

1 Einleitung

Datawarehousing hat sich in den letzten Jahren zu einem der zentralen Themen der Informationstechnologie entwickelt. Es wird als strategisches Werkzeug zur Bereitstellung von Informationen für Kontroll- und Entscheidungsprozesse genutzt. Darüber hinaus bildet es die Basis für andere Technologien, zum Beispiel in den Bereichen strategische Unternehmensführung, Kundenmanagement oder Prozessanalyse.

Der Bedarf an Systemen zur Bereitstellung und Analyse von Informationen ist so alt wie die Informationstechnologie selbst. Die Datenanalyse wurde jedoch anfangs nur als Appendix operativer Systeme verstanden. Mitte der 90er-Jahre setzte sich der Begriff des Data Warehouse durch, formte seine speziellen Konzepte und wurde zu einem eigenen Bereich in der Informationstechnologie. Denn Data-Warehouse-Systeme schaffen Möglichkeiten, um unternehmensweit auch große Datenmengen zu analysieren und den Aufwand für die Bereitstellung von Daten erheblich zu verkürzen.

Der Begriff *Data Warehouse* bezeichnet ein System zur zentralen Bereitstellung von Informationen für Kontroll- und Entscheidungsprozesse. Dies bringt technische und funktionale Besonderheiten mit sich, durch die sich ein Data-Warehouse-System von anderen Systemen unterscheidet.

Zur Charakterisierung von Data-Warehouse-Systemen kann die Menge aller Anwendungssysteme in zwei Kategorien unterteilt werden:

- betriebswirtschaftlich administrative Systeme
- entscheidungsunterstützende Systeme

Die Kategorie der *betriebswirtschaftlich administrativen Systeme* (auch als *operative Systeme* bezeichnet) stellt Funktionen zur Verfügung, mit deren Hilfe die geschäftlichen Transaktionen eines Unternehmens verwaltet und durchgeführt werden können. Typische Aufgaben für administrativ betriebswirtschaftliche Systeme liegen in der

*Betriebswirtschaftlich
administrative Systeme*

Auftragerfassung, Rechnungsstellung, Lagerverwaltung, Personalverwaltung, Lohnbuchhaltung etc. Im Falle der sogenannten Enterprise-Resource-Planning-Systeme (ERP) unterstützen diese Systeme nicht nur einzelne Bereiche, sondern alle Funktionen der betriebswirtschaftlichen Wertschöpfungskette. Als ERP-System gilt zum Beispiel das SAP ECC bzw. dessen Vorgänger, das SAP R/3, die nachfolgend nur unter dem Begriff SAP ERP subsummiert werden.

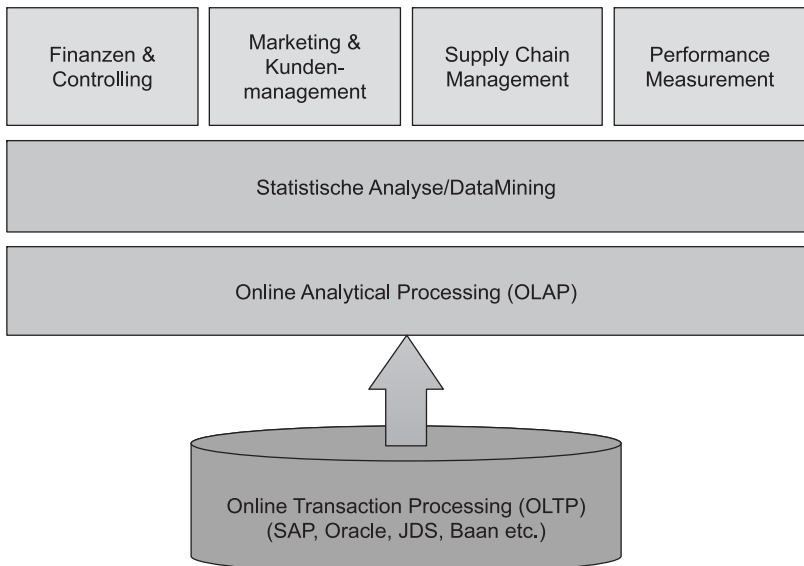
Da sich betriebswirtschaftlich administrative Systeme aus technischer Sicht vor allem dadurch charakterisieren, dass ihre Funktionen auf einzelnen Transaktionen (Aufträge, Buchungssätze etc.) basieren, werden sie auch als *OnLine-Transaction-Processing-Systeme* (OLTP-Systeme) bezeichnet.

Entscheidungsunterstützende Systeme

Administrativ betriebswirtschaftliche Systeme sind aufgrund ihrer Ausrichtung auf einzelne Transaktionen nicht dafür geeignet, als Grundlage komplexer betriebswirtschaftlicher Entscheidungen eingesetzt zu werden. Aus diesem Grund existieren als Gegenpart die *entscheidungsunterstützenden Systeme* (Decision-Support-Systeme, DSS).

Als DSS gelten das klassische Berichtswesen, die interaktive Datenanalyse (*OnLine Analytical Processing*, OLAP) sowie Systeme zur Suche nach komplexen Zusammenhängen und unbekanntem Datenmustern (Data Mining).

Abb. 1-1
Entscheidungsunterstützende Systeme



Neben diesen Tools, die allgemeine Funktionen zur Datenanalyse bereitstellen, existieren prozessorientierte Tools, die auf die Analyse bestimmter Prozessdaten spezialisiert sind. Typische Einsatzgebiete

dieser Tools liegen in den Bereichen Finanzen & Controlling, Marketing & Kundenmanagement, Supply Chain Management und Performance Measurement (siehe Abb. 1–1).

