

XML-Datenimport in das SAP Business Information Warehouse bei Bayer MaterialScience

Michael Hahne

cundus AG

Lothar Burow

Bayer MaterialScience AG

Torben Elvers

Bayer Business Services GmbH

Das SAP Business Information Warehouse (BW) ist die technologische Basiskomponente zum Aufbau von Business Intelligence Anwendungen mit SAP Produkten. Während für das Laden von Daten aus SAP R/3-Systemen über generische und Standard-Extraktoren eine direkte Verbindung zur Verfügung steht, weist der ETL-Prozess für Nicht-R/3 Quellsysteme eine deutlich höhere Komplexität auf. Ein Kernproblem ist dabei, dass Fehler oftmals erst während oder nach einem versuchten Ladevorgang feststellbar sind. Hier bietet das SAP BW seit der Version 3.0 erweiterte Möglichkeiten auf Basis der XML-Technologie, mit der Daten über einen SOAP Webservice an das System übergeben werden können. Diese neue Technologie nutzt Bayer MaterialScience zur Integration von Daten der Nicht-R/3-Gesellschaften des Konzerns. Eine umfangreiche Business Server Page (BSP) Applikation dient dabei der Verwaltung umfangreicher Datenlieferungen als reine Web-Anwendung. Auf Basis dieser Technik erhält Bayer MaterialScience bereits heute umfangreiche Forecast-Daten und wird zukünftig auf die bisherigen komplexen hostbasierten Integrationsanwendungen komplett verzichten können.

-
- ♦ Veröffentlicht als: Hahne, Michael; Burow, Lothar; Elvers, Torben: XML-Datenimport in das SAP Business Information Warehouse bei Bayer MaterialScience, in: Schelp, Joachim; Winter, Robert (Hrsg.): Auf dem Weg zur Integration Factory, Physica-Verlag, Heidelberg, 2005, S. 231-251. Siehe auch http://www.springer.de/cgi/svcat/search_book.pl?isbn=3-7908-1564-0.

1 Einleitung

Im Umfeld des Data Warehousing, OLAP und Data Mining positioniert die SAP AG ihre Lösung mySAP Business Intelligence für vielfältige betriebliche Anwendungen von Fach- und Führungskräften zur Analyse und Entscheidungsunterstützung. Technologische Basiskomponente ist das Produkt Business Information Warehouse (BW), das den Aufbau von Systemen mit einer konsistenten und einheitlichen Datenbasis ermöglicht.

Beim Import von Daten in das SAP Business Information Warehouse ist grundsätzlich zwischen SAP R/3 und Nicht-SAP R/3 Systemen zu unterscheiden. Während für erstgenannte Standard- und generische Extraktoren im Business Information Warehouse (BW) existieren und somit eine direkte Verbindung zwischen dem BW (Zielsystem) und dem R/3 (Quellsystem) hergestellt werden kann, gestaltet sich der Import aus Nicht-R/3 Systemen zumeist komplexer.

Einerseits sind die Systeme selbst nicht in der Lage unmittelbar miteinander zu kommunizieren, andererseits müssen die Daten in ein für das Business Information Warehouse geeignetes Format transformiert und vor dem Import validiert werden. Oft wird aus den Quellsystemen ein sogenanntes Flatfile (flache Datei) extrahiert und dem BW über Zwischenschritte bzw. Zwischensysteme periodisch zum Import zur Verfügung gestellt.

Folgende Schritte sind daher im Regelfall durchzuführen:

- Extraktion aus dem Quellsystem
- Übermittlung an ein Schnittstellensystem
- Transformation der Daten in ein geeignetes Format
- Validierung der Daten
- Übertragen der Daten zum Business Information Warehouse
- Laden der Daten in das Business Information Warehouse
- Ggf. Rückmeldung(en) an das Quellsystem

Eine Validierung der Daten schon im Quellsystem findet im Regelfall nicht statt. Oft werden Fehler in der Datei erst auf dem Zwischensystem oder im schlechtesten Fall beim Importieren bemerkt und der entsprechende Ladejob im Business Information Warehouse bricht ab. Erst durch manuelles Überprüfen des Monitors im BW wird in diesem Fall ein Abbruch erkannt und daraufhin ggf. die anliefernde Gesellschaft benachrichtigt. Die betroffene Nicht-R/3 Gesellschaft erfährt erst sehr spät, dass Probleme beim Datenladen aufgetreten sind. Eventuell werden Fehler sogar erst bemerkt, wenn Daten in den Datenzielen fehlen.

Mögliche Fehler sind:

- Fehler bei der Datenübertragung
- Fehler in der Struktur der Datei
- Fehler im Format der einzelnen Datensätze
- Inhaltliche Fehler (Stammdaten)

Bayer MaterialScience hat ein mehrstufiges Projekt zum Datenimport von XML Daten in das BW aufgesetzt. Hierbei ist die Verwendung eines XML/SOAP-Webservices Bestandteil einer komplexeren Service Applikation für Nicht-R/3 Gesellschaften basierend auf der Business Server Page Technologie. Die aktuellen Techniken und Möglichkeiten des SAP Business Information Warehouse werden dabei voll ausgeschöpft. Mit dieser Technologie gestaltet sich der gesamte Integrationsprozess für Daten von Tochtergesellschaften ohne R/3-System deutlich einfacher, kostengünstiger und weniger fehleranfällig. Ziel des Projektes ist die mittelfristige Ablösung des traditionellen Verfahrens der Datenlieferungen von Nicht-R/3 Gesellschaften an das Business Information Warehouse. Diese sind klassifiziert durch einzelne Datenpakete, in denen die Zyklizität und Struktur festgelegt sind. Neben monatlich zu transferierenden Paketen für den Forecast gibt es für die verschiedensten Gruppen von Kennzahlen in den diversen Anwendungsbereichen u. a. die täglich zu übermittelnden Datenpakete der Vertriebsicht. Die größten einzelnen zu übermittelnden Datenpakete umfassen dabei einige hunderttausend Datensätze.

In Abschnitt 2 erfolgt zunächst eine Darstellung der grundsätzlichen Möglichkeiten des Datenimports in das SAP Business Information Warehouse über eine XML-Schnittstelle. Das über dieses Verfahren abzulösende traditionelle hostbasierte Verfahren, das sich bei Bayer MaterialScience im Einsatz befindet, ist Gegenstand des Abschnittes 3. Die Darstellung des realisierten neuen Verfahrens steht in Abschnitt 4 im Vordergrund. Der Artikel schließt mit einem Ausblick in Abschnitt 5.

2 XML-Datenimport im SAP Business Information Warehouse

Der wachsenden Bedeutung von XML als Technologie im Kontext von Data Warehouse Lösungen (vgl. Schwalm, Bange 2004) trägt auch die SAP AG mit der zunehmenden Integration von XML-basierten Techniken in das Business Information Warehouse, dessen allgemeine Architektur in Abschnitt 2.1 kurz beschrie-

ben wird, Rechnung. Zwei wesentliche Bereiche der Verwendung von XML-Techniken sind dabei zu nennen: Der Metadatenaustausch und Datenladevorgänge. Die Spezifikation der Metadaten basiert auf einer Erweiterung des Common Warehouse Metamodels und ermöglicht den Austausch von BW Metadaten über XML. Für den Einsatz im Bereich der ETL Prozesse können XML Dateien über einen SOAP-Anschluss direkt in das Business Information Warehouse geladen werden. Nach einer kurzen Darstellung der zugrunde liegenden Technologien in Abschnitt 2.2, befasst sich der Abschnitt 2.3 mit diesem Aspekt detaillierter.

