

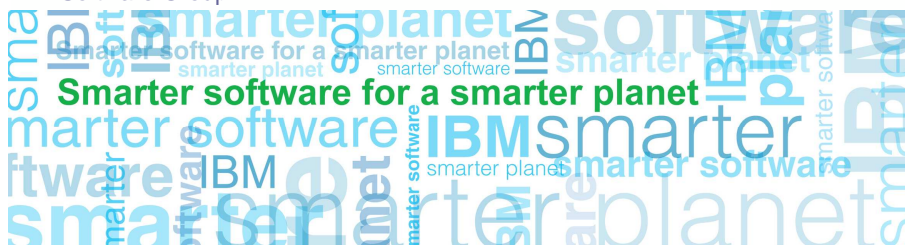


From Zero to z/Hero Workshops 2012

Information Management

High Quality Daten-Management auf nachvollziehbaren Wegen

Fritz Oosterbeek
Client Technical Professional
IBM Software Group



© 2010 IBM Corporation



Ein essenziell wichtiger Punkt bei der Gewinnung von Informationen sind die Qualität der zu Grunde gelegten Quelldaten und die Nachvollziehbarkeit der Datenwege.

Business Analytics und die Kunst Geschäftssteuerung &-optimierung auf Basis qualitativ hochwertiger Daten ist nicht nur die rein grafische Auswertung in einem Balken- oder Kuchendiagramm. Business Analytics ist vielmehr ein Prozess bei dem zu jeder Zeit nachvollziehbar sein muss, wie Datenaggregate & Auswertungen entstanden sind bzw. im Umkehrschluss welche Folgen die Änderung einer Tabellenstruktur zur Folge hat.

Erfahren Sie in dieser Session, wie Sie

- mit IBM Tools auch in Zeiten enormen Datenwachstums konstant hohe Datenqualität sicherstellen,
- Nullwerte filtern sowie Redundanzen unterbinden können
- und darüber hinaus mit einem lückenlosen Metadatenmanagement den Werdegang ihrer Kennzahlen von der Datenquelle bis zum Report nachvollziehen können.

© 2012 IBM Corporation

Disclaimer

© Copyright IBM Corporation [current year]. All rights reserved.
U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.

IBM, the IBM logo, ibm.com, DB2, InfoSphere, Cognos, and InfoSphere Warehouse on System z are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml

Other company, product, or service names may be trademarks or service marks of others.

© 2012 IBM Corporation

Die Menge der Informationen explodiert!



Quelle: Diverse IBM und öffentlich zugängliche Studien

© 2012 IBM Corporation

Entscheider und deren heutige Herausforderungen



1 von 3

Geschäftsführern trifft **kritische Entscheidungen auf Basis nicht valider** oder fehlender Informationen

1 von 2

Geschäftsführern sagt **er habe nicht den Zugang zu** den für seinen Job notwendigen Informationen

19+ Stunden

werden wöchentlich zum **Suchen und Verstehen** von benötigten Informationen aufgewendet

5

bration

Der Druck wächst...

Nie zuvor standen Organisationen unter so grossem Druck, den Wert von Informationen wirksam zu nutzen, um intelligente Entscheidungen zu treffen

Menge digitaler Daten

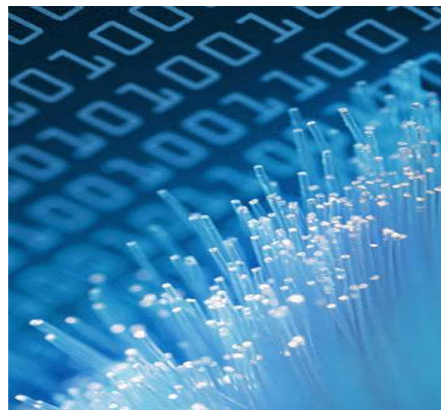
- Verzehnfachung der Datenmenge in den kommenden drei Jahren erwartet
- Automatisiert generierte Daten: Sensoren, RFID, GPRS

Vielfalt der Informationen

- 80% des Datenwachstums sind unstrukturierte Daten
- Emails, Bilder, Audio, Video..

Geschwindigkeit von Entscheidungen

- Dramatisch schnelle Änderungen am Geschäftsklima
- Vorausschauen wird lebensnotwendig: Probleme frühzeitig abschätzen und lösen



Information ist für moderne Unternehmen so wichtig, wie Wasser für das Leben auf der Erde

Rohstoff: Wasser

- Häufig vorkommender Rohstoff, der sogar durch Erderwärmung wächst
- Circa ¾ der Erde sind bedeckt mit Wasser

Trotzdem verdursten Menschen.

Rohstoff: Daten und Informationen

- Überall und immer mehr. Datenflut!
- Es wird eine Verzehnfachung der Daten bis 2011 erwartet (1.800 Exabytes)*

Trotzdem fehlen Unternehmen Informationen, um effizient zu steuern, zu vermarkten und zu entwickeln



» Die bloße quantitative Existenz eines Rohstoffs bringt noch keinen Vorteil ... und wird eventuell sogar zum Problem und unkalkulierbaren Kostenfaktor!

* IDC Whitepaper, March 2008: „The Diverse and Exploding Digital Universe“, An Updated Forecast of Worldwide Information Growth Through 2011

© 2012 IBM Corporation

Die Vision der IBM: Das ‚Information on Demand‘ Ökosystem für einen Smarter Planet



Information

- Keine Dürre, Keine Überflutung
- Rein und klar

On Demand

- Fließt ohne großes Zutun
- Genau passend zur Nachfrage
- Kommt in der erforderlichen Qualität aus der Leitung



» IBM hat mit ‚Information on Demand‘ eine Strategie zum Aufbau zukunftsfähiger und Wert schöpfender Informationsarchitekturen.

