

SYBASE®

Sybase IQ

Schneller und sicherer Zugriff auch bei exponentiellem Datenwachstum

Macht es mir Linux schneller und einfacher?

THE
ENTERPRISE.
UNWIRED.



Kia Khosroschahli

Business Consultant

INNOVATION • INTEGRITÄT • KUNDENORIENTIERUNG

- Gegründet 1984, Hauptsitz in Dublin, Kalifornien, USA
- 3.600 Mitarbeiter in 60 Ländern
- 95 der *Fortune* 100 Firmen setzen Sybase-Technologie ein
- Maßgebliche Durchdringung in wichtigen vertikalen Märkten, u. a. Finanzsektor, Telekommunikation, Gesundheitswesen, öffentliche Verwaltung, Homeland Security
- Umsatz 2003: US\$778.1 Millionen
- Barguthaben von mehr als US\$550 Millionen
- 23 Quartale mit pro forma Profitabilität
- Erweiterung strategischer Partnerschaften mit: SAP, Sun, HP, Intel, NEC



■ Mobile & Wireless

- Nr. 1: Mobile Middleware
- Nr. 1: Mobile Datenbank – 73% Marktanteil
- Nr. 1: Wireless Portal (MyAvantGo)

■ Integration

- Nr. 1: größtes B2B Portal in den U.S.A.
- Nr. 1: Lieferant für HIPAA Compliance
- Nr. 1: Online Banking Lösung für Tier 1 Banken

■ Datenbanken

- Nr. 1: Daten Modellierung
- Nr. 1: Produktionsdatenbank an der Wall Street, mehr als die Hälfte aller Transaktionen läuft über Sybase
- Nr. 1: Mobile Datenbank – 73% Marktanteil
- Nr. 1: Telekommunikation (42%) und Transport (36%) Datenbank in China

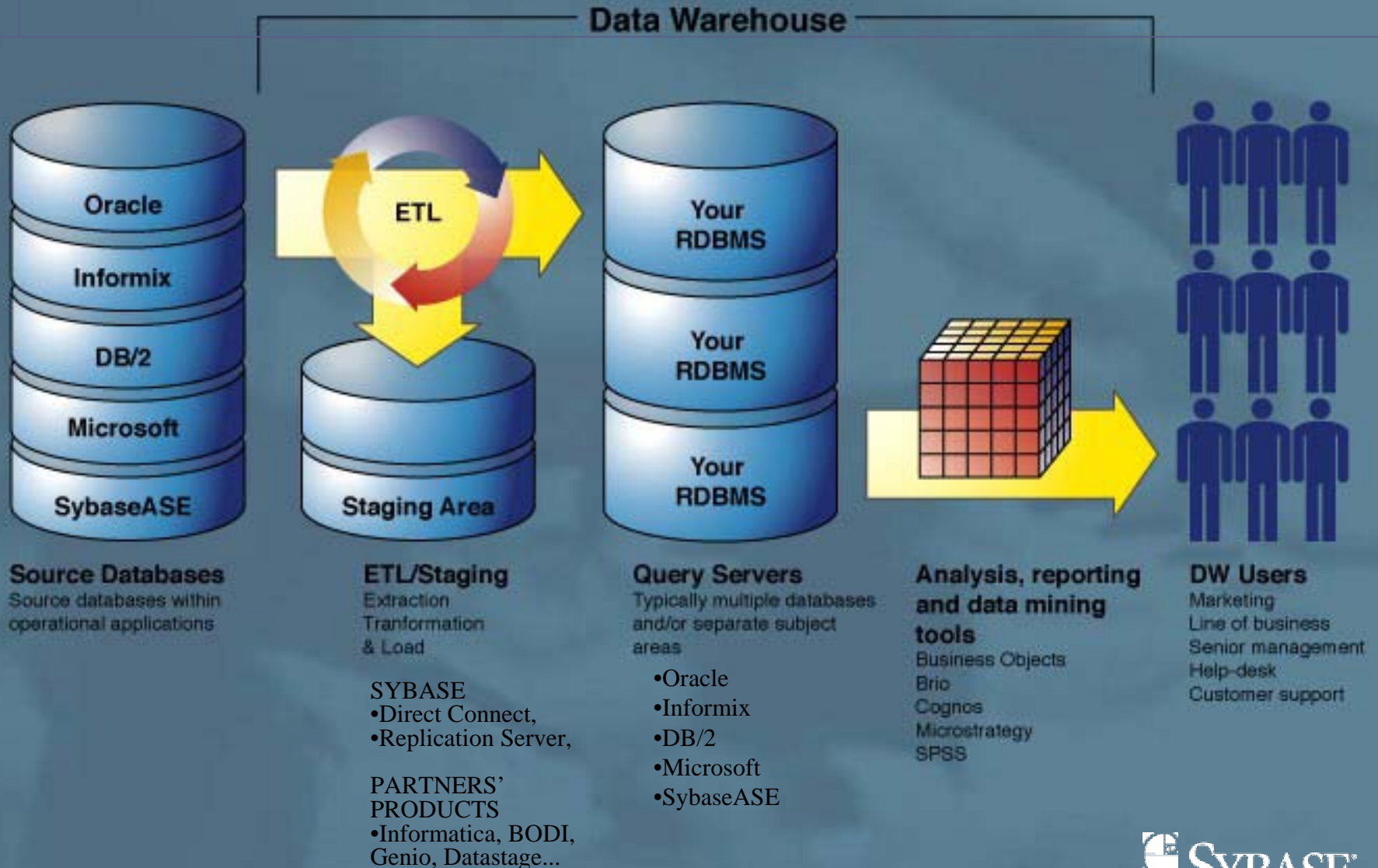
Sybase iAnywhere Solutions ist als das führende Unternehmen in mobiler Middleware Technologie mit höherem Umsatz und Marktanteil als die annähernd 20 Wettbewerber anerkannt.