Alle entscheidungsunterstützenden Systeme benötigen eine Datenbasis als Grundlage ihrer analytischen Funktionen. Da OLTP-Systeme aufgrund ihrer Funktionen und insbesondere aufgrund ihrer Datenstrukturen als Datenbasis nicht geeignet sind und DSS in der Regel über keine eigene Datenhaltung verfügen, werden Data-Warehouse-Systeme als gemeinsame Datenbasis aller Decision-Support-Systeme eingesetzt.

Data-Warehouse-Systeme

Diese sind vor allem durch ihr Datenmanagement charakterisiert, das die analytische Betrachtung des Datenbestandes besonders unterstützt. Der Datenbestand eines Data Warehouse wird aus den Daten der OLTP-Systeme übernommen (extrahiert), im BW in Datenmodelle überführt, die für die Analyse großer Datenbestände optimiert sind, und dort redundant zu den Daten der Quellsysteme vorgehalten.

Ein Data Warehouse stellt damit nicht per se ein vollständiges Decision-Support-System dar, sondern ist lediglich die Grundlage eines solchen und stellt entsprechende Schnittstellen zur Verfügung. In der Regel werden Data-Warehouse-Systeme aber bereits durch den Hersteller zumindest mit einem OLAP-Tool gekoppelt. Aus dieser Gewohnheit heraus werden Data-Warehouse-Systeme oftmals auch als OLAP-Systeme bezeichnet, was jedoch nicht richtig ist, sondern lediglich die Verkaufsgewohnheiten der Anbieter beschreibt.

Im Falle des SAP BW wird mit den Tools des Business Explorers gleich eine ganze Produktsuite an entscheidungsunterstützenden Systemen an das Data Warehouse gekoppelt, die teilweise sogar Bestandteil der BW-Implementierung sind (z.B. die Data Mining Workbench und große Teile der integrierten Planung).

Wie der Titel bereits ausdrückt, beschäftigt sich dieses Buch mit dem SAP BW als Data Warehouse, lässt also die mit dem BW ausgelieferten entscheidungsunterstützenden Werkzeuge so weit außen vor, wie dies möglich ist. Nicht etwa, weil diese Werkzeuge nicht von Bedeutung wären, doch würde eine angemessene Beschreibung einiger dieser Werkzeuge bereits ein eigenes Buch füllen und den hier zur Verfügung stehenden Rahmen sprengen.

Inhalt

Das Buch ist in fünf aufeinander aufbauende Abschnitte unterteilt:

- Architektur
- Datenmodell
- Analytical Engine
- Extraktion & Staging
- BW-Design
- BW im Produktivbetrieb

Im Rahmen der *Architektur* werden die an einer BW-Installation beteiligten Komponenten, ihre Funktionen und ihr Zusammenwirken konzeptionell erläutert. Dabei werden neben dem Metadaten-Repository vor allem auch die Applikationsplattformen des Basissystems und ihre Kommunikationstechnologien umrissen.

Im Anschluss daran wird zunächst auf die spezifischen Datenmodelle im Data Warehousing im Allgemeinen eingegangen, um daraus die spezifische Implementierung des *Datenmodells* im SAP BW abzuleiten und zu erklären. Dabei wird insbesondere auf die für die Performance relevanten Gesichtspunkte der Datenmodellierung eingegangen und praxisnahe Grundregeln zur Modellierung gegeben.

Zwischen den entscheidungsunterstützenden Werkzeugen und den Datenbeständen des BW steht die sogenannte *Analytical Engine*, die mit den an sie gerichteten Abfragen den Datenbestand des BW liest, auf die individuelle Datenmodellierung des BW eingeht und dabei auch Statusinformationen aus Extraktion und Staging berücksichtigt. In dem Abschnitt zur Analytical Engine werden die Arbeitsweise und das Tuning bzw. Monitoring der Engine behandelt.

Der Abschnitt über *Extraktion & Staging* beschreibt, wie sich der Datenfluss von den Quellsystemen bis ins BW definieren lässt. Die Beschreibung orientiert sich an einer Referenzarchitektur, die den Fluss der Daten in logische Ebenen untergliedert, in denen sie validiert, transformiert, fehlerbereinigt und integriert werden. Des Weiteren werden die Themen Datenqualität und Performance-Tuning aus Sicht des Staging betrachtet.

In einem eigenen Abschnitt über das *BW-Design* werden typische Abwandlungen der Referenzarchitektur erläutert, die im vorangegangenen Abschnitt entwickelt wurde. Dabei wird insbesondere auf Partitionierungstechniken und Large-Scale-Architekturen eingegangen.

Im Anschluss an die vorangehenden Abschnitte, die das BW in einer statischen Betrachtung erläutern, wird im Rahmen des *Regelbetriebs* eine Zusammenstellung aller regelmäßig auszuführenden Prozesse gegeben und aus organisatorischen wie auch aus technischen Gesichtspunkten erläutert, wie die Automatisierung und das Monitoring dieser Prozesse zu realisieren sind.

Den Abschluss des Buchs bildet ein umfangreicher *Anhang*, der Spezialthemen wie z.B. Währungsumrechnung, Transportwesen und Übernahme und Entwicklung von Metadaten-Content behandelt.