

Die Data Warehousing Kolumne

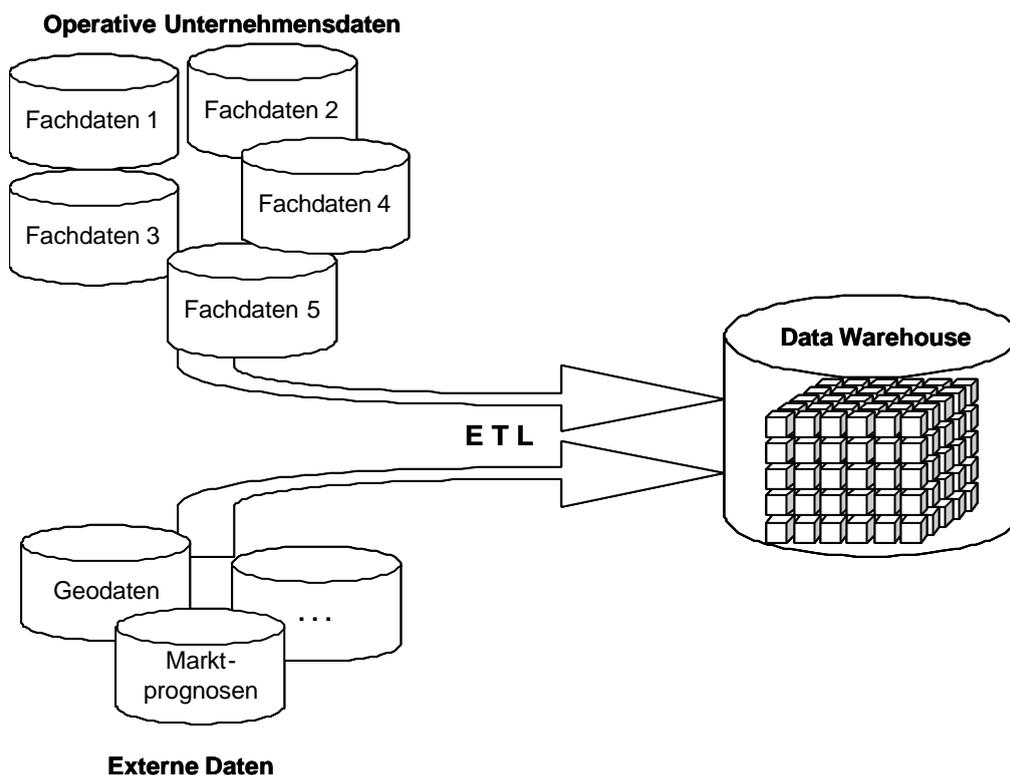
Diese Kolumne stellt regelmäßig unterschiedliche Aspekte aus dem Bereich Data Warehousing zur Diskussion. Beiträge bitte direkt an die Betreuer dieser Kolumne: Manfred Jeusfeld oder Heinrich Jasper (Adressen siehe Umschlaginnenseite).

Der folgende Beitrag befasst sich mit Trends bei der Anwendung von Data Warehousing, die der Autor während seiner Tätigkeit als Berater eines großen internationalen Medienunternehmens beobachten konnte. Anschließend gibt es eine kurze Zusammenfassung von einem Workshop zum Thema „Design und Management von Data Warehouses“.

Data Warehousing: Anwendungstrends in der Praxis

Das Ziel des Einsatzes eines Data Warehouses ist die integrierte Bereitstellung von Informationen in einem Unternehmen. Üblicherweise stehen diese Daten in den operativen IT-Systemen zur Verfügung, die tagesaktuelle Informationen über angefallene Geschäftsprozesse verwalten. Da in der Regel diverse dieser sogenannten Quellsysteme existieren, sind die Daten durch entsprechende Verfahren zu extrahieren, zu konsolidieren, zu transformieren und in ein Data Warehouse zu integrieren. Dieses gilt im übrigen auch (noch) beim Einsatz von Standardsoftware, da die vollständig integrierte Behandlung von Informationen bisher von keinem System in ausreichendem Maße geboten wird.

