



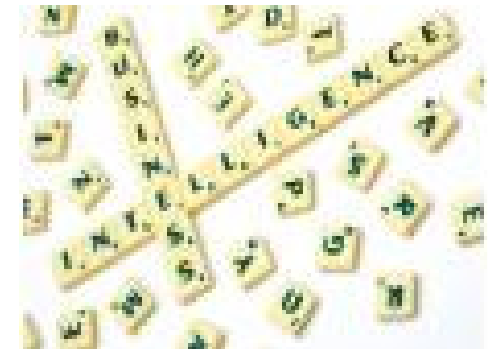
Datenschaufler für Java

Welche ETL-Werkzeuge gibt es
und wie kann ich sie einsetzen?

Frank Pientka, Senior Consultant
[E2] 3.7.2008, 9.50 - 10.35 Uhr



Pientka - Juli 2008



ETL mit
Java

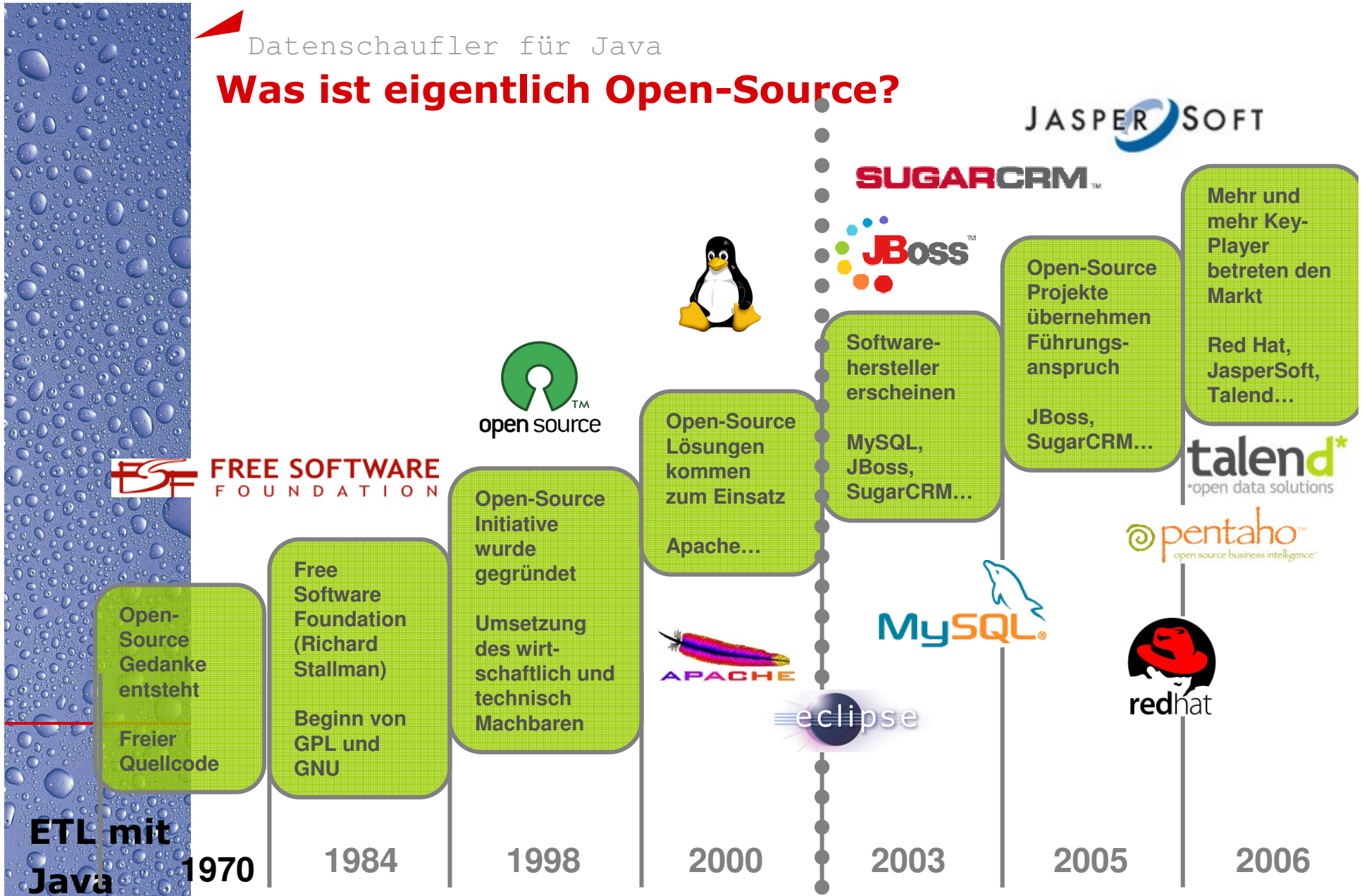
Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**

Was ist eigentlich Open-Source?



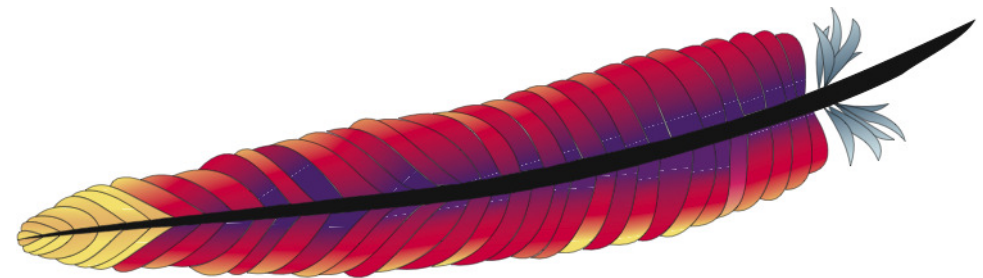
Zunehmende Professionalisierung der Produkte, Geschäftsmodelle

Was ist Open Source Software?



Open Source ist eine Software, wenn sie unter einer anerkannten Lizenz der Open Source Initiative steht.

(<http://www.opensource.org>)



ETL mit
Java

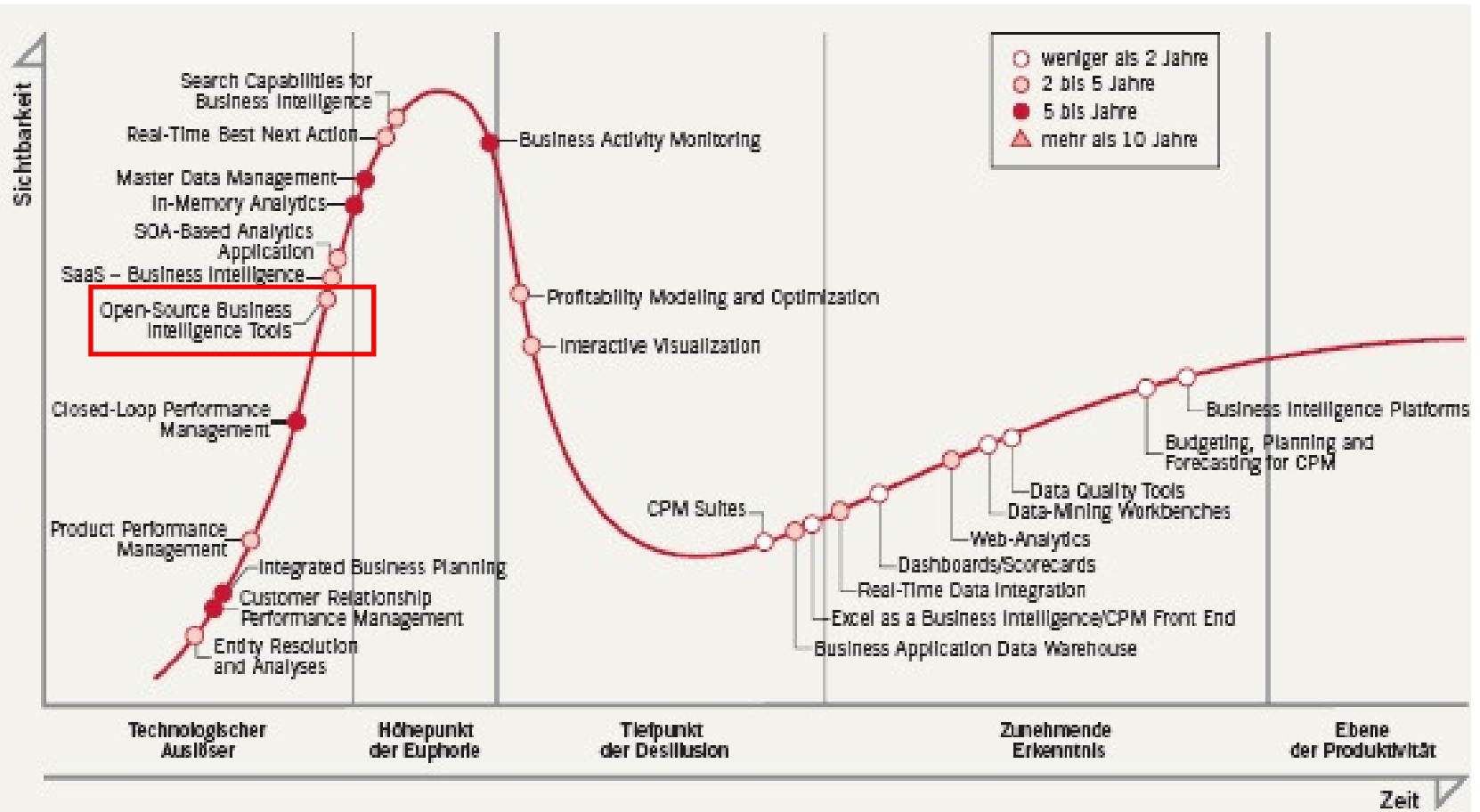
Open Source vs. Closed System

- Best-of-breed
- Wahlfreiheit
- herstellerunabhängig
- Innovation
- Flexibilität
- Standards
- Alles-aus-Einer-Hand
- Vorauswahl
- Vendor Lock-in
- Integration

