

Einsatz von Data Federation für den schnellen Aufbau eines BI-Systems

Dr. Nick Golovin, Eligent Data GbR

Der Beitrag schildert Erfahrungen aus einem Business-Intelligence-Projekt bei einem internationalen mittelständischen Medien-Unternehmen, in dem Data Federation für den schnellen Aufbau eines BI-Systems in einer stark heterogenen, internationalen IT-Landschaft zum Einsatz kam. Die Highlights des Ansatzes sind die schnellen ersten Ergebnisse, dank Data Federation kombiniert mit steigender Performance durch den graduellen Aufbau von einem Data Warehouse im Hintergrund.

Die zahlreichen Probleme klassischer Data-Warehousing-Lösungen sind bereits seit Langem gut bekannt. Dazu gehören unter anderem die langwierige Konzipierungsphase, aufwändige und teure technische Implementierung und eine schlechte Anpassbarkeit an die Änderungen in den Geschäftsprozessen. Diese Herausforderungen können von den großen Konzernen dank deren reichlichen Ressourcen in der Regel bewältigt werden, Mittelstandsfirmen haben aber ihre Probleme mit den Kosten und Aufwänden, welche für den fachgerechten Aufbau eines Data Warehouse benötigt werden. An dieser Stelle wird häufig gespart und auf Halblösungen ausgewichen – mit negativen Konsequenzen für die Zufriedenheit der internen Kunden.

Auch die Art der Projekt-Einführung ist bei einem klassischen Data Warehouse wenig mittelstandsfreundlich. Die häufig angewendete Wasserfall-Methode mit ausführlicher Spezifikationsphase ist für die Großkonzerne, bei denen die Geschäftsprozesse relativ formalisiert sind und die Key-User es gewohnt sind, die Anforderungen im Detail niederzuschreiben, besser geeignet. Im Mittelstand, wo eher die „hands-on“-Mentalität herrscht, ist die lückenlose Formalisierung der Anforderungen schwierig. Auch sind beim Mittelstand die Geschäftsprozesse viel flexibler, deshalb fällt es den Key-Usern schwierig, bereits heute die Anforderungen von übermorgen präzise zu definieren. Die iterative Projekt-Einführung, bei der jeder Schritt von den Key-Usern kontrolliert und gegebenenfalls angepasst werden kann, wäre eine geeignetere Methode, insbesondere im Mittelstandsbereich.

Trotz dieser Schwierigkeiten wächst der Bedarf nach Daten-Integrations-Lösungen für Business Intelligence auch im Mittelstand, wo immer neue Anforderungen entstehen. So werden etwa häufiger Echtzeit-Daten aus den Quellsystemen für operatives Reporting gebraucht (operatives BI, real-time BI). Für tiefere Analysen auf der Ebene der Fachabteilungen brauchen die Endbenutzer auch Kontrolle über die Extraktion von Daten aus den Quellsystemen (Self-Service-BI, In-Memory-Analyse). Diese Anforderungen werden von klassischen Data-Warehouse-Lösungen in der Regel nicht erfüllt.

Als eine mögliche Antwort auf die Problematik der Daten-Integration und -Analyse für den Mittelstand haben sich seit einiger Zeit die sogenannten „Self-Service-BI-Tools“ und „In-Memory-Analyse-Tools“ etabliert, um Analysen der großen Datenmengen direkt beim Endanwender zu ermöglichen. Diese Tools können zwar den unmittelbaren Bedarf der Fachabteilungen nach der flexiblen Datenanalyse stillen, werden aber aus strategischer Sicht als Rückschritt angesehen. Solche Lösungen weichen von dem Single-Source-of-Truth-Konzept ab und führen in gewisser Hinsicht zurück in die Zeiten vor Business Intelligence, in denen jeder Mitarbeiter seine Daten-Analysen in den eigenen, unzähligen Excel-Dateien betrieben hatte.

Aus diesen Gründen werden neue, strategisch tragfähige Ansätze für die Integration der Daten aus unterschiedlichen Datenquellen gefordert, insbesondere im Hinblick auf die speziellen Anforderungen des Mittelstands bezüglich Kosten und Flexibilität. Ein solcher Ansatz wurde bei der Koch Media GmbH zur Einführung des Business-Intelligence-Projekts erfolgreich angewendet.

Dabei ist eine BI-Lösung entstanden, die kostengünstig, flexibel und dennoch performant und strategisch tragfähig ist.

Kundenszenario und verwendete Technologien

Die Koch Media GmbH wurde 1994 gegründet und ist spezialisiert auf Distribution, Publishing und Producing von Entertainment-Produkten (Computerspiele, Videospiele, Online-Spiele, DVD/Blu-ray-Filme, Consumer-Software). Die Firma zählt 2011 ca. 350 Mitarbeiter weltweit (Europa und USA), der jährliche Umsatz liegt bei mehr als 300 Millionen Euro. Die Entwicklung des Unternehmens in der Zeit vor der Einführung des Business Intelligence wurde durch starkes Wachstum und Expansion ins Ausland gekennzeichnet. Dabei wurden einige ausländische Unternehmen mit vorhandener IT-Infrastruktur übernommen. Des Weiteren waren einige ausländische Niederlassungen historisch bedingt im IT-Bereich relativ eigenständig. Dadurch ist eine stark heterogene Systemlandschaft mit vier großen ERP-Systemen und zahlreichen anderen kleineren Systemen und Datenbanken entstanden.

Es wurde notwendig, der Konzern-Geschäftsführung und der Leitung der zentralen Fachabteilungen wie Controlling, Finanzbuchhaltung und internationales Marketing einen Überblick und Einblick in die Daten in allen diesen Systemen zu verschaffen. Unter diesen Systemen und Datenquellen befinden sich neben den vier ERP-Systemen noch mehrere Internet-Shops, Produkt-Datenbanken, Projektmanagement-Systeme, Groupware-Systeme, externe Partner und Marktforschungs-Da-

ten. Die hinter diesen Systemen liegenden Datenbank-Systeme umfassen Oracle, Informix, IBM DB2/AS400, MS SQL, MySQL, PostgreSQL und MS Access. Die Anzahl der Datensätze in den größten Tabellen liegt bei mehreren zehn Millionen. Die gesamten Datenbestände umfassen mehrere zehn Terabyte. Die Dokumentation der Daten und Strukturen war leider in großen Teilen lückenhaft. Vor der Einführung von Business Intelligence wurden unterschiedlichste Werkzeuge verwendet, um Daten aus den Systemen zu extrahieren und Berichte zu generieren (RPG, Access, Excel, SQL Server, Lotus, AS400-Data Transfer, ASP, JSP/Jasper, siehe Abbildung 1). Ein solches dezentrales Berichtswesen war natürlich schwer zu unterstützen, es gab viele redundante Auswertungen und die Diskrepanzen in den Auswertungen waren sehr schwer zu verfolgen. Es konnte auch keine Rechteverwaltung eingesetzt werden.

Die Verwendung unterschiedlicher Technologien erschwerte den Aufbau von Know-how. Bei den Anforderungen für die zu implementierende Business-Intelligence-Architektur wurden zwei große Aufgabenbereiche definiert:

- *Operatives Berichtswesen für jedes Land*
Berichte sowohl mit Echtzeit-Daten (Verkaufsdaten, Finanzdaten, Produktdaten) als auch mit historischen Daten
- *Analytisches Berichtswesen, vor allem länderübergreifend*
Überblick über alle Koch-Firmen, interaktive Analyse mit Drill-Down, Ad-hoc-Berichte, Analyse von historischen Daten. Die Endkunden für solch analytisches Berichtswesen sind Top-Management, Controlling, internationales Marketing, internationales Producing/Publishing (Produktionsplanung)

Bei der Analyse der Anforderungen wurde schnell klar, dass ein klassisches Data Warehouse keine passende Lösung bietet. Ausschlaggebend waren dabei die folgenden Kriterien:

- In vielen Bereichen wurden Echtzeitdaten aus mehreren Systemen benötigt, teilweise in Kombination mit historischen Daten
- Einige der sehr wichtigen Berichte aus mehreren Datenquellen waren von

