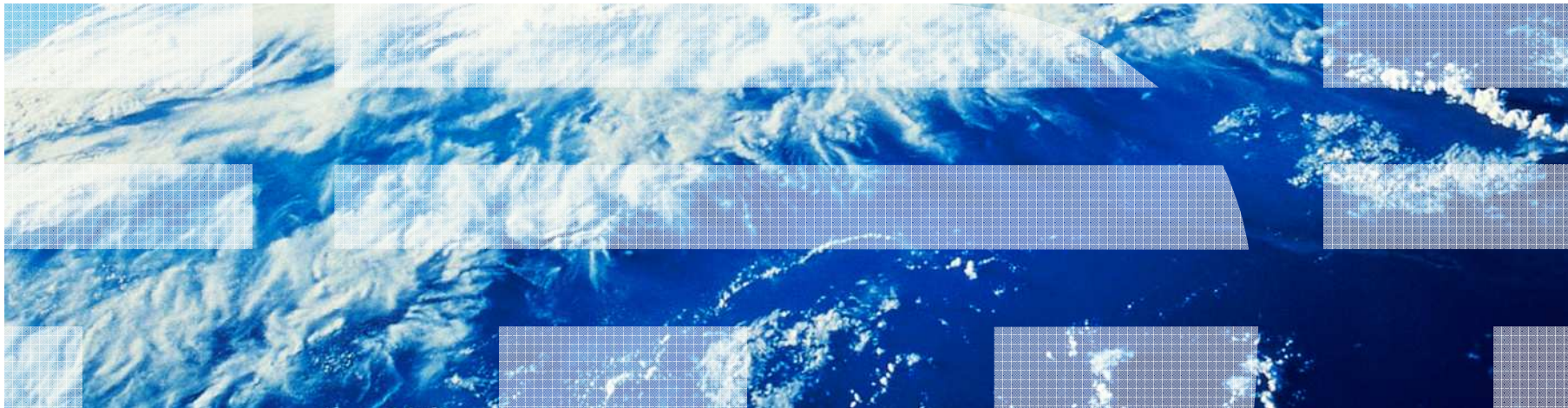


Infosphere Information Server

Trusted Information

- oder -

Können Sie sich auf Ihre Daten verlassen?



Harald Gröger
Consulting IT Specialist InfoSphere

Was genau bedeutet ein Geschäftsbegriff? Unternehmensweite Definition für jeden leicht zugreifbar

Total Account Balance by Gender & Marital Status

| Expected Loss % | Auto Finance | Credit Cards | Home Equity |
|-----------------------------|--------------|--------------|-------------|
| 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 | 23.5130494 |
| 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 |

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

Term Details

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

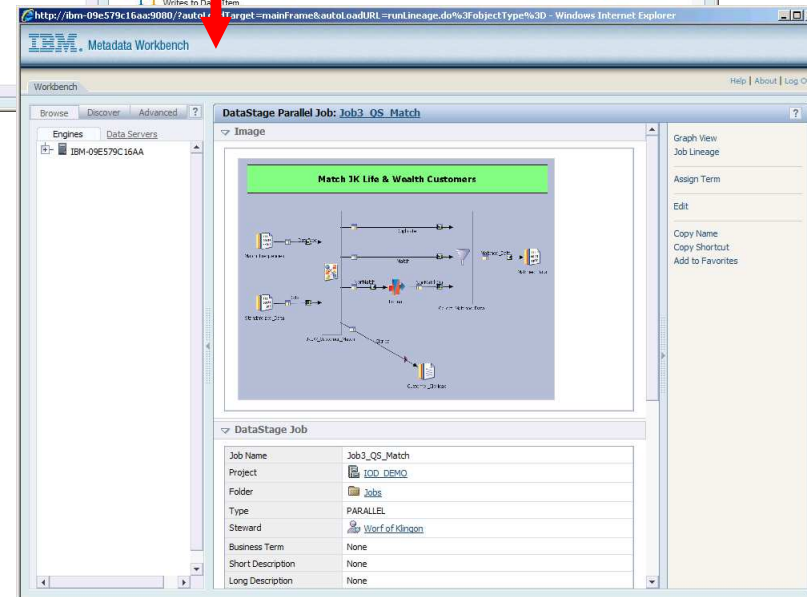
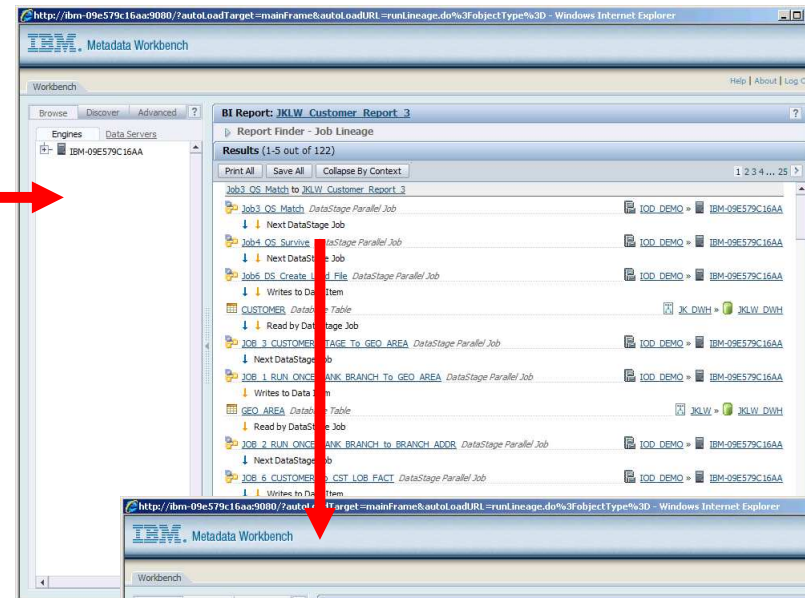
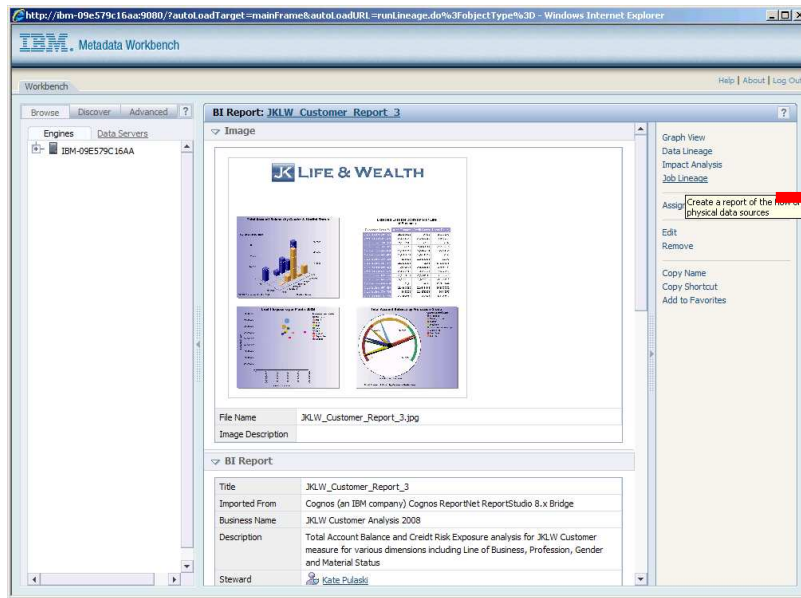
Status: Accepted

Related IT Assets

- JK_BANK1
- IBM-09E579C16AA
- JKLW_DB

- Markieren
- Übersicht
- Details

Stimmen die Zahlen in meinem BI Report? Zurückverfolgung der BI Zahlen über alle Teilschritte



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

Metadata Workbench

Information On Demand

Informationen sind strategisch für Geschäfts-Optimierung



- Information in Unternehmensprozesse nahtlos integrieren
- ‚Echtes‘ und korrektes Wissen für Entscheidungen anwenden

A dark blue banner with a background of financial data. On the left, the word 'COGNOS' is written in large, red, 3D-style letters. To its right, the text 'Business Intelligence & Performance Management' is written in white, sans-serif font.

- Daten standardisieren, bereinigen und integrieren
- Daten unternehmensweit effizient bereitstellen und verwalten

A dark blue banner with a background of a person working at a computer. On the left, the word 'InfoSphere' is written in large, green, 3D-style letters. Below it, the phrase 'Trusted Information' is written in a smaller, white, italicized font. To the right, the text 'Information Integration, Warehousing, & Management' is written in white, sans-serif font.

