

Begriff Datawarehouse

Der Begriff Data Warehouse beschreibt eine Plattform, die Daten aus verschiedenen Datenquellen sammelt, verdichtet, sie langfristig sichert und nachgelagerte Analysesysteme versorgt. Vorteil des Datenlagers ist, dass eine globale Sicht auf Daten aus unterschiedlichen Datenbeständen entsteht.

Das Data Warehousing ist in vier Teilprozesse aufteilbar:

Anzeige

- Datenbeschaffung: Beschaffung und Extraktion der Daten aus verschiedenen Datenbeständen
- Datenhaltung: Speicherung der Daten im Datenlager inklusive Langzeitarchivierung
- Datenversorgung: Versorgung der nachgelagerten Systeme mit den benötigten Daten, Bereitstellung von Data Marts
- Datenauswertung: Analysen und Auswertungen der Datenbestände

Architektur und Prozesse des Data Warehouse

Die Prozesse des Data Warehouse lassen sich in einem Architekturschaubild vier verschiedenen Bereichen zuordnen. Diese vier Bereiche sind:

- die Quellsysteme,
- die Data Staging Area,
- die Data Presentation Area sowie
- die Data Access Tools.

OLAP/OLTP

OLAP

Online Analytical Processing ist eine Kategorie von Softwaretools, die Daten für Geschäftsentscheidungen analysieren. OLAP-Systeme ermöglichen es Benutzern, Datenbankinformationen aus mehreren Datenbanksystemen gleichzeitig zu analysieren. Die Daten werden bei OLAP subjektbezogen und verdichtet, unter spezieller Berücksichtigung historischer Verlaufsdaten, vorgehalten. Datenaktualisierungen finden nur periodisch durch Abzüge (Snapshots) operativer Systeme statt.

OLTP

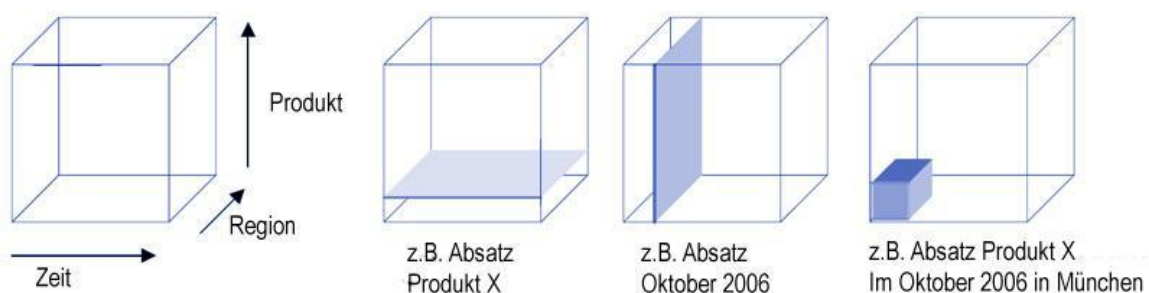
Die Online-Transaktionsverarbeitung kurz OLTP unterstützt transaktionsorientierte Anwendungen in einer 3-Teil-Architektur. OLTP verwaltet die tägliche Transaktion einer Organisation. Bei OLTP wiederholen sich die Datenbankprozesse ständig, sind strukturiert und bestehen aus isolierten, atomaren Transaktionen. Diese arbeiten mit aktuellsten Daten und greifen lesend sowie schreibend meist nur auf wenige Datensätze über Primärschlüssel zu.

DER OLAP-CUBE

Die Daten werden für OLAP-Analysen nicht wie im relationalen Datenmodell als flache Tabellen, sondern in einem multidimensionalen Datenwürfel, dem „Cube“, präsentiert. Auf diese Weise können die Daten aus unterschiedlichen Perspektiven und in diversen Detaillierungsstufen betrachtet werden. Betriebswirtschaftliche Kennzahlen wie Umsatz oder Kosten lassen sich so anhand von Dimensionen wie Kunden, Regionen oder Zeit mehrdimensional analysieren.

Mit Hilfe von OLAP-Funktionen kann der Würfel analysiert werden. Die erforderlichen Datensätze können durch verschiedene Operationen erzeugt werden, etwa indem man einen Teilwürfel ausschneidet („Slicing“) oder den Datenwürfel rotiert („Dicing“).

Bestimmte Sichten, etwa das Betrachten eines bestimmten Produktes oder eines bestimmten Jahres, erhält man durch das Herausschneiden von Ebenen.



Die Analyseergebnisse lassen sich auch schrittweise verfeinern bzw. verdichten, was als „Drill down“ bzw. „Roll-up“ bezeichnet wird. Beim Roll-Up werden alle Einzelwerte zu einem weiter oben liegenden Hierarchieattribut verdichtet. Die Funktion Drill-Down bewirkt eine detailliertere Darstellung der Daten.

Mit Hilfe der beschriebenen Operationen kann der Benutzer interaktiv durch Drag & Drop der Dimensionsschaltflächen beliebige Auf- und Einsichten in den Datenwürfel erreichen.

STAR SCHEMA

Das Star-Schema ist, wie der Name schon beschreibt, sternförmig aufgebaut und auch für analytische Anwendungen im Data Warehouse-Umfeld geeignet.

Die Fakten (engl. Measures) in einem Star-Schema sind auch in diesem Modell das zentrale Element der Datenanalyse. Sie haben die Aufgabe wichtige Zusammenhänge in quantitativ messbarer und verdichteter Form wiederzugeben.

Die Dimensionen (engl. Dimensions) in einem Star-Schema ermöglichen unterschiedliche Sichten auf die Fakten. Sie liefern einen fachlichen Bezug auf die quantitativen Werte in einem Sternschema.

Die **Fakten** können in **einem Star-Schema gruppiert und analysiert** werden. Die Betrachtung von Verdichtungsstufen ermöglicht die sogenannten Hierarchisierungen. Da im Star-Schema keine direkte Hierarchisierung von Dimensionen, wie im Snowflake-Schema, möglich ist, wird über die „kontrollierte Redundanz“ in den **Dimensionen eine Hierarchisierung ermöglicht**.

Das **Star-Schema** setzt sich aus **Fakten- und Dimensionstabellen zusammen**. Die Faktentabelle enthält einerseits die wichtigen Kennzahlen und andererseits speichert sie die Fremdschlüssel der Dimensionstabellen.

VORTEILE

Die **Vorteile des Star-Schema** sind das **intuitive Datenmodell** und eine **geringe Anzahl an Join-Operationen**, um an betriebswirtschaftliche Informationen zu gelangen.

NACHTEILE

Ein **großer Nachteil** ist das **schlechte Antwortverhalten bei großen Dimensionstabellen**, das aber unter Einhaltung von wichtigen Modellierungsregeln beherrschbar ist