



# ETL-Prozess mit **talend\*** im Produktivbetrieb

*Christian Kolodziej*



## *Zu meiner Person*

- Studium Informatik (Diplom) bis Januar 2009
  - Vorlesung Business Intelligence im 8. Semester
  - Diplomarbeit „Analytisches CRM im Bereich der Online-Medien“  
beim Huber Verlag für Neue Medien GmbH
- Studium Wirtschaftsinformatik (Master)
  - Im 2. Semester
  - Master-Thesis im Frühjahr 2010
- Sonstiges: Fußball-Schiedsrichter, Internet-Programmierung, Mac-Nutzer, Halbmarathon-Läufer, Jugendarbeit





## *Huber Verlag für Neue Medien GmbH*

- Anbieter für Direkt-Informationendienste und Pressedienstleistungen
- Ziel:
  - Unterstützung der Kunden bei deren Pressearbeit
  - Aktive Verbreitung von Pressemeldungen
- 50 Mitarbeiter am Standort Karlsruhe

*aktiv*  
**verzeichnis**

**seminar** SPIEGEL

 **PRESSEBOX**<sup>®</sup>

**itit**  
ititpro.com

**life** 

*initiative*  
**mittelstand**

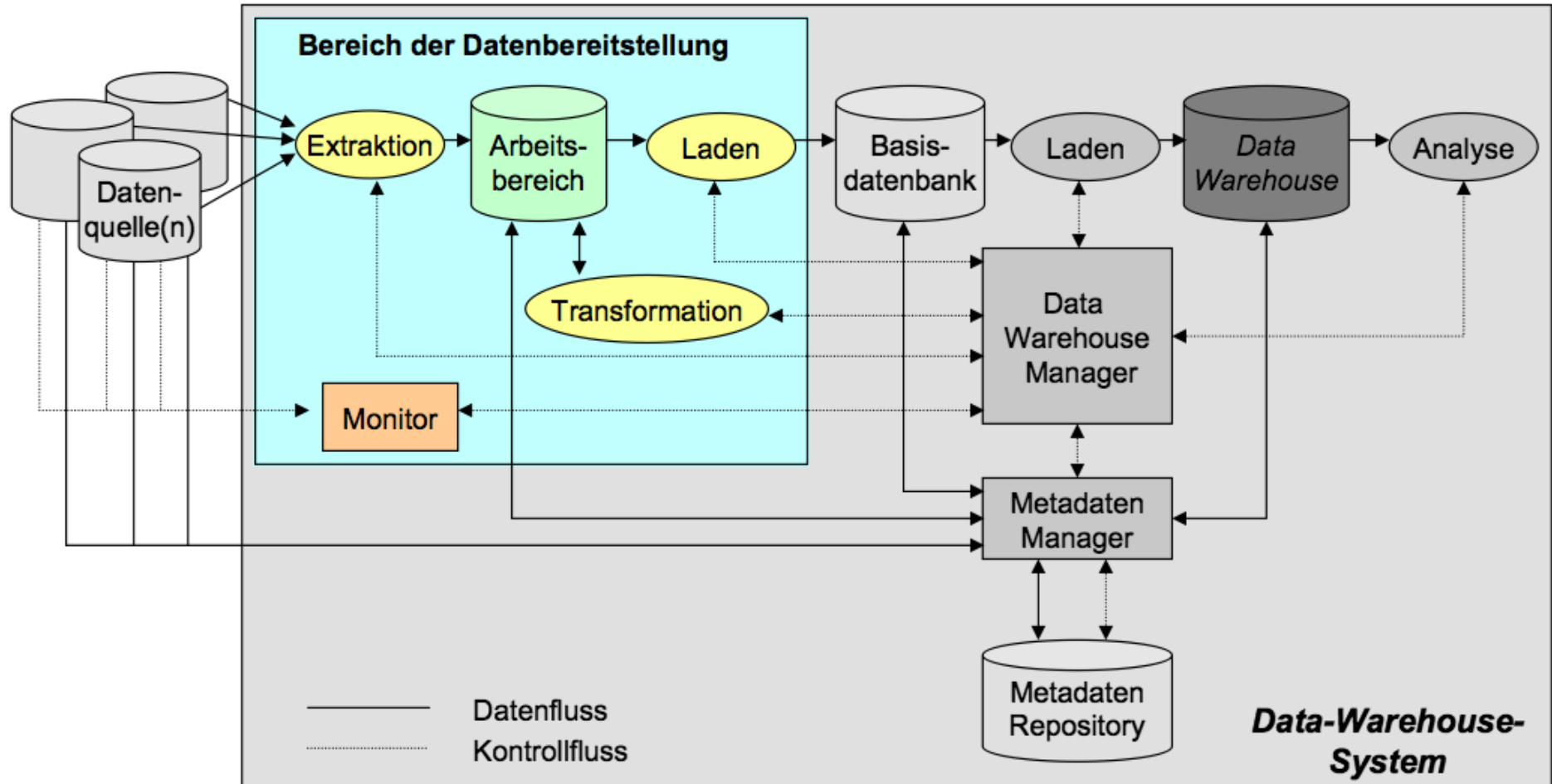


## *Agenda*

- Einführung: ETL-Prozess in einem Data Warehouse System
- Die Firma Talend
- Das ETL-Tool Talend Open Studio
  - Features
  - Live-Demo
  - Stärken und Schwächen
- Einsatz und Ausblick beim Huber Verlag für Neue Medien



*Einführung: Referenzarchitektur nach Bauer/Günzel (2004)*





## *Die Firma Talend*

- Marktführer im Bereich Open-Source-Datenintegration
- Weltweite Niederlassungen und Kunden
- Deutsche Niederlassung in Nürnberg
  
- Produkte
  - Talend Open Profiler (Datenqualitätsbestimmung)
  - **Talend Open Studio (ETL-Tool)**
  - Talend On Demand

**talend\***  
\*open data solutions

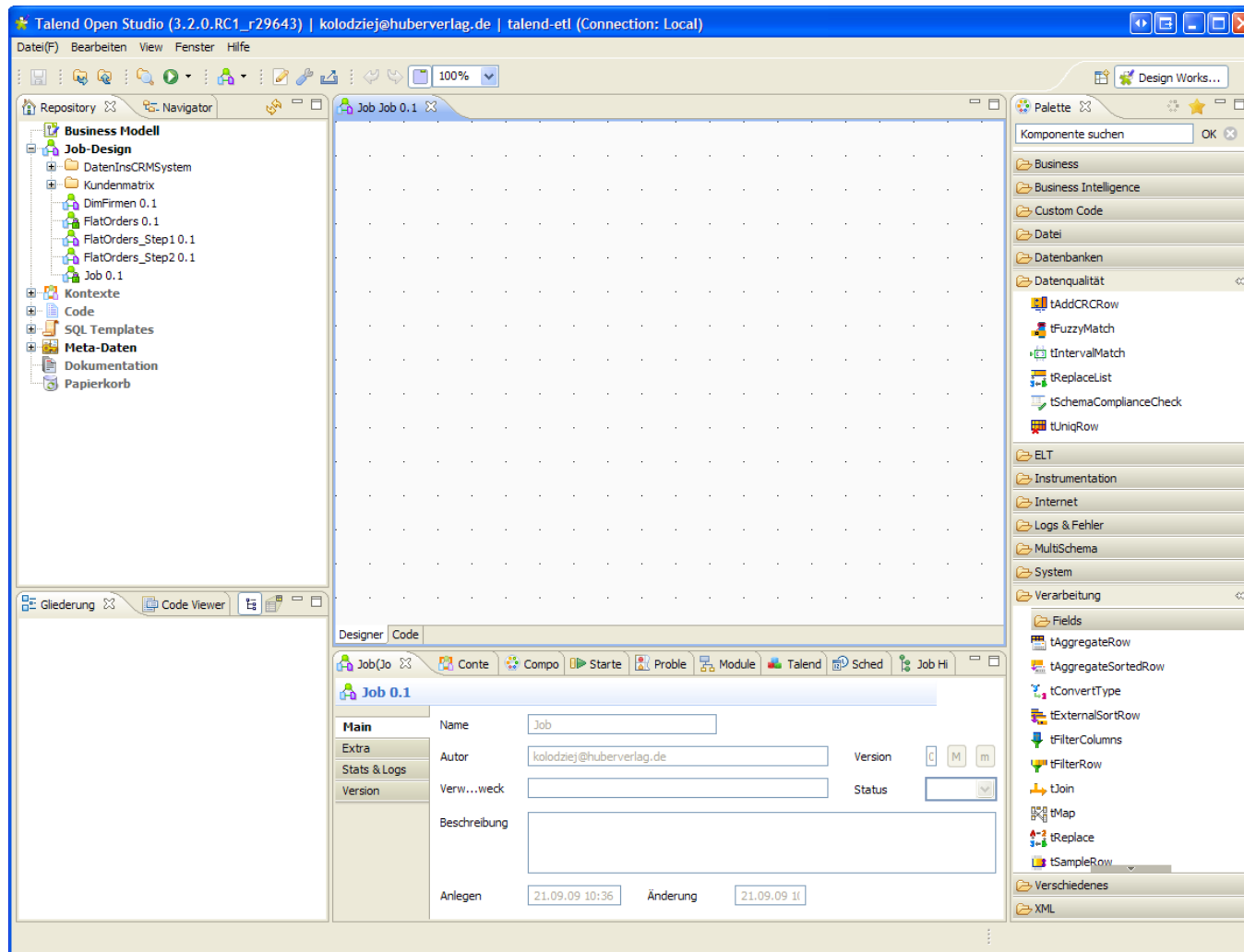


## *Talend Open Studio (1)*

- Auf Eclipse-Basis und damit plattform-unabhängig
- Aktuell in der Version 3.1.4 vom 17.09.2009
- Konzept: Commercial Open Source
  
- Features, über 200 Komponenten, z.B. für
  - Prozesssteuerung
  - Datenmanipulation
  - Input/Output
  - Debugging
  - Dokumentation



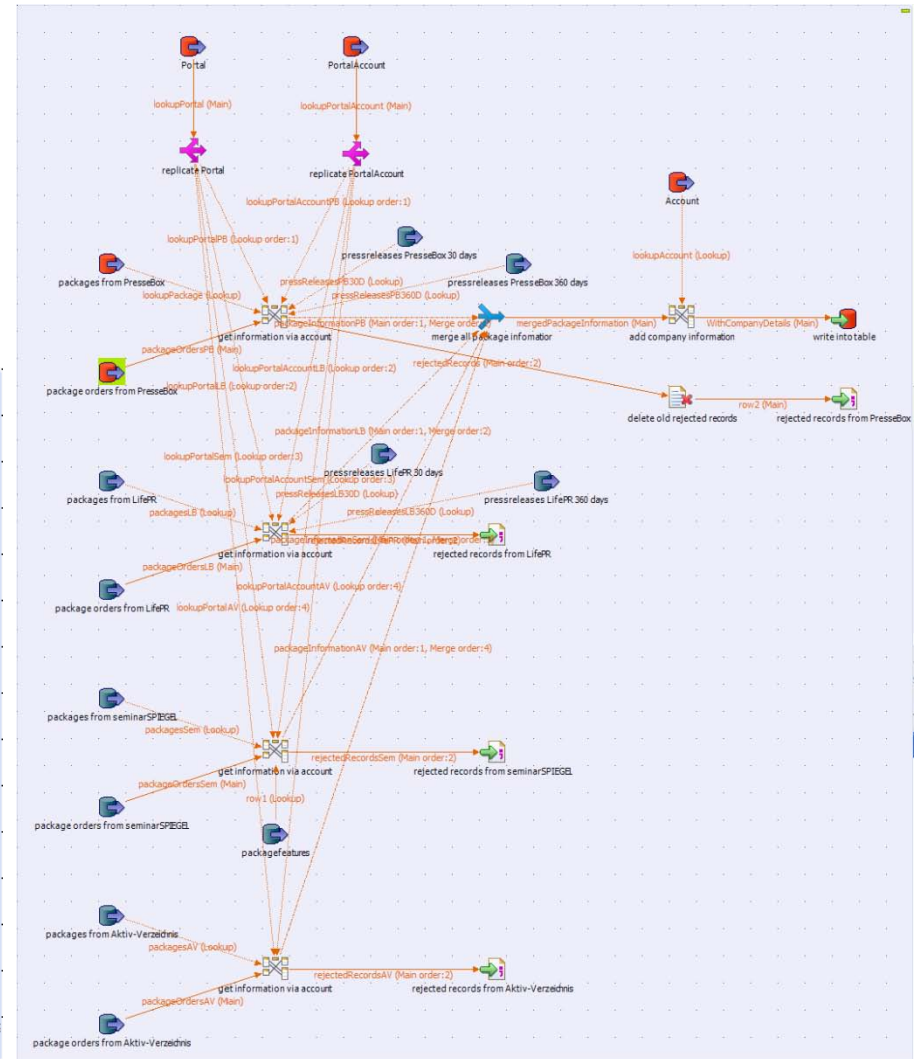
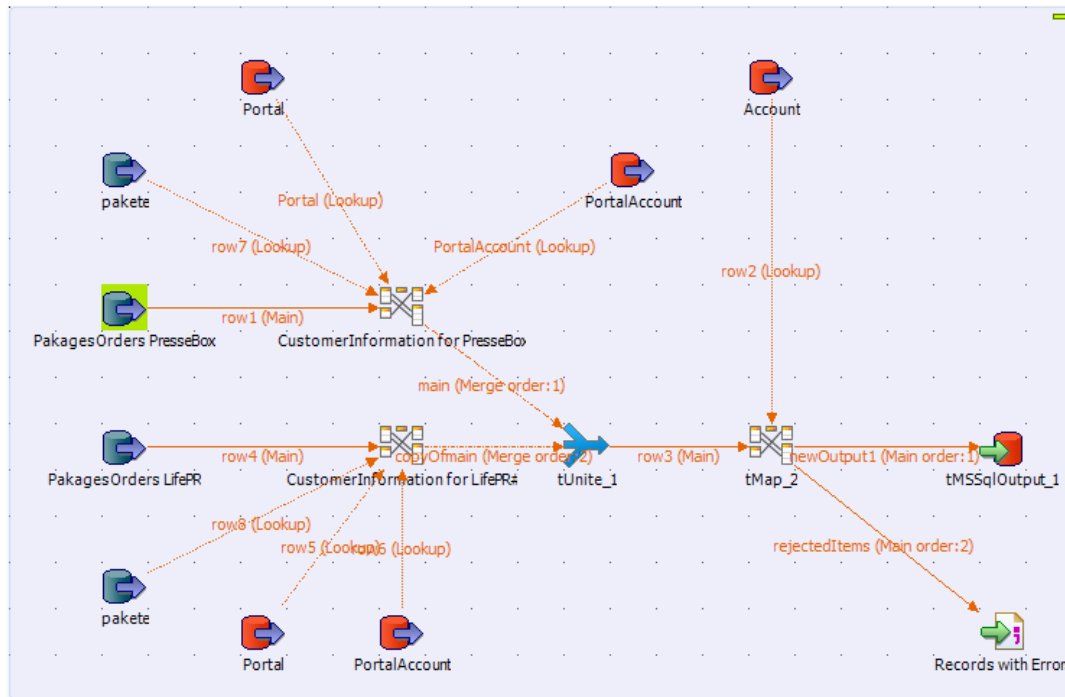
## Talend Open Studio (2) – Die Benutzeroberfläche