2.1 Architektur des Business Information Warehouse

Die Architektur des Business Information Warehouse ist an die allgemeine Referenzarchitektur für ein Data Warehouse angelehnt. Ausgehend von verschiedenen möglichen Quellsystemen erfolgt ein im Allgemeinen periodischer Datenladevorgang in die zentrale Datenhaltung des Business Information Warehouse (BW). R/3-Systeme haben eine besondere Bedeutung als vorgelagertes OLTP-System, aber auch die Anbindung von R/2-, Nicht-SAP- und anderen BW-Systemen ist neben dem Import über flache Dateien möglich.

Eine grobe Übersicht der Architektur und des Zusammenspiels der beteiligten Komponenten des Business Information Warehouse ergibt sich aus der folgenden Abbildung 1 (vgl. Hahne 2003, S. 632).

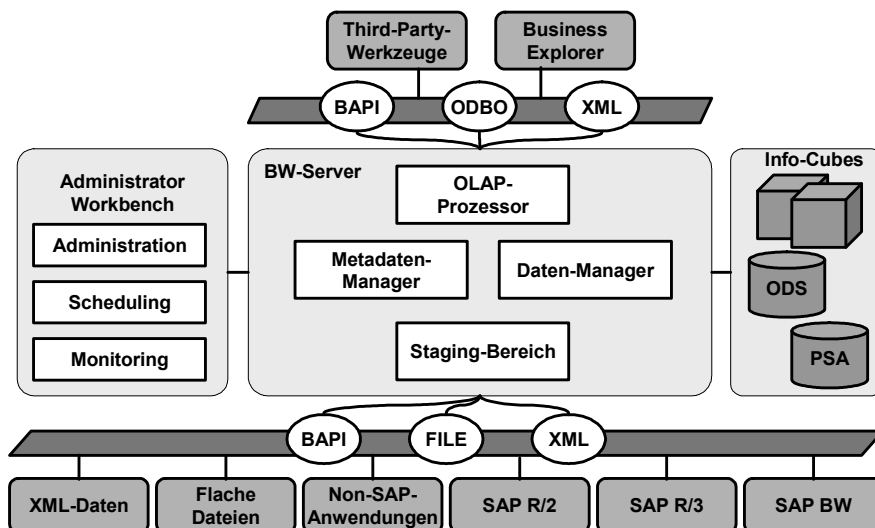


Abb. 1: Architektur des SAP Business Information Warehouse 3.x

Die Verwaltung des BW-Systems erfolgt in der Administrator Workbench, einer Administrationskonsole, in der die Steuerung, Überwachung und Pflege des gesamten Extraktions- und Lademanagements von Quelldaten, die Benutzerverwaltung sowie die Modellierung der Datenstrukturen im BW erfolgt. Im Mittelpunkt der Architektur steht der eigentliche Business Information Warehouse-Server, der für die Ablage, Aufbereitung und Abfrage der Daten im BW verantwortlich ist. Die wesentlichen Objekte zur Datenablage sind dem Datenfluss folgend zunächst die Persistent Staging Area (PSA), der Operational Data Store (ODS) und die Info-Cubes. Während die PSA temporären Charakter hat, sind die Daten im ODS und in den Info-Cubes für die dauerhafte Ablage gedacht. In den relational strukturierten ODS-Objekten erfolgt die Speicherung von belegnaher Information, die Info-Cubes bilden die mehrdimensionalen Strukturen als Grundlage für OLAP-Analysen ab.

Basis für Auswertungen und die Erstellung von Berichten mit verschiedenen Front-End-Werkzeugen sind die im Business Information Warehouse abgelegten Datenstrukturen. Mögliche Kategorien von Anwendungen basieren zum einen auf Werkzeugen zur Unterstützung des Managements, die von einfachen Berichts- und Abfragesystemen über leistungsstarke Werkzeuge zur Entscheidungsunterstützung bis zu den Führungsinformationssystemen reichen. Zum anderen handelt es sich um die unter den Stichworten Data Mining und OLAP diskutierten Anwendungsbereiche. Eine einheitliche Metadatenbasis ist für alle auf das Business Information Warehouse zugreifenden Front-End-Werkzeuge essentiell.

Integraler Bestandteil des Business Information Warehouse ist der Business Explorer Analyzer als Excel-basiertes OLAP-Werkzeug. Der Bereich des Webreportings wird durch die neue Technologie auf Basis des Web Application Servers und dem Web Application Designer adressiert.

Entlang des in Abbildung 2 zusammenfassend dargestellten Datenflusses ausgehend vom Quellsystem ist zunächst die Persistent Staging Area zu nennen. Dieser fungiert als Sammler der von Quellsystemen gelieferten nicht aufbereiteten und nicht modifizierten Daten. Der so genannte Operational Data Store ist im Kern eine relationale Speicherkomponente, der in einer Referenzarchitektur auf der Ebene der Integration und Transformation angesiedelt wird und dient in der Darstellung der Konsolidierung verschiedener Datenquellen. Im Business Warehouse sind diese Objekte zwar ebenfalls als Informationsspeicher für Auswertungen und Analysen optional verfügbar, hierfür sind aber die mehrdimensional strukturierten Info-Cubes dediziert konzipiert. Diese sind die wichtigsten Objekte für Analysen und Auswertungen. In einer weiteren Stufe erfolgt die betriebswirtschaftliche Konsolidierung mit dem Aufbau von Info-Cubes (vgl. hierzu Chamoni et. al. 2004, S. 50f.).

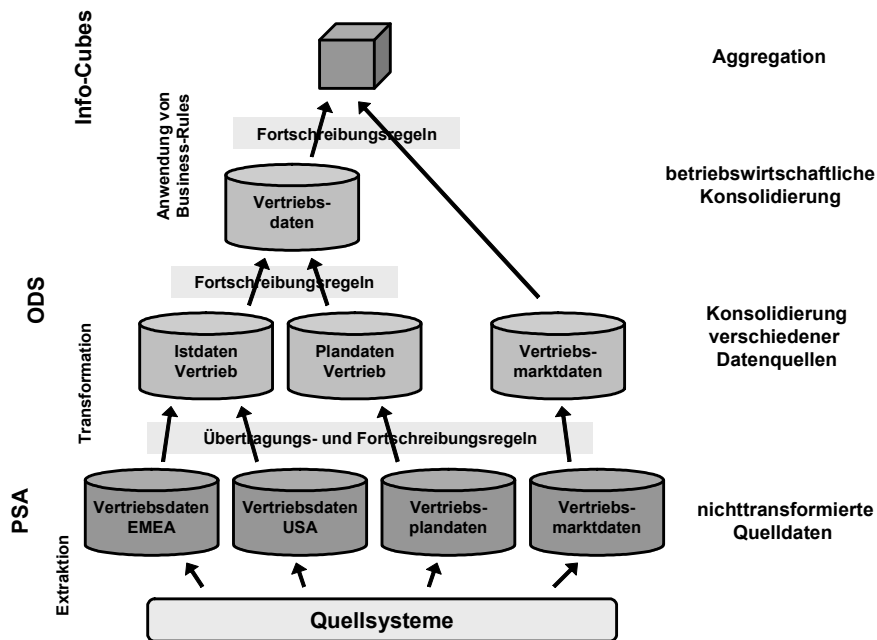


Abb. 2: Datenfluss und Integrationsarchitektur des SAP BW

Diese skizzierte Architektur hat einen gewissen Referenzcharakter, da dies eine optimale Ausgestaltung repräsentiert. In der Praxis sind natürlich verschiedenste Varianten denkbar, gerade in Fällen geringer Anforderungen an die Transformation ist etwa die Konsolidierungsebene der ODS-Objekte obsolet.

