



IBM Software Partner Academy Program

**Telefonkonferenz am 05.02.2010**

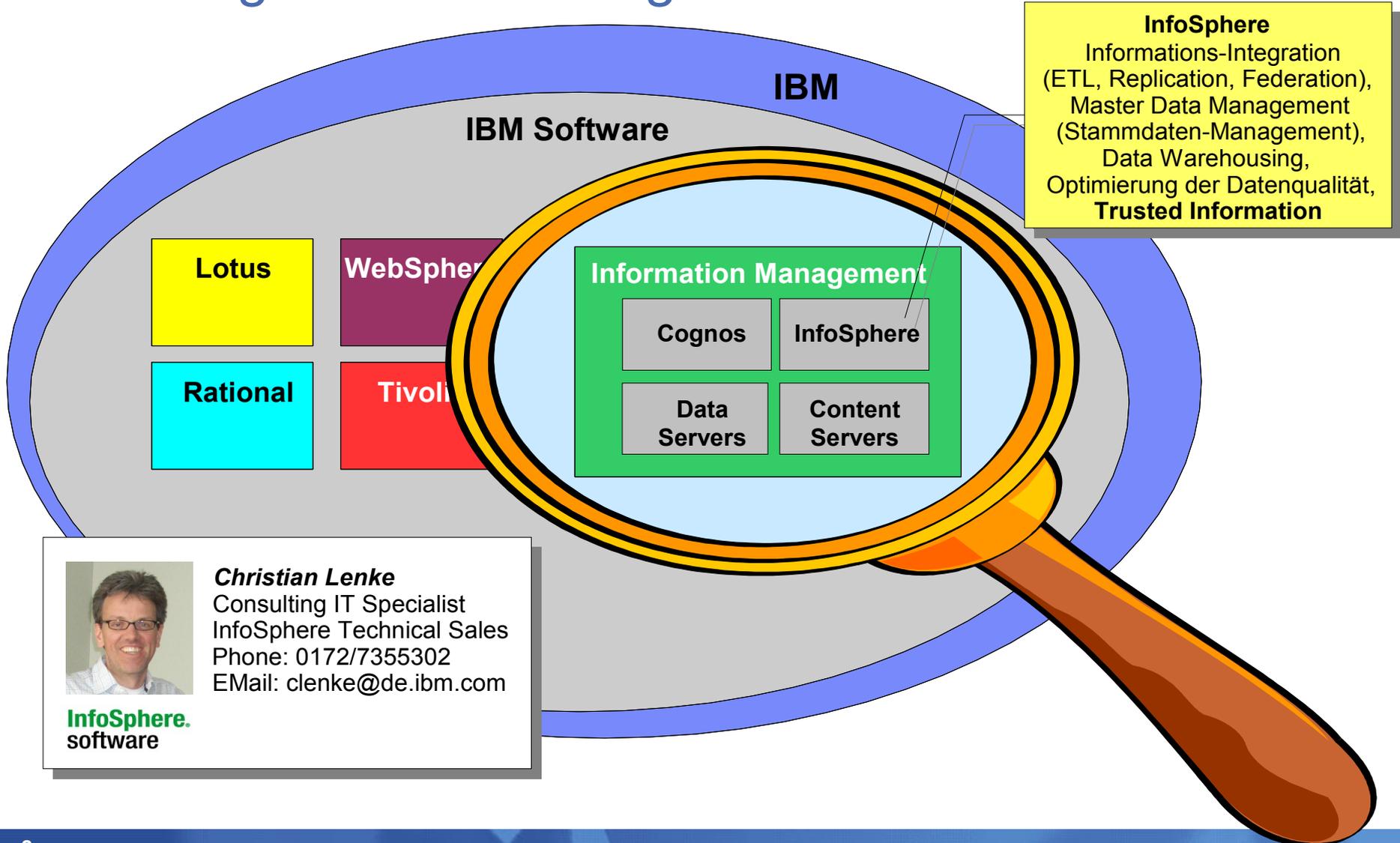
**IBM InfoSphere Information Server**

Christian Lenke

*IT Specialist for InfoSphere Software*

*Information Integration, Data Replication and Data Management*

# Einleitung und Vorstellung



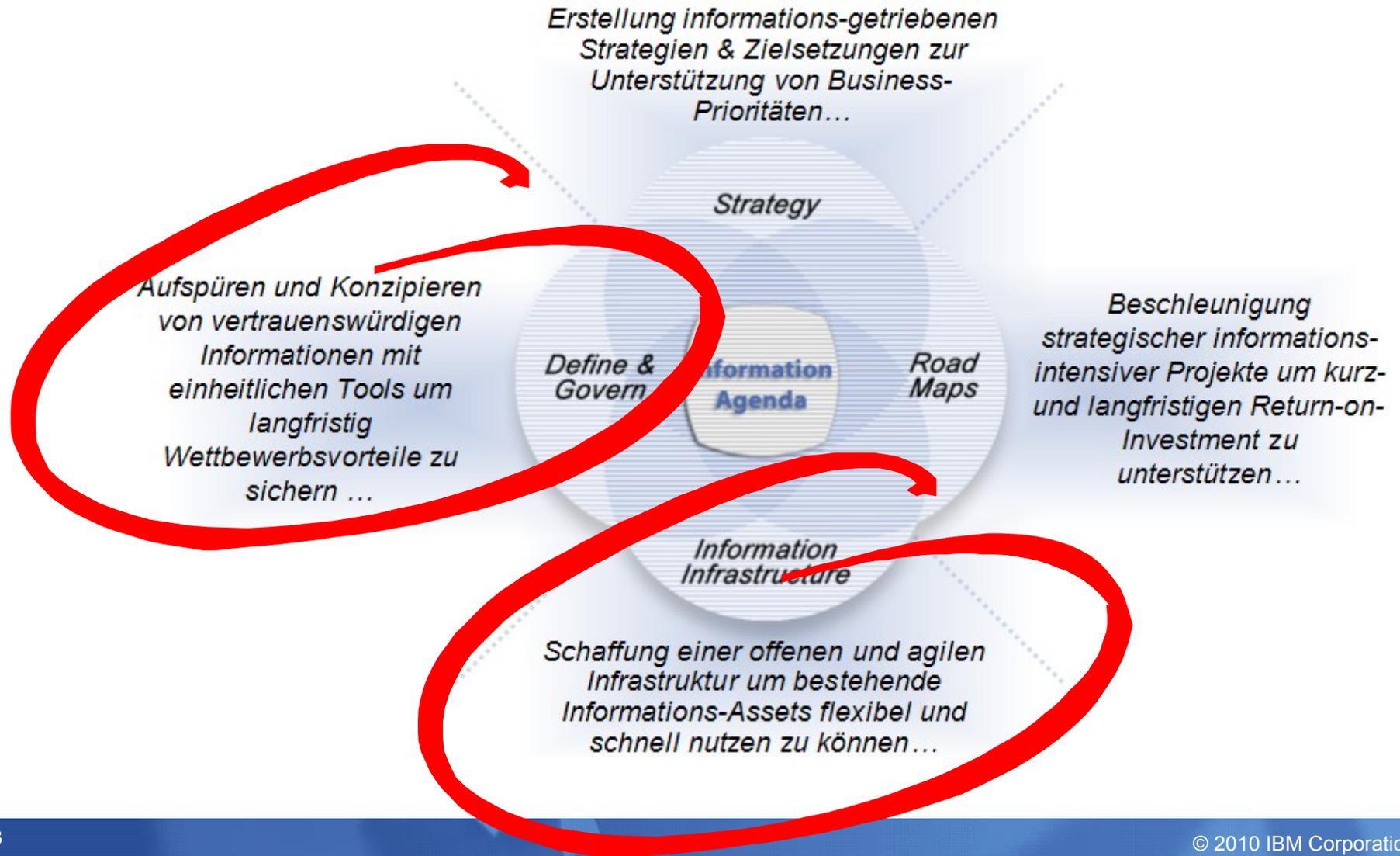
**InfoSphere**  
 Informations-Integration  
 (ETL, Replication, Federation),  
 Master Data Management  
 (Stammdaten-Management),  
 Data Warehousing,  
 Optimierung der Datenqualität,  
**Trusted Information**



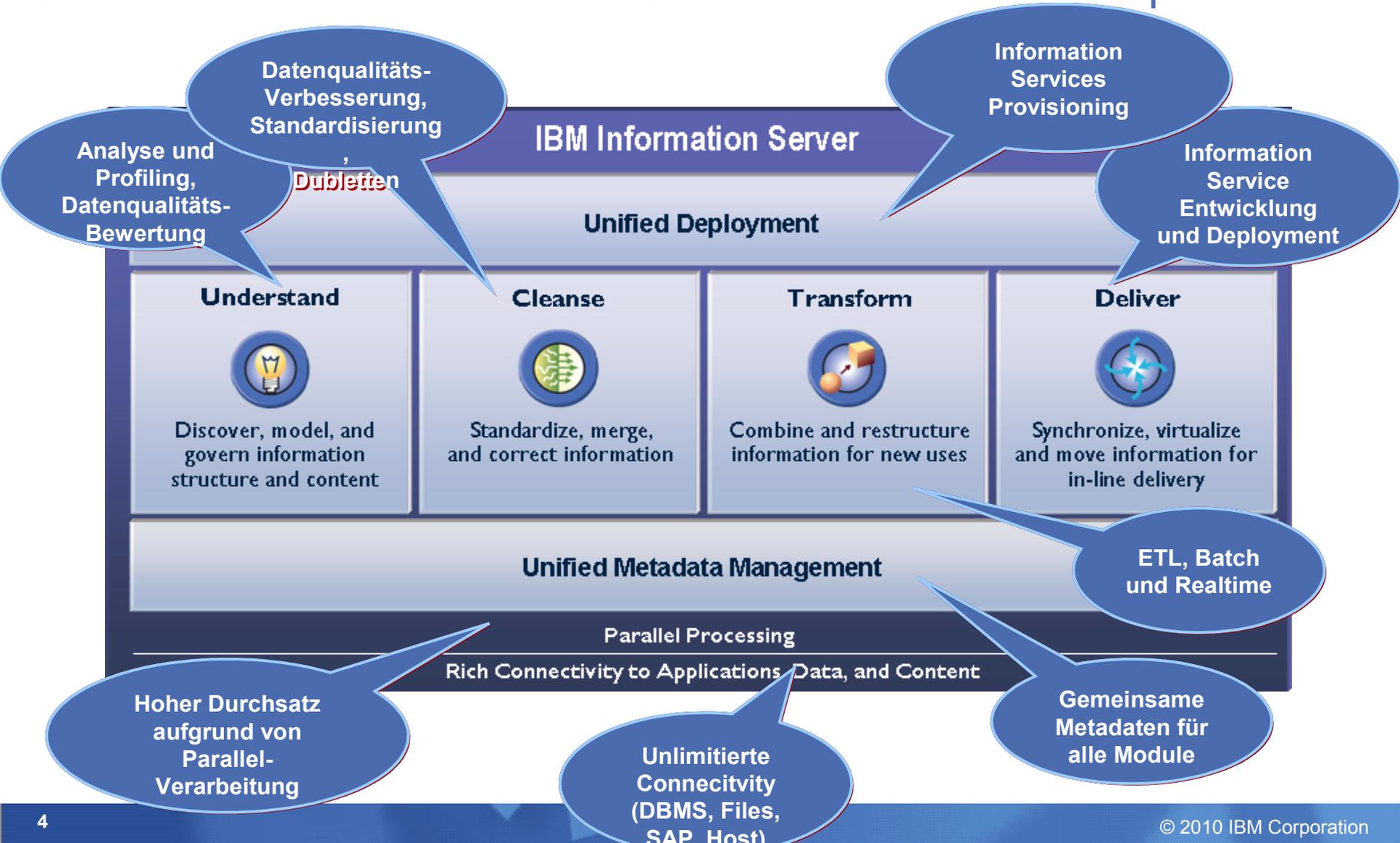
**Christian Lenke**  
 Consulting IT Specialist  
 InfoSphere Technical Sales  
 Phone: 0172/7355302  
 EMail: clenke@de.ibm.com

**InfoSphere.**  
 software

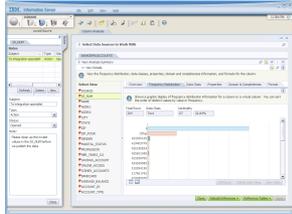
# Positionierung: Information Server und Information Agenda



# Information Server – Trusted Information for the Enterprise

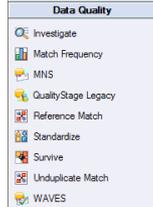


# Information Server - Komponenten



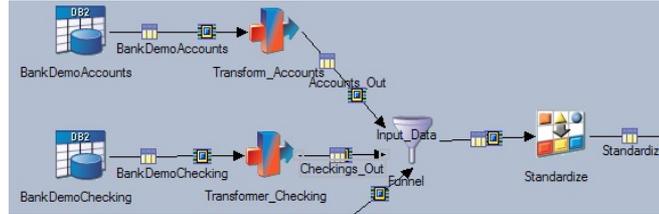
**Erkennen / Überwachen von Informationsstruktur /-inhalt**

**Verstehen**



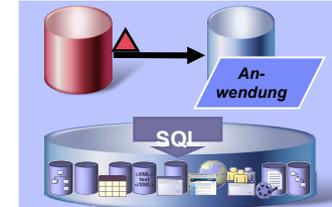
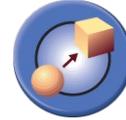
**Standardisieren / Korrigieren von Informationen**

**Bereinigen**



**Kombinieren / Restrukturieren von Informationen**

**Transformieren**



**Synchronisieren / Virtualisieren von Informationen, Information as a Service**

**Liefern**

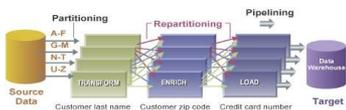


## Plattform Services

**Parallelität**



**Variable Parallelisierung Festlegung zur Laufzeit**



**Verbindung**



**Datenbanken, Files, Echtzeit, ERP, CRM, ...**



**Metadaten**



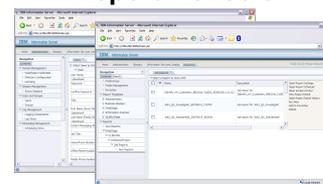
**Geschäfts-Definitionen Abhängigkeits-Analysen**



**Administration**



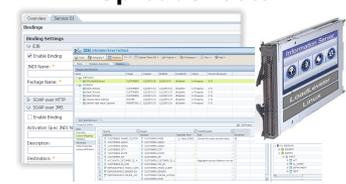
**Admin-Konsole Report-Konsole**



**Deployment**



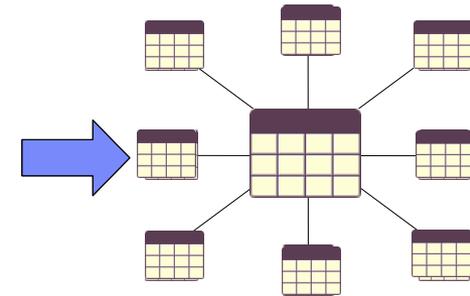
**Services / Blades / Spreadsheets**



# Welche Projekte profitieren vom Einsatz des Information Servers

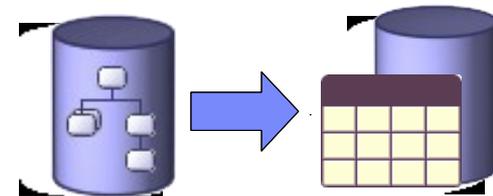
## DataWarehousing / Business Intelligence / Business Performance Management

- ◆ **Erstellung von Transformationsregeln** um Legacy-Daten (DBMS, File, SAP, Cobol, ...) in DWH-Formate und Datenmodelle umzusetzen
- ◆ **Mapping** von IT- und Fachvokabular
- ◆ **Vernetzung** von Reporting- und Transformations-Metadaten
- ◆ **High Performance ETL**



## Daten-Migrationen

- ◆ **Analyse** der Quelldaten
- ◆ **Bewertung** der Migrationsrisiken
- ◆ **Mapping** von Quell- und Zielstrukturen
- ◆ **Klassifizierung** von Feldern (Key, Konstante, Wert, ...)
- ◆ Durchführung der stichtagsbezogenen **Bulk-Migration**



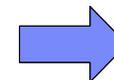
## Anwendungs-Modernisierung

- ◆ **Aufspüren** von Objektbeziehungen
- ◆ **Mapping** von Quell- und Zielstrukturen



## Datenqualitäts-Offensiven

- ◆ **Identifikation** von Dubletten
- ◆ **Periodische Prüfung** von Datenqualitätsregeln
- ◆ **Visualisierung** von Inkonsistenzen (Formate, Inhalte)



**Paul Müller**  
=  
**Dr. P. Mueller ???**

## Information Server ist nicht bloß ein „ETL-Tool“

- ETL ist eine wesentliche Komponente. *DataStage* und *QualityStage* sind bewährte Produkte im Markt
- Der Schwerpunkt wurde von IBM aber auf frühe Projektphasen verlagert (Requirements-Analyse und Dokumentation, Metadata, Datenqualität)
- Information Server ist eine Plattform. Alle Komponenten sind über das zentrale Metadata Repository integriert
- Aus der Requirements-Dokumentation können ETL-Datenflüsse generiert werden, was zu Produktivitätsvorteilen führt
- Über das Metadata Repository können Lineage und What-if Analysen erstellt werden

## Probleme ohne den Einsatz des Information Servers



Ohne Information Server werden heute vielfach die benötigten Informationen in unverbundenen Werkzeugen ohne jede Transparenz gespeichert (Excel, Word, Visio, etc.). Bei der Umsetzung in ETL-Tooling führt dieser Medienbruch zu Reibungsverlusten und Fehlern.

# Information Server unterstützt den gesamten Projekt Lifecycle

## Analyse

- Analyse und Dokumentation einheitlicher Fachbegriffe (fachliche Metadaten)
- Generierung technischer Metadaten
- Data Profiling
- Beurteilung der Datenqualität

## Design

- Erstellung von Mapping-Regeln
- Entwurf von Data Cleansing-Prozessen
- Dokumentation von Business-Regeln
- Generierung von Transformationen aus Mapping-Regeln

## Entwicklung

- Graphische Entwicklung von Transformationen
- Optimierung der Datenqualität: Standardisierung, Cleansing und Dubletten-Matching
- Automatische Dokumentation
- Effiziente Realisierung durch Kapselung und Wiederverwendbarkeit
- Integriertes Error Handling und Reporting
- Analyse von Impact und Data-Lineage

## Durchführung

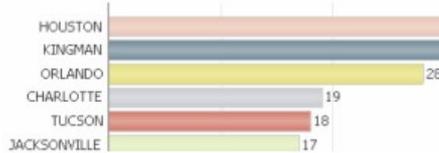
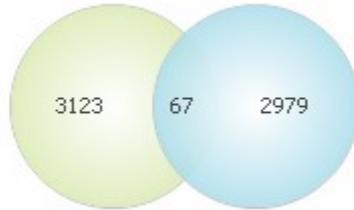
- Deployment der Entwicklung
- Effiziente Erstellung von Ergebnissen durch:
  - die massiv Parallele Engine
  - Flexible Parameterisierung
- Auswertung von Datenqualität und entsprechender Trendverläufe
- Aktivitätsüberwachung mit der Admin-Console
- Erzeugen von Runtime-Metadaten

# Beispiel: Bewertung der Datenqualität (z.B. Der Quellsysteme)

Die Bewertung der Datenqualität der Quellsysteme und Schnittstellen führt zu einer besseren Risiko- und Aufwand-Abschätzung.

## Data Profiling

General Format		
General Format	Count	Percent
:AAAAAAA, AAAAA	3267	10.22
:AAAAAA, AAAAAA	2853	8.93
:AAAAAAA, AAAA	2655	8.31
:AAAAAAA, AAA	2452	7.67
NA	2442	7.64



## Data Auditing

Carrier ID Validation							
View Results							
<input type="checkbox"/> Include Tests							
Job Log							
Type	Timestamp	Validity			Contact	Sample	
		Severity	# Pass	# Fail			
Run	01/01/09	10.8%	792	208	Mim Foster		
	12/01/08		904	96	Mim Foster		
	01/01/08		926	74	Mim Foster		
	08		500	0	Mim Foster		
			500	0	Mim Foster		

Wie kann ich Daten, Risiken identifizieren und mildern, wenn ich nicht weiss, welche Daten ich habe?

### Beispiel – Basic Level:

Füllgrad einer Spalte (wieviel NULLs, wie viele Werte)

### Beispiel Medium – Level :

Verteilung der Werte, Cardinality, Primary Key Kandidaten

### Beispiel Advanced – Level:

Identifikation von Foreign Keys

### Beispiel – Basic Level:

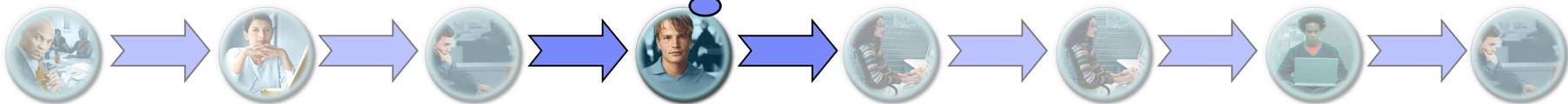
BIRTH\_DATE ist ein Datum

### Beispiel Medium – Level :

BIRTH\_DATE > 1900-01-01

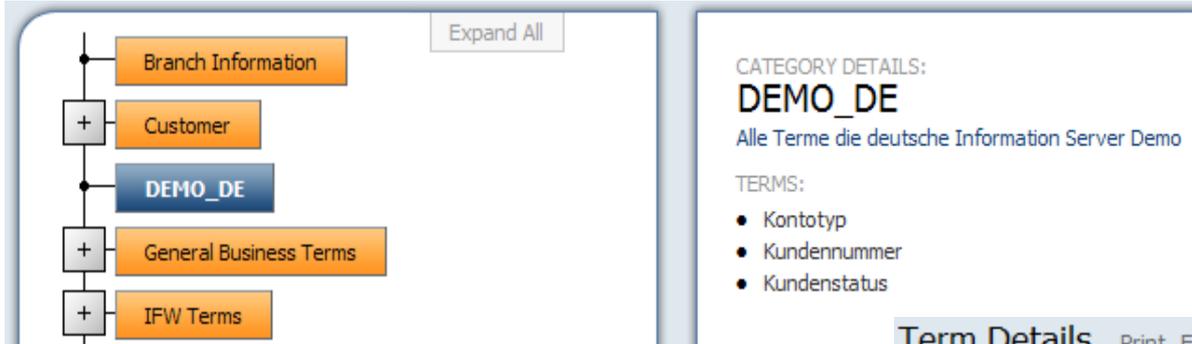
### Beispiel Advanced – Level:

PROFESSION = 'STUDENT' AND AGE < 40



Source Data Assessment

# Beispiel: Business Glossar. Zentrale Definition von Termen und Vernetzung mit IT-Assets



- Funktionen (Beispiele):**
- Hierarchische Darstellung von Kategorien und Termen
  - Advanced Search (Freitext)
  - Änderungshistorie
  - Verlinkung von Termen mit IT Assets
  - Dokumentation des Stewardship

Das Business Glossary kann als zentrale Dokumentation von fachlichen und technischen Begriffen eingesetzt werden und im Projekt verwendet werden, um beispielsweise unterschiedliche Begriffe für das selbe Objekt (z.B. Unterschiedliche fachliche und technische Beschreibung) zu verlinken.

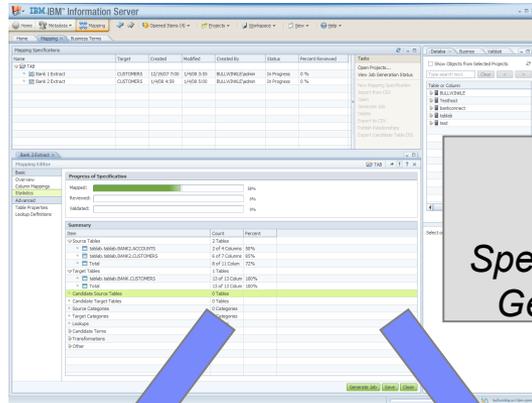
Beispiel: Begriff Kundenstatus  
Der Begriff ist an zwei Spalten des Schemas DEMO\_DE gebunden.

The 'Term Details' window shows the term 'Kundenstatus' under the category 'DEMO\_DE'. It includes a 'Short Description' (Status eines Kunden), a 'Long Description' (Status eines Kunden aufgrund seines Kontosaldos), and metadata such as 'Steward: Undefined', 'Assigned to terms: Undefined', and 'Status: Candidate'. A 'General Information' section lists fields like Abbreviation (KDSTATUS KSTATUS), Example (ENTRY, SILVER, GOLD, PLATINUM), Created By (isadmin), Created On (2009-07-31 16:41:15), Modified By (isadmin), and Modified On (2009-08-03 16:38:37).

Assigned Assets

1.	KDSTATUS	( IBM-01CEFA5FFF9 » SAMPLE » DEMO_DE » KDSTATUS )
2.	KSTATUS	( IBM-01CEFA5FFF9 » SAMPLE » DEMO_DE » KUNDE_TRUSTED )

# Beispiel: Identifikation und Dokumentation von Requirements. Automatische Generierung von ETL-Abläufen



**Definition der Mapping-Spezifikationen in der Geschäftssprache**

**Mapping Specification Implementation Details**

Project: Tablab  
Report Name: Bank 2 Extract  
Executed: 07-09-2007 14:31:45  
Generated By: Sam Gampage  
Description: Extracts info from bank 2

Specification Name: Bank 2 Extract  
Owner: In Progress

**Details**

Source Table:  
tablab.tablab.BANK2\_ACCOUNTS  
tablab.tablab.BANK2\_CUSTOMERS

Target Table:  
tablab.tablab.BANK2\_CUSTOMERS

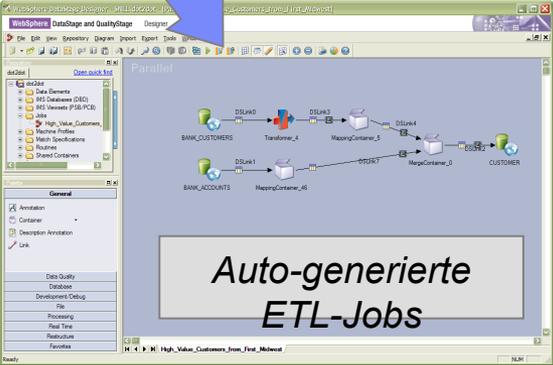
Filter:  
Join:  
Case:  
Reviewed:  
Documentation:

**Details**

Source Column:  
tablab.tablab.BANK2\_ACCOUNTS.ACCOUNT\_BALANCE  
tablab.tablab.BANK2\_CUSTOMERS.CUSTOMER\_ID

Target Column:  
tablab.tablab.BANK2\_CUSTOMERS.ACCOUNT\_BALANCE

**Flexibles Reporting und Tracking**



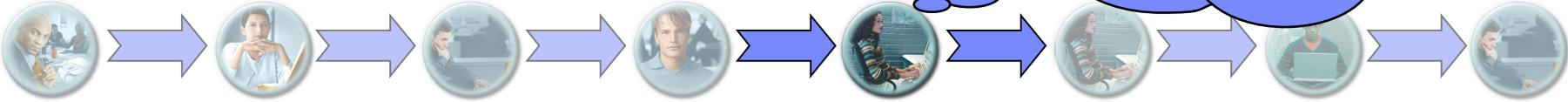
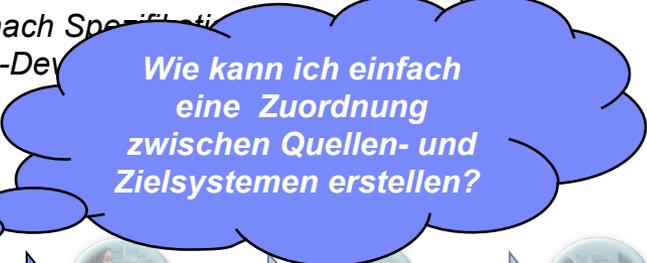
Auf Basis von Metadaten der Quell- und Zielsysteme wird in tabellarischer Form ein Mapping definiert, incl.:

- Filter
- Lookup
- Join
- Transformation

Transformationen können als DataStage Function oder in Geschäftssprache hinterlegt werden.

Mappings können aus Excel-Tabellen importiert werden.

Aus den Mappings werden lauffähige ETL-Jobs oder Job-Gerüste automatisch generiert (je nach Spezifikation)  
ETL-Dev

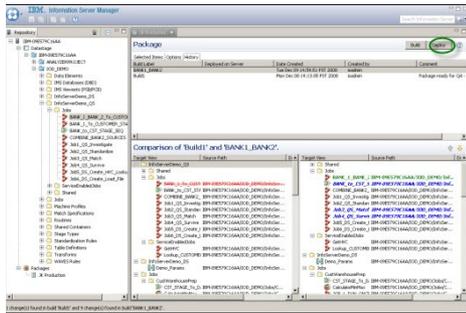


Source / Target Mapping

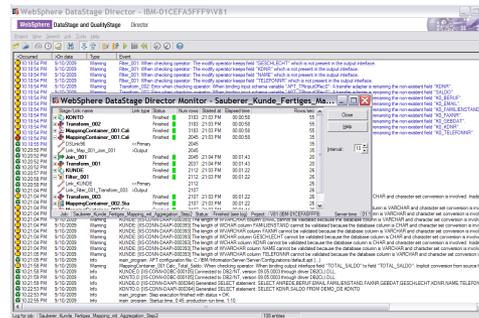


# Beispiel: Operation und Analyse

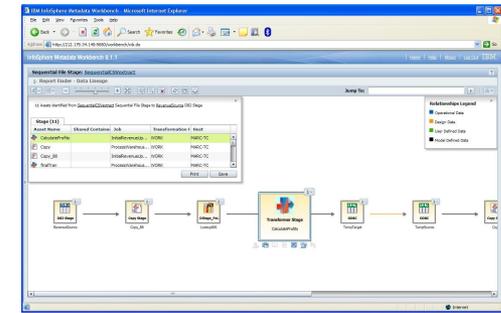
## Manager



## Director



## Metadata Workbench



### Deployment:

Management von Information-Server-Applikationen in einer Umgebung mit mehreren Instanzen (z.B. Test, QA, Prod).

### Version-Management:

Vergleich von Versionen, Roll-Back zu einer früheren Version eines Packages.

### Scheduling:

Integriertes Job- und Sequence-Scheduling. Externe Scheduler können problemlos integriert werden.

### Überwachung:

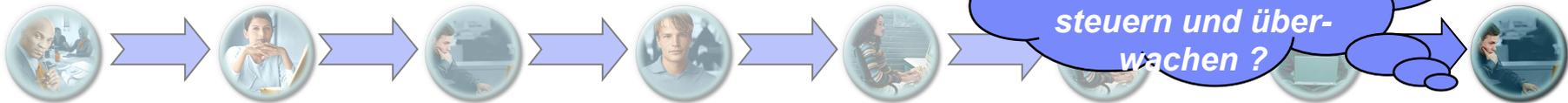
Drill-Down in die Stages eines Jobs. Protokollierung von Fehlersituationen und von operativen Metadaten.

### Metadata Interface:

Browser-basierter Zugriff auf alle Design-Metadaten und alle operationalen Metadaten. Drill-Down zu Spalten und transformationen.

### Lineage:

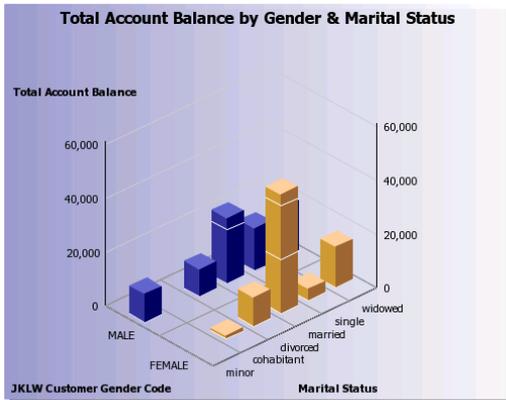
Wie kann ich die Umgebung aus dutzenden Jobs steuern und überwachen?



Operations

# Integration mit Cognos: Was genau bedeutet ein Geschäfts-Begriff. Link zum Information Server Business Glossary

Cognos Viewer - JKLW\_Customer\_Report\_3



**Expected Loss per Score Range / Line of Business**

Expected Loss %	Auto Finance	Credit Cards	Home Equit
Credit Risk Score 350 - 374	26.78205128	26.3556962	24.297297
Credit Risk Score 375 - 399	25.00943396	25.31162791	24.4576923
Credit Risk Score 400 - 424	24.68421053	27.5825	22.1
Credit Risk Score 425 - 449	24.756	24.81489362	22.7938775
Credit Risk Score 450 - 474	26.27428571	24.78414634	25.38437
Credit Risk Score 475 - 499	27.275	27.44074074	22.2
Credit Risk Score 500 - 524	22.69393939	27.04444444	25.17
Credit Risk Score 525 - 549	23.53333333	26.58333333	23.5130434
Credit Risk Score 550 - 574	25.996875	26.42359551	23.8588235
Credit Risk Score 575 - 599	26.81764706	23.84901961	27.4207547
Credit Risk Score 600 - 624	28.56428571	24.34492754	24.8378378
Credit Risk Score 625 - 649	27.20877193	27.764	23.8226415

IBM Information Server Business Glossary Anywhere

Business

Open in Browser

JK Wealth »

**Line of Business**

**Short Description:** JK Life & Wealth division

**Long Description:** BANK1 and insurance are called divisions of JK Life & Wealth. A division is responsible for the creation of the organization's products and/or services. It is not IT, HR and Accounting

**Steward:** Data Analyst (Lt. Commander) Data Soong

**Defined by:** Undefined

## Integrierte Metadaten:

Browser-basierter Zugriff auf die im Information Server Repository gespeicherten Business-Metadaten, um die Bedeutung von KPIs oder Geschäftsbegriffen ohne Wechsel des Tools anzuzeigen.

IBM InfoSphere Business Glossary Browser

Welcome to JKLW Business Glossary

Search • Category Tree • Stewards • About • Help • Log Out

Term Details [Print](#) [Feedback](#)

JK Wealth » Line of Business

**Line of Business**

**Short Description:** JK Life & Wealth division

**Long Description:** BANK1 and insurance are called divisions of JK Life & Wealth. A division is responsible for the creation of the organization's products and/or services. It is not IT, HR and Accounting

**Steward:** Data Analyst (Lt. Commander) Data Soong

**Defined by:** Undefined

**Status:** Accepted

**Created By:** Isadmin  
**Created On:** 2008-12-04 18:56:30  
**Modified By:** picard  
**Modified On:** 2008-12-10 13:46:56

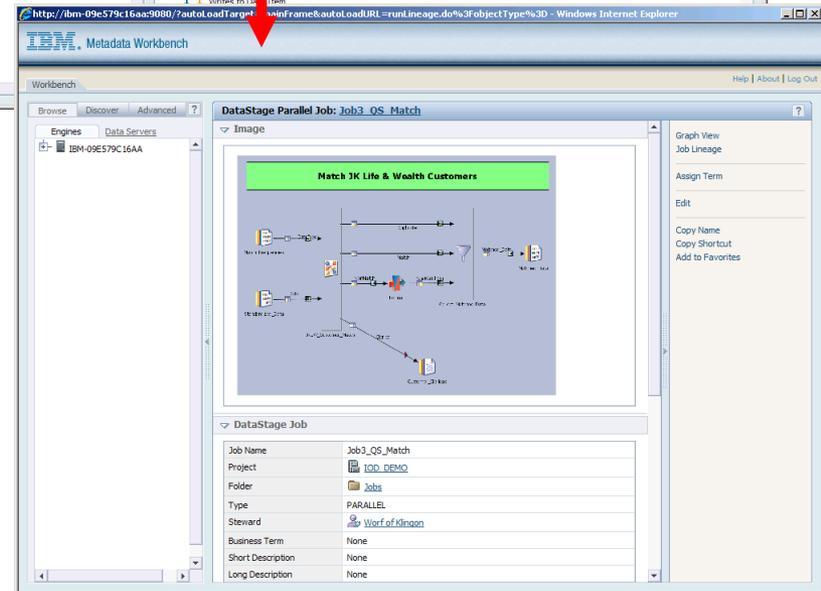
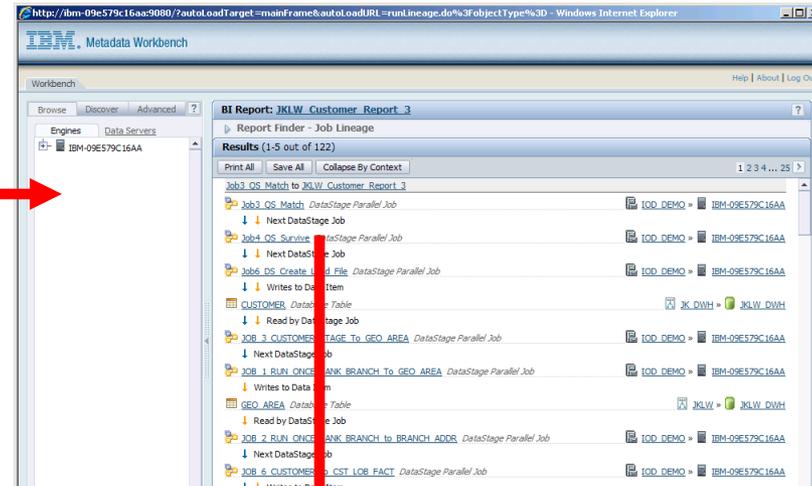
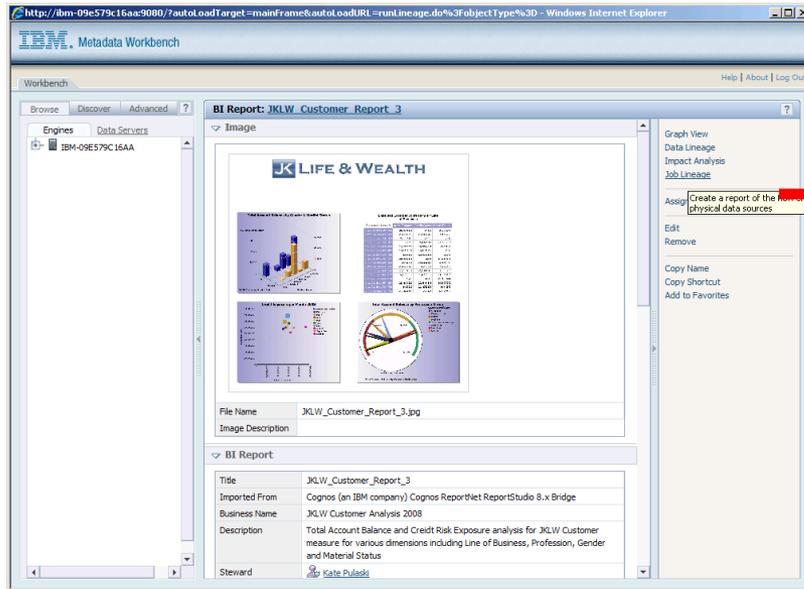
**General Information**

**Related IT Assets**

- JK\_BANK1
- IBM-09E579C16AA
- JKLW\_DB

Notes

# Integration mit Cognos: Rückverfolgung der BI-Zahlen über ein detailliertes Lineage (Verkettung von ETL-Abläufen)



- Details des BI Reports
- Übersicht des Berechnungs-Ablaufs
- Teilschritte mit Laufzeitinformationen

# Die Stärke der IBM (speziell bezogen auf Information Server)

## Technologisch

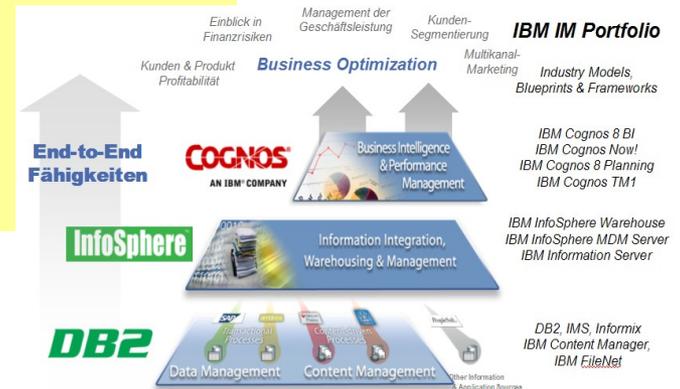
- ◆ Information Server liefert “Trusted Data!” mit integrierten Werkzeugen
- ◆ Datenqualität und Cleansing ist integraler Bestandteil der Suite
- ◆ Unsere Werkzeuge sind integriert. Alle benutzen die gleiche Engine, alle benutzen das gleiche Metadata Repository
- ◆ Wir unterstützen offene Standards und alle kommerziell relevanten Betriebs-Plattformen
- ◆ Performance: Aufgrund der parallelen Engine werden auch größte Datenvolumina bewältigt

## Synergien – Business-Intelligence-Lösungen End-to-End

- ◆ DB2 als skalierbare Datenbanken für Datawarehousing und BI
- ◆ Datenintegration und Migration incl. Datenqualität sowie operationales Stammdatenmanagement mit InfoSphere
- ◆ Cognos als BI-Plattform

## Know-How vor Ort

- ◆ Deutschsprachiger Support
- ◆ Produktspezialisten in Software Group, Consulting und bei Partnern
- ◆ Entwicklung zu großen Teilen in Deutschland (Böblingen)





## IBM Software Partner Academy Program

### Kontakt Daten:

# Christian Lenke



***Christian Lenke***

Consulting IT Specialist

InfoSphere Technical Sales

Phone: 0172/7355302

E-Mail: [clenke@de.ibm.com](mailto:clenke@de.ibm.com)



**InfoSphere.**  
software

**Vielen Dank für Ihre Aufmerksamkeit!**