UNIVERSITÄT LEIPZIG

Mainframe Internet Integration

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

SS2013

Hybrid Computing Teil 4

Netezza

© copyright W. G. Spruth,

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.

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.

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“.

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**
- **Business Intelligence**

Im Internet:

- **Angriffserkennung (Security Attacks)**
- **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.**

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 Kassenbons**

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.

Data Mining

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

ETL

Die Abkürzung ETL steht für „Extraktion, Transformation, Laden“, ein Prozess, der Daten aus mehreren Datenquellen in einer 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 Istsituation 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?**

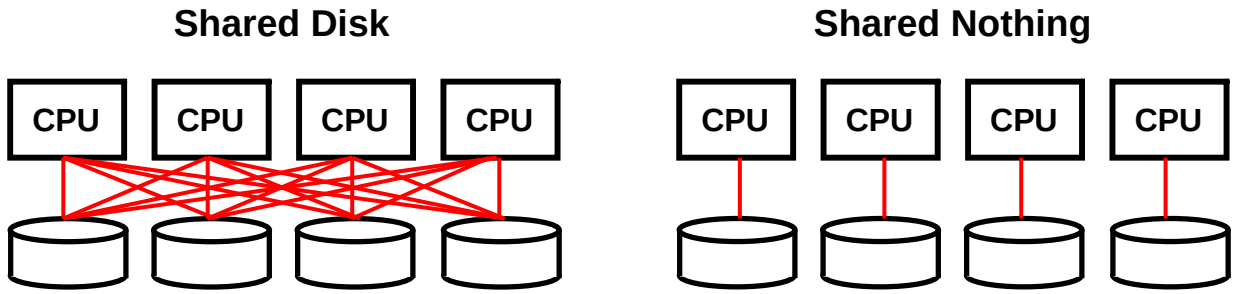
Optimierung

- **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?**

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 mehren 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.



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

Shared Nothing

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, 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 Dara 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 In-Memory Computing Produkt.**

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.

IBM PureData System for Analytics (Netezza)

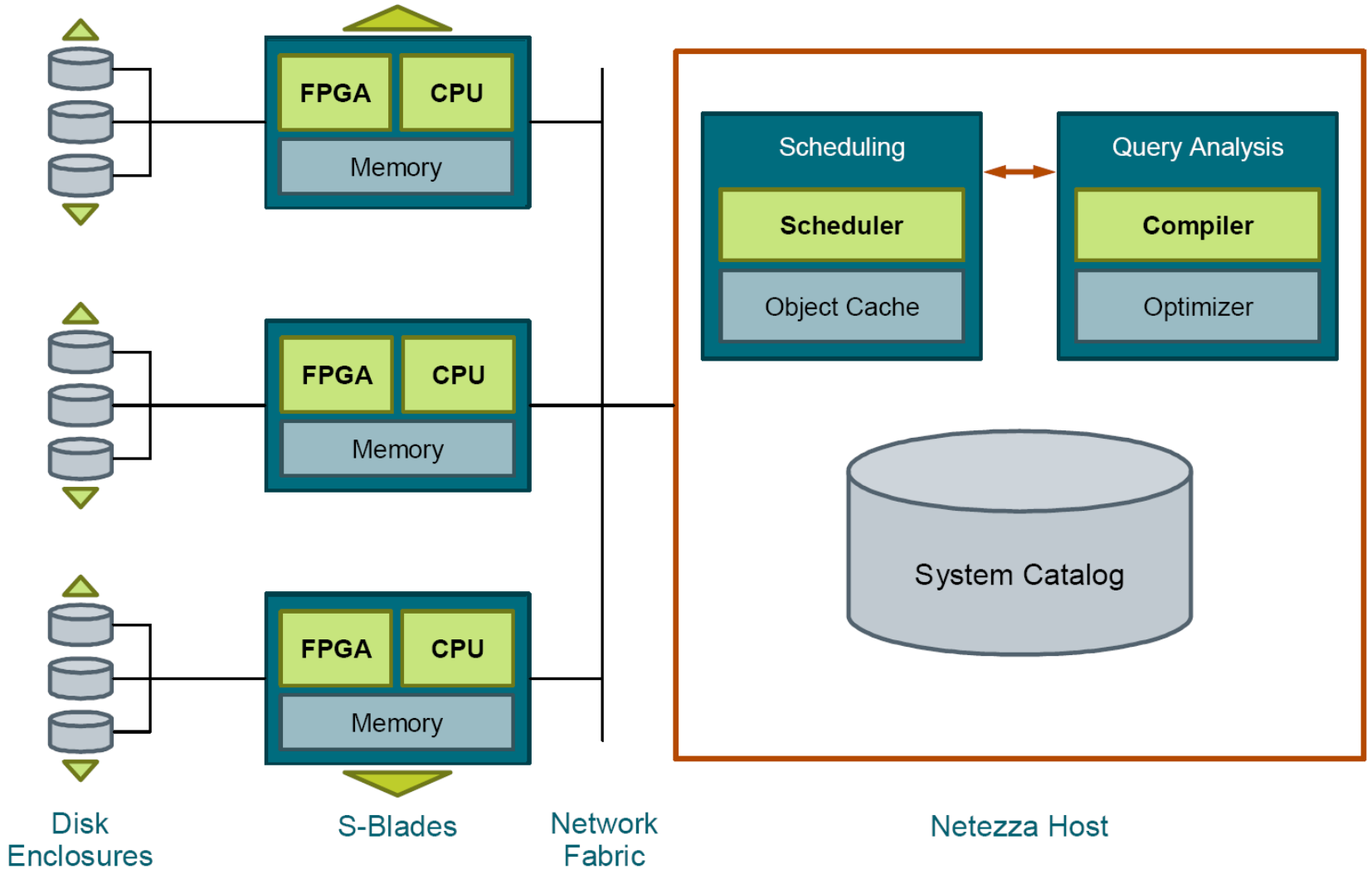
Das IBM PureData System for Analytics (früher: Netezza, englische Aussprache: „netiesa“) ist eine Data-Warehouse-Appliance, d.h. ein kombiniertes System aus Hardware und speziell darauf abgestimmter Software, welches sich von alternativen Produkten wie z. B. Teradata oder Exadata durch zwei Merkmale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

- **Ein PureData System for Analytics ist zwar unabhängig von zBX. Der „IBM DB2 Analytics Accelerator for z/OS“ 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.

