

Open Source Planung für ein Oracle Data Warehouse

Hans-Peter Weih
Standard Life Versicherung
Frankfurt

Michael Weiler
PROMATIS software GmbH
Ettlingen

Schlüsselworte

Business Intelligence (BI), Open Source, OLAP, Planung, Oracle, Data Warehouse (DWH)

Einleitung

Standard Life plc ist ein führender britischer Lebensversicherer und Vermögensverwalter mit weltweit über 6 Mio. Kunden und mehr als 250 Mrd. Euro Anlagevermögen. Seit 1996 ist Standard Life auch auf dem deutschen Markt mit einer Tochtergesellschaft und seit 1999 mit einer österreichischen Vertriebsdirektion präsent und hat sich hier mit mehr als einer halben Million Kunden als Maklerversicherer etabliert.

Aus Frankfurt werden alle Kunden und Makler in Deutschland und Österreich betreut, die Vertragsbestände verwaltet, die Anwendungsprogramme entwickelt sowie die notwendige Hardware betrieben.

Schon sehr früh erkannte man die Notwendigkeit eines Data Warehouses in der deutschen Niederlassung, um die Daten der verschiedenen Anwendungen zu integrieren und ein effizientes und schnelles Reporting zu ermöglichen. Die strategische Plattform für die Datenbanken aller Anwendungen war von Beginn an Oracle, so dass Oracle für das Data Warehouse gesetzt war. Die zunehmend umfangreichere Budgetplanung wurde jedoch weiterhin mit Excel betrieben, so dass eine softwareunterstützte Lösung immer dringender wurde.

Der Beitrag beschreibt die Ausgangssituation, die Toolwahl, den Entscheidungsprozess sowie die Einführung und die weitere Entwicklung zum strategischen OLAP Tool für die Niederlassung im Zusammenspiel mit Oracle und SAP.

Open Source für Business Intelligence (BI)

Während Open Source schon sehr lange ein heiß diskutiertes Thema für Betriebssysteme und vor allem Standardsoftware wie Büro-, oder Zeichenprogramme ist, ist Open Source in der BI erst seit wenigen Jahren im Einsatz. Hier gibt es im Wesentlichen zwei Modelle:

- Klassische Open Source von einer Entwicklergemeinde getrieben wie Linux oder
- kommerzielle Open Source von einem Unternehmen entwickelt und getrieben.

Hinter der Entwicklung eines Unternehmens steht oft der Gedanke, eine Grundidee freizugeben und einzusetzen, den Gewinn jedoch durch Supportleistungen sowie zusätzliche kostenpflichtige Features zu erwirtschaften. Hierbei entspricht sehr häufig die Open Source Version nicht dem aktuellen Entwicklungsstand, sondern dem Stand einer vorherigen Version.

Kommerzielle Open Source Produkte haben den Vorteil, dass ein Unternehmen mit Entwicklungsplan und zentraler Codeentwicklung im Hintergrund die gezielte Weiterentwicklung eher garantieren kann als eine Entwicklergemeinde. Im Umfeld von BI existieren beide Modelle, z.B. Pentaho als Beispiel für eine von der Entwicklergemeinde bereitgestellten Software sowie PALO oder Talend als Beispiele für „kommerzielle“ Open-Source-Produkte.

Ausgangssituation

Mit dem Ansatz „Best of Breed“ hatte man sich in den vergangenen Jahren bei Standard Life eine sehr heterogene BI Landschaft aufgebaut. Neben der Oracle Datenbank, das ETL Tool Data Manager von

Cognos, WebFOCUS von Information Builders als Web basiertes Tool sowie PowerPlay von Cognos als OLAP Datenbank (siehe Abbildung 2 ohne PALO Komponente).

Das Data Warehouse beinhaltet aktuell Daten aus den unterschiedlichsten Quellsystemen wie der Bestandsverwaltung, der Telefonanlage, dem Workflow, den Leistungsdatenbanken, der Personaldatenbank und SAP. Neben Oracle werden auch Daten aus Access, MySQL und anderen Quellen eingelesen. Der wöchentlich verarbeitete Datenauszug der Bestandsverwaltung entspricht mit 100 GB dem größten Datenvolumen. Dieser besteht aus mehreren Tabellen mit jeweils mehreren hundert Millionen Datensätzen.

Das DWH ist in den Jahren auf ein Volumen von 250 GB angewachsen und beinhaltet über 100 Dimensions- und über 300 Fakttabellen. Es sind drei virtuelle Server im Einsatz, Windows und Linux noch mit 32 Bit, jeweils 3,5 GB Hauptspeicher und 2 Prozessoren, AIX mit 0,75 CPU garantiert und maximal 2 CPU. Der Festplattenplatz wird den Systemen über ein SAN zur Verfügung gestellt.

Zu Beginn des Projekts waren alle notwendigen Daten bereits im Data Warehouse enthalten bzw. existierten Schnittstellen zu allen notwendigen Systemen, da der bisherige Planungsprozess bereits auf den Daten des Data Warehouses aufgebaut war.

Der Planungsprozess bei Standard Life wurde mit Hilfe mehrerer Spreadsheets realisiert. Abbildung 1 zeigt einen Überblick über den Planungsprozess.

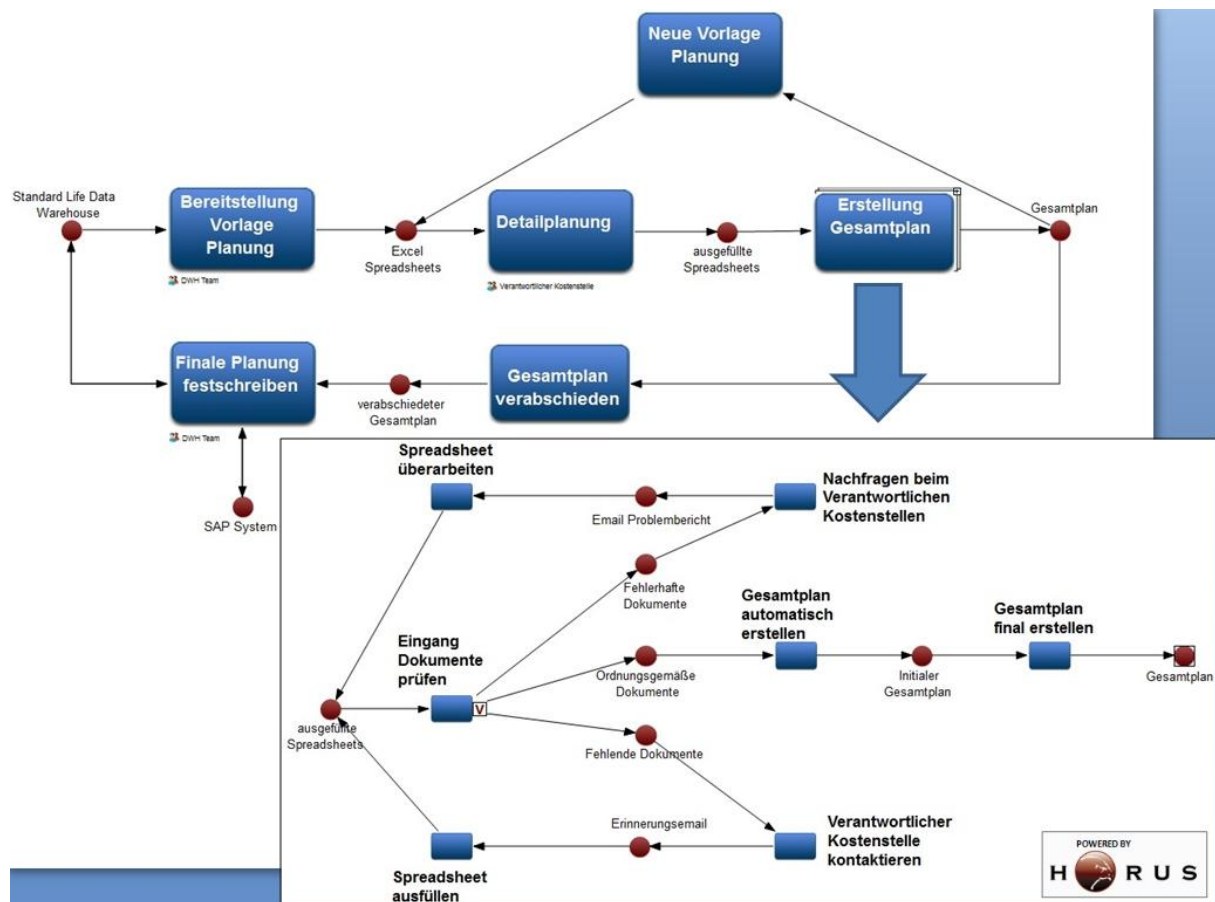


Abbildung 1: Planungsprozess mit Detailablauf vor Einführung der Open Source Planung