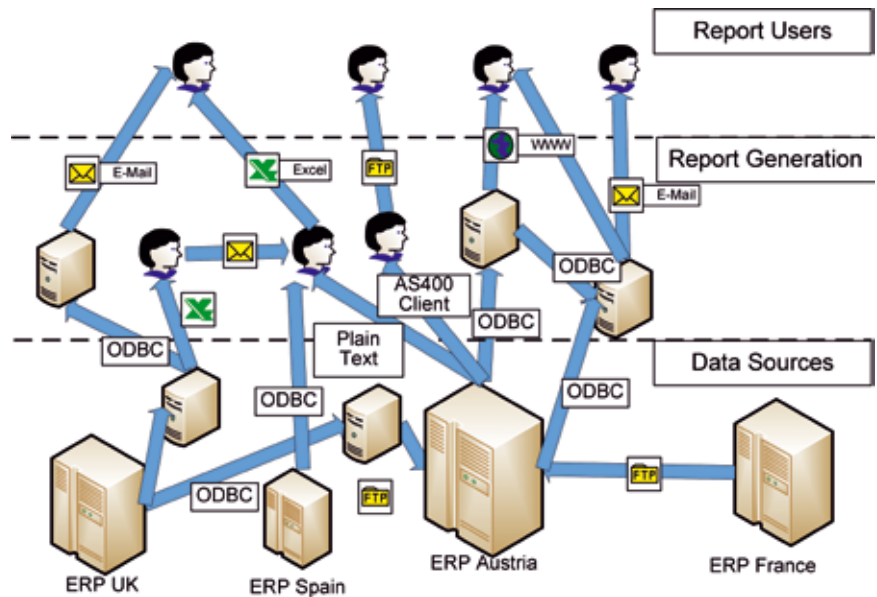


Abbildung 1: Reporting-Situation vor der Einführung von Business Intelligence

der Geschäftsleitung schnellstmöglich benötigt, das Abwarten, bis das Data Warehouse fertig ist, stand nicht zur Diskussion

- Es war abzusehen, dass die Geschäftsprozesse sich in der absehbaren Zukunft stark ändern werden, sodass keine zuverlässige Konzipierung möglich war

Im Hinblick auf diese speziellen Anforderungen fiel die Entscheidung, die Data-Federation-Technologie einzusetzen. Deren Einsatz lässt alle an das Data-Federation-Tool angeschlossenen Datenquellen wie eine einzige große, relationale Datenquelle aussehen. Die Datenquellen können über ODBC, JDBC oder Web-Services angeschlossen sein, das Data-Federation-Tool selbst kann ebenfalls über JDBC, ODBC oder Web-Service angesprochen werden. Es wandelt die empfangene SQL-Abfrage in mehrere Teil-Abfragen um, die es an die angeschlossenen Datenquellen schickt, bekommt die Ergebnisse zurück und führt die Teilergebnisse zum endgültigen Ergebnis zusammen. Das Ganze geschieht „on-the-fly“, ohne dass sich der Entwickler um die Speicherung der Zwischenergebnisse kümmern muss. Ein Data-Federation-Tool ist in der Lage, auch große Datenmengen mit mehreren Millionen Datensätzen problemlos zu bewältigen, es versucht je-

doch, einen möglichst großen Anteil der Abfragen-Bearbeitung an die Datenquellen auszulagern, was die Menge der von den Datenquellen zu übertragenden Daten im Vorfeld minimiert. Beispiele für Data-Federation-Tools sind Oracle Data Service Integrator, Business Objects Data Services von SAP und InfoSphere Federation Server von IBM.

Durch Einsatz von Data Federation können die Datenquellen sehr schnell angebunden und die ersten Ergebnisse bereits am ersten Projekttag präsentiert werden – schneller als das mit einem klassischen Data Warehouse möglich wäre. Allerdings gibt es auch Nachteile, die gegen den alleinigen Einsatz von Data Federation bei den Business-Intelligence-Projekten sprechen. Er werden zum Beispiel keine historischen Daten gespeichert. Da Data Federation direkt auf die Datenquellen zugreift, werden diese unter Umständen stärker belastet. Auch die Geschwindigkeit der Ausführung der Abfragen, die über das Data-Federation-Tool an die Datenquellen geschickt werden, ist langsamer als bei einem Data Warehouse. Das liegt jedoch nicht an dem Data-Federation-Tool, sondern daran, dass die dahinterliegenden Datenquellen eventuell eine höhere Netzwerk-Latenz haben und für die Ausführung von analytischen Abfragen nicht ausgelegt sind. Aus diesen Gründen hat man sich für eine hybride Ar-

chitektur bei der Durchführung des Projekts entschieden.