© 2012 IBM Corporation

smarter planet 

smarter software for a smarter planet

Funktionierende Infrastruktur ist Voraussetzung für effiziente Versorgung mit frischem Wasser



The diagram illustrates a water supply system using various infrastructure components. It features a central map of Germany with several images placed over it, each labeled with a component of the water infrastructure. The components are: Pumpen (Pumps), Kanalsystem (Canal system), Beregnung (Irrigation), Wasserhähne (Faucet), Quellen (Sources), Filter (Filters), Stauseen (Reservoirs), and Drainagen (Drainages).

» Wie Bewässerungsprojekte letztlich durch den schrittweisen Einsatz von Spitzentechnologie erfolgreich werden ...

© 2012 IBM Corporation

smarter planet 

smarter software for a smarter planet

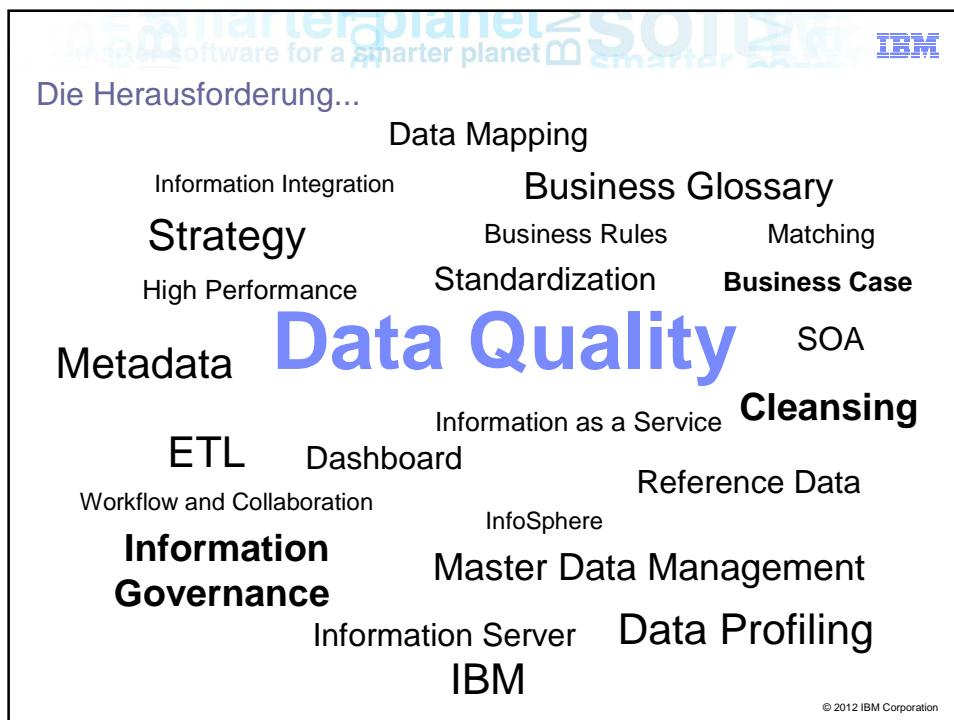
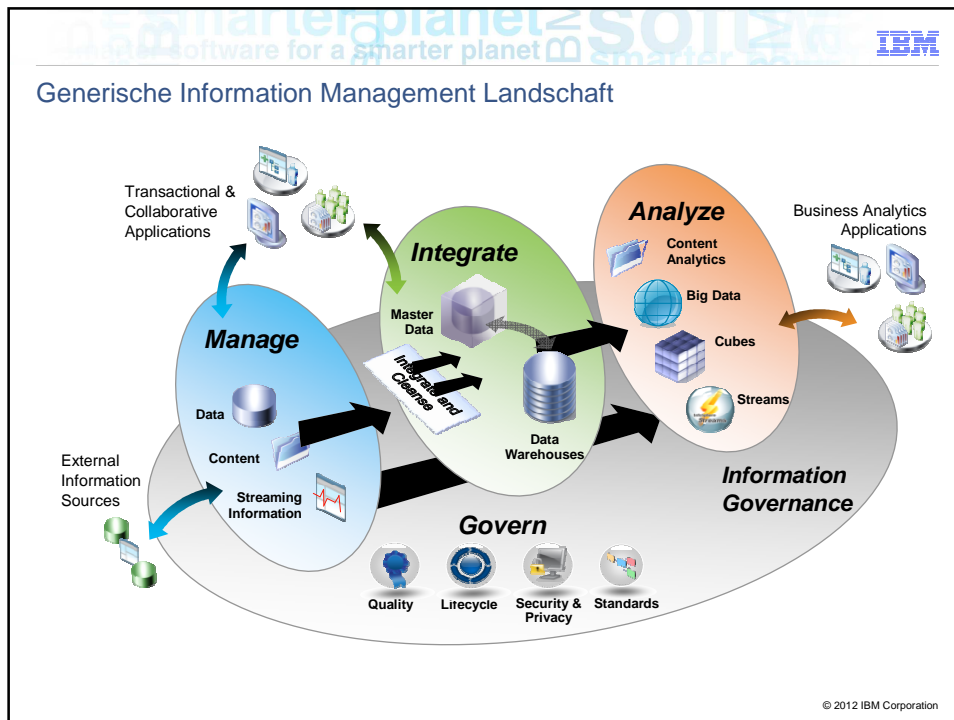
Funktionierende Infrastruktur ist Voraussetzung für effiziente Versorgung mit sauberen Daten



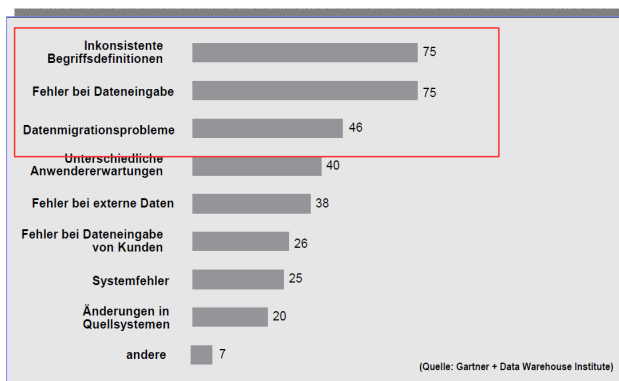
The diagram illustrates a data supply system using various infrastructure components. It features a central map of Germany with several images placed over it, each labeled with a component of the data infrastructure. The components are: ETL Scheduler Message Broker, Datentransport Service Bus, BI-Systeme, Applikationen, Datenbanken, Datenbereinigung Harmonisierung, Master Data Management Content Management, and Archivierung.