Die ODS-Objekte erfüllen im Business Information Warehouse die zentrale Funktion der Speicherung auch historischer transformierter und aufbereiteter Daten zur weiteren Verwendung in Info-Cubes. Diese können dann unabhängig vom gesamten Datenaktualisierungsprozess neuen Anforderungen flexibel angepasst und neu aufgebaut werden. Der Aufbau von Info-Cubes ist im Wesentlichen von den Berichts- und Analyseanforderungen determiniert. Die ODS-Objekte stellen die Basisdaten für verschiedene Szenarien und Reporting-Anforderung zu Verfügung.

2.2 XML, SOAP und Webservices

Die eXtensible Markup Language (XML) ist eine Auszeichnungssprache und optisch ähnlich der Hypertext Markup Language (HTML). Diese ist nur ein Teil der auch als „X-Technologie“ bezeichneten verschiedenen Spezifikationen, nutzt eine höhere Abstraktionsebene und ermöglicht die strikte Trennung von Inhalt, Logik und Präsentation. Die Inhalte werden in XML Dokumenten abgelegt. Deren

Validierung erfolgt durch Referenz auf (webbasierte) Document Type Definitions (DTD) bzw. XML Schemata (vgl. van der Vlist 2003). Mittels der eXtensible Stylesheet Language können XML-Dateien in verschiedene Zielformate transformiert werden (vgl. Floyd 2002). Dieser Zusammenhang ist in Abbildung 3 kurz zusammengefasst.

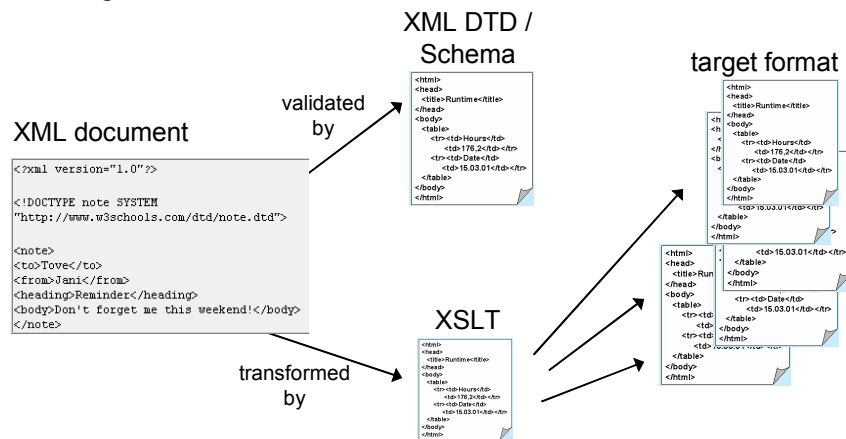


Abb. 3: Übersicht über die „X-Technologie“

Transaktionen im Internet wurden bisher überwiegend durch den User manuell im Browser angestoßen und durch proprietäre Programmierung realisiert. Webservices sind eine neue „Evolutionsstufe“ und ermöglichen nicht nur manuellen Informationsabruf, sondern bieten eine Standardschnittstelle für „Dienste“ im Internet. Die Webservice Schnittstelle ist mit Simple Object Access Protocol (SOAP) und XML standardisiert und wird somit auch für Anwendungen mit der Möglichkeit der Maschine-Maschine-Kommunikation nutzbar. XML und SOAP setzen auf dem weit verbreiteten HTTP Protokoll auf. Ein Beispiel einer SOAP-Nachricht ist in Abbildung 4 dargestellt.

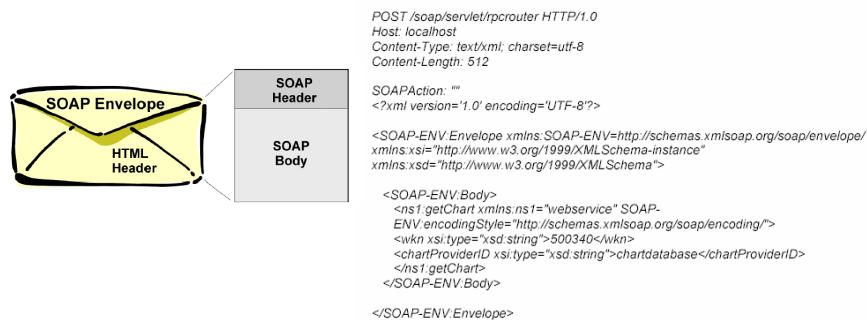


Abb. 4: Beispiel einer SOAP-Nachricht

Das SAP Business Information Warehouse nutzt diese Technologien, wie im nächsten Abschnitt beschrieben wird. Gerade die Webservices entwickeln sich zu einer zentralen strategischen Komponente der Produkte der SAP AG.

2.3 Bedeutung des Einsatzes von XML in ETL-Prozessen mit dem Business Information Warehouse

Seit dem BW Release 3.0 gibt es die Möglichkeit, Daten über die neue SOAP / XML Schnittstelle direkt in das BW zu laden (vgl. SAP 2001). Vorher konnte vom BW aus nur auf Daten aus direkt angeschlossenen Quellsystemen zugegriffen werden.

Die SOAP / XML Schnittstelle ist ein sogenannter Webservice. Dieser Webservice ist als Empfänger anzusehen, der ständig auf Datenanlieferungen wartet. Kommt auf dem entsprechend konfigurierten Empfangskanal ein Datenpaket an, für das eine Info-Source im BW definiert wurde, so gelangen die Daten in die Delta Queue des BW.

Die Delta Queue ist in diesem Zusammenhang als Datenzwischenspeicher anzusehen. Für das Business Information Warehouse verhält sich die Delta Queue so, als seien die dort geladenen Daten in einem angeschlossenen R/3 System. Aus der Delta Queue können die Daten vom BW abgeholt, verarbeitet und in die Datenziele fortgeschrieben werden, befinden sich also quasi schon im BW-System.

Der Datenimport per XML ermöglicht daher das aktive Laden von Daten direkt in das Business Information Warehouse. SOAP ist ein Protokoll, das mit XML realisiert ist und auf HTTP basiert. Für viele Quellsysteme ist damit eine Verbindung über HTTP direkt mit dem BW ohne Zwischensysteme denkbar. Folgende Abbildung 5 veranschaulicht die neue Importschnittstelle. Dabei befindet sich die Delta Queue auf dem Web Application Server (WAS), von wo die Daten über ein Info-Package (IP) in die Persistent Staging Area (PSA) gelangen. Die weiteren Schritte im Datenfluss führen über die ODS-Objekte in die Info-Cubes als zentrale persistente mehrdimensionale Datenspeicher.

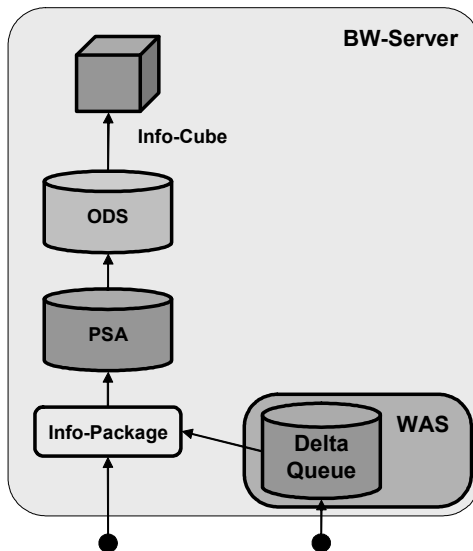


Abb. 5: Datenimport über den Web Application Server

XML ist nicht unbedingt das Format der Wahl zur Übertragung von Massendaten, denn es bläht das Datenvolumen durch die Auszeichnungselemente auf und die XML Daten müssen „eingepackt“, übertragen und wieder „ausgepackt“ werden (Performance-Aspekt). Diese Nachteile treten allerdings mit steigender Hardwareperformance und Hardwarekapazität in den Hintergrund. Allgemein ist XML auf dem besten Wege die Standardsprache für Datenaustausch und Schnittstellenformate in inhomogenen Systemlandschaften zu werden.

Auch bei der SAP AG und dem SAP BW sind XML Schnittstellen im Vormarsch, neben der oben bereits vorgestellten XML / SOAP Schnittstelle auch beim Import und Export von Datenmodellen bzw. Metadaten. SOAP ermöglicht die direkte Kommunikation zwischen Quell- und Zielsystem. Das Quellsystem kann über das weit verbreitete HTTP Protokoll aktiv Daten direkt in das BW schicken. Dies reduziert allgemein die Komplexität und die Anzahl der beteiligten (Schnittstellen-)Systeme und resultiert schließlich in besserer Wartbarkeit der Systeme und Kostensenkung für den Kunden.

Durch Verwendung des XML / SOAP Formats ist eine Validierung der Daten schon vor dem Abschicken möglich. Eine Prüfung findet dabei gegen eine im Web (oder auf dem Web Application Server) zentral zur Verfügung gestellte Schablone in Form eines XML Schemas statt. Geprüft werden kann auf eine gültige Struktur der Datei, gültige Formate und Typen der Inhalte sowie ggf. auch gegen Stammdaten. Diese Möglichkeit der Validierung verdeutlicht die folgende Abbildung 6.

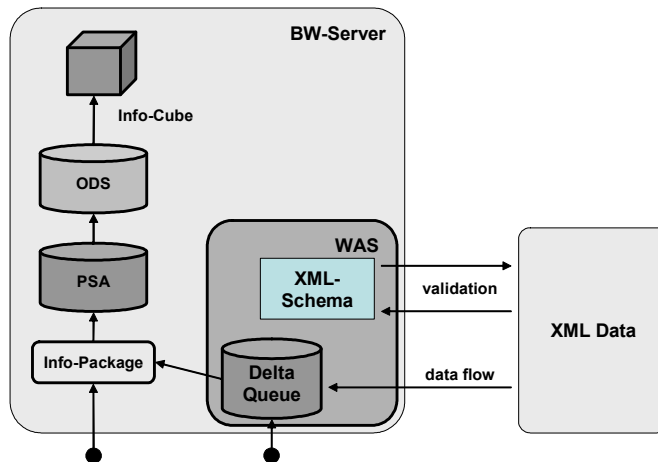


Abb. 6: Validierungsablauf beim XML basierten Datenupload

Hierbei wird das XML-Schema, gegen das die abzuschickenden Datenpakete geprüft werden können, vom BW Team erstellt und zentral gehostet. Das XML Schema braucht nicht an alle datenliefernden Organisationen verschickt zu werden, sondern steht zentral auf dem Server zur Verfügung.

3 Traditionelles Host basiertes Verfahren der Datenbereitstellung für SAP BW bei Bayer Material Science

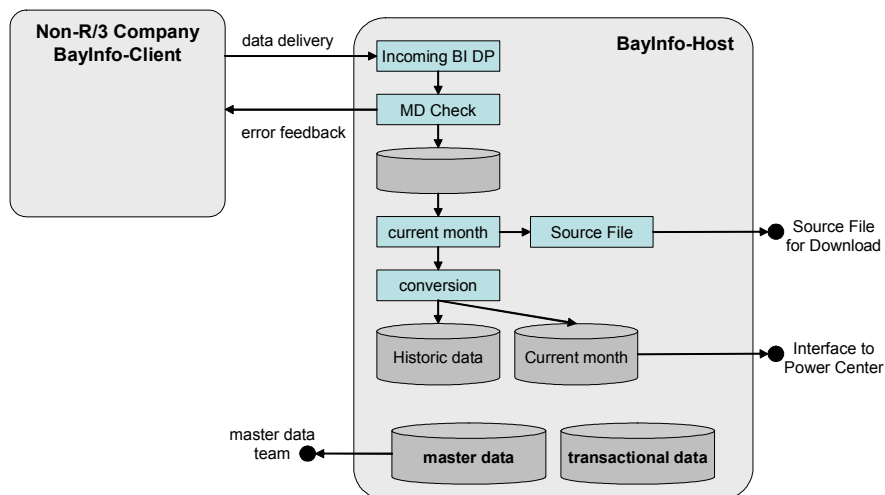
In einem großen Konzern wie Bayer Material Science mit seinem Geflecht an beteiligten Tochtergesellschaften und weltweit verteilten dezentralen Einheiten ist es nicht verwunderlich, eine Vielzahl an verschiedenen Systemen zur Abwicklung der Geschäftsprozesse vorzufinden. Trotz einem sehr weit verbreiteten Einsatz von SAP R/3 Systemen bei Bayer sind viele Teilgesellschaften mit proprietären Lösungen in ein Gesamtkonzept zur Datenintegration mit aufzunehmen. Eine wesentliche Aufgabe besteht dabei in der Zusammenführung verschiedener Datenlieferungen der einzelnen Gesellschaften für das zentrale Berichtswesen und die Gesamtplanung.

Im Bayer Konzern spielt das SAP Business Information Warehouse eine wesentliche Rolle zur Datenintegration im Rahmen einer unternehmensweiten Data Warehouse Strategie. Während die Bereitstellung von Daten aus SAP R/3 Systemen an

das SAP BW relativ problemlos implementiert ist, stellt die Zusammenführung der Daten proptiefärer Systeme eine besondere Herausforderung dar.