**ETL mit
Java**

Gartner BI-Hype-Cycle 2007

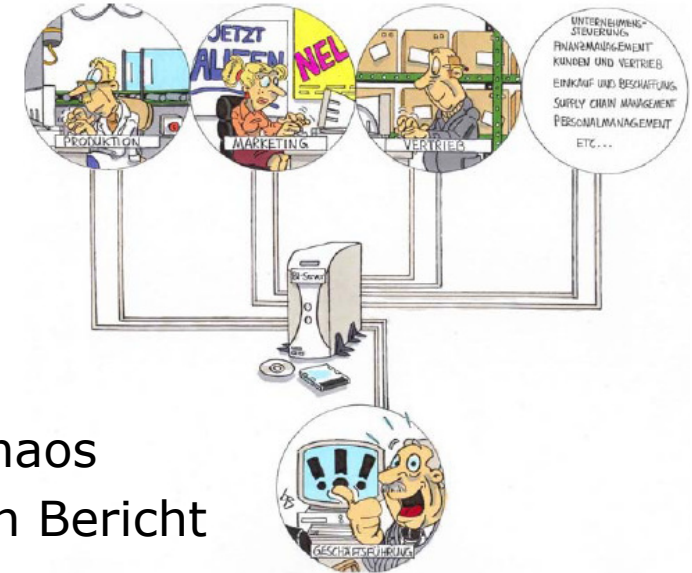
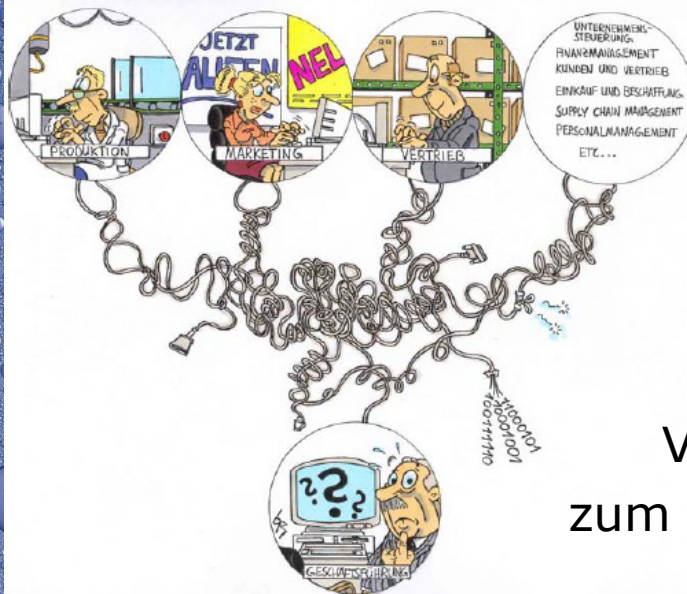
Das erste mal Open-Source (BIRT, Jasper, Pentaho) werden im magischen BI-Quadranten im **Februar 2008** berücksichtigt .



ETL mit
Java

Datenschaufler für Java

Was ist Business Intelligence (BI)? Informationen intelligent nutzbar machen



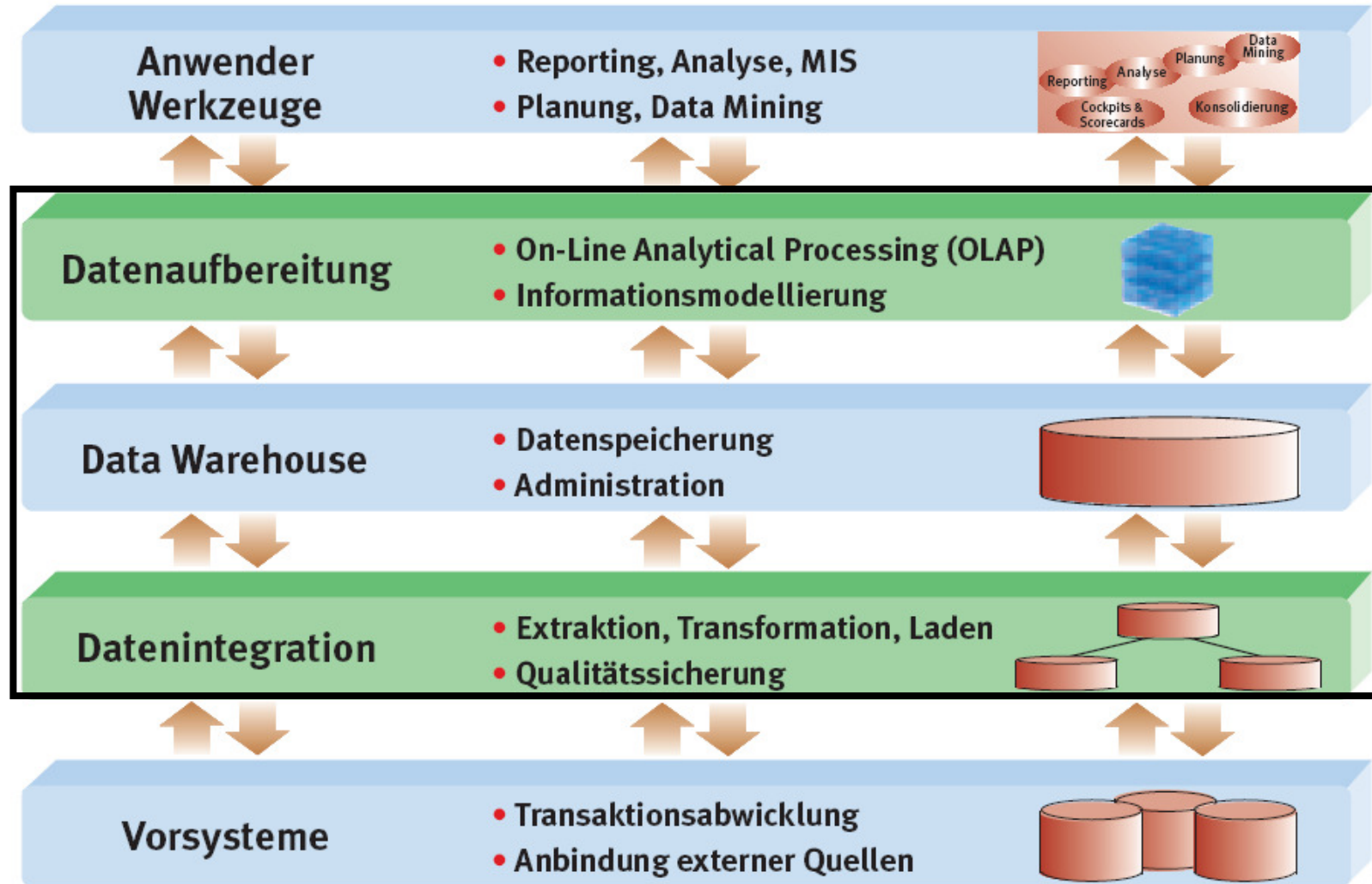
Vom Datenchaos
zum einheitlichen Bericht

ETL mit
Java

Quelle:
Management-Brevier der Lünendonk GmbH



Eine BI-Referenzarchitektur (BARC)



ETL mit
Java

60-80% der Aufwände in einem Data Warehouse fallen bei der Datenintegration an

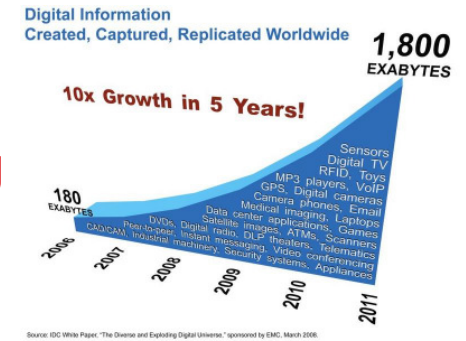
Datenintegration als Herausforderung

■ Situation

- Datenvolumen wächst *exponential*
- Unterschiedliche und *heterogene* Datenquellen
- Schnittstellen-Änderungen
- *Daten haben langes Leben*, sind verteilt, inkonsistent

■ Anforderungen

- Daten müssen in immer kürzerer Zeit verfügbar sein
- *Datenqualität* ist kritisch (Wertschöpfungsfaktor)
- Kontrolle und Steuerung der Datenflüsse
- Aus Unternehmensdaten Informationen gewinnen
- *Metadaten* sind wichtig (*single version of the truth*)



ETL mit
Java

Welche Herausforderungen hat Datenintegration?

Daten sind **unternehmenskritisch**, operativ
Weg vom Batch (**ETL**) hin zu Realtime (**EAI**)

- Performance, Skalierbarkeit

Herausforderungen im Datenmanagement:

- Datenqualität, Metadaten, Steuerung, Umgang mit Fehlern

Datenintegration ist **das schwarze Loch** der IT (Kosten, Schmutzdecke)

Oft **brisante Mischung** aus **Programmiersprachen-Mix**

Wartungs-/Testproblem

Unterschätzt **Komplexität**, Zeit, Kosten



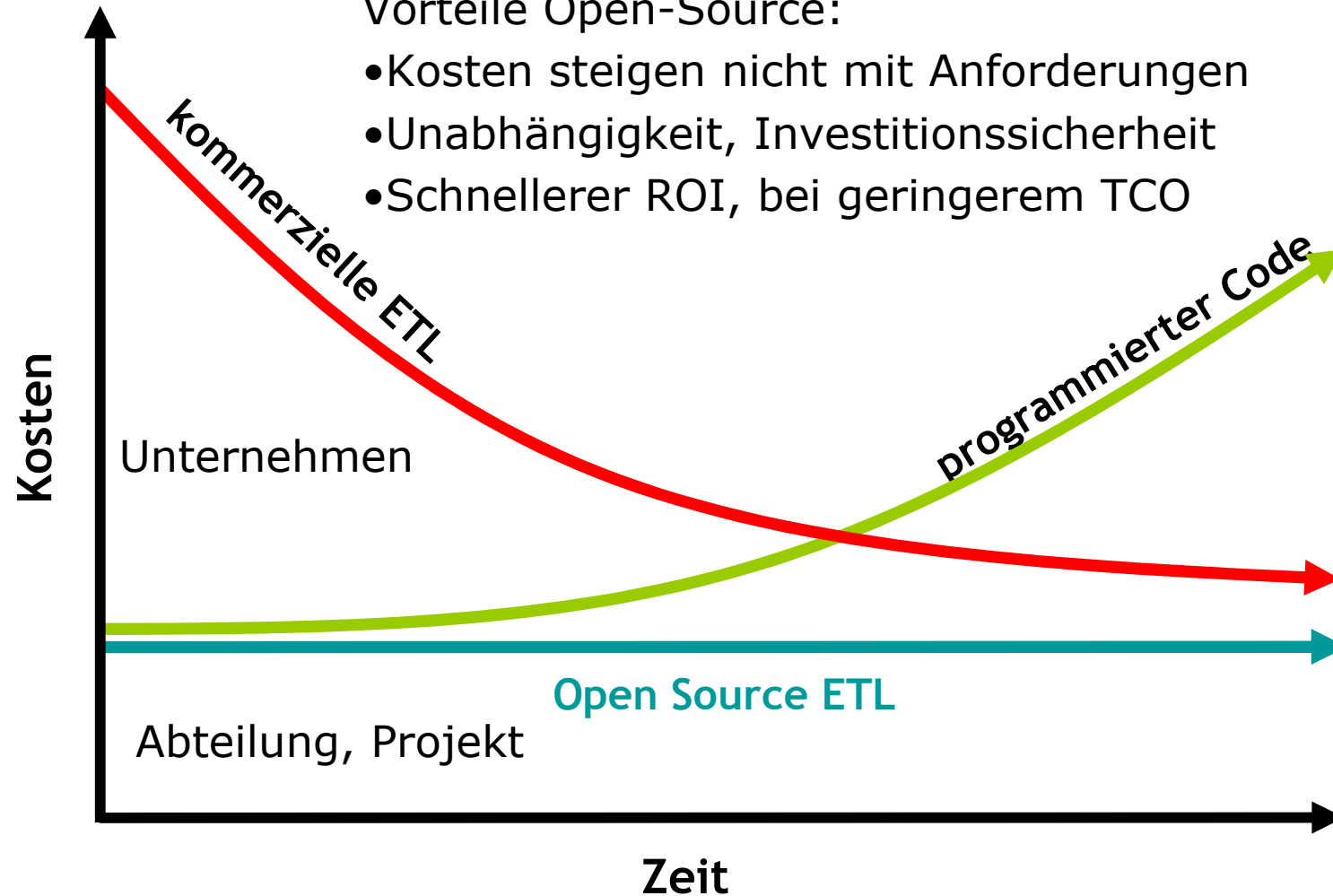
ETL mit
Java



Daten-Integrationsdilemma: Build Versus Buy

Vorteile Open-Source:

- Kosten steigen nicht mit Anforderungen
- Unabhängigkeit, Investitionssicherheit
- Schnellerer ROI, bei geringerem TCO



ETL mit
Java

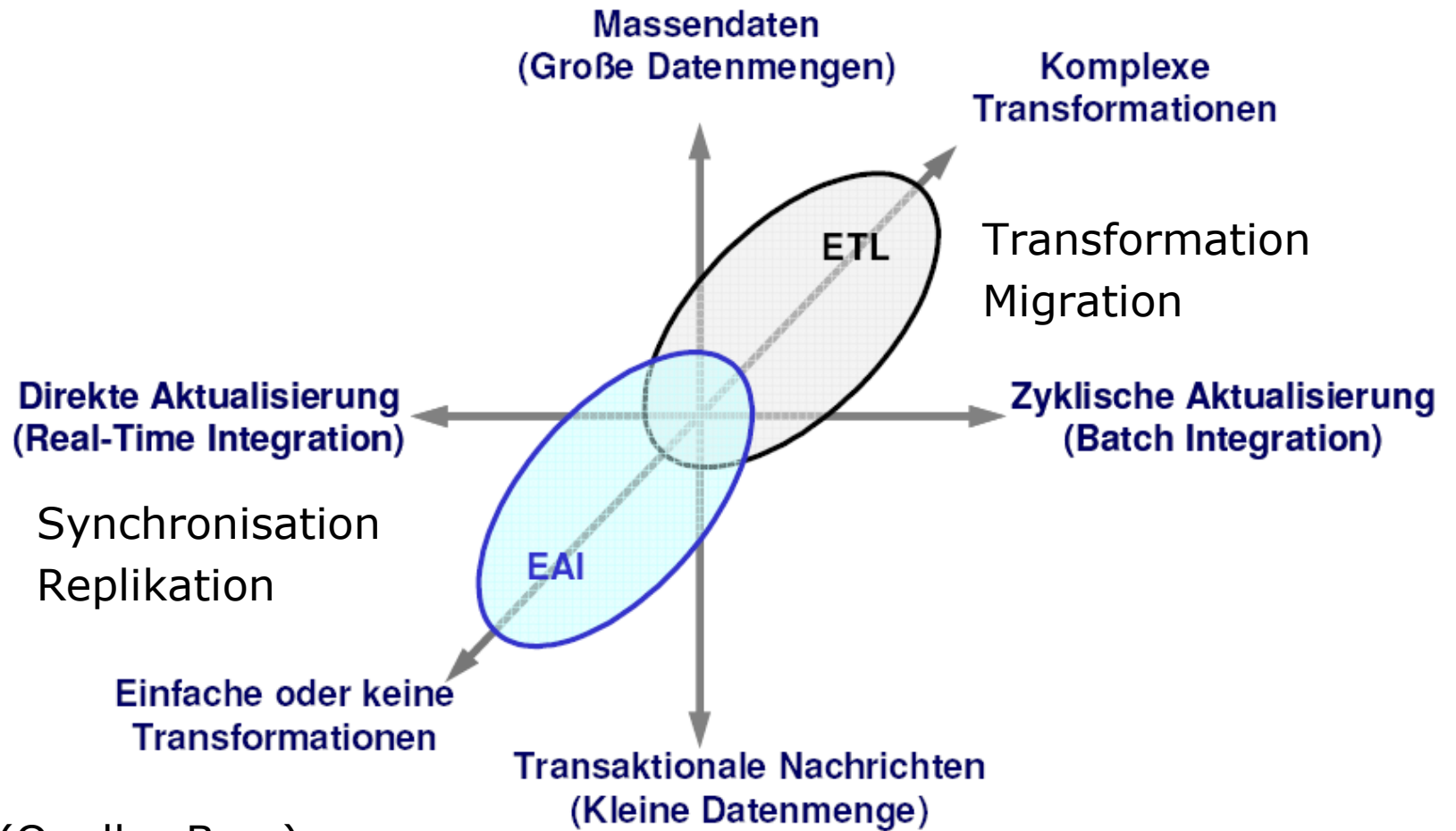
Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**

Wie passen ETL und EAI zusammen?



(Quelle: Barc)

ETL mit
Java

Methoden der Datenintegration

Extraktion, Transformation, Laden (ETL)

- **Ziel:** Datenintegration
- Zyklische Integration großer Datenmengen
- Umfangreiche Transformation
- ETL-Engines & Repositories

Enterprise Information Integration (EII)

- **Ziel:** Föderation von Datenquellen
- Erstellung von Views und Abfrage auf Anforderung
- Strukturierte u. unstrukturierte Daten
- EII Server für Query-Handling

Enterprise Application Integration (EAI)

- **Ziel:** Geschäftsprozessintegration
- Near-Time Verbindung operativer & dispositiver Systeme
- Formattransformation von Nachrichten
- Messaging-Middleware & Broker

ETL mit
Java

Vergleich ETL und ELT

ETL: Extraktion, Transformation und Laden in die Zieldatenbank

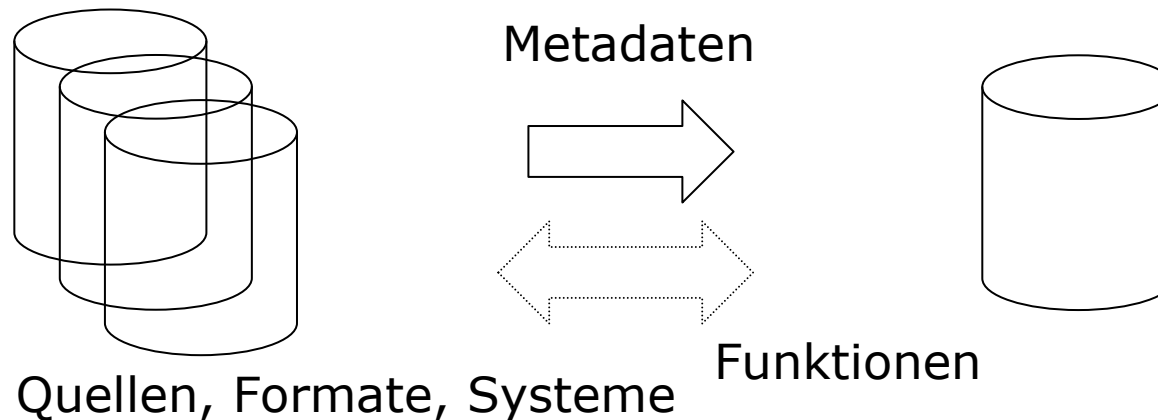
ELT: Extraktion, Laden und Transformation in der Zieldatenbank

ELT-Werkzeuge

- nutzen mit SQL die Leistungsfähigkeit der Datenbank und haben keine eigene Transformationsengine
- kommen von Datenbank- oder Spezialanbietern

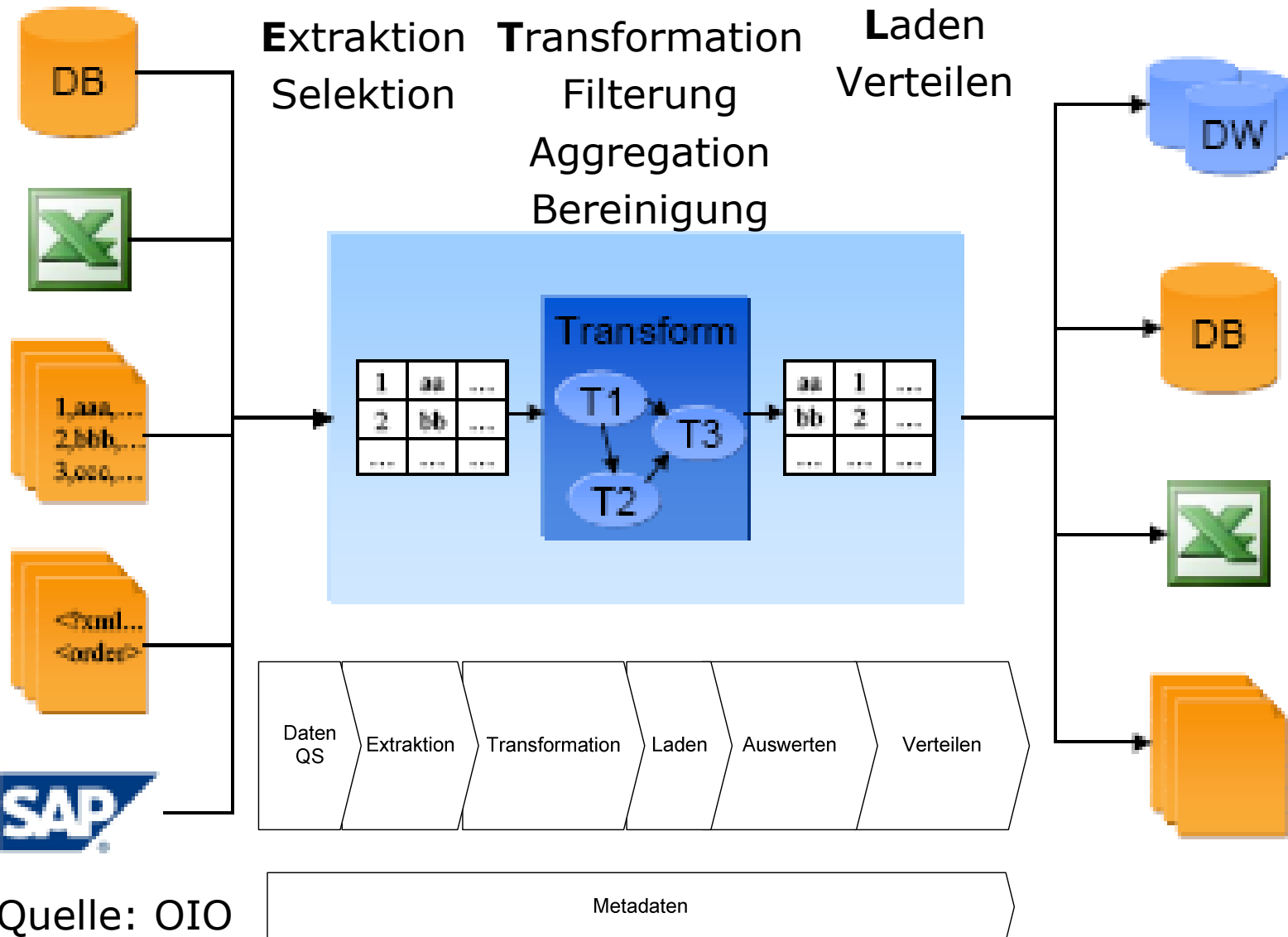
Datenintegration: Breites Einsatzspektrum