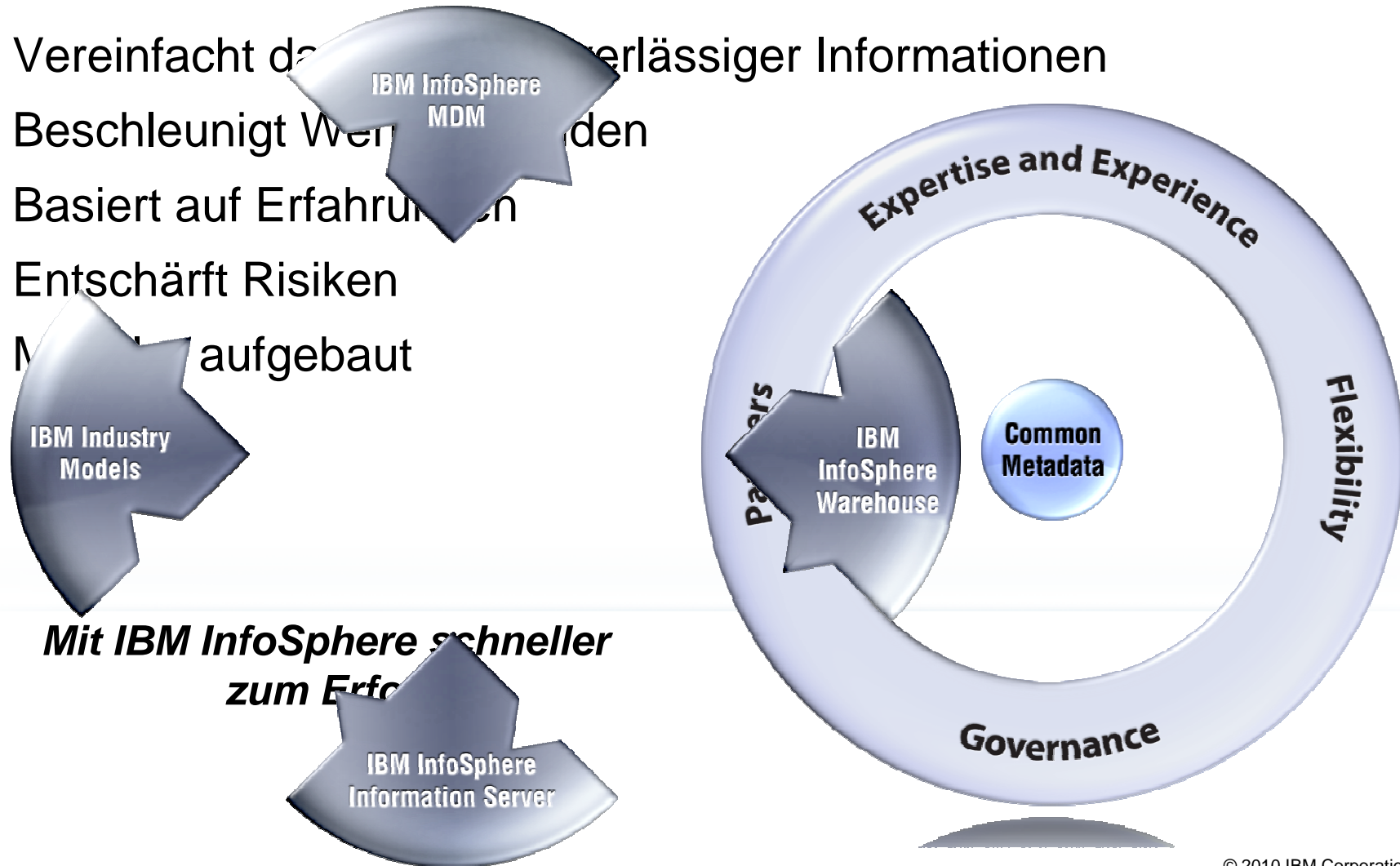
- Neue Quellen schneller erschließen
- Daten und Dokumente sicher und lückenlos vorhalten

A dark blue banner with a background of server racks. On the left, the text 'Data Management' is written in white, sans-serif font, with an icon of a server rack above it. On the right, the text 'Content Management' is written in white, sans-serif font, with an icon of a server rack and a document above it.

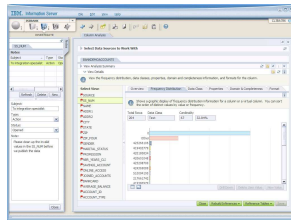
IBM InfoSphere™

Umfassende Informations-Plattform

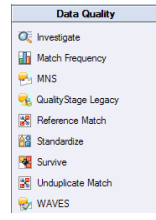
- Vereinfacht das Auffinden von zuverlässiger Informationen
- Beschleunigt Wert aus Daten
- Basiert auf Erfahrungen
- Entschärft Risiken
- Mit MDS aufgebaut



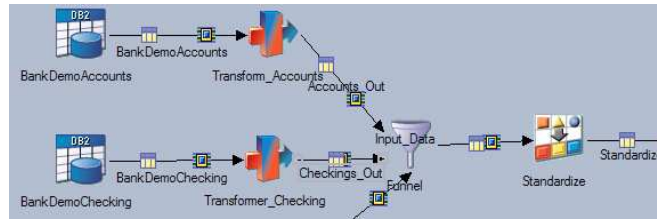
IBM InfoSphere Information Server Übersicht Module



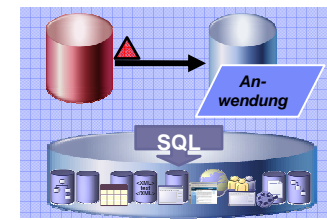
Erkennen / Überwachen von Informationsstruktur /-inhalt



Standardisieren / Korrigieren von Informationen



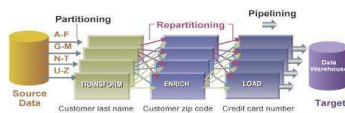
Kombinieren / Restrukturieren von Informationen



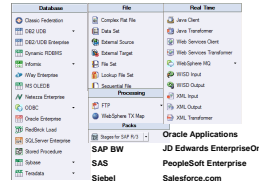
Synchronisieren / Virtualisieren von Informationen

| | | | |
|---------------------------|-----------------------|---------------------------|---------------------------|
| Verstehen | Bereinigen | Transformieren | Liefern |
| Plattform Services | | | |
| Parallelität | Verbindung | Metadaten | Administration |
| Deployment | | | |

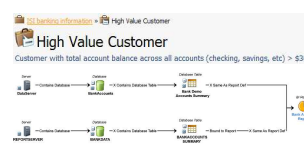
Variable Parallelisierung
Festlegung zur Laufzeit



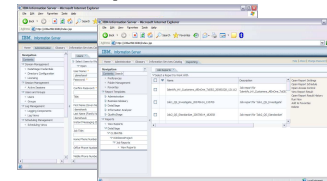
Datenbanken, Files, Echtzeit, ERP, CRM, ...



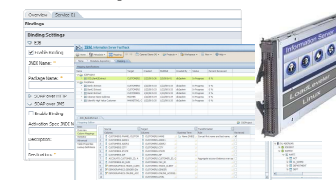
Geschäfts-Definitionen
Abhängigkeits-Analysen