Das bisher eingesetzte Verfahren zur Integration von Nicht-R/3-Gesellschaften in den ETL-Prozess zur Versorgung des zentralen Data Warehouse ist hostbasiert. Die verschiedenen Quellsysteme erzeugen Datenpakete im CSV-Format, die allgemeinen Spezifikationen genügen – die Festlegungen für die so genannten Bay-Info Datenpakete (BI DP) – und im Rahmen eines periodischen Verfahrens an den Host zur weiteren Bearbeitung übergeben werden. Auf diesem sogenannten Bay-Info-Host erfolgt dann zunächst eine Prüfung auf vorhandene und korrekte Stammdaten (der MD Check), bevor eine weitere Bearbeitung möglich ist. Die sendenden Einheiten erhalten dabei eine Nachricht über evtl. auftretende Fehler.

Da einzelne Datenlieferungen jeweils eine einzelne Periode betreffen und auch kumulierte Werte beinhalten können, ist u. U. eine Phase der Berechnung anhand historischer Werte nachzuschalten. Diese sind zusammen mit den einzelnen angelieferten Datenpaketen auf dem Host für die spätere Verwendung und als Backup vorhanden. Eine Übersicht über diesen Schritt im ETL-Prozess ist in Abbildung 7 zusammengestellt.



stehen damit für das Berichtswesen und weitere Applikationen zur Verfügung. Dieser Verarbeitungsschritt ist in Abbildung 8 zusammenfassend dargestellt.

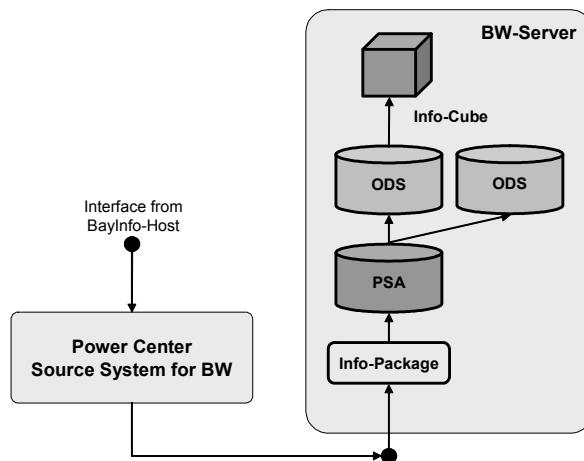


Abb. 8: Altes hostbasiertes Verfahren des Datenimports

Eine zentrale Zielsetzung bei Bayer Material Science besteht in der mittelfristigen Ablösung dieser Hostanwendung, welche die hiermit verbundenen Kosten reduziert. Neben einem möglichst umfassenden Einsatz von R/3-Systemen, wobei zu berücksichtigen ist, dass die Tochtergesellschaften einen hohen Grad an Autonomie aufweisen, sind für die proprietären Systeme spezielle Lösungen zu finden. Die Darstellung des hierzu konzipierten Verfahrens auf Basis des Business Information Warehouse ist Gegenstand des folgenden Abschnittes.

4 Neues Datenimportverfahren auf Basis von XML und SOAP

In der ersten Projektphase werden die auf dem Host bereits mit BayInfo validierten Datenpakete vom Host abgeholt und mit einem Konverter in XML umgewandelt. Der Upload dieser XML Files in die Delta Queue des BW erfolgt dabei zunächst manuell. Somit ändert sich für die Nicht-R/3 Gesellschaften in der ersten Phase Nichts. Ein entsprechendes XML Schema ist hierbei die Grundlage.

Weiterhin befindet sich auf dem BW System eine Business Server Page (BSP) Applikation, d. h. eine auf dem Web Application Server (WAS) des BW gehostete

Internetanwendung. Die auf dem WAS gehosteten Web Seiten können per JavaScript oder ABAP mittelbar auf sämtliche Inhalte, Metadaten und Funktionsbausteine des BW zugreifen. Diese Applikation ermöglicht ein Monitoring für die in das BW geladenen XML Datenpakete via Internet bzw. Intranet.

In einer frühen Pilotphase erfolgte zunächst nur die Anzeige von Informationen über geladene Requests. Später kam im BW weiterhin als Teil des Datenladeprozesses auch eine Stammdatensvalidierung hinzu. Ergebnisse dieser Validierung werden dann in Tabellen im BW abgelegt und als Protokoll in der Webapplikation zur Verfügung gestellt. Letztendlich ist auch das Freigeben und Sperren von bereits geladenen Datenpaketen webbasiert möglich. In der letzten Projektphase sollen die Nicht-R/3 Gesellschaften ihre Datenpakete direkt durch eine Maschine-Maschine Verbindung per HTTP abgeben. Kleinere Gesellschaften könnten ihre Datenpakete auch weiterhin per Uploadfunktion in der Webapplikation hochladen.

In den folgenden Abschnitten erfolgt zunächst die Verfahrensbeschreibung in der Übersicht bevor auf die detaillierte Darstellung des Versands der XML-Pakete und der Verwaltung der gelieferten Pakete in der BSP-Anwendung eingegangen wird. Die Darstellung des Datenflusses innerhalb des Business Information Warehouse schließt sich daran an.

4.1 Darstellung des Verfahrens in der Übersicht

Über den Java-Client validiert der Benutzer, sofern er hierzu berechtigt ist, ein abzulieferndes Datenpaket gegen die auf dem Web Application Server befindlichen Schemata und schickt dieses nach erfolgreicher Validierung zum BW-Server. Für die unterschiedlichen zu übertragenden Datenpakete kommen verschiedene Schemata zum Einsatz, die ihrerseits über eine Versionierung in verschiedenen Varianten vorgehalten werden. Dabei wird festgehalten, welche sendenden Einheiten mit welchen Schemaversionen arbeiten, um auch ggf. notwendige Anpassungen an nicht abwärtskompatible neue Schemata zu ermöglichen. In einer Business Server Page (BSP) Anwendung überprüft er die im BW angekommenen Daten in der Delta-Queue und startet von dort die weitere Datenübernahme ins BW oder die Rücknahme der Datenlieferung durch Löschen des Paketes.

Nach erfolgreichem Ladevorgang stehen die übernommenen Daten im ersten ODS-Objekt zur Verfügung und können nun gegen die Stammdaten im System geprüft werden. Dieser ebenfalls als Validierung bezeichnete Prozessschritt wird auch über die BSP initiiert. Evtl. Fehler der Stammdatenverprobung sind dann in einem Fehler-Log einsehbar und können von dort auch zwecks weiterer Verarbeitung in eine Excel oder CSV-Datei exportiert werden.

Die validierten Pakete können im weiteren Ablauf vom Benutzer in der BSP freigegeben werden. Nur die auf diese Art fachlich freigegebenen Pakete können mit dem nächsten Teilschritt im gesamten Ablauf in den Ladeprozess für die Info-

Cubes eingehen und stehen danach für das Berichtswesen zur Verfügung. Der Gesamtprozess ist in Abbildung 9 zusammengefasst.