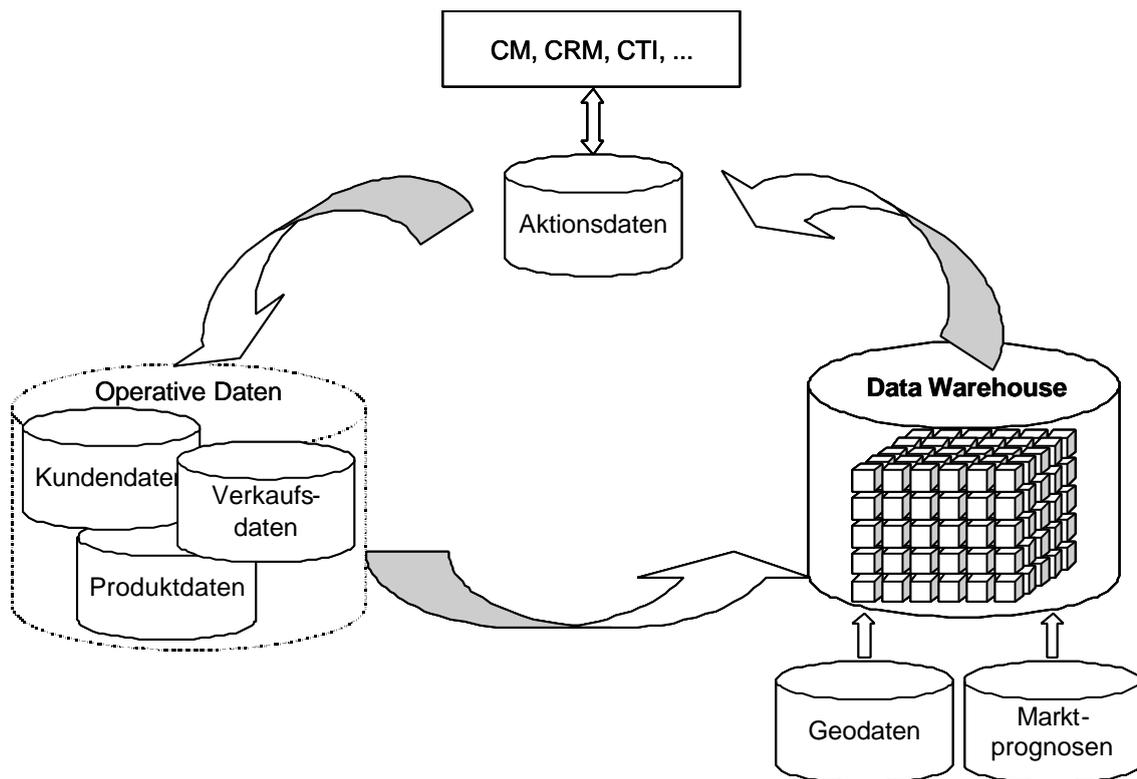
Werkzeuge zur Unterstützung des Integrationsprozesses – sogenannte ETL Tools (Extraction, Transformation, Loading) – werden in unterschiedlichen Ausprägungen am Markt angeboten. Vor der Anwendung eines ETL-Tools ist in jedem Fall die exakte Analyse und Definition der in das Data Warehouse zu integrierenden Daten notwendig. Schließlich ist die Integration von Daten in einem Data Warehouse kein Selbstzweck, sondern dient der Unterstützung von geschäftsrelevanten Tätigkeiten im jeweiligen Unternehmen.



Wesentliche inhaltliche Zielsetzung des Data Warehousing war und ist in den meisten Fällen die Bereitstellung unternehmensweit konsolidierter Daten für das interne Berichtswesen (BDR: Business Decision Reporting) und der online Zugriff auf diese Daten für strategische Analysen (OLAP: Online Analytical Processing). In diesem Umfeld hat sich Data Warehousing etabliert und es steht eine Reihe von Werkzeugen zur Verfügung, mit denen geeignete Berichte und Analysen erstellt werden können.

Die Erfolge bei der konsolidierten Verwaltung wesentlicher Datenbestände eines Unternehmens führen zu neuen Ideen der Datennutzung und ermöglichen damit letztendlich neue Geschäftsprozesse. Mit den Zielsetzungen „one face to the customer“ und „E-Commerce“ sind neuartige Architekturen von Geschäftsprozessen unter den Schlagworten „supply chain management“ und „customer relationship management“ bekannt geworden. Nur mit einer integrierten Datenverwaltung lassen sich die damit verknüpften, modernen Geschäftsprozesse realisieren.