Admin-Konsole
Report-Konsole

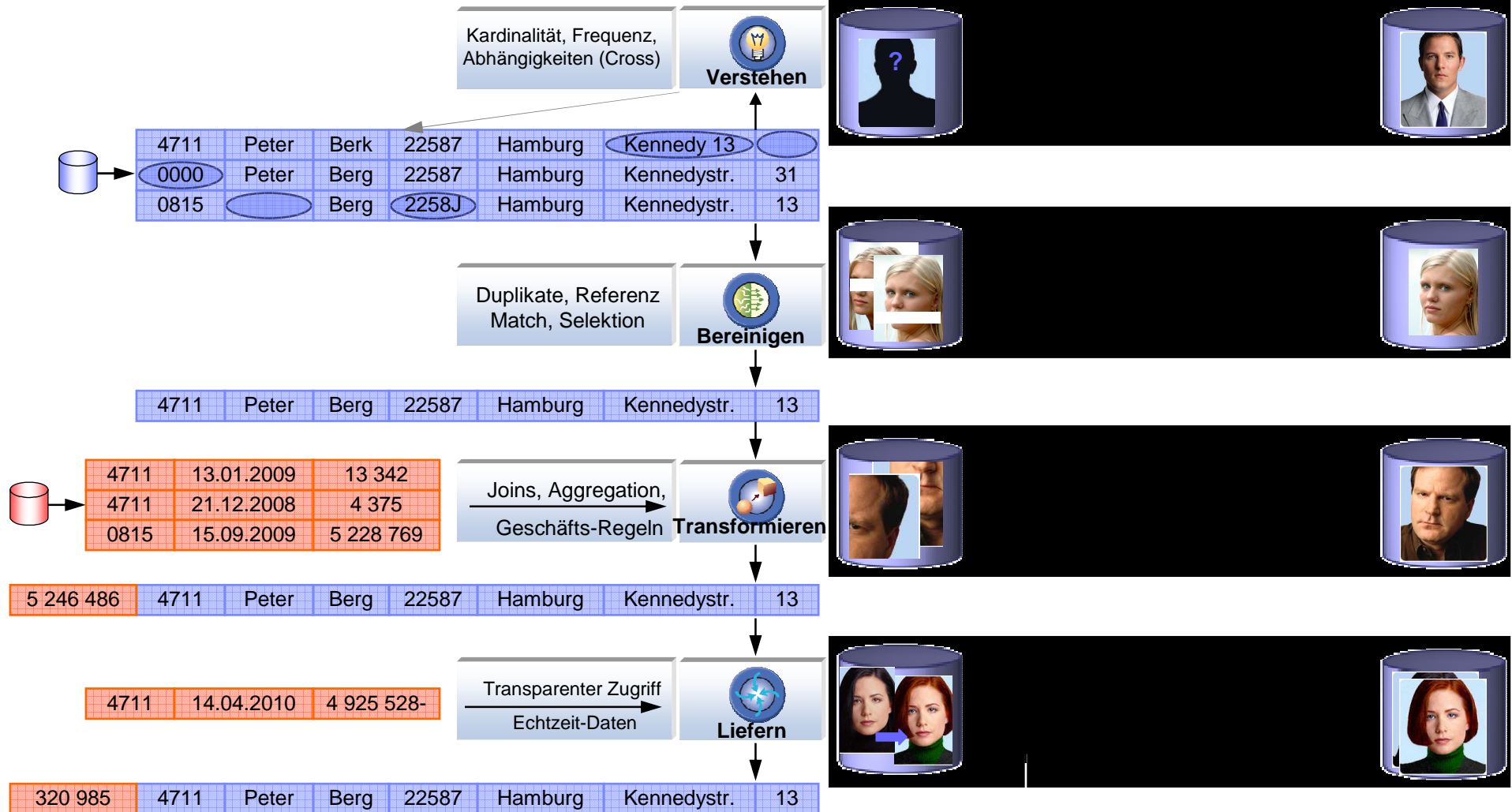
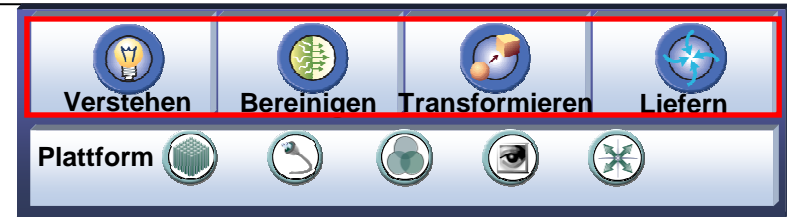


Services / Blades / Spreadsheets



Datenintegration

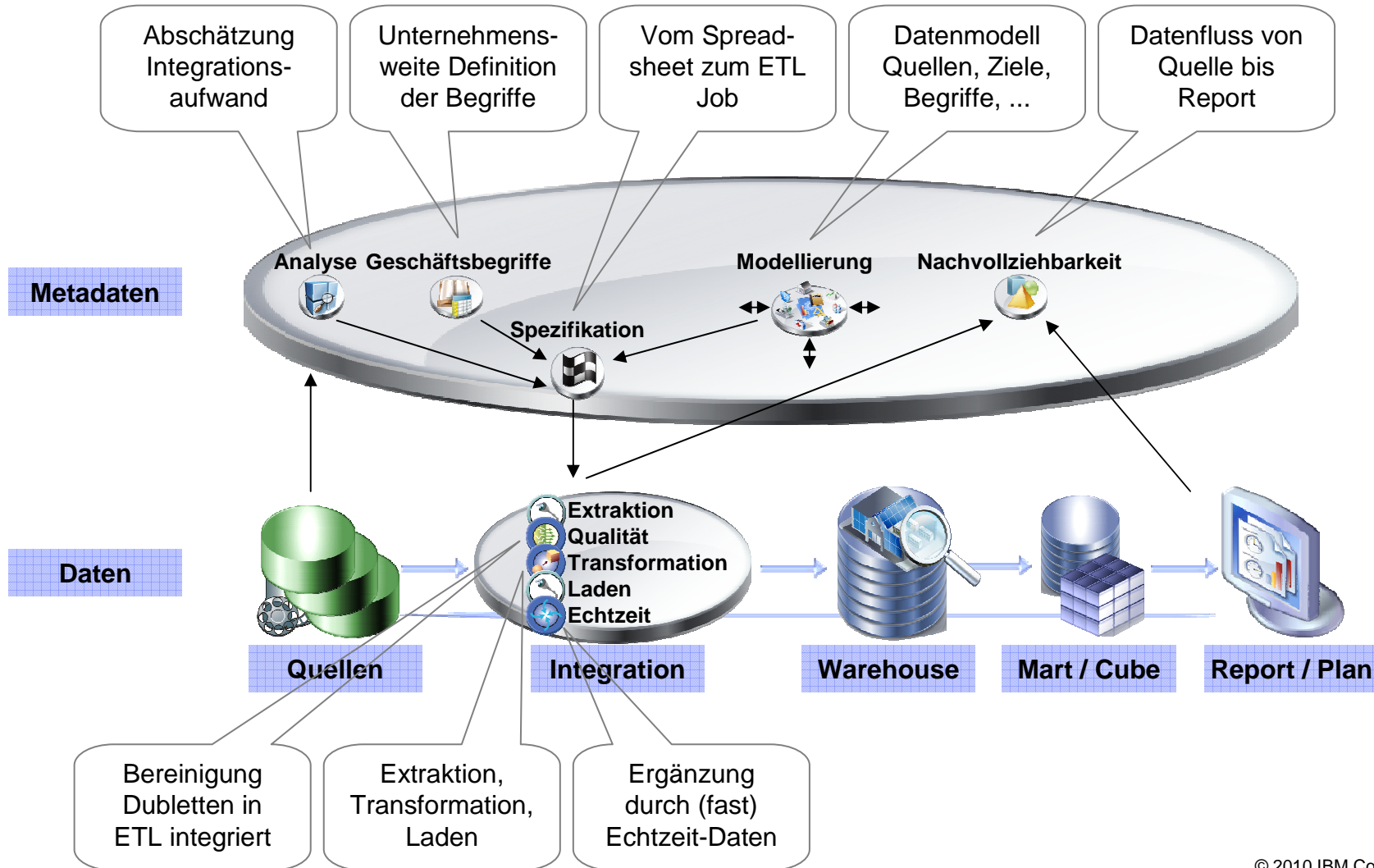
Beispiel: Datenfluss von Kundendaten



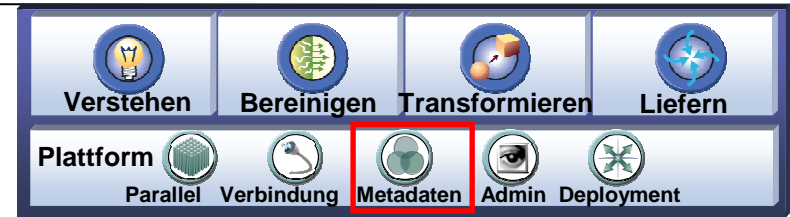
IBM InfoSphere Information Server

Daten- und Metadaten-Integration

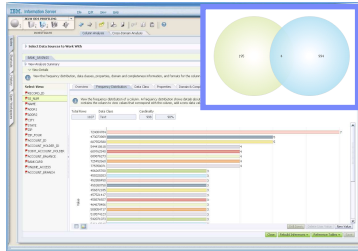
| | | | |
|------------------|-------------------|-----------------------|------------------|
| Verstehen | Bereinigen | Transformieren | Liefern |
| Plattform | Parallel | Verbindung | Metadaten |
| Admin | Deployment | | |



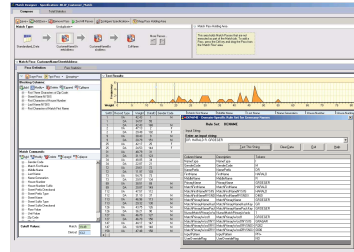
Verschiedene Nutzer von Metadaten Zentrales Metadaten-Repository