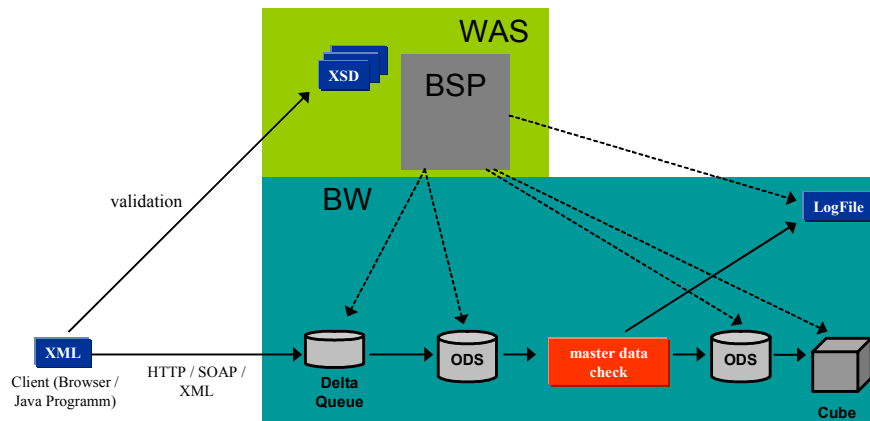


Abb. 9: XML-basiertes Verfahren des Datenimports

Der folgende Abschnitt befasst sich mit dem ersten Teilschritt im gesamten Ablauf, in dessen Fokus der Versand der Datenpakete liegt.

4.2 Versand von XML-Datenpaketen

Die datenliefernden Einheiten generieren aus ihren operativen Systemen in festgelegten periodischen Zyklen einen Datenextrakt, der an die zentrale Einheit zu berichten ist. Dieser Extrakt, deren Basis im Allgemeinen eine relationale Datenstruktur ist, besteht im alten Verfahren aus einer einfachen CSV-Datei. Nach dem neuen Verfahren ist aber eine XML-Datei notwendig. Nun wird teilweise bereits beim Generieren der Extraktdatei das gewünschte XML-Format generiert, teilweise aber auch nur die erzeugte flache Datei anschließend konvertiert.

Der Prozess der Datenübergabe an das Business Information Warehouse beginnt für den Benutzer im Aufruf des JavaClients der XML-Anwendung, in der nach der erfolgreichen Authentifizierung gegenüber dem BW-Server zunächst die Auswahl der zu übermittelnden XML-Datei erfolgt. In einem zweiten Schritt erfolgt die Validierung der Datei gegen die festgelegten XML Schemata auf dem zentralen Server. Dabei erfragt der Client über einen Webservice auf dem empfangenden BW-System die dort als gültig für diese sendende Einheit vorliegenden Schemata und vergleicht diese mit den in der XML-Datei benutzten. Nur im Falle der erfolgreichen Überprüfung ist es möglich, den dritten Prozessschritt der Datenübermittlung zu initiieren. Diese Abfolge ist auch an der Reihenfolge der Schaltflächen wie in Abbildung 10 dargestellt erkennbar.

Eine wesentliche Komponente im gesamten Datentransferkonzept ist die automatische Übergabe eines eindeutigen Zeitstempels beim Versand. Dieser wird von der Applikation in die XML-Datei beim Versand eingefügt. Hierzu wird quasi der SOAP-Umschlag geöffnet, die Eintragung des Zeitstempels in einem entsprechenden Tag vorgenommen und dann der SOAP-Umschlag wieder verschlossen. Die Anwendung erhält den Zeitstempel von einem zentralen Webservice auf dem BW Server, der diesen Sekundengenau erstellt.

Alle einzelnen im JavaClient ausgeführten Schritte sind im Nachhinein von Administratoren nachvollziehbar, da die Applikation diese Informationen über einen Webservice in ein Meldungs-Log auf dem BW-System schreibt.

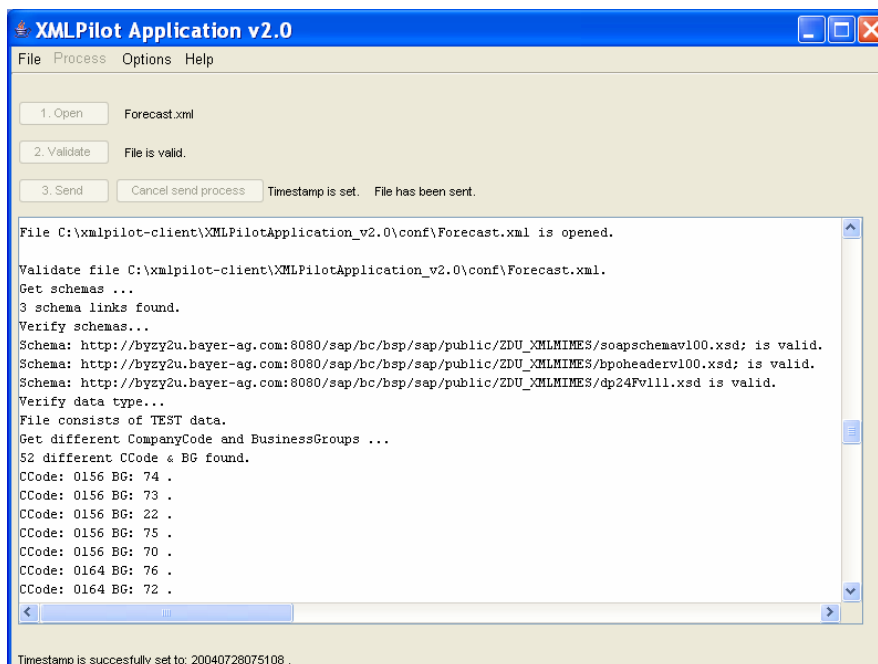


Abb. 10: Validierung und Versand der XML-Dateien über den Java-Client

In Lasttests bewährte sich dieses Verfahren bis zu einer Größenordnung von mehreren hunderttausend Items in einer zu übertragenden XML-Datei. Dies korrespondiert mit dem zu erwartenden Maximalwert in der konkreten Anwendungssituation bei Bayer MaterialScience. Zukünftig wird verstärkt statt der manuellen Datenübermittlung per Java Client auf die direkte Systemkopplung über HTTP gesetzt. Hierdurch verschlankt sich das Verfahren nochmals und die Grenzen des zu übertragenden Datengerüsts werden signifikant nach oben verschoben.

4.3 Verwaltung der per XML gelieferten Datenpakete in einer Web-Applikation

Die gesamte Service-Applikation ist auf der Grundlage der Business Server Page Technologie entwickelt und basiert auf einem Framework zur Gestaltung und Navigation. Der Bildschirmaufbau ist dreigeteilt. Im Kopfbereich findet sich neben den gestalterischen Elementen der Logos die Navigation für die Applikation. Im Fußbereich befindet sich die Statuszeile mit Informationen zum angemeldeten Benutzer und dem System.

In der ersten Übersichtsseite der Anwendung, wie exemplarisch in Abbildung 11 dargestellt, findet sich eine Übersicht über alle Interfaces, die Daten per XML empfangen können, sowie deren genaue Liste der zu übermittelnden Felder.