- **Business Intelligence**/Data Warehousing
- **Migration** von Altsystemen
- **Systemkonsolidierung**
- **Synchronisation** zwischen Systemen (Replikation)



**ETL mit
Java**

Datenintegration: Funktionen, Prozess



ETL mit Java



Quelle: OIO

Auswahlkriterien ETL-Werkzeuge

„Wer nicht weiß, wohin er will, darf sich nicht wundern, wenn er woanders ankommt.“ Mark Twain

- Keine allgemeine Wunschliste
- Wenige, aussagekräftige Kriterien
- Evaluation der Anforderungen
- Wähle die Lösung, nicht den Anbieter

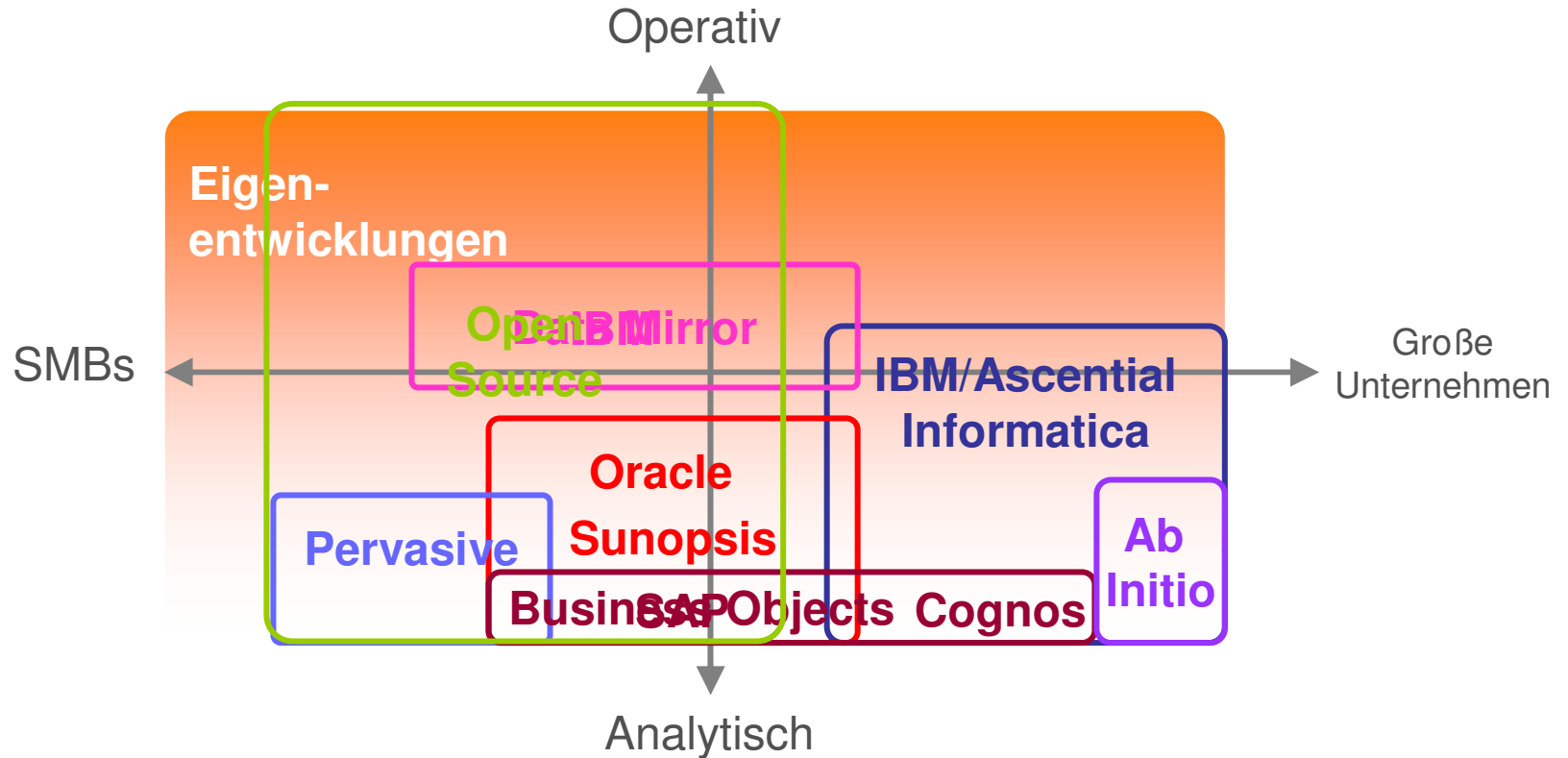
ETL-Kriterienkatalog

- Anbieter und Produkt
- Bedienung
- Dokumentation
- Architektur
 - (ETL, ELT, RT/Codegenerator)
 - Multithread, Cluster, Parallelisierung
- Datenquellen, Datenformate
- Konnektivität, Schnittstellen
- Funktionen, Transformation
- Modellierung, Metadatenverwaltung
- Entwicklung, Fehlersuche
- Überwachung, Steuerung
- Betrieb, Skalierbarkeit