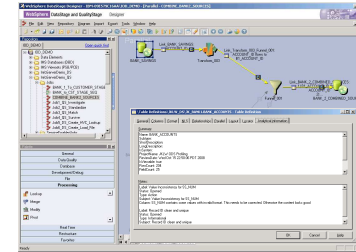
Daten-Analyst



Analyse / Überwachung
(Verteilung / Qualität)



Qualität
(Standards, Duplikate)



Transformationen
(Abhängigkeit, Analyse)



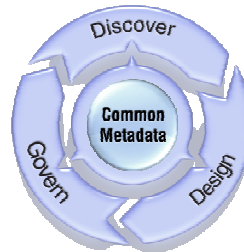
Entwickler



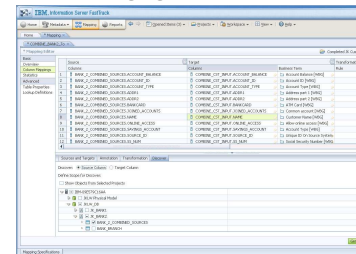
Endanwender



Geschäftsbegriffe
(Kontakt, Beziehungen)



Zentrales Repository
(intern und extern)



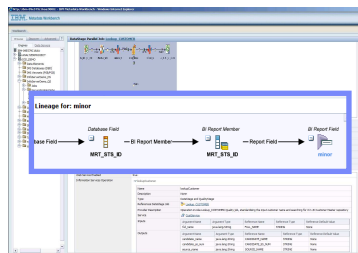
Spezifikation
(Spreadsheet => ETL)



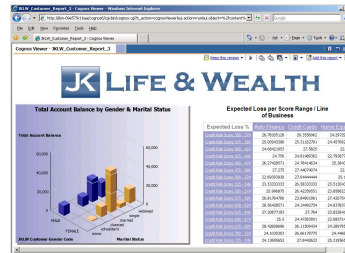
Spezialist
Fachabteilung



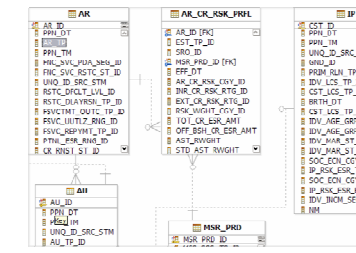
Metadaten-Analyst



Nachvollziehbarkeit
(Abhängigkeit, Laufzeit)



Externe Tools
(BI, Modellierung)



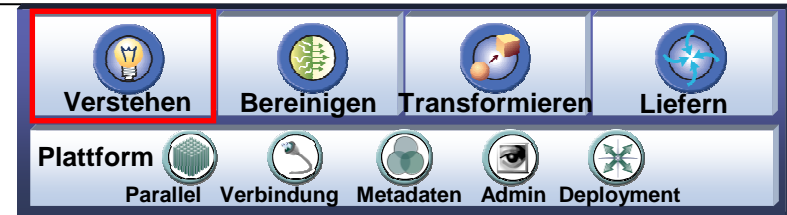
Modellierung
(Entitäten, Beziehung)



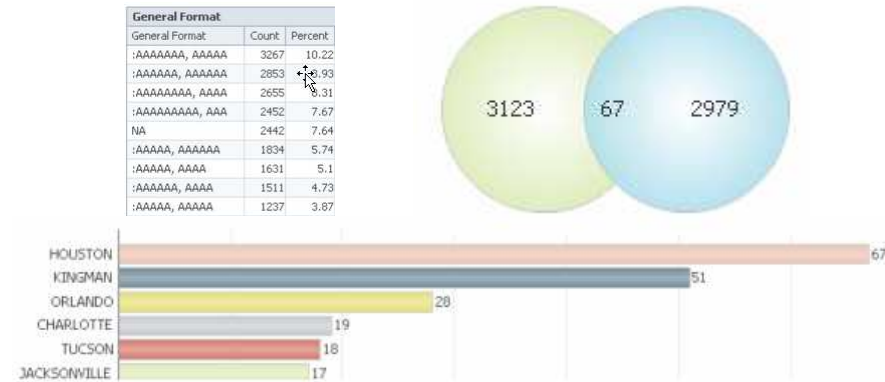
Modellierer

Verstehen

Schrittweise Analyse der Quelldaten



- **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



Carrier ID Validation

View Results

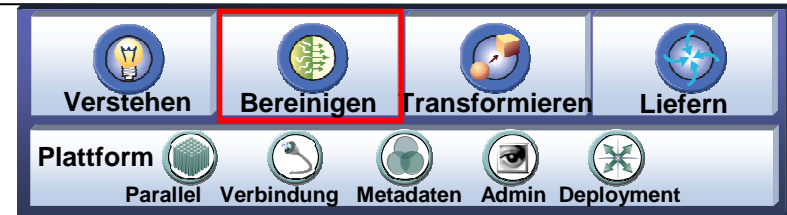
Include Tests

| Type | Timestamp | Validity | Validity | | Trend | Contact | Sample |
|------|-----------|----------|----------|--------|-------|--------------|--------|
| | | | Severity | # Pass | | | |
| Run | 01/01/09 | ✖ | 10.8% | 792 | 208 | ✖ Mim Foster | |
| Run | 12/01/08 | ✔ | | 904 | 96 | ✖ Mim Foster | |
| Run | 11/01/08 | ✔ | | 926 | 74 | ✖ Mim Foster | |
| Run | 10/01/08 | ✔ | | 500 | 0 | Mim Foster | ✔ |
| Run | 09/01/08 | ✔ | | 500 | 0 | Mim Foster | ✔ |

Information Analyzer

Bereinigen

Erhöhen der Datenqualität



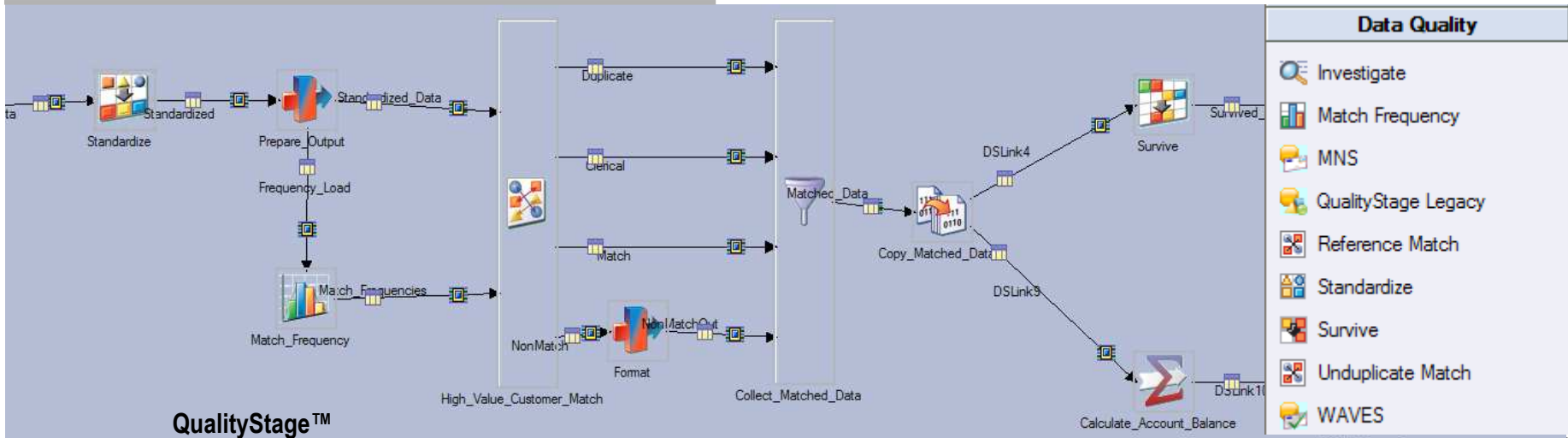
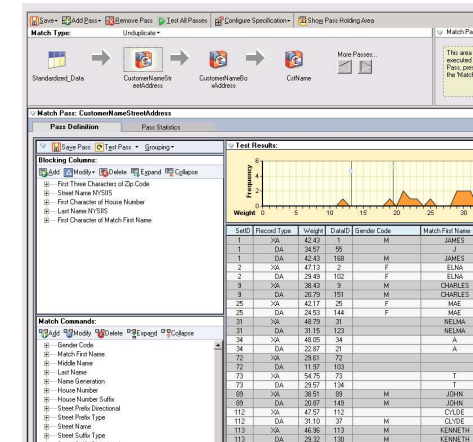
▪ Bereinigung der Qualität

- Unerwartete / versteckte Daten
- Fehlende Daten
- Duplikate

| | | |
|---------------|-----------------|---------------|
| Peter A. Berk | Kennedy 13, | Hamburg |
| Peter Berg | Kennedystr. 13, | 22587 Hamburg |
| Hr. Berg | Kennedystr. 31, | 2258J Hamburg |

