

Data Warehouse als unternehmensweiter Informationspool

Unter einem Data Warehouse wird eine Architektur verstanden, die die unternehmensweite Versorgung mit entscheidungsorientierten Informationen gewährleistet. Dazu wird das Data Warehouse getrennt von operativen Vorsystemen aufgebaut und betrieben. Nur so lässt sich eine konsistente unternehmensweite Datenbasis etablieren, in die selektierte und verdichtete Informationen anwendungsgerecht aufbereitet einfließen und auf die interaktiv und intuitiv zugegriffen werden kann. Für die gespeicherten Dateninhalte ist deren thematische Ausrichtung sowie Vereinheitlichung, Dauerhaftigkeit und Zeitorientierung charakteristisch.

Data-Warehouse-Lösungen existieren heute in einer Vielzahl von Unternehmen. Das Zusammenführen von relevanten Daten für die Unternehmensführung und -steuerung zu einer harmonisierten und zentralen Basis ist gängige Praxis. Dabei geht es keinesfalls mehr nur noch darum, die Daten entscheidungsorientiert dem Management für Berichtswesen und Analyse aufzubauen und die Data-Warehouse-Datenbank als Einbahnstraße zu betrachten. Vielmehr werden basierend auf Data-Warehouse-Technologie inzwischen auch Lösungen für die Integration von operativen Systemen bzw. zur Erfüllung operativer Aufgaben bereitgestellt. Die „Operationalisierung“ des ursprünglichen Data-Warehouse-Konzepts spiegelt sich in den Anforderungen des Realtime Warehousing und des Closed-Loop-Ansatzes wider.

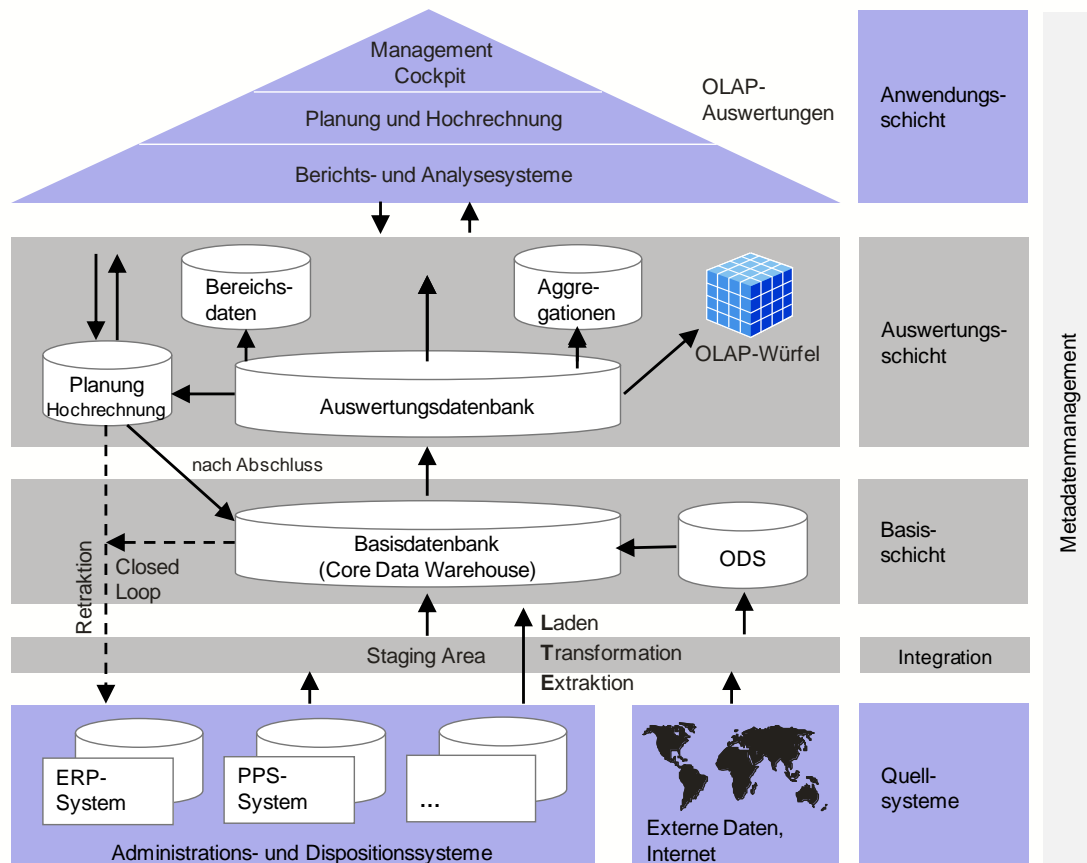
Konzept und Lösung:

- Data Warehouse als unternehmensweiter Informationspool
- Abgestimmte, harmonisierte und konsistente Inhalte
- Vielfältige Analysefunktionalität
- Reduktion benötigter Schnittstellen
- Verbesserung der Entscheidungsqualität
- Realtime Warehousing
- Closed-Loop-Ansatz

Die cundus AG hat ihre Kompetenz bei der Implementierung von Data Warehouses in vielen Projekten bewiesen. Die Kunden profitieren insbesondere vom Einsatz eines konzeptionellen Vorgehensmodells in Verbindung mit Frameworks bei der Realisierung.

Auswertungsschicht

Die Auswertungsschicht einer Data-Warehouse-Lösung besteht aus einer Kombination von relationaler und multidimensionaler Datenhaltung. Verschiedene funktionale Erweiterungen der marktgängigen Datenbanken tragen dazu bei, dass die spezifischen analytischen Anforderungen auch bei großen Datenmengen erfüllt werden können.



cundus
Germany
Duisburg
Frankfurt
Munich

cundus
Switzerland
Basel
Zurich

cundus
United Kingdom
London

cundus
United States of America
Washington D.C.

cundus
Canada
Toronto

Zugunsten der erforderlichen Zugriffperformance wird häufig auf eine konsequente Normalisierung – wie bei operativen Systemen üblich – verzichtet. Vielmehr erfolgt der Aufbau denormalisierter Datenmodelle, die als Star-Schema bezeichnet werden und aus Fakten- und Dimensionstabellen bestehen. Die Faktentabellen beinhalten die betriebswirtschaftlich relevanten und durch mehrere sachliche Identifikationskriterien bzw. Dimensionen (wie Region, Kunde oder Artikel) beschriebenen numerischen Messgrößen. Dimensionstabellen teilen mindestens eine Attributspalte mit den zugehörigen Faktentabellen und speichern weitere Angaben zu den Dimensionselementen. Eine besondere Herausforderung bei der Gestaltung der Auswertungsschicht besteht in der anwendungsbezogenen Strukturierung von relationalen Data Marts und Aggregaten sowie multidimensionalen OLAP-Würfeln.

Basisschicht

Die Basisdatenbank ist der eigentliche Kern der langfristigen Datenspeicherung in einem Data Warehouse. Sie wird daher auch als Core Data Warehouse bezeichnet. Beim Übergang von der vorgelagerten Staging Area in die Basisdatenbank werden die in verschiedenen Formaten und Bedeutungen angelieferten Daten vereinheitlicht. Die Basisdatenbank wird auch oft als (Daten-) Konsolidierungsebene benutzt, und über sie werden vor allem die Dateninhalte und -formate standardisiert.

In einem Data Warehouse, das mehrmals täglich mit Daten aus Ländern mit verschiedenen Zeitzonen beladen wird, muss die Konsistenz des Datenbestandes durch die Basisdatenbank sichergestellt werden. Zu definierten Zeitpunkten werden die Daten in die Auswertungsdatenbank übernommen und stehen dann sofort für Analysen zur Verfügung. Zeitliche Inkonsistenzen lassen sich durch dieses Vorgehen vermeiden.

