

Data Integration Features Comparison Matrix

Talend Open Studio

vs

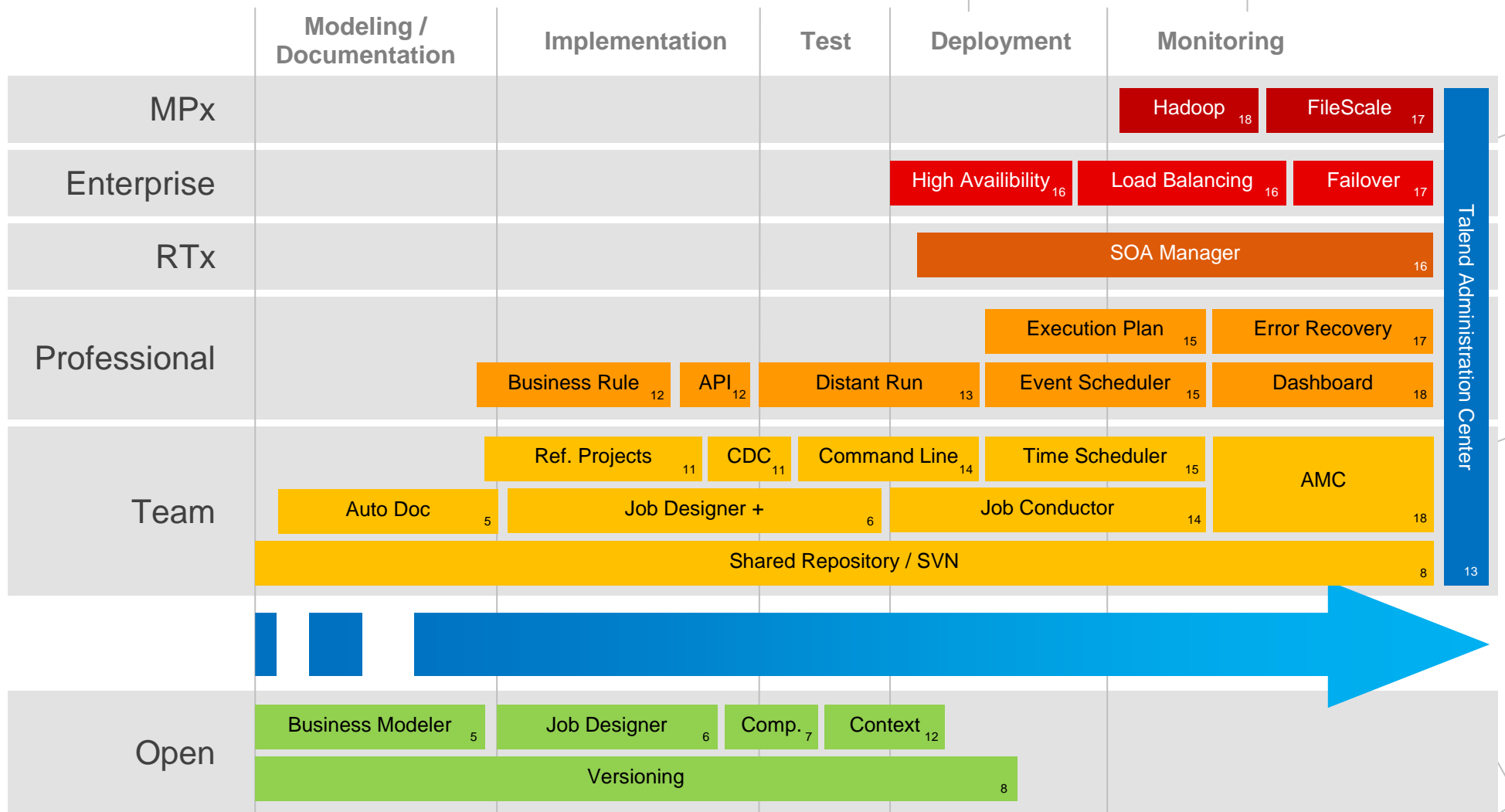
Talend Integration Suite

Team, Professional, RTx, Enterprise & MPx editions

<http://de.talend.com/products-data-integration/matrix.php>

Q1 2011

Feature details



Lizenzart und Haftungsausschluss

Open Source-GPL-Lizenz

*dieses Feature
ist enthalten in:*

Open

Team

Professional

RTx

Enterprise

MPx

Die GNU General Public License regelt die rechtlichen Bedingungen für die Verteilung kostenloser Software im Rahmen des GNU-Projekts.

Mit der GNU GPL-Lizenz werden dem Benutzer der Software die folgenden Rechte erteilt:

- das Recht zur Ausführung der Software zu jedem beliebigem Zweck und ohne jede Einschränkung
- das Recht, die Funktionsweise der Software zu untersuchen und die Software den eigenen Bedürfnissen gemäß anzupassen

Nimmt ein Entwickler Änderungen an einer Open Source-Software vor und möchte die angepasste Software weiterverteilen, muss er dies unter der GPL-Lizenz tun.

Den kompletten Text der GPL-Lizenz können Sie hier einsehen: <http://www.opensource.org/licenses/gpl-2.0.php>

Zugriff auf Quellcode

*dieses Feature
ist enthalten in:*

Open

Team

Professional

RTx

Enterprise

MPx

Der Quellcode zu Talend Open Studio steht auf der folgenden Webseite zur Verfügung: <http://www.talendforge.org/trac/tos/>

Um größtmögliche Transparenz und Einheitlichkeit zu gewährleisten, stellt Talend Kunden, die den Quellcode anfordern, auch sämtliche in der kommerziellen Edition enthaltenen Tools zur Verfügung.

Subskriptionslizenz

*dieses Feature
ist enthalten in:*

Open

Team

Professional

RTx

Enterprise

MPx

Die "Enterprise"-Versionen der Talend-Software enthalten zusätzlich zum Leistungsumfang der Open Source-Produkte weitere Mehrwert-Funktionen (siehe unten) und – Services und werden unter einer kommerziellen Lizenz vertrieben.

Talends Preismodell garantiert Transparenz und Planungssicherheit: Der Preis richtet sich nicht nach dem Datenvolumen oder einem möglichen zusätzlichen Bedarf an Konnektoren oder CPUs, sondern orientiert sich an der Anzahl der Entwickler (Studio), dem Funktionsumfang (ausgewählte Edition) und der Laufzeit.

Das Subskriptions-Modul garantiert dem Kunden einen größtmöglichen ROI: Die Anzahl der Lizenzen kann jährlich nach oben oder unten korrigiert werden, so dass die Kosten an den Entwicklungsstand eines Projekts und die Mitarbeiterzahl angepasst werden können.

Die Lösungen von Talend sind 50 bis 80 % billiger als vergleichbare proprietäre Lösungen und lassen sich mit deutlich geringerem Kostenaufwand implementieren, pflegen und betreuen.

Haftungsausschluss

*dieses Feature
ist enthalten in:*

Open

Team

Professional

RTx

Enterprise

MPx

Open Source-Softwareprodukte sind das Ergebnis gemeinschaftlicher Entwicklungsanstrengungen, und der Code geht auf Beiträge aus verschiedensten Quellen zurück. Wird die Integration der einzelnen Code-Beiträge nicht sorgfältig kontrolliert und verwaltet, können durch die Verwendung der Software die Rechte einzelner Beitragender verletzt werden.

In einem solchen Fall hat der Endanwender möglicherweise mit einer strafrechtlichen Verfolgung wegen Verletzung geistiger Eigentumsrechte zu rechnen, selbst wenn kein vorsätzliches Handeln vorlag.

Talend bietet seinen Subskriptions-Kunden eine Haftungsfreistellungsklausel. Damit garantiert Talend dem Anwender, dass er vor strafrechtlicher Verfolgung und finanziellen Forderungen geschützt ist, falls der Code von Talend die Eigentumsrechte Dritter verletzen sollte.

Support und Dokumentation

Support durch Community: Foren, Bugtracker...

*dieses Feature
ist enthalten in:*

Open

Team

Professional

RTx

Enterprise

MPx

Die Talend Anwender-Community umfasst zehntausende Fachexperten, die sich aktiv an der Weiterentwicklung der Software beteiligen. Zu den wichtigsten Beiträgen dieser Community gehören:

- Testen und Qualitätssicherung neuer Versionen
- Ausarbeitung von Feature Requests
- Übersetzung und Lokalisierung von Produkten
- Support und Austausch über Foren
- Entwicklung und Weitergabe neuer Komponenten, Konnektoren, Jobs, Modelle und anderer Plug-Ins.

In der Talend Exchange können Community-Mitglieder eigene Plug-Ins veröffentlichen und sie so anderen Anwendern zur Verfügung stellen. Die meisten Beiträge werden – im Anschluss an detaillierte Tests und Verbesserungen durch unser internes F&E-Team – in das Produkt integriert.

Talend leistet seinerseits Beiträge zu zahlreichen großen Open Source-Projekten und ist Mitglied der Eclipse Foundation und der Apache Software Foundation.

Enterprise-Support mit SLAs

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Mit einer Talend Support Services-Subskription profitieren Sie von den Erfahrungen unserer internen Fachleute, die sich täglich mit Talends F&E-Team austauschen. Dieses Angebot wurde geschaffen, um Subskriptions-Kunden eine effektive Betreuung mit einem Höchstmaß an Sicherheit zu bieten. Die Support Services stehen in den drei Kategorien Silver, Gold und Platinum zur Verfügung. Zu jedem dieser drei Levels gibt es feste Zusagen, innerhalb welcher Zeit beispielsweise eine Fehlermeldung beantwortet oder ein Patch bereitgestellt wird.

<http://de.talend.com/professional-support/support.php>

Benutzer-, Referenzhandbuch

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Die Dokumentation zu Talend Open Studio steht auf Englisch und Französisch zum kostenlosen Download zur Verfügung. Die beiden Handbücher im PDF-Format, der User Guide (276 Seiten) und der Components Reference Guide, sind auf der folgenden Webseite hinterlegt: <http://www.talend.com/resources/documentation.php>

Eine Druckversion dieser Handbücher können Sie bei Amazon erwerben:

http://www.amazon.com/Talend-Open-Studio-Guide-Version/dp/1453632077/ref=sr_1_1?ie=UTF8&s=books&qid=1296040035&sr=1-1

Nutzen Sie außerdem unsere Tutorials mit wertvollen Informationen zu unseren Lösungen!

Die Tutorials finden Sie unter: <http://www.talendforge.org/tutorials/menu.php>

Modellierung/Dokumentation

Business Modeler

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Der Business Modeler ist ein Tool, das keine technischen Vorkenntnisse erfordert (wie z. B. Microsoft Visio). Es hilft Ihnen dabei, die relevante Dokumentation und die technischen Elemente zu strukturieren. Gleichzeitig unterstützt es den Datenintegrationsprozess anhand eines anwenderfreundlichen Diagramms und ermöglicht es unterschiedlichen Teams (Design, Entwicklung, Test, Produktion etc.) ein gemeinsames Tool für die Arbeit an einem gemeinsamen Modell zu verwenden.

Business-Anwender arbeiten z. B. mit Geschäftsmodellen, um ihre Datenintegrationsanforderungen darzustellen. Auf diese Weise können IT-Entwickler und -Fachkräfte die Geschäftsanforderungen besser verstehen und in technische Prozesse (Jobs) umsetzen. Sobald jede einzelne technische Implementierungsphase (Jobs) abgeschlossen ist, lässt sich das Geschäftsmodell einfach aktualisieren, um den Entwicklungsfortschritt anzuzeigen und alle Beteiligten auf dem neuesten Stand zu halten.

Datenbankadministratoren (DBAs) können anhand von Geschäftsmodellen gemeinsam auf die erforderlichen Metadaten für die DB-Anbindung zugreifen. Der Systemarchitekt erhält so einen Überblick über die Datenintegrationsanforderungen.

Die Entwicklung von Geschäftsmodellen gehört zu den Best Practices, die ein Unternehmen bereits in einer frühen Phase seines Datenmanagement- oder -integrationsprojekts umsetzen sollte, um den Erfolg seines Projekts zu gewährleisten. Da sich Engpässe und Schwachstellen durch den Einsatz von Geschäftsmodellen normalerweise schnell erkennen und beheben lassen, können diese dazu beitragen, Budgetüberschreitungen zu minimieren und/oder Vorabinvestitionen zu reduzieren.

Auto Doc

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Mithilfe dieser Funktion kann auf Wunsch eine detaillierte technische Dokumentation zu all Ihren Jobs erstellt werden. Hierbei werden Metadaten zum Job (Autor, Version, Status, Aktualisierungsdatum usw.) sowie eine grafische Sicht zum jeweiligen Job mit allen Parametern der beteiligten Komponenten in einem interaktiven anwenderfreundlichen Format (HTML/XML) präsentiert werden.

Die Dokumentation lässt sich ganz einfach mit personalisierten Kommentaren ergänzen.

Auto Doc+

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Mit AutoDoc+ wird die technische Dokumentation (siehe oben) automatisch für jede einzelne Version eines Jobs erstellt. Wenn Sie einen Job speichern, wird die zugehörige Dokumentation aktualisiert und im Repository abgelegt. Damit wird sie automatisch zum Zugriff für alle Anwender freigegeben.

AutoDoc+ erlaubt Ihnen auch, die grafische Darstellung der Dokumentation anzupassen und beispielsweise das eigene Logo und einen Firmennamen einzufügen oder mittels personalisierter CSS die Farben zu ändern.

Implementierung

Job Designer

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Mit seiner grafischen Palette an Komponenten und Konnektoren bietet der Job Designer sowohl eine grafische als auch eine funktionale Sicht der eigentlichen Integrationsprozesse.

Um die Integrationsprozesse zu erstellen, werden die Komponenten und Konnektoren einfach per Drag & Drop in einen grafischen Arbeitsbereich gezogen, Verbindungen und Beziehungen untereinander festgelegt und entsprechende Einstellungen vorgenommen.

Mit dem Job Designer erhalten Sie über eine umfangreiche Komponentenbibliothek Zugriff auf alle Arten von Quellen und Zielen, die für Datenintegrations-, Datenmigrations- oder Synchronisierungsprozesse benötigt werden. Es gibt Komponenten und Konnektoren für alle Arten von Tasks und Operationen zu den Daten selbst, sowie zum Datenmanagement und zur Ablaufsteuerung des Datenflusses. Konnektoren unterstützen den Zugriff und das Lesen/Schreiben aller Quell- und

Zielsysteme für die Datenintegration, Datenmigration und Datensynchronisierung. Die Parameter werden zentral in einer Sicht konfiguriert, wenn die einzelnen am Job beteiligten Komponenten ausgewählt werden. Sie können auch vom Metadaten-Manager (Repository) vererbt werden.

Komplexe Komponenten sind mit einer speziellen intuitiven grafischen Benutzeroberfläche ausgestattet oder verfügen über integrierte Assistenten, die bei der Erstellung von Jobs helfen.

Damit das Jobdesign übersichtlich bleibt, kann das Jobdiagramm in Unterjobs gegliedert und anschließend in Child- und Parent-Jobs aufgeteilt werden, um den Ausführungsablauf festzulegen. Orchestrierungskomponenten sowie verschiedene Beziehungsarten helfen dem Benutzer dabei, die Reihenfolge der Prozessausführung festzulegen. Über eine integrierte Konsolensicht können Anwender schnell die Ausführung überwachen sowie direkt im Studio die Performance prüfen und nachverfolgen.

Komponenten

dieses Feature ist enthalten in:

Open

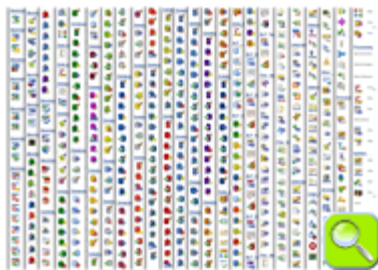
Team

Professional

RTx

Enterprise

MPx



Talend bietet native Open Source-Konnektoren für den technischen und geschäftlichen Bereich, um alle IT-Umgebungen einzubinden. Die große Anzahl an Konnektoren gewährleistet die Interoperabilität zwischen Anwendungen und Datenbanken; sie erlaubt die Überbrückung heterogener Datenstrukturen bei unübertroffener Performance. Durch die ständig wachsende Zahl an Konnektoren werden außerdem die Funktionen der Open Source-Datenintegrations-, Datenmanagement- und Datensynchronisierungslösungen von Talend erweitert. Über 550 Komponenten stehen kostenlos zur Verfügung. 60 % davon wurden von der Talend-Community konzipiert und entwickelt.

Extern entwickelte Konnektoren und Komponenten können über die Talend Exchange-Plattform (<http://talendforge.org/exchange/>) gemeinsam genutzt werden. Eine gewisse Anzahl der zur Verfügung gestellten Komponenten werden von Talend validiert und optimiert, bevor sie nativ integriert und unterstützt werden.

Eine umfassende Liste der unterstützten Konnektoren erhalten Sie unter <http://www.talendforge.org/components>.

ETL-Unterstützung

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

ETL (Extraktion, Transformation und Laden) ist der Standardmodus der Talend Datenintegrationslösungen. Dabei werden Datenzeilen einzeln nacheinander gemäß einem Flussmodell verarbeitet. Der Modus wurde speziell für die Verwendung in heterogenen Umgebungen angepasst und erlaubt die Integration verschiedenster Technologien in den Quell- und Zielsystemen (Web Service, Dateien, Datenbanken, MOM, Geschäftsanwendungen usw.).

Der ETL-Modus steht sowohl bei der Batch-Verarbeitung als auch bei der Echtzeit-Verarbeitung zur Verfügung. Um die Ausführung zusätzlich zu beschleunigen, können auch mehrere ETL-Prozesse parallel ausgeführt werden.

Die einzigartige Architektur von Talend ist nicht auf eine bestimmte Ausführungsengine beschränkt, da bei Talend eigenständige Prozesse generiert werden, die auf jedem beliebigen Server (innerhalb oder außerhalb des Unternehmens) ausgeführt werden können. Darüber hinaus werden ETL-Prozesse so nah wie möglich an den Datenquellen ausgeführt, wodurch die Zugriffszeit und der Bandbreitenverbrauch reduziert und mögliche Engpässe beseitigt werden.

Innerhalb eines Jobs kann dieser Ansatz auch mit dem ELT-Verfahren (siehe unten) kombiniert werden, um eine bestmögliche Leistung ohne Einschränkung durch die Architektur zu erzielen.

ELT-Unterstützung

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Die Datenintegrationslösungen von Talend unterstützen zusätzlich den ELT-Modus (Extraktion, Laden und Transformation), bei dem Daten in einer Mengenoperation (mithilfe der Operatoren Union, Except und Intersect) direkt im DBMS der Zieldatenbank verarbeitet werden.

Dieser Modus ist ausschließlich für die Verwendung in homogenen Umgebungen (eine einzige Datenbank) vorgesehen. Der Vorteil liegt hier darin, dass die vorhandenen Hardware-Ressourcen voll ausgenutzt werden können, weswegen sich das Verfahren besonders für die Verarbeitung großer Datenvolumen in "Data Warehouse Appliance"-Umgebungen wie beispielsweise Teradata oder Netezza anbietet.

Innerhalb eines Jobs kann dieser Ansatz auch mit dem ETL-Verfahren (siehe oben) kombiniert werden, um eine bestmögliche Leistung ohne Einschränkung durch die Architektur zu erzielen.

Versionierung

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Die Versionierung von Objekten lässt sich im Talend Studio einfach über die native manuelle Versionierungsfunktion verwalten.

Bei der Erstellung eines Jobs wird automatisch eine Haupt- und eine Nebenversionsnummer generiert, die sich mit der Zeit bzw. wenn Updates durchgeführt werden, einfach erhöhen lässt. Dazu gibt es direkt in der Designer-Ansicht vom Talend Studio ein eigenes Fenster für die Versionskontrolle.

Alle Objekte, die in Talend Studio erstellt wurden, können auf diese Weise versioniert werden: Geschäftsmodelle, Jobs, Routinen, Metadaten, Dokumentationsmaterialien etc.

Die Versionierung ist ein genereller Bestandteil unseres Best Practice-Ansatzes und soll die Wiederverwendung von Objekten und bei Bedarf die Rückkehr zu einer vorherigen Entwicklungsstufe erleichtern.

Shared Repository

dieses Feature
ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Das **Shared Repository** (Metadaten-Manager) wurde konzipiert, um alle Projektinformationen und Metadaten eines Unternehmens in einem zentralen Repository zu konsolidieren, auf das alle, die am Integrationsprozess beteiligt sind, gemeinsam zugreifen können.

In der Studio-Ansicht können alle Anwender entsprechend ihrer Funktion und ihren im Talend Administration Center definierten Berechtigungen auf die Projekte zugreifen.

Dieses gemeinsame Repository ermöglicht Teamarbeit und Kollaboration zwischen allen Mitarbeitern, die an einem Integrationsprojekt beteiligt sind. Es hilft Ihnen dabei, alle Talend-Objekte zu speichern und gemeinsam zu nutzen: Geschäftsmodelle, Jobs (Prozesse), Joblets, Routinen, Metadaten Definitionen (z. B. Anbindungen an Quell-/Zielsysteme) etc.

Hinter dem gemeinsamen Repository steht ein Industriestandard-Source-Manager (Subversion), mit dem sich sämtliche Versionen aller Objekte speichern und verwalten lassen.

Ein automatisches Sperrsystem garantiert, dass Jobs beim Erstellen effektiv geblockt werden, so dass kein anderer Benutzer gleichzeitig Änderungen am gleichen Job vornehmen kann.

Ab Version 4.0 bieten wir die volle Versionierungsleistung von Subversion, so dass Sie mit verschiedenen Verzweigungen, Check-Ins/Check-Outs, sowie manuellen oder automatischen Commits und Kommentaren arbeiten können.

Data Viewer

dieses Feature
ist enthalten in:

Open

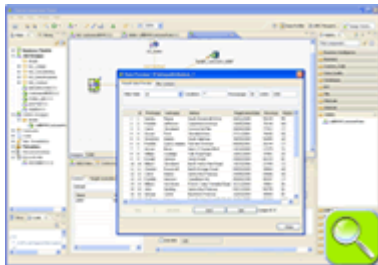
Team

Professional

RTx

Enterprise

MPx



Bei der Entwicklung von Jobs mit Talend müssen Sie möglicherweise den Inhalt verschiedener Quell- oder Zielsysteme (Dateien, Datenbanken, etc.) einsehen. Der Data Viewer erlaubt Ihnen einen Drill-Down in die Datenquell-/zielsysteme, unabhängig von der Anwendung, die normalerweise zum Öffnen benötigt wird: Notepad für .txt- und .csv-Dateien, ein SQL-Query-Browser für Datenbanktabellen, MS Excel für .xls-Dateien, ein HTML-Browser, etc.

Sie benötigen keine unterschiedlichen Tools und sie müssen auch keine Systeme durchsuchen, um per Drill-Down zu den Daten zu gelangen. Der Data Viewer verwendet die definierten Quell-/Zielpfadeinstellungen, um direkt zu den gewünschten Daten zu gelangen.

Der Data Viewer kann Ihnen viel Zeit sparen, da Sie direkt im Studio auf das Tool zugreifen können. Klicken Sie dazu einfach mit der rechten Maustaste auf eine beliebige Komponente. Sie können ganz bequem und unabhängig vom Format (Excel, Datenbanktabelle, CSV, etc.) die Daten in Ihren Quell-/Zielsystemen einsehen, während Sie Ihre Integrationsprozesse entwickeln.

Assistenten

dieses Feature
ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Dynamisches Schema

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Mithilfe dynamischer Schemata lassen sich Jobs mit unbekannter Spaltenstruktur und –anzahl erstellen. Je nach Auswahl des Entwicklers können dynamische Spalten unter Verwendung eines Pass-Through-Modus direkt dem Ziel zugeordnet werden.

Bevorzugte Anwendungsbereiche für diese Funktion sind Replikationsszenarien oder einfache Eins-zu-Eins-Zuordnungen mehrerer Spalten. Die Erstellung entsprechender Jobs gestaltet sich mit dieser Funktion denkbar einfach. Muss ein Entwickler beispielsweise eine komplette Datenbank mit hunderten Tabellen migrieren, kann er dies im Rahmen eines einzigen Jobs tun, ohne die Tabellenstrukturen im Detail kennen zu müssen.

Impact Analysis (Wirkungsanalyse)

dieses Feature ist enthalten in:

Open

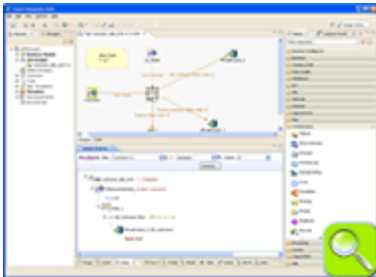
Team

Professional

RTx

Enterprise

MPx



Mit der Funktion Impact Analysis können Sie herausfinden, welche Konsequenzen eine Veränderung haben wird. Die Funktion steht Ihnen im Metadaten-Manager zur Verfügung. Sie können eine Wirkungsanalyse mit jeder beliebigen Spalte mit beliebigen Metadaten (Datenbank, Datei, etc.) durchführen. Das Ergebnis der Wirkungsanalyse wird Ihnen in einem grafischen und interaktiven Bericht angezeigt, in dem Sie die Spalte zurückverfolgen und alle angewendeten Operationen von der Quelle bis zum Ziel, während des gesamten Jobs, einsehen können.

Der Bericht kann als HTML-Datei exportiert werden.

Data Lineage (Datenherkunft)

dieses Feature ist enthalten in:

Open

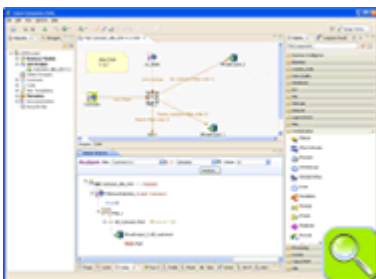
Team

Professional

RTx

Enterprise

MPx



Mit der Funktion Data Lineage können Sie sehen, wo eine Änderung vorgenommen wurde.

Die Funktion steht Ihnen im Metadaten-Manager zur Verfügung und kann auf jede beliebige Spalte mit beliebigen Metadaten (Datenbank, Datei, etc.) angewendet werden. Das Ergebnis wird in einem Bericht angezeigt, in dem Änderungen von der Zielendkomponente eines Jobs bis zum Quellende nachverfolgt werden.

Der Bericht kann als HTML-Datei exportiert werden.

Job Compare (Jobvergleich)

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Mit der Funktion Job Compare können Sie Unterschiede zwischen zwei Jobversionen oder unterschiedlichen Jobs ermitteln.

Job Compare ist vollständig in das Talend Integration Suite Studio integriert. Das Ergebnis eines Jobvergleichs wird in einem grafischen und interaktiven Bericht in HTML- oder XML-Format angezeigt. Die Unterschiede werden hervorgehoben dargestellt.

Der Vergleichsbericht in diesem Beispiel zeigt, dass das Trennzeichen-Feld der tFileInputDelimited-Komponenteneigenschaften in den beiden verglichenen Jobs nicht gleich definiert ist: In Version 3.2 ist das Trennzeichen „\t“, in Version 4.2 dagegen „\n“.

Joblets

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Mit Joblets können Sie einen Jobteil (Subjob) in eine Joblet-Komponente zerlegen. Wählen Sie dazu die Komponenten aus, die den Jobteil bilden, den Sie wiederverwenden möchten oder den Sie zerlegen möchten, indem Sie auf den Menüpunkt „Refactor to Joblet“ (in Joblet zerlegen) klicken.

Das Jobdesign wird automatisch einfacher, wenn die ausgewählten Komponenten in einer einzigen Joblet-Komponenten zusammengefasst werden. Diese Joblet-Komponente kann über spezielle Joblet-Ordner in der Komponentenpalette gemeinsam genutzt werden und lässt sich damit einfach für beliebige andere Jobs verwenden.

Joblets erleichtern die Pflege redundanter und komplexer Jobs ganz erheblich.

Darüber hinaus können Sie mit einem „Impact Analysis“-Mechanismus herausfinden, welche Jobs ein definiertes Joblet verwenden.

Referenzprojekte

dieses Feature ist enthalten in:

Open

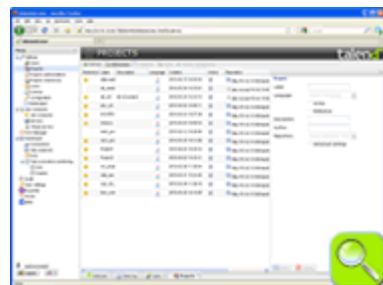
Team

Professional

RTx

Enterprise

MPx



Referenzprojekte helfen Ihnen dabei, die Duplizierung (durch Copy & Paste) von Objekten (Jobs, Routinen, Dokumentationsmaterialien, Metadaten, etc.) zwischen Projekten zu vermeiden.

„Slave“-Projekte werden mit einem (oder mehreren) „Master“-Projekt(en) durch eine Referenz verknüpft und bekommen auf diese Weise Objekte von dem/den Parent-Objekt(en) vererbt.

Die Ressourcen vom Master-Projekt erscheinen im Slave-Projekt in schreibgeschütztem Modus: Sie können wiederverwendet und ausgeführt, aber nicht verändert werden.

Da es eine starke Verbindung zwischen Slave- und Master-Projekten gibt, werden bei einer Änderung an einem Objekt im Master-Projekt alle Slave-Projekte automatisch aktualisiert.

Die Referenzprojekte teilen alle redundanten Projekt-Objekte (Jobs, Vorlagen, Metadaten), um sie anderen Projekten zur Verfügung zu stellen. Mit dieser Funktion können Sie die 30 % der Objekte nutzen und wiederverwenden, die normalerweise in allen Datenintegrationsprojekten vorkommen, so dass sich die Pflege drastisch reduzieren lässt.

Change Data Capture

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Beim Data Warehousing werden Daten von einer oder von mehreren Datenbanken in eines oder mehrere Zielsysteme zu Analysezwecken extrahiert oder transferiert. Dies bedeutet, dass bei der Extraktion und beim Transfer enorme Datenmengen aufkommen, was sowohl einen hohen Zeit- als auch Kostenaufwand verursachen kann.

Die Möglichkeit, nur veränderte Daten in Echtzeit zu erfassen, ist unter dem Begriff Change Data Capture (CDC) bekannt. Die Erfassung von Änderungen verringert das Datenaufkommen zwischen Systemen und trägt dazu bei, ETL-Zeit zu sparen.

Die Talend CDC-Architektur basiert auf einem Publisher-/Subscriber-Modell. Der Publisher erfasst die Datenveränderungen und stellt sie dem Subscriber (Talend-Jobs) zur Verfügung. Subscriber verwenden dabei die Datenänderungen vom Publisher.

Diese Funktion erkennt veränderte Einträge in Echtzeit, so dass die geänderten Daten unmittelbar an Subscriber-Jobs übermittelt werden können. Auf diese Weise kann Zeit gespart werden, die für das Laden und Aktualisieren von Daten während ETL- oder operativen Datenintegrationsprozessen benötigt wird.

Talends Change Data Capture-Funktion bietet die am häufigsten verwendeten Modi: Trigger und Redo Logs. Welcher Modus verfügbar ist, hängt von der eingesetzten Datenbankart ab.

Geschäftsregeln

dieses Feature ist enthalten in:

Open

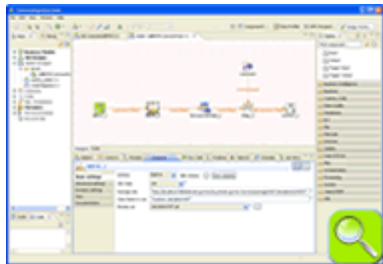
Team

Professional

RTx

Enterprise

MPx



Geschäftsregeln werden normalerweise von Business-Anwendern durch Spezifikationen definiert, die anschließend von den technischen Mitarbeitern interpretiert und implementiert werden.

In der Talend Integration Suite ist eine Engine für Geschäftsregeln integriert, die den Benutzern dabei hilft, eigene Geschäftsregeln zu konfigurieren. Auf diese Weise können Anwender Marktsegmentkriterien definieren (nach Alter, Region etc.) und ihre Geschäftsregeln mit einer Excel-Tabelle oder einer Drools Guvnor-Oberfläche direkt im webbasierten Talend Administration Center festlegen.

Die Drools Guvnor-Oberfläche ermöglicht es Business-Anwendern Regeln mit einem grafischen Editor schnell und direkt zu erstellen und zu editieren. Sie können den Zugriff auf die Regeln und andere Funktionen kontrollieren und Regelversionen und -änderungen verwalten. Die Regeln können von den entwickelten Jobs getestet und aufgerufen werden (siehe nächste Folie).

API

dieses Feature ist enthalten in:

Open

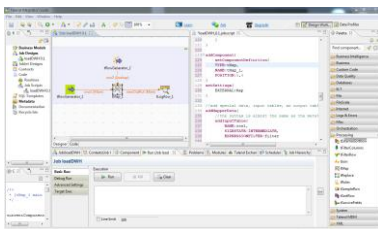
Team

Professional

RTx

Enterprise

MPx



Sophisticated but simple to grasp scripting language which enables programmatic job generation

Fully featured dedicated editor including syntax highlight, code completion aid & warning system

Automatic reverse engineering from an existing Job design to produce the corresponding script

Test

Context Management (Kontextverwaltung)

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Mithilfe von Kontexten lassen sich so gut wie alle Parameter von Komponenten/-jobs externalisieren. Dies hilft beispielsweise Anwendern, die Parameter „on-the-fly“ während der Laufzeit definieren möchten, oder Anwendern, die unterschiedliche Einstellungen für die Tests/Produktion verwenden möchten.

Kontexte können nach Bedarf für alle Umgebungen (Entwicklung, Test, Produktion, etc.) definiert werden, ohne Einschränkung bei der Anzahl der erstellten Kontexte.

Anwender können jederzeit – in der Entwicklungsphase oder zur Laufzeit - Kontexte wechseln, um die definierte Einstellung zu verwenden.

Die Parameterwerte können auch über eine Dialogbox in der Entwicklungs- und Testphase verändert werden. Darüber hinaus können mit einer speziellen Komponente, die dafür ausgelegt ist Parameter zu laden, sämtliche Werte dynamisch überschrieben

werden.

Distant Run (Fernausführung)

dieses Feature ist enthalten in:

Open

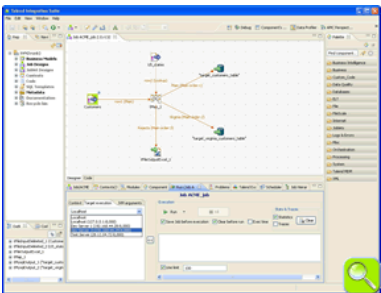
Team

Professional

RTx

Enterprise

MPx



Mit der Funktion **Distant Run** lassen sich Jobs auf beliebigen Servern direkt vom Studio remote ausführen.

Dies kann äußerst nützlich sein, wenn Sie Jobs testen müssen, z. B.:

- in einer der Produktionsumgebung ähnlichen Konfiguration,
- auf verschiedenen Betriebssystemen,
- nach Anfrage auf spezifischen Systemen,

da keine komplexen Implementierungsprozesse durchlaufen werden müssen.

Das Zielsystem kann dynamisch zur Laufzeit direkt vom Studio ausgewählt werden. Alle regulären Fehlerbehebungs-,

Nachverfolgungs- und Echtzeitstatistik-Optionen sind auch in diesem Remote-Modus verfügbar.

Implementierung

Talend Administration Center

dieses Feature ist enthalten in:

Open

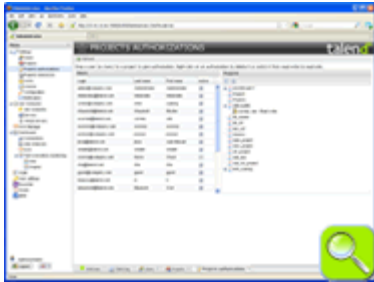
Team

Professional

RTx

Enterprise

MPx



In allen Subskriptionsangeboten ist ein Studio (oder mehr je nach Anwenderzahl) und eine Software-Komponente enthalten, die auf einem Server installiert und über eine Weboberfläche, dem Talend Administration Center, verwaltet werden kann.

Die Studios sind also nicht mehr länger im lokalen Modus, sondern remote an die im Talend Administration Center definierten Projekte angebunden.

Talend Administration Center ist eine schlanke Anwendung (Browser-basiert, keine Implementierung erforderlich), die den Projektverantwortlichen eines Integrationsprojekts dabei unterstützt, Anwender, Projekte, Benutzerrechte, Lizenzen, etc. zu verwalten.

Die Projektautorisierung lässt sich mühelos pro Benutzer zuweisen (Unterstützung von LDAP-Verzeichnissen). Auf diese Weise erhalten Anwender entsprechend ihrer Funktion Berechtigungen für den Projektzugriff, z.B.: Keine Berechtigung, nur Leserechte, Lese- und Schreibrechte, etc.

Lese- und Schreibrechte, etc.

Anwender können anschließend Repository-Objekte (Jobs, Geschäftsmodelle, Metadaten zur DB-Anbindung, etc.), für die Projekte für die sie autorisiert sind, gemeinsam mit anderen Anwendern direkt im Studio nutzen. Weitere Informationen zum gemeinsamen Repository erhalten Sie auf den folgenden Folien.

Je nachdem welche Talend Integration Suite Edition Sie abonniert haben, stehen zahlreiche zusätzliche Plug-Ins auf der linken Navigationsleiste (Dashboard, SOA-Manager, Server-Manager etc.) zur Verfügung.

Job Conductor

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Der Job Conductor koordiniert die Ausführung von Datenintegrationsjobs. Er bietet eine zentrale Ausführungsschnittstelle, von der alle Jobs entweder auf Anfrage oder Zeit-basiert (mit der Team Edition) bzw. Event-basiert (mit der Professional Edition) gestartet werden können.

Das Job Conductor-Modul stützt sich auf „JobServers“ oder Agents (kleine Anwendungen, die auf jedem Server installiert sind, auf dem Jobs ausgeführt werden).

Wenn Sie Ihre Agents eingerichtet haben, können Sie mit dem Job Conductor alle Hardware-Ressourcen (verfügbare CPU-, RAM-, HD-Kapazitäten, etc.) in Echtzeit überwachen, so dass Sie die Ausführung Ihrer Jobs im Netz verteilen können, je nachdem welcher der verfügbaren Server am besten geeignet ist. Dank nativer JMX-Unterstützung können Sie über 40 Indikatoren überwachen. Somit lässt sich jeder beliebige Job auf jedem beliebigen Server mit nur einem Klick implementieren!

Befehlszeilen

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Sie können den mit dem Job Designer-Tool entwickelten Integrationsprozess außerhalb der grafischen Benutzeroberfläche von Talend Studio mit dem Befehlszeilen-Modul implementieren und aktualisieren.

Das Talend Befehlszeilen-Modul bietet einen Satz an Befehlszeilen-Optionen, mit denen Entwickler und Administratoren ganz einfach Batch-Operationen durchführen können.

Nahezu alle Jobverwaltungsfunktionen in Talend Studio und im Talend Administration Center sind auch über die Befehlszeile abrufbar. Dazu zählen Funktionen wie die Aktualisierung von Job-Eigenschaften, der Übergang von Jobs in die Produktionsphase und der Export/Import einzelner oder mehrerer Jobs.

Mit der Befehlszeilenfunktion lassen sich zahlreiche und komplexe Job-Implementierungen und Ausführungen einfach und schnell realisieren, wobei Abhängigkeiten und Metadaten für die Ausführung berücksichtigt werden.

Die integrierte Befehlszeilenhilfe bietet eine umfangreiche Liste mit allen verfügbaren Befehlen sowie einer kurzen Funktionsbeschreibung.

Time Scheduler (Zeit-basierte Disposition)

dieses Feature ist enthalten in:

Open

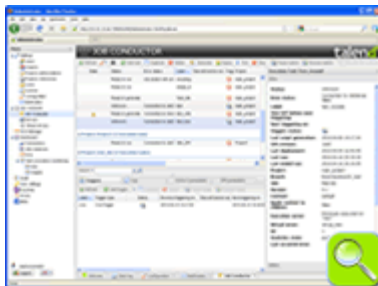
Team

Professional

RTx

Enterprise

MPx



Der **Time Scheduler** hilft Ihnen dabei, die Ausführung eines Jobs zu einer festgelegten Zeit und an einem bestimmten Datum (z. B. am ersten Montag eines Monats, jeden Dienstag, etc.) oder regelmäßig über einen gewissen Zeitraum hinweg zu realisieren. Mithilfe einer Task werden alle Informationen, die für die Jobausführung erforderlich sind (Projektname, Jobname, Jobversion, Server etc.) zentralisiert.

Die Task wird anschließend gemäß Zeitplan getriggert. Daraufhin wird der Job automatisch auf dem festgelegten Server zur festgelegten Zeit bereitgestellt und ausgeführt. Ein nützliches Statussystem hilft Ihnen bei der Überwachung des Triggering-Zustands und der Ausführung (Erfolg/Fehler) direkt im Job Conductor.

In der Professional Edition ist eine zusätzliche Funktion zur Event-/Datei-basierten Disposition verfügbar. (Siehe Folie zum Event Scheduler).

Event Scheduler (Event-basierte Disposition)

dieses Feature
ist enthalten in:

Open

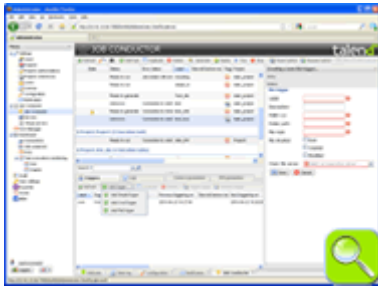
Team

Professional

RTx

Enterprise

MPx



Der **Event Scheduler** erweitert die Zeit-basierten Dispositionsfunktionen für die Echtzeit-Integration.

Mit dem Event Listener kann bei der Ausführung von Prozessen eine bedarfsorientierte oder Event-basierte Ausführung getriggert werden.

Events können Datei-abhängig sein, wie z. B. Erscheinen, Verschwinden oder Ändern einer Datei, oder SQL-basiert mit „Wait for“-Bedingung. Sobald das erwartete Event identifiziert ist, wird die Ausführungsaufgabe getriggert und die Job-Implementierung und der Job-Rollout werden durchgeführt.

Sie können ganz einfach zu einer beliebigen Aufgabe neue Event-Trigger hinzufügen und die Industrialisierung der automatischen Ausführungen erweitern.

Execution Plan (Ausführungsplan)

dieses Feature
ist enthalten in:

Open

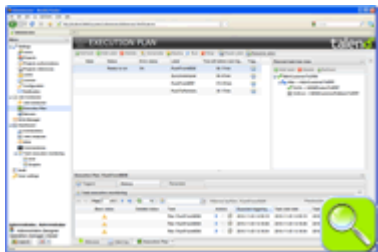
Team

Professional

RTx

Enterprise

MPx



Die Funktion **Execution Plan** hilft Ihnen bei der Ablaufsteuerung und Orchestrierung verschiedener Job-Ausführungen und erleichtert die Fehlerbehebung direkt im Job Conductor. Die Funktion ist Task-basiert. Sie stellt Abhängigkeiten zwischen unterschiedlichen Tasks dar und orchestriert die Ausführungsabfolge.

Die Abhängigkeiten der Tasks werden über eine hierarchische Sicht der Haupt- und Child-Task definiert, wobei jede Task eine untergeordnete Task besitzen kann.

Execution Plans können geplant und getriggert werden. Sie können durch die Umgebung definierte Ausführungsparameter aus dieser einzigen Job Conductor-Sicht verwenden.

SOA Manager / Webservices

dieses Feature
ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Der SOA-Manager bietet eine webbasierte und vollständig grafische Oberfläche zur Bereitstellung eines oder mehrerer Datenintegrationsjobs als Services (Web Services). Dies ermöglicht die automatische Implementierung in heterogenen Anwendungen und Systemen, die SOAP verwenden (RPC oder dokumentbasiert). Ein eigener Assistent hilft Ihnen dabei, WSDL-Deskriptoren zu generieren, um Jobs als Web Services zusammenzufassen und entsprechende UDDI-Einträge zu finden, um Web Services einzubinden.

Der SOA-Manager verfügt über eine erweiterte Verarbeitungsfunktion für eingehende Anfragen und stützt sich dabei auf ein optimiertes System, das Aufgaben zusammenfasst und der Reihe nach bearbeitet. Die benutzerdefinierte Auswahl aktiver Services verarbeitet eine gewisse Anzahl von Anfragen in Echtzeit, während der Queue-Manager zusätzliche Anfragen abwickelt, indem er die Daten zur asynchronen Verarbeitung in einem Puffer zwischenspeichert.

Mit dem SOA-Manager können Sie die in Talend Integration Suite Studio erstellten Jobs in einer Webservice-Umgebung bereitstellen, so dass die Interoperabilität zwischen unterschiedlichen Anwendungen und Systemen gewährleistet ist.

Lastverteilung

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Das (über den Job Conductor aufrufbare) Grid Conductor-Modul optimiert die Skalierbarkeit und die Verfügbarkeit von Integrationsprozessen, da es für eine optimale Auslastung des Ausführungsnetzes sorgt. Der Grid Conductor stützt sich auf die Definition virtueller Server, die verfügbare Ressourcen unabhängig von der Systemart (CPU, Betriebssystem, etc.) gruppieren.

Tasks werden an virtuelle Server des Grid Conductors statt an einen einzigen Ausführungsserver verteilt.

Dank einer konstanten Überwachung der verfügbaren Ressourcen auf den Ausführungsservern garantiert das Grid Conductor-Modul, dass alle Jobs reibungslos zur Trigger-Zeit ausgeführt werden und die verfügbaren Ressourcen optimal ausnutzen. Auf diese Weise werden Engpässe vermieden, die beim herkömmlichen Single-Server-Ansatz entstehen können.

Damit müssen Sie sich keine Gedanken mehr über die Rückgabe von Ressourcen (Resource Preemption) machen, wenn eine große Anzahl von Jobs gleichzeitig ausgeführt werden, oder wenn keine dedizierten Server eingesetzt werden. Der Grid Conductor bietet außerdem ein automatisches Failover falls Ressourcen ausfallen.

Hochverfügbarkeit

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Hochverfügbarkeit wird gewährleistet, da mehrere Job Conductors und Job-Ausführungsserver implementiert werden können.

Darüber hinaus gewährleisten Datenbank-Cluster ein Failover und verhindern Unterbrechungen bei der Ausführung.

Failover

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

FileScale

dieses Feature
ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx

Benchmark Results

Configuration

- 2 CPU Xeon 5520 Quad Core / 2,26GHz / 24 GB RAM
- Sun StorageTek 2540 Fiber Channel: 12 HD SAS 136Gb

Results

	1	10	100	300	550
# rows	6 million	60 million	600 million	1.8 billion	3.3 billion
data volume	740 MB	7.4 GB	74 GB	225 GB	415 GB
sorting time	6 sec	1 min 02 sec	21 min 30 sec	2 hr 01 min	5 hr 17 min

Independent benchmark made by Sun Microsystems

Die Talend Integration Suite MPx profitiert von Multi-Server-, Multi-CPU- und Multicore-Architekturen, bei denen sowohl Code als auch separate Unterprozesse parallel ausgeführt werden, um das Maximum aus der Architektur herauszuholen. Diese massiv-parallele Funktion maximiert die Leistungsfähigkeit der Unternehmensserver und die Anzahl verfügbarer Prozessoren, so dass die Verarbeitungszeit enorm verbessert wird.

Die Talend Integration Suite MPx bietet mit ihrer **FileScale**-Funktion eine einzigartige Technologie, mit der die Hardware des Ausführungsservers optimal ausgelastet und die Leistung von Low-Level-Sortieralgorithmen maximiert wird. Die FileScale-Technologie arbeitet bei (sehr) großen Dateien im Bulk-Modus und nutzt ihre effiziente Architektur voll aus, da sie nicht den Beschränkungen der JVM oder Ausführungsebene herkömmlicher Datenintegrationsarchitekturen unterliegt.

Mit der FileScale-Technologie lassen sich Daten anhand eines innovativen mathematischen Hochleistungsalgorithmus zur Datenverarbeitung sortieren und transformieren. Dabei kommt die MapReduce-Architektur zum Einsatz, die automatisch jede

Datenverarbeitungsoperation in einzelne kleinere Prozesse unterteilt und auf diese Weise eine hohe Leistung erzielt. Siehe auch Sun Microsystems Workbench: http://blogs.sun.com/aja/entry/talend_s_new_data_processing

Hadoop

dieses Feature
ist enthalten in:

Open

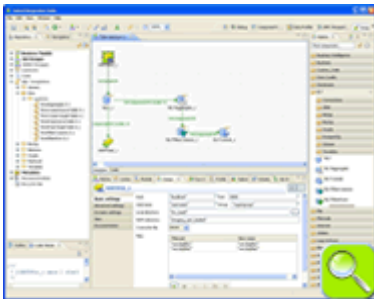
Team

Professional

RTx

Enterprise

MPx



Apache Hadoop ist ein Open Source Java Software Framework, das datenintensive verteilte Applikationen unterstützt. Es nutzt die Map Reduce-Architektur und erlaubt Anwendungen mit tausenden von Nodes und Petabytes von Daten und einem großen Netz preiswerter Server zu arbeiten. Die Talend Integration Suite MPx bietet eine native Unterstützung für Hadoop. Damit ist eine nahezu unbegrenzte Skalierbarkeit und Unterstützung beliebiger komplexer Datentypen möglich, so dass Unternehmen ihre Hadoop-Cluster für höchste Datenaufkommen und komplexe Transformationen nutzen können.

Ein spezieller Satz an Komponenten, der in der Komponentenpalette verfügbar ist, hilft beim Lesen und Schreiben von HDFS und Hive-Systemen und umfasst ELT- und SQL-Vorlagen.

Überwachung

AMC

dieses Feature ist enthalten in:

Open

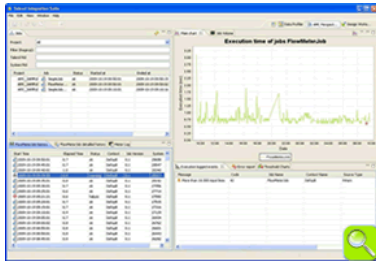
Team

Professional

RTx

Enterprise

MPx



Die **Talend Activity Monitoring Console** bietet eine komfortable grafische Oberfläche und ein zentrales Überwachungstool.

Sie ermöglicht eine detaillierte Überwachung, um die gesammelten Protokoll-Daten zu konsolidieren, die zugrunde liegenden Job-Interaktionen zu verstehen, unerwartete Fehler zu verhindern und Entscheidungen hinsichtlich der Systemverwaltung zu unterstützen.

Die Activity Monitoring Console überwacht Job-Events (Erfolge, Fehler, Warnungen, etc.), Ausführungszeiten und Datenvolumina über eine einzige Konsole von einer zentralen Stelle aus.

Sie ist als eigenständiges Tool verfügbar oder kann vollständig in Talend Studio integriert werden.

Dashboard

dieses Feature ist enthalten in:

Open

Team

Professional

RTx

Enterprise

MPx



Beim Dashboard handelt es sich um eine webbasierte Version der AMC-Konsole, das ganz einfach über einen Webbrowser aufgerufen werden kann.

Das Dashboard bietet Diagramme zur Ausführungsperformance und Statusindikatoren, mit denen alle Beteiligten sowohl den aktuellen als auch einen historischen Status der Integrationsprozesse einsehen können.

Es bietet außerdem detaillierte Überwachungsfunktionen, um gesammelte Protokoll-Daten zu konsolidieren, die zugrunde liegenden Komponenten- und Job-Interaktionen zu verstehen, Informationen zur Taskausführung schnell bereitzustellen, unerwartete Fehler zu verhindern und Entscheidungen hinsichtlich der Systemverwaltung zu unterstützen.

Error Recovery (Fehlerbehandlung)

*dieses Feature
ist enthalten in:*

Open

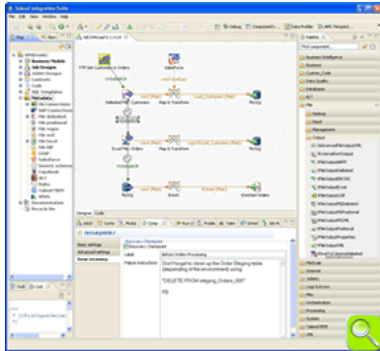
Team

Professional

RTx

Enterprise

MPx



Jobausführungsprozesse, wie auch Backup- und Wiederherstellungsoperationen können sehr langwierig sein. Talend Integration Suite Studio bietet Ihnen die Möglichkeit bei der Erstellung von Jobs Wiederherstellungs-Checkpoints einzurichten.

Mit dieser Funktion können die Prozesse von einem der Checkpoints wieder aufgenommen werden. Die Job-Entwickler können ein spezielles Fehlermanagement erstellen und integrieren und mithilfe der Checkpoint-„on-Failure“-Anweisung auf bestimmte Fehlerbedingungen reagieren.

Checkpoints können an bestimmten Punkten des Datenflusses (bei Trigger-Verbindungen) in passenden Abständen gesetzt werden. Damit können sie Zeit und Aufwand reduzieren, wenn ein Job aufgrund eines Fehlers erneut ausgeführt werden muss.

Mithilfe der Error Recovery-Checkpoint-Funktion kann der Prozess vom letzten erfolgreich passierten Checkpoint vor dem Abbruch (oder von jedem anderen Checkpoint vor dem Abbruch) wieder gestartet werden und muss nicht komplett neu ausgeführt werden.