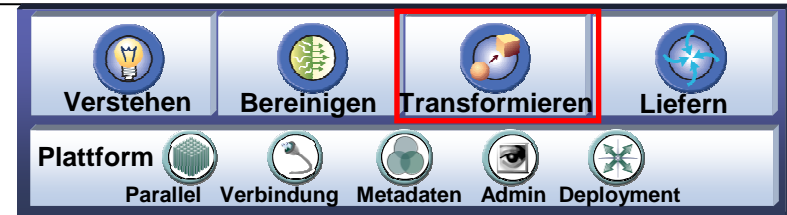
▪ Ablauf integriert in Transformations-Funktionalität

- Untersuchung
- Standardisierung
- Match Frequenz
- Match
- Survive

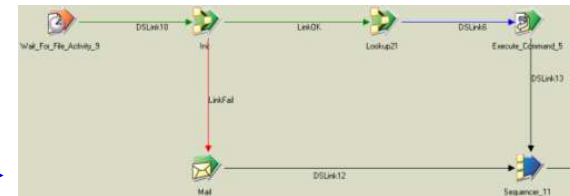
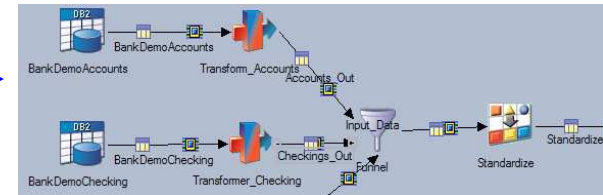


Transformieren

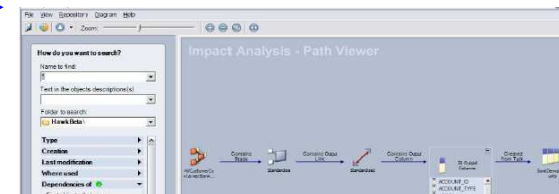
Extrahieren, Transformieren und Laden



- **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
- **Qualität:** Analyse und Bereinigung integriert
- **Metadaten:** Übergreifende Information / Analyse

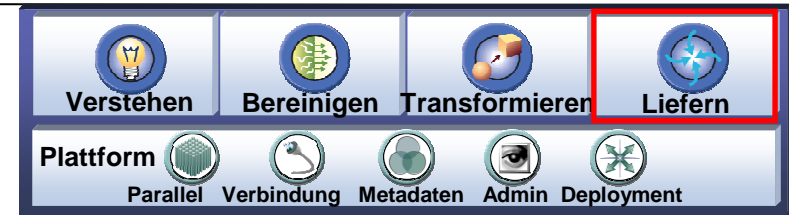


| >Occured | >On date | Type | Event |
|----------|------------|---------|---|
| 12:16:48 | 17.10.2006 | Control | Starting Job Demo1. |
| 12:16:48 | 17.10.2006 | Info | Environment variable settings: (...) |
| 12:16:48 | 17.10.2006 | Info | Demo1: Set NLS locale to US-ENGLISH,US-ENGL |
| 12:16:48 | 17.10.2006 | Info | Demo1..Transformer: DSD StageRun Active stage |
| 12:16:48 | 17.10.2006 | Info | Demo1..Transformer: Set NLS locale to US-ENGL |
| 12:16:48 | 17.10.2006 | Info | Demo1..Employee: Using NLS map MS1252 |
| 12:16:50 | 17.10.2006 | Info | Demo1..Department: Using NLS map MS1252 |
| 12:16:50 | 17.10.2006 | Info | Demo1..Demo1: Using NLS map MS1252 |
| 12:16:50 | 17.10.2006 | Warning | Demo1..Demo1: SQLExecDirect: Returned succes |



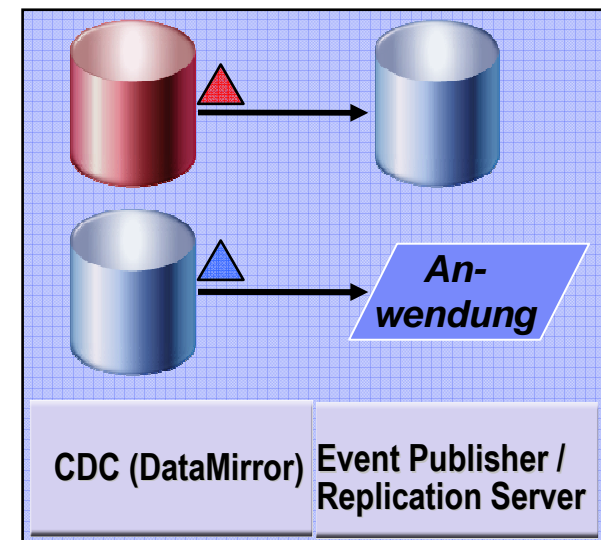
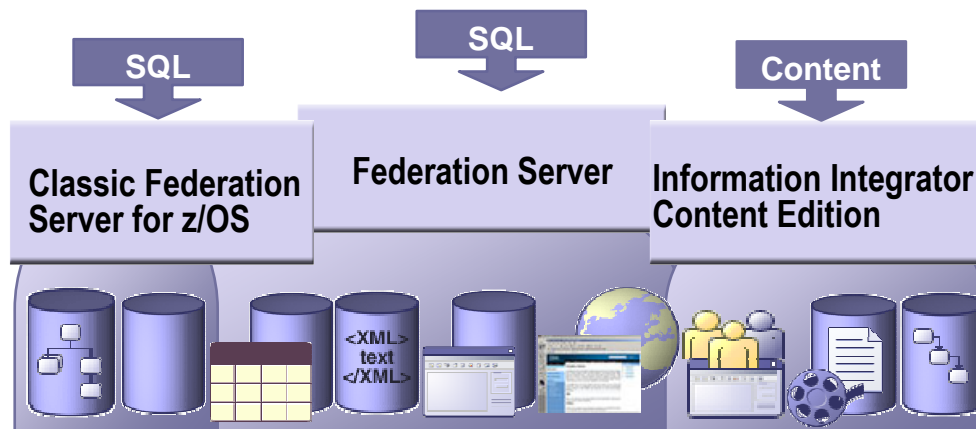
Liefern

Verfügbarkeit aktuellster Daten

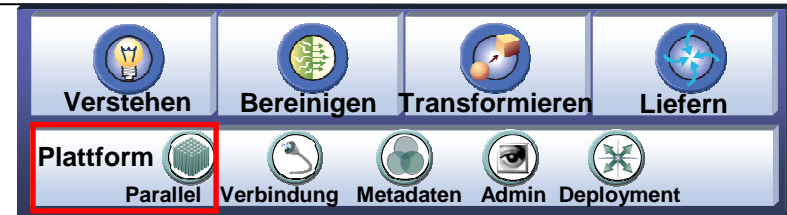


- **Föderation: Eine grosse virtuelle Datenbank**
 - Zugriff auf unterschiedliche und verteilte Informationen, als befänden diese sich in einem einzigen System
 - Echtzeitzugriff von Analyseanwendungen auf integrierte Informationen

- **Synchronisierung: Konsistente Informationen**
 - Abgleich zwischen Datenbanken, Hochverfügbarkeit, Migration
 - Datenänderungen starten Anwendungen

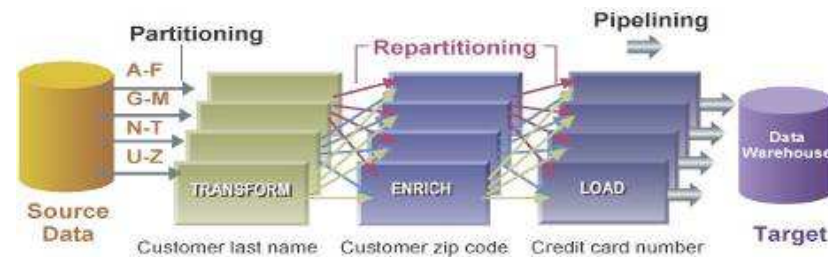


Parallelität Performance und Skalierbarkeit



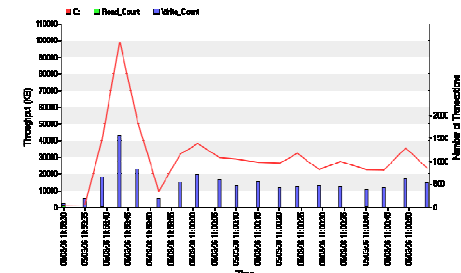
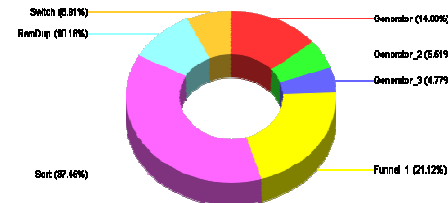
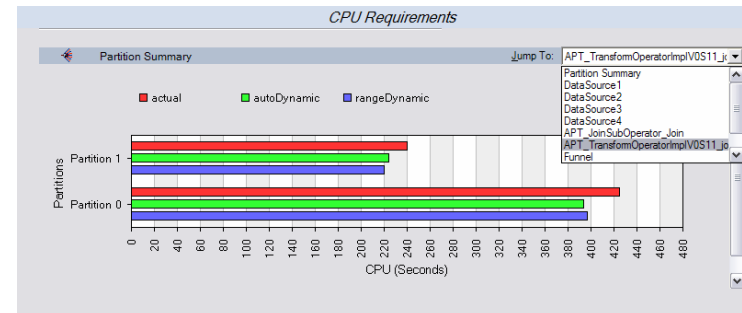
Parallelität und Skalierbarkeit