“[T]he company with the strongest financial and marketing muscle...”

-IDC Report, July, 2003

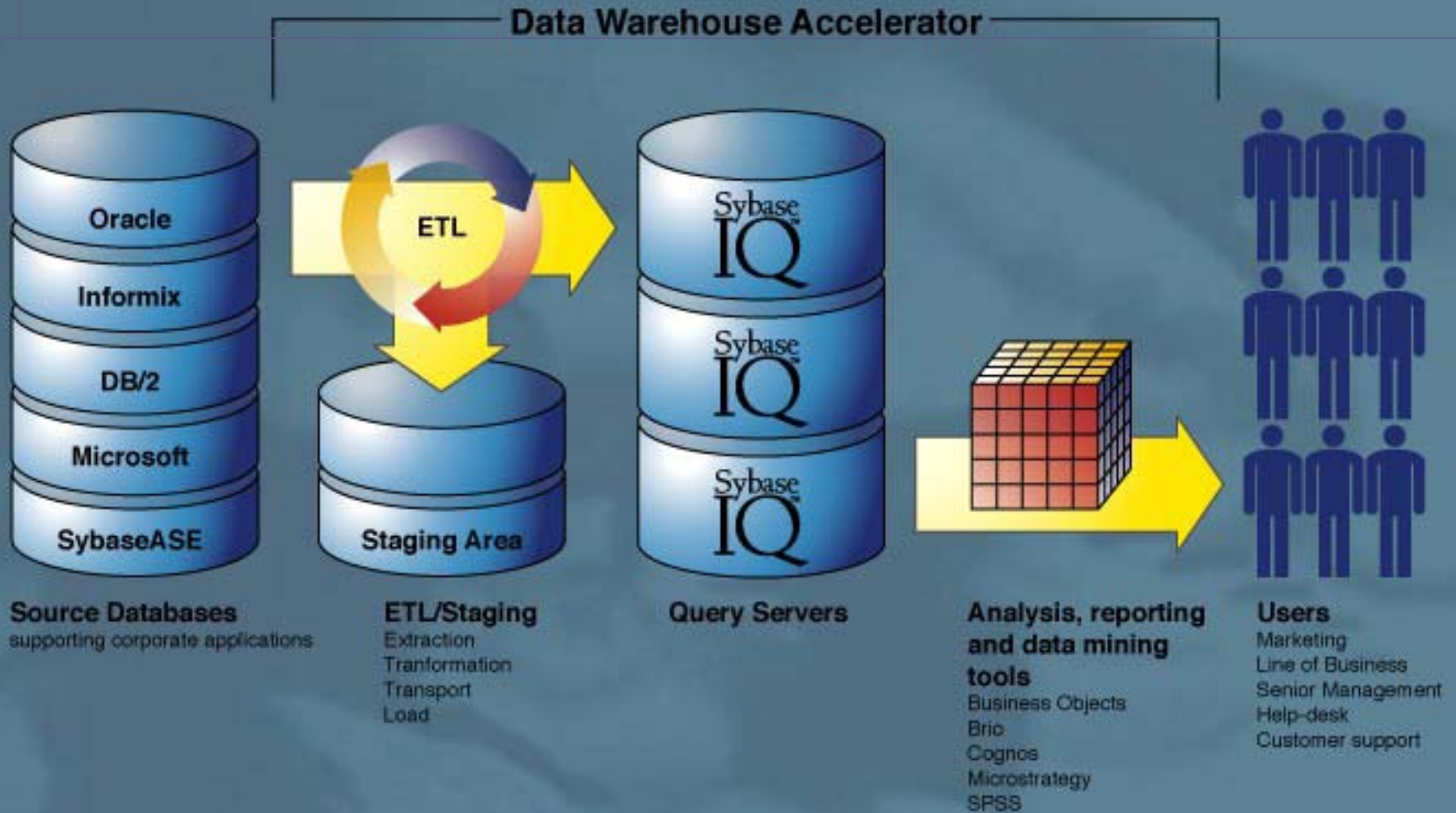


Eine typische Analyse Architektur