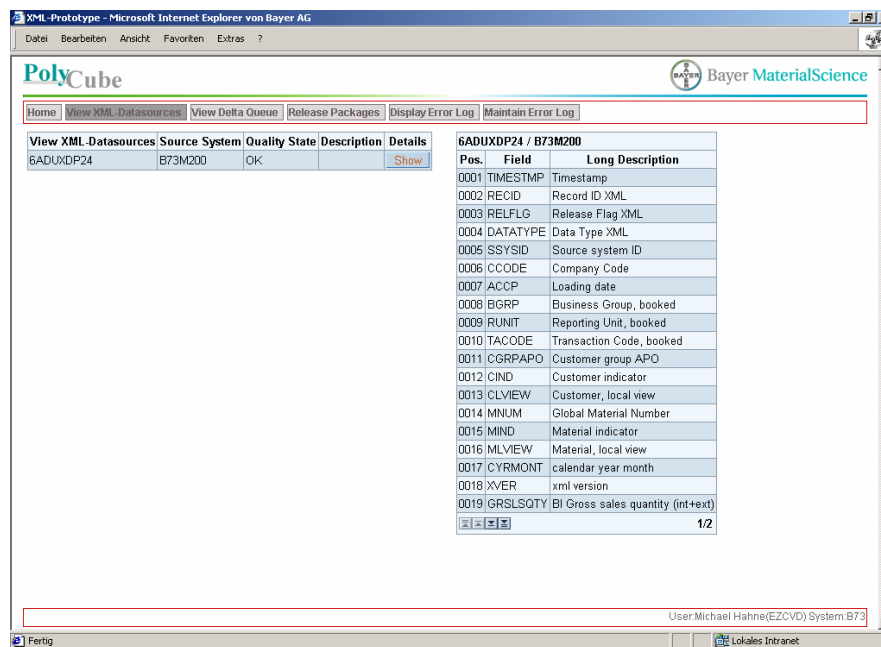


Abb. 11: Übersicht über die Feldstruktur einer XML Datenquelle

Die Feldnamen in der Detailsicht zu einer spezifischen gewählten XML Data Source korrespondieren mit den Namen der Tags in der XML-Datei. In der Kopfzeile der Anwendung sind die einzelnen Module wählbar. Neben der dargestellten Data-Source Ansicht ist der Aufruf des Moduls für die Darstellung des Inhaltes

einer Delta Queue, der Freigabe und Validierung von Datenpaketen und für die Log-Analysen möglich.

In der jetzigen Ausbaustufe der Anwendung (Stand Mai 2004) ist lediglich eine Data-Source produktiv geschaltet, so dass unter dem nächsten Menüpunkt zur Delta Queue Anzeige nur diese eine Delta Queue ausgewertet wird. Hierzu sind in Abbildung 12 insbesondere die technischen Felder des Zeitstempels (Timestamp) und XML spezifische Felder erkennbar.

Timestamp	Record ID XML	Release Flag XML	Data Type XML	Source system ID	Company Code	Loading date	Business Group	booked	Reporting L
20040317170535	0000000001	R	P	A6	0005	20030202	20		070
20040317170535	0000000002	R	P	A6	0005	20030202	20		070
20040317170535	0000000003	R	P	A6	0005	20030202	20		070
20040317170535	0000000004	R	P	A6	0005	20030202	20		070
20040317170535	0000000005	R	P	A6	0005	20030202	20		070
20040317170535	0000000006	R	P	A6	0005	20030202	20		070
20040317170535	0000000007	R	P	A6	0005	20030202	20		070
20040317170535	0000000008	R	P	A6	0005	20030202	20		070
20040317170535	0000000009	R	P	A6	0005	20030202	20		070
20040317170535	0000000010	R	P	A6	0005	20030202	20		070
20040317170535	0000000011	R	P	A6	0005	20030202	20		070
20040317170535	0000000012	R	P	A6	0005	20030202	20		070
20040317170535	0000000013	R	P	A6	0005	20030202	20		070
20040317170535	0000000014	R	P	A6	0005	20030202	20		070
20040317170535	0000000015	R	P	A6	0005	20030202	20		070
20040317170535	0000000016	R	P	A6	0005	20030202	20		070
20040317170535	0000000017	R	P	A6	0005	20030202	20		070

Abb. 12: Übersicht gelieferter Datensätze in der Queue-Anzeige

Durch Interaktion des Benutzers erfolgt die weitere Bearbeitung der Daten in der Delta-Queue. Diese stehen danach in einer paketweisen Darstellung zur weiteren Analyse zur Verfügung (siehe Abbildung 13). Zunächst erfolgt die Validierung einzelner Datenpakete. Bei auftretenden Fehlern kann von dieser Stelle aus direkt in die Anzeige der Fehlermeldungen abgesprungen werden.

The screenshot shows a web browser window titled 'XML-Prototype - Microsoft Internet Explorer von Bayer AG'. The page header includes the 'PolyCube' logo and 'Bayer MaterialScience'. Below the header is a navigation bar with links: Home, View XML-Datasources, View Delta Queue, Release Packages, Display Error Log, and Maintain Error Log. There are two control sections: 'Set Release Flag to:' with buttons for Preliminary, Released, and Invalid; and 'Set Package Status:' with buttons for Validate Package and Refresh Table. The main content is a table with the following data:

	Timestamp	Company Code	Source System	XML Data Package Status	Loading Date	Release Flag	Error Log
C	20040317170535	0005	A6	Delivered	02.02.2003	Delivered	Show
C	20040317170543	0005	A6	Delivered	02.02.2003	Delivered	Show
C	20040317170554	0005	A6	Delivered	02.02.2003	Delivered	Show

At the bottom right of the table, there is a page indicator '1/1'.

Abb. 13: Geladene Datenpakete in der Übersicht

Nach erfolgreicher Validierung sind die Datenpakete vom Benutzer freizugeben, damit eine weitere Bearbeitung erfolgen kann. Die freigegebenen Datenpakete stehen nach Aufruf der Laderoutine in den Info-Cubes für das Berichtswesen zur Verfügung.

Der zur Abbildung dieses Prozesses benötigte Datenfluss ist Gegenstand der Darstellungen im nachfolgenden Abschnitt.

4.4 Implementierung des Datenflusses im BW

In der hier beschriebenen Version der Anwendung ist nur eine Data-Source berücksichtigt. Für diese steht auf dem BW-Server eine Info-Source-Definition und ein darauf basierender SOAP-Anschluss zur Verfügung, der auf ankommende Datenlieferungen wartet und diese in einer Delta-Queue speichert. Dies ist in Abbildung 14 dargestellt. Für alle zum Transfer zugelassenen Datenpakete sind bei Bayer MaterialScience eigene SOAP-Anschlüsse definiert, die entsprechende Daten empfangen können. Diese unterscheiden sich hinsichtlich der Datenstruktur und der Zyklizität der übertragenen Daten.

