer und Anwendungen sehen nur eine DB2-für-z/OS Schnittstelle.

IDAA nutzt Kopien der DB2 -Tabellen, die im PureData-System gespeichert werden. Dies bietet das Beste aus beiden Welten: die gute Leistung von DB2 für z/OS für transaktionale Abfragen und branchenführende Leistung für analytische Abfragen. Abfragen werden von IDAA an das PureData-System weitergeleitet und von diesem verarbeitet.

14.4.9 IDAA – z/OS Integration

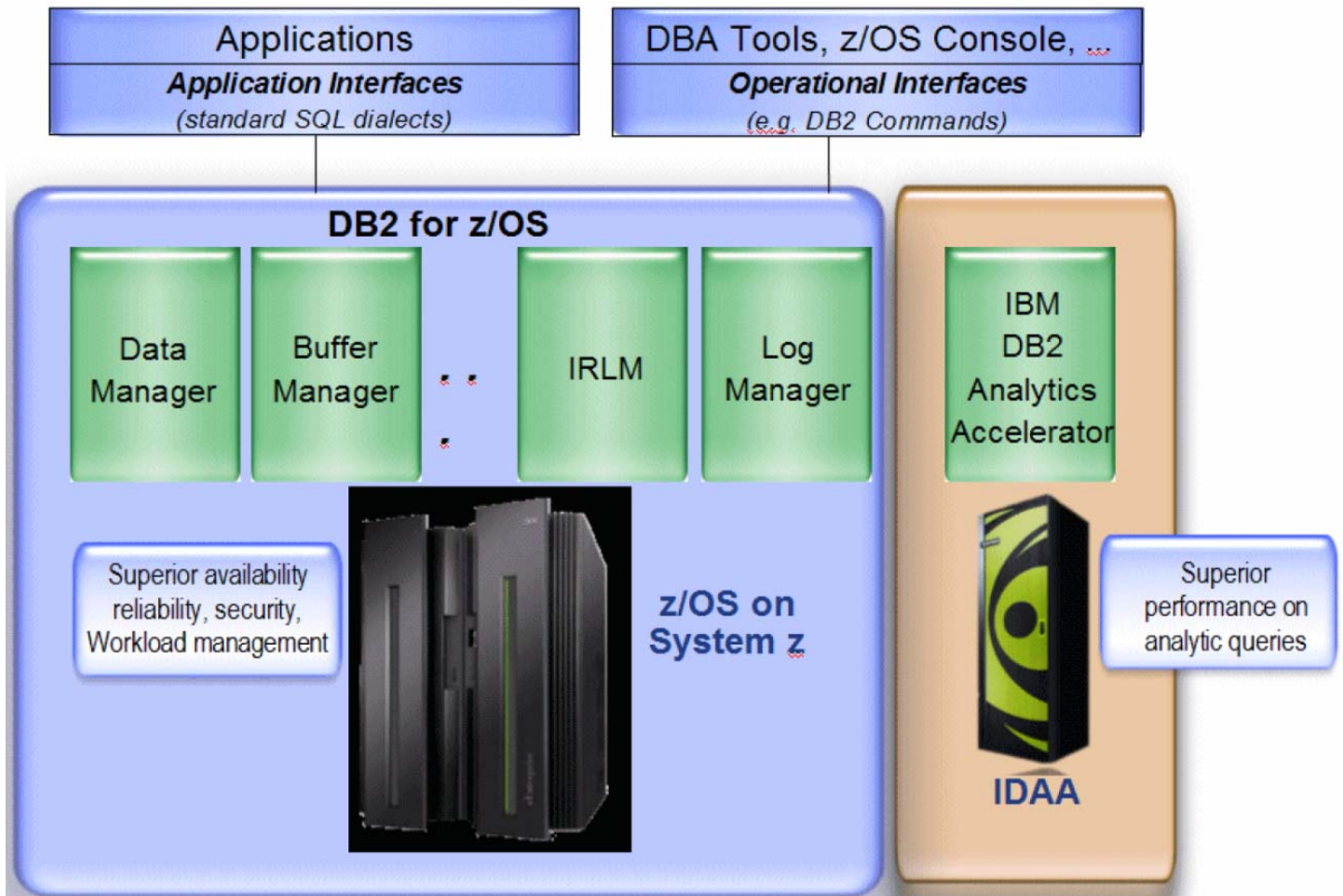


Abb. 14.4.11

Der DB2 Analytics Accelerator ist eine zusätzliche DB2 Komponente

DB2 besteht aus mehreren internen Komponenten, den Ressourcenmanagern: Buffer Manager für Daten-Cache-Management, Interne Ressourcen-Manager (IRLM) für Locking, usw. (Band 1, Abschnitt 7.1.2). Diese Komponenten sind für die Außenwelt von Anwendungen und Datenbank-Administration nicht sichtbar, wenn