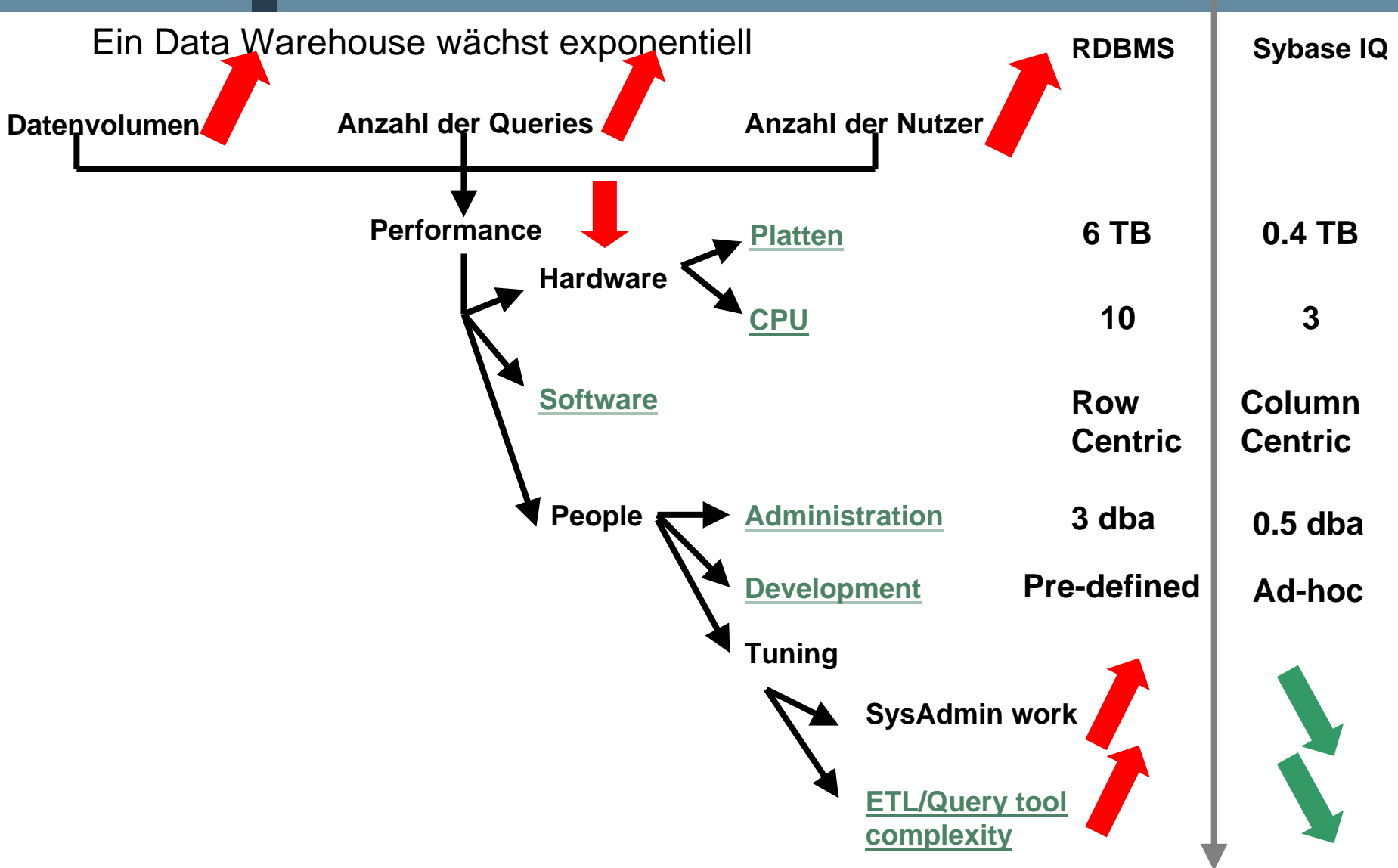
- Infrastruktur ist teuer
- Daten sind nicht aktuell genug
- Schlechte Performance
- Anwender können ihre Abfragen nicht sofort erweitern

Sybase IQ: Spezielle analytische Datenbank



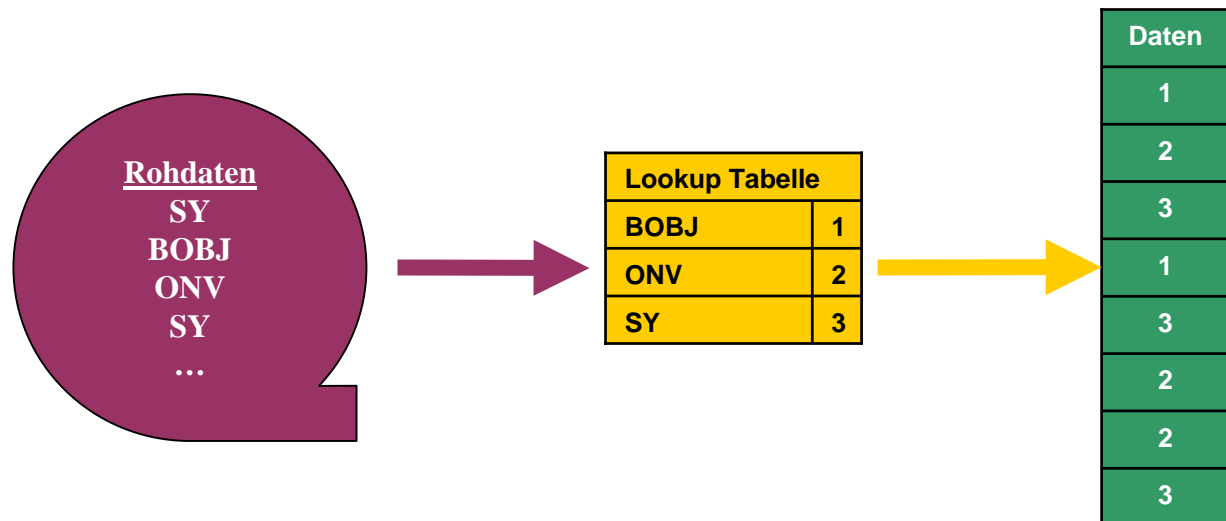
Gegenüberstellung

Ein Data Warehouse wächst exponentiell



- **Schnellste Abfragen, da für Analysen optimiert**
- **Daten werden zu 30% bis 95% komprimiert**
- **Anwender Skalierbarkeit für 100te bis 1000de Anwender**
- **Abfragen Skalierbarkeit für 100te bis 1000de Anwender**
- **Minimaler Administrationsaufwand, da keine Indizes gepflegt werden müssen, weniger Aggregate notwendig sind und jedes Schema möglich ist**
- **Während Laden kann von Tabelle gelesen werden**

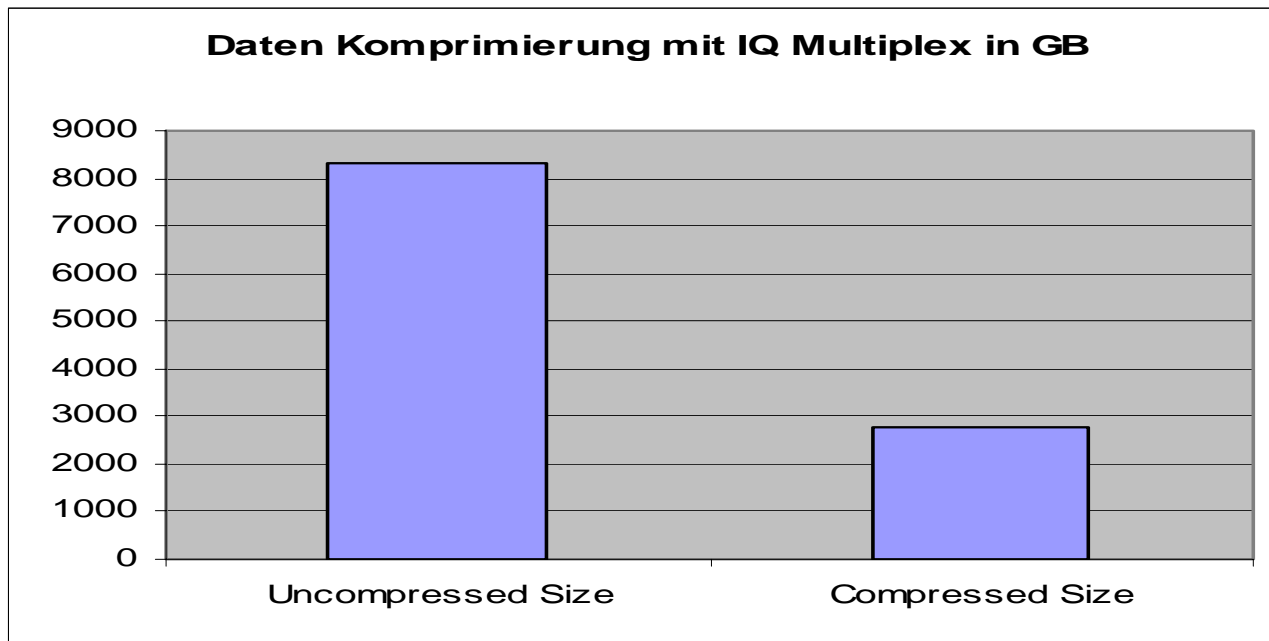
- Unter 65.000 Ausprägungen in einer Spalte wird jeder Wert in einer Lookup Tabelle nur einmal gespeichert
- Bei über 65.000 Ausprägungen werden Rohdaten komprimiert abgespeichert
- Die Lookup Tabelle arbeitet mit einem oder zwei Bytes für Spalten < 256 Ausprägungen oder < 65.000 Ausprägungen



- Nur relevante Daten werden gelesen
- Bei Wertebereich < 65.000 werden Werte nur einmal gespeichert
- JEDE Spalte ist indiziert
- Indizes sind für Analysen optimiert
- Bitmuster sind komprimiert
- IO und Cache wird um 90% reduziert

Date	Store	State	Class	Sales
[Hatched]	[Hatched]	1	1	1
		0	0	0
		0	0	0
		0	0	0
		0	0	0
		1	1	1
		0	0	0
		1	1	1
		0	0	0
		0	0	0
		1	1	1
		0	0	0
		1	1	1
		0	0	0
		0	0	0

- **Alle Bitmuster sind komprimiert**
- **Daten nehmen üblicherweise 0,5 – 1,0 fachen Platz der Rohdaten ein. Bei voller Indizierung!**



Zehn Indextypen stehen derzeit zur Auswahl

- Fast Projection Alle Spalten
- Low Fast: Kardinalität ≤ 1.000
- High Group: Kardinalität > 1.000)
- High Non Group: Dimensions- und Faktenspalten
- Compare: Redundante Spalten in einer Tabelle
- Join: Fremdschlüssel, künstl. Prim Key
- Word: Textspalten
- Date/Time/Datetime Datumsspalten

Mehrere Indizes pro Spalte sind möglich und üblich!