## Talend Open Studio (3) – Beispiel-Jobs





*Talend Open Studio (4) –  
Genug der Theorie ...*

**talend\*** in Aktion





## *Talend Open Studio (5) – Schwächen*

- Es ist nicht alles Gold was glänzt
  - Metadatenmanagement
  - Typumwandlungen durch unterschiedliche Datentypen
  - Fehlermeldungen ohne Aussagekraft
  - Teilweise nur Grundfunktionalität, z.B. Palo
  - Unvollständige Übersetzung



## *Talend Open Studio (6) – Stärken*

- Leistungsfähiges Open-Source-Programm
  - Großer Funktionsumfang durch über 200 Komponenten
  - Leichte Erweiterbarkeit durch Java-Code
  - Eigene Komponenten möglich durch Quelloffenheit
  - Automatisierbarkeit
  - Aktive Weiterentwicklung und aktive Community#
  - Automatische Dokumentation
- **Summa summarum:** Das Talend Open Studio ist schon in der kostenlosen Open-Source-Variante ein ausgereiftes und stabiles Tool, das die wesentlichen Funktionen enthält und somit den grundsätzlichen Anforderungen gerecht wird.



## *Talend Open Studio im Produktivbetrieb*

- Einsatz der Open-Source-Variante ohne kommerziellen Support beim Huber Verlag für noch Neue Medien
  - Einsatz des Talend Open Studio als reines ETL-Tool
  - ETL-Jobs laufen als Cronjobs auf Linux-Server
  - Excel Pivot-Tabellen als Frontend statt vollwertige BI-Lösung
  - Kennzahlen für CRM-System generieren und aktualisieren
- Zukunft?!
  - Weiteres Vorgehen noch nicht beschlossen
  - Talend soll unabhängig vom zukünftigem Data Warehouse und Frontend weiterhin den ETL-Part übernehmen



## *Nützliche Daten und Links*

- Weitere Links und Infoquellen zum Talend Open Studio
  - Talend: [www.talend.com](http://www.talend.com)
  - Webinare (auch On-Demand): [www.talend.com/webinar](http://www.talend.com/webinar)
  - Talend Roadshows
  - Forum zum Talend Open Studio: [www.talendforge.org](http://www.talendforge.org)
  - Web 2.0ig: Blog ([www.talend.com/blog](http://www.talend.com/blog)) und Twitter (@talend)
- Meine Kontaktdaten
  - eMail: [mail@christian-kolodziej.de](mailto:mail@christian-kolodziej.de)
  - Internetseite: [www.christian-kolodziej.de](http://www.christian-kolodziej.de)
  - ... oder suchen Sie mich bei **XING** 



*... für Ihre Aufmerksamkeit!*

Fragen?

Domandas?

Preguntas?

Any questions?

Förfrågan?

?

Perguntas?

Vraagen?

D'autres questions?