

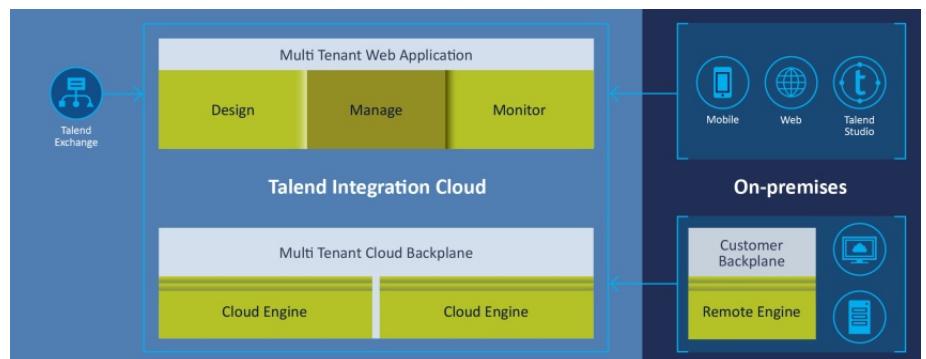
Cloud-basierte Datenintegrationslösung mit leistungsfähiger Entwicklungsumgebung

Dr. Götz Gütlich

Mit der Talend Integration Cloud stellt Talend eine sichere, Integrationsplattform auf Cloud-Basis zur Verfügung, mit der die Anwender in die Lage versetzt werden, ihre Anwendungen in der Cloud oder on-premise miteinander zu verbinden und Daten zwischen ihnen zu übertragen. Die Lösung bietet im Browser nutzbare grafische Entwicklungswerzeuge sowie vorgefertigte Integrationsaktionen, Flow Templates, Komponenten und Konnektoren. Diese machen die Integration der Daten einfach. Für anspruchsvollere Entwicklungsaufgaben steht zudem das "Talend Studio for Cloud" bereit, das lokal auf dem Rechner des Benutzers läuft und direkt mit der Cloud kommuniziert. Wir haben uns im Testlabor angesehen, wie die Arbeit mit der Cloud von Talend abläuft.

Die Talend Integration Cloud eignet sich nicht nur dazu, um Daten von einem System – zum Beispiel einer Datenbank – zu einem anderen – wie etwa einem Cloud-Dienst wie Salesforce – zu übertragen. Sie kann auch zum Einsatz kommen, um Informationen neu zu organisieren (beispielsweise durch das Aufteilen eines Namensfeldes in Vor- und Nachnamen) und anzureichern, also beispielsweise automatisch Adresseinträge anhand der Städte- und Straßennamen um Postleitzahlen zu ergänzen und ähnliches. Darüber hinaus lässt sich die Lösung nutzen, um Daten zu standardisieren und zu deduplizieren.

Das Produkt besteht im Wesentlichen aus zwei Komponenten: Zunächst der Cloud-Lösung selbst, die den Anwendern diverse Funktionen beziehungsweise Aktionen auf Browserbasis zur Definition ihrer Datenquellen, Da-



tensenken, Flows, Mappings und so weiter zur Verfügung stellt. Mit dem Web-Interface ist es auch möglich, die generierten Flows (also Datenübertragungen) zu testen und in den Produktivbetrieb zu überführen. Darüber hinaus besteht die Option, die durchgeführten Aktivitäten zu überwachen und die vorhandenen Aktionen, also die Bauteile, aus denen ein Flow zusammengesetzt wird, zu verwalten. Ein Scheduler, der dafür sorgt, dass bestimmten Jobs zu bestimmten Zeiten automatisch ablaufen, schließt den Leistungsumfang des Web-Interfaces ab.

Bei der zweiten Komponente handelt es sich um das Talend Studio for Cloud. Das ist eine Entwicklungsumgebung auf Basis von Eclipse, die die Anwender in die Lage versetzt, Komponenten für die Talend Integration Cloud zu entwickeln, die im Web-Interface noch nicht vorhanden sind.

Damit lassen sich zum Beispiel neue Datenquellen erschließen, für die noch kein Konnektor existiert, oder auch neue Funktionen implementieren. Die neu entwickelten Software-Lösungen können dann direkt in die Cloud

hochgeladen werden und anschließend dort zum Einsatz kommen. Mit "Talend Exchange" existiert zudem eine Plattform, über welche die Benutzer ihre Entwicklungen anderen Anwendern im Netz zur Verfügung stellen. Out-of-the-Box bietet das Studio bereits mehr als 900 Konnektoren und Komponenten, mit

Software, die auf einem Server im Unternehmensnetz installiert wird und dort die Aufgabe übernimmt, die Integration Flows abzuarbeiten. Das ergibt Sinn, wenn große Datenvolumen in die Cloud übertragen werden müssen oder wenn Daten, welche die Cloud-Lösung verarbeiten soll, das Haus nicht verlassen dürfen.

gaben benötigen – auch im Big Data-Umfeld. Dabei werden sämtliche Sicherheitsanforderungen auf Unternehmensebene erfüllt.

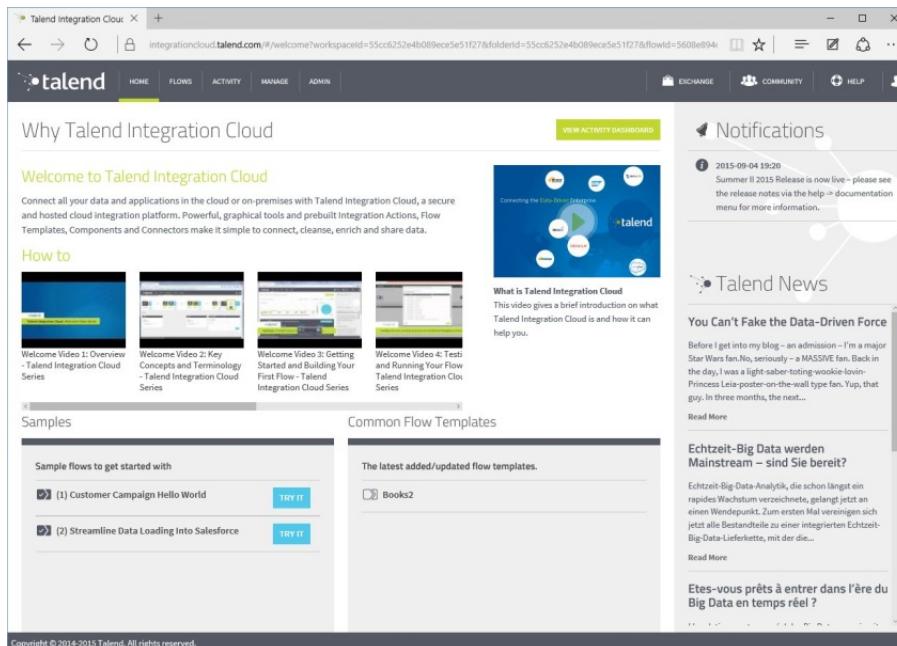
Der Test

Für den Test stellte uns Talend einen Cloud-Zugang vom Typ "Elastic Edition" zur Verfügung. Dieser wurde für den Enterprise-Einsatz im Unternehmen und in der Cloud konzipiert und verfügt über alle Features, die die Talend Integration Cloud bereitstellt. Für Umgebungen mit geringeren Anforderungen bietet Talend auch noch die Editionen "Hybrid" und "SaaS".

Diese wenden sich an Benutzer, die Cloud-to-Ground- und B2B-Integration betreiben wollen (Hybrid) oder die mit einer reinen Cloud-to-Cloud-Konnektivität (SaaS) auskommen. Diese Editionen verfügen folglich nicht über den vollen Leistungsumfang der Lösung, sind dafür aber günstiger.

Nachdem wir von Talend die Zugangsdaten zu unserem Cloud-Account erhalten hatten, machten wir uns zunächst einmal mit dem Web-Interface der Lösung vertraut und arbeiteten die Beispiele ab, die Talend für neue Benutzer zur Verfügung stellt. Danach ging es daran, Daten von einem System in ein anderes zu übertragen. Dabei behielten wir den Entwicklungs- und Arbeitsaufwand sowie die Bedienerführung im laufenden Betrieb im Auge.

Im nächsten Schritt installierten wir bei uns im LAN eine Remote Engine Probe und verwendeten diese, um unsere bis dahin erzeugten Flows abzuarbeiten. Zum Schluss setzten wir das Talend



Der Willkommensbildschirm der Talend Integration Cloud mit den einführenden Videos

denen die zuständigen Mitarbeiter ihre Datenintegrationsprojekte vereinfachen können.

Dabei werden unter anderem Datenübertragungen zwischen Marketo, NetSuite, Redshift, SAP und Salesforce unterstützt. Genauso arbeitet die Lösung mit relationalen Datenbanken wie Oracle, MySQL oder dem Microsoft SQL-Server zusammen. Das gleiche gilt auch für nicht SQL-basierte Datenbanken wie Cassandra und MongoDB sowie für Data-Warehouses wie AWS Redshift, Terradata und ExaSol.

Neben den genannten Entwicklungswerkzeugen der Integration Cloud bietet Talend auch noch eine "Remote Engine". Das ist eine

Mit der Talend Integration Cloud sind die Anwender dazu in der Lage, sowohl Integration Flows zu definieren, die von einem Cloud-Dienst in einen anderen laufen, als auch hybride Flows zu erzeugen, die als gesteuerte Integrations-Dienste zum Einsatz kommen. Durch das Verschieben der Workloads von den Servern on-premise in eine gesicherte Cloud-Umgebung mit Cloud-to-Cloud- und Cloud-to-Ground-Konnektivität lässt sich eine Vielzahl der Aufgaben nach Außen verlagern, was Ressourcen im Unternehmensnetz freimacht.

Talend stellt seinen Cloud-Kunden dabei jederzeit die Cloud-Ressourcen zur Verfügung, die diese für die Erfüllung ihrer Auf-

Studio for Cloud ein, um eigene Komponenten zu entwickeln, in die Cloud hochzuladen und in unseren Flows zu verwenden.

Der Beginn der Arbeit

Um mit der Talend Integration Cloud arbeiten zu können, müssen die Anwender über einen ak-

Was die benötigte Hardware angeht, so dürfte für die allermeisten Anforderungen ein Standard-Server vollkommen ausreichen.

Nach dem Login beim Webinterface der Talend Integration Cloud findet sich der Anwender zunächst einmal auf einem Will-

entwickelten Komponenten, schließen den Leistungsumfang der Willkommensseite ab.

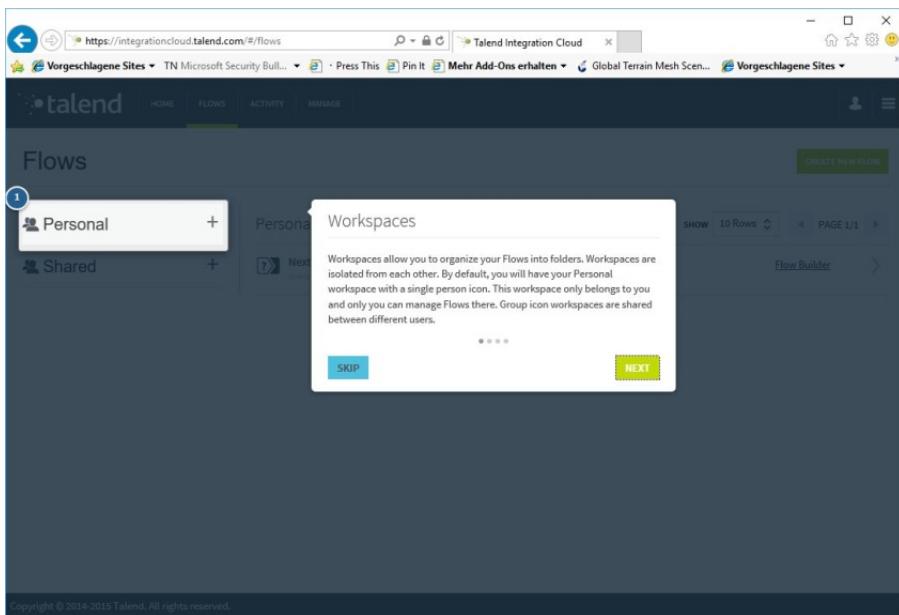
Ruft ein Benutzer eine Seite zum ersten Mal auf, so präsentiert die Cloud-Lösung ihm übrigens umfassende Hilfetexte, die ihn auf die richtige Art und Weise, mit dem Interface zu arbeiten, hinweisen. So werden auch Anfänger schnell in die Lage versetzt, die Umgebung effizient zu nutzen.

Besonders interessant ist die zweite Seite des Web-Interfaces. Sie nennt sich "Flows" und umfasst alle im System vorhandenen Data Flows. Die Talend Integration Cloud stellt den Anwendern immer zwei Workspaces zur Verfügung. Die unter "Personal" abgelegten Projekte gehören nur dem jeweiligen User, die unter "Shared" befindlichen Daten können auch von anderen Mitarbeitern des Unternehmens genutzt werden. Innerhalb der Workspaces lassen sich auch Ordner anlegen, um die Umgebung übersichtlicher zu gestalten.

Wechseln die Anwender auf einen Flow-Eintrag, so zeigt das System ihnen an, wie oft der Flow abgearbeitet wurde, ob er gerade aktiv ist und ob alles glatt lief oder ob es zu Rejections oder Failures kam. An gleicher Stelle lassen sich auch die "Run Details" einsehen, also die verwendete Datenquelle, die Zielumgebung und der Zeitplan für automatische Läufe. Darüber hinaus haben die Benutzer hier auch die Option, Läufe zu starten und zu stoppen.

Der Flow Builder

Um die Flows zu bearbeiten steht der so genannte Flow Builder be-



Umfassende Hilfetexte sorgen beim ersten Aufruf einer Seite dafür, dass die Anwender schnell in die Lage versetzt werden, effizient mit der Cloud-Lösung zu arbeiten

tuellen Browser verfügen. Offiziell unterstützt Talend den Internet Explorer 10 und 11, Firefox 38 oder neuer, Chrome 41 oder neuer und Apple Safari 8. Was mobile Geräte angeht, so empfiehlt Talend die Arbeit mit dem iPad 3 mit iOS 7 oder neuer, dem Samsung Galaxy Tab 3 10.1 mit Android 4.2.x oder dem Samsung Galaxy Note 10.1 mit Android 4.4.

Das Studio läuft auf Workstations unter Linux, MacOS und Windows. Die Software benötigt Java sowie mindestens drei GByte RAM (vier GByte werden empfohlen) und mehr als drei GByte Speicherplatz auf der Festplatte. Die Remote-Engine lässt sich ebenfalls unter Linux, MacOS und Windows benutzen.

kommensbildschirm wieder, der etliche Einführungsvideos mit Informationen über die Integrationsumgebung enthält. Diese englischsprachigen Videos geben einen Überblick über den Aufbau und das Konzept der Cloud-Lösung, zeigen exemplarisch, wie sich ein Flow aufbauen lässt, demonstrierten die Installation und Inbetriebnahme einer Remote Engine und geben einen Einblick in die Arbeit mit dem Talend Studio for Cloud.

Darüber hinaus finden sich auf dieser Seite auch Beispiele, die den Einstieg in die Arbeit mit dem Produkt erleichtern. Aktuelle News, eine Aktivitätsübersicht und ein Hinweis auf Talend Exchange, die Online-Börse zum Austausch der von den Usern

reit. Dieses Tool ermöglicht es den Anwendern, dem Flow Datenquellen und -senken hinzuzufügen und Funktionen wie Mapper oder auch Daten-Standardisierer und Validier einzubinden. Damit ist es möglich, die Daten an das Format des Zielsystems anzupassen und um Informationen wie beispielsweise die zuvor genannten Postleitzahlen zu ergänzen.

Da es sich beim Flow Builder um die webbasierte Entwicklungs-umgebung der Integration Cloud handelt, ist es sinnvoll, an dieser Stelle etwas genauer auf die Arbeit mit dem Werkzeug einzugehen. Möchte ein Anwender einen neuen Flow erzeugen, so hat er mit dem Tool die Möglichkeit, dem Flow zunächst einmal einen Namen zu geben. Im Test wollten wir zu diesem Zeitpunkt die Inhalte einer Excel-Tabelle in unseren Salesforce-Account hochladen.

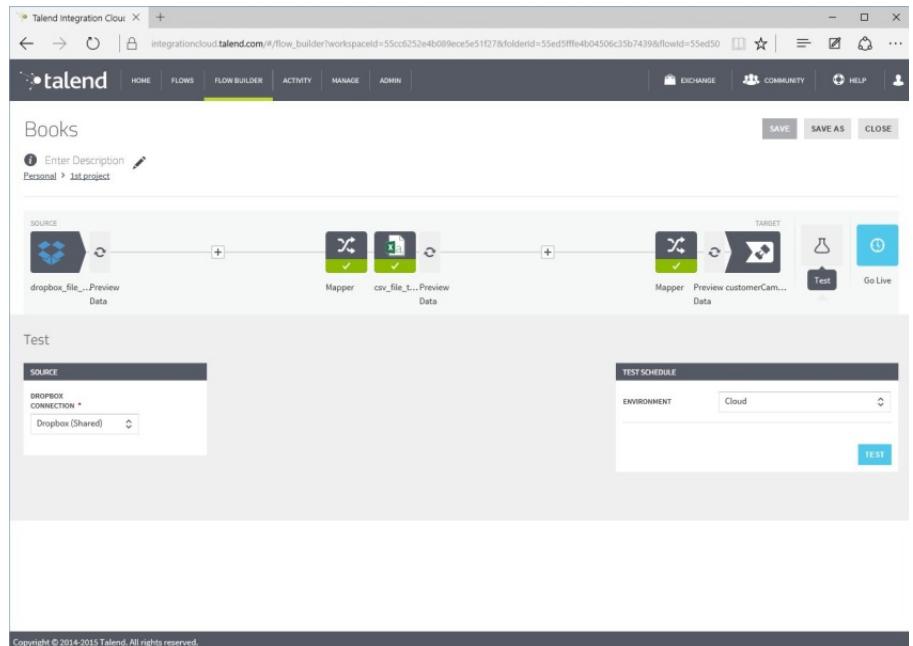
Wir nannten den Flow also "Migration from Excel to Salesforce" und gaben ihm bei der Gelegenheit auch gleich eine kurze Beschreibung mit. Anschließend wählten wir unter "Choose a Source" unsere Excel-Datei als Datenquelle aus. Da sich diese in einem Dropbox-Konto befand, selektierten wir als Quell-Icon den Eintrag "dopbox_file_download_source" und gaben unserer Quelle den Pfad zur Datei innerhalb der Dropbox und das passende Dropbox Access-Token an.

Danach ging es daran, die Informationen aus der Excel-Datei so umzuwandeln, dass das System sie modifizieren konnte. Dazu kam der Eintrag "xlsx_file_toColumns_process_step" zum Einsatz. Dieser wandelte, wie der

Name schon sagt, die in dem Excel-File vorhandenen Einträge in Spalten um. Zu diesem Zeitpunkt konnten wir unter "Preview Data" bereits sehen, welche Daten in dem Sheet vorhanden waren. Bei den genannten Spalten handelt es sich um das interne Format der Integration Cloud. In die-

Salesforce- Datenbankeinträgen zu, damit das System wusste, wo es die Daten genau hinschreiben sollte.

Alternativ ist es auch möglich, das Schema über die Funktion "Guess Schema" in Erfahrung zu bringen. Diese fragt nach der zu



Ein einfacher Flow, der Daten aus einer in der Dropbox liegenden CSV-Datei ausliest

sem Format werden die Daten im Speicher gehalten und gestreamt.

Im nächsten Schritt definierten wir unser Ziel, also unseren Salesforce-Account. Hierfür verwendeten wir die Aktion "salesforce_contact_upsertBulk_target". Diese benötigt neben den Kontoangaben auch noch ein Salesforce-Sicherheitstoken, damit sie richtig funktioniert. Sobald die erforderlichen Einträge vorgenommen worden waren, konnten wir über den Mapper die Schemadaten anlegen, die wir zuvor über den "Preview Data"-Eintrag herausgefunden hatten, also Felder für den Firmennamen, den Namen des Ansprechpartners, die Adresse und ähnliches. Zum Schluss wiesen wir diese Felder den dazugehörigen

verwendenden Quelldatei, öffnet diese und liest dann das Schema daraus aus.

Mit dem Zuweisen der Inhalte zu Feldern war die Definition des Flows abgeschlossen und wir konnten einen ersten Testlauf vornehmen. Dabei zeigte uns das System an, wie viele Datensätze es aus der Datei auslas und wie viele Datensätze hochgeladen wurden. Bei unserem einfachen Beispiel waren diese Zahlen identisch, es ist aber auch möglich, kompliziertere Flows zu definieren, die beispielsweise sämtliche Ländernamen in allen Datensätzen vereinheitlichen.

Letzteres ergibt beispielsweise Sinn, wenn in den Quelldaten unterschiedliche Bezeichnungen

für ein und dasselbe Land Verwendung finden, wie zum Beispiel "United Kingdom" und "Great Britain". Kommen solche Modifikationsmethoden zum Einsatz, so kann es bei der Namenserkennung oder bei der Umwandlung zu Fehlern kommen, so dass einige Einträge auf der Strecke bleiben.

Die "Preview Data"-Funktion zeigt in diesem Fall genau an, welche Informationen bei welchem Schritt hängen bleiben und helfen so bei der Fehlersuche. Genauso ist es auch möglich, mit einem Flow gezielt Daten auszufiltern. In diesem Fall dürfen die Zahlen der Ein- und der Ausgehenden Datenfelder ebenfalls nicht übereinstimmen und auch hier lässt sich die Vorschau nutzen, um festzustellen, ob alles so funktioniert wie erwartet.

Hat sich bei dem Testlauf herausgestellt, dass alles ordnungsgemäß läuft, so haben die Anwender über den Button "Go Live" die Option, den Flow mit einem Zeitplaner zu versehen, der ihn in regelmäßigen Abständen automatisch startet und so dafür sorgt, dass Änderungen in der Quelle zum Beispiel täglich automatisch zum Ziel übertragen werden. Dabei gibt es nicht nur die Möglichkeit, den Flow einmal, täglich, wöchentlich oder monatlich abzuarbeiten, sondern auch einen Punkt für die Auswahl der Ausführungsumgebung.

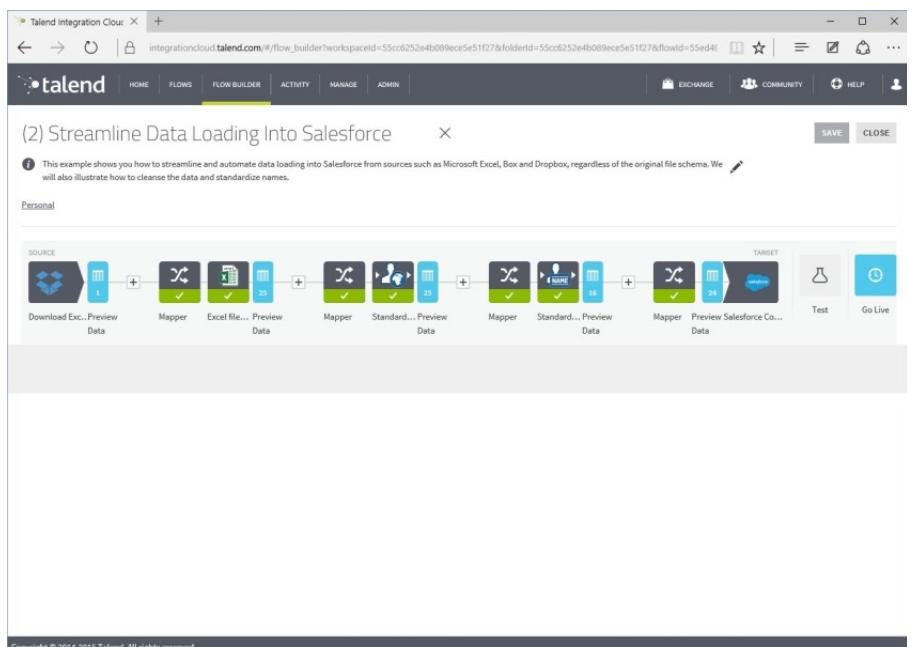
Dabei stehen die Cloud, eventuell vorhandene Remote Engines, sowie "Cloud Exclusive" und "Cloud Sandbox" zur Verfügung. Bei Cloud Exclusive erhält der Flow eine eigene Arbeitsumgebung, die er mit nichts anderem teilen muss und bei Cloud Sand-

box läuft die Arbeit so ähnlich ab wie bei Cloud Exclusive, die Cloud-Umgebung sorgt in diesem Fall aber zusätzlich dafür, dass das System Produktionsdaten nicht gleichzeitig mit Tests abarbeitet.

Im Test ergaben sich mit unserem ersten Flow keine Probleme und wir konnten ihn sofort "Live" schalten. Dabei richteten wir ihn

Aufgaben geben. Diese lassen sich entweder für benutzerdefinierte Zeiträume oder für den letzten Tag, die letzten zwei Tage, die letzte Woche beziehungsweise den letzten Monat definieren. Außerdem ist es noch möglich, die Anzeige nach Benutzern und Workspaces zu filtern.

Unter "Manage" finden sich im Gegensatz dazu die von dem je-



Dieser Flow liest Daten aus einer in der Dropbox liegenden Excel-Datei aus, standardisiert den Ländernamen und teilt das Feld "Name" in Vor- und Nachnamen auf. Anschließend überträgt er die Daten in Salesforce.

so ein, dass er täglich ablief und das funktionierte von Anfang an wie erwartet. Treten Probleme auf und müssen die zuständigen Mitarbeiter Änderungen an den Flows vornehmen, so hilft ihnen eine leistungsfähige Versionsverwaltung dabei, die Übersicht zu behalten.

Der sonstige Funktionsumfang des Web-Interfaces

Wenden wir uns nun aber wieder dem Funktionsumfang des Web-Interfaces zu. Unter "Activity" stehen den Anwendern Dashboards zur Verfügung, die Aufschluss über die gerade laufenden und die bereits durchgeführten

weiligen Benutzer verwendeten Aktionen, also die Quellen, Ziele, Umwandlungsaktionen und ähnliches. Diese lassen sich hier auch zu Talend Exchange exportieren und bei Bedarf besteht auch die Möglichkeit, weitere Aktionseinträge von dieser Plattform zu importieren.

Der Eintrag "Connections" bietet den Anwendern zusätzlich die Option, die bestehenden Verbindungen – beispielsweise zu Salesforce, zur Dropbox oder zu anderen Diensten – zu verwalten. Analog dazu existieren auch noch Einträge zu den Flow Templates, also Flow-Einträgen, die als

Muster für andere Flows vordefiniert wurden, Ressourcen (wie beispielsweise Datenbankschemata) und Remote Engines mit ihrem aktuellen Status. Alle unter Manage vorhandenen Komponenten stehen sowohl für den persönlichen als auch für den gemeinsamen Workspace zur Verfügung.

Der letzte Punkt des Web-Interfaces nennt sich "Admin". Er dient zur Verwaltung des Abonnements. Zunächst einmal können die Benutzer an dieser Stelle ein Dashboard einsehen, das ihnen Aufschluss darüber gibt, wie lange ihre Lizenz noch gültig ist, wie viele Benutzerkonten angelegt wurden und welche neuen Nachrichten es von Talend gibt.

Copyright © 2014-2015 Talend. All rights reserved.

Die Einträge der einzelnen Flows umfassen unter anderem auch die "Run History", die Aufschluss darüber gibt, ob es beim Abarbeiten der Daten zu Problemen kam

Die Benutzerverwaltung ermöglicht das Anlegen neuer User-Konten. Dazu müssen die Administratoren dem Anwender einen Login-Namen und ein Passwort geben sowie den Vor- und den Nachnamen, die E-Mail-Adresse und die Telefonnummer festlegen. In diesem Zusammenhang

ist es wichtig zu wissen, dass das Passwort sowohl Klein-, als auch Großbuchstaben sowie Zahlen und Sonderzeichen enthalten muss.

Der nächste Konfigurationsdialog bietet den Administratoren die Option, den Benutzerkonten Rechte zuzuweisen. Normale Benutzer erhalten lediglich das Recht sich einzuloggen, Administrationsbenutzer bekommen zusätzlich noch einen Haken in der Spalte "Account Admin". Nur sie haben Zugriff auf den Administrationsbereich.

Auch für die Workspaces lassen sich Rechte vergeben. So legen die Administratoren für jeden Workspace (also für die persönli-

ten von beziehungsweise zu Talend Exchange zu erlauben oder zu untersagen. Im Test traten dabei keine Probleme auf.

Der Punkt "Subscription" gibt Aufschluss über das aktuelle Abonnement. Hier sehen die User, welches Abo sie abgeschlossen haben, wie lange es gültig ist, wie viele Engines zur Verfügung stehen und wie viele Benutzerkonten sie einrichten können. Unter "Remote Engines" lassen sich im Gegensatz dazu weitere Remote Engines hinzufügen.

Das funktioniert über einen so genannten Remote Engine Key, den das Web-Interface an dieser Stelle erzeugt und den die Administratoren dann bei der Installation der Remote Engine angeben. Daraufhin weiß diese, mit welchem Konto sie sich verbinden muss und steht dann sofort nach dem Setup bereit. Da die Remote Engines lediglich eine ausgehende SSL-Verbindung benötigen, um den Kontakt zur Talend Integration Cloud herzustellen, wird in den meisten Umgebungen dazu nicht einmal eine Neukonfiguration der Firewall erforderlich sein. Im Test traten beim Einrichten der Remote Engine auf jeden Fall keine Probleme auf.

"Downloads" stellt den Anwendern schließlich Download-Links für die zusätzliche Software zur Verfügung. Neben dem Talend Studio for Cloud und den Remote Engines ist das noch der SAP RFC Server, der es bei Bedarf möglich macht, SAP IDocs zu empfangen und zu bearbeiten.

Diverse Links zur User-Community, zu Talend Exchange und eine Hilfefunktion mit Support und

Dokumentation gehören ebenfalls noch zum Funktionsumfang des Web-Interfaces. Das gleiche gilt für einen Konfigurationsdialog zum Bearbeiten des aktuellen Benutzerprofils. In diesem sind

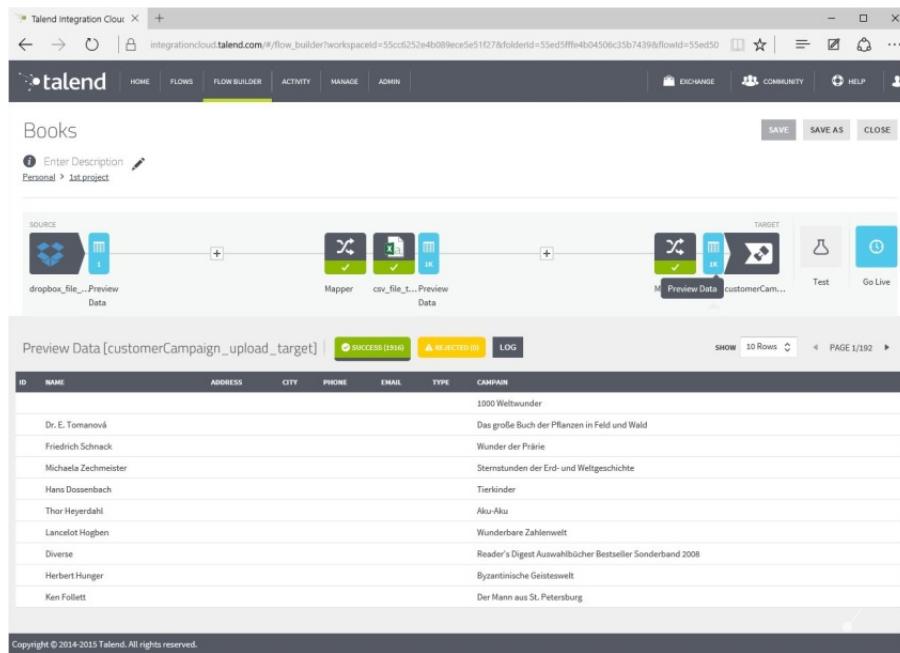
Interessanter ist das zweite Beispiel, das ähnlich wie unser zuvor beschriebenes Template Kundendaten aus einer Excel-Datei ausliest und diese dann in Salesforce exportiert. Anders als unse-

genannte Studio mit. Dabei handelt es sich um eine Eclipse-basierte Entwicklungsumgebung, die dazu in der Lage ist, die Verantwortlichen beim Erstellen der von ihnen benötigten Aktion zu unterstützen und die dann neu generierten Programme in die Integration Cloud hochzuladen, damit diese innerhalb der Flows zum Einsatz kommen können.

Anders als die meisten anderen Entwicklungsumgebungen arbeitet das Studio for Cloud mit einer Sammlung von Icons, die bestimmte Funktionen repräsentieren. Möchte ein Benutzer mit dem Studio eine Software entwickeln, so muss er lediglich die dazugehörigen Icons in den Arbeitsbereich ziehen, die benötigten Konfigurationsparameter – wie zum Beispiel erforderliche Variablen oder Pfade – angeben und die Icons anschließend so verbinden, dass die Darstellung eines Datenflusses erscheint. Die Verbindungen definieren folglich, in welcher Reihenfolge die einzelnen Funktionen abgearbeitet werden.

Zuvor hatten wir bereits mehrmals die Aktion zum Herunterladen einer Datei aus der Dropbox genutzt. Öffnet man diese im Studio for Cloud, so finden sich im Arbeitsbereich der Entwicklungsumgebung mehrere Icons, die diese Aktion definieren. Konkret handelt es sich dabei um die Arbeitsschritte "Connection Configuration", "Download File", "Rename Variables" und "Action Output Data".

Die Verbindungskonfiguration umfasst in diesem Zusammenhang die Definition der Verbindung zum Quelldienst, "Download File" erklärt sich selbst und



Über die "Preview Data"-Funktion können die Anwender jederzeit einsehen, in welcher Form die Daten – hier unsere Bucheinträge – bei welchem Arbeitsschritt ankommen

die User unter anderem dazu in der Lage, ihren Namen und ihre Mail-Adresse zu aktualisieren, ihr Passwort zu verändern und ihre Zeitzone zu setzen.

Die Arbeit mit den Sample Flows

Wie bereits erwähnt, stellt Talend den Anwendern auch Sample Flows zur Verfügung, mit denen sie sich mit der Arbeitsweise der Integration Cloud vertraut machen können. Konkret handelt es sich dabei um zwei Beispiele. Das erste ist eine einfache "Hello World!"-Anwendung.

Diese weist Kunden anhand ihres in einer Liste definierten Typs eine Kundenkampagne zu. Anschließend lässt die den Kunden – abhängig von ihrer aktuellen Kampagne – eine Nachricht, wie "Hello World!" zukommen.

re einfache Lösung standardisiert dieses Template dabei aber noch zusätzlich die Ländernamen und trennt die Vor- von den Nachnamen. Damit liefert Talend ein sehr praxisbezogenes Beispiel mit, das sich mit Datenintegration befasste Administratoren auf jeden Fall einmal ansehen sollten.

Talend Studio for Cloud

Gehen wir nun noch kurz auf das Talend Studio for Cloud ein. Obwohl Talend bereits viele Aktionen für die Integration Cloud vordefiniert hat und in Talend Exchange noch eine Vielzahl andere Features verfügbar sind, wird es vorkommen, dass die zuständigen Mitarbeiter auch eigene Datentransformationen, Datenquellen oder auch Datenziele definieren müssen. Zu diesem Zweck liefert Talend das eben

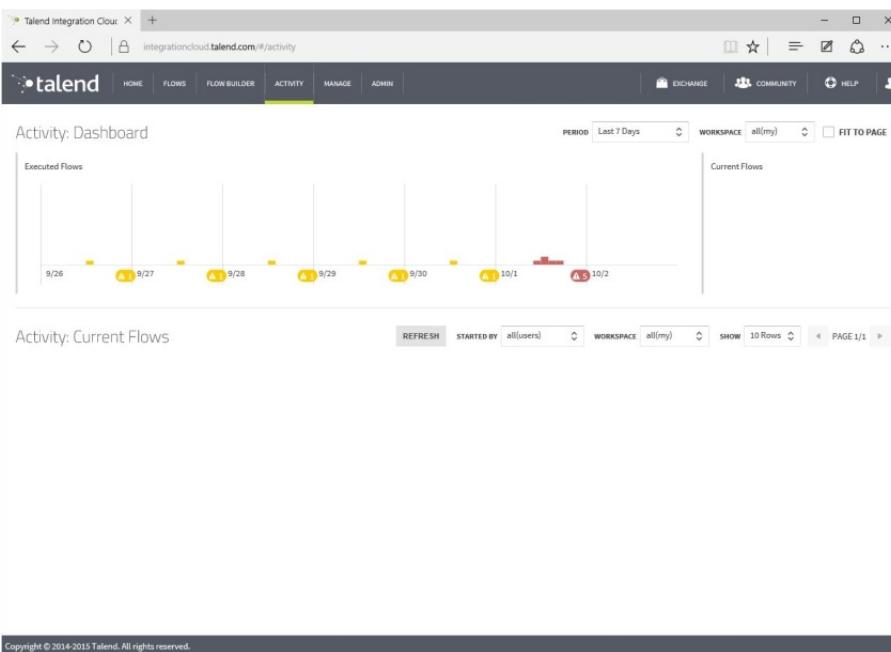
"Rename Variables" kommt zum Einsatz, um die File-Zugriffe zu vereinheitlichen. "Action Output

Gruppen wie "Big Data", "Cloud", "Datenbanken", "ELT" und "System" aufgeteilt, damit

Vielzahl der vordefinierten Funktionen nur in Einzelfällen nötig sein.

Im Test machten wir uns zu diesem Zeitpunkt daran, ein Eingabeelement zu erzeugen, das sehr spezifische Anforderungen erfüllte. Wir hatten zuvor mit Hilfe einer Android-App eine Datenbank mit den bei uns im Haus vorhandenen Büchern erstellt und wollten diese Datenbank in ein anderes System importieren. Leider war die Android-App lediglich dazu in der Lage, die Inhalte in Form einer CSV-Datei zu exportieren, die in keiner Hinsicht – weder was die Trennzeichen, noch was das verwendete Encoding betraf – standardkonform war. Sie ließ sich also nicht mit einer der vorgefertigten Importfunktionen der Integration Cloud importieren. Das lag an dem – wie gesagt sehr seltsamen – Format der Datenquelle, nicht an der Talend-Lösung selbst, machte aber manuelle Entwicklungsarbeiten nötig.

Um die CSV-Datei erfolgreich einlesen zu können, legten wir zunächst im Browser einen neuen Flow an und verwendeten dabei als Datenquelle unsere, in der Dropbox liegende CSV-Datei. Diese wandelten wir im nächsten Schritt mit der Aktion "csv_file_toColumns_process_step" so um, dass das System die enthaltenen Daten erkennen konnte. Dazu mussten wir diese Aktion innerhalb des Studio for Cloud so modifizieren, dass sie unsere etwas seltsame CSV-Datei richtig interpretieren konnte. Dabei war es erforderlich, manuell das Encoding, den Field-Separator und ähnliche Parameter, die in unserer Datei nicht den üblichen Erwartungen entsprachen, an die



Copyright © 2014-2015 Talend. All rights reserved.

Das "Activity Dashboard" zeigt auf einen Blick, welche Flows wann gelaufen sind und ob es dabei zu Schwierigkeiten kam

Data" übernimmt die Aufgabe, die Informationen an den nächsten Arbeitsschritt weiter zu leiten. Parallel dazu finden sich auch noch zwei andere Einträge, die mit dem eigentlichen Workflow nichts zu tun haben. "Catch Error" erkennt auftretende Fehler und leitet sie an "Log Error" zur Aufzeichnung weiter.

Auf diese Weise lassen sich also nicht nur neue Funktionen implementieren, sondern auch bereits vorhandene Aktionen modifizieren, erweitern und an die jeweiligen Anforderungen anpassen. In der Praxis läuft die Arbeit mit dem Studio folgendermaßen ab: Nach dem Anlegen oder Öffnen eines Projekts haben die zuständigen Mitarbeiter die Möglichkeit, auf der rechten Seite des Arbeitsfensters auf eine Art "Werkzeugkasten" zuzugreifen, der alle Icons mit den vordefinierten Funktionen enthält. Dieser lässt sich einfach nach Schlagworten durchsuchen, wurde aber auch in

die Verantwortlichen ihre Tools problemlos finden. Kennen sie den Namen der von ihnen gerade benötigten Funktion, so reicht es sogar, diesen Namen in Arbeitsbereich einzutippen, danach bietet das Studio direkt das entsprechende Icon an. Greifen die Mitarbeiter auf den Werkzeugkasten zurück, so können sie die Icons einfach per Drag-and-Drop platzieren.

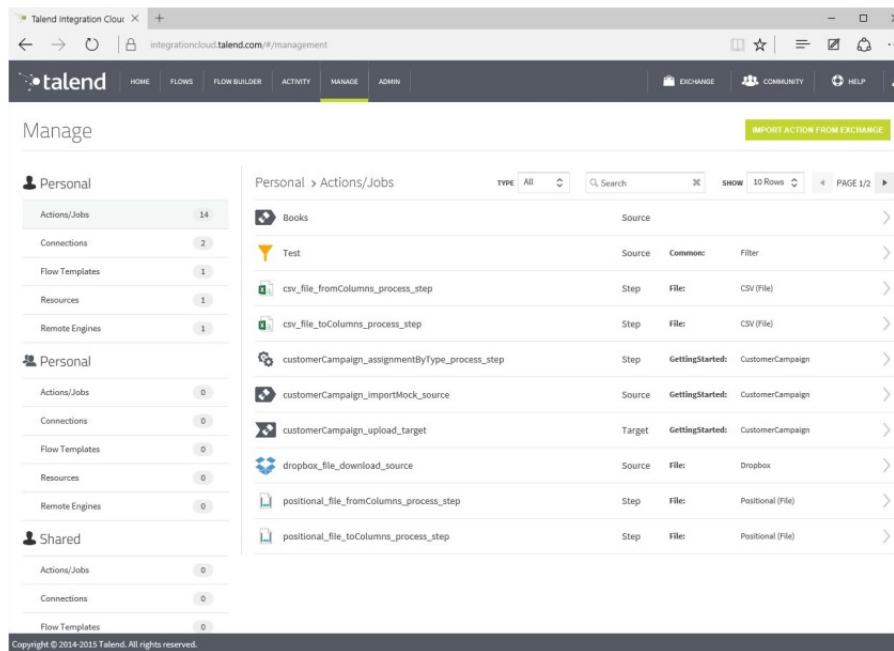
Wurde die Funktion mit allen Icons und dem Datenfluss definiert, so haben die zuständigen Mitarbeiter die Möglichkeit, sie direkt im Studio zu testen. Läuft der Task ohne Schwierigkeiten durch, so lässt er sich im nächsten Schritt in die Integration Cloud hochladen und dort benutzen. Treten Fehler auf, so bietet das Studio umfassende Optionen zum Debuggen. Falls erforderlich, sind die Entwickler auch jederzeit dazu in der Lage, eigenen Code in ihre Projekte einzubinden. Das wird aber aufgrund der

aktuellen Gegebenheiten anzupassen. Nachdem wir die modifizierte Aktion über einen Rechtsklick auf den Eintrag und

Trotz der großen Flexibilität und des großen Funktionsumfangs ist die Lösung relativ einfach zu bedienen und ermöglicht es Unter-

nisse über die Arbeit mit dem Web-Interface und dem Flow Builder benötigen.

Auch die IT-Administration wird durch die Integration Cloud deutlich entlastet, macht die Cloud es doch überflüssig, die Umgebung zum Abarbeiten der Datenintegrationsprojekte mit Hard- und Software, Hochverfügbarkeit und Sicherheit on-premise zu implementieren und zu warten. Das spart vor allem im Big Data-Umfeld viel Geld und Arbeit. Damit stellt die Talend Integration Cloud für viele Unternehmen eine interessante Alternative zu klassischen Datenintegrationsprojekten dar. Auch andere Organisationen, die sich bislang nicht wirklich an das Thema "Datenintegration" herangetraut haben, erhalten über die Cloud eine einfach zu nutzende, risikolose und komfortable Einstiegsoption.



The screenshot shows the Talend Integration Cloud Management interface. The left sidebar has sections for Personal (Actions/Jobs: 14, Connections: 2, Flow Templates: 1, Resources: 1, Remote Engines: 1), Personal (Actions/Jobs: 0, Connections: 0, Flow Templates: 0, Resources: 0, Remote Engines: 0), and Shared (Actions/Jobs: 0, Connections: 0, Flow Templates: 0). The main area is titled 'Personal > Actions/Jobs' and shows a list of actions. The list includes: 'Books' (Source), 'Test' (Source, Common: Filter), 'csv_file_fromColumns_process_step' (Step, Files: CSV (File)), 'csv_file_toColumns_process_step' (Step, Files: CSV (File)), 'customerCampaign_assignmentByType_process_step' (Step, GettingStarted: CustomerCampaign), 'customerCampaign_importMock_source' (Source, GettingStarted: CustomerCampaign), 'customerCampaign_upload_target' (Target, GettingStarted: CustomerCampaign), 'dropbox_file_download_source' (Source, Files: Dropbox), 'positional_file_fromColumns_process_step' (Step, Files: Positional (File)), and 'positional_file_toColumns_process_step' (Step, Files: Positional (File)). There are buttons for 'Import Action from Exchange' and 'PAGE 1/2'.

Unter "Manage" lassen sich Aktionen, Connections, Flow Templates, Ressourcen und Remote Engines verwalten

die Auswahl des Befehls "Publish to Cloud" in unseren Workspace in der Cloud hochgeladen hatten, konnten wir sie zu unserem Flow hinzufügen.

Um die Bücherliste zu exportieren genügte es nun, im Flow Builder der Integration Cloud die Zieldatenbank anzugeben und das Mapping vorzunehmen. Danach migrierte die Cloud-Lösung alle Informationen in die Datenbank. Bei uns funktionierte das zwar nicht auf Anhieb, da wir mit den Parametern zur Interpretation der CSV-Datei etwas herumprobieren mussten, dank der integrierten Versionsverwaltung war es aber im Test kein Problem, mit dem Studio so lange herum zu probieren, bis das Ergebnis unseren Erwartungen entsprach.

Fazit

Mit der Integration Cloud liefert Talend ein sehr leistungsfähiges Produkt zur Datenintegration aus.

nehmen beispielsweise "Do it yourself"- Datenintegrationsprojekte zu realisieren.

In vielen Fällen kann es dazu erforderlich sein, dass die Entwicklungsabteilung mit dem Studio für Cloud die im jeweiligen Unternehmen benötigten Aktionen vordefiniert. Wurde das einmal erledigt, so können die Datenintegrationsverantwortlichen als eine Art Endbenutzer über das Web-Interface die von ihnen benötigten Flows selbst zusammenstellen und modifizieren.

Das Studio brauchen sie dazu nicht und ebenfalls keine Schulung für die Entwicklung mit dem Studio. Das spart sowohl Zeit für die Entwicklungsabteilung, die sich nach der erstmaligen Einrichtung der benötigten Aktionen nur noch in Sonderfällen mit der Datenintegration befassen muss, als auch bei den "Endbenutzern", die nur Kennt-

Gehen wir zum Schluss noch kurz auf die Sicherheit der Talend Integration Cloud ein. Die Lösung setzt auf Amazon Web Services auf und arbeitet mit einer doppelten Verschlüsselung der Passwörter und Verbindungsdaten. Außerdem wird jeder Zugang zur Cloud mit einem Schlüssel abgesichert. Die Cloud speichert übrigens im Betrieb keine Nutzdaten, diese werden einfach durch das System "durchgestreamt". Lediglich Metadaten, wie zum Beispiel Verbindungsinformationen, liegen in verschlüsselter Form dauerhaft in der Cloud.

Aufgrund des hohen Sicherheitsniveaus, der vielen Einsatzmöglichkeiten und des großen Leistungsumfangs verleihen wir der Talend Integration Cloud die Auszeichnung "IAIT Tested and Recommended".