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



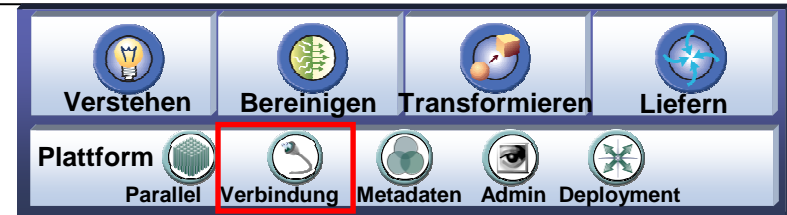
Performance

- 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



Verbindung

Anbindung aller wichtigen Datenquellen und -ziele

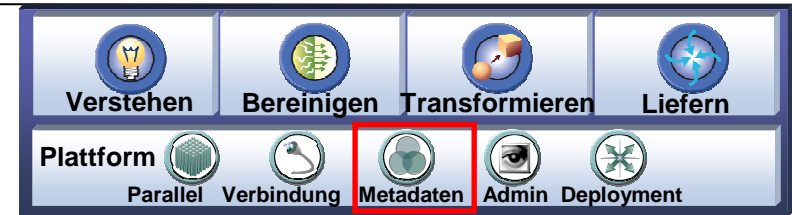


- **Datenbanken**
- **Files**
- **Processing**
- **Echtzeit**
- **Packs**

| Database | File | Real Time | Packs |
|--|--|---|---|
| <ul style="list-style-type: none"> 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 ▾ | <ul style="list-style-type: none"> Complex Flat File Data Set External Source External Target File Set Lookup File Set Sequential File <li style="text-align: center;">Processing FTP ▾ WebSphere TX Map | <ul style="list-style-type: none"> Java Client Java Transformer Web Services Client Web Services Transformer WebSphere MQ ▾ WISD Input WISD Output XML Input XML Output XML Transformer | <div style="border: 1px solid gray; padding: 2px;"> Stages for SAP R/3 ▾ </div> <ul style="list-style-type: none"> SAP R/3 SAP BW SAS Oracle Applications Siebel JD Edwards EnterpriseOne PeopleSoft Enterprise Salesforce.com Essbase |

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

Business Glossary
 Business Glossary Browser
 Business Glossary Anywhere

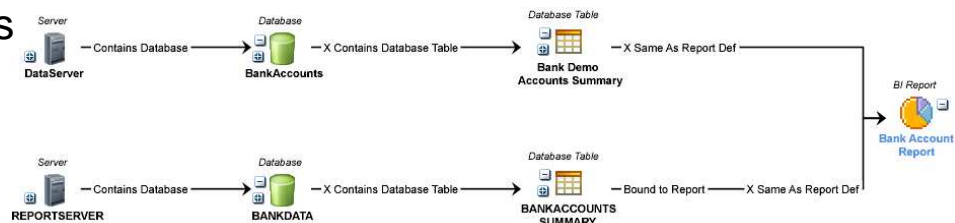
Logische Metadaten

- Daten Modellierung / Abhängigkeiten
- Data Architect

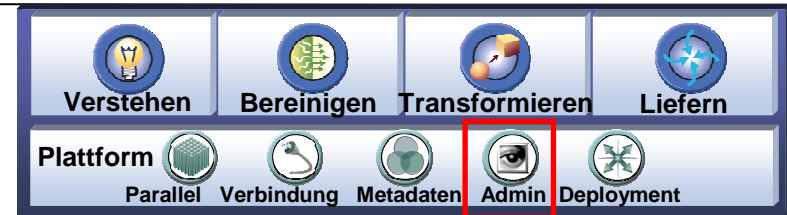
Erkunden der Metadaten

- Analysieren, Verstehen und Reports
- Metadatenaustausch mit anderen Tools
- Abhängigkeitsanalyse, Datenfluss
- Metadata Workbench®

The screenshot shows a 'Term Details' window for 'High Value Customer'. It includes a 'Long Description', 'Abbreviation' (LHVC), and 'Related IT Resources' (HV_CUSTOMER). Below is a 'Database Table Details' section showing a dependency diagram with tables: AR, AR_CR_RSK_PRF, TP, AU, and MSR_PRD. The diagram illustrates relationships between columns in these tables, such as AR_ID [FK] in AR_CR_RSK_PRF linking to AR_ID in AR.



Administration Verwaltung und Reports

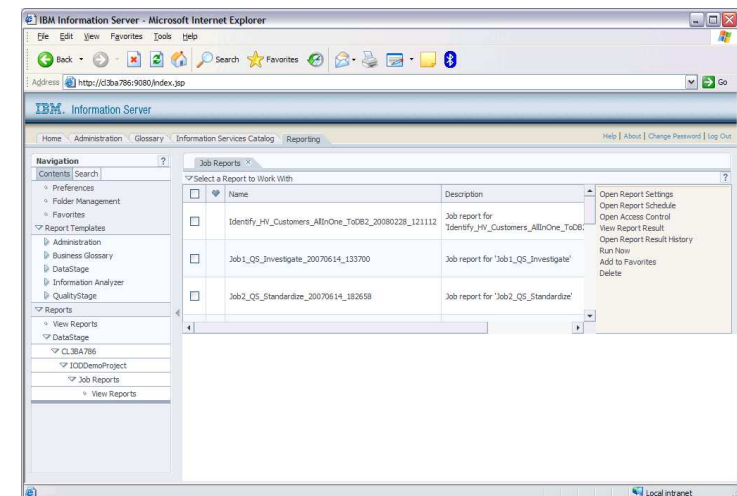
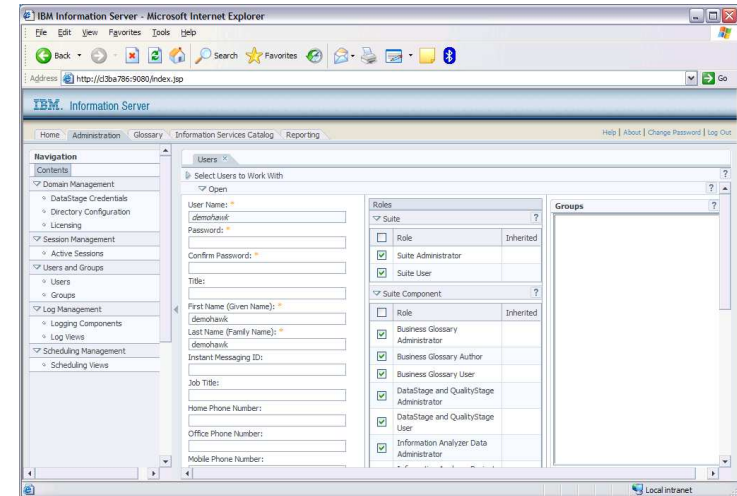


■ Administrations Konsole

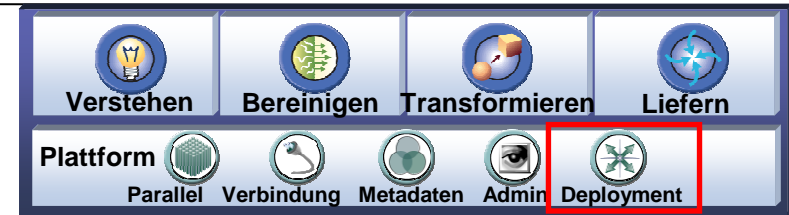
- Domain
- Session
- Users / Groups
- Log
- Scheduling