Gerade um qualitativ hochwertige Kundenbeziehungen zu unterstützen müssen die Daten eines Kunden konsistent und aktuell am jeweiligen Ort zur Verfügung stehen. Beispielsweise bieten Unternehmen gleichzeitig Dienstleistungen über eine individuelle Bedienung in Geschäften vor Ort als auch mittels neuartiger Services des electronic shopping im Internet an. Dazu sind integrierte Daten in der Filiale zu persönlichen Kundenbedienung, im Call Center für die telefonische Betreuung und Bearbeitung kundenindividueller Vorgänge und im Internet für das persönliche Kundenkonto konsistent und aktuell bereitzustellen. Hier werden zur Zeit Data Warehouses als Mittel zur Realisierung gesehen, die unabhängig von den existierenden operationale Altsystemen schnell eingesetzt werden können.



nen.

Während in der ursprünglichen Ausrichtung das Data Warehousing im wesentlichen ein Einbahnstraße Richtung Managementinformationen war, ist es aufgrund dieser neuen Anforderungen in die operativen Geschäfte vollständig zu integrieren. Das Szenario für die genannten Bereiche der Kundenbetreuung findet sich in der obigen Skizze wieder (hier konzentriert auf Marketingaktionen; CM = Campaign Management, CRM = Customer Relationship Management, CTI = Computer Telefon Integration).

Damit ist das gängige Verständnis eines Data Warehouses als materialisierte Sichten und spezielle Ausprägung von föderierten Datenbanken nicht mehr ausreichend. Vielmehr ist die Einbindung von Data Warehouses in die operativen Systeme zwingend notwendig („Operatives Data Warehouse“). Dieses lässt sich z. B. daraus ableiten, dass Geschäftstransaktionen wie Verkäufe an Endkunden direkt das Verhalten des Kundenbetreuungsprozesses beeinflussen müssen, somit die operativen Daten im Idealfall bei Abschluss der Transaktion schon in diesem „operativen Data Warehouse“ stehen und dadurch im dargestellten Anwendungsfall die Kundenbetreuungsprozesse direkt beeinflussen.

In vielen Unternehmen werden zur Zeit IT-Systeme nach dem oben dargestellten Lösungsszenario realisiert bzw. sind aktuell schon in Betrieb. Die Aktualitätslücke bei den vom Data Warehouse bereitgestellten Daten – die sich aufgrund der üblichen, einmal täglichen Ladevorgänge ergibt – wird zur Zeit noch in Kauf genommen. Es besteht aber der dringliche Wunsch, dass die Daten im Data Warehouse zu jedem Zeitpunkt mit den operativen Daten übereinstimmen.

Daraus ergeben sich Fragestellungen für die weitere Forschung in diesem Bereich da das Vorgehen zur Beseitigung dieser Aktualitätslücke nicht klar ist: Wie kurzfristig bzw. in welchen Zyklen können z. B. inkrementelle Ladeprozesse durchgeführt werden: Minutenweise oder sogar noch kürzer? Hier würden mich z. B. Meinungen von Tool-Herstellern interessieren.

Oder ist hier vielmehr die Gesamtarchitektur in Frage zu stellen? Die Data Warehouse basierte Architektur ist in allen mir bekannten Fällen im dargestellten Anwendungsszenario üblich. Dieses liegt insbesondere an den großen Datenmengen, die für die Kundenbetreuungsprozesse zu analysieren sind. Gibt es hier alternative Architekturen und wie sehen diese aus?

Über Diskussionsbeiträge würde ich mich sehr freuen!

Heinrich Jasper

Workshop „Design und Management of Data Warehouses“

Im Juni 2000 fand im Rahmen der CAiSE-Konferenz in Stockholm der zweite internationale Workshop „Design and Management of Data Warehouses“ statt. Die Vortragenden berichteten unter anderem über Techniken für verbesserte multidimensionale Indexierung, Entwurfsmethodiken zur bestimmter einer minimalen Anzahl zu unterstützender Dimensionen.

Ein interessante Diskussion entspann sich um die Frage, ob das anspruchsvolle Qualifikationsprofil eines Data-Warehouse-Administrators nicht den Einsatz von Data Warehouses in der Praxis behindert? Die Forschung solle sich um Entwurfs- und Evolutionsmethoden bemühen, die ein Data-Warehouse zu einem weitgehend sich selbst administrierendem Werkzeug machen.

Die Papiere des Workshops sind online über die Adresse <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/> abrufbar. Im Rahmen der CAiSE-Konferenz 2001 im schweizerischen Interlaken ist eine Fortsetzung dieser Workshopserie geplant.

Manfred Jeusfeld