

Das Data Warehouse als Business-Integrations-Plattform

Alfred Schlaucher, ORACLE Deutschland B.V. & Co. KG

Trotz des Wunsches vieler Firmen nach Standardisierung wächst die Komplexität der IT-Systeme kontinuierlich. Es kommt zwar immer mehr Standardsoftware zum Einsatz und der Anteil der Eigenprogrammierung sinkt, aber die vielen Standardpakete sollten in einer vernetzten Unternehmens-Sys-temlandschaft ihre Daten austauschen und sich gegenseitig verstehen können. Damit wachsen die Herausforderungen, denn mit zusätzlichen Anwendungen potenziert sich die mögliche Schnittstellenanzahl und zwangsläufig die Komplexität.

In folgendem Beispiel haben fünf Anwendungen zehn potenzielle Schnittstellen, wenn man eine Schnittstelle bidirektional definiert (siehe Tabelle 1).

	A	B	C	D	E
A	x	1	2	3	4
B	1	x	5	6	7
C	2	5	x	8	9
D	3	6	8	x	10
E	4	7	9	10	x

Tabelle 1

Kommt eine sechste Anwendung hinzu, springt die Zahl der Schnittstellen von zehn auf fünfzehn (siehe Tabelle 2).

	A	B	C	D	E	F
A	x	1	2	3	4	5
B	1	x	6	7	8	9
C	2	6	x	10	11	12
D	3	7	10	x	13	14
E	4	8	11	13	x	15
F	5	9	12	14	15	x

Tabelle 2

Nicht alle Anwendungen fordern eine Verbindung zu allen anderen, aber ein Trend dorthin ist zu erkennen und bei mittelgroßen Unternehmen bestehen in der Regel

mehr als hundert Anwendungen, von denen ein großer Teil miteinander kommuniziert. Konzepte und Technologien zur Anwendungsintegration und zur Beherrschung dieser Komplexität gibt es seit Längerem. Allen voran regeln SOA-Konzepte den Datenaustausch zwischen Anwendungen mittels standardisierter Nachrichten, die man über einen generischen Bus verschickt (siehe Abbildung 1).

Integration ist mehr als Daten von A nach B schicken

Bei genauer Betrachtung ist die Kommunikation per Bus jedoch lediglich eine intelligente Schnittstellensteuerung, bei der man sehr flexibel Daten mechanisch kopiert. Über die Qualität der Daten, über ihren fachlichen Hintergrund und interpretierbare Bedeutung sagt das Verfahren nichts. Die Problematik der geschilderten Komplexität liegt jedoch in der unterschiedlich verstehbaren, semantischen Bedeutung der ausgetauschten Daten: Unterschiedliche Anwendungen nutzen Daten unterschiedlich. Missverständnisse sind unvermeidbar. Allerdings bewegen wir uns hier in einem rein operativen Umfeld, mit dem einen wichtigen Ziel, Transaktionen schnell abzuwickeln.

Bestehen weiterführende Anforderungen wie nach einer Interpretation oder Analyse der Daten oder nach taktischen und strategischen Auswertungen, so können Data-Warehouse-Systeme neben ihrer Integrations-Funktion auch zur Verbesserung der Semantik der auszutauschenden Daten beitragen.

In einem Data Warehouse legen Firmen mehr oder weniger langlebige Informationen ab, um sie entweder über Berichte auszuwerten, oder sie wieder anderen operativen Anwendungen zur Verfügung zu stellen. Data-Warehouse-Systemen übernehmen damit einen Teil des Datenaustauschs auch für operative Systeme.

Anmerkung: An dieser Stelle sind Data-Warehouse-Systeme nur ergänzend und nicht alternativ zu SOA-Konzepten beschrieben. Sie können und sollen auch nicht EAI-Systeme ersetzen, da die Konzepte und die Ziele andere sind. Die Anwendungsgebiete überschneiden sich jedoch, und so kann man je nach Fall ein Data-Warehouse-System auch für den operativen Datenaustausch nutzen. Es wird damit zur Schnittstellen-Plattform.

Mehr als drei Viertel aller Oracle-basierten Data Warehouses stehen heute auch im

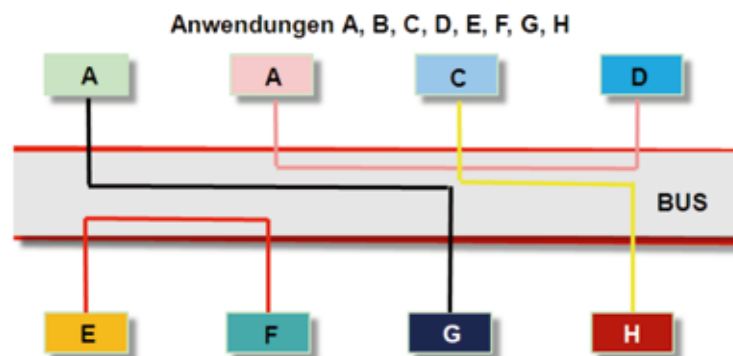


Abbildung 1: Kommunikation per Bus

operativen Einsatz. Sie liefern aufbereitete Daten nicht nur für Business-Intelligence-Anwendungen, sondern versorgen auch operative Online-Systeme. Anders als bei den oben genannten Bus-Systemen überführt der Integrationsprozess in einem Data Warehouse die Daten zunächst in eine anwendungsneutrale Form. Das schafft die Grundlage für eine flexible Weiterverarbeitung. Allerdings kann der hier nötige Prüf- und Transformationsprozess die Zeit zur Bereitstellung der integrierten Daten verlängern. Sekundenschnelle Durchlaufzeiten sind nicht zu erwarten, dafür aber stimmige, geprüfte, ergänzte und verständliche Daten.

Ein weiterer Unterschied zu Bus-Systemen ist die Persistenz der auszutauschenden Daten. Informationen bleiben für einen definierten Zeitraum in einer Datenbank gespeichert. Während sich in operativen Systemen die Daten in der Regel permanent verändern, lässt sich die Aktualisierungssequenz in einem Data Warehouse nach Bedarf festlegen – ob im Sekunden-, Ta-

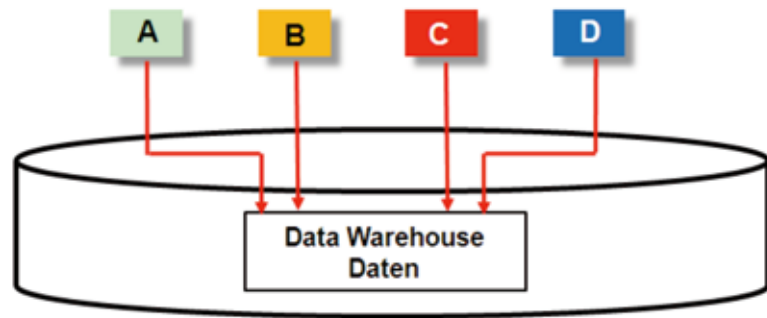


Abbildung 2: Kommunikation per Data Warehouse

ges-, Wochen- oder Monatsabstand (siehe Abbildung 2). Das ermöglicht zeitbezogene Vergleiche und lässt Trends erkennen. Daten können vervollständigt und verglichen werden. Auch aufwändige Plausibilitätsprüfungen sind machbar.

Aufbau von geprüften und integrierten Datenbeständen

Die besondere Aufbereitung von Daten in einem Data Warehouse hat jedoch seinen Preis. Die Daten durchlaufen mehrere Sta-

dien, bis sie sich in einer verwendungsfähigen Form befinden. Das Kopieren in ein Data Warehouse ist nur ein erster Schritt. Danach folgen regelgesteuerte Prüfverfahren, das Granularisieren (Zerlegen der Daten in Einzelinformationen, „Normalisierung“) und schließlich das Anreichern mit Ergänzungen aus Referenzdaten (Look-ups). In einem letzten Schritt fügt man die Daten unter neuen Kriterien dann meist wieder zu neuen Informationseinheiten zusammen (siehe Abbildung 3).



Vorbeugen ist besser als notretten: Remote Administration Service (RAS) von Hunkler

ORACLE Platinum Partner

- Optimale Konfiguration Ihrer Oracle-Datenbanken
- Fernüberwachung der Performance
- Früherkennung und Behebung von Fehlerquellen und Systemstörungen
- Zugang über VPN
- Telefonischer Support
- Individuell gestalteter Leistungsumfang

Best Solutions Based on Oracle
HUNKLER
GmbH & Co. KG

Hauptsitz Karlsruhe
Geschäftsstelle Bodensee

Bannwaldallee 32
Fritz-Reichle-Ring 2

76185 Karlsruhe
78315 Radolfzell

Tel. 0721-490 16-0
Tel. 07732-939 14-00

Fax 0721-490 16-29
Fax 07732-939 14-04

info@hunkler.de
www.hunkler.de

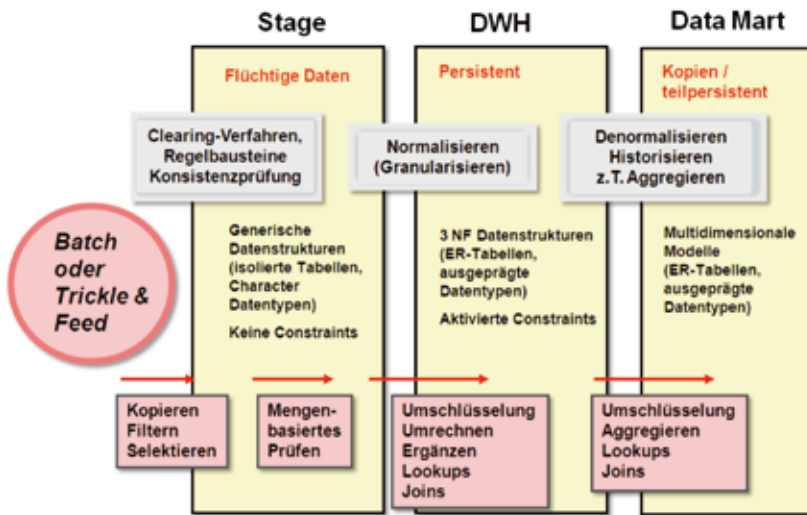


Abbildung 3: Aufbereiten der Daten

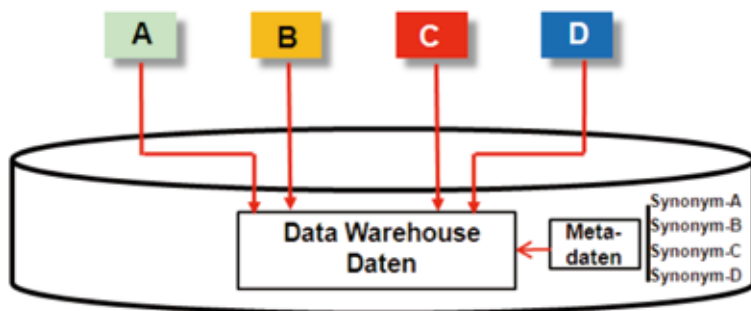


Abbildung 4: Der Einsatz von Metadaten

	Golden Gate	ODI	OWB
System-Modellierung	-	-	++
Heterogene Quellen	++	++	+
Heterogene Ziele	++	++	0
Deltadaten	++	+	+
Lade Performance	+	+	++
Wiederverwendung DB-Skills	-	+	++
Visualisierung Datenfluss	-	+	++
Metadaten	-	0	+

Abbildung 5: Die Tools im Vergleich:

++ sehr gut 0 ausreichend
+ gut - nicht relevant

Semantischer Abgleich über Metadaten

Data-Warehouse-Daten sind einmalige Daten, das heißt sie sind nur einmal vorhanden. Sie können jedoch in den diversen operativen Anwendungen auseinanderdriftende Ausprägungen besitzen. Unterschied-

de können in der Benennung, in Maßeinheiten oder in der Art von Berechnungen bestehen, es kann auch Zeitabhängigkeiten geben und Ähnliches.

Metadaten zu den einzelnen Daten beschreiben diese unterschiedlichen Ausprägungen. Diese Vorgehensweise ist umso wichtiger, je mehr Daten und Systeme betroffen sind. Die Metadaten erhalten sogenannte „Synonym-Attribute“ für die jeweiligen operativen Anwendungen, in denen die Daten Verwendung finden (siehe Abbildung 4).

Die technische Umsetzung

Zur Realisierung des genannten Szenarios sind drei Komponenten erforderlich:

- Eine Datenbank zur persistenten Datenvorhaltung. Hier bietet sich Oracle 11g R2 an.
- Ein Repository für die Metadaten. Hier gibt es eine Reihe von funktionsträchtigen Werkzeugen am Markt, die nicht im-

mer passen, aber oft sehr viel Geld kosten. Es stellt sich die Frage, ob man so viele Funktionen benötigt oder ob eine einfache Lösung ausreicht. Der Autor verschickt auf Anfrage an alfred.schlaucher@oracle.com eine Dokumentation über eine sehr einfache, aber ausreichende und kostenneutrale Lösung zum Selbermachen.

- Tools zum Extrahieren, Transportieren und Transformieren der Daten. Hier bietet Oracle gleich drei Werkzeuge an, die nachfolgend näher betrachtet werden (siehe Abbildung 5).

Neben der Eigenentwicklung „Oracle Warehouse Builder“ (OWB) hat sich Oracle in den vergangenen Jahren durch Zukäufe in diesem Segment verstärkt. Die Lösung der jüngsten Akquisition „GoldenGate“ ist besonders gut geeignet, wenn operative Systeme so schonend wie möglich „angepapft“ werden sollen. Das Tool liest die Log-Files der jeweiligen Datenbanken aller verbreiteten Hersteller, ohne die Leistungs-Ressourcen des operativen Systems zu belasten.

Oracle Data Integrator (ODI) hat seine Stärken in dem Bereitstellen von Agenten auf unterschiedlichen Plattformen. Über diese Agenten kann man die Steuerkontrolle zum Datentransport flexibel über mehrere Rechner-Knoten asynchron und asymmetrisch verteilen.

Sind Daten einmal in der Oracle-Datenbank angekommen, kann man mit OWB komplexe Transformationen Daten-nah in der Datenbank durchführen und die Daten „veredeln“. Mit OWB lässt sich das System gut modellieren und eine passende Metadaten-Auswertung aufbauen.

Fazit

Alle Werkzeuge sind natürlich in Kombination miteinander einsetzbar. So nutzt das jüngste OWB-Release (11.2) bereits die Agenten-Technik von ODI, und ODI kann Komponenten von GoldenGate einbinden. Mittelfristig hat Oracle die Verschmelzung aller Komponenten in eine neue Data Integration Suite angekündigt, die ab einer Version 12 verfügbar sein soll.

Kontakt:

alfred.schlaucher@oracle.com