Ausgehend vom Data Warehouse wurden mehrere Excel Spreadsheets mit den Kostenarten an die Kostenstellenverantwortlichen verteilt. Die Vorlagen enthielten die Istwerte des aktuellen Jahres, die auf den Rest des Jahres projiziert wurden. Auf dieser Basis sollten die Kostenstellenverantwortlichen

ihr Budget für das kommende Jahr planen. Für die Rücksendung der Planung wurde eine Frist gesetzt, die Spreadsheets wurden eingesammelt und in einen Gesamtplan konsolidiert. Dabei gab es immer wieder Verzögerungen durch fehlerhafte oder verspätete Lieferungen. Die automatische Konsolidierung durch ein Visual Basic-Makro konnte den Gesamtplan nur unregelmäßig vollautomatisch erstellen, so dass manuelle Nacharbeiten zur Erstellung des finalen Plans notwendig waren. Dieses Bottom-up Budget wurde im Anschluss durch Vorgaben der Geschäftsleitung eingeschränkt, basierend auf dem erwartenden Neugeschäft und der Profitabilität. Es erfolgt eine zweite Planungsrunde nach dem gleichen Prozedere und eventuell auch eine dritte Planungsrunde mit dem Resultat, dass das Budget oft erst im folgenden Jahr verabschiedet wurde. Neben der langen Dauer des Prozesses wurde mit jeder Planungsrunde die Anzahl der Dateien, die zum Nachweis der Planung gespeichert und versioniert werden mussten, vervielfacht.

Das finale Ergebnis wurde dann in SAP übertragen und von dort ins Data Warehouse überspielt. Um die Ist- bzw. Soll-Vergleiche für das aktuelle Jahr bereit zu stellen, wurden die Istwerte des laufenden Jahres auf Monatsbasis aus SAP ins DWH übernommen und den Kostenstellenverantwortlichen in einem weiteren Excel-Bericht zur Verfügung gestellt mit Hinweisen zu problematischen Kostenarten bzw. der Möglichkeit, die Istwerte, die über oder unter Plan liegen, zu kommentieren. Die Ergebnisse wurden wiederum eingesammelt und abgespeichert. Problematisch war hier vor allem die Planung einer Kostenart über mehrere Kostenstellen (z.B. IT-Kosten). Zusammenfassend war der Planungsprozess aufwändig, Fehler anfällig und langsam. Die Verbesserung der Prozesse war das zentrale Ziel der Einführung einer neuen Planungslösung.

Auswahlprozess

In einem Auswahlprozess wurden mehrere Software-Anbieter betrachtet. Rasch zeichnete sich ab, dass der Aufbau einer zentralen OLAP Datenbank mit Anbindung an SAP und das Oracle-basierte Data Warehouse eine deutliche Vereinfachung des Planungsprozesses ermöglicht.

Aufgrund der vielen Kostenstellenverantwortlichen waren die Lizenzkosten einer kommerziellen Lösung enorm. Daher wurden auch Open Source Produkte evaluiert. Mit der Open Source OLAP Datenbank PALO konnte ein „Proof of Concept“ realisiert werden, der eine deutliche Vereinfachung des Planungsprozesses ermöglichte. PALO nutzt als Frontend unter anderem Microsoft Excel. Die Umstellung war die für die Benutzer deshalb einfacher, da das bisherige „Look and Feel“ erhalten blieb. Auch passte sich PALO in die bestehende Architektur optimal ein und so erkannte man das Potential zum weiteren Ausbau. Eine kurze Berechnung zeigte, dass sich PALO schon nach einer Planungsrunde amortisiert hatte.

Der neue Prozess

Zu Beginn des neuen Planungsprozesses werden zuerst die Istwerte als Grundlage aus SAP mittels PALO ETL in die Oracle Datenbank geladen und von dort aus die OLAP Datenbank gefüllt. Somit stehen den Planenden die Werte des aktuellen Jahres erweitert um Schätzwerte des Budgets bis zum Jahresende als Grundlage zur Verfügung.

Mit PALO gibt es jetzt lediglich ein Excel Spreadsheet das mittels "Excel Plug-In" mit der OLAP Datenbank kommuniziert. Es wird von den angemeldeten Benutzern aufgerufen und durch ein Rechtekmanagement Zugriff auf Ihre Kostenstellen gewährt. Somit können sie auf Kostenarten- und Monatebene planen und die Werte direkt in die Datenbank zurückschreiben. Die Werte werden sofort konsolidiert und stehen ihnen als Ergebnis direkt zur Verfügung. So gibt es nach Planungsende keine Konsolidierung mehr, und die Ergebnisse können direkt weiterverarbeitet werden. Durch eine einfache Kopierfunktion wird eine neue Version der Planung angelegt, so dass Top-Down mittels PALO Funktionen das vorgegebene Budget verteilt werden kann und die Kostenstellenverantwortlichen darauf erneut planen können.

Das finale Budget wird ins Data Warehouse zurückgespielt, in SAP übertragen und dient als Grundlage für das weitere Ist- bzw. Soll-Reporting.

Mit Beginn des neuen Budgetjahres werden nun Monatsreports erstellt mit Ist- und Plan-Werten. Da alle Daten aus SAP ins DWH übertragen werden, können einerseits Berichte über das Web basierte Reporting-Tool abgerufen, andererseits die Daten weiter in PALO übertragen und den Benutzern dort zur Kommentierung zur Verfügung gestellt werden.

Erfahrungen bei der Einführung der neuen Planungslösung

Mit einem Aufwand von nur 15 Beratertagen konnte die Lösung innerhalb von sechs Wochen neben dem normalen Tagesgeschäft eingeführt werden. Anpassungen waren nur bei der Installationsroutine für die Clients nötig um eine Installation im Hintergrund ausführen zu können. Alle weiteren Komponenten konnten unverändert benutzt werden.

Der bestehende virtuelle Windows Server konnte als Plattform für PALO benutzt werden, die 3,5 GB Hauptspeicher waren ausreichend für die In-Memory OLAP-Datenbank.

Bis zu 50 Kostenstellenverantwortliche planten zeitweise auf der Datenbank monatlich auf Kostenarten- und Kostenstellenebene. Die Personalplanung der einzelnen Kostenstellen wurde sogar durch einzelne Personen durchgeführt.

Die bisherige Datenübertragung von SAP in Oracle erfolgte über in SAP programmierte Berichte, die als Textfiles abgespeichert und erneut in Oracle eingelesen wurden. Allein die Erstellung der Berichte war sehr zeit- und kostenintensiv. Dank PALO ETL können nun die Daten direkt aus SAP über das Dictionary angesprochen und eingelesen werden, was zu einer deutlichen Performance-Steigerung führte. Komplexere Abfragen in SAP werden durch Definition von Sichten in SAP oder über Weiterverarbeitung von SAP Tabellenkopien im DWH performant umgesetzt.

Ein Problem stellte die Synchronisation zwischen PALO und Oracle dar. Dies wurde so gelöst, dass in PALO die Primärschlüssel und Datentypen des Oracle Data Warehouses übernommen wurden. Dadurch wurde sichergestellt, dass das Oracle Data Warehouse als zentrales System und die PALO Daten jederzeit konsistent sind. Oracle ist das führende System, aber die PALO Würfel können jederzeit aus dem DWH wieder aufgebaut werden. Mit PALO konnte die bisherige Data Warehouse Architektur unverändert weiter benutzt werden (siehe Abbildung 2). Dies war ein wichtiger Erfolgsfaktor für das Projekt.