Abbildung 2 zeigt, wie bei der eingesetzten Architektur alle analytischen Abfragen grundsätzlich durch den Data-Federation-Layer bearbeitet werden. Als eine der angeschlossenen Datenquellen fungiert ein klassisches Data Warehouse, das mithilfe von ETL-Tools erstellt und befüllt wird. Wichtig dabei ist jedoch, dass nicht alle für die Analyse benötigten Daten in diesem Data Warehouse gespeichert sein müssen. Nur in folgenden Fällen ergibt die Unterbringung von Tabellen in einem Data Warehouse Sinn:

- Die Tabelle ist besonders groß oder kommt von einem System, das über ein langsames Netzwerk angebunden ist
- Es ist eine spezielle Verarbeitung/Bereinigung von Daten gewünscht, die mit SQL nicht möglich oder besonders langsam ist
- Es muss eine Historie für die Daten aufbewahrt werden, die aus den Quellsystemen regelmäßig gelöscht wird

In diesem Fall mussten insgesamt nur etwa zehn Prozent der Tabellen, die für die Analysen erforderlich sind, auch tatsächlich in dem Data Warehouse physisch abgespeichert werden. Deshalb ist hier eher von einem „Teil-Data-Warehouse“ die Rede. Durch den Wegfall der Notwendigkeit, alle Daten im Data Warehouse zu haben, ist eine geschätzte 90-prozentige Ersparnis bei der Entwicklung von ETL-Prozessen und Datenstrukturen im Data Warehouse erzielt worden.

Eine wichtige Eigenschaft der vorgestellten Architektur besteht darin, dass die Nutzerberichte nicht wissen müssen, ob bestimmte Daten direkt aus dem Quellsystem oder aus dem Teil-Data-Warehouse kommen. Dadurch wird eine iterative Entwicklung des BI-Systems in engem Kontakt mit den Endnutzern ermöglicht. Diese Architektur erlaubt, die ersten Bericht-Prototypen ausschließlich basierend auf Data Federation zu entwickeln. Ein solcher Prototyp kann extrem kurzfristig entstehen und schnell in Rücksprache mit dem Endbenutzer nachgebessert und finalisiert werden. Erst wenn der Bericht fertig ist, kann man dessen Performance analysieren und ge-

gebenfalls entsprechende Datenstrukturen im Teil-Data-Warehouse vorbereiten. Danach wird der Data-Federation-Layer so umkonfiguriert, dass die relevanten Abfragen aus dem Teil-Data-Warehouse kommen. Dabei müssen der Bericht und dessen SQL-Abfragen nicht geändert werden. Im Laufe des Projekts ist es dank dieser Architektur möglich:

- Alle Daten im Unternehmen zu integrieren und für die Berichte zur Verfügung zu stellen, ohne sie komplett in einem Data Warehouse abspeichern zu müssen
- Bei der Projekt-Einführung eine sehr signifikante Kostenersparnis zu erreichen, dank des verringerten Aufwands für den Aufbau des Data Warehouse
- Der Geschäftsleitung die notwendigsten Berichte schnellstmöglich zu präsentieren, was mit einem klassischen Data Warehouse unmöglich gewesen wäre
- Eine iterative Vorgehensweise bei der Projekteinführung anzuwenden. Dabei

war man imstande, den Endnutzern die jeweils nächsten Versionen der Bericht-prototypen in kurzen Zeitabständen präsentieren zu können, sodass der Fortschritt sichtbar wird und das Feedback besser berücksichtigt werden kann. Damit war die Zufriedenheit der internen Kunden aufrechterhalten.

Ursprung und Ausblick

Die bei dem beschriebenen Projekt angewendeten Ideen und Ansätze basieren auf Forschungsergebnissen, die am Lehrstuhl für Datenbanken an der Universität Leipzig entstanden sind (Lehrstuhl-Inhaber Prof. Dr. E. Rahm). Diese Forschung wird im Rahmen des Projekts „Eligent Data“ an der Universität Leipzig fortgeführt. Dabei entsteht ein Daten-Integrations-Produkt, das den hier beschriebenen Ansatz auf die nächste Stufe bringt. Dank der vorgelagerten Data-Federation-Schicht ist das neue Produkt imstande, die für die Berichte verwendeten Daten und deren Nutzungsmuster zu sammeln

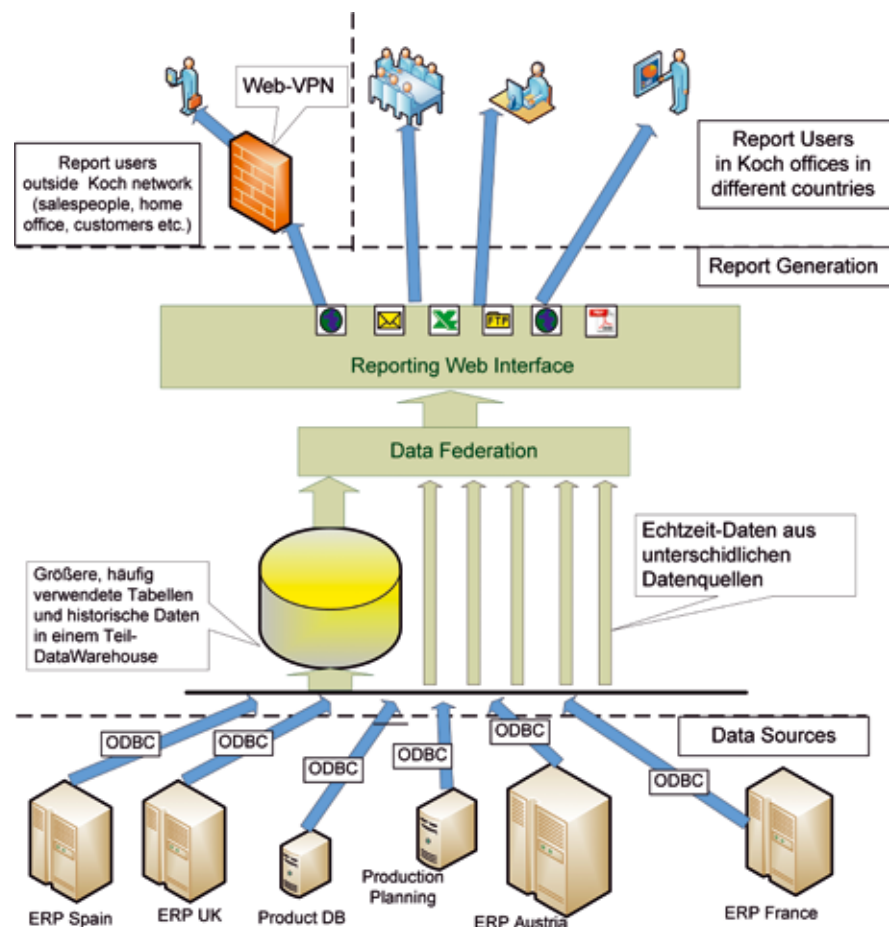


Abbildung 2: Business Intelligence mit Data Federation und (Teil)Data Warehouse

und zu analysieren. Basierend auf dieser Analyse kann die Daten-Integrations-Software die Struktur des Teil-Data-Warehouse vorschlagen und sie per Knopfdruck samt der nötigen ETL-Prozesse automatisch implementieren, was zur weiteren Senkung der Kosten und des Aufwands für den Aufbau von BI-Projekten führt. Das Projekt an der Universität Leipzig sucht momentan nach Kooperationspartnern und potenziellen Kunden für das innovative Daten-Integrations-Produkt.