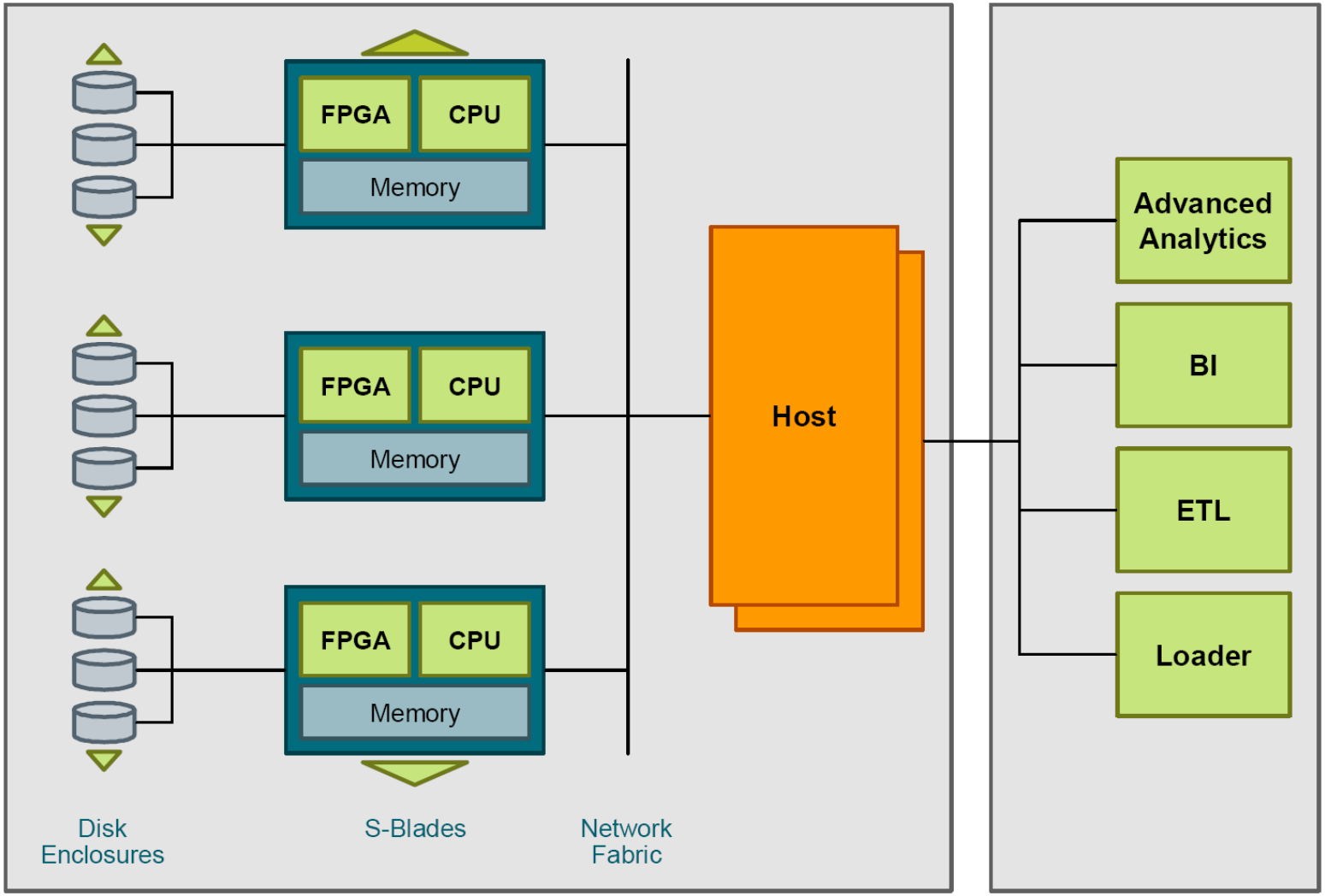
Mit der Einführung der FPGAs wurden kompakte, anwenderspezifische Schaltungen in geringen Stückzahlen ermöglicht. Heutzutage gestatten sie z. B. die preiswerte und flexible Fertigung komplexer Systeme wie Mobilfunk-Basisstationen als Alternative zur teureren Auftragsfertigung durch Halbleiterhersteller.

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



Die Basiskonfiguration eines PureData System for Analytics besteht aus 144 SAS2 600-GB Festplatten (bei vierfacher Komprimierung stehen dem Benutzer 96 TB 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.

