

Data Warehouse

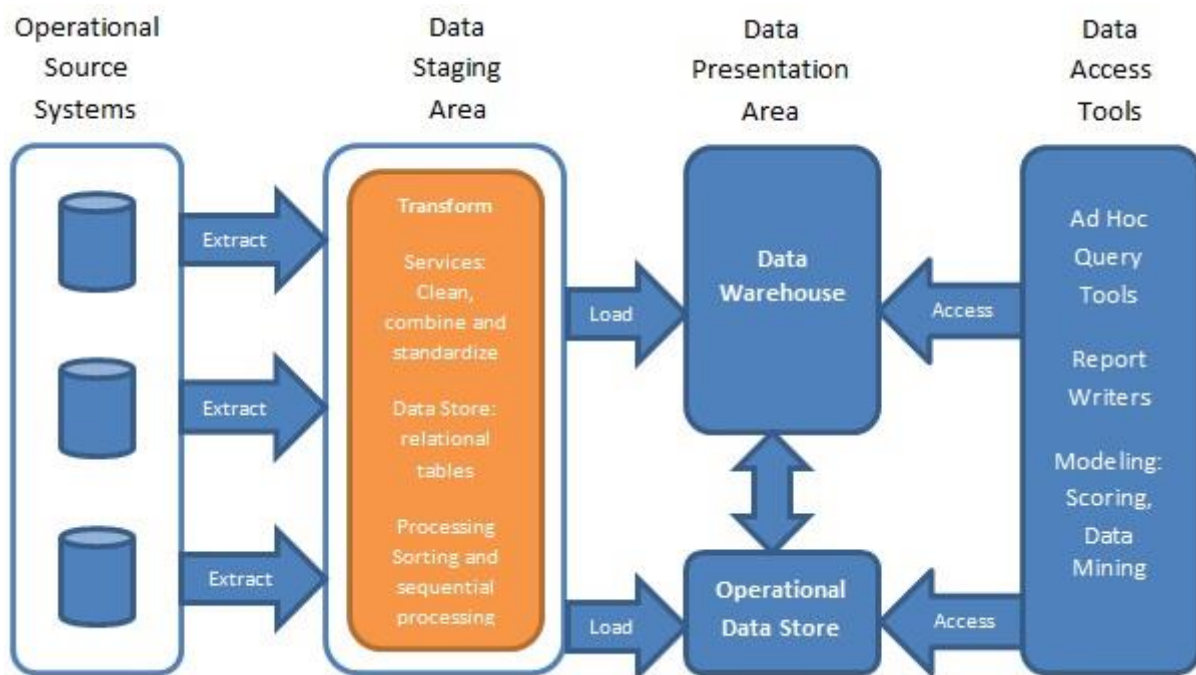
Ein Data Warehouse ist eine themenbezogene und dauerhaft gespeicherte Ansammlung von Daten. Die Speicherung erfolgt strukturiert in denormalisierten relationalen Datenbanken.

Das Data Warehouse dient der abteilungsübergreifenden Auswertung von Informationen und unterstützt den Entscheidungsprozess in einem Unternehmen.

Ein Data Warehouse führt die Daten aus unterschiedlichsten Quellen zusammen und stellt logische Verbindungen zwischen allen diesen Daten her. Dazu müssen die Daten ihren Quellort nicht zwangsläufig verlassen. Sie können auch als verknüpfte Datenverbindung im Data Warehouse vorliegen.

Das Data Warehouse-System wird von operativen Systemen mit Informationen befüllt, die in der Data Staging Area gesichert werden.

Mittels ETL-Prozesse (Extract, Transform, Load) werden die Informationen bereinigt, konsolidiert und aggregiert in ein Data Warehouse überführt. Data Access Tools haben damit den Zugriff auf ein komplettes Data Warehouse bzw. auf einzelne Sichten, die wiederum durch sogenannte Data Marts repräsentiert werden.



OLAP / OLTP

OLAP ist die Abkürzung von Online Analytical Processing. Man versteht darunter einen bestimmten Datenbanktyp, bei dem – im Gegensatz zu transaktionsorientierten Systemen (z.B. Buchungen oder Bestellungen durchführen) – das Auswerten vorhandener Datenbestände im Vordergrund steht.

Es dient zur Betrachtung von Informationen in einem multidimensionalen Datenraum.

OLAP-Systeme beziehen ihre Daten entweder direkt aus operativen Vorsystemen. Dies geschieht vor allem dann, wenn nur wenige Daten verarbeitet werden sollen. Bei größeren Datenbeständen importiert man aus einem zentralen Data Warehouse. Dorthin werden periodisch die Daten aus meistens mehreren Vorsystemen übernommen und redundant gehalten.

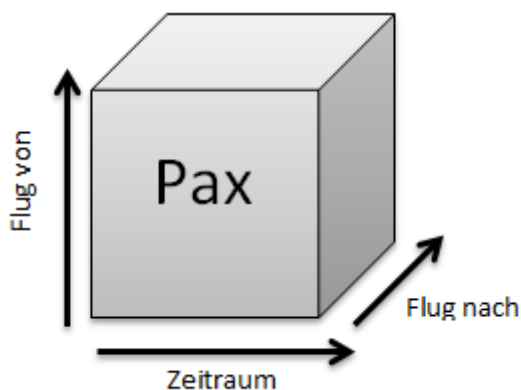
Online Transaction Processing (OLTP) oder Online-Transaktionsverarbeitung, bezeichnet eine Klasse von Programmen, die in der Lage sind, transaktionsorientierte Anwendungen im Internet auszuführen.

In der Regel werden OLTP-Systeme für Auftragserfassung, Finanztransaktionen, Customer Relationship Management und Einzelhandelsumsätze eingesetzt. Diese Systeme haben eine große Anzahl von Nutzern, die kurze Transaktionen ausführen. Die Datenbankabfragen sind meist einfach, erfordern Antworten in Sekundenbruchteilen und liefern relativ wenig Datensätze.

Ein wