

Falk Lehmann ist Market Unit Manager
of BI & Reporting bei metafinanz Informations-
systeme GmbH, München

Datenqualitäts-Firewall

Datenprüfung für Konzerncontrolling

Data Governance-Prozesse gelten im Finanz- und Versicherungssektor als geschäftskritisch, weil bei ungenügender Qualität von Daten enorme finanzielle und nicht-monetäre Risiken drohen. Banken und Versicherungen implementieren daher engmaschige Prüfnetze mit Data Quality (DQ)-Routinen - doch Entwicklung und Wartung dieses Instrumentariums verschlingt oft enorme Ressourcen. In einem Kundenprojekt entwickelte man daher auf Grundlage einer serviceorientierten Architektur (SOA) zentrale DQ-Services, die sowohl in den weltweit verteilten heterogenen Anwendungssystemen als auch von den klassischen ETL-Batch-Routinen genutzt werden. Datenqualität wird an einem zentralen Punkt fokussiert – bündelt damit die Kompetenzen und verringert die IT-Aufwände.

Spätestens seit dem Zusammenbruch von Lehman Brothers und der weltweiten Finanzkrise rückt das Thema Data-Governance in das Blickfeld des Risikomanagements der Banken- und Versicherungswirtschaft. Zwar sah sich der Bankensektor bereits mit Basel II einer Vielzahl von regulatorischen Anforderungen ausgesetzt, doch werden die Bestimmungen auf breiter Front ausgeweitet. Nun kommen auch auf die Versicherungswirtschaft unter dem Stichwort Solvency II neue Regeln für das Risikomanagement zu, die bis 2012 umgesetzt sein müssen.

Für Finanzdienstleister hat dies vor allem in Konzerncontrolling und -bilanzierung signifikant höheren Aufwand für das Bereitstellen und Aufbereiten von qualita-

tiv hochwertigen Daten für das Risikomanagement zur Folge. Insbesondere im Hinblick auf die dritte Säule von Solvency II – der Berichterstattungspflichten von Versicherungsunternehmen - benötigt man konsistente Informationen für Öffentlichkeit und Aufsichtsbehörden, wobei im Rahmen der implementierten Data-Governance die Datenqualität für das interne und das externe Reporting sichergestellt werden muss. Im Falle mangelhafter Datenqualität steigt das Risiko, Reputationsschäden zu erleiden oder auch hohe Schadensersatzforderungen leisten zu müssen.

Steigender Reporting-Aufwand

In der Praxis ist das Streben nach hoher Datenqualität allerdings mit einigem Aufwand verbunden, wie das Projekt eines Datenproviders für das konzernweite Asset-Liability-Management eines großen Finanzdienstleisters zeigte. Für das Controlling der Asset- und Investmentmanagement-Prozesse müssen tagtäglich eine große Anzahl von Reports und Analysen erstellt werden, die eine essentielle Bedeutung für das Compliance- und Risikomanagement des Konzerns haben. Die Daten für dieses Reporting werden von einer Vielzahl unterschiedlicher Liefersysteme aus dem gesamten weltweiten Konzernumfeld sowie externer Marktdatenprovider bereitgestellt. Üblicherweise werden die Daten als CSV-Dateien oder proprietäre Datafeeds in das Datawarehouse geladen. Da die fachlichen Inhalte der Dateien sich stark unterscheiden, muss für jedes Liefersystem eine eigene Datenqualitäts-Prüfroutine innerhalb des ETL-Prozesses entwickelt werden.

Neben der technischen Komplexität ergeben sich auch fachlich diverse Herausforderungen: So bestehen beispielsweise bei manchen Reports Abhängigkeiten zwischen unterschiedlichen Datenlieferungen - im Portfoliomanagement betrifft das etwa die Composites. Bei Composites handelt es sich um eine Bündelung mehrerer Investmentportfolios gleichen Gestaltungstyps. Als Datenlieferung kommen hier einerseits Marktdaten, wie Aktien- oder Terminkurse herein, die den Wert eines gesamten Composites darstellen, gleichzeitig laufen auch Daten jedes einzelnen inhärenten Investmentfonds von den Fondsgesellschaften ins zentrale Datawarehouse. Am Ende muss die AUM-Summe („Asset-under-Management“: Wert der Aktie, des Fonds oder des Composites) der Einzelwerte identisch sein mit dem Gesamtwert des alle Investmentportfolios beinhaltenden Composites.

Um die Datenqualität sicherzustellen, kamen in der Vergangenheit verschiedene Datenqualitäts-Prüfungen zum Einsatz, die sich jedoch auf einzelne oder Gruppen von Lieferungen bezogen. Durch diese Heterogenität entstehen nicht nur zeitliche Verzögerungen beim Reporting, sondern auch unnötige Mehrkosten.

Portal und Webservices mit zentralem ETL-Prozess

Aufgrund der vielschichtigen Anforderungen entschied man sich, die gesamten Qualitätsprüfungsprozesse zu zentralisieren. Die Idee dabei war, auf der bestehenden Informatica Powercenter-Architektur mit ihren klassischen ETL-Schnittstellen und den bereits implementierten DQ-Checks aufzusetzen, gleichzeitig aber diese DQ-Funktionen durch Einsatz der Informatica Realtime-Option auch weiteren Applikationssystemen – wie JEE-Portalen (Java Enterprise Edition) oder Rich-Client-Applikationen als zentrale DQ-Services für eine direkte Nutzung zur Verfügung zu stellen – als singulären Punkt der Datenprüfung in einem mehrstufigen ETL-Prozess.

Die neue Architektur bietet in Richtung Anwender/Datenlieferant in der aktuellen Ausbaustufe zwei Schnittstellen: Zum einen entwickelte man für den allgemeinen Zugang ein universelles User-Interface auf Basis eines Java-Web-Clients, die auf den Webservice zugreift. Der Kunde meldet sich dabei über eine SSL-gesicherte Verbindung an, um seine Daten ins Datawarehouse zu laden. Am Client startet nun eine zweistufige Datenprüfung, die per Informatica-Webservice auf die ETL-Mechanismen zugreift. Es handelt sich hierbei um ein interaktives Di-

alogverfahren, das über Anzeige und Editierfunktionen verfügt. Zunächst findet im ersten Schritt eine Vorprüfung der gelieferten Daten statt. Diese basiert auf über Metadaten konfigurierbaren generischen Checks, so dass für neue Liefersysteme keine Programmierung, sondern lediglich eine Konfiguration erforderlich ist. Implementiert sind diese generischen Checks mittels regulärer Ausdrücke, welche durch die in Informatica vorhandene Perl-kompatible Regular-Expression-Engine ausgeführt werden. Durch diese Metadaten-basierte Architektur ist das System nicht auf ein bestimmtes CSV-Datenformat festgelegt, sondern kann flexibel je nach fachlichem Kontext konfiguriert werden.

Die zu prüfenden Daten könnten beispielsweise nach folgendem Muster aufgebaut sein: In der ersten Spalte der Lieferung steht das Lieferdatum, in der zweiten der Fondsname, in der dritten die ISIN-Kennzahl (International Security Identification Number; die ISIN dient der eindeutigen internationalen Identifikation von Wertpapieren), in der vierten der Anlagetyp – zum Beispiel EQ für Equity oder FI für Fixed Income – und so weiter. Per Feldprüfung wird nun zunächst nach syntaktischen oder logischen Fehlern gesucht. Tauchen in einer oder mehreren Feldprüfungen Fehler auf, markiert das System die entsprechenden Zeilen und Datenfelder und stellt die Daten mit Status-

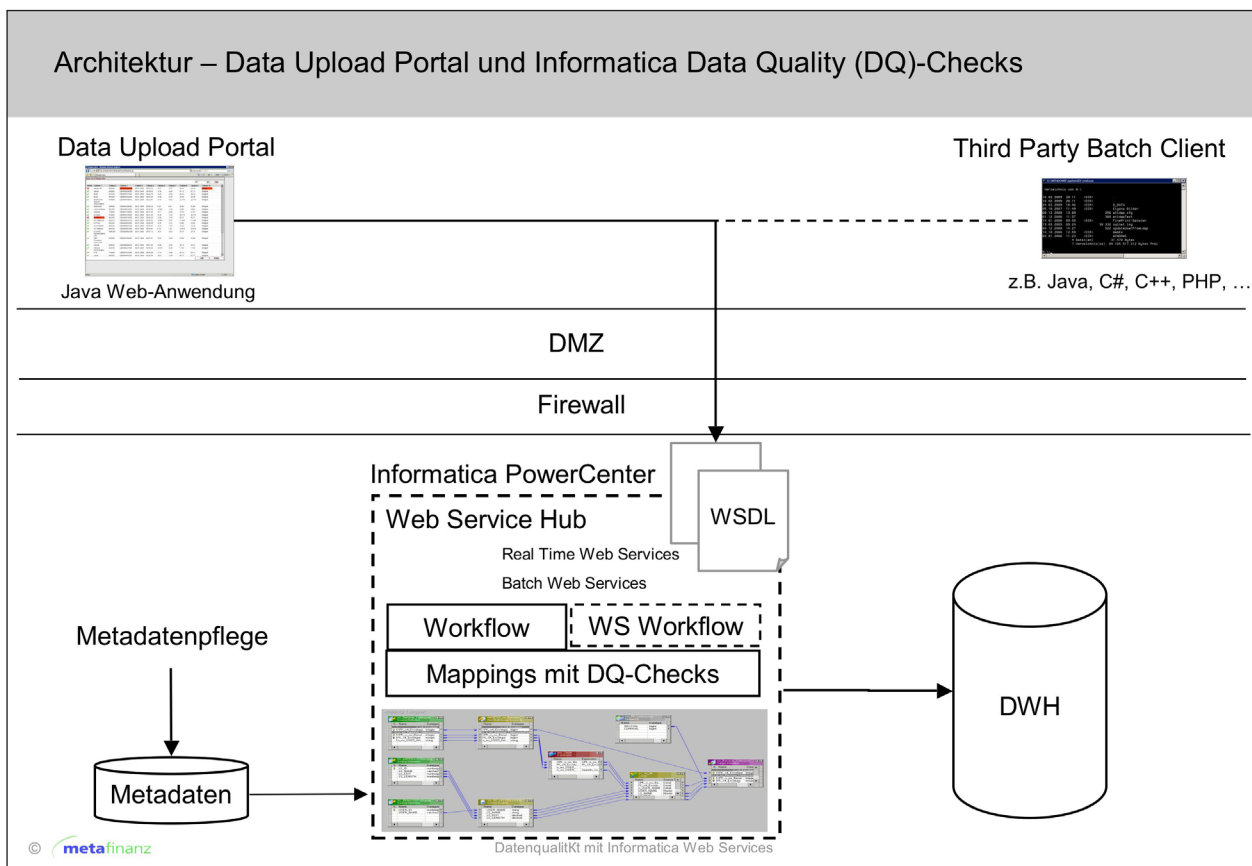


Abb. 1: WSDL ist die XML-Beschreibung einer Webservice-Schnittstelle. Die Grafik beschreibt, welche Daten(typen) in den Service hinein und hinaus gehen – als eine Art Kontrakt zwischen Datenlieferant und Informatica-Service.

anzeigen an der Benutzerschnittstelle dar, wo sie der Anwender umgehend interaktiv ausbessern kann. Die korrigierten Daten werden nun automatisch per Asynchronous Javascript and XML (AJAX) für eine erneute Prüfung an das Datawarehouse zurückgespielt und das Ergebnis der erneuten Prüfung wird dem Anwender ohne Page-Refresh des User-Interfaces dargestellt. Sobald die Daten alle konfigurierten generischen Prüfungen erfüllen, kann in einem zweiten Schritt eine in Informatica hinterlegte Spezialprüfung durchgeführt werden. Hierbei ist es notwendig, eine der Fachlichkeit der Lieferung entsprechende ETL-Prüfroutine zu schreiben. Diese ETL-Prüfroutine wird als klassisches Mapping implementiert. Denkbar wäre beispielsweise die zu Beginn des Artikels skizzierte fachliche Anforderung nach Gleichheit von AUM-Summe des Composites und der AUM-Summen der im Composite enthaltenen Einzelportfolios. Hierzu würde im Mapping die AUM-Summe des Composites mit einer bereits früher erfolgten und deshalb bereits im Datawarehouse vorhandenen Lieferung der enthaltenen Fonds verglichen.

Als zweiten Zugang zur ETL-Logik können die Webservices über beliebige Programmiersprachen als Batch-Schnittstelle genutzt werden. Auf der Seite der dezentralen Liefersysteme eröffnet sich damit ein breites Spektrum für Individuallösungen, die alle stets mit der zentralen Ladeinstanz von Informatica kommunizieren. Nützlich ist das etwa für Lieferanten mit hoher Frequenz: Statt die täglichen Fonds-Schlusskurse immer händisch über das Portal zu laden, ist es möglich einen Adapter schreiben, der die Zahlen automatisch liefert.

Fazit: Self-Service und Zentralisierung der Kompetenzen

Mit dieser zentralisierten Lösung profitiert der Kunde in mehrfacher Hinsicht. Das Portal steht den Anwen-

dern weltweit rund um die Uhr zur Verfügung und folgt damit der Self-Service-Philosophie. Gleichzeitig kommt nun gewissermaßen das Verursacherprinzip zum Tragen: Datenfehler müssen dort korrigiert werden, wo sie entstehen, nämlich beim Lieferanten und in dessen Quellsystemen. Die Lieferanten wiederum haben nun die Möglichkeit, Fehler bereits direkt während der Datenladung in Echtzeit angezeigt zu bekommen und interaktiv zu beheben.

Auf der anderen Seite wird mit dieser Architektur auch ein Single Point of Responsibility geschaffen. Hatte man in der Vergangenheit verschiedene Datenprüfungs-Firewalls entwickelt und damit das Rad immer wieder neu erfunden, konzentriert sich diese Aufgabe und das damit verbundene fachliche Knowhow jetzt direkt am Datawarehouse. Webentwickler an externen Standorten müssen sich nun nicht mehr mit speziellen fachlichen Fragen auseinandersetzen, sondern können die vorhandenen SOA-Komponenten direkt in ihre Portale einbauen.

Schließlich spricht auch noch ein ständiger Aktualisierungsbedarf für eine konsequente Zentralisierung. Sobald etwa Gesetzesänderungen vorliegen, müssen diese nur noch einmal im ETL-Tool umgesetzt werden, alle liefernden Systeme in den externen Organisationen bleiben davon zukünftig unberührt.

BI-SPEKTRUM ist eine Fachpublikation des Verlags:
SIGS DATACOM GmbH | Lindlaustraße 2c | 53842 Troisdorf
Tel.: +49 (0) 22 41.2341-100 | Fax: +49 (0) 22 41.2341-199
E-mail: info@sig-datacom.de
www.bi-spektrum.de

SIGS DATACOM
FACHINFORMATIONEN FÜR IT-PROFESSIONALS