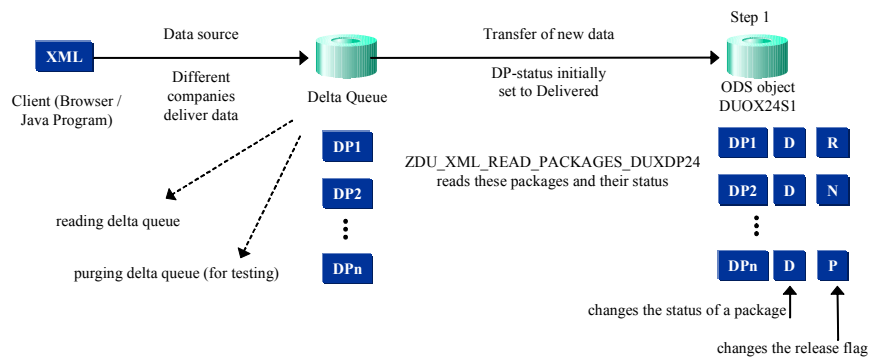


Abb. 14: Anlieferung von XML-Datenpaketen

Der Versand der XML-Dateien kann per JavaClient oder über eine spezielle Upload-Seite auf dem Web Application Server erfolgen. Die einzelnen Datenpakete in der Delta-Queue werden dann an das ODS-Objekt weitergereicht, auf dem die Verwaltung der Datenpakete erfolgt. Dieses dient als Speicher aller neuen Datenpakete, die noch weiter zu bearbeiten sind. In einem ersten Schritt erfolgt die Validierung der Datenpakete, d. h. die Stammdatenverprobung (siehe Abbildung 15). Die Prüfung erfolgt dabei u. a. dahingehend, ob die verwendeten Stammdaten im System vorhanden sind. Ist dies nicht der Fall, erfolgt ein Eintrag mit den Fehlerwerten in die Log-Datenbank und das Datenpaket wird als nicht-valide gekennzeichnet.

In einer späteren Version der Anwendung erfolgt der Upload der Daten im XML-Format nicht mehr manuell auf Basis eines Clients, sondern über eine direkte Maschine-Maschine-Schnittstelle auf der Grundlage einer HTTP-Verbindung. Dies ermöglicht dann einen noch mehr automatisierten Datentransfer.

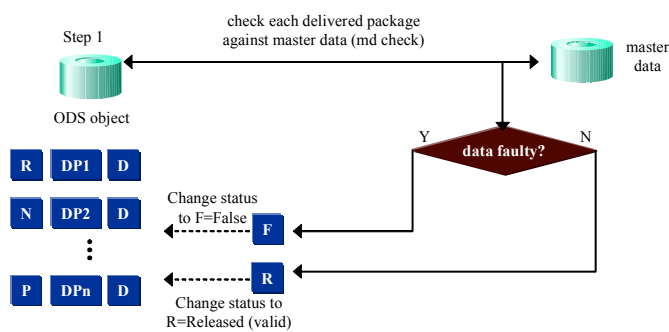


Abb. 15: Überprüfung und Freigabe gelieferter XML-Datenpakete

Nach dem Prozess der Stammdatenvvalidierung hat jedes Datenpaket im Step 1 ODS-Objekt den definierten Status geladen (L), fehlerhaft (F) oder valide (R) und die fehlerhaften Pakete werden zurückgewiesen und sind erneut anzuliefern. Die Weitergabe der freigegebenen wie auch der bereits geladenen Pakete erfolgt dann in einem weiteren Schritt (siehe Abbildung 16).

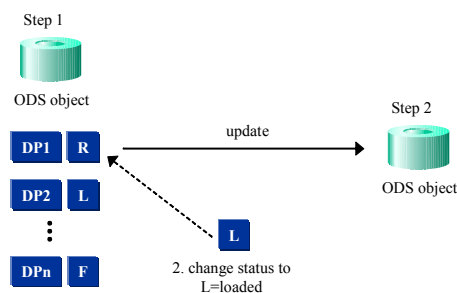


Abb. 16: Weitergabe der XML-Datenpakete für das Reporting

Im Step 2 ODS-Objekt erfolgt die Zusammenfassung der gelieferten Datenpakete und die paketspezifischen Informationen fallen weg. Es ist der „Sammelbehälter“ aller gelieferten, transformierten und zusammengefassten Datenpakete. Von hier aus erfolgt die Weitergabe an die Info-Cubes auf deren Basis die Auswertungen stattfinden.

Sind für die angelieferten Daten weitere Transformationsschritte notwendig, bevor diese in die Objekte für das Reporting überführt werden können, so ist u. U. eine weitere Zwischenschicht in Form eines ODS-Objektes einzuziehen. Dies ist bei einigen Datenflüssen, die kumulierte Daten anliefern, der Fall. Diese Daten sind für das Business Information Warehouse in die tatsächlichen Periodenwerte zu überführen, es findet eine so genannte Dekumulierung statt.

5 Ausblick

Mit dem realisierten Verfahren der Datenanlieferung per XML über einen Webservice hat Bayer MaterialScience einen Weg eingeschlagen, der den bisherigen Prozess des Ladens von Daten von Nicht-R/3-Gesellschaften deutlich schlanker macht. Die XML-Technologie ermöglicht es den dezentralen Organisationen, ihre Datenlieferungen schon vor dem Versand auf strukturelle wie auch partielle inhaltliche Konformität zu überprüfen, da eine Übertragung nur nach erfolgreicher Validierung gegen die zentral gepflegten XML Schemata möglich ist.

In der jetzigen Version der Anwendung verbleibt die Stammdatenverprobung allerdings auf dem BW-System, also nach dem Empfang der Datenpakete. In einer zukünftigen Version soll auch die Stammdatenvalidierung über einen Web-Service bereits vor dem Versand erfolgen, so dass im Allgemeinen nur noch gültige Pakete an das BW geschickt werden und der Validierungs- und Freigabeprozess noch weiter von manuellen Schritten befreit werden kann.

Bereits die implementierte Version macht die alte hostbasierte Anwendung für die betroffenen Datenpakete obsolet und mit einem Einsatz der XML-Technologie für die Übertragung aller Datenpakete kann die Hostanwendung, die im PowerCenter die Daten dem BW zur Verfügung stellt, ganz wegfallen. Der gesamte Prozess ist dann schlanker und kostengünstiger. Des Weiteren hat sich herausgestellt, dass die zu übertragenden Datenmengen mit dem Verfahren effizient handhabbar sind.

6 Literatur

Chamoni, P., Gluchowski, P., Hahne, M.: SAP Business Information Warehouse, Perspektiven betrieblicher Informationsversorgung und Entscheidungsunterstützung, Springer-Verlag, 2004.

Floyd, M. et. al.: Special Edition Using XSLT, Que Corporation, 2002.

Hahne, M.: Logische Datenmodellierung zur Abbildung mehrdimensionaler Datenstrukturen im SAP Business Information Warehouse, in: Weikum, Gerhard; Schöning, Harald; Rahm, Erhard (Hrsg.): BTW 2003 Datenbanksysteme für Business, Technologie und Web, Köllen Druck + Verlag, Bonn, 2003, S. 630-647.

SAP: How to... Send XML Data to BW, ASAP for BW Accelerator, 2001.

Schwalm, S., Bange, C.: Einsatzpotenziale von XML in Business-Intelligence-Systemen; Wirtschaftsinformatik 46 (2004) 1, S. 5-14.

Van der Vlist, E.: XML Schema, O'Reilly, 2003.