**ETL mit
Java**

Der ETL-Markt heute



Marktforscher wie Gartner, Forester, BARC
sehen gute Chancen für Open-Source-BI

ETL mit
Java

Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**

Open-Source-BI-Werkzeuge nach Schichten

Berichte

Jasper Report

BIRT

Pentaho Reporting

JPivot

Analyse

WEKA

MONDIRAN

YALE

PALO

Datenintegration

Kettle

Clover.
ETL

Talend
Jasper.ETL

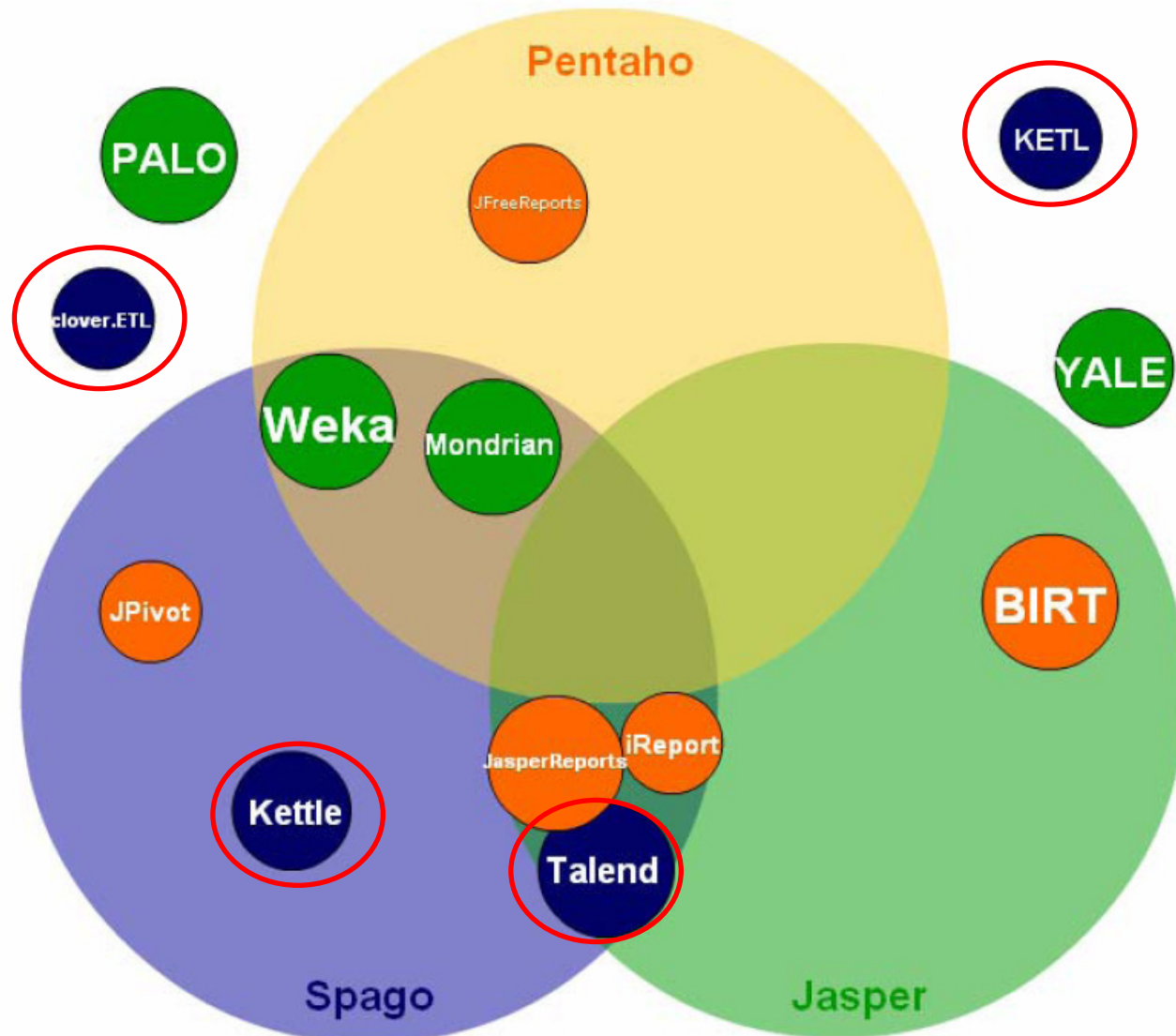
Ketl

ETL mit
Java

Open-Source-BI-Suiten

- Pentaho
- JasperSoft
- SpagoBI

BI Open-Source-Suiten: Gemeinsamkeiten

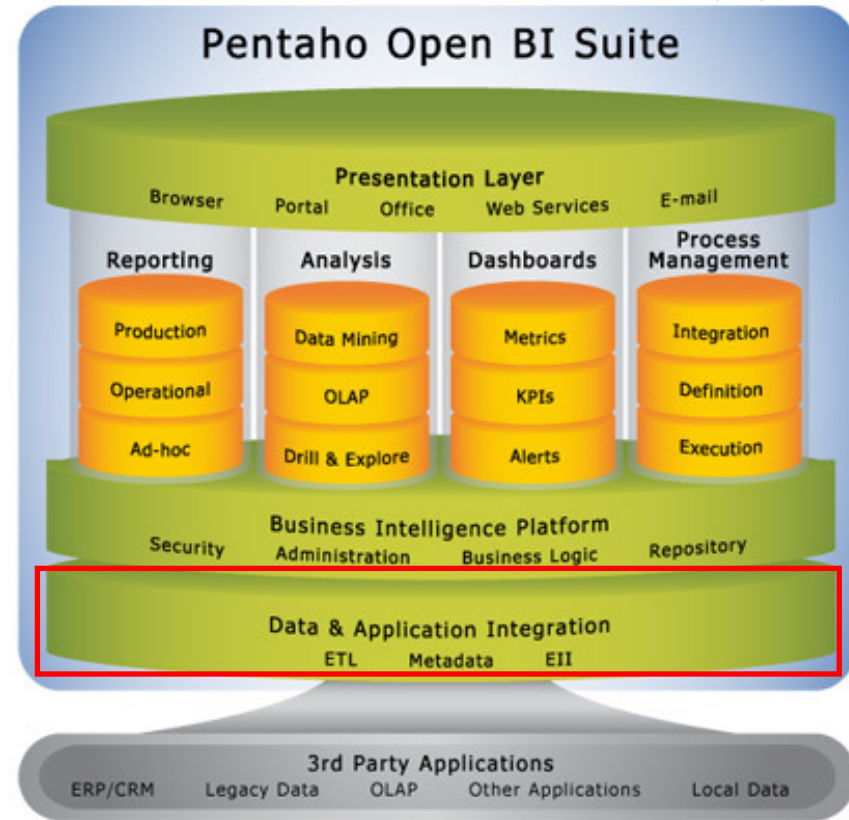


ETL mit
Java



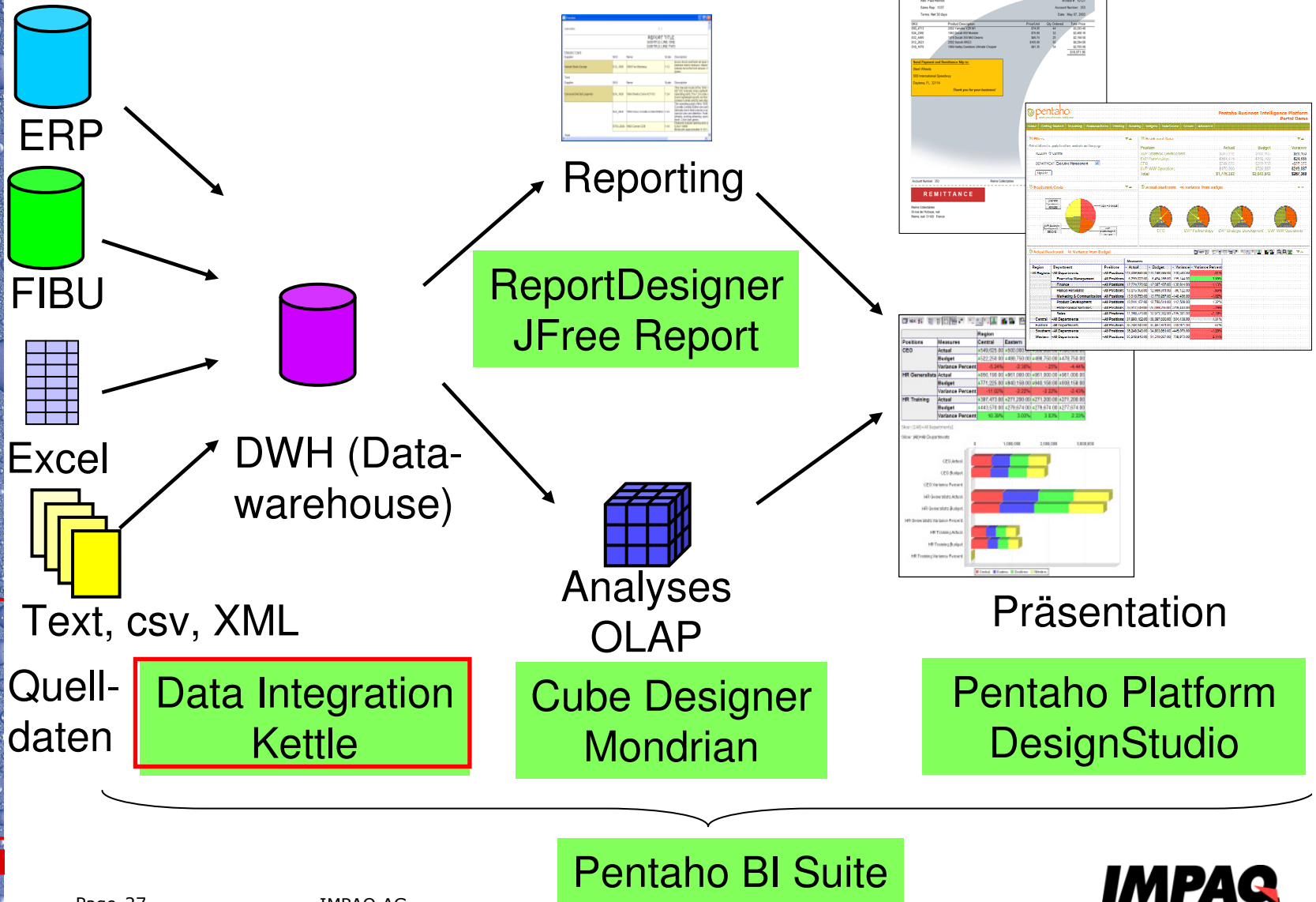
Pentaho Business Intelligence-Suite

- Metadata
- Data Integration (PDI – ehemals KETTLE)
- Reporting (ehemals JFreeReport)
- Analysis Services (Mondrian und WEKA)
- Dashboards



ETL mit
Java

Pentaho Business Intelligence-Suite



ETL mit Java

JasperIntelligence BI-Suite



JasperReports

Pixel-perfect reports for screen or print



Java Developers



JasperServer

Interactive and managed reporting



Business Users



iReport

Powerful graphical interactive & production report designer



BI Developers



JasperAnalysis

Interactive data analysis / OLAP



Power Business Users



JasperETL

High-performance data integration



ETL Developers

Kommt von Talend

ETL mit
Java

Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**

Open-Source-ETL im Überblick: untersucht



- Talend Open Studio (LGPL) 2.4
- Pentaho Data Integration 3.0 (ehemals KETTLE) (LGPL)
- Octopus Together Data Transformer 3.6 (LGPL)
- DBUNIT 2.2 (LGPL)



Open-Source-ETL im Überblick

Nicht betrachtet:

- Clover.ETL 2.4 (GPL für Engine)
- Apatar Data Integration/ETL 1.1 (GPL)
- SUN ETL Integrator (Open ESB JBI, NB-Plugin, GNU Classpath)





Was ist DBUnit?

Manuel Laflamme Feb 2002

Aktuelle Version 2.2.3 (LGPL) <http://dbunit.sourceforge.net>

Stellt Testdaten plattform- und datenbankunabhängig zur Verfügung (auch zur Migration geeignet)

Funktionen:

- Erzeugung und Validierung von DTD's für die Datenbank-Schemas
- Import/Export/ Vergleich von XML-Datasets
- CRUD-Operationen auf Basis von Datasets
- Ant-Tasks (operation type, Export, compare)

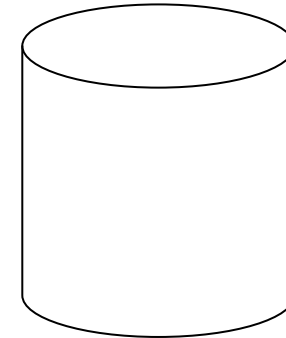
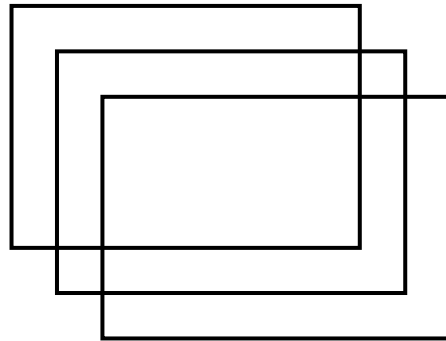
Verwendung:

- Testdatenbestand
- Regressionstests
- Migration im Kleinen

ETL mit
Java



DBunit - Ablauf



DataSets

- FlatXmlDataSet
- XmlDataSet
- XLS/CSV

Werkzeuge

- ANT
- MAVEN 1/2

JDBC-Datenbank

- mySQL
- MS SQL
- DB2, IFX, DERBY
- ORACLE
- PostgreSQL
- HSQL, H2
- Sybase

Datenbankverbindung einstellen:

profile.properties

```
dbunit.profile.driver.class = COM.ibm.db2.jdbc.app.DB2Driver
dbunit.profile.url = jdbc:db2:sample
dbunit.profile.user = db2admin
dbunit.profile.password = db2admin
```


Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**



Enhydra Octopus

- Gründung: 2002, Sinisa Milosevic, Together Teamlösungen EDV-Dienstleistungen GmbH
- Aktuelle Version: **3.6.5**
- Release: 25.10.2007
- Lizenz: LGPL v2.1
- Website:
forge.objectweb.org/projects/octopus/
- Einsatzgebiet: Migration, Testdaten
- Datenquellen: LDAP, CSV, XML, JDBC (MSSQL, MySql, Access, Excel, PostgreSQL, Qed, InstantDB, XML, BorlandJDataStore, Oracle, HSQL, McKoi, DB2, Sybase und Paradox)

ETL mit
Java



Enhydra Octopus

The screenshot shows the 'Together Data Transformer Application' window. The 'ExampleGenerator' tab is active, displaying configuration for a JDBC source and target database. The source database is configured as HypersonicSQL with driver 'hsqldb' and user 'SA'. The target database is configured as Csv. The output directory is set to 'C:\tdt-3.6\bin'. The log window at the bottom shows the completion of the 'LoaderJob.o1j' file generation.