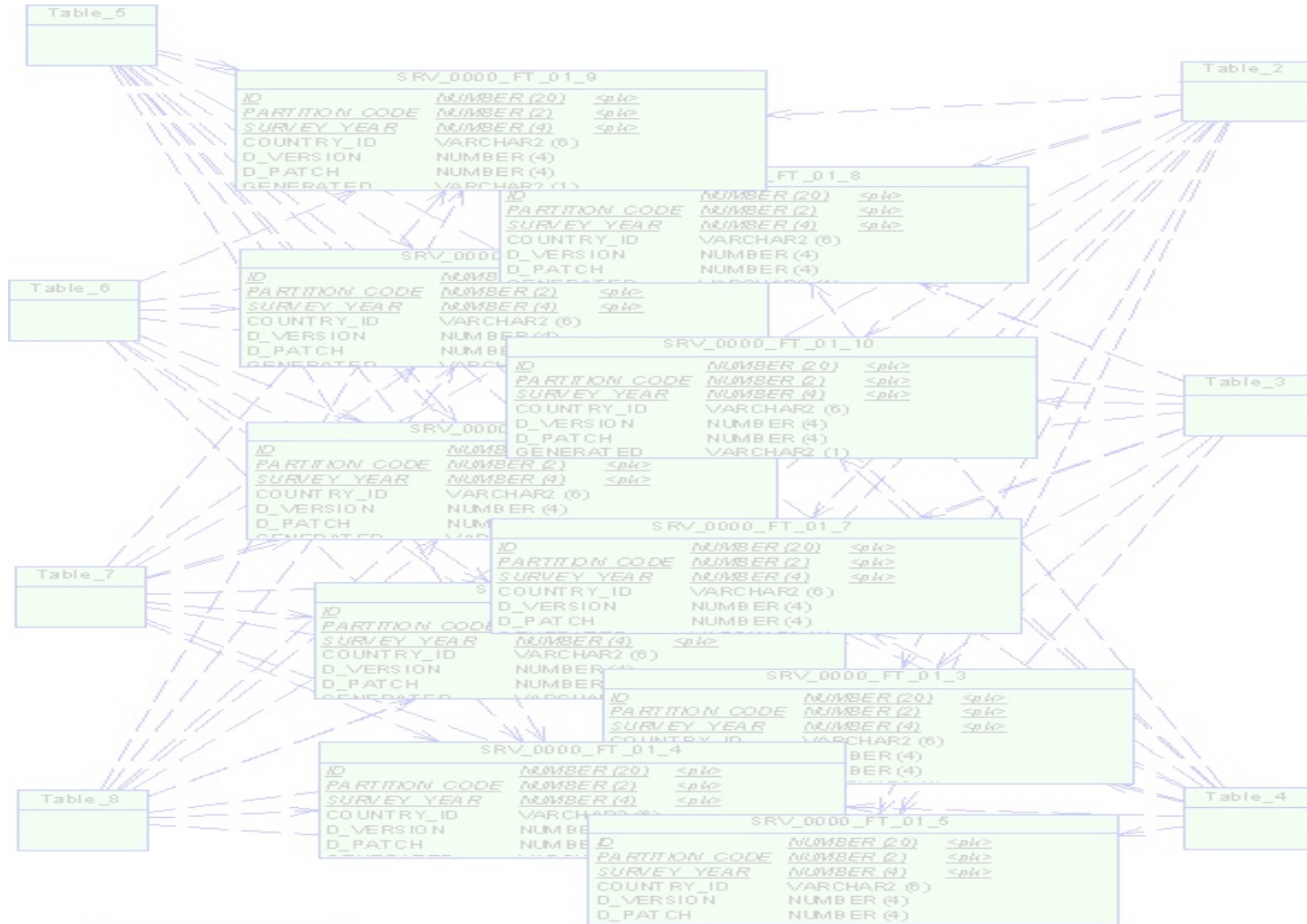
- Tabellen Partitionierung (horizontale Partitionierung)
 - By Year, By Month, By Week...oder anders
- Tabellen Splitting (vertikale Partitionierung)
 - erzeugen von mehreren Tabellen mit einer “one-one” Beziehung
- Summen/Aggregate/vorberechnete Tabellen