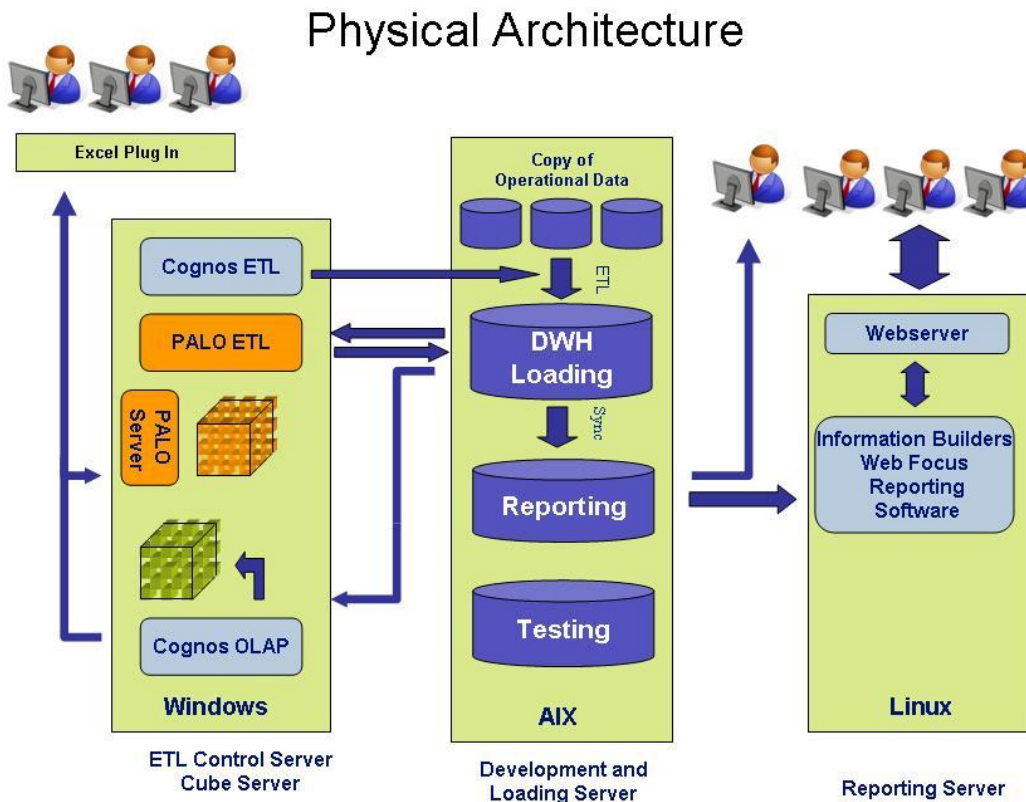


Abbildung 2: BI Architektur nach Einführung der Open Source Planung

Weiterer Ausbau

Nach erfolgreicher Einführung von PALO und der vielfältigen Schnittstellen u.a. zu SAP und Oracle war schnell klar, dass PALO als strategisches OLAP-Tool ausgebaut werden sollte. Ziel war vor allem die bisherigen manuellen Prozesse zu automatisieren und Benutzern Statistiken komfortabler zur Verfügung stellen zu können. Da die Lizenzkosten hier keine Rolle mehr spielen, war eine Erweiterung des Nutzerkreises ohne Schwierigkeiten und Einschränkungen möglich.

So wurde als erstes Projekt ein Buchungsabgleich zwischen dem Bestandsführungssystem und SAP eingeführt, um nachzuweisen dass beide Systeme identisch sind. Hierbei werden die Daten mittels PALO ETL ins Data Warehouse auf Einzelpostenebene geladen und dann zu Berichtszwecken im Würfel aggregiert. Bei Abweichungen können Abweichungsanalysen auf Einzelpostenebene erstellt werden.

Die Erstellung des Business Plans wurde komplett auf PALO und das Data Warehouse migriert. Eine sehr komplexe Excel-Anwendung mit vielen Berechnungsroutinen und Verknüpfungen wurde im Data Warehouse implementiert und PALO als Grundlage der Ergebnispräsentation benutzt. Dies war so erfolgreich, dass inzwischen auch der Business Plan der irischen Niederlassung im deutschen DWH berechnet und die Darstellung mit einer lokalen PALO Installation vor Ort durchgeführt wird.

Fazit

BI Open Source kann es durchaus mit kommerziellen Tools aufnehmen. Die Funktionalität ist oft nicht so umfangreich wie bei professionellen Tools, aber für bestimmte Problemstellungen ausreichend. Die Kosten sind daher oft das Hauptargument, wenngleich man erwähnen muss, dass nach den ersten Schritten ein Supportvertrag oder eine Lizenzierung der kommerziellen Lösung

durchaus sinnvoll ist, um erweiterte Funktionalitäten, wie zum Beispiel den SAP Adapter von PALO, zu erhalten. Zu Beginn des Projekts existierte bereits die datei-basierte Schnittstelle zu SAP, so dass alle Daten im DWH vorhanden waren und sie mit dem kostenfreien ETL in PALO geladen werden konnten. Der Vorteil überwog den der datei-basierten Schnittstelle vor allem in Hinblick auf den weiteren Ausbau, so dass diese lizenziert wurde.

Eine weitere wichtige Erkenntnis aus dem Projekt ist, dass Open Source keine Insellösung darstellt. Die Integration in gängige BI Architekturen stellt in der Regel kein Problem dar.

Kontaktadressen:

Hans Peter Weih
Teamleiter Management Information
Standard Life Versicherung
Lyoner Str. 15
60528 Frankfurt/Main

Telefon: +49(0) 69 66572-240
Fax: +49(0) 69 66572-110
E-Mail: hanspeter.weih@standardlife.de
Internet: www.standardlife.de

Michael Weiler
Division Manager
PROMATIS software GmbH
Pforzheimer Str. 160
76275 Ettlingen

Telefon: +49(0)7243 2179-0
Fax: +49(0)7243 2179-99
E-Mail: michael.weiler@promatis.de
Internet: www.promatis.de