Together Data Transformer Application
Project IDEOptions Help

ExampleGenerator

Application

TDT Generator TDT Loader

JDBC Output options Advanced

JDBC Source Database

Source database type: HypersonicSQL

Source database URL: C:\tdt-3.6\examples\ObjectLoader\LoaderInput\Database\TestDB

Source database driver name: hsqldb

Source database user: SA

Source database password:

JDBC Target Database

Target database type: Csv

Target database URL: C:\tdt-3.6\examples

Target database driver name: csv

Target database user:

Target database password:

Value mode: copy

Generator Output: C:\tdt-3.6\bin

Tables List:

Add new drivers: C:\tdt-3.6\lib\hsqldb.jar

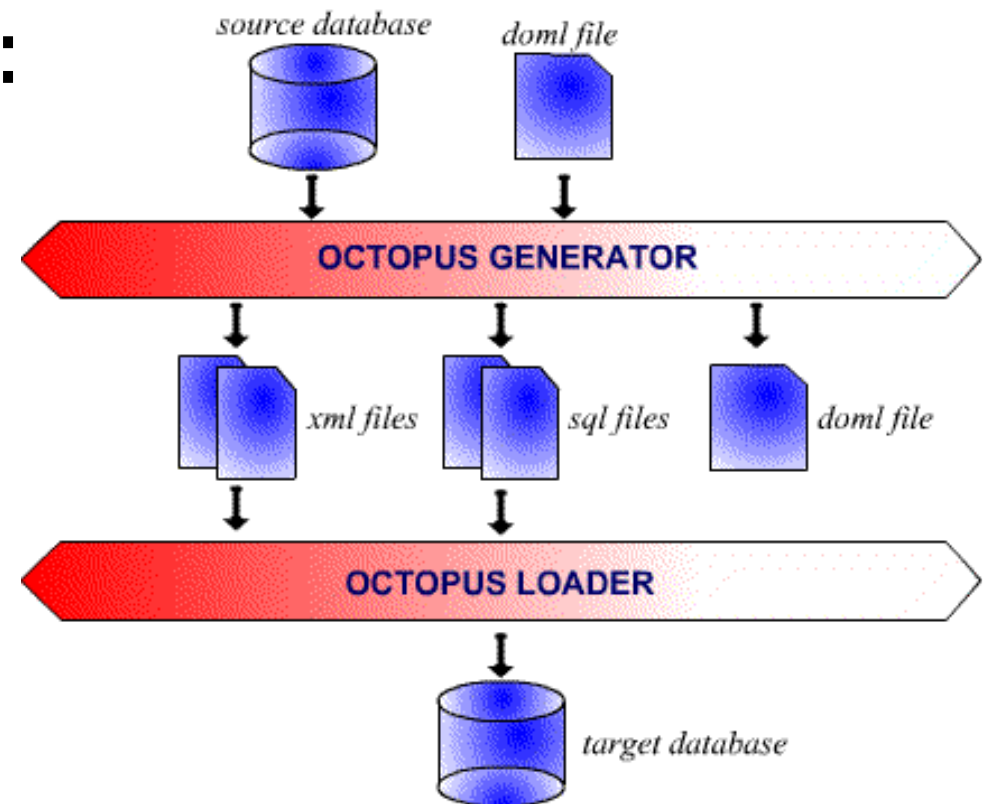
Log TDT Generator output ...

```
Generating of LoaderJob.o1j file is finished.  
  
Duration: 0,688 seconds  
  
OctopusGenerator finished.
```

ETL mit
Java

Enhydra Octopus: Überblick

- Together Data Transformer
- Bibliotheken:
 - Ant
 - FOP
 - HSQMLDB
 - Rhino
 - JUnit
 - log4j
 - P6Spy



Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



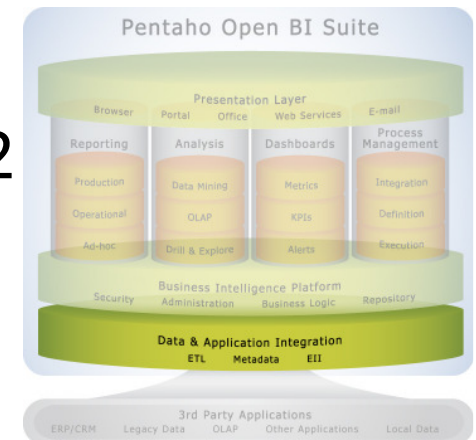
**ETL mit
Java**



Pentaho Data Integration (PDI) - Kettle

- Gründung: 2003 von Matt Casters
- Dezember 2005 Kettle wird Open-Source
- April 2006 Kettle-Übernahme durch Pentaho

- Aktuelle Version: 3.0.3/3.1.0-M2
- Release: 24.4.2008
- Lizenz: LGPL v2.1
- Website: kettle.pentaho.org



- Produkte: Pentaho Data Integration (ETL)
- - Kettle: (K) Extract Transport Transform Load (E)
- - Pentaho Meta Data

ETL mit
Java



Pentaho Data Integration - Funktionen

- 35 verschiedene Transformationsschritte und 10 Jobtypen
- Über 25 Datenbanken: MySQL, SQL Server, Oracle, DB2, PostgreSQL, MS Access, Sybase, Informix, MaxDB, Firebird, AS/400, Ingres, Caché usw.
- BulkLOADER: mySQL, (ORACLE, DB2 geplant)
- in einem Transformationsschritt: Slowly Changing Dimensions (Kimbal Typen I, II and III) und Junk Dimensions unterstützt
- Partitioning und Cluster je Schritt
- Keine parallelen Abläufe innerhalb von Jobs
- Keine Unterstützung für Wiederanlauf

ETL mit
Java



Pentaho Data Integration Bestandteile

- **Spoon:** Editor zur Definition von Schritten oder Transformationen
- **Kitchen:** Ausführung von Jobs
- **Pan:** Ausführung von Transformationen
- **Carte:** startet einen Remote-Server zur Transformationsausführung im Cluster
- Repository-basiert, Metadaten, Wiederverwendung, Steuerung



Pentaho Data Integration - Spoon

- Grafischer **SWT**-Editor
- Definition von Schritten oder Transformationen
- Konfiguration der Schritte
- Modellierung des Datenfluss durch Sprünge
- Metadaten basiert XML-Dateien oder DB-Repository
- Fehlerkorrektur

Pentaho Data Integration - Spoon

- Run



- Startet Schritt oder Transformation

- Preview



- Startet Transformation im Vorschaumodus

- Angabe Zeilenanzahl je Schritt

- Replay



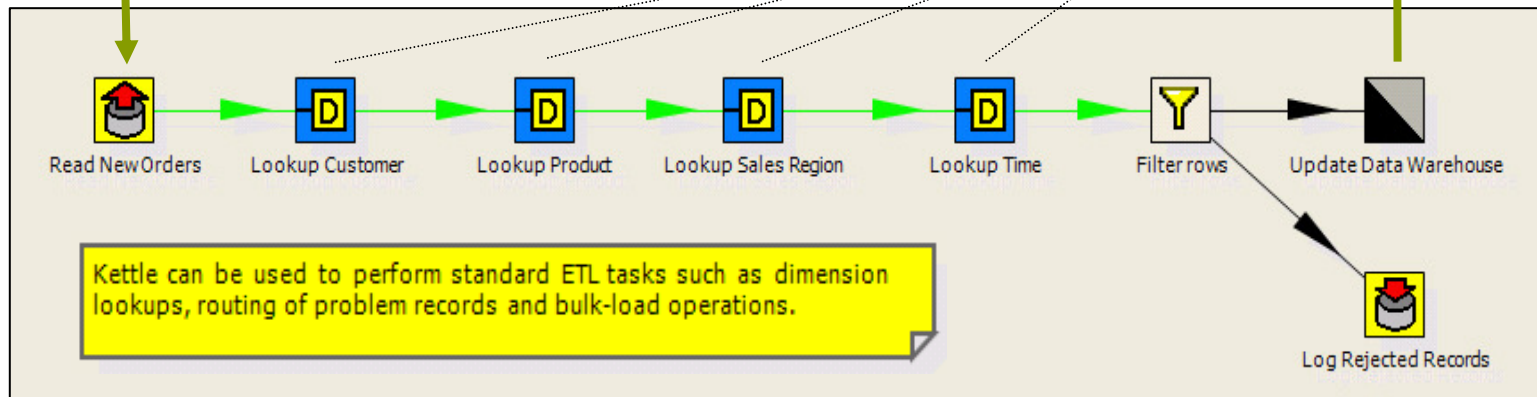
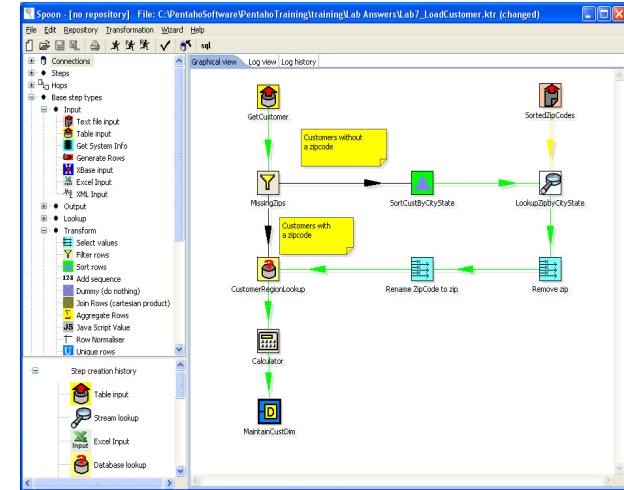
- Erneutes Ausführen eines zeitabhängigen Schrittes

- Debug



- Ermöglicht Debugging einer Transformation durch Haltebedingungen

Pentaho Data Integration Job Schritte



ETL mit Java

Agenda

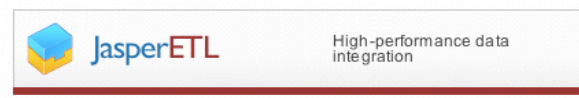
- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**