» ...so sichern best-in-class Lösungen den dauerhaften Erfolg von Information on Demand durch schrittweise Einführung.

© 2012 IBM Corporation



Die Ursachen für Datenqualitätsprobleme sind vielfältig



13

© 2012 IBM Corporation

Datenqualitätsprobleme haben vielfältige Auswirkungen im operativen Geschäft

Entgangenes Geschäftspotenzial



Compliance; Risiko-Management



Überbordende Kostenflut



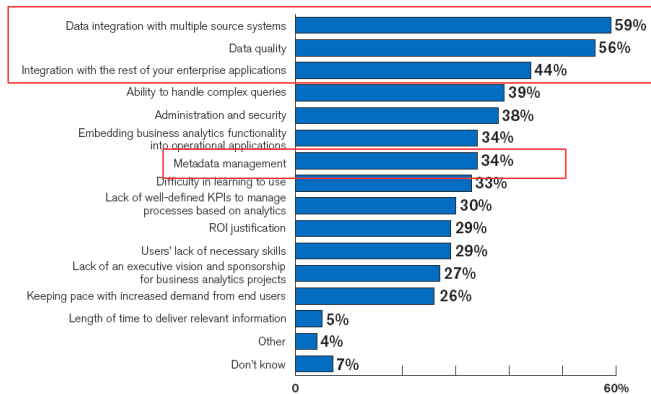
Verfehlte Business Transformation



14

© 2012 IBM Corporation

Wichtigkeit der Datenqualität für das Reporting und somit für die organisatorische Entscheidungsfindung



Note: Multiple responses allowed.

"Defining Business Analytics and Its Impact on Organizational Decision-Making," February 2009, ComputerWorld

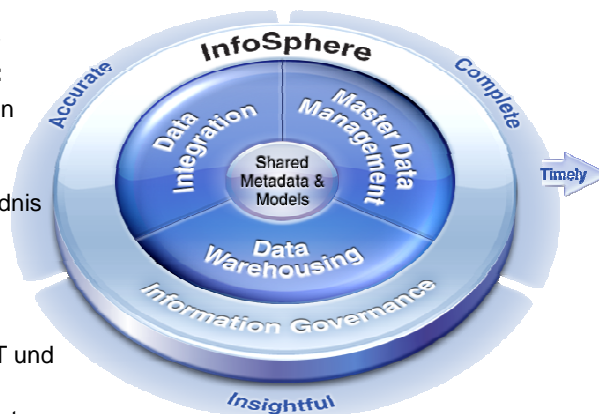
15

© 2012 IBM Corporation

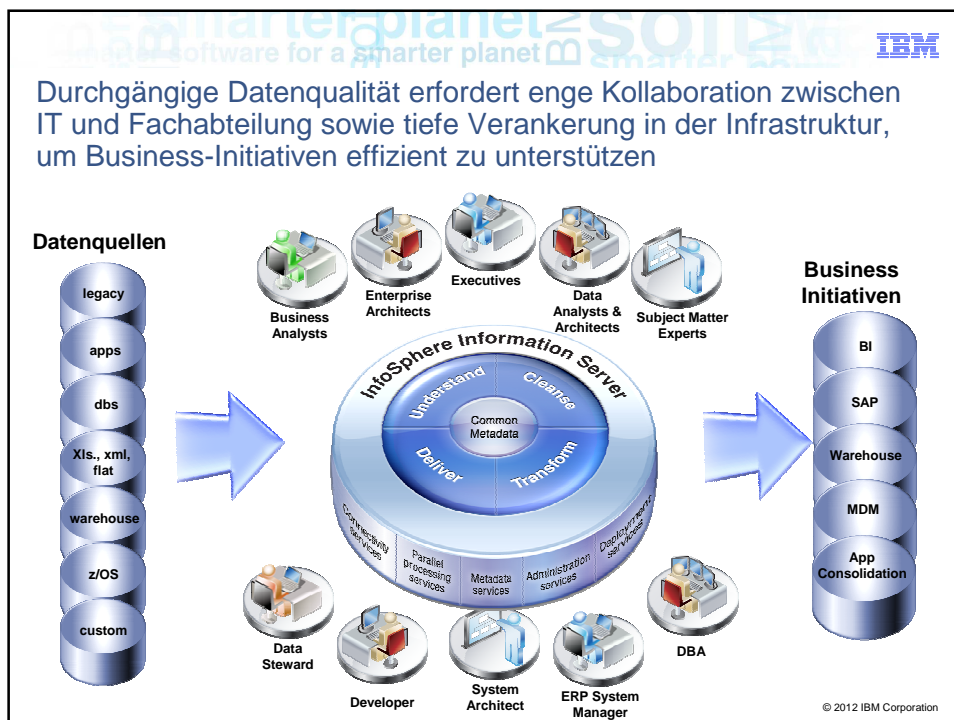
IBM hat die wesentlichen Fähigkeiten einer "Trusted Data" - Architektur im InfoSphere Portfolio zusammengefasst

Schnelle Realisierung der "Business Values" durch:

- Tiefgreifende Integration über die Komponenten hinweg
- Gemeinsames Verständnis von gemeinsamen Bestandteilen
- Kollaborative Elemente fördern effiziente Zusammenarbeit von IT und Fachbereichen
- Vorkonfigurierte Elemente dienen der Beschleunigung in der Umsetzungsphase