diese sich mit DB2 durch Anwendungsschnittstellen wie SQL oder operative Schnittstellen wie Commands und Utilities verbinden. Die interne Verarbeitungssteuerung unter den Komponenten ist vollständig transparent für die Außenwelt. Wir können damit einen anderen Resource Manager hinzufügen unter Beibehaltung der bestehenden Schnittstellen. Keine Änderung einer Anwendung oder eines Datenbank Administrations-Verfahrens ist erforderlich.

Die tiefe Integration des PureData-Systems in DB2 für z/OS unterscheidet IDAA von anderen Data-Warehouse-Lösungen.

14.4.10 DB2 Optimizer

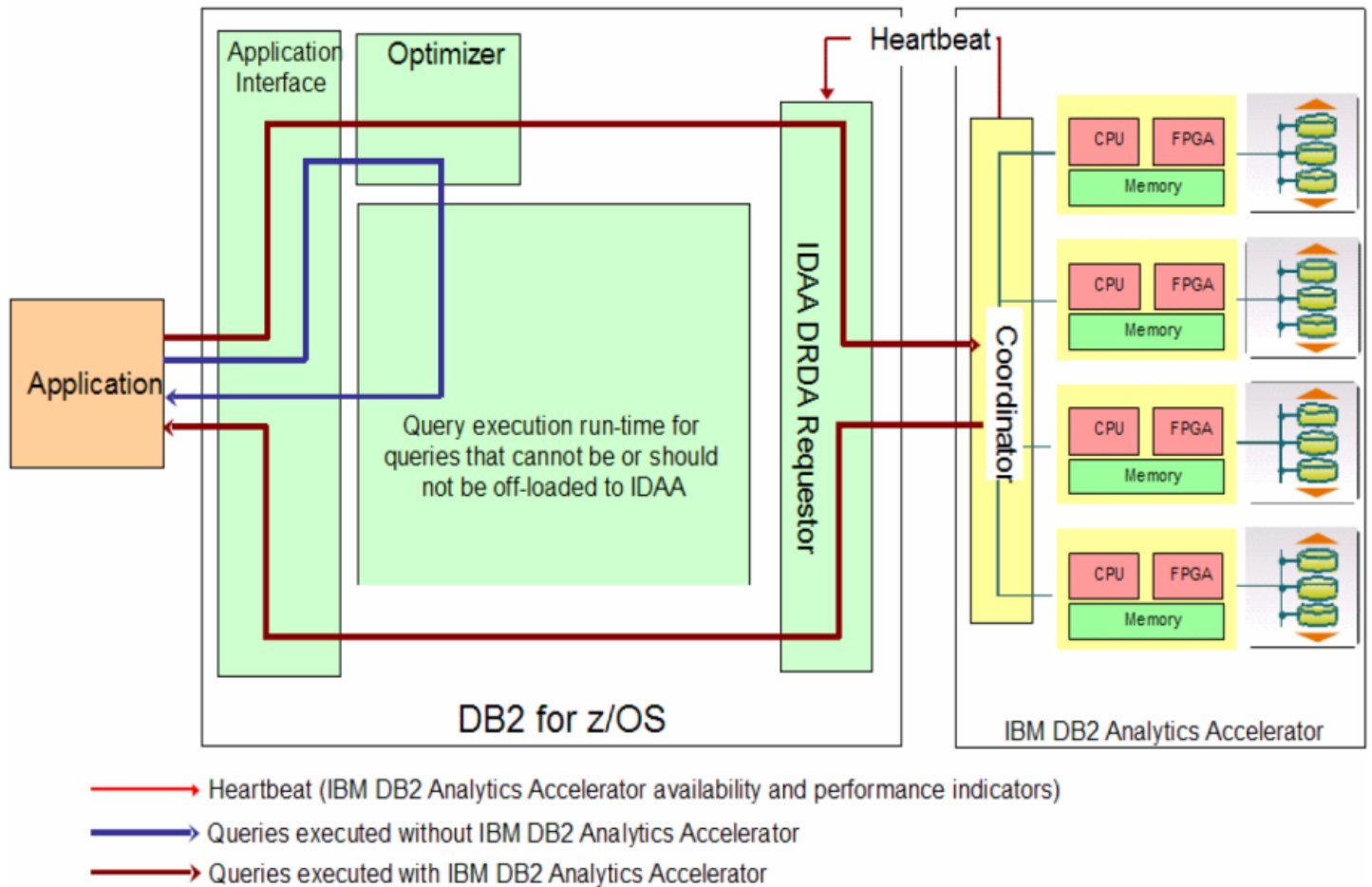


Abb. 14.4.12
Anschluss des DB2 Analytics Accelerators

Die Optimizer-Komponente von DB2 für z/OS entscheidet, ob eine SQL-Anfrage unter z/OS verarbeitet wird, oder an IDAA und letztendlich an das PureData-System weitergeleitet wird.

