

# Talend Open Studio

# v3

Talend vereinfacht die Datenintegration.

Mittels einer offenen und innovativen Plattform bietet Talend eine einfache und leistungsfähige Lösung.

Durch die hohen Kosten von proprietären Lösungen wurden eine Reihe von Unternehmen gezwungen, ihre Datenintegrationsbedürfnisse mittels Eigenentwicklungen und Routinen zu stillen. Meistens jedoch decken diese Integrationsroutinen nur eine ganz spezielle Anforderung ab, ohne einen globalen Ansatz oder eine konsolidierte Entwicklung zu fokussieren und ohne Berücksichtigung bereits existierender Prozesse. Dieses Verfahren erzeugt einen hohen Wartungsaufwand, beeinträchtigt die Stabilität, gefährdet die Integrität und Konsistenz von Informationssystemen und spiegelt sich in hohen Wartungskosten wieder.

## Eine vielseitige Datenintegrationslösung

Talend Open Studio ist eine leistungsstarke und vielseitige Open Source Lösung, welche alle Belange der täglichen Datenintegration abdeckt:

- Synchronisation oder Replikation von Datenbanken
- Pünktlicher Austausch und Verteilung von Daten innerhalb der IT-Umgebung
- ETL (Extraktion, Transformation, Laden) für analytische Systeme
- Datenmigration
- Aufwendiges Daten-Transformieren und Laden
- Datenqualität

Talend Open Studio besteht aus drei Modulen (Business Modeler, Job Designer und Metadata Manager), welche mittels der einheitlichen graphischen Entwicklungsumgebung, auf Eclipse basierend, sehr einfach zu bedienen sind. Anwender erhalten aktive Unterstützung von der Talend Community und können auf Contributions direkt zugreifen.

## Geschäftsorientierte Prozessmodellierung

Talend Open Studios Business Modeler unterstützt den Top-Down-Ansatz. Dieser ermöglicht die Einbindung von Geschäftsanwendern in das Design des Integrationsprozesses und zeigt die Fortschritte während der Umsetzung an.

Der Business Modeler verbindet alle für den Datenintegrationsprozess relevanten Dokumentationen in einer anwenderfreundlichen Übersicht. Diese ermöglicht ein äußerst effizientes Überwachen und Analysieren von Jobs, falls einmal ein Fehler auftreten sollte.

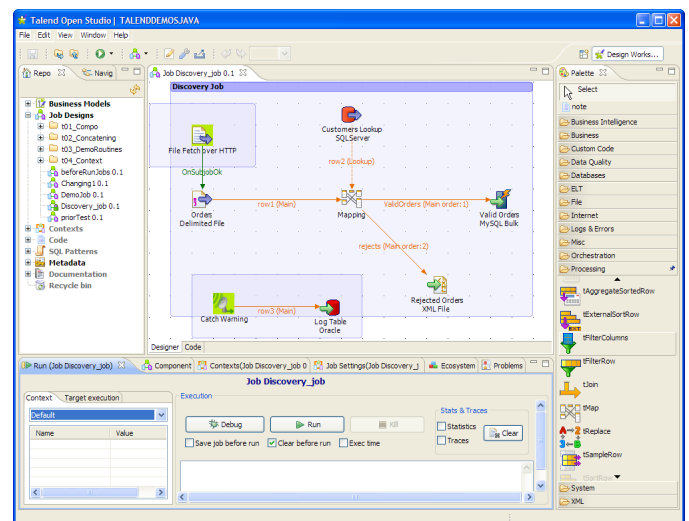
## Graphische Entwicklung

Talend Open Studios Job Designer bietet sowohl eine graphische als auch eine funktionale Sichtweise des aktuellen

Integrationsprozesses. Der Job Designer enthält eine graphische Darstellung aller verfügbaren Komponenten und Konnektoren – die Komponentenbibliothek.

Integrationsprozesse werden mittels einfacher Drag-and-Drop-Operationen erstellt. Hierzu werden Komponenten und Konnektoren auf den Arbeitsbereich gezogen, die Verbindungen graphisch definiert und die Eigenschaften festgelegt.

Komponenten und Konnektoren bieten eine Basis für die jeweilige Datentransformationsaufgabe, können aber auch in der Gesamtprozesssteuerung eingesetzt werden.



Talend Open Studio macht keine Unterscheidung zwischen Transformation und Ablaufsteuerung. Eigenschaften von Komponenten werden direkt bei der Auswahl oder durch den Metadata Manager festgelegt. Um die Darstellung von Jobs zu verbessern, können Prozessschritte in einzelne Subjobs unterteilt und in einer Parent-Child Beziehung ausgeführt werden. Verschiedenste Log- und Protokollierungsmechanismen können direkt an den jeweiligen Komponenten aktiviert werden.

## Metadatengestütztes Design und Ausführung

Talend Open Studio ist eine, von allen Modulen verwendete, metadatengestützte Lösung, in der alle Informationen in einem Repository gespeichert und verwaltet werden. Dieses zentrale Metadatenrepository verbindet alle Projektinformationen und gewährleistet die Konsistenz über alle Integrationsprozesse hinweg.

Der für die Integrationsprozesse notwendige Bezug von Quell- und Ziel-System wird durch eine Reihe von Wizards

unterstützt. Hierbei werden die Metadaten durch erweiterte Datenbank- und Dateianalysen einfach ins Metadatenrepository geladen.

Das Metadatenrepository beruht auf einem offenen relationalen Modell und gewährleistet die Wartbarkeit durch die Analyse von möglichen Abhängigkeiten und Beeinträchtigungen.

### Zukunftsweisende und vielseitige Konnektivität

Talend Open Studio beinhaltet native und geschäftliche Konnektoren für alle IT-Umgebungen. Dieses breite Spektrum von verschiedensten Treibern ist der Schlüssel für ein erfolgreiches Zusammenarbeiten von Applikationen und Datenbanken. Es erlaubt die Verbindung von unterschiedlichsten und heterogenen Datenstrukturen ohne Geschwindigkeitsverluste. Durch kontinuierliche Erweiterungen wächst der Umfang der Lösung weiter an.

#### Einige Beispiele der enthaltenen Konnektoren:

**Dateien:** CSV, Excel, positional getrennt, XML, LDIF, unstrukturiert, etc.

**Datenbanken:** MySQL, MS SQL, DB2, Oracle, Ingres, PostgreSQL, Sybase, Vertica, ParAccel, AS series, JDBC, etc.

**Applikationen:** SAP, SugarCRM, MS Dynamics, MS AX, etc.

**Log & Error:** info, warning, statistics, filtering, etc.

**Datenqualität:** deduplication, fuzzy logic matching, CRC, etc.

**Transformationen:** filtering, mapping, reference, aggregation, lookup, XSLT, (de)normalizing, etc.

**Internet & Protokolle:** Web Services, FTP, HTTP, SSH, POP, etc.

**Steuerung:** iteration, loop, replication, unification



Eine Übersicht aller verfügbaren Konnektoren ist unter <http://www.talendforge.org/components> abrufbar.

Talend Open Studio beruht auf den standardisierten Sprachen: Java, Perl und SQL. Diese bieten Kunden eine einfache Möglichkeit, die bestehenden Komponenten zu verändern oder die Lösung um Eigenentwicklungen direkt zu ergänzen. Talend Open Studio bietet einen direkten Zugriff auf Talend Exchange, eine Plattform auf der neu entwickelte Komponenten, Programme und Codeteile frei zugänglich heruntergeladen und verwendet werden können

### Real-time debugging

Talend Open Studio beinhaltet leistungsfähige Test-, Debugging- und Tuning-Funktionen, welche einen sofortigen Einblick über den gesamten Ablauf zur Laufzeit ermöglichen. Neben einem erweiterten Trace-Mode werden auch direkt Statistiken über die eigentliche Ausführung erzeugt.

Und natürlich ist der von Talend Open Studio erzeugte Programmcode, unabhängig von der Sprache, in der Entwicklungsumgebung immer zugänglich und einsehbar.

### Verteilung und Wartung

Ein spezielles Kontextmanagement (Test, Staging, Produktion, etc.) unterstützt bei der Umsetzung und Entwicklung von Integrationsprozessen. Das direkte Laden der Kontextpara-

meter im Job Design vereinfacht die Entwicklung und Steuerung der unterschiedlichen Ausführungsumgebungen. Auch die Prozessverteilung innerhalb von Unternehmen kann mit Hilfe eines Export Tools einfach durchgeführt werden.

Die automatische und vollständige Dokumentierung (in XML und HTML) liefert eine immer aktuelle technische Beschreibung und hilft sowohl Nutzern als auch Geschäftsanwendern bei der Wartung und Aktualisierung von weiteren Prozessen. Die Folgeanalyse ermöglicht Nutzern die Identifikation der jeweiligen Beziehungen zwischen den mit Talend Open Studio entwickelten Integrationsprozessen und vereinfacht ein globales Update der im Repository zentral gespeicherten Transformationen.

### Stabile und skalierbare Ausführung

Während viele Integrationslösungen auf einem zentralen Integrationsserver beruhen oder die Datenverarbeitung nur innerhalb der Datenbank ausführen können, verteilt Talend Open Studio die Prozesse je nach Verfügbarkeit über den gesamten Rechnerverbund. Der Einsatz von dedizierten Systemen ist hierbei nicht notwendig. Stattdessen werden von Talend Open Studio alle vorhandenen Ressourcen genutzt. Talend Open Studio ist die einzige Datenintegrationslösung, welche sowohl den traditionellen ETL (Extraktion, Transformation, Laden) als auch den bekannten ELT (Extraktion, Laden, Transformation) Ansatz unterstützt. ELT nutzt die Stärken relationaler Datenbanksysteme, indem die Transformationen direkt innerhalb der Datenbank ausgeführt werden. Dadurch können große Datenmengen in ungeahnter Geschwindigkeit verarbeitet werden.

*Neben der großen Flexibilität und Zeiteinsparung, ermöglicht uns die Nutzung von Talend Open Studio die Wiederverwendung von Integrationsprozessen bei nachfolgenden Projekten. Das bedeutet für uns, dass wir nach einer Erstinvestition die bereits getane Arbeit nutzen können, um Projekte, die sich aufgrund der Quelldaten und Struktur unterscheiden, schneller ausführen zu können, als wenn wir wieder von vorne begonnen hätten.*

Bei jedem Teilschritt eines Prozesses kann das beste Verfahren für höchste Geschwindigkeit und Skalierbarkeit gewählt werden. Dieses Verfahren, welches speziell für die Verwendung von kostengünstigen Grid-Systemen (Groß oder Klein) entwickelt wurde, erlaubt die Verarbeitung nahe dem Quellsystem und maximiert die Verwendung von Rechenleistung.

### Eine aktive Anwender-Community

**Christian Roszak**  
CEO, Theis Consult

Die Anwender von

Talend Open Studio profitieren einer großen aktiven Community und haben Zugang zu folgenden TalendForge Online Community Tools:

- Forum: <http://www.talendforge.org/forum/>
- Wiki: <http://www.talendforge.org/wiki/>
- Bugtracker: <http://www.talendforge.org/bugs/>
- Exchange: <http://talendforge.org/exchange/>
- Babili: <http://www.talendforge.org/babili/>

Für weitere Details: [www.talend.com](http://www.talend.com) | [info@talend.com](mailto:info@talend.com)