Erhöht DB Schema Komplexität

→ erhöht ETL Komplexität

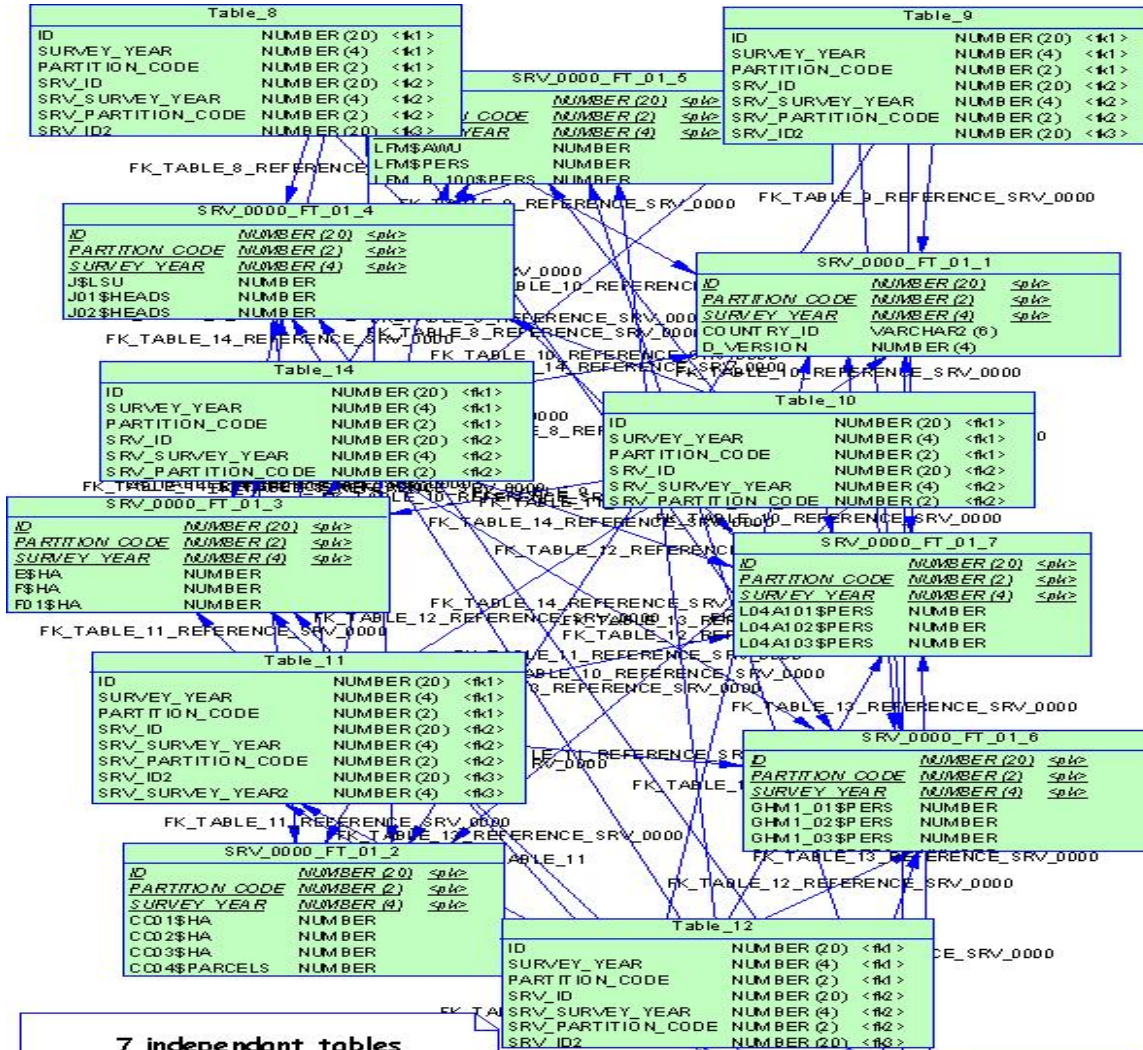
→ Erhöht Query Tool Komplexität

→→→→ Erhöht DWH Kosten !!

