
1 Einleitung

Der Begriff »Data Warehouse« bezeichnet ein System zur zentralen Bereitstellung von Informationen für Kontroll- und Entscheidungsprozesse. Das Sammeln von Informationen einerseits und das Bereitstellen dieser Daten andererseits bringt technische und funktionale Besonderheiten mit sich, durch die sich Data-Warehouse-Systeme von anderen Systemen abgrenzen.

Ein wesentliches Merkmal von Data-Warehouse-Systemen ist ihre Optimierung auf die Verarbeitung und Speicherung von Massendaten – entweder um große Mengen von Transaktionsdaten zu verarbeiten und zu speichern, wenn sie aus operativen Systemen gesammelt werden (bspw. Aufträge, Fakturen, Bestandsveränderungen usw.), oder um die gespeicherten Daten über lange Zeiträume zu selektieren, zu lesen und (aggregiert) bereitzustellen.

Die Grundlagen zur Modellierung geeigneter Datenstrukturen für ein Data-Warehouse-System werden im Teil I (Datenmodell) behandelt. Dabei werden auch die Besonderheiten herausgearbeitet, die der Betrieb des SAP BW auf unterschiedlichen Datenbankplattformen mit sich bringt. Unterschieden wird dabei zwischen den gängigen relationalen Datenbanksystemen (Oracle, DB2, MS SQL Server, Max DB, Informix) und der hauptspeicherresistenten HANA-Datenbank, die zusätzlich zu den bekannten zeilenbasierten Speicherstrukturen der relationalen Datenbanksysteme auch eine spaltenbasierte Speicherung bietet und damit vollständig andere Ansätze bei der Modellierung von Datenstrukturen ermöglicht. BW-Systeme, die auf Basis der HANA-Datenbank betrieben werden, werden von der SAP auch als »BW powered by SAP HANA« bezeichnet.

Eine ähnliche Sonderstellung kommt dem Business Warehouse Accelerator (BWA) zu, der über vergleichbare Technologien wie die HANA-Datenbank verfügt, jedoch nicht das jeweils eingesetzte relationale Datenbanksystem ersetzen, sondern nur in Teilbereichen ergänzen kann.

Im Anschluss an die Grundlagen zur Datenmodellierung werden in diesem Abschnitt der Umgang mit den einzelnen Objekttypen erläutert, die die unterschiedlichen Datenmodelle im BW abbilden. Die Erläuterungen reflektieren die unterschiedlichen Datenbanktechnologien und gehen implizit auf performancerelevante Aspekte ein.

Ein weiteres Merkmal von Data-Warehouse-Systemen ist ihre Fähigkeit, Daten aus operativen Systemen zu sammeln (zu »extrahieren«) und in der individuellen Form, in der sie durch die unterschiedlichen Quellsysteme bereitgestellt werden, zu verarbeiten, zu vereinheitlichen und zusammenzuführen. Zur Extraktion von Daten aus Quellsystemen können Tools von Drittherstellern genutzt werden, die sich auf den Zugriff auf unterschiedlichste Systeme spezialisiert haben. Im Falle des SAP BW ist der Einsatz derartiger Tools in der Regel nicht erforderlich, da das BW bereits umfangreiche Funktionen zur Extraktion aus SAP-ERP- und SAP-BW-Systemen, aus Datenbanksystemen, Dateien, JDBC-, XML/A- und ODBO-Datenquellen sowie aus XML-Datenströmen bereithält.

Wie Datenquellen an das BW angebunden, Daten im BW verarbeitet und in den unterschiedlichen Datenmodellen gespeichert werden, wird im *Abschnitt Datenbewirtschaftung* ab Seite 199 detailliert erläutert. Der Fokus liegt dabei auf der Beschreibung, wie der Datenfluss mit den Funktionen des BW-Release 7.3 definiert werden kann. Die »alten« Verfahren zur Datenbewirtschaftung, die bis zum BW-Release 3.5 bereitstanden, werden auf die Verarbeitung von Stammdaten reduziert und in Kapitel 19 gesondert erläutert.

Alle Prozesse zur Datenbewirtschaftung und zum Management von Datenbeständen können im SAP BW mithilfe sogenannter Prozessketten automatisiert werden. Die Beschreibung der Steuerungsmechanismen und der Gestaltungsmöglichkeiten von Prozessketten liefert der *Abschnitt Automatisierung* ab Seite 475.

Zur Verwertung der in einem Data Warehouse bereitgestellten Daten können unterschiedlichste entscheidungsunterstützende Systeme eingesetzt werden. Das SAP BW wird zu diesem Zweck mit dem sogenannten Business Explorer ausgeliefert, der einfache OLAP- und Reporting-Werkzeuge bereitstellt. Darüber hinaus können weitere entscheidungsunterstützende Systeme auf das SAP BW zugreifen, beispielsweise die diversen SAP-BusinessObjects-Tools innerhalb der SAP BI Suite, aber auch Tools von Drittherstellern.

Zwischen den entscheidungsunterstützenden Werkzeugen und den Datenbeständen des BW steht die sogenannte Analytical Engine. Diese nimmt alle Abfragen auf den Datenbestand entgegen und bildet daraus spezifische Zugriffe für die jeweiligen physischen Modelle. Sie berück-

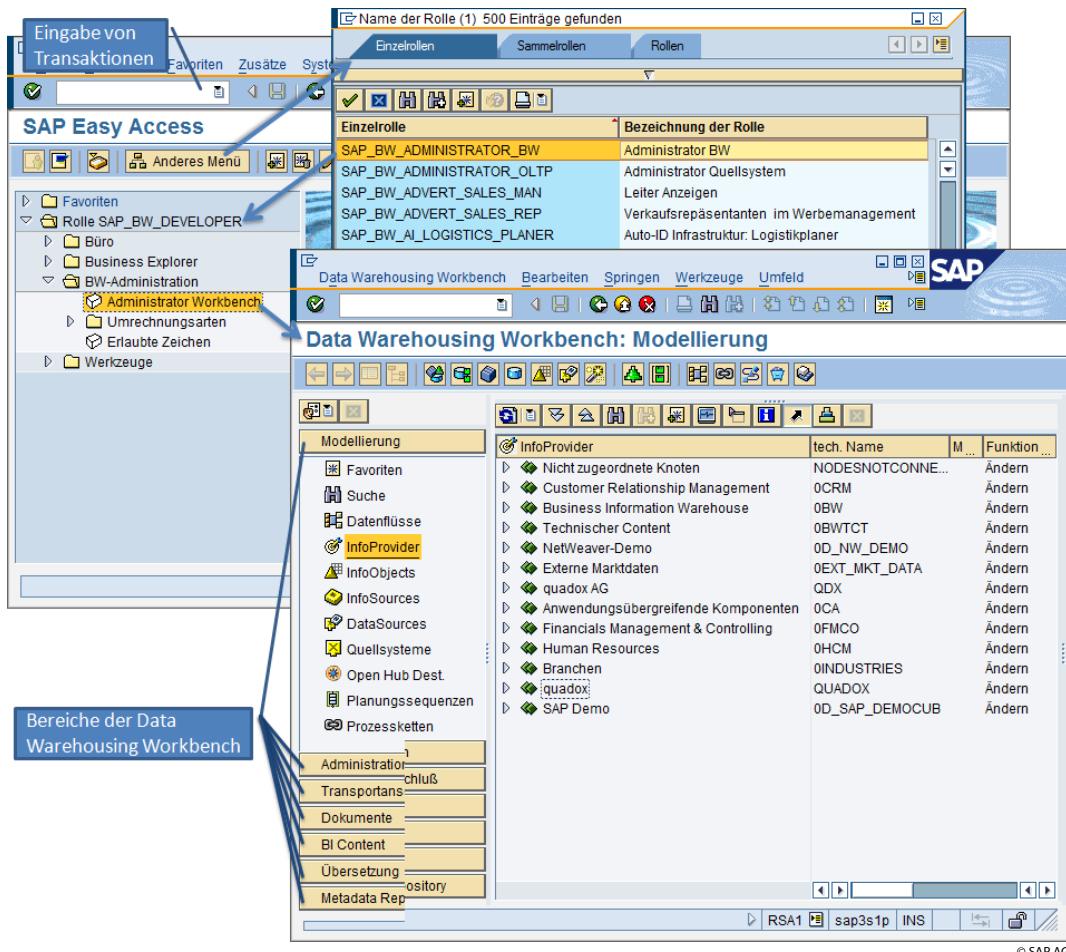
sichtigt dabei auch Statusinformationen aus Extraktion und der weiteren Verarbeitung. Im *Abschnitt Analytical Engine(s)* ab Seite 523 werden die Arbeitsweise und das Tuning bzw. Monitoring der Analytical Engine behandelt. Die Beschreibung umfasst ferner solche Datenmodelle, die nicht physisch modelliert werden, sondern lediglich als logisches Modell vorliegen und von der Analytical Engine verarbeitet werden (bspw. MultiProvider zur Zusammenführung von Daten aus mehreren physisch vorhandenen Datenbeständen).

Der gesamte Prozess des Sammelns, Aufbereitens, Speicherns und der Bereitstellung wird im Rahmen dieses Buchs unter dem Begriff Data Warehousing zusammengefasst. Durch die Vielzahl von Objekten und Möglichkeiten zur Verarbeitung und Bereitstellung von Daten lassen sich Anforderungen auf unterschiedlichste Weise im SAP BW implementieren. Im *Abschnitt BW-Design* ab Seite 583 wird abschließend ein Referenzdesign beschrieben, in dem typische Problemstellungen behandelt werden. Hierdurch soll dem Leser ein strukturierter Ansatz zur Implementierung des SAP BW geboten werden.

Die Funktionen zur Modellierung, Steuerung und Überwachung von Datenbeständen und Datenflüssen nimmt der sogenannte SAP NetWeaver Application Server wahr. Dieser besteht gleich aus zwei Entwicklungs- und Laufzeitumgebungen, die jeweils über alle Komponenten eines typischen Applikationsservers, wie z.B. Datenbank- und Kommunikationsschnittstellen, Sperrverwaltung, Job- und Prozesssteuerung oder Tools zur Systemadministration, verfügen: ein ABAP- und ein Java-EE-Applikationsserver.

Für das SAP BW (und viele andere Produkte der SAP) ist der ABAP-Applikationsserver von zentraler Bedeutung. Dieser nimmt alle Funktionen im Bereich Modellierung, Datenbewirtschaftung und Datenmanagement wahr. Die Entwicklung und Steuerung dieser Funktionen erfolgt im SAP Graphical User Interface (SAPGUI). Der zentrale Einstieg ist hierbei das sogenannte Easy-Access-Menü, von dem aus in weitere Menüs verzweigt werden kann. Welche im Easy-Access-Menü angeboten werden, richtet sich nach der jeweiligen Rolle, die ein Anwender hat.

In Abbildung 1–1 wird das Menü und die Auswahl der Rolle SAP_BW_ADMINISTRATOR_BW dargestellt, die über den Menüpunkt *BW-Administration→Administrator Workbench: Modellierung* die Data Warehousing Workbench aufrufen kann. Die Data Warehousing Workbench steht im Mittelpunkt von Entwicklung, Administration und Monitoring von BW-Funktionen. Alternativ zur Nutzung des Easy-Access-Menüs kann die Data Warehousing Workbench über die direkte Eingabe der Transaktion RSA1 aufgerufen werden.

**Abb. 1-1**

SAP Easy-Access-Menü und Data Warehousing Workbench

Dem Java-EE-Applikationsserver war ursprünglich eine wachsende Bedeutung für das BW zugedacht. Mit dem Release 7.3 des BW wurden jedoch keine neuen Funktionen vom ABAP- in den Java-EE-Applikationsserver verschoben, sondern bestehende Entwicklungen sogar wieder zurückgenommen. Der Java-EE-Applikationsserver wird aktuell vor allem zum Betrieb von Portalanwendungen wie dem BI Administration Cockpit oder dem Information Broadcaster sowie für den Universal Data Connect (mit dessen Hilfe Daten aus JDBC-, ODBO- und XML/A-Quellen extrahiert werden können) eingesetzt.