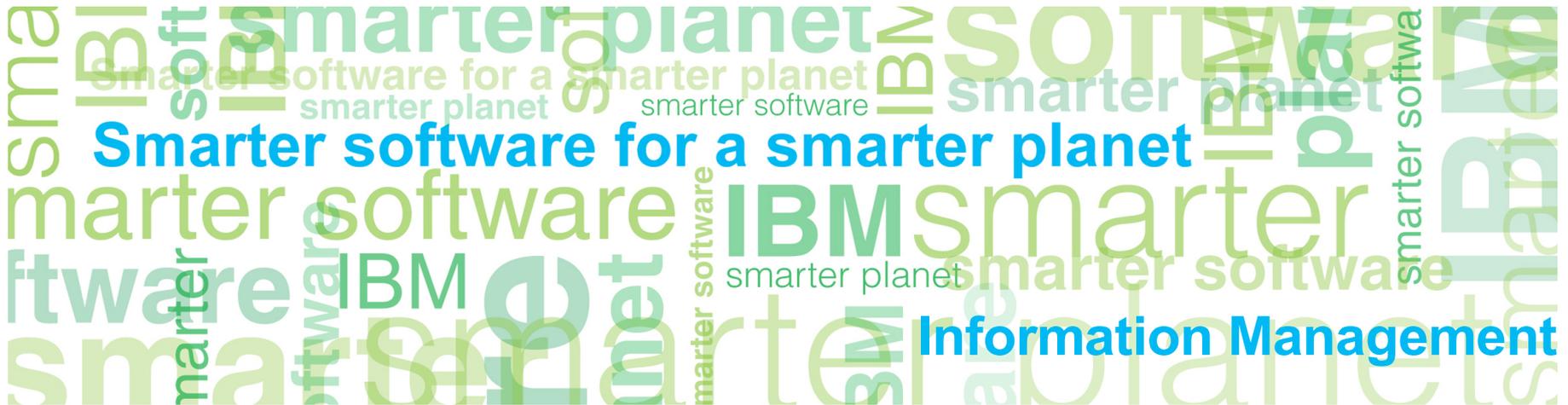




Bahnbrechende In-Memory-Technologie sorgt für Geschwindigkeitsrekorde beim Data Warehousing

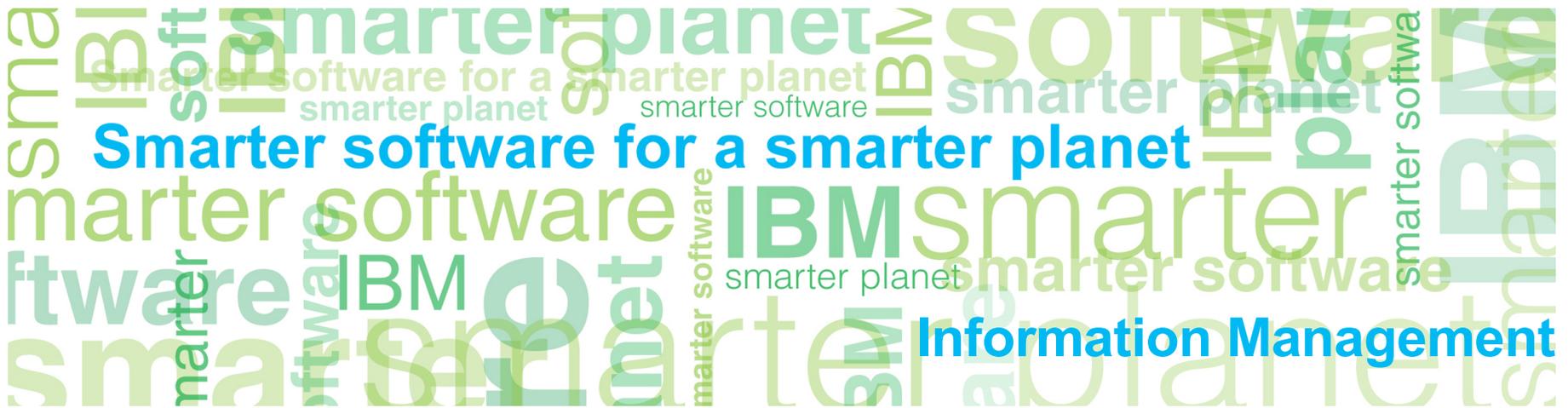


Agenda

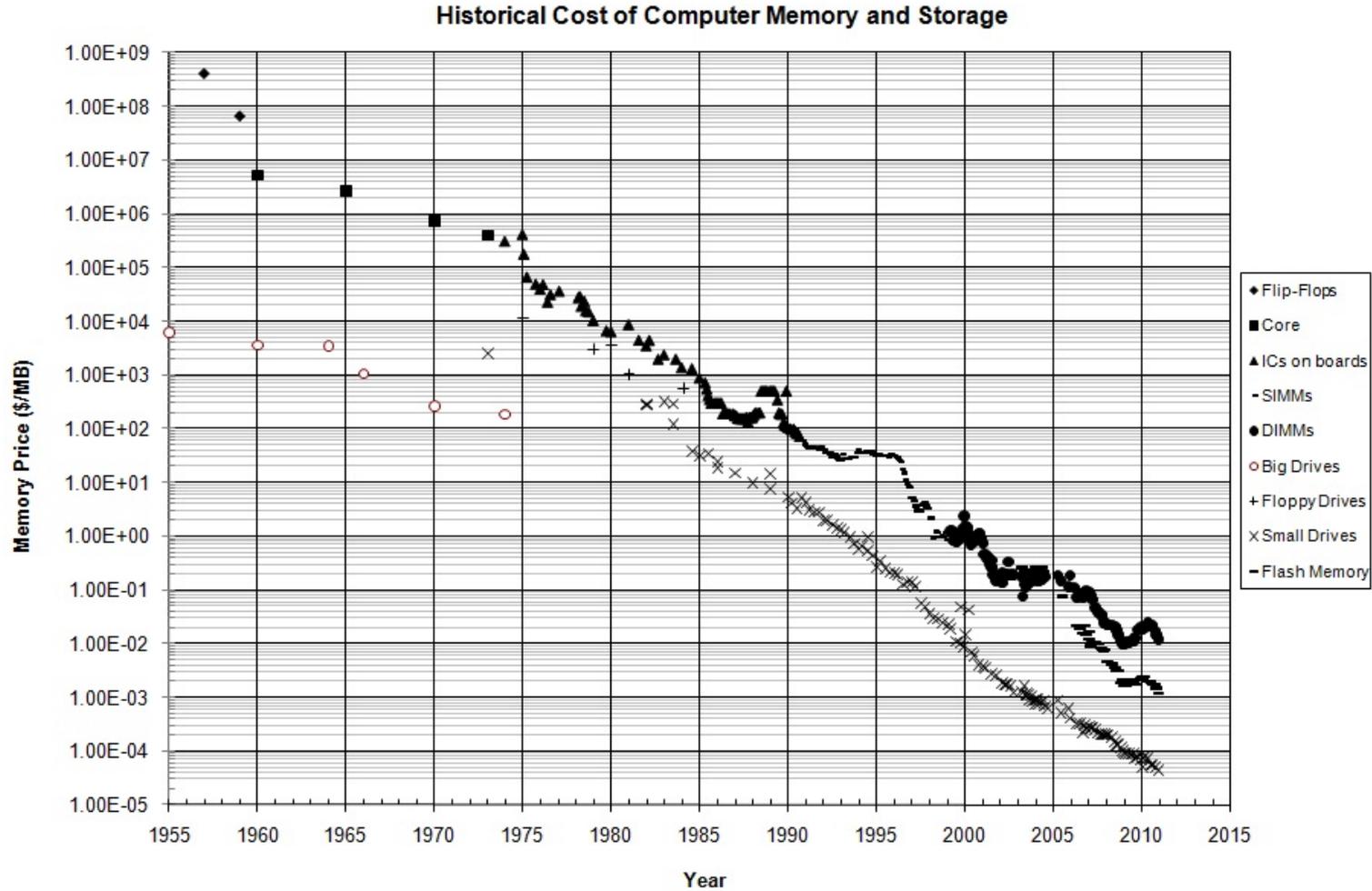
- Markttrends
- Was ist IWA?
- Plattformverfügbarkeit und Packaging für IUWE und IWA
- IWA im Detail
- IWA Performance
- Wann ist IWA die perfekte Lösung?
- Fragen für Kunden
- Zusammenfassung



Markttrends



Preistrends bei Speichermedien



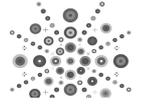
George C. McCullum <http://www.jcmit.com/mem2010.htm>

Hardware Trends

- Preis in \$/MB ist jetzt (2011) für Hauptspeicher so hoch wie für Festplatten vor 10 Jahren
- Random Zugriffe:
 - Die durchschnittliche Zugriffszeit für Festplatten hat sich in den letzten 20 Jahren nur relativ wenig verbessert.
 - Zugriff auf Hauptspeicher hat sich mehr verbessert
 - Lücke bei Zugriffszeiten von Hauptspeicher und Platten hat sich damit vergrößert
- Performance-Lücke zwischen Hauptspeicher und Cache
- Lokalität der Hauptspeicherzugriffe deshalb wichtig
- Erhöhter Parallelitätsgrad bei CPUs:
 - mehr SIMD Parallelismus
 - mehr Cores

Die dritte Generation der Datenbanktechnologie

DATABASE



In einem IDC Artikel vom Feb 2010 schreibt Carl Olofson:

1. Generation DBMS:

- Proprietäre DBMS wie IMS, IDMS, Datacom

2. Generation DBMS:

- RDBMS für offene Systeme hängen von Disk Layout ab, Begrenzungen bzgl. Skalierbarkeit und Disk I/O

3. Generation DBMS: IDC sagt voraus, dass innerhalb von 5 Jahren:

- die meisten Data Warehouses einen Column Store verwenden
- die meisten OLTP Datenbank entweder durch eine In-Memory-Datenbank ergänzt werden, oder selbst vollständig im Hauptspeicher liegen
- die meisten Datenbank Server für große Systeme horizontale Skalierbarkeit durch Clustering erreichen werden

IWA adressiert alle für ein DBMS der 3. Generation genannten Punkte!

Marktdaten: Wieso Änderungen im DWH?

What problems will eventually drive you to replace your current primary data warehouse platform?
(Select nine or fewer.)

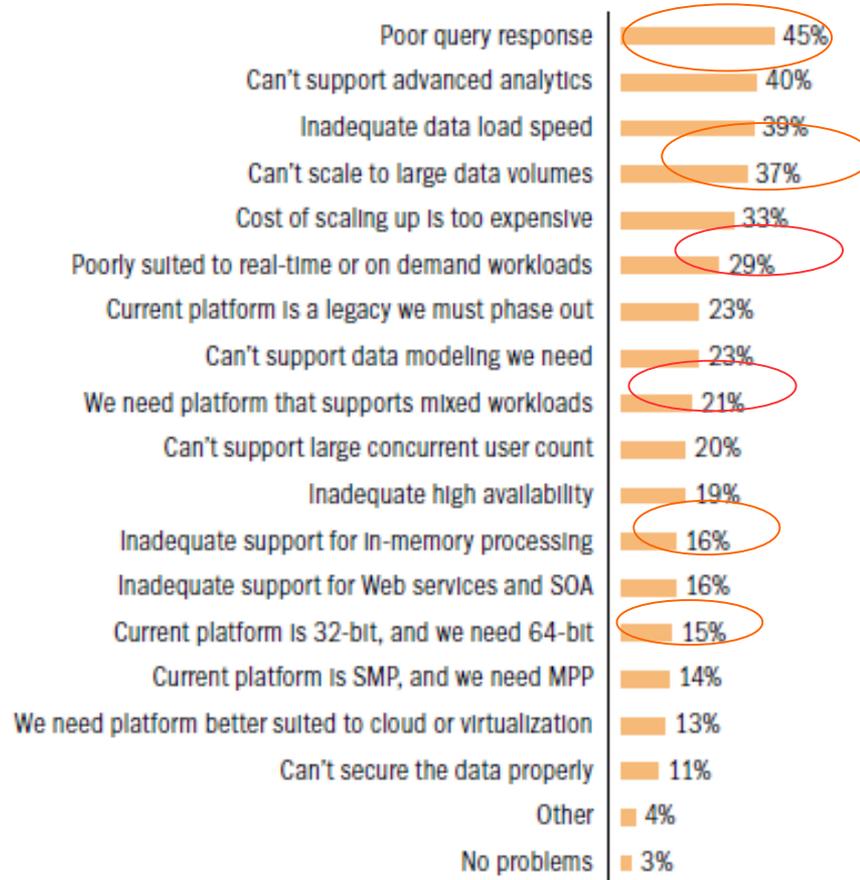


Figure 3. Based on 1,752 responses from 417 respondents; 4.2 responses per respondent on average.

Informix Warehouse Accelerator

die 3. Generation der Datenbanktechnologie ist schon jetzt verfügbar

Was steckt dahinter?

Der Informix Warehouse Accelerator (IWA) ist ein workload-optimiertes, appliance-ähnliches, Add-On, das die Integration von analytischen Auswertungen in operative System ermöglicht. Er beschleunigt Queries zu bisher noch nie dagewesenen Antwortzeiten.



Wieso ist das etwas völlig Neues?

- **Performance:** Die um viele Faktoren verbesserte Performance ermöglicht Analysen, die bisher noch nicht vorstellbar waren.
- **Integration:** Tiefe Integration in Informix so dass für Anwendungen völlige Transparenz besteht.
- **Selbstverwaltung der Last:** Queries werden ohne Tuning-Maßnahmen auf die effizienteste Art ausgeführt
- **Transparenz:** Anwendungen konnektieren sich zu Informix und müssen nichts über IWA wissen
- **Vereinfachte Administration:** appliance-ähnliche Bedienung

**Ein technologischer Durchbruch,
der grundlegende neue Möglichkeiten eröffnet**

Technologiedurchbrüche für extreme Performance

Extreme Kompression

Erforderlich, da RAM der begrenzende Faktor ist.

kombinierte Row & Columnar Datenbank

zeilenorientierte Speicherung in IDS für transaktionsorientierte Workloads und spaltenweise Speicherung im Accelerator für OLAP Queries.

Für multi-Core und Vektorverarbeitung optimierte Algorithmen

Vermeidung von Locking und unnötiger Synchronisation

In Memory Datenbank

Datenbanktechnologie der 3. Generation vermeidet IO. Kompression und erlaubt es grosse Datenbanken vollständig im Memory zu halten

Prädikatauswertung auf komprimierten Daten

Häufig Auswertung der Daten in Scans ohne Dekomprimierung möglich

Häufigkeitspartitionierung

ermöglicht den parallelen Scan der partitionierten Daten sowie horizontale und vertikale Partitionseliminierung.

Massive Parallelität

Alle verfügbaren Cores werden für die Queries benutzt

Vieles ist nicht mehr notwendig:



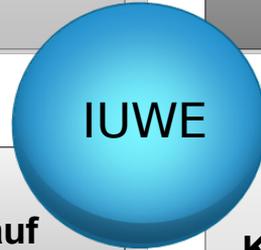
Was sind die Hauptunterschiede zu anderen Produkten?

Hybride spaltenweise Speicherung

Daten werden spaltenweise gespeichert und zugegriffen

In-Memory

Da immer die gesamte Datenmenge im Hauptspeicher gehalten wird, wird IO vermieden und es können entsprechend spezialisierte Algorithmen eingesetzt werden



IUWE

Prädikatauswertung auf komprimierten Daten

Häufig Auswertung der Daten in Scans ohne Dekomprimierung möglich

Kein Tuning

Keine Notwendigkeit für Indizes oder Datenbank-Tuning

330 mal

900 mal

450 mal

1350 mal

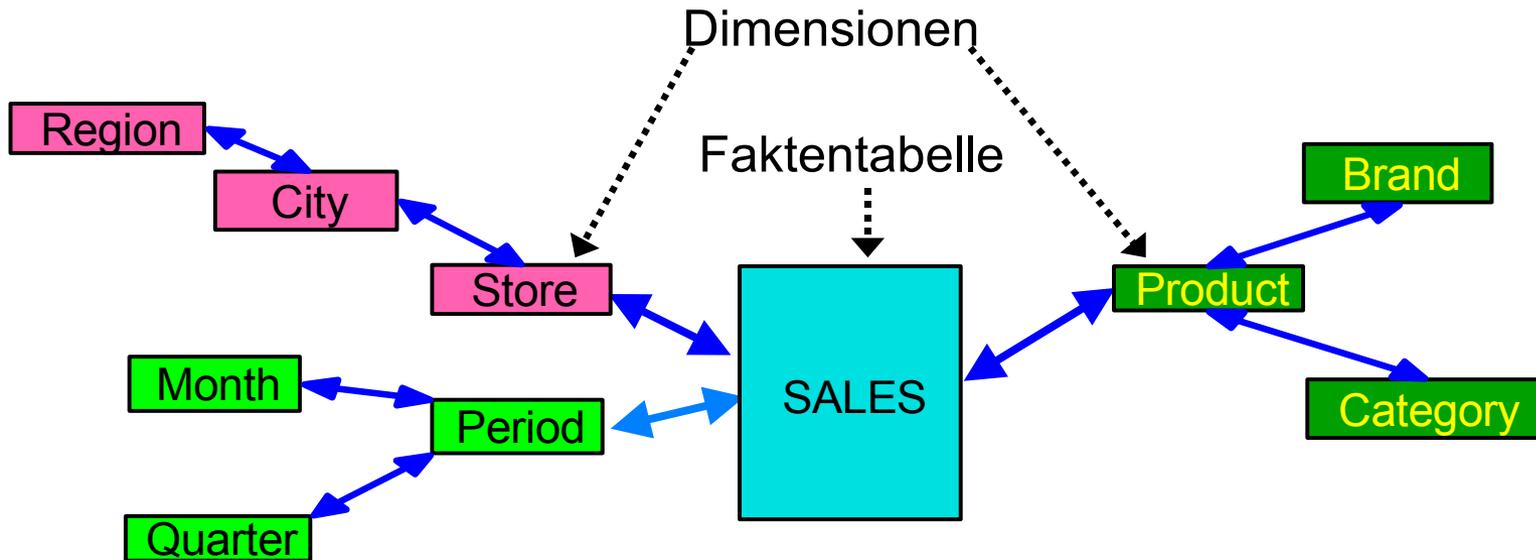
Mehrere Größenordnungen an PERFORMANCE-VERBESSERUNGEN!!

IWA - Charakteristika

- Ein dediziertes System (Linux auf Intel x86_64)
- Keine Anwendungsänderungen
 - Applikationen verbinden sich weiter an Informix
 - Wenn Queries auf dem Accelerator ausgeführt werden können, erledigt Informix das transparent für die Anwendung
 - Informix selbst ist unabhängig von möglichen Ausfällen des Accelerator (Verfügbarkeit!)
- Performance Verbesserung um Größenordnungen
- Reduktion des in Informix notwendigen Tunings
- Appliance-ähnliches Verhalten
 - so gut wie kein spezielles Know-How für Betrieb des Accelerators notwendig
- Deutlich verbessertes PreisLeistungsverhältnis und der TCO als ein kombinierter Effekt von:
 - Beschleunigung von komplexen analytischen Queries
 - Performance-Verbesserungen um mehrere Größenordnungen bei beschleunigten Queries
 - Reduzierter DBA Aufwand für das Tuning von beschleunigten Queries

Zielmarkt: Business Intelligence (BI)

- Charakterisiert durch:
 - “Star” oder “snowflake” Schema:



Komplexe, ad hoc Queries, die typischerweise

- Trends und Ausnahmen suchen um daraus Geschäftsentscheidungen ableiten zu können
- Auf große Teilmengen einer Datenbank in den einzelnen Queries zugreifen (im Gegensatz zu OLTP)
- Gruppierung und Aggregationsfunktionen werden verwendet (wie COUNT, SUM, AVG,...)
- **Das ist das optimale Einsatzgebiet für den IWA!**

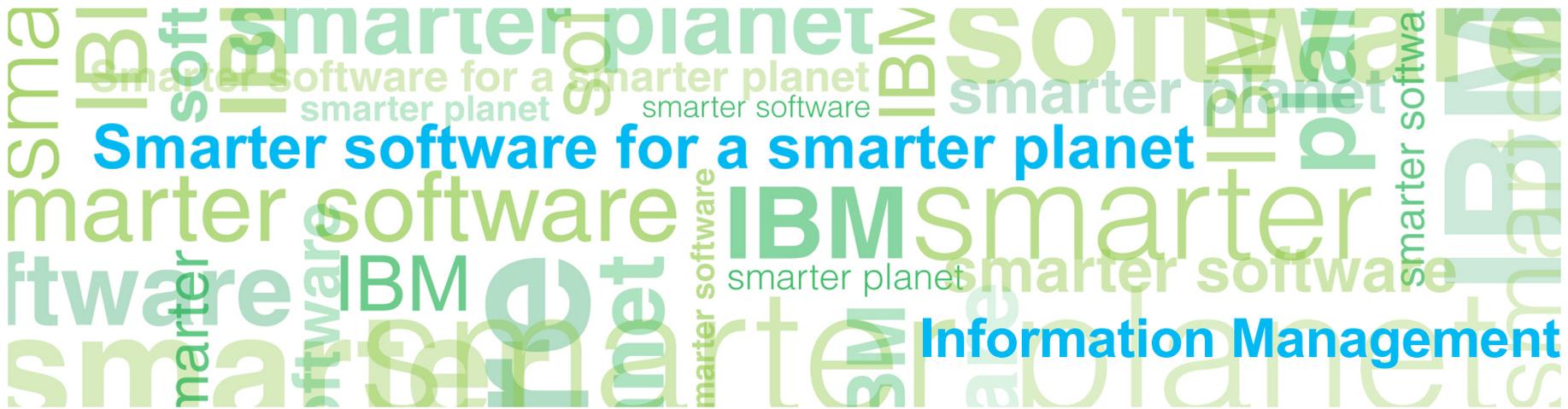
IWA wurde entworfen für:

- Selektive, schnelle Scans über große (Fakten-)tabellen
- Joins mit kleineren Dimensionstabellen
- OLAP-artige Queries über großen Faktentabellen in relationalen Star Schemata die Gruppierung und Aggregation involvieren

```
SELECT PRODUCT_DEPARTMENT, REGION, SUM(REVENUE)
FROM FACT_SALES F
    INNER JOIN DIM_PRODUCT P ON F.FKP = P.PK
    INNER JOIN DIM_REGION R ON F.FKR = R.PK
    LEFT OUTER JOIN DIM_TIME T ON F.FKT = T.PK
WHERE T.YEAR = 2009
    AND R.GEOID = 17
    AND P.TYPEID = 3
GROUP BY PRODUCT_DEPARTMENT, REGION
```



Plattformverfügbarkeit und Packaging für IUWE und IWA



Packaging und Verfügbarkeit

- Informix Ultimate Warehouse Edition (IUWE) besteht aus:
 - Informix Ultimate Edition
 - Storage Optimization Feature
 - Informix Warehouse Accelerator
- als Lizenzmetrik nur Processor Value Units (PVU) verfügbar
- IUWE verfügbar auf
 - AIX 64 Bit
 - Solaris Sparc 64 Bit
 - HPUX Itanium
 - Linux x86 64 Bit
- Aber IWA selbst läuft nur auf x86 64 bit Linux
- Auf einem x86 64 Bit Linux Server können Informix und IWA auf der gleichen Maschine laufen

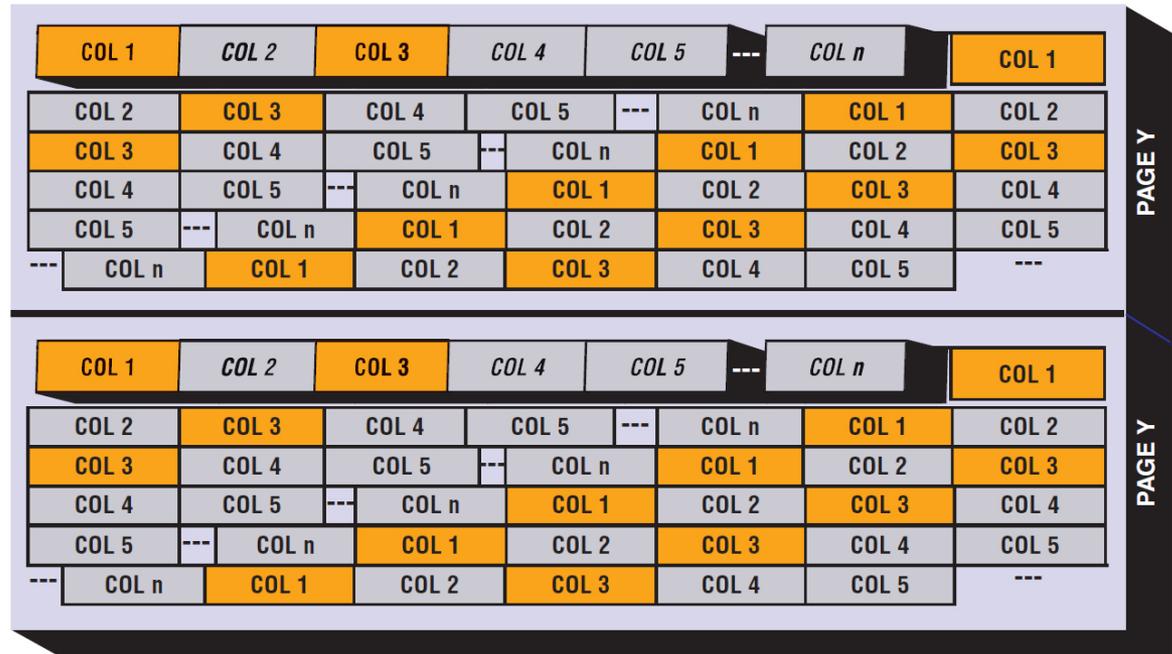
Produktnummern und Preise

- Informix Ultimate Warehouse Edition (IUWE)
 - Informix Ultimate Whse Ed Per Processor Value Unit (PVU) Lic + SW S&S 12 Mo
 - Part Number: D0IJGLL
 - Listenpreis: 438 EUR pro PVU
- Tradeup von Informix Ultimate Edition auf IUWE
 - Part Number: D0IJILL
 - Listenpreis: 152 EUR pro PVU
 - Preis kaum höher als für Storage Optimization Feature alleine
- Weitere Details im Announcement Letter:
 - http://www-01.ibm.com/common/ssi/ShowDoc.jsp?docURL=/common/ssi/rep_ca/8/897/ENUS211-148/index.html&breadCrum=DET001PT022&url=buttonpressed=DET002PT005&specific_in dex=DET001PEF502&DET015PGL002=DET001PEF011&submit.x=7&submit.y=8&lang=en_US

Row Oriented Data Store

Jede Row wird sequentiell gespeichert

- Optimiert für Record I/O
- jedesmal Holen und Dekomprimieren der ganzen Row
- Ergebnis –
 - Sehr effizient für transaktionale Workloads
 - Nicht immer effizient für analytische Workloads



Auch wenn nur ein paar Spalten benötigt werden, wird immer die gesamte Row geholt und entkomprimiert

Column Data Store

Daten werden für jede Spalte sequentiell gespeichert

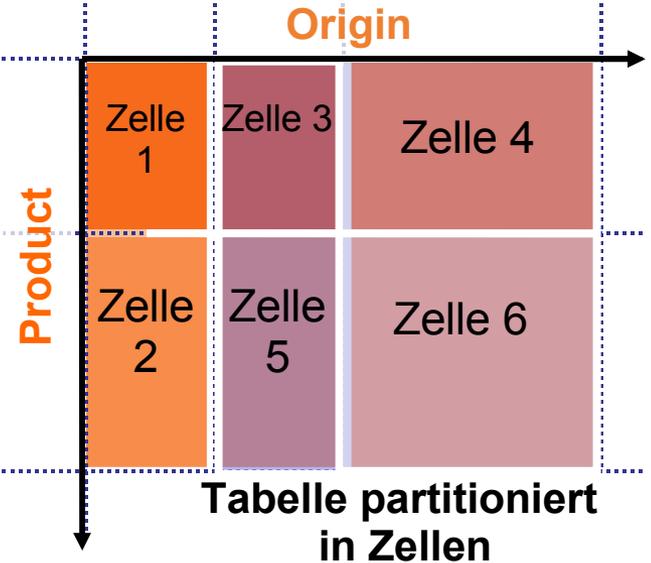
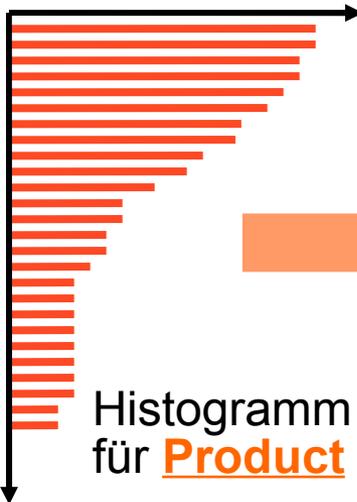
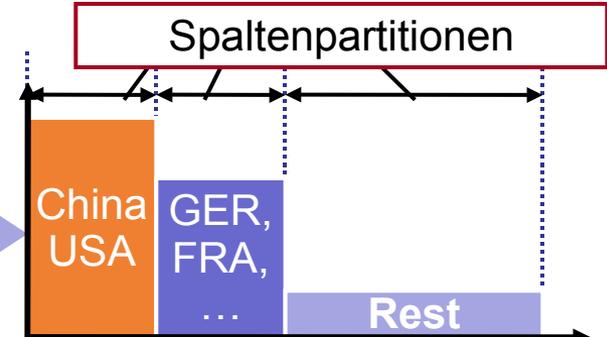
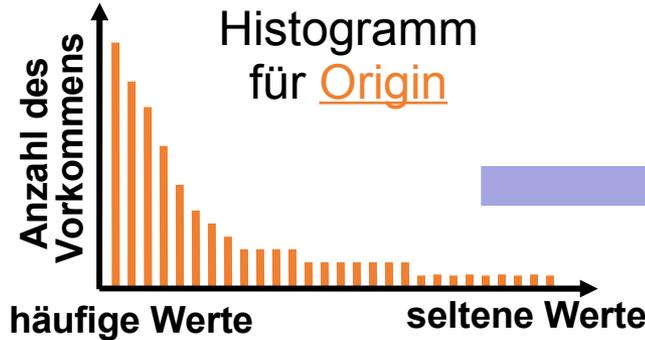
- Die Daten werden sequentiell für jede Spalte komprimiert:
 - Hilft bei sequentiellen Scans
 - Bremst bei random Zugriffen

COL 1					
COL 1	COL 1	COL 1	COL 1	COL 2	COL 2
COL 2					
COL 2	COL 2	COL 3	COL 3	COL 3	COL 3
COL 3					
COL 4					
COL 4	COL 4	COL 4	COL 5	COL 5	COL 5
COL 5					
COL 5	COL 5	---	COL n	COL n	COL n
COL n					
COL n					

Attribute, die für eine Query-Ausführung nicht benötigt werden, werden vollständig übersprungen.

Komprimierung: Häufigkeitspartitionierung

Trade Info (volume, product, origin country)



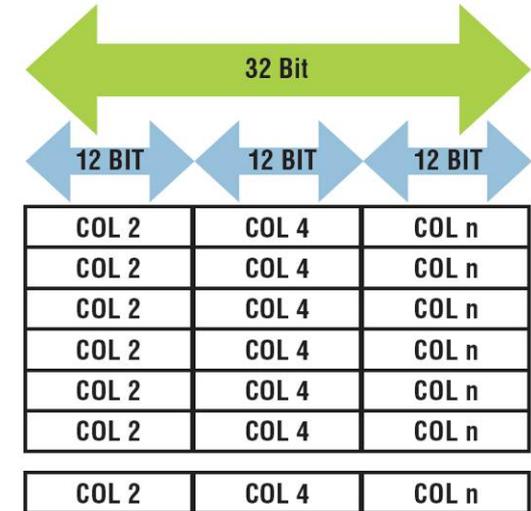
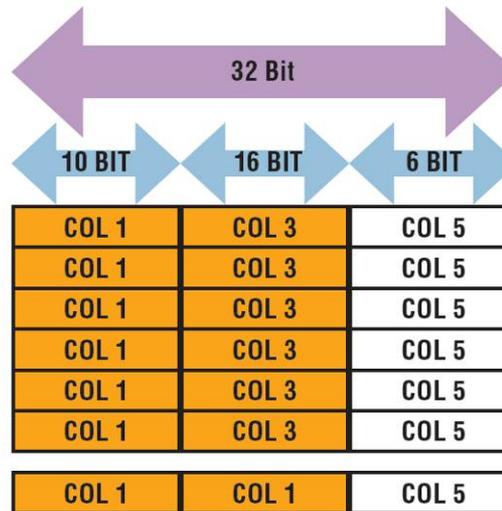
- **Feldlänge unterschiedlich zwischen Zellen**
 - Größere Häufigkeit → kürzere Codes (ähnlich Huffman Kodierung)
- **Feldlänge fest innerhalb von Zellen**

Verarbeitung von Daten in komprimierter Form

- Mehrere Spalten werden zusammengruppiert innerhalb eines Registers gespeichert

- Die Summe der Breiten der komprimierten Spalten darf nicht die Breite eines vorhandenen Registers überschreiten.

- Ausnutzung der vollen Möglichkeiten eines 64 Bit Systems unabhängig davon wieviele Spalten in einem Register sind.



Pädikatsauswertung erfolgt auf komprimierten Daten!

Der **Register-Store** ist eine Optimierung des Column-Store-Ansatzes, wobei versucht wird, die HW bestmöglich zu nutzen.

Gleichzeitige Auswertung von mehreren Prädikaten

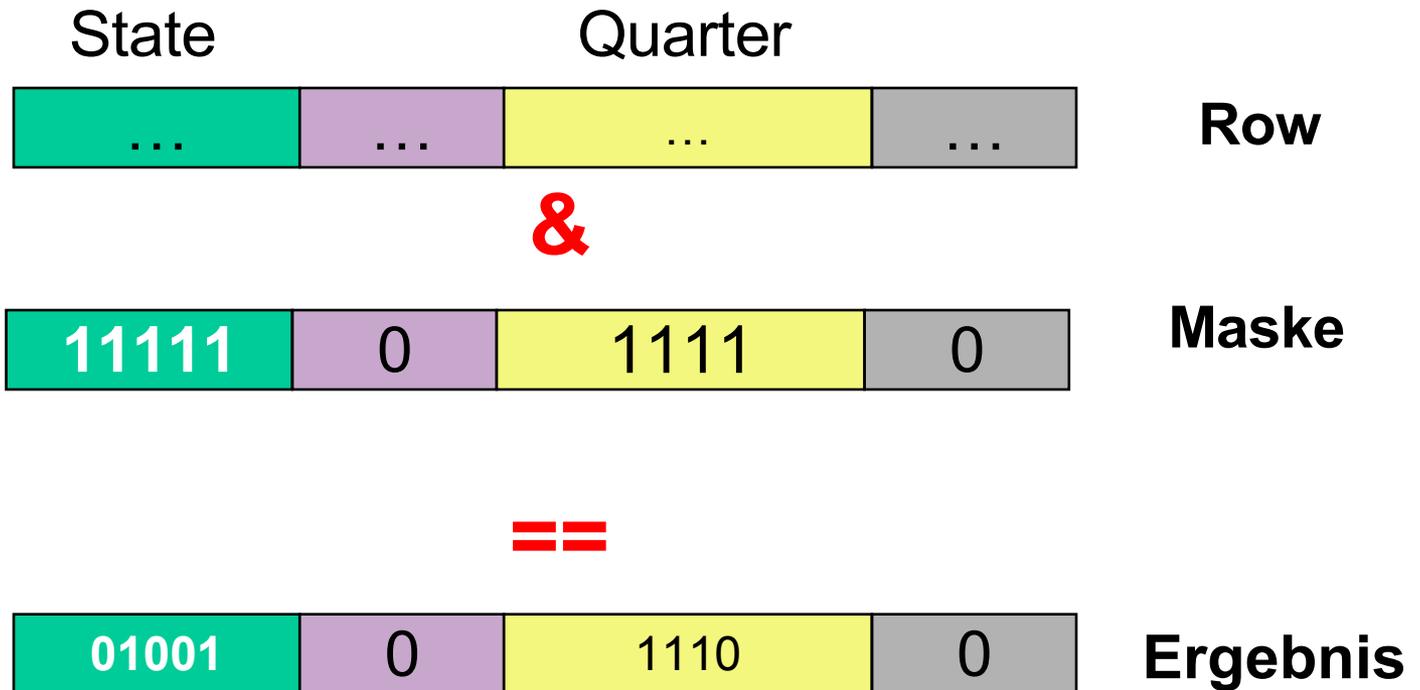
- CPU arbeitet mit 128-Bit Einheiten
 - Viele Felder passen in 128 Bit
- Die Felder fangen bei festen Offsets an
- Anwendung der Prädikate auf alle Spalten gleichzeitig!

State=='CA' && Quarter == 'Q4'



Übersetzung von "Wert-Query"
in "Code-Query"

State==01001 && Quarter==1110



Kunden-Test-Szenarien

Bundesbehörde Deutschland

- 537 Queries ausgeführt auf 30 GB großem Datamart
- 432 Microstrategy Queries wurden mit IWA beschleunigt
- Performance mit IDS+IWA war **90 Mal schneller** als das aktuelle System

Bundesbehörde - USA

- Beispieldaten aus 2 TB Warehouse
- 8 repräsentative Queries liefen **127 Mal schneller**

Großer Schuheinzelhändler

- BI Queries auf 150 GB Daten
- Faktentabelle – 1 Milliarde Rows
- Queries gegen Informix – 22 Minuten
- Queries gegen Informix + IWA : **4 Sekunden, 330 Mal schneller**

Großer globaler Einzelhändler

- 10 GB Beispiel-Datamart
- Faktentabelle – 25 million Rows
- Queries gegen Informix + IWA: **4 Sekunden**
- **5 bis 275 Mal schneller als Teradata**

Beispiel:

Detaillierte Ergebnisse – Großer Schuheinzelhändler

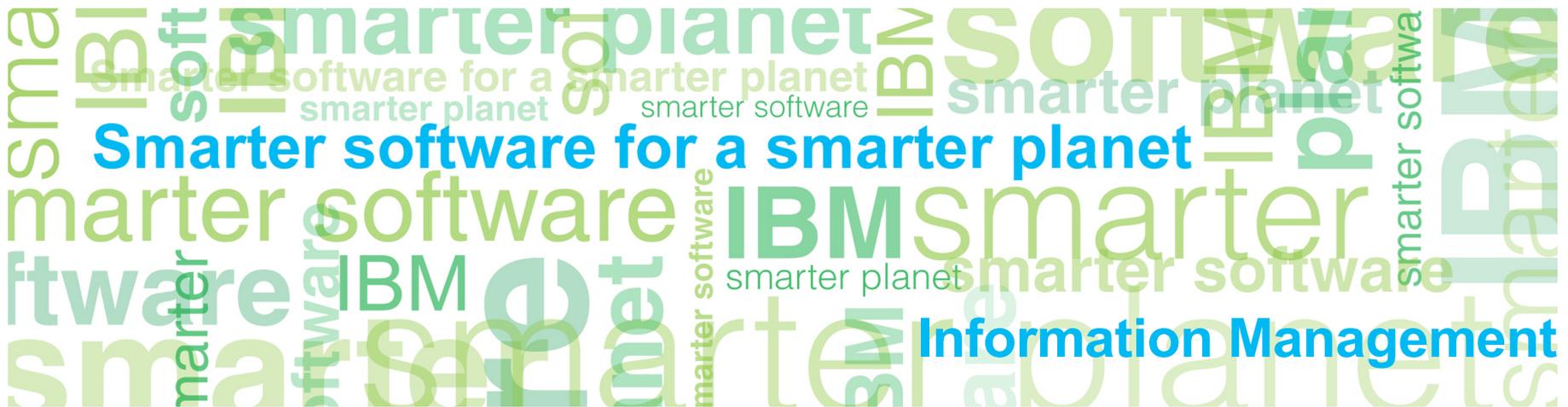
Die Zahlen sagen alles!

Query	Informix [s]	Informix mit IWA [s]	Verbesserung
1	1320	4	330 Mal
2	63	2	31 Mal
3	220	2	110 Mal
4	1800	4	450 Mal
5	120	2	60 Mal
6	1800	2	900 Mal
7	2700	2	1350 Mal

geringe Laufzeitschwankungen
= vorhersehbare Antwortzeiten



Wann ist IWA die perfekte Lösung?



Was sind typische Kundensituationen

Fall 1: Kunde hat Performance Problem mit Fremddatenbank (Oracle, MS SQL Server, Teradata...)

- Performance Probleme bei Queries mit BI Tools wie Cognos, BO oder Microstrategy auf vorhandenem Star oder Snowflake Schema
- Neues IUWE System:
 - Datenaufbereitung zunächst weiter auf Altsystem
 - Tabellen werden 1:1 von Altsystem übernommen
 - regelmäßiges Laden der Tabellen aus Altsystem
 - BI Tools laufen nun auf IUWE System
 - Keine Administration z.B. Backup/Restore auf IUWE
- Vorteile:
 - Gewaltige Performance Verbesserung
 - Kein Migrationsaufwand
 - Kein tiefes Informix Know-How erforderlich

Was sind typische Kundensituationen

Fall 2: IWA als Ergänzung zu bestehendem Informix System

- Kunde nutzt IWA auf bestehender HW (Linux 64bit) oder neuer HW
- Trade-Up auf IUWE oder Neukauf IUWE
- Vorgehen:
 - Tabellen für IWA werden identifiziert und in IWA geladen
 - Data Warehousing/Reporting Queries, ob auf OLTP oder Data Warehouse System, werden automatisch beschleunigt
- Entlastung von Altsystem
- Queries deutlich schneller, damit mehr und neue Fragestellungen beantwortbar

Was sind typische Kundensituationen

Fall 3: Fachabteilung möchte Ihr eigenes Query-System haben

- Fachabteilung nutzt bestehendes von IT administriertes System, hat aber folgende Schmerzen:
 - Query Performance schlecht und unverhersehbar
 - Fachabteilung möchte gerne schnell und flexibel eigene Änderungen vornehmen
- Lösung wie bei Fall 1
- Vorteile:
 - Zentrales Altsystem mit ETL wird weiter von IT gepflegt
 - Fachabteilung kann auf IUWE System schnell eigene Erweiterungen implementieren ohne tiefes IT Know-How haben zu müssen

Fragen an den Kunden

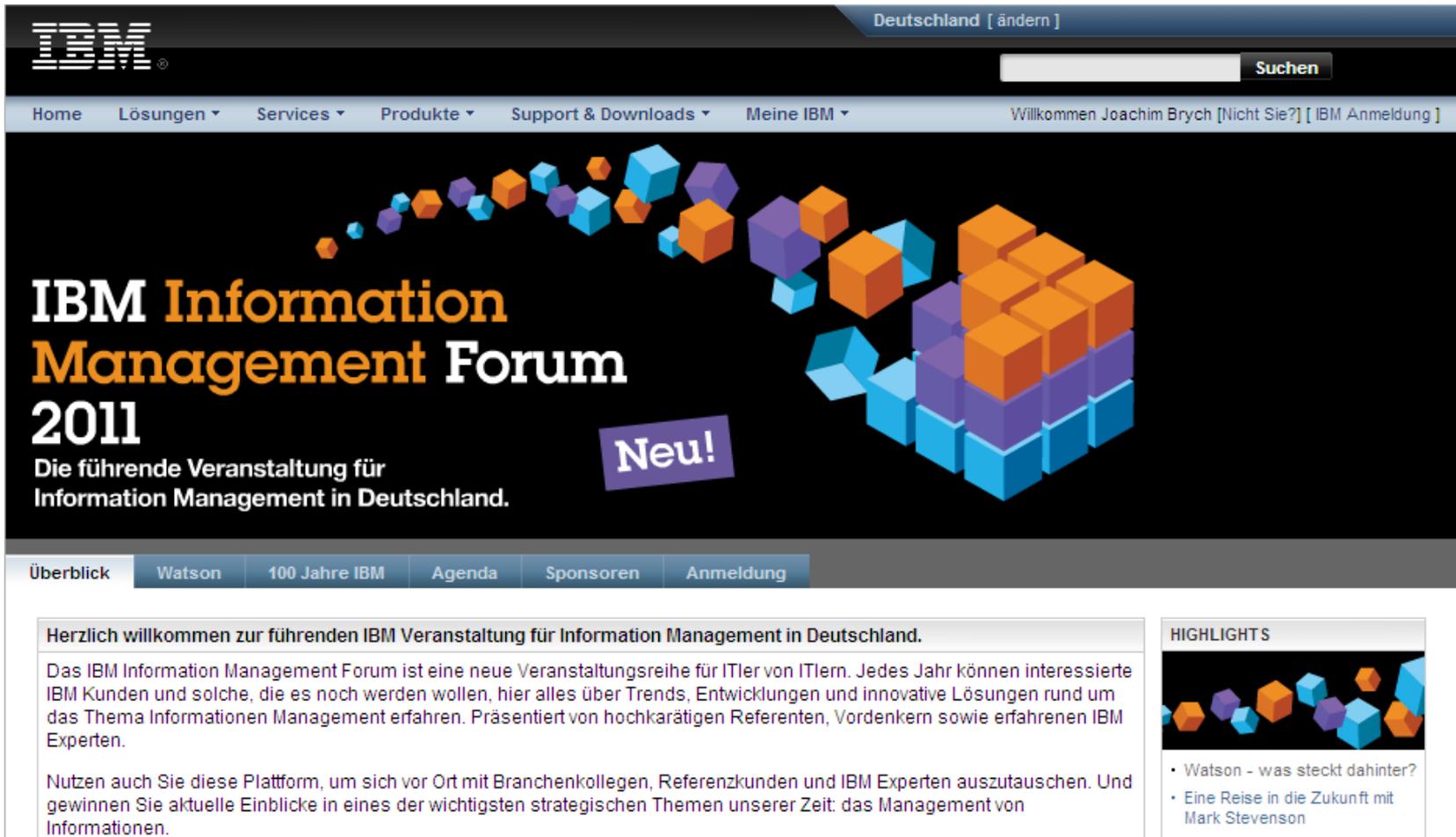
- Haben Sie Performance Probleme mit den Abfragen in Ihrem Data Warehouse / Data Mart?
- Möchten Sie als Fachabteilung Ihr eigenes Analysesystem unabhängig von der IT haben, ohne gleichzeitig tiefes IT Know-How haben zu müssen?
- Möchten Sie OLTP und OLAP in einem gemeinsamen System machen?
- Möchten Sie sich die aufwendige Datenaufbereitung für Ihr MOLAP System sparen und gleichzeitig bessere Performance haben?
- Haben Sie Skalierungsprobleme mit Ihrem MOLAP System?
- Möchten Sie Ihr bestehendes Informix OLTP System um performante Analysen und Berichte ergänzen ohne zusätzliche Last auf das OLTP System zu bringen?

Zusammenfassung

- Gewaltige Performance-Vorteile (normalerweise große Faktoren) im Vergleich zu traditionellen relationalen Datenbank Management Systemen
 - skalierbare shared-nothing Architektur
 - speziell für in-memory Verarbeitung optimiert
 - speziell für moderne Prozessorarchitekturen optimiert
 - spaltenorientierte Komprimierung und Speicherung der Daten
 - Verarbeitung von komprimierten Daten
- keine Anwendungsänderungen notwendig
- kein Tuning notwendig
- Kombination von OLTP and Data Warehousing in einem System möglich

Information Management Forum 2011

Der deutsche **Information On Demand** Event , 23. – 24. Mai 2011



Deutschland [ändern]

Suchen

Home Lösungen Services Produkte Support & Downloads Meine IBM Willkommen Joachim Brych [Nicht Sie?] [IBM Anmeldung]

IBM Information Management Forum 2011

Die führende Veranstaltung für Information Management in Deutschland.

Neu!

Überblick **Watson** 100 Jahre IBM Agenda Sponsoren Anmeldung

Herzlich willkommen zur führenden IBM Veranstaltung für Information Management in Deutschland.

Das IBM Information Management Forum ist eine neue Veranstaltungsreihe für ITler von ITlern. Jedes Jahr können interessierte IBM Kunden und solche, die es noch werden wollen, hier alles über Trends, Entwicklungen und innovative Lösungen rund um das Thema Informationen Management erfahren. Präsentiert von hochkarätigen Referenten, Vordenkern sowie erfahrenen IBM Experten.

Nutzen auch Sie diese Plattform, um sich vor Ort mit Branchenkollegen, Referenzkunden und IBM Experten auszutauschen. Und gewinnen Sie aktuelle Einblicke in eines der wichtigsten strategischen Themen unserer Zeit: das Management von Informationen.

HIGHLIGHTS

- Watson - was steckt dahinter?
- Eine Reise in die Zukunft mit Mark Stevenson

auch mit Vortrag zu In-Memory-Datenbanken!