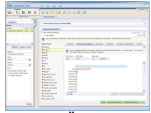
© 2012 IBM Corporation



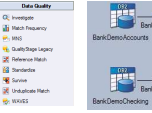
Software for a smarter planet **IBM**

IBM InfoSphere Information Server

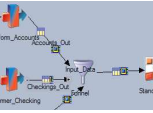
Bereitstellung von zuverlässigen Informationen



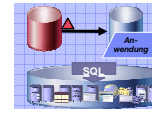
Erkennen / Überwachen von Informationsstruktur /-inhalt




Standardisieren / Korrigieren von Informationen




Kombinieren / Restrukturieren von Informationen




Synchronisieren / Virtualisieren von Informationen




Verstehen



Bereinigen




Transformieren




Liefern

Plattform Services




Parallelität

Variable Parallelisierung
Festlegung zur Laufzeit




Verbindung

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




Metadaten

Geschäfts-Definitionen
Abhängigkeits-Analysen



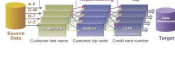
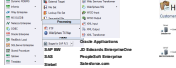


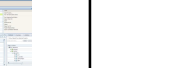
Administration

Admin-Konsole
Report-Konsole



Deployment

Services /
Spezifikation








© 2012 IBM Corporation


Software for a smarter planet **IBM**

Verstehen


Analyse der Quelldaten und Qualitätsregeln




Verstehen



Bereinigen




Transformieren




Liefern


Plattform




Parallel




Verbindung



Metadaten



Admin



Deployment

Analysieren der Quelldaten


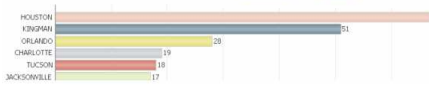
- Analysieren, Verstehen und Reports
- Generierung von Metadaten


Schrittweise Analyse

- Spalten (Datentypen, Füllungsgrad)
- Primary Key (eindeutige Schlüssel)
- Foreign Key (Abhängigkeiten)
- Tabellen übergreifend (Abhängigkeiten)

Überwachung der Qualität

- Definition von Regeln, z.B.
 - BIRTH_DATE ist ein Datum
 - BIRTH_DATE > 1900-01-01
 - PROFESSION = 'STUDENT' AND AGE < 40
- Historie der Ergebnisse



© 2012 IBM Corporation

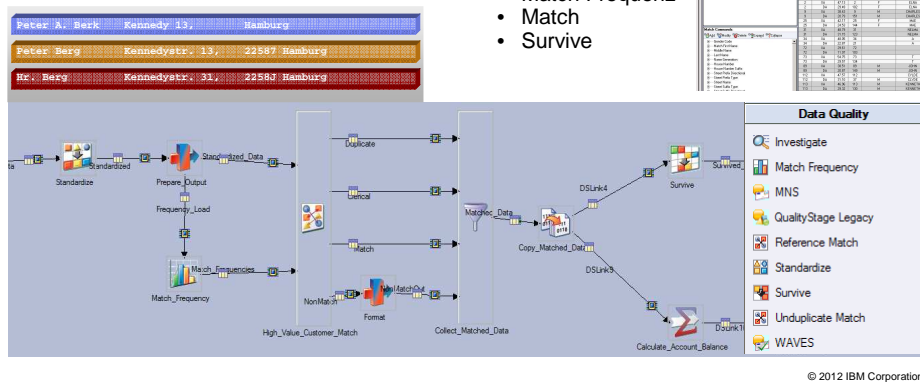
Bereinigen

Erhöhte Datenqualität

- Bereinigung der Qualität
 - Unerwartete / versteckte Daten
 - Fehlende Daten
 - Duplikate

Ablauf integriert in Transformations-Funktionalität

- Untersuchung
- Standardisierung
- Match Frequenz
- Match
- Survive

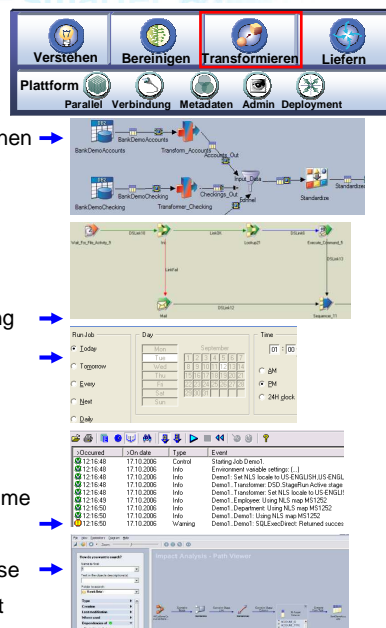


© 2012 IBM Corporation

Transformieren

Extrahieren, Transformieren und Laden

- Transformation: Umfassende vordefinierte Funktionen →
- Performance: Parallelisierung und Skalierbarkeit
- Anbindung: Unterschiedlichste Quellen / Ziele
- Wiederverwendbarkeit: Parametrisierung und Kapselung
- Sequenzen: Prozessabläufe / Fehlerbehandlung →
- Scheduling: Definierte Termine und Zeiten →
- Latenzzeit: Batch, Echtzeit und Services
- Reports: Prozessabläufe / automatisiert
- Standards: Information Services / SOA
- Erweiterbarkeit: C++, Java, Kommandos, Programme →
- Überwachung: GUI mit Monitor / Log / Grafik →
- Metadaten: Übergreifende Information / Analyse →
- Qualität: Analyse und Bereinigung integriert



© 2012 IBM Corporation

Software for a smarter planet

Liefen (Föderieren / Synchronisieren)

Verfügbarkeit aktuellster Daten

Föderation:

Eine grosse virtuelle Datenbank

- Zugriff auf unterschiedliche und verteilte Informationen, als befänden diese sich in einem einzigen System
- Echtzeitzugriff von Analyseanwendungen auf integrierte Informationen

Synchronisierung:

Konsistente Informationen

- Abgleich zwischen Datenbanken, Hochverfügbarkeit, Migration
- Datenänderungen starten Anwendungen

© 2012 IBM Corporation

Software for a smarter planet

Von Daten zu Informationen

Beispiel Kundendaten

Verstehen

Kardinalität, Frequenz, Abhängigkeiten (Cross)

4711	Peter	Berk	22587	Hamburg	Kennedy 13
0000	Peter	Berg	22587	Hamburg	Kennedystr.
0815		Berg	22587	Hamburg	Kennedystr.

Bereinigen

Duplikate, Referenz Match, Selektion

4711	Peter	Berg	22587	Hamburg	Kennedystr.	13
------	-------	------	-------	---------	-------------	----

Transformieren

Joins, Aggregation, Geschäfts-Regeln

4711	13.01.2009	13 342
4711	21.12.2008	4 375
0815	15.09.2009	5 228 769

Liefen

Transparenter Zugriff, Echtzeit-Daten

4711	07.05.2012	4 925 528
------	------------	-----------

Analysieren der Quelldaten

Analysieren, Verstehen und Reports
Quellen-übergreifend

Verbesserung der Qualität

Fehlende Informationsstandards
Fehlende Daten / Duplikate

Extraktion, Transformation, Laden

Daten → Geschäftsinformationen
Vordefiniert, wiederverwendbar

Aktuellste Daten

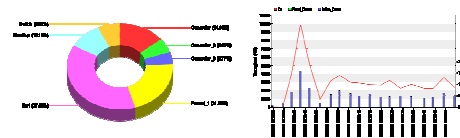
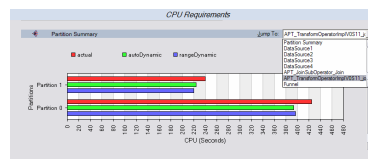
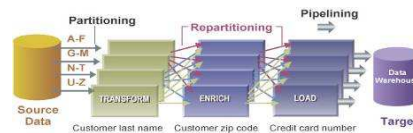
Echtzeitzugriff durch Föderation
Konsistente Daten durch Replikation

© 2012 IBM Corporation

Performance und Skalierbarkeit

Parallelität und Skalierbarkeit

- Automatische Partitionierung durch native parallele Engine, Daten werden im Hauptspeicher gehalten
- Entwicklung ohne Festlegung der Parallelisierung
- Definition des Parallelisierungsgrads zur Laufzeit



© 2012 IBM Corporation

Performance

- Performance Analyse: Schrittweise Anzeige von Zeit- und Ressourcenverbrauch für Bereinigung und Transformation
- Resource Estimation: Abschätzen der Performance von Datenflüssen zur Bereinigung und Transformation basierend auf Beispieldaten und Metriken

Anbindung aller wichtigen Quellen und Ziele



	Database	File	Real Time	Packs
Datenbanken	Classic Federation DB2 UDB DB2/UDB Enterprise Dynamic RDBMS Informix iWay Enterprise MS OLEDB Netezza Enterprise ODBC Oracle Enterprise RedBrick Load SQLServer Enterprise Stored Procedure Sybase Teradata	Complex Flat File Data Set External Source External Target File Set Lookup File Set Sequential File <div>Processing</div> FTP Web Sphere TX Map	Java Client Java Transformer Web Services Client Web Services Transformer WebSphere MQ WISD Input WISD Output XML Input XML Output XML Transformer	<div>Real Time Stages for SAP R/3</div> SAP R/3 SAP BW SAS Oracle Applications Siebel JD Edwards EnterpriseOne PeopleSoft Enterprise Salesforce.com Essbase

© 2012 IBM Corporation

Metadaten

Angepasste Sichten auf ein zentrales Repository

Business Metadaten

- Abgleich von IT und Business
- Business Kontext für IT Objekte
- Zuständigkeit und Verantwortlichkeit
 - Geschäftsmetadaten
 - Datenbankobjekte
 - Analysemetadaten

Logische Metadaten

- Daten Modellierung / Abhängigkeiten

Erkunden der Metadaten

- Analysieren, Verstehen und Reports
- Metadaten austausch mit anderen Tools
- Abhängigkeitsanalyse, Datenfluss



© 2012 IBM Corporation

Administration

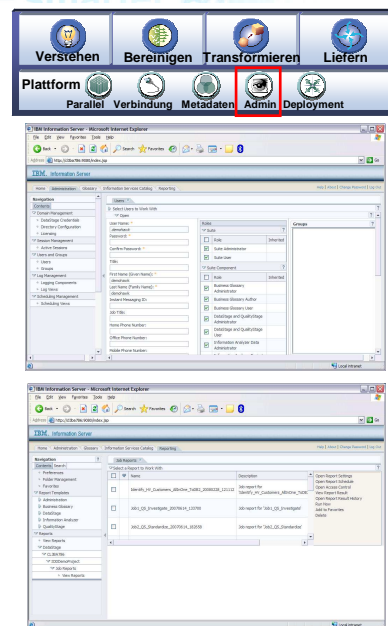
Verwaltung und Berichte

Administrations Konsole

- Domain
- Session
- Users / Groups
- Log
- Scheduling

Reporting Konsole

- Templates
- Reports



© 2012 IBM Corporation



Deployment

Informations-Services




Schnelle SOA Entwicklung

- Infrastruktur für Informationsservices
- EJB, JMS oder Web-Services, REST, RSS
- Bereinigen, Transformieren und Föderieren





© 2012 IBM Corporation



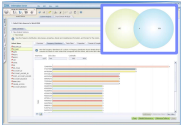
Zuverlässigkeit durch Metadaten

Zentrales Repository






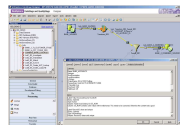
Daten-Analyst




Analyse / Überwachung
(Verteilung / Qualität)




Qualität
(Standards, Duplikate)




Transformationen
(Abhängigkeit, Analyse)




Entwickler



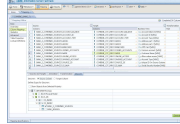
Endanwender




Geschäftsbegriffe
(Kontakt, Beziehungen)




Zentrales Repository
(intern und extern)




Spezifikation
(Spreadsheet => ETL)




Spezialist Fachabteilung



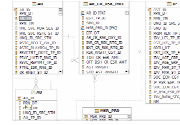
Metadaten-Analyst




Nachvollziehbarkeit
(Abhängigkeit, Laufzeit)



Externe Tools
(Business Intelligence)

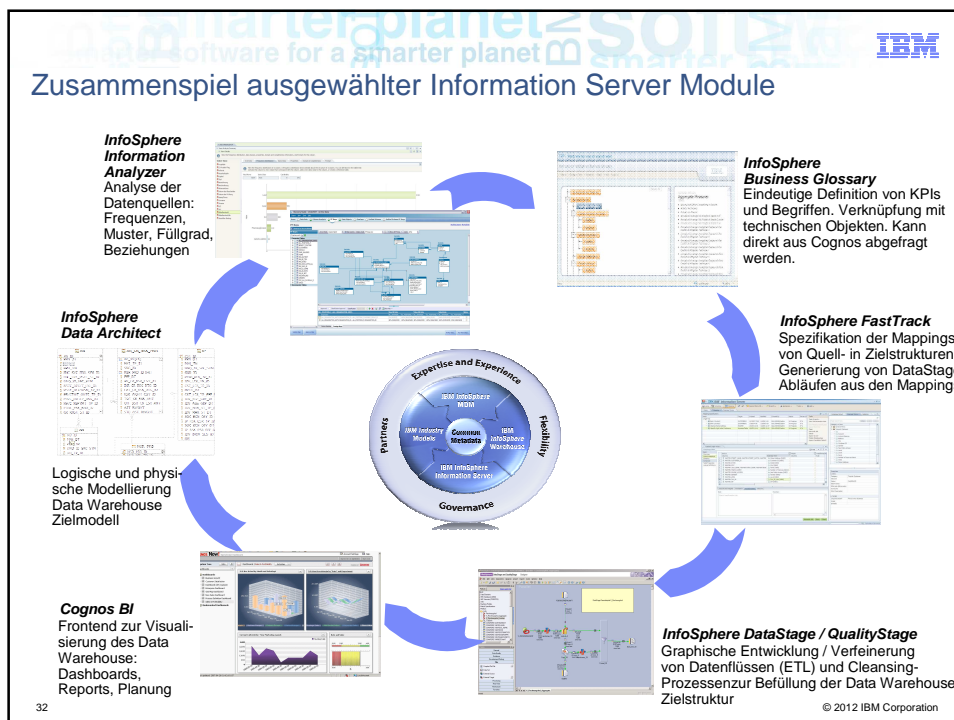
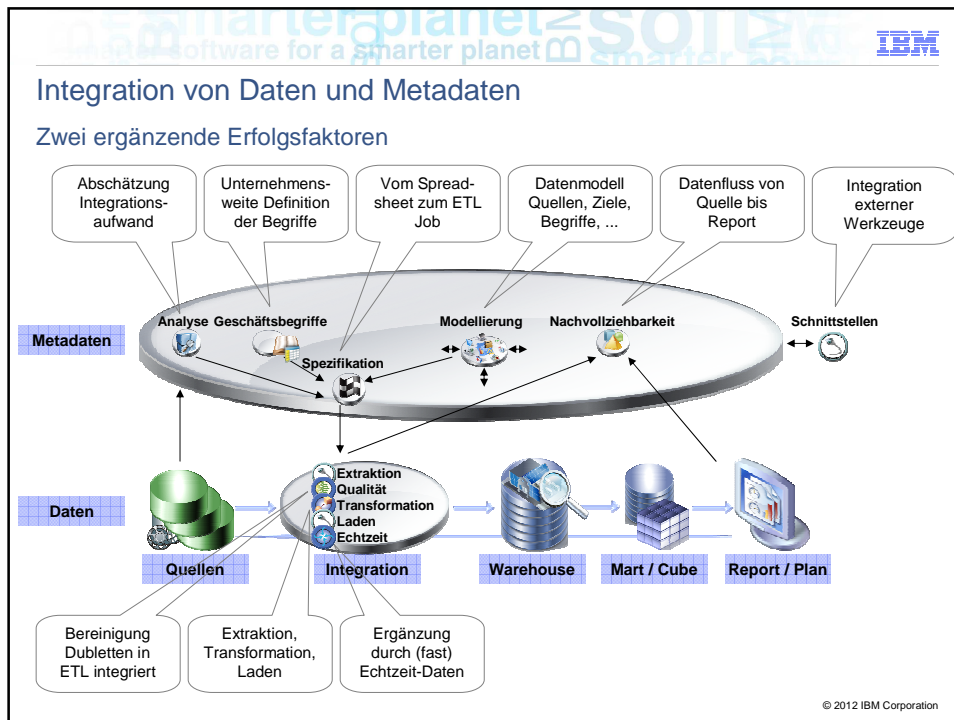


Modellierung
(Entitäten, Beziehung)



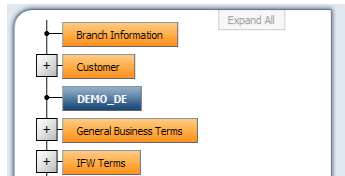
Modellierer

© 2012 IBM Corporation



Business Glossary

Glossary Terms und Kategorien, Verlinkung zu IT-Assets



CATEGORY DETAILS:
DEMO_DE
Alle Terme die deutsche Information Server Demo

TERMS:

- Kontotyp
- Kundennummer
- Kundenstatus

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.

Assigned Assets

1. KDSTATUS (IBM-01CEFA5FF9 > SAMPLE > DEMO_DE > KDSTATUS)
2. KSTATUS (IBM-01CEFA5FF9 > SAMPLE > DEMO_DE > KUNDE_TRUSTED)

Term Details Print Feedback

DEMO_DE >

Kundenstatus

Short Description: Status eines Kunden
Long Description: Status eines Kunden aufgrund seines Kontosaltdos
Stewards: Undefined
Assigned to terms: Undefined
Status: Candidate

General Information

Abbreviation — KDSTATUS KSTATUS
Example — ENTRY, SILVER, GOLD, PLATINUM
Created By — isadmin
Created On — 2009-07-31 16:41:15
Modified By — isadmin
Modified On — 2009-08-03 16:38:37

33

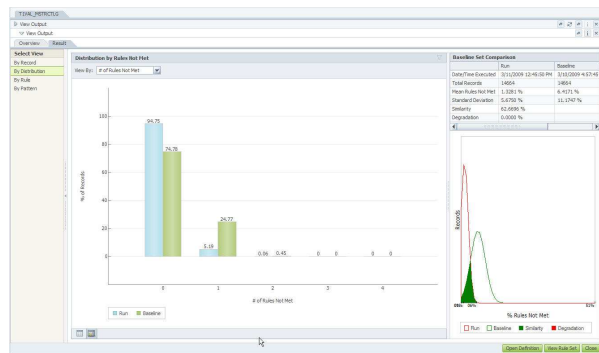
© 2012 IBM Corporation

Datenprofiling mit dem InfoSphere Information Analyzer



InfoSphere Information Analyzer

'Menschwürdige' Annäherung an die Daten durch Datenprofiling sowie Messung der Datenqualität über die Zeit



34

Requirements

- Durchführung von 'Data Quality Assessments'
- Definition von Regelwerken - Aufbereitung und Überwachung der Datenqualität

Benefits

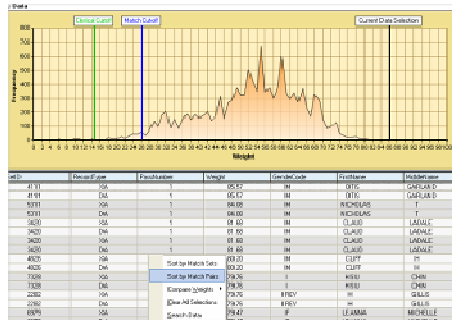
- Frühzeitige Identifikation der herrschenden Datenqualität - Minderung von Projektrisiken
- Überwachung der DQ über die Zeit - Einhaltung der DQ Metriken
- Einigkeit über und Vertrauen in die Daten herstellen
- Analyseergebnisse stehen allen InfoSphere Komponenten für die Weiterverarbeitung zur Verfügung

Datenaufbereitung mit InfoSphere QualityStage®



InfoSphere QualityStage

Standardisierung und hochqualitative
Aufbereitung der Daten mit dem Ziel
der Herstellung einer 'Ein-
Eindeutigen' Datenbasis



Requirements

- Bedarf nach hochqualitativen Daten
- Standardisierung von Formaten
- Datenbereinigung
- Dublettenauffindung und Elimination
- Datenqualität für Realtime-Anwendungen

Benefits

- Entfernt Dublikate
- Spürt vergleichbare Datensätze auf
- Automatische Herstellung des 'Golden Record'
- Validierung und Anreicherung der Daten

35

© 2012 IBM Corporation

InfoSphere and Cognos

InfoSphere bildet die Basis für vertrauenswürdige Informationen als Grundlage für
Cognos Business Intelligence & Performance Management Lösungen

InfoSphere™
software



Cognos.
software

- **Business Metadata sharing**
- **End-to-end Data Lineage**
- **Data Profiling and Quality integration**
- **InfoSphere DataStage with Cognos Data Manager**
- **InfoSphere Change Data Capture with Cognos Now!**

© 2012 IBM Corporation

IBM Cognos BI - Lineage

*Making information
...usable!*

Data Lineage – das “wer, was, wann, wo und warum” –
Ist z. B. ein Hauptaugenmerk im Finanzwesen

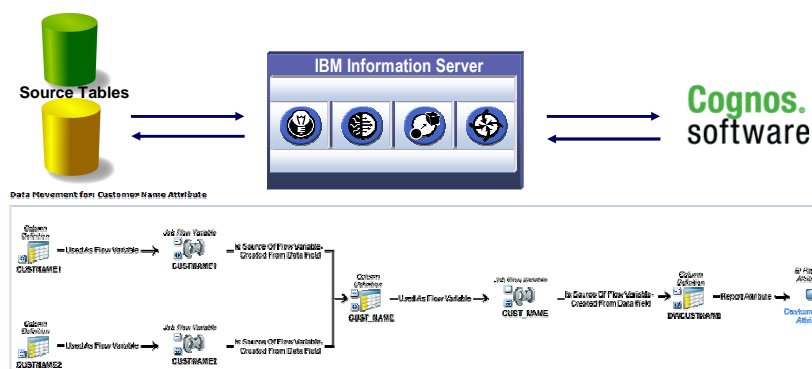


- **Lineage aus technischer Sicht:**
 - Tracking der Daten – z. B. Ursprungs-Applikation / System, Datum, Uhrzeit
 - Welche Änderungen wurden vorgenommen
 - Wer ist/war verantwortlich für Zugriff und Verarbeitung der Daten
- **Lineage aus geschäftlicher Sicht:**
 - Beschreibt auf Metadaten-Basis, wo Information herkommen
 - Beinhaltet Informationen über Autor, Business-Beschreibung des Models

© 2012 IBM Corporation

End-to-End Data Lineage mit Cognos Woher kommen die Werte in einem Report?