■ Reporting Konsole

- Templates
- Reports



Deployment Services und Appliance



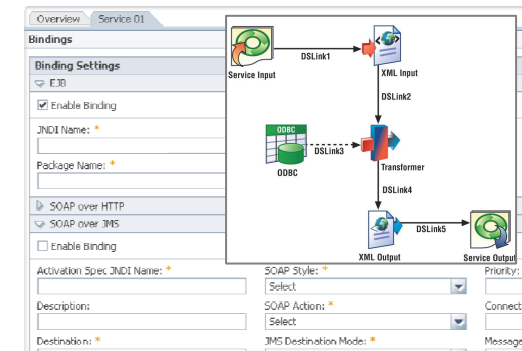
- **Schnelle SOA Entwicklung**
 - Infrastruktur für Informationsservices
 - EJB, JMS oder Web-Services, REST, RSS
 - Bereinigen, Transformieren und Föderieren

Information Services Director

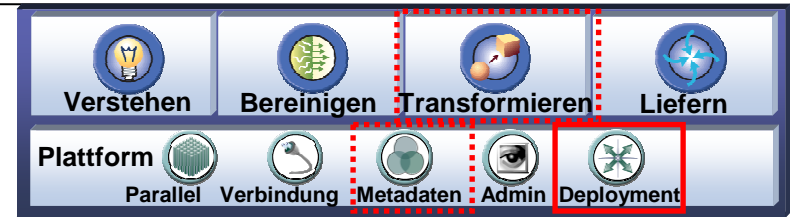
- **Integration von Hardware, Software und Service**

- Grid-basierte Integrationslösung für Unternehmensdaten
- IBM HS21 BladeCenter
- 2 dual core Intel Prozessoren pro Blade
- 4GB RAM, 60GB Storage pro Blade
- Red Hat Linux
- Tivoli Workload Schedule LoadLeveler
- DataStage, QualityStage, Metadata Server
- Optimale Skalierbarkeit, Reduzierte Kosten

Information Server Blade



Metadaten generieren ETL Von Spreadsheet Definitionen zu ETL Jobs



- Excel mit Geschäftsregeln
- Komplettierung in FastTrack
- ETL Job Generierung

Specifications of Generated Job: CleanseNameAddress, IdentifyHighValueCustomers
Date of Generated Job: 2/13/08 4:50 PM

| Source Column(s) | Target Column(s) | Transformation Rule(s) |
|--------------------------------------|---|---|
| DSLink2.C0_ADDR1 DSLink2.C0_ADDR2 | M0_STREET_NAME M0_STREET_SUFFIX M0_HOUSE_NUMBER | Standardize with USADDR.SET ruleset. Use first character of house number as blocking column in match. Street name very important. |
| DSLink2.C0_STATE | M0_STATE | Standardize with USAREA.SET. |
| DSLink2.C0_ZIP | M0_ZIP | Standardize with USAREA.SET. Add first three digits as blocking column for match. |
| DSLink2.C0_TAX_ID | M0_TAX_ID | Standardize with USTAXID.SET. Very important for match. |
| DSLink2.C0_NAME | M0_MIDDLE_NAME M0_FIRST_NAME M0_LAST_NAME | Standardize with USNAME.SET ruleset. Use first character of first name as blocking column for match. First name and last name are very important for match. |
| DSLink2.C0_CITY | M0_CITY | Standardize with USAREA.SET. |

Tip: All other columns can be mapped with the column auto match option in the corresponding stage.

| 1 | Source Columns | Source Business Term | Target Columns | Target Business Term | Business Rule | Function | Annotation |
|----|--|----------------------|--|----------------------|---|----------|---|
| 2 | tablab.tablab.BANK.CUSTOMERS.HA ME | Customer | tablab.tablab.BANK.MASTER.FIRST_NA ME; tablab.tablab.BANK.MASTER.MIDDL E_NAME; tablab.tablab.BANK.MASTER.L AST_NAME | High Value Customer | Standardize with USNAME.SET ruleset. Use first character of first name as blocking column for match. First name and last name are very important for match. | | |
| 3 | tablab.tablab.BANK.CUSTOMERS.ADD R1; tablab.tablab.BANK.CUSTOMERS. ADDRESS | Street_Addr | tablab.tablab.BANK.MASTER.HOUSE_N UMBER; tablab.tablab.BANK.MASTER.ST REET_NAME; tablab.tablab.BANK.MAST ER.STREET_SUFFIX | Street | Standardize with USADDR.SET ruleset. Use first character of house number as blocking column in match. Street name very important. | | |
| 4 | tablab.tablab.BANK.CUSTOMERS.CITY tablab.tablab.BANK.CUSTOMERS.STA TE | | tablab.tablab.BANK.MASTER.CITY tablab.tablab.BANK.MASTER.STATE | City State | Standardize with USAREA.SET. Standardize with USAREA.SET. | | |
| 6 | tablab.tablab.BANK.CUSTOMERS.ZIP | | tablab.tablab.BANK.MASTER.ZIP | Zip | Standardize with USAREA.SET. Add first three digits as blocking column for match. | | |
| 7 | tablab.tablab.BANK.CUSTOMERS.ACC OUNT_BALANCE | | tablab.tablab.BANK.MASTER.LEVEL | Combined Balance | Standardize with USAREA.SET. Add first three digits as blocking column for match. | | If ACCOUNT_BALANCE > 30000 then 'A' else if ACCOUNT_BALANCE > 10000 then 'B' else 'C' |
| 8 | tablab.tablab.BANK.CUSTOMERS.TAX ID | SS_Num | tablab.tablab.BANK.MASTER.TAX_ID | | Standardize with USTAXID.SET. Very important for match. | | |
| 9 | tablab.tablab.BANK.CUSTOMERS.YEA RS_CLIENT | | tablab.tablab.BANK.MASTER.YEARS_C LIENT | | | | |
| 10 | tablab.tablab.BANK.CUSTOMERS.GEN DER | | tablab.tablab.BANK.MASTER.GENDER | | | | |
| 11 | tablab.tablab.BANK.CUSTOMERS.ONL INE_ACCESS | | tablab.tablab.BANK.MASTER.ONLINE_A CCESS | | | | |
| 12 | tablab.tablab.BANK.CUSTOMERS.CUS TOMER_ID | | tablab.tablab.BANK.MASTER.CUSTOME R_ID | | | | |

Mapping Specifications

| Name | Target | Created | Modified | Created By | Status | Percent Reviewed |
|------------------------------|-------------|--------------|--------------|------------|-------------|------------------|
| ICDProject | | | | | | |
| ICD_BankExtract | CUSTOMERS | 1/22/08 5:35 | 2/5/08 8:41 | dbzadmin | In Progress | 8 % |
| ICD_Demo | | | | | | |
| Bank1 Extract | CUSTOMERS | 1/22/08 5:33 | 1/22/08 5:33 | dbzadmin | In Progress | 0 % |
| Bank2 Extract | CUSTOMERS | 1/22/08 5:33 | 1/22/08 5:33 | dbzadmin | In Progress | 0 % |
| Bank3 Extract | CUSTOMERS | 1/22/08 5:33 | 1/22/08 5:33 | dbzadmin | In Progress | 0 % |
| Cleanse Name Address | MASTER | 1/22/08 5:34 | 1/22/08 5:34 | dbzadmin | In Progress | 0 % |
| Identify High Value Customer | MARKETING_C | 1/22/08 5:35 | 1/22/08 5:35 | dbzadmin | In Progress | 0 % |

