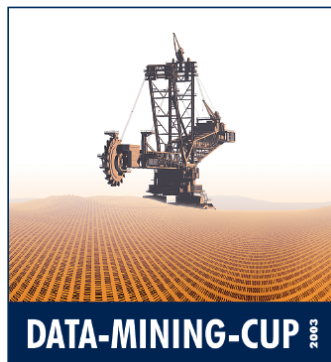


---

**Vortrag auf den 4. DATA-MINING-CUP Anwendertagen  
(Chemnitz, 18.-20. Juni 2003)**



<http://www.data-mining-cup.de>

Copyright-Hinweis:

Das Urheberrecht des folgenden Vortrags liegt beim Author. Verbreitung, Vervielfältigung und Kopie, auch auszugsweise, ist nur mit schriftlicher Genehmigung des Authors erlaubt.



IBM Software Group

## Data Mining – Leicht zu benutzen und integriert

Gerd Piel

gerd\_piel@de.ibm.com

Leiter Data Mining Entwicklung  
IBM Deutschland Entwicklung GmbH, Böblingen

**DB2** Information Management Software

 business on demand software

IBM Software Group | DB2 Information Management Software



## Agenda

### Ø IBMs Data Mining Entwicklung in Böblingen

#### § Data Mining – Easy to use und integriert – heute

- 4 BI Plattform – DB2 Data Warehouse Edition
- 4 Anwendungsintegration

#### § Data Mining – Easy to use und integriert – in der Zukunft

- 4 Business on demand

## IBM Development Lab @ Böblingen bei Stuttgart



IBMs größtes Entwicklungslabor ausserhalb der USA

## IBM Development @ Böblingen

### § DB2 Information Management

- 4 Data Mining
- 4 Information Mining / Text Search
- 4 Content Management
  - CommonStore
- 4 Information Integration
- 4 DB2 Performance Tools
- 4 SAP DB2 Multiplattform

### § e-business Integration

- 4 WebSphere Portal Server
- 4 WebSphere Workflow
- 4 Pervasive Computing

### § Betriebssysteme

- 4 Linux für zSeries
- 4 VSE
- 4 Systems Management

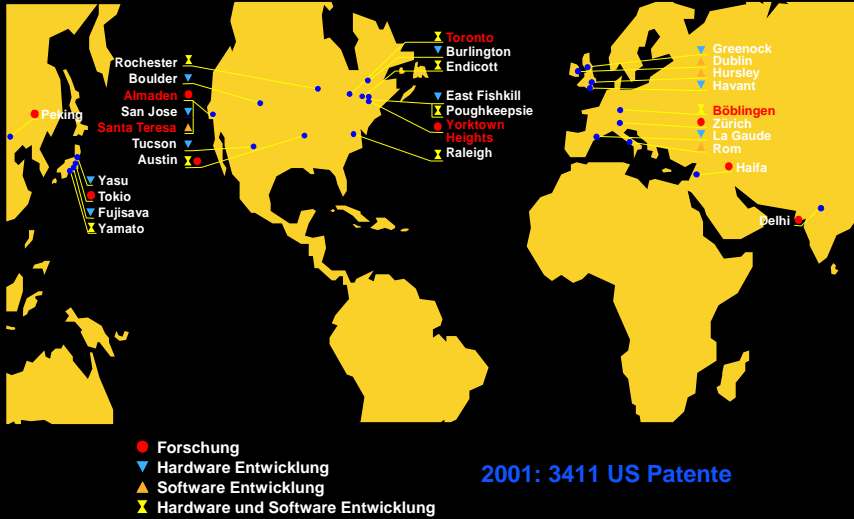
### § Industrie-Lösungen

- 4 Finanzanwendungen
- 4 Life Science

### § Hardware Entwicklung

- 4 zSeries Prozessoren
- 4 OEM Micorprozessoren

## IBM hat 29 Forschungs- und Entwicklungsstandorte



## DB2 Information Management Software

*Information Management... mehr als eine Datenbank*

Enterprise Content Mgmt.

Business Intelligence

DB2 Content Manager

DB2 for BI

Information Integration

Database Servers & Tools

DB2 Information Integrator

DB2 UDB



Entwicklung	3.000+ Mitarbeiter
Vertrieb / Marketing	3.000+ Mitarbeiter
Partner	16.000+ Partner

## Intelligent Miner Produktfamilie V 8.1

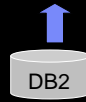
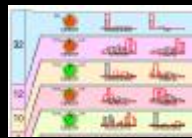
Mining Workbench / Expertentool

DB2 Erweiterung / Middleware

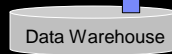
### IM for Data



Statistiker / Experte



Extrahieren



Anwendungsprogrammierer



IM Modeling

IM Scoring

### IM Visualization

für PMML / XML Modelle



## Agenda

ü IBMs Data Mining Entwicklung in Böblingen

∅ Data Mining – Easy to use und integriert – heute

∅ BI Plattform – DB2 Data Warehouse Edition

∅ Anwendungsintegration

§ Data Mining – Easy to use und integriert – in der Zukunft

4 Business on demand

## Was ist die DB2 Data Warehouse Edition?

### §Eine Strategie, um mehr BI Funktionen mehr Kunden zur Verfügung zu stellen

f Vereinfachung: leicht zu benutzen, leicht zu kaufen, leicht zu verstehen

### §Eine Sammlung von Tools, um Data Warehouses und analytische Anwendungen zu erstellen:

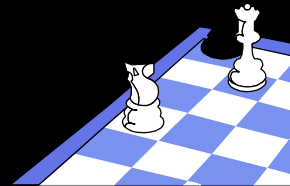
f DB2 UDB Data Warehouse **Enterprise** Edition

f DB2 UDB Data Warehouse **Standard** Edition

### §Sehr attraktive Preisgestaltung

### §Eine Grundlage für *Information On Demand*

f über das Echtzeit-Data-Warehouse

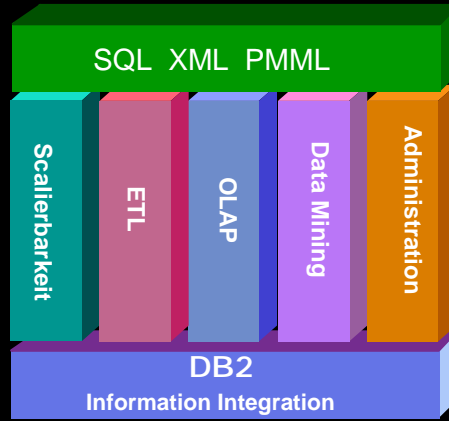


## Die DWE Plattform Komponenten

DB2 Enterprise Server Edition	Basis der BI Plattform
Database Partitioning Feature	“Shared-nothing” Skalierbarkeit
DB2 Cube Views	OLAP Prozessor in DB2
<b>Intelligent Miner Scoring, Modeling &amp; Visualization</b>	<b>Data Mining in DB2</b>
Warehouse Center & Manager	ETL Transformationen & Terminierung
Office Connect Enterprise Edition	Excel Connectivity zu DB2
Query Patroller	Query Management
Information Integrator	Zugriff auf heterogene Daten



## Das DB2 Rahmenwerk für Business Intelligence



BI Funktionen in der Datenbank als Teil einer integrierten BI Plattform mit offenen, standardisierten Schnittstellen.

Kooperation mit Partnern für andere Ebenen der BI Architektur.

## Die DWE bringt die DB2 BI Vision



## DB2 Data Warehouse Standard Edition Was ist enthalten?



	Wert
DB2 UDB Workgroup Unlimited V8.1	Query Performance
DB2 Cube Views V8.1	OLAP Summary Reporting
DB2 Intelligent Miner Modeling V8.1	Data Mining Modelle erstellen
DB2 Intelligent Miner Visualization V8.1	Data Mining Modelle visualisieren
DB2 Intelligent Miner Scoring V8.1	Score von Daten mit Mining Modellen
Office Connect Professional Edition V4.0	Query & reporting mit Spreadsheets
DB2 Warehouse Center V8.1	Extract, Transform und Load von Produktionsdaten ins DB2 Warehouse

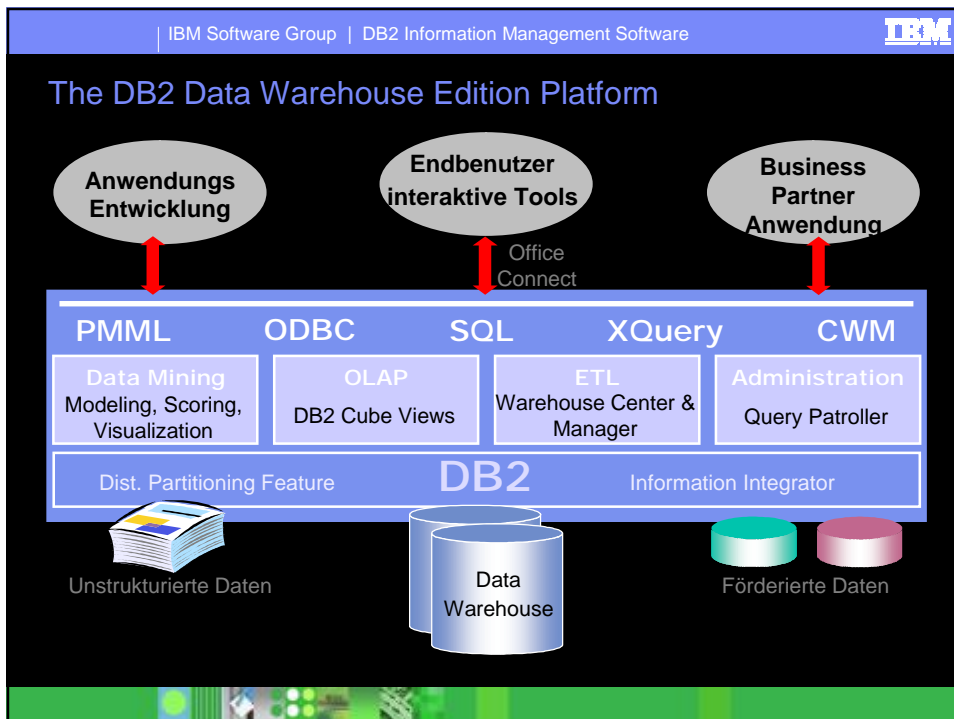
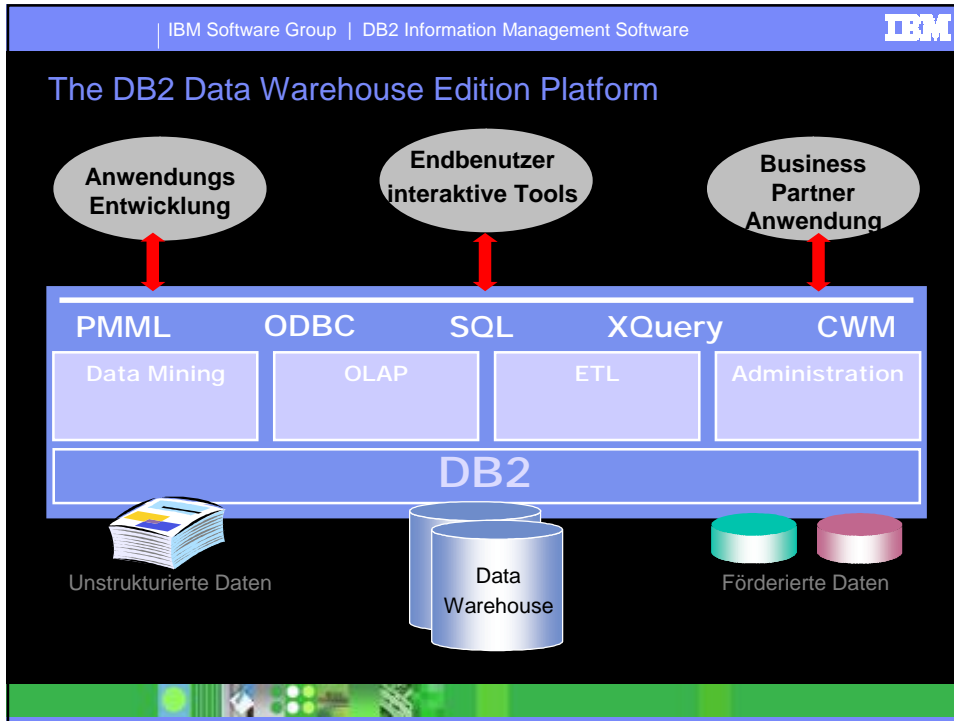
Hauptsächlich für Mittelstandskunden, Fachbereichs- und Entwicklungswarehouses

## DB2 Data Warehouse Enterprise Edition Was ist enthalten?



	Value
DB2 UDB Enterprise Server Edition V8.1	Query Performance
DB2 Cube Views V8.1	OLAP summary reporting
DB2 Intelligent Miner Modeling V8.1	Data Mining Modelle erstellen
DB2 Intelligent Miner Visualization V8.1	Data Mining Modelle visualisieren
DB2 Intelligent Miner Scoring V8.1	Score von Daten mit Mining Modellen
Office Connect Web Enterprise Edition V4.0	Query & reporting mit Spreadsheets
DB2 Warehouse Manager Standard Edition V8.1	Extract, Transform und Load von Produktionsdaten ins DB2
DB2 Query Patroller V8.1	Manage Queries performant
DB2 Data Partitioning Feature V8.1	Skalierbarkeit durch Aufteilung
DB2 Information Integrator Standard Edition V8.1	Zugriff auf heterogene Daten

Hauptsächlich für große Data Warehouses oder Data Marts



## Beispiel DB2 Data Warehouse **Standard** Edition Preis

	metric	Unit price \$K
DB2 UDB Workgroup Unlimited V8.1	CPU	7.5
DB2 Cube Views V8.1	CPU	7.5
DB2 Intelligent Miner Modeling V8.1	server	45.0
DB2 Intelligent Miner Visualization V8.1	server	15.0
DB2 Intelligent Miner Scoring V8.1	CPU	15.0
Office Connect Professional Edition V4.0	10 users	2.5
DB2 Warehouse Center V8.1	CPU	0
Total der Einzelpreise		92.5
DWE Preis	CPU	15.0
Total Ersparnis für eine CPU / Einzelner Server		\$77.5

Hauptsächlich für Mittelstandskunden und Data Warehouses einzelner Bereiche

## Der Data Mining Prozess mit DB2 DWE

Analyst definiert Modelle und macht die Analyse ...



XML



XML

Modelle werden in Echtzeit in Anwendungen angewendet



... durch Nutzung der Daten im Data Warehouse



Modelling

Modelle können über das Web angesehen werden



Visualization

DB2 OLTP oder DW oder ODS

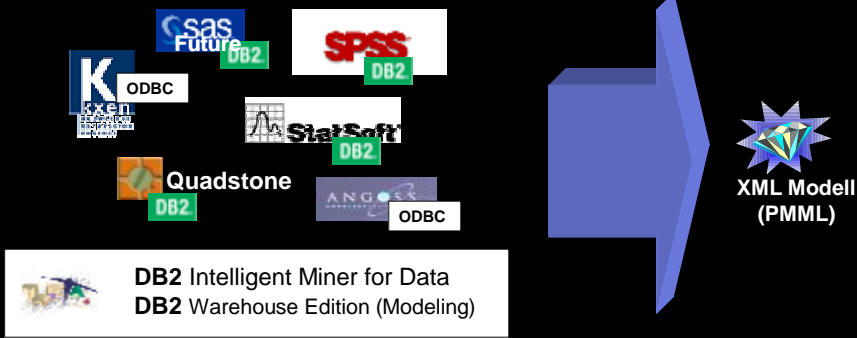


Scoring

## Der Data Mining Prozess mit DB2 DWE

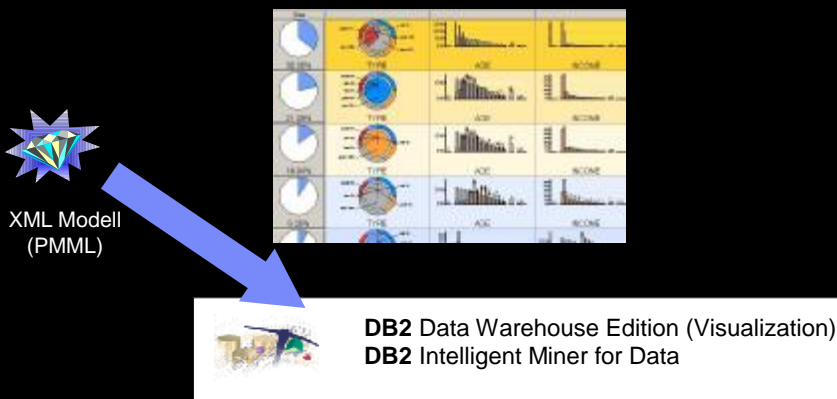
- § Modeling
- § Visualization
- § Scoring

IBM oder andere Tools erstellen Modell  
(jedes Tool, das XML/PMML Modelle ausgibt, kann Modelle für DB2 produzieren)

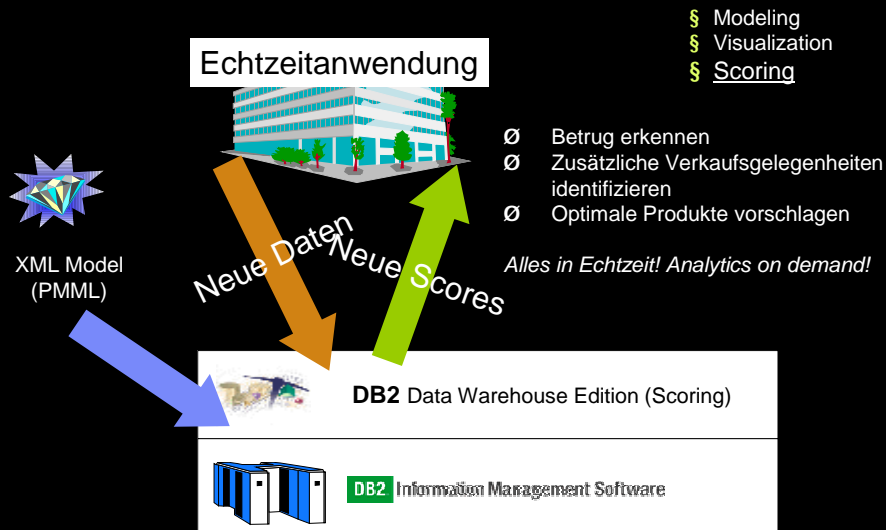


## Der Data Mining Prozess mit DB2 DWE

- § Modeling
- § Visualization
- § Scoring



## Der Data Mining Prozess mit DB2 DWE



## Beispiel IM Scoring SQL Statement

```

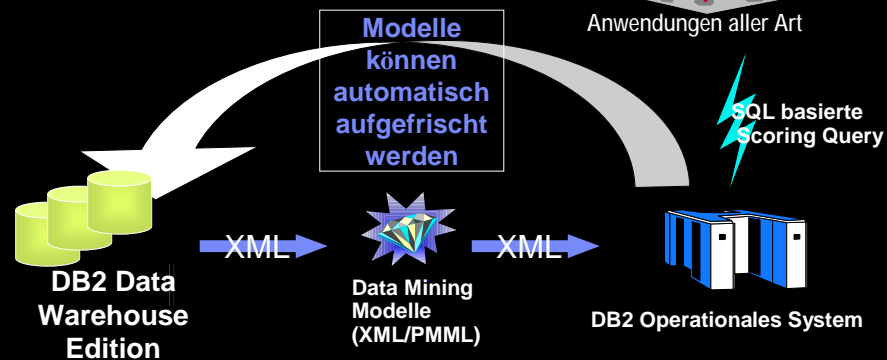
WITH clusterView(clusterResult )AS
(
  SELECT IDMMX.DM_applyClusModel(C.MODEL,
    IDMMX.DM_impApplData(
      rec2xml(1.0,'COLATTVAL ','',B.TYPE,B.AGE,
        B.SIBLINGS, B.INCOME )))
  FROM BANKING_SCORING B, IDMMX.ClusterModels C
  WHERE C.MODELNAME='DemoBanking '
)
SELECT IDMMX.DM_getClusterID(clusterResult ),
  IDMMX.DM_getClusScore(clusterResult )
FROM clusterView ;

```

## Aktuelle Modelle mit DB2 DWE (Modeling)

Aktuelle Modelle bedeuten schnellere Antwort auf Änderungen in der Geschäftsumgebung

Alle Funktionen werden mit SQL kontrolliert



## Standards

### § SQL/MM: Data Mining

- 4 SQL Erweiterung für Data Mining
- 4 IBM, Oracle

### § PMML 2.0

- 4 Predictive Model Markup Language
- 4 [www.dmg.org](http://www.dmg.org)
- 4 Mining Modelle in XML
- 4 Austausch der Modelle zwischen verschiedenen Data Mining Anbietern
- 4 IBM, SAS, SPSS, NCR, Microsoft, Oracle, Angoss, SIGKDD

## Data Mining Funktionen in Anwendungen

- § Gute Beschreibung in Redbook: Enhance Your Business Applications:
  - 4 Einfache Integration von umfassenden Data Mining Funktionen
  - 4 <http://publib-b.boulder.ibm.com/cgi-bin/searchsite.cgi?query=SG24-6879-00>



## Agenda

- ü IBMs Data Mining Entwicklung in Böblingen
  - Ø Data Mining – Easy to use und integriert – heute
    - ü BI Plattform – DB2 Data Warehouse Edition
    - Ø Anwendungsintegration
- § Data Mining – Easy to use und integriert – in der Zukunft
  - 4 Business on demand

## Warum ist Anwendungsintegration von Data Mining wichtig?

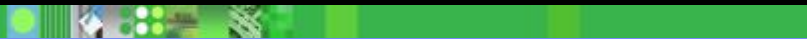
- § „manuelles“ Data Mining mit einer „Workbench“ ist teuer
  - 4 hochbezahlte Data Mining Spezialisten, falls überhaupt verfügbar
  - 4 oft aufwendige Datenvorverarbeitung, Bilden von „ad hoc“ Datenmodellen
  - 4 Umsetzen der Erkenntnisse aus den Mining-Modellen in den Produktionsablauf ist oft sehr zeitaufwendig und kostenintensiv
- § „Quick Time to Value“ und „Return on investment“ sind hier wichtige Schlagworte
- § Data Mining muss automatisiert werden
  - Dies betrifft sowohl
    - 4 das Anwenden der Mining-Modelle als auch
    - 4 das Trainieren/Erstellen der Mining-Modelle

## Dilemma bei der Anwendungsintegration

- § Data Mining Spezialisten können oft nicht programmieren.
- § Programmierer verstehen oft nichts oder nicht viel von Data Mining.

## Lösen des Dilemmas: 2 Ansätze

1. Entwicklungsumgebungen für Data Mining Lösungen, in denen man ohne Programmierung auskommen kann
    - 4 Intelligent Miner Solution Framework
  2. Vereinfachen von Data Mining für den Anwendungsentwickler
    - 4 Einfache Programmierschnittstelle
    - 4 Höherwertige Funktionen
    - 4 Integration in Entwicklungsumgebungen, z.B. Eclipse/WebSphere Studio Application Developer
- è „Quicker Time to Value“



## Höherwertige Funktionen

- § Zusammenfassung typischer Schritte innerhalb des Mining-Prozesses
  - 4 z.B. Entdecken von Ausreißern
- § Anwendungsnahe Funktionen
  - 4 Erzeugen von Verkaufsempfehlungen aus Bondaten
  - 4 Erzeugen von Erklärungen für Produktionsfehler, Garantiefälle, ..



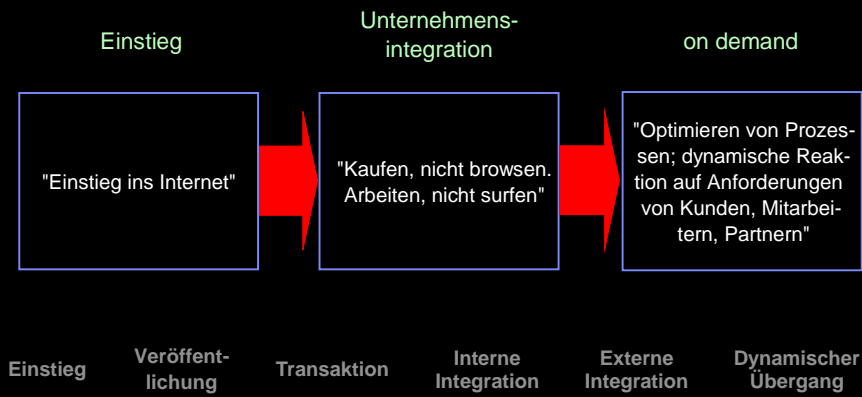
## Agenda

- ü IBMs Data Mining Entwicklung in Böblingen
- ü Data Mining – Easy to use und integriert – heute
  - ü BI Plattform – DB2 Data Warehouse Edition
  - ü Anwendungsintegration
- Ø Data Mining – Easy to use und integriert – in der Zukunft
  - Ø e-business on demand

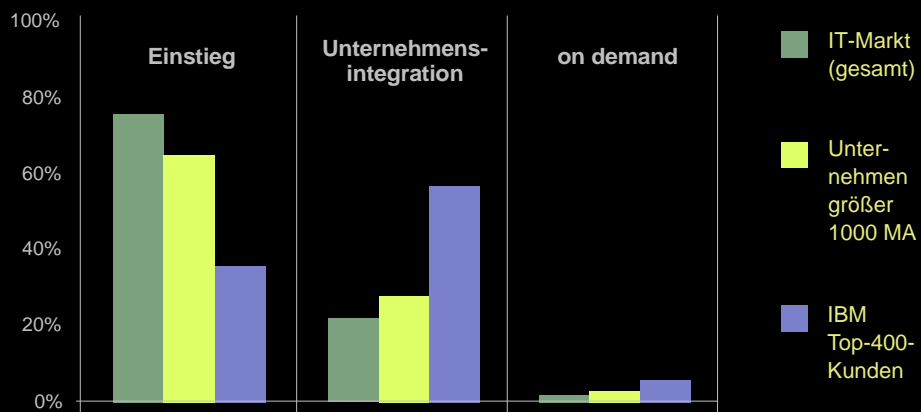


**e** business on demand


## Phasen des Übergangs zum e-business



## Unternehmen auf dem Weg zum on demand Business



Source: e-business Adoption Tracking Study, 1Q02

 business on demand

*Ein Unternehmen, dessen gesamte Geschäftsprozesse – integrierte, unternehmensweite End-To-End-Prozesse sowie Prozesse mit zentralen Partnern, Lieferanten und Kunden – so ausgelegt sind, dass schnell auf Kundenanforderungen, Marktchancen oder externe Risiken reagiert werden kann.*

## Die on demand Welt

1. Das on demand Business
2. Lösungen für on demand Betriebsumgebungen
3. Neu definiert: Erwerb und Management von IT-Assets

## Das on demand Business

- ✓ Reaktionsfähig: Reaktionen in Echtzeit
- ✓ Variabel: Nutzungsbasierte Kostenstrukturen
- ✓ Zielgerichtet: Konzentration auf das Wesentliche
- ✓ Widerstandsfähig: Verfügbarkeit – weltweit, rund um die Uhr

## Das on demand Business – Beispiele



Marktplatz ohne  
Ladenschlusszeiten



Zugriff auf Brustkrebsdiagnose-  
daten zusammen mit dem  
behandelnden Arzt



Mehr geschäftliche Flexibilität  
durch virtuelle IT-Ressourcen

## Lösungen für on demand Betriebsumgebungen

*Eine auf offenen Standards  
basierende, heterogene Welt –  
integrationsfähig und mit  
individuellen Fähigkeiten  
ausgestattet*



## Lösungen für on demand Betriebsumgebungen

**Anforderung: Integriert**



## Lösungen für on demand Betriebsumgebungen

Anforderung: Offen



## Lösungen für on demand Betriebsumgebungen

Anforderung: Virtuell



## Lösungen für on demand Betriebsumgebungen

Anforderung: **Autonom**

selbst-schützend

selbst-optimierend

selbst-heilend

selbst-konfigurierend



## Die on demand Welt

### on demand Business

- Reaktionsfähig
- Variabel
- Zielgerichtet
- Widerstandsfähig

### Lösungen für on demand Betriebsumgebungen

- Integriert
- Offen
- Virtuell
- Autonom



## Die ersten Schritte

- Ist Ihr Unternehmen in der Lage, dynamisch und in Echtzeit auf ein sich ständig änderndes Geschäftsumfeld zu reagieren?
- Welche Geschäftsprozesse sehen Sie als Ihre Kernkompetenzen, die dazu beitragen, sich von Ihren Mitbewerbern zu unterscheiden?
- Verfügen Sie über "Best Practices" bei der Prozessoptimierung und branchenführendes Know-how beim Prozessdesign?
- Sehen Sie Möglichkeiten, die Ausnutzung Ihrer IT-Ressourcen entscheidend zu verbessern?

## DB2 Information Management Software

*Information Management... Mehr als eine Datenbank*

### Enterprise Content Mgmt.

- § Optimierte Ressourcenmanager für Rich Content
- § Management der digitalen Rechte
- § e-Records Management

**DB2 Content Manager**

### Information Integration

- § Heterogene Föderation, Search, Replication

**DB2 Information Integrator**



### Business Intelligence

- § Integrierte analytische Funktionen
- § Lösungsintegration

**DB2 for BI**

### Database Servers & Tools

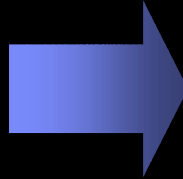
- § OLTP & Analytische Anwendungen
- § Hohe Performance & Skalierbarkeit
- § Hohe Verfügbarkeit, Sicherheit
- § Plattform Optimierung
- § Reduzierte Administrationskosten durch Automation & Tools

**DB2 UDB**

## Business Intelligence Angebote

On Demand Business

- **Reaktionsfähig in Echtzeit**
- **Variabel nutzenbasierte Kostenstrukturen**
- Zielgerichtet  
Konzentration auf das Wesentliche
- Widerstandfähig  
Verfügbarkeit weltweit, rund um die Uhr



### Erkenntnis getriebene Entscheidungen

Rendite  
Outsourcing zur Fokussierung

### Risiko Minimierung

Geschäftsoptimierung

### Technologie-Optimierung

**Integriert**

**Offen**

Virtuell

Autonom

## Was bedeutet e-business on Demand für Data Mining?

### § Integriert

- 4 Data Mining integriert in e-business Anwendungen
- 4 Data Mining Anwendungen übers Web (e-Utility)
- 4 Real Time Data Warehousing / Data Mining

### § offene Standards

- 4 SQL/MM, PMML, JSR 31
- 4 Bald: Standard für Mining Web-Services

### § Virtuell

- 4 Data Mining mit Grid Computing

### § Autonom

- 4 Verstärkte Automation von Mining-Funktionen
- 4 Data Mining als integraler Bestandteil des "Autonomic Computing" für automatisches Konfigurieren und Lösen von Problemen

## Zusammenfassung

- § Data Mining verändert sich stark von isolierten Consulting/Experten Analysetools hin zu wichtigen und differenzierenden Bestandteilen von Anwendungen
  - 4 Data Mining unter der Decke und automatisiert
  - 4 Data Mining ist integriert und einfach zu benutzen
  - 4 Die meisten Benutzer wissen gar nicht, dass sie Data Mining nutzen
- § Data Mining wird fester Bestandteil von Business Intelligence Plattformen
  - 4 Das Zusammenspiel der Komponenten und damit die Integration wird immer wichtiger
- § Neue Herausforderungen für Data Mining in der neuen e-business on Demand Ära