Im Gegensatz zur Auswertungsdatenbank liegen die Daten in der Basisdatenbank meist in normalisierter Form und teilweise sogar in einer der Original-Struktur des Quellsystems ähnlichen Form vor. Die Detailtiefe der Daten in der Basisschicht kann fallbezogen höher sein, als in der Auswertungsschicht.

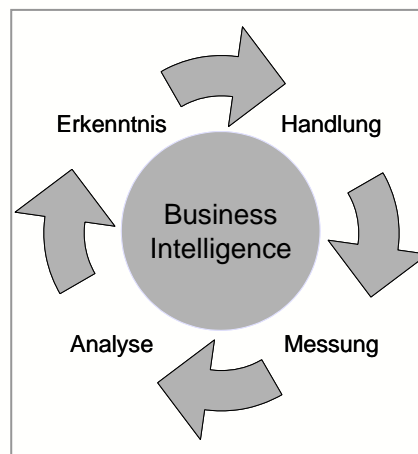
Integrationschicht – Extraktion, Transformation, Laden (ETL)

Die Etablierung von ETL-Prozessen ist in vielen Projekten der aufwändigste Teil der gesamten Implementierung. Daten, die in unterschiedlichen Formaten und Definitionen vorliegen, müssen aus den operativen Systemen extrahiert, dann in das Zielformat transformiert und geladen werden. Die Transformation beinhaltet dabei nicht nur die Konvertierung von Formaten, sondern auch die Vereinheitlichung von Dateninhalten und -schemata. Für diesen Vorgang sind komplexe Transformationsregeln notwendig, wie beispielsweise das Aufsplitten oder Zusammenführen von Attributen und verschachtelte SQL-Joins.

Realtime Warehousing und Closed-Loop-Ansatz

Realtime Warehousing und der Closed-Loop-Ansatz sind Anforderungen, die aus der praktischen Data-Warehouse-Nutzung entstanden sind und in den frühen Konzepten noch keine Rolle gespielt haben. Damals wurden die Komponenten vorwiegend für die Management-Unterstützung konzipiert. In vielen Unternehmen ist der Einsatz von Data-Warehouse-Technologie so erfolgreich verlaufen, dass die Anwender den Einsatz nun auch für den operativen Bereich fordern. Darüber hinaus möchte man die wertvolle Datenbasis für weitere Zwecke einsetzen. Unter Realtime Warehousing wird die zeitnahe Verarbeitung und Bereitstellung von Daten aus Geschäftsprozessen direkt nach einer Transaktion verstanden. Analysen sind damit auf Basis gerade beendeter oder sogar noch laufender Geschäftsprozesse möglich. Beispielsweise beobachtet ein großer Flughafen in Deutschland kontinuierlich den Status aller gerade laufenden Bodenoperationen mit Hilfe eines Reportingtools basierend auf einem Realtime Warehouse.

In der Vergangenheit wurde ein Data Warehouse konzeptionell oft als Einbahnstraße verstanden, in das zwar Daten automatisiert eingespielt werden, das allerdings keine Daten automatisiert an operative Systeme zurückgibt. Dies hat sich geändert, und inzwischen ist es üblich, dass Planzahlen oder Hochrechnungen, die mit Unterstützung von Ist-Zahlen aus einem Data Warehouse entstehen, in die operativen Systeme zurückgespielt werden. In den operativen Systemen werden sie weiter heruntergebrochen, und aus ihnen können sogar konkrete Soll-Wert-Vorgaben für den Produktionsprozess abgeleitet werden. Der Closed-Loop-Ansatz geht aber darüber noch hinaus: Aufgrund der umfangreichen und vereinheitlichten Datenbasis in einem Data Warehouse ist es möglich, komplexe Regeln zu definieren, die bei Vorliegen bestimmter Datenkonstellationen direkten Einfluss auf die operativen Geschäftsprozesse nehmen.



cundus
Germany
Duisburg
Frankfurt
Munich

cundus
Switzerland
Basel
Zurich

cundus
United Kingdom
London

cundus
United States of America
Washington D.C.

cundus
Canada
Toronto

Discover more at:
www.cundus.com