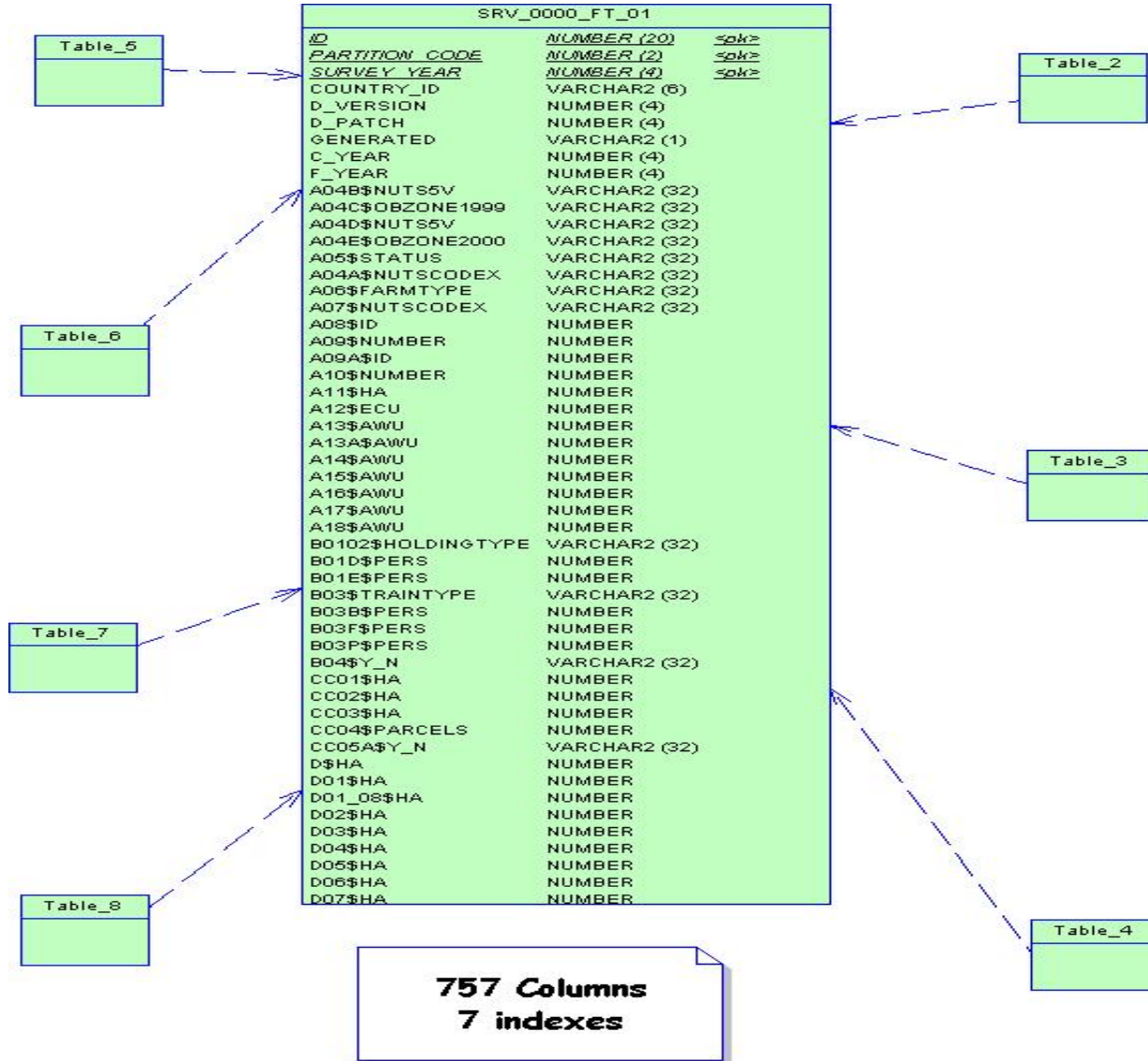


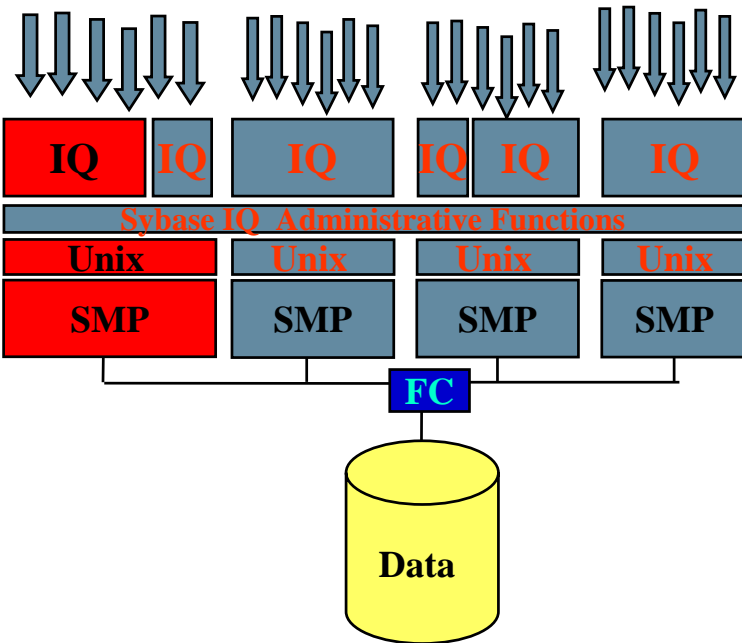
10 partitions
757 Columns
7 indexes

Vertikale Partitionierung



7 independant tables
Need to define links with dimensions
How to define the partitioning, as query need all columns ?





Unterstützte Betriebssysteme:

- Sun Solaris (64-bit)
- **Linux** (32 Bit/64-bit) Redhat, Suse, Redflag
- IBM RISC System/6000 AIX (32-bit/64-bit)
- HP-UX (64-bit)
- HP Itanium (64-bit)
- Windows NT /Windows 2000 /Windows 2003

- Niedrige Kosten
- Mehr Performance durch Skalierbarkeit
- Alle relevanten Datenbanksicherheitskonzepte
- Sicherheit durch Linux und Datenbanksicherheitsstufen
- Mehrere Maschinen parallel (Multiplex)
- Benutzer Skalierung
- Abfrageskalierung
- Einfachste Migration bereits bestehender Systeme

AT&T	26.269	94.305	Daytona	SMP	AT&T	Sun	Sun
Amazon.com	13.001	34.219					
France Telecom	29.232	29.735					
Health Insurance Review Agency	11.942	29.299	IQ	MPP/Cluster	Sybase	HP	Hitachi
Barclays Bank	6.408	24.756	Teradata	MPP/Cluster	Teradata	NCR	LSI
FedEx Services	9.981	14.745	Teradata	MPP/Cluster	Teradata	NCR	EMC
Samsung Card.	7.684	14.567	IQ	MPP/Cluster	Sybase	HP	HP
Kmart	12.592	13.874	Teradata	MPP/Cluster	Teradata	NCR	LSI
Cho-Hung Bank	3.361	12.350	IQ	MPP/Cluster	Sybase	Sun	Hitachi
LG Card	6.336	12.313	IQ	MPP/Cluster	Sybase	Sun	EMC

Tabelle: Auszug !