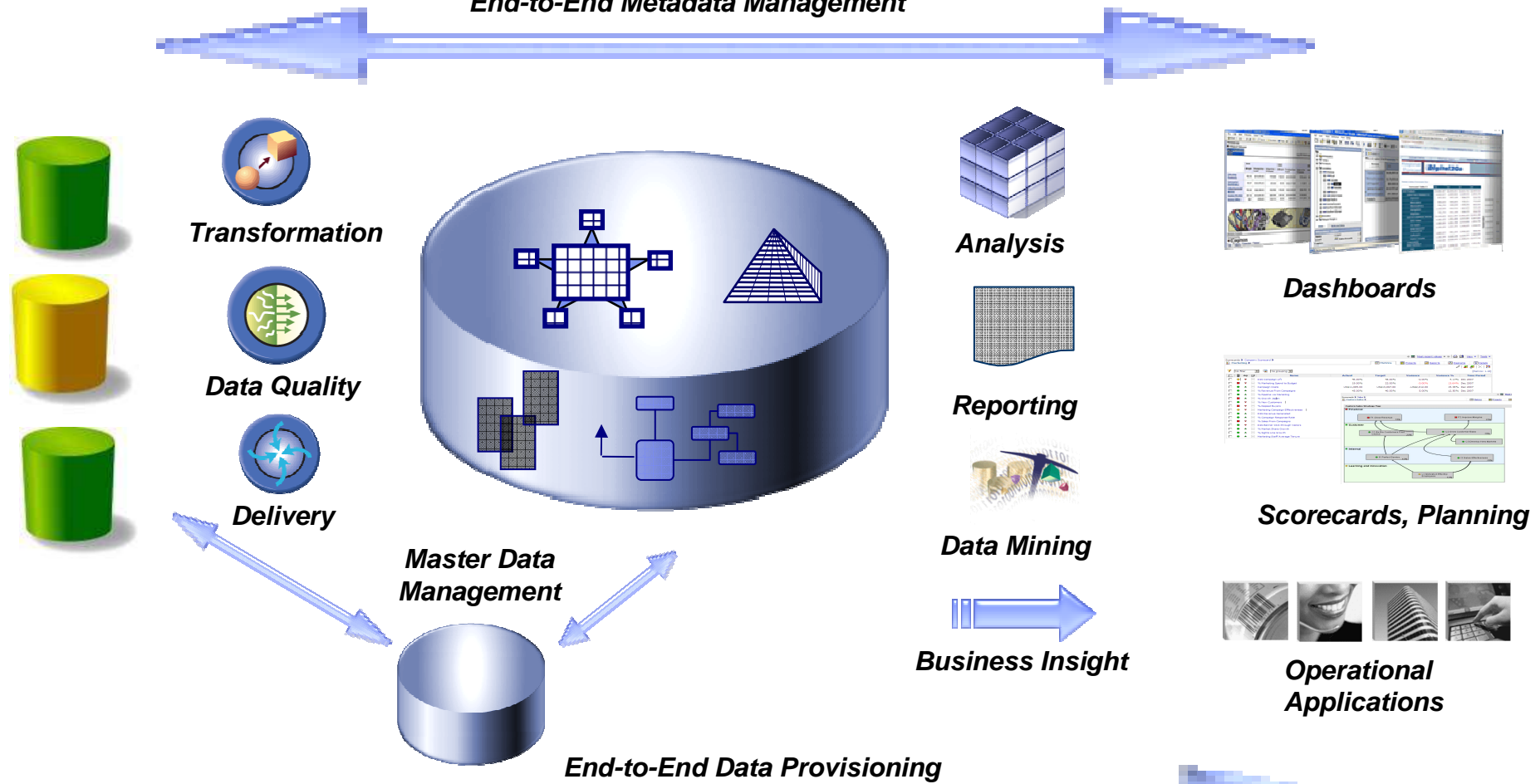
Column Mappings

| Source Columns | Target Columns | Business Term | Rule | Reviewed |
|-----------------------------|--------------------------|---------------|-----------------------------------|-------------------------------------|
| 1 CUSTOMERS.FNAME, CUSTOM | CUSTOMERS.NAME | | Concat first name and last name. | <input checked="" type="checkbox"/> |
| 2 CUSTOMERS.ADDR1 | CUSTOMERS.ADDR1 | | | <input type="checkbox"/> |
| 3 CUSTOMERS.ADDR2 | CUSTOMERS.ADDR2 | | | <input type="checkbox"/> |
| 4 CUSTOMERS.CITY | CUSTOMERS.CITY | | | <input type="checkbox"/> |
| 5 CUSTOMERS.STATE | CUSTOMERS.STATE | | | <input type="checkbox"/> |
| 6 CUSTOMERS.ZIP | CUSTOMERS.ZIP | | | <input type="checkbox"/> |
| 7 ACCOUNTS.CUSTOMER_ID, A | CUSTOMERS.CUSTOMER_ID, C | | Aggregate account balance over sa | <input type="checkbox"/> |
| 8 CUSTOMERS.SS_NUM | CUSTOMERS.TAX_ID | | | <input type="checkbox"/> |
| 9 DEMOGRAPHICS.YEARS_CLIENT | CUSTOMERS.YEARS_CLIENT | | | <input type="checkbox"/> |
| 10 DEMOGRAPHICS.GENDER, Dn | CUSTOMERS.GENDER | | | <input type="checkbox"/> |
| 11 DEMOGRAPHICS.ONLINE_ACC | CUSTOMERS.ONLINE_ACCESS | | | <input type="checkbox"/> |
| 12 DEMOGRAPHICS.LEVEL | CUSTOMERS.LEVEL | | | <input type="checkbox"/> |

InfoSphere Informations-Architektur

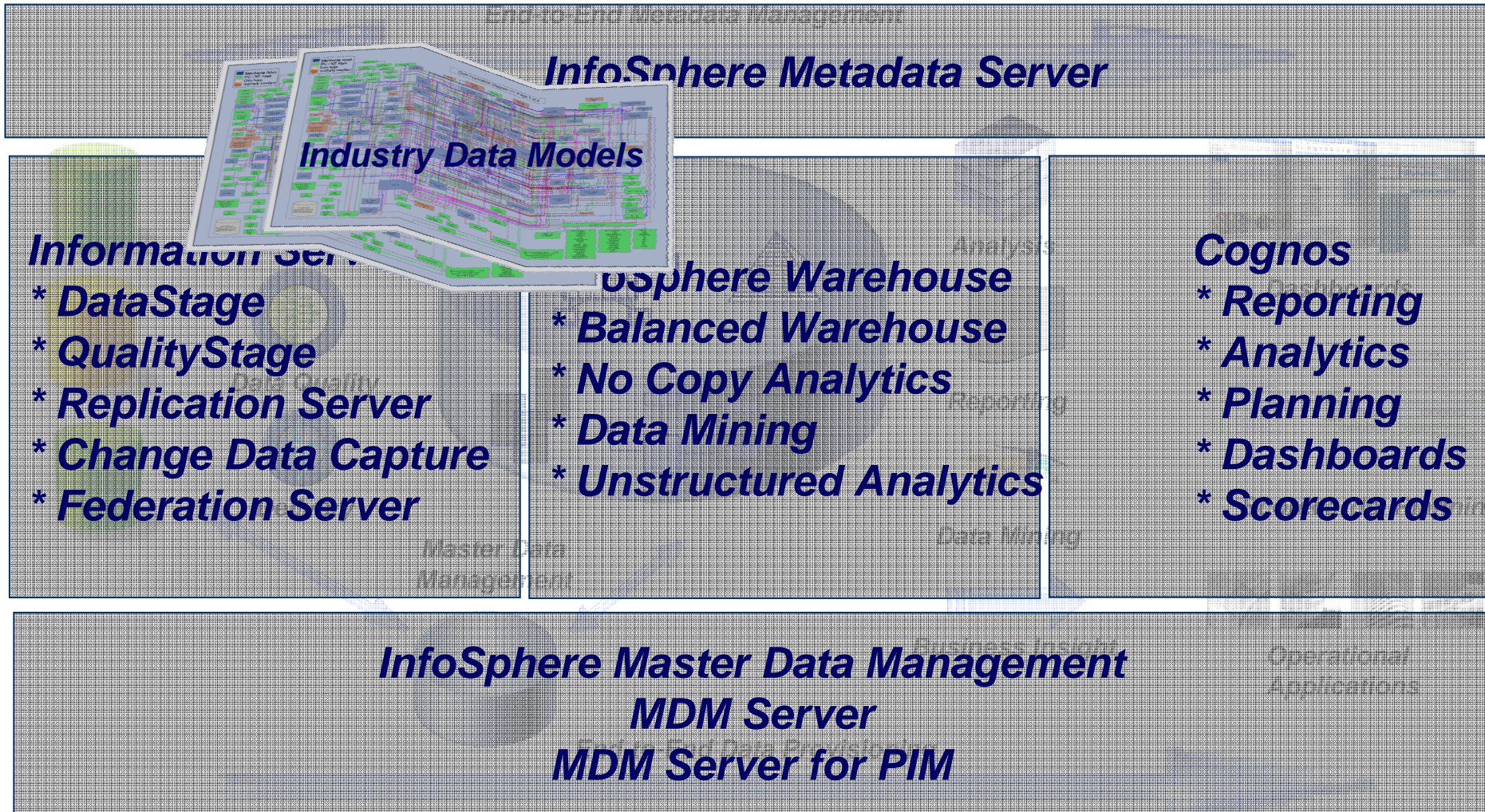
Konsistente, gültige Information von der Entstehung bis zur Analyse

End-to-End Metadata Management



InfoSphere Produkte – Trusted Information

Zuverlässige Informationen von der Entstehung bis zur Auswertung



Operational Data Sources

Data Integration

Enterprise Data Warehouse

Information Delivery

Dispositive Data