Fazit

Der Artikel hat einen innovativen Ansatz für den schnellen und kostengünstigen Aufbau eines BI-Systems unter Verwendung von Data Federation skizziert. Der geschilderte Ansatz stammt aus den Forschungsergebnissen des Lehrstuhls für Datenbanken an der Universität Leipzig und wurde bei einem international aufgestellten, mittelständischen Medienunternehmen zur Einführung eines Business-Intelligence-Projekts erfolgreich angewendet. Die resultierende Lösung

ist durch geringe Kosten, kurze Einführungszeit, hohe Flexibilität und Performanz sowie den problemlosen Zugriff auf Echtzeitdaten aus den Quellsystemen gekennzeichnet. Dieser Ansatz wird im Rahmen des Projekts „Eligent Data“ an der Universität Leipzig erweitert und weitergeführt, sodass eine nahtlose Integration von Data Warehousing und Data Federation ermöglicht wird.

Dr. Nick Golovin

n.golovin@kochmedia.com

nick.golovin@eligentdata.de



Dick Faris, Senior Vice President
Oracle Primavera Global Business Unit

„Primavera ist weltweit als Standard in der Industrie-Branche etabliert ...“

Als die Projektmanagement-Software Primavera 1983 auf den Markt kam, veränderte die Desktop-basierte Lösung die IT-Landschaft in puncto Projektmanagement nachhaltig und behauptete sich dann als eine der führenden Lösungen. Primavera wird auch nach der Übernahme durch Oracle immer noch von Dick Faris, Mitgründer von Primavera und jetziger Oracle Primavera Global Business Unit Senior Vice President, verantwortet. Am Mittwoch, 14. September 2011, stand der Primavera-Guru in einer Q&A-Session beim 3. Primavera Community Day den Teilnehmern Rede und Antwort. Christian Körner und Sebastian Hunke, Leiter der DOAG BSC Primavera Community, haben mit Dick Faris über die Entwicklung der Software gesprochen.

Wie würden Sie die Herangehensweise von Oracle in Bezug auf Primavera beschreiben?

Faris: Oracle hat Primavera als eine globale Geschäftseinheit mit separatem Vertrieb, Entwicklung und Produktstrategie etabliert, um sicher zu stellen, dass ihre Identität und Produkte weiterhin den Bedürfnissen ihres traditionellen Kundenstamms gerecht wird und ihre Präsenz in diesen Industrie-Branchen ausgebaut.

Hat Oracle die Absicht, Primavera in das Oracle Fusion Project Portfolio Management zu integrieren? Könnte es das Ende von Primavera als allein stehendes Tool bedeuten?

Faris: Primavera P6 ist bereits in Oracle Fusion PPM integriert, wie alle anderen Oracle Enterprise Resource Planning (ERP)-Lösungen auch (E-Business Suite, JD Edwards,

PeopleSoft und SAP). Oracle plant, die Primavera-Lösung weiterhin als eine starke EPPM-Lösung zu entwickeln, die mehrere ERP-Systeme integriert.

Welche Rolle wird die Primavera Client-Version in Zukunft haben?

Faris: Oracle plant eine kontinuierliche Verbesserung der P6-Client-Version, um sicher zu stellen, dass sie ein wichtiges Instrument für Projekt-Experten bleibt.

Was sind die Stärken der Primavera-Lösung im Vergleich zu anderen führenden Projektmanagement-Tools?

Faris: Primavera hat sich weltweit als Standard in der Industrie-Branche etabliert. Sie ist zudem in der Lage, einen unternehmensweiten Überblick über alle Projekte

innerhalb einer Organisation zu liefern. Primavera kann beliebig skaliert werden, um alle Projektarten innerhalb einer Organisation zu bedienen, von dem kleinsten bis zum größten Projekt. Das heißt, die Primavera-Lösung eignet sich für Unternehmen jeder Größe.

In welchen Bereichen muss Primavera Ihrer Meinung nach verbessert werden?

Faris: Ich glaube, Primavera muss weiterhin ihr Footprint erweitern, um eine breitere Gruppe von Projektteilnehmern, Führungskräften und anderen Stakeholdern bedienen zu können. Die Primavera-Lösung muss auch mit Workflow und BI weiter angepasst werden, um die spezifischen Bedürfnisse der vertikalen Industrien und der einzelnen Anwender besser zu bedienen.