- Import & Browse Full BI Report Metadata
- Navigate through report attributes
- Visually navigate through data lineage across tools
- Combines operational & design viewpoint

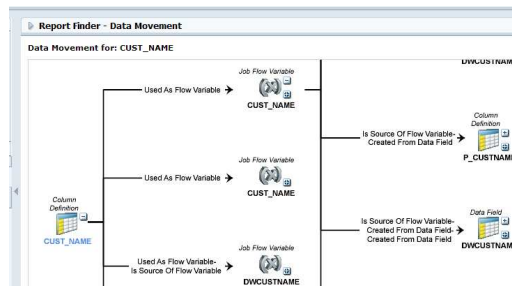


© 2012 IBM Corporation

End-to-End Data Lineage mit Cognos

Welche Auswirkungen hat die Änderung einer Spalte?

- Show complete change impact in graphical or list form
- Includes impact on reports in BI tools
- Visually navigate through impacted objects across tools
- Allows impact analysis on any object type

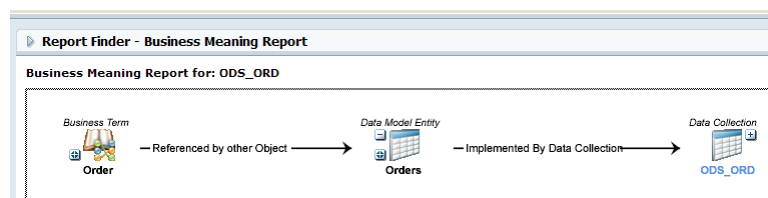


© 2012 IBM Corporation

End-to-End Data Lineage mit Cognos

Welche Bedeutung hat ein Feld?

- Show relationships between business terms, data model entities, and technical and report fields
- Provides cross-tool mapping of business terms
- Allows field meaning to be understood
- Allows business term relationships to be understood

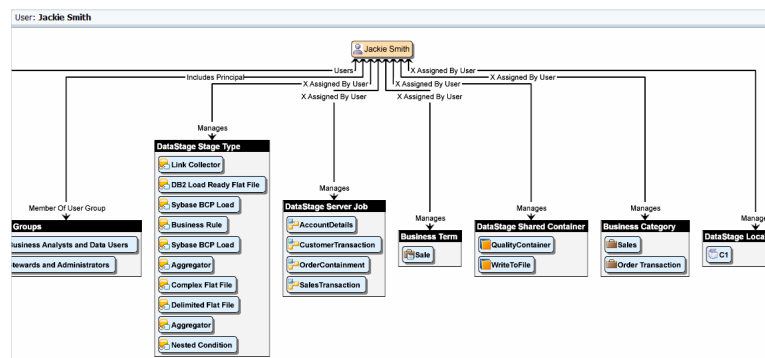


© 2012 IBM Corporation

Welche Objekte sind diesem Benutzer zugeordnet?

Welche Objekte sind diesem Benutzer zugeordnet?

- Shows objects that user manages
- Shows stewardship relationships on business terms
- Shows user group associations

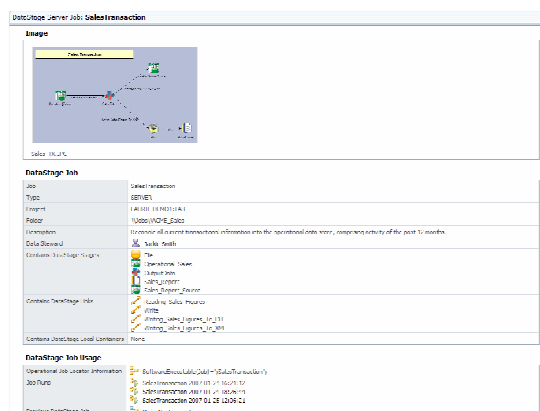


© 2012 IBM Corporation

Was geschah bei der letzten Durchführung eines Jobs?

Was geschah bei der letzten Durchführung eines Jobs?

- *Navigation through complete job details*
- *Navigation of complete operational metadata*

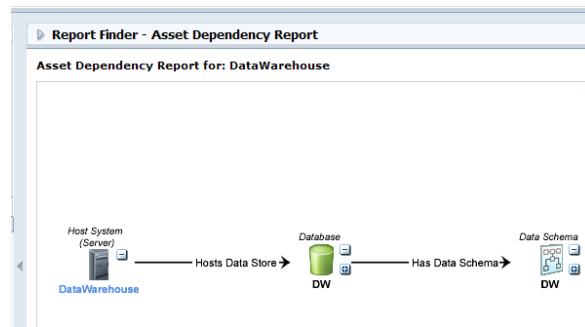


© 2012 IBM Corporation

End-to-End Data Lineage mit Cognos

Welche Systeme sind von Änderungen betroffen?

- Visibility into databases on each host system
- Know who to notify and coordinate upgrades with for system downtime



© 2012 IBM Corporation

Die Vorteile einer zentralen IBM Data Warehouse Infrastruktur

- Single Version of the Truth
 - Konsistente, historisierte Datenbasis
 - flexibel, skalierbar und offen für zukünftige / zusätzliche Anforderungen
 - Transformation und Vereinheitlichung der Daten an einer Stelle
 - Zentral definierte Kennzahlen mit einheitlicher Bedeutung
- Trusted Data
 - Datenqualität durch technische und fachliche Validierungen
 - Prozeßqualität durch Monitoring und Validierung
 - Zentrale Metadaten und Business Glossary
- Flexible, skalierbare und bewährte Data Warehouse Architektur
 - Schnelle Entwicklung von Reports und Ad hoc Analysen
 - Hohe Wiederverwendung von Stammdaten durch „Conformed Dimensions“
 - offen für zukünftige / zusätzliche Anforderungen
 - Einsatz über (Fach-)Abteilungsgrenzen hinweg

44

© 2012 IBM Corporation

Thank
You