z/OS und der IBM DB2 Analytics Accelerator for z/OS repräsentieren derzeit die technologische Sperspitze auf dem Gebiet Data Mining.

z/OS Appliances sind Acceleratoren für z/OS. Neben den Power und x86 Blades in einer zBX spielen „WebSphere DataPower Appliances“ (Abschnitt 14.3.3) eine wichtige Rolle. Sie sind ebenfalls eng in die zBX Infrastruktur integriert.

Ein PureData System for Analytics kann zunächst in beliebigen Umgebungen eingesetzt werden. Es kann aber auch in Verbindung mit IDAA als eine z/OS Appliance eingesetzt werden. In dieser Rolle benutzt es heute nicht die zBX und Unified Resource Manager Infrastruktur, und dafür eine eigene DB2 Infrastruktur Integration. Wir spekulieren, dass die zBX und Unified Resource Manager Infrastruktur Integration zu einem späteren Zeitpunkt ebenfalls erfolgen wird.

In der Zwischenzeit sind das PureData System for Analytics und IDAA weitere Beispiele, wo sich das z/Ost-Betriebssystem von anderen Plattformen durch seine zahlreichen technologisch führenden Eigenschaften abhebt.

4.5 Weiterführende Information

Ein technisches Hewlett Packard Video

<http://www.youtube.com/watch?v=PpbEeW7ltHs&feature=endscreen&NR=1>

Eine detaillierte zBX Darstellung ist enthalten in

<http://www.cedix.de/VorlesMirror/Band2/EMA02.pdf>

Ein vertriebsorientiertes Video über zBX und den Unified Resource Manager existiert unter

<http://www.youtube.com/embed/CVPS1vkCaBs>

und

<http://www.engadget.com/2010/07/23/ibms-zenterprise-architecture-makes-mainframes-cool-again-also/>

Ein Netezza Video ist zu finden unter

<http://www.youtube.com/watch?v=iO-Mq5AhmXw>

<http://www.cedix.de/VorlesMirror/Band2/zBXdetail.pdf>

enthält eine detaillierte Präsentation zum Thema zBX

<http://www.cedix.de/VorlesMirror/Band2/PDSA01.pdf>

beschreibt die PureData System for Analytics N2001 Ankündigung Februar 2013

FPGA Tutorials sind zu finden unter

http://www.cs.ucr.edu/~vahid/courses/122b_w02/dt96_fpgatutorial.pdf

http://www.altera.com/literature/tt/tt_my_first_fpga.pdf

<http://www.eetimes.com/design/programmable-logic/4015129/How-to-design-an-FPGA-from-scratch>

<http://www.mu21.de/Development//users/.michael/studium/vhdl/2-praktikum/dokumentation/1-calculator.pdf>

Ein FPGA Video Tutorial ist zu finden unter

<http://www.youtube.com/watch?v=wwkGmDaa4oM>

Beispiele für Spezialelektronik unter Benutzung von FPGAs sind zu finden unter

<http://www.gbm.de/fpga-systeme/?gclid=CPLD1pergbQCFUy5zAodCkgAYQ>

Ein Home Computer (nicht x86 kompatibel) mit FPGAs implementiert ist zu sehen in

<http://www.youtube.com/watch?v=dPuSGyrbd6c>

Ein weiterer Home Computer

http://www.youtube.com/watch?v=GK6QzsSut_E

Ein FPGA PC in Aktion ist zu sehen unter
<http://www.youtube.com/watch?v=V0jXQXZHToE>

Interessante Details in:

IBM Redbook: Optimizing DB2 Queries with IBM DB2 Analytics Accelerator for z/OS

<http://www.redbooks.ibm.com/abstracts/sg248005.html>

Andersartige Data Mining Ansätze von Oracle und SAP, siehe

<http://www.cedix.de/VorlesMirror/Band2/SAP01.pdf>

Trend: Zentralisierung auf dem Mainframe

Siehe:

<http://www.cedix.de/VorlesMirror/Band2/Rubin01.pdf>