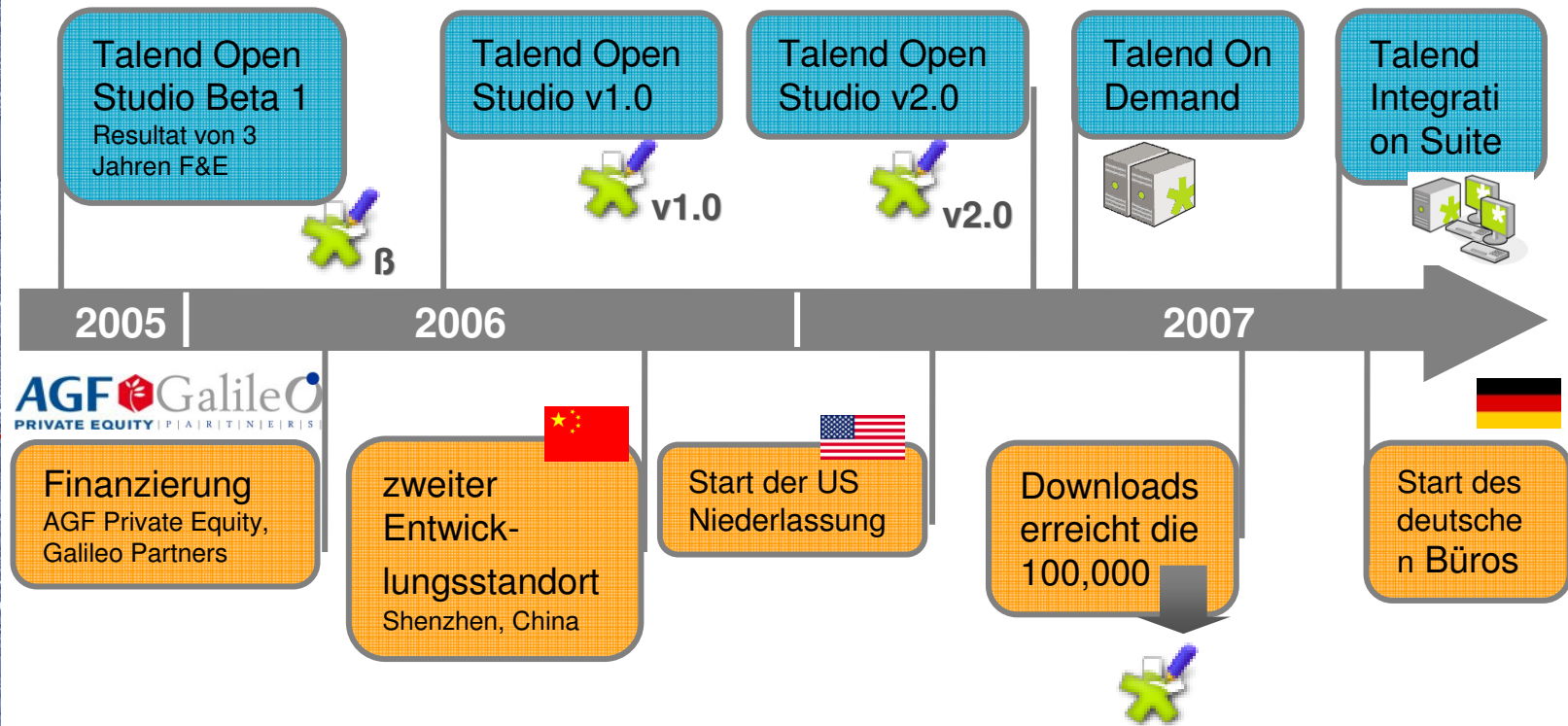
Talend – ETL-Generator

- Gründung: 2005 von Bertrand Diard und Fabrice Bonan (ehemals Informatica, Ascential)
- Orte: Suresnes, Niederlassungen in Nürnberg, Kalifornien und China
- Aktuelle Version: 2.4
- Release: 7.6.2008
- Lizenz: GPL v2
- Website: www.talend.com
- Produkte: Open Studio (auch als JasperETL), SaaS-Version 'Talend On Demand', Talend Integration Suite ist die Erweiterung von Talend Open Studio als Abonnement



ETL mit
Java

Talend's Geschichte



ETL mit Java

Produktphilosophie Talend

OpenSource Datenintegration mit neuesten Technologien

- Standards: Eclipse RCP, Java, Perl, SQL, SOA
- GPL v2 und Subskriptionen



Investitionsschutz

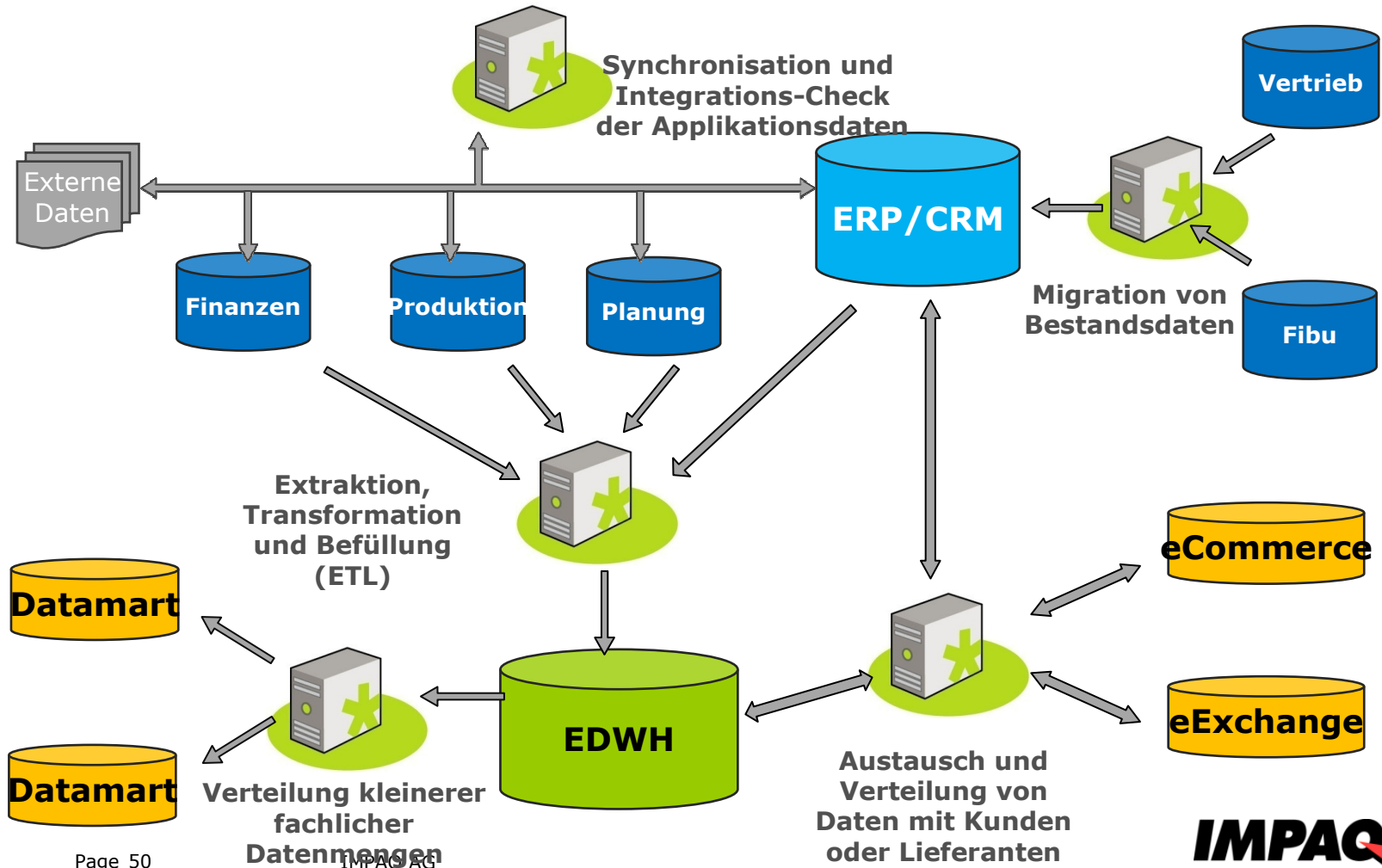
- Technischer Support, Expert Consulting und Schulung
- Kontinuierliche Investitionen in R&D

Sofortiger Nutzen

- Bessere Geschwindigkeit, hohe Skalierung
- Hohe Benutzerfreundlichkeit und umfangreiche Schnittstellen
 - Zweckmäßiger Einsatz
 - Kürzere Entwicklungszeiten
- Risikominimierung

**ETL mit
Java**

Datenintegration: Einsatz von Talend

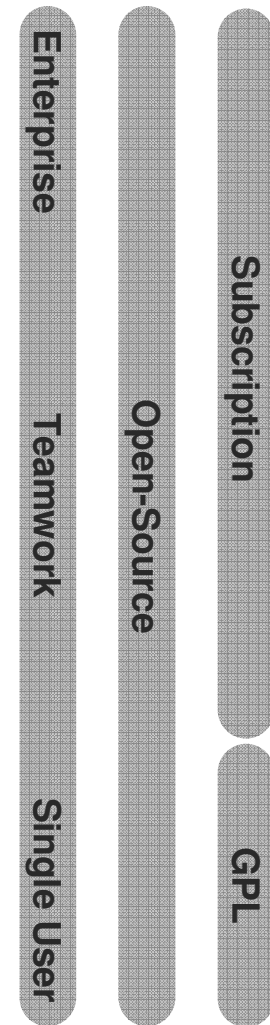


ETL mit Java

Talend's Lösungen



**ETL mit
Java**



Vielseitige Möglichkeiten / Schnittstellen

- Zahlreiche Schnittstellen zur Unterstützung diverser Systeme
 - **150+** direkt einsetzbare Konnektoren
- Geschäftsanwendungen:
 - SugarCRM, Salesforce.com, LDAP, Sage, Microsoft Dynamics, SAS, SAP (Q4/08)...
- Datenbanken / RDBMS/ Reports/ (B)-ulkLOADER (ELT)/ SCD:
 - Oracle (B; SCD), MS SQL (SCD), MySQL, DB2 (AS/400, SCD), SQL Server, Sybase (SCD), Ingres (SCD), PostgreSQL (SCD), Terradata (B), Palo, Mondrian, JasperSoft ...
- Web:
 - Web Services, FTP, HTTP, POP3, SMTP...
- Dateiformate:
 - Delimited, positional, XML (SAX), Excel...
- keine Unterstützung von parallelen Abläufen innerhalb eines ETLJobs
- Debuggen des Java-Codes



ETL mit
Java

Grafisches Design der Datenabbildung

The screenshot displays the Talend Open Studio interface for designing a data mapping job. The main workspace is divided into several panels:

- Schema editor:** Shows two source schemas, 'row1' and 'row2', with columns like 'id', 'name', 'quantity', 'flag', 'creation', and 'delivery'. 'row2' has an 'Inner join' checkbox checked.
- Expression editor:** A central table for defining expressions and variables:

Expression	Variable
<code>\$row1[flag]</code>	flag
<code>uc \$row1[name]</code>	name
- Output table:** A table defining the output columns and their corresponding expressions:

Expression	Column
<code>\$row1[flag]</code>	flag
<code>\$row1[id]</code>	id
<code>\$Var[name]</code>	name
<code>\$row2[id]</code>	id_1
<code>\$row2[name]</code>	name_1
- RejectInnerJoin component:** A table for defining inner join rejection logic:

Expression	Column
<code>'lookup inner join reject'</code>	RejectFlag
<code>\$Var[flag]</code>	flag
<code>\$row1[id]</code>	id
<code>\$Var[name]</code>	name
- RejectFilter component:** A table for defining filter rejection logic:

Expression	Column
<code>'Reject by filter'</code>	flagReject
<code>\$Var[flag]</code>	flag
<code>\$row1[id]</code>	id
<code>\$Var[name]</code>	name

At the bottom, there are two tables for schema details:

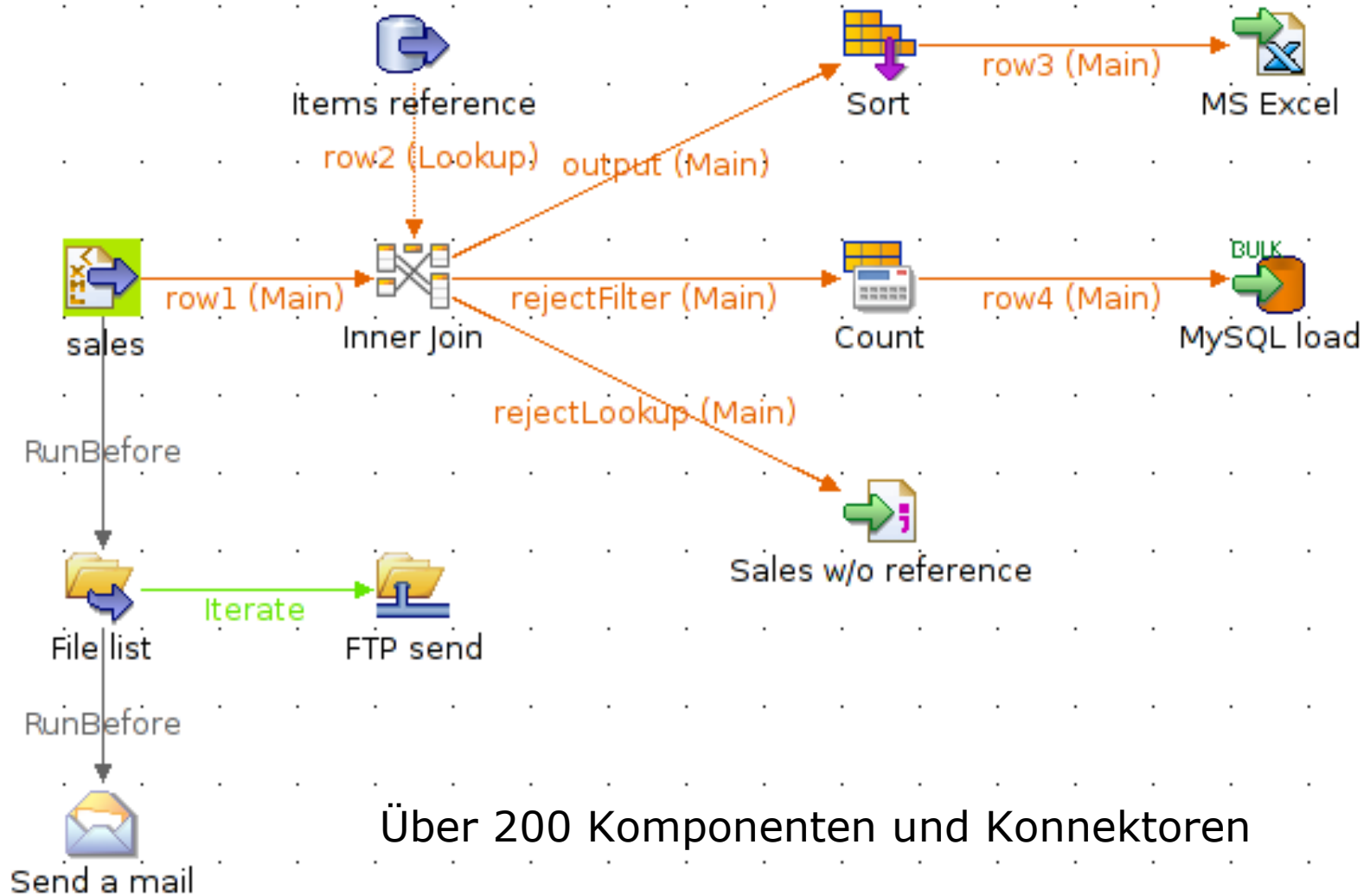
- row1 schema:**

Column	Key	Type	Length	Precision	Nullable	Co...
id	<input checked="" type="checkbox"/>	int	11	0	<input checked="" type="checkbox"/>	
- Output schema:**

Column	Key	Type	Length	Precision	Nullable	Co...
flag	<input type="checkbox"/>	int	1	0	<input type="checkbox"/>	

ETL mit
Java

Grafisches Design der ETL-Schritte



Über 200 Komponenten und Konnektoren

**ETL mit
Java**

Metadaten werden im Metadata Repository abgelegt

Schema Table

New Schema Table in connection "demoConnection"

Add a Schema on repository

Schema

- customer
- customer1
- demotable
- majuscule
- state2

Name: customer1

Comment:

Based on table: customer

Retrieve Schema

Use the "Retrieve Schema" button to replace the current Schema by the table based Schema

Schema

Column	Key	DB Type	Type	Nullable	Length	Precision	D...	Co...
id	<input type="checkbox"/>	INT	int	<input checked="" type="checkbox"/>	2			
Total	<input type="checkbox"/>	INT	int	<input type="checkbox"/>	20		0	
CustomerName	<input type="checkbox"/>	VARCHAR	String	<input checked="" type="checkbox"/>	30			
CustomerAddress	<input type="checkbox"/>	VARCHAR	String	<input checked="" type="checkbox"/>	25			
idState	<input type="checkbox"/>	INT	int	<input checked="" type="checkbox"/>	2			
id2	<input type="checkbox"/>	INT	int	<input checked="" type="checkbox"/>	2			
RegTime	<input type="checkbox"/>	VARCHAR	String	<input checked="" type="checkbox"/>	14			
RegisterTime	<input type="checkbox"/>	VARCHAR	String	<input checked="" type="checkbox"/>	23			
Somme1	<input type="checkbox"/>	FLOAT	float	<input checked="" type="checkbox"/>	12			
Somme2	<input type="checkbox"/>	FLOAT	float	<input checked="" type="checkbox"/>	12			

Add Schema

< Back Next > Finish Cancel

ETL mit
Java

Agenda

- Einleitung:
 - Was ist eigentlich Open-Source?
 - Was ist Business Intelligence?
 - Datenintegration als Herausforderung
 - Was ist ETL (Extract, Transform, Load)?
- Anforderung an ETL
 - Unterschied zu EAI
 - Prozess, Funktionalitäten
 - Einsatzgebiete
 - Kriterienkatalog
- Der BI-Markt und Open-Source
- Open-Source-BI-Suiten von Pentaho, JasperSoft
- Produkte im Einzelnen
 - DBunit
 - Octopus
 - Kettle
 - Talend
- Bewertung, Ausblick



**ETL mit
Java**

Fazit Open-Source-ETL-Werkzeuge

- Octopus: für Datenmigration, einfache Replikation, Integration geeignet
- DBUNIT: sehr gut für kleinere Datenmengen, Testdaten geeignet
- Talend Open Studio:
 - Für operationales BI/ELT geeignet
 - Eine große Anzahl Konnektoren
 - Java-Code-Generator
 - Kommerzielle Zusatzprodukte
- Pentaho Data Integration:
 - Integrierte sich sehr gut in die Pentaho BI-Suite
 - Modell wird zur Laufzeit interpretiert
 - SAP-Konnektor
 - Ausgereifte ETL-Lösung zur Erstellung, Ausführung, hoch skalierbar

**ETL mit
Java**

Fazit Open-Source-ETL-Werkzeuge

- Modellieren, statt programmieren
 - Grafische Abbildung des Prozesses
 - Metadaten zur Überwachung des Prozesses
- Mit ETL wird Datenintegration beherrschbar
- Fertige Schnittstellen erleichtern Integration
- Unterstützung vieler Datenquellen, -Formate, Schnittstellen (ERP- und CRM-Systeme, WebServices)

- **Vorteile:**
 - Gute Integrierbarkeit
 - Flexible Einsatz-, Anpassbarkeit
 - Unabhängigkeit
 - geringe TCO und schneller ROI

- **Aktuelle Nachteile:**
 - Mangelnde Parallelisierung
 - Keine Restartfähigkeit
 - Dokumentation ist nur auf Englisch vorhanden

ETL mit
Java

Fazit Open-Source-BI-Markt

- Der BI-Softwaremarkt bleibt weiter in Bewegung
- Auch bei einer SOA-Architektur bleiben Daten wichtig
- In Deutschland ist eine hohe Akzeptanz für Open-Source Produkte
- Durch JasperSoft und Pentaho wird der OS-BI-Markt professionalisierter
- Europa scheint ein guter Herkunfts- und Einsatzort für Open-Source-ETL-Tools zu sein
- Migration und Integration bestehender Daten ist auch für Java-Anwendungen ein wichtiges und kein triviales Thema
- Einfache Integrier-, Portierbarkeit durch Open-Source Lizenz und Java

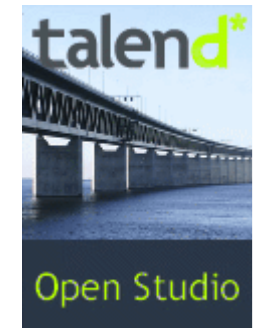


**ETL mit
Java**



Referenzen: Produkte

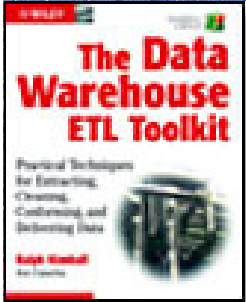
- Kettle Projektseite:
http://www.javaforge.com/proj/summary.do?proj_id=318
- Blog von Matt Casters:
<http://www.ibridge.be>
- Talend: <http://de.talend.com>






Referenzen: allgemein

- The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data, Ralph Kimball: John Wiley & Sons, 2004
- <http://www.etltool.com//etltoolslist.htm>
- ETL tools running in Linux: <http://linuxetl.com/>



ETL mit
Java

Referenzen: Analysten

- Forrester Research, November, 2007: Market Overview: Open Source ETL Tools, An Attractive Alternative To Custom Code 
- Gartner 2008: Magic Quadrant for Business Intelligence Platforms 
- Aberdeen Group Research, Dezember, 2007: The TCO of Business Intelligence - Open Source Takes on Traditional BI 
- Bloor: Talend Open Studio - an independent analyst review 

ETL mit
Java

Datenschaufler für Java

Vielen Dank für Ihre Aufmerksamkeit



OS macht Freude !!!

IMPAQ

IMPAQ AG

Emil-Figge-Str. 85
44227 Dortmund

www.impaqgroup.com

Mobil +49 175 1613922

Phone +49 (231) 9704-218

Fax +49 (231) 9704-199

Frank Pientka

Senior Berater

frank.pientka@impaqgroup.com

**ETL mit
Java**