(Exploded) DB Size, All Environments & UNIX							
Company	Input data(GB)	DB size	DBMS	Syst Arch.	DBMS vendor	server	storage
France Telecom	29.735	29.232		SMP			
AT&T	94.305	26.269	Daytona	SMP	AT&T	Sun	Sun
SBC	NULL	24.805	Teradata	MPP/Cluster	Teradata	NCR	LSI
Anonymous	8.591	16.191	DB2	MPP/Cluster	IBM	IBM	IBM
Amazon.com	34.219	13.001	Oracle	SMP	Oracle	HP	HP
Kmart	13.874	12.592	Teradata	MPP/Cluster	Teradata	NCR	LSI
Claria Corporation	4.361	12.100	Oracle	SMP	Oracle	Sun	Hitachi
Health Insurance Review Agency	29.299	11.942	IQ	MPP/Cluster	Sybase	HP	Hitachi
FedEx Services	14.745	9.981	Teradata	MPP/Cluster	Teradata	NCR	EMC
Vodafone D2 GmbH	NULL	9.108	Teradata	MPP/Cluster	Teradata	NCR	LSI

DB ROWS All Environments & UNIX							
Company	rows (M)	DBMS	Syst Arch.	DBMS vendor	server	storage	
AT&T	496.041	Daytona	SMP	AT&T	Sun	Sun	
Nielsen Media Research	250.918	IQ	MPP/Cluster	Sybase	Sun	Sun	
France Telecom	156.788	Oracle	SMP	Oracle	HP	HP	
SBC	144.417	Teradata	MPP/Cluster	Teradata	NCR	LSI	
Kmart	133.079	Teradata	MPP/Cluster	Teradata	NCR	LSI	
FedEx Services	89.343	Teradata	MPP/Cluster	Teradata	NCR	EMC	
Kroger	77.837	DB2	MPP/Cluster	IBM	IBM	IBM	
LG Card	74.703	IQ	SMP	Sybase	Sun	EMC	
Health Insurance Review Agency	72.093	IQ	MPP/Cluster	Sybase	HP	Hitachi	
AT&T Wireless Services	64.534	Teradata	MPP/Cluster	Teradata	NCR	NCR	

DB ROWS Windows							
Company	input size GB	rows (M)	DBMS	Syst Arch.	DBMS vendor	server	storage
ComScore Networks Inc.	14.464	61.361	IQ	MPP/Cluster	Sybase	HP	EMC
ComScore Networks Inc.	19.604	44.195	IQ	MPP/Cluster	Sybase	Dell	EMC
ComScore Networks Inc.	14.636	40.523	IQ	MPP/Cluster	Sybase	Dell	EMC
Ordina		6.402	SQL Server	SMP	Microsoft	IBM	IBM
Arclight Systems LLC		6.246	SQL Server	SMP	Microsoft	IBM	EMC
Edgars Consolidated Stores		5.543	SQL Server	SMP	Microsoft	IBM	IBM
Hewlett Packard Company		3.704	Oracle	SMP	Oracle	HP	HP
Gomez		3.418	SQL Server	SMP	Microsoft	HP	other
DataQuick		3.273	SQL Server	SMP	Microsoft	Dell	EMC
NAREX Inc		3.228	SQL Server	SMP	Microsoft	Dell	Dell

Top 10	
DBMS	# spots
IQ	14
Teradata	11
Oracle	10
AT&T	3
DB2	2
Microsoft	10
Total =>	50

- SUN-Sybase DW Referenz Architektur in Menlo Park (CA)
- Das größte Datawarehouse läuft auf Sybase IQ mit SUN
- 1 Billion Zeilen !!!
- 155 TB Inputdaten
- Kompression der Inputdaten v. **155 TB** auf **55 TB**

- Vodafone D2 GmbH
- 1&1 Internet AG
- Bertelsmann Music Group
- EMI Electrola
- RTL Television
- Allianz-Dresdner Bausparkasse
- Dresdner Bank
- Deutsche Bank
- Citibank
- DEVK Allgemeine Versicherungen AG
- Risk Consulting
- Raiffeisen Hauptgenossenschaft Nord
- Müller (Drogeriemärkte)
- European Southern Observatory
- DekaBank

Sybase IQ ist eine Datenbank, die speziell für DSS entwickelt wurde:

- Performante Abfragen (10X – 1000X schneller als RDBMS)
- Komprimierte Spaltenweise Speicherung der Daten
 - Plattensparnis (1:3 – 1:15)
 - Weniger CPUs
 - Kein teures IO
- Erstellen der Indizes zur Designzeit
 - geringer administrativer Aufwand
 - Performante Ad-hoc-Queries
- Reduktion der Komplexität v. ETL Prozessen
 - geringer administrativer Aufwand
- Parallele Loads (auch online)
- Beste Analyse und Performance Ergebnisse auch auf Linux

- Taten sagen mehr als Worte, lassen Sie sich überzeugen:
 - Mit Ihren Daten
 - Bei Ihnen
 - Auf Ihrer Hardware
 - In 2 bis 3 Tagen
 - Kostenlos!

Wir belegen unsere Behauptungen, gemeinsam mit Ihnen.
Ohne Verpflichtung Ihrerseits.



Kia Khosroschahli
Business Consultant
Sybase GmbH
kkhosros@sybase.com