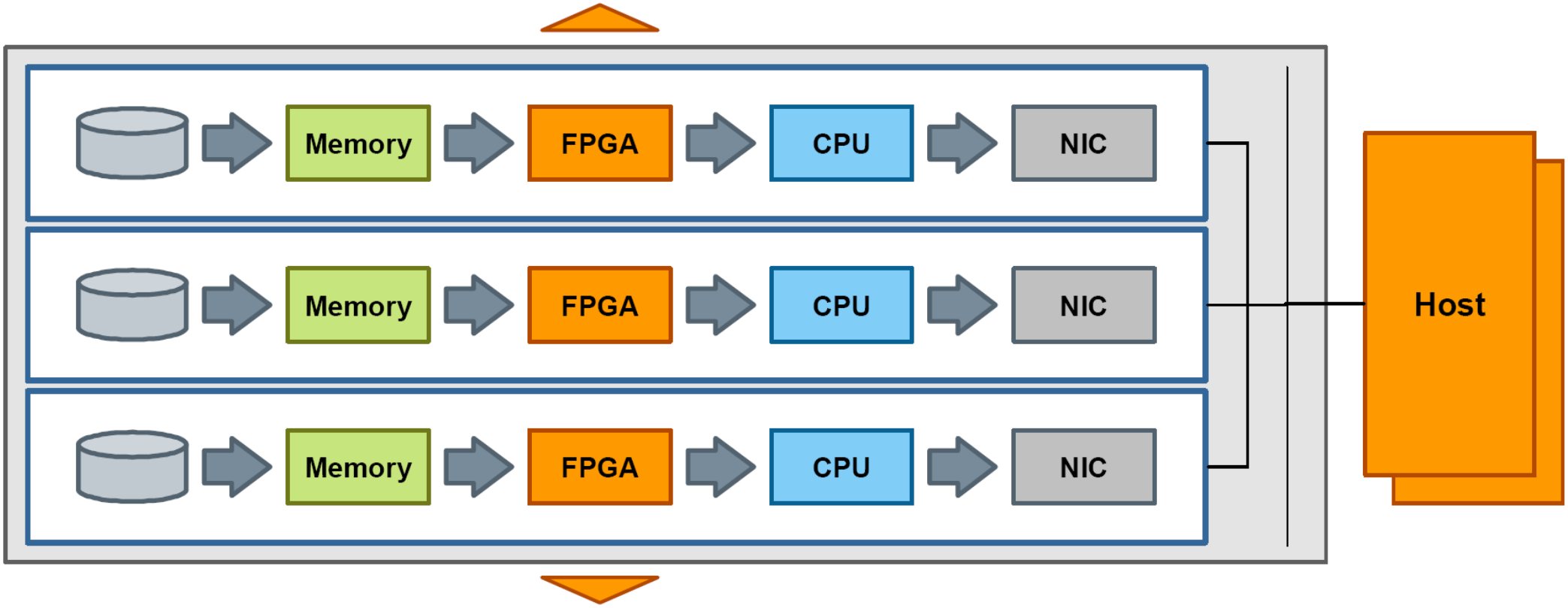
Der Host ist ein selbständiger Hochleistungs-Linux-Server, der aus Verfügbarkeitsgründen doppelt vorhanden ist. Er kompiliert SQL-Abfragen von Advanced Analytics, BI oder ETL-Anwendungen in ausführbare Code-Segmente und verteilt diese zur Ausführung auf die S-Blades.



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.

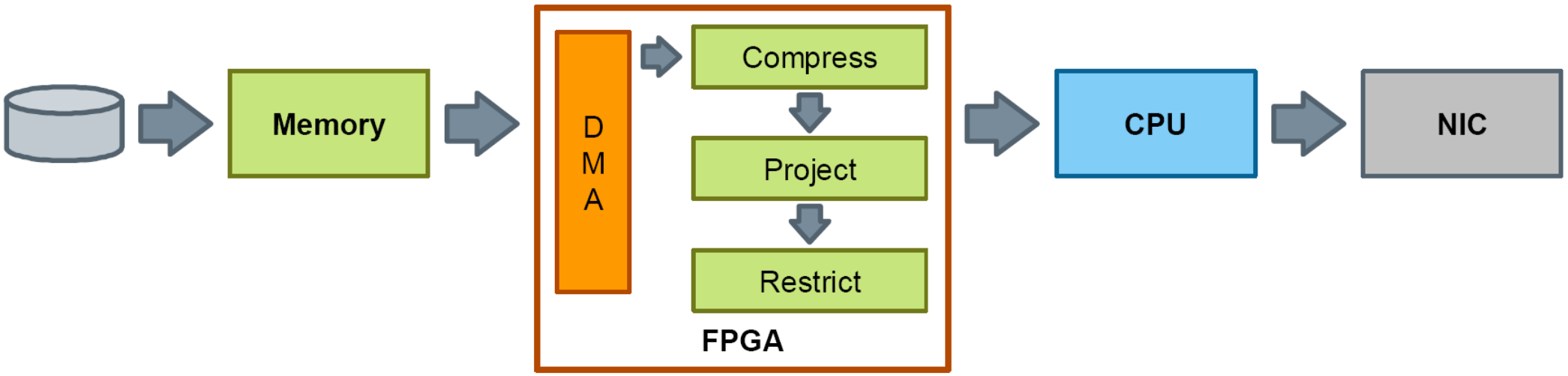
S-Blade



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.

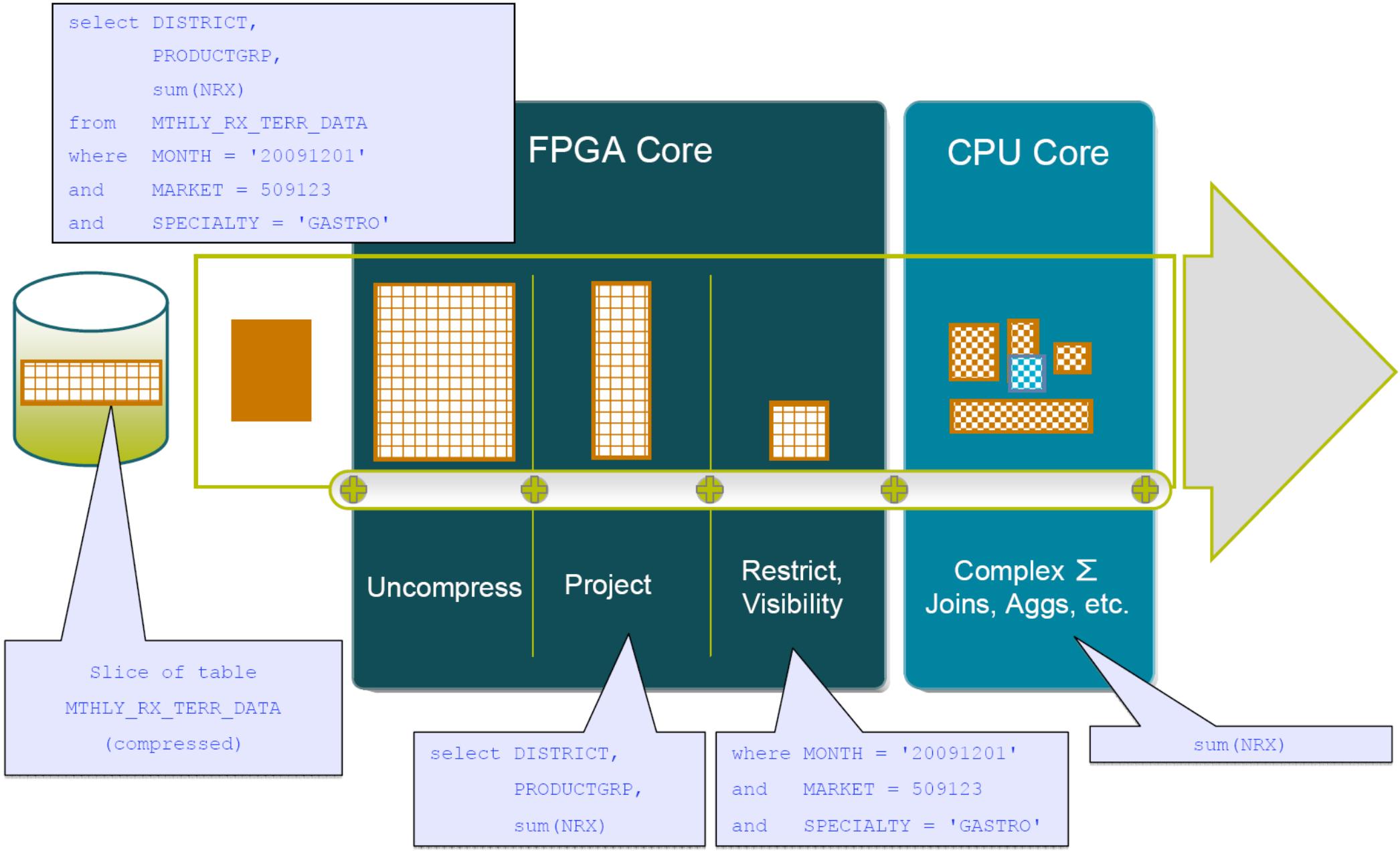
Anfragen werden im SQL Format gestellt (mit einigen Erweiterungen).

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



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.**


select DISTRICT,
PRODUCTGRP,
sum(NRX)
from MTHLY_RX_TERR_DATA
where MONTH = '20091201'
and MARKET = 509123
and SPECIALTY = 'GASTRO'
FPGA Core
CPU Core
Uncompress
Project
Restrict, Visibility
Complex Σ Joins, Aggs, etc.
Slice of table MTHLY_RX_TERR_DATA (compressed)
select DISTRICT,
PRODUCTGRP,
sum(NRX)
where MONTH = '20091201'
and MARKET = 509123
and SPECIALTY = 'GASTRO'
sum(NRX)

Disk Enclosures
- Slices of User Data
- Swap and Mirror Partitions
- High-Speed Data Streaming

SMP Hosts
- SQL, Query Plan, Optimize, Admin

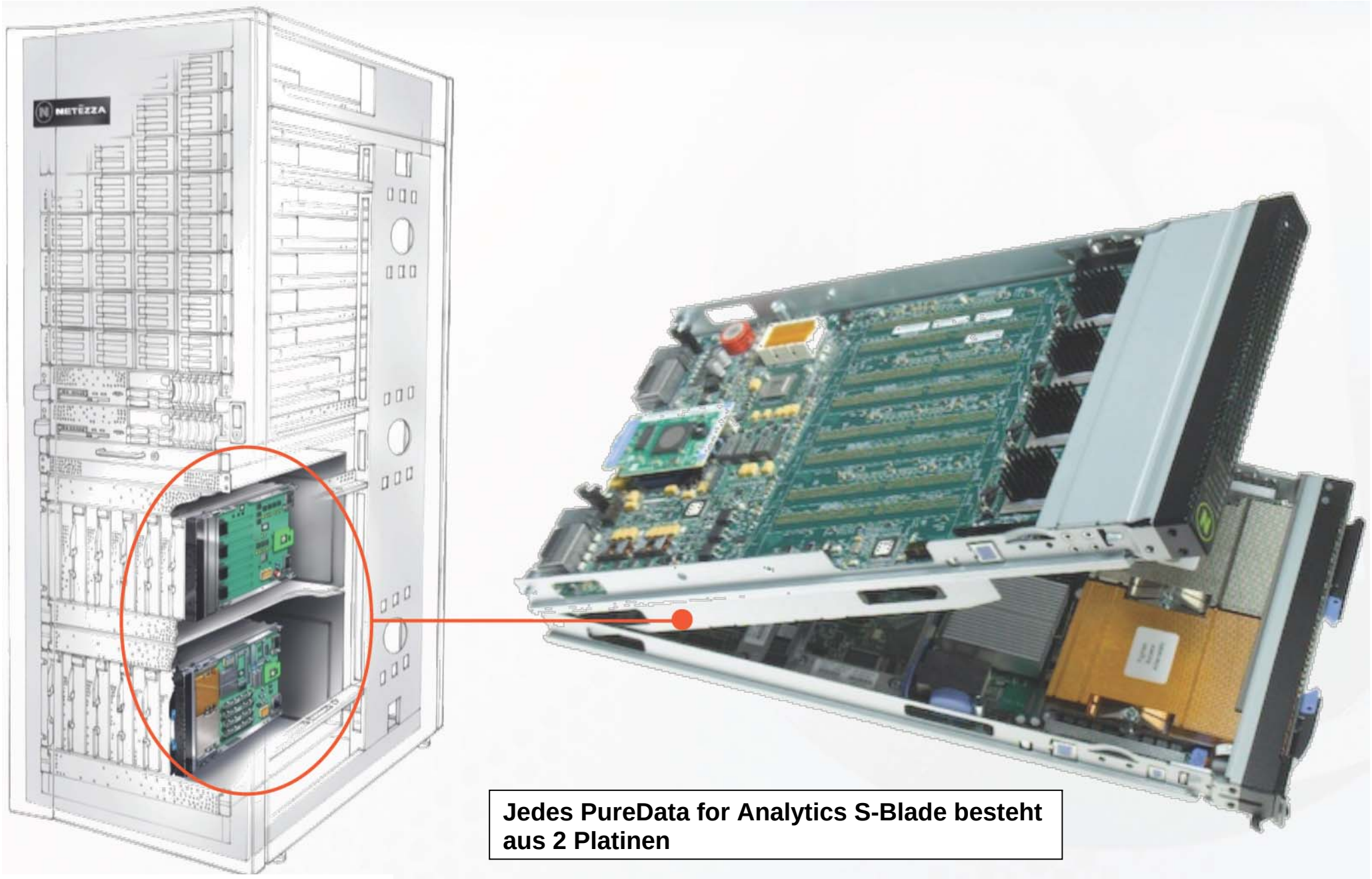
Snippet Blades (S-Blades)
- Processor & Streaming DB Logic
- High-performance database engine, complex analytic processing, streaming joins, aggregations, sorts, etc …

IBM PureData System for Analytics uses Field Programmable Gate Arrays (or FPGAs) that are programmed to handle large volumes of data very efficiently. These FPGAs eliminate all unnecessary I/O calls, and then further reduce and filter the data as fast as it streams off the disk. This removes I/O bottlenecks and frees up downstream components such as the CPU, memory and network from processing unnecessary data, creating a significant turbocharger effect of system performance.

Complex analytics are performed on powerful multi-core CPUs, where database primitives and the analytical processing are executed on the filtered data stream. Analytic tasks are run as independent processes operating on data streams on each S-Blade. The IBM Netezza Analytics platform harnesses the power of all computational cores and the MPP scalability of the PureData system for Analytics architecture for advanced analytics, while presenting an abstracted single view to simplify their deployment.

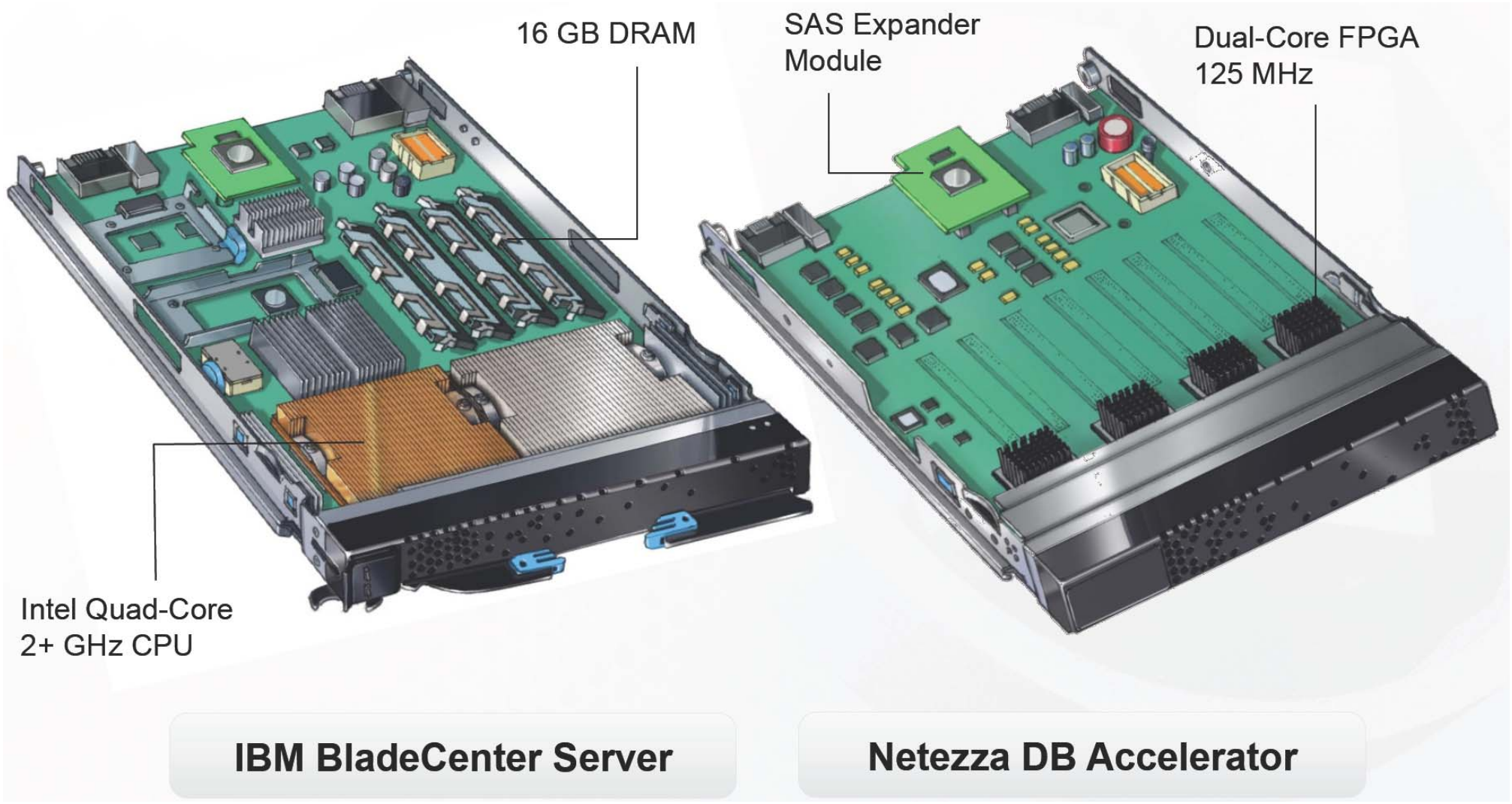
PureData System for Analytics N2001-010: Hardware

Single-Rack-Systeme sind in einem 2 Meter hohen, 65 cm breiten Gehäuse untergebracht.



Jedes PureData for Analytics S-Blade besteht aus 2 Platinen

http://www.monash.com/uploads/netezza.pdf



Eine der beiden Platinen ist ein Standard-x86-Blade, wie es von IBM auch in anderen Blade-Center-Produkten eingesetzt wird.

Die zweite Platine ist proprietär. Sie enthält die FPGA-Bausteine, die für den großen Performance-Gewinn verantwortlich sind.

User Defined Functions

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.

UDX (UDF und UDA)

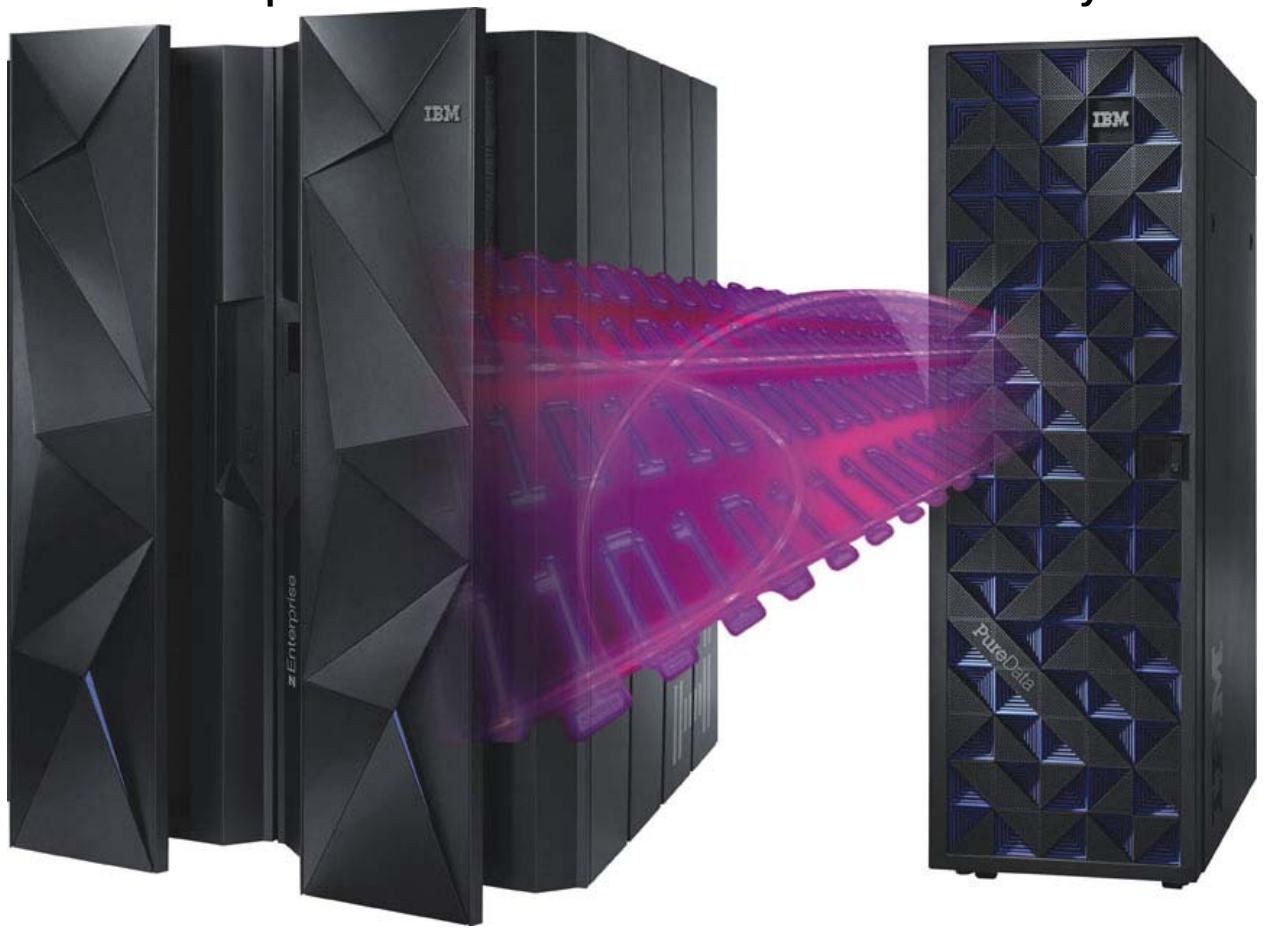
Netezza Benutzer können UDFs (User Defined Functions) und UDAs (User Defined Aggregates) schreiben (beide als UDxs bezeichnet). UDAs geben einen Wert für mehrere Zeilen zurück, während UDFs einen Wert pro Zeile zurückgeben. Auf UDAs-und UDFs kann mit einer einfache SELECT-Anweisung zugegriffen werden.

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.

zEnterprise

IBM PureDataSystem for Analytics



IBM DB2 Analytics Accelerator for z/OS

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.

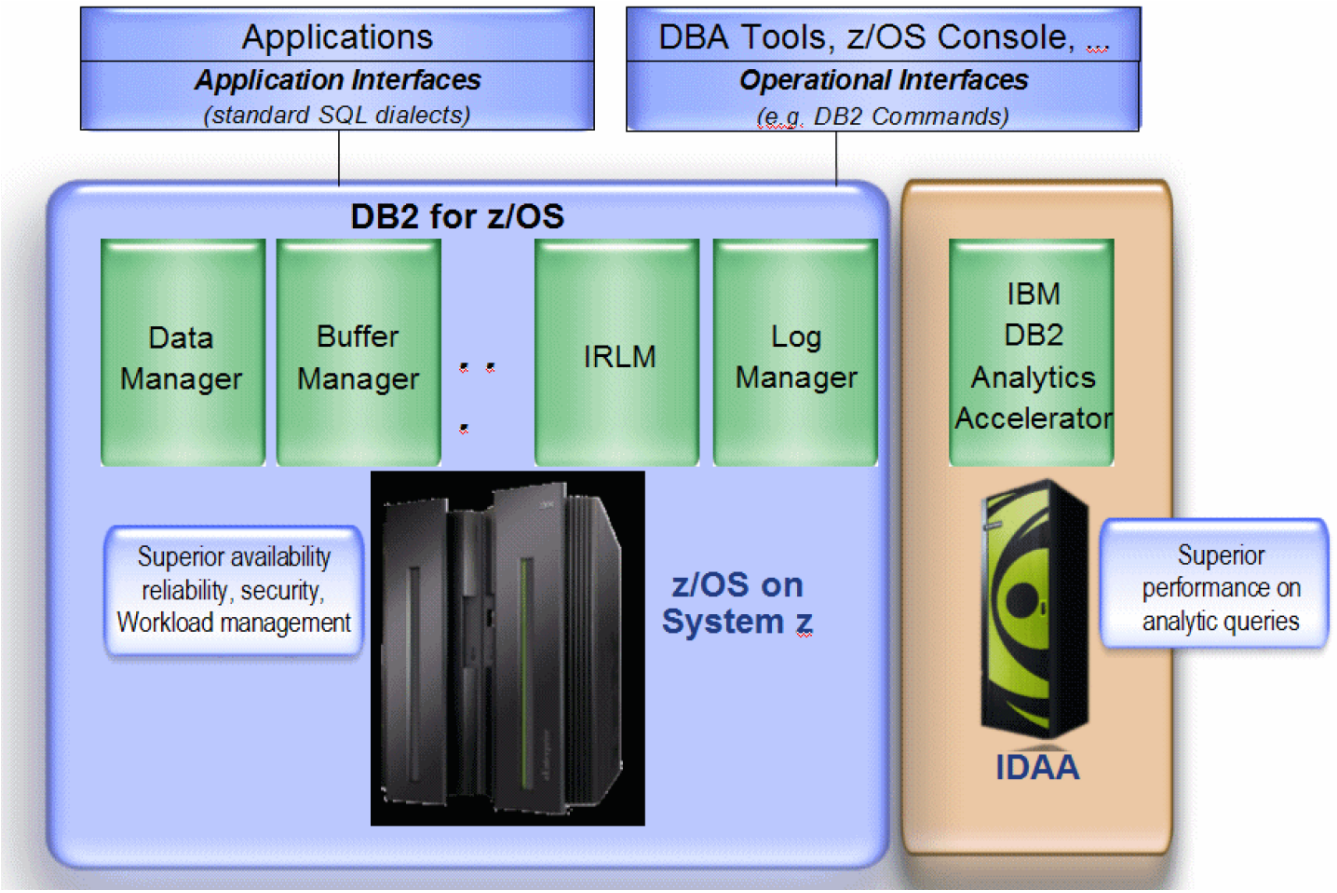
IBM DB2 Analytics Accelerator for z/OS

Der “IBM DB2 Analytics Accelerator” for z/OS (IDAA) 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.

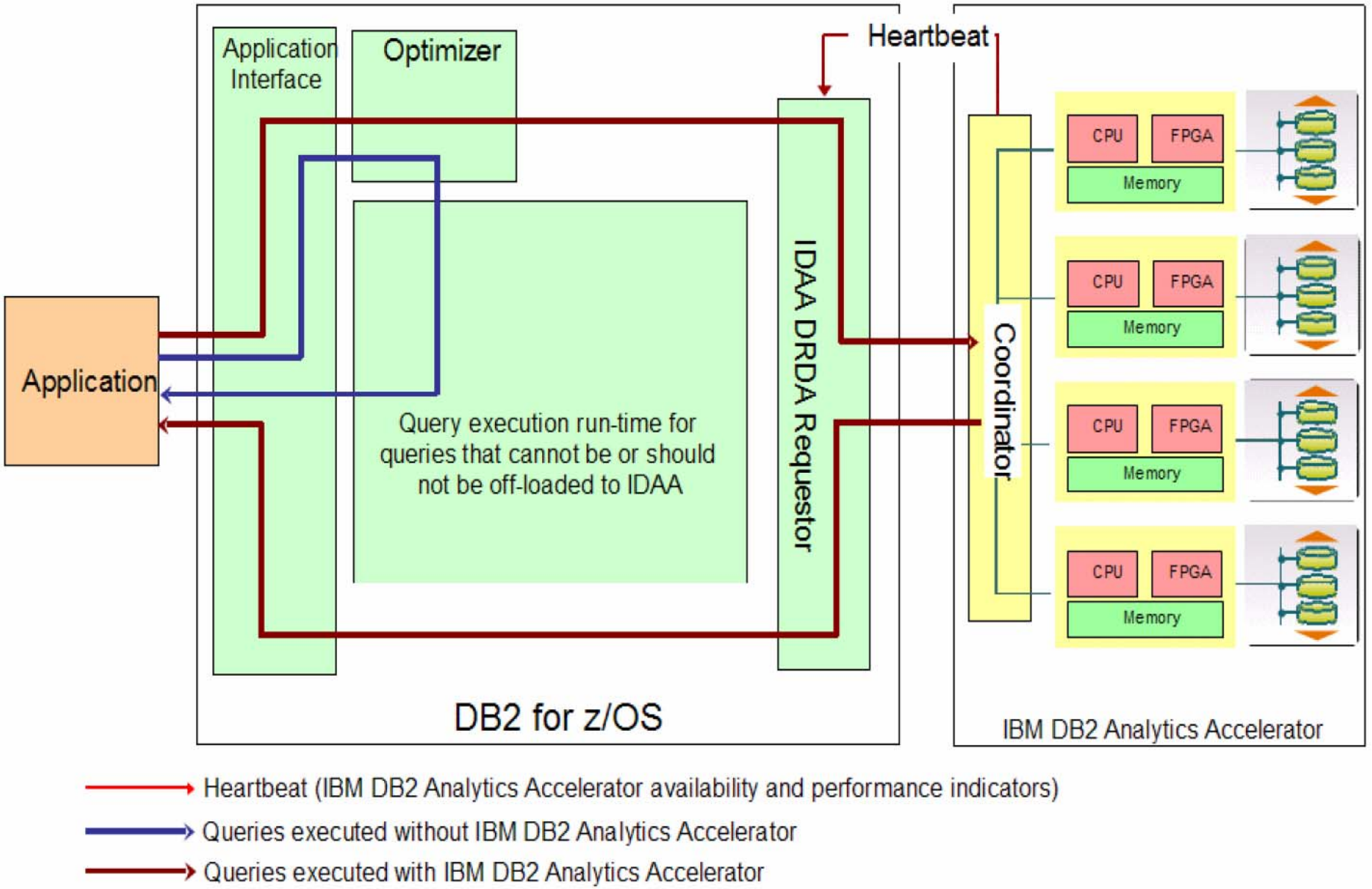
IDAA nutzt Kopien der DB2 -Tabellen, die im PureData-System gespeichert werden. Dies bietet das Beste aus beiden Welten: die excellente Leistung von DB2 für z/OS für transaktionsionale Abfragen und branchenführende Leistung für analytische Abfragen. Abfragen werden von IDAA an das PureData-System weitergeleitet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

Literatur:
IBM Redbook: Optimizing DB2 Queries with IBM DB2 Analytics Accelerator for z/OS
http://www.redbooks.ibm.com/abstracts/sg248005.html



DB2 besteht aus mehreren internen Komponenten, den Ressourcenmanagern: Buffer Manager für Daten-Cache-Management, Interne Ressourcen-Manager (IRLM) für Locking, usw. Diese Komponenten sind für die Außenwelt von Anwendungen und Datenbank-Administration nicht sichtbar, wenn sie sich mit DB2 durch Anwendungs- und operative Schnittstellen wie SQL oder commands und Utilities verbinden. Die interne Verarbeitungssteuerung unter den Komponenten ist vollständig transparent für die Außenwelt. Wir können damit einen anderen Ressource 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.



Die neue 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.

Die Zukunft

z/OS und der IBM DB2 Analytics Accelerator for z/OS repräsentieren derzeitig die technologische Speerspitze 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 Integration Appliances“ eine wichtige Rolle. Sie sind ebenfalls eng in die zBX Infraruktur 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/OS-Betriebssystem von anderen Platformen durch seine zahlreichen technologisch führenden Eigenschaften abhebt.

Andere Hersteller, vor allem Oracle und SAP versuchen andersartige Ansätze. Siehe z. B. http://www.wissenstransfertag-mrn.de/download/IT-Vortrag-DrCzekalla.pdf, Mirror at http://jedi.informatik.uni-leipzig.de/de/VorlesMirror/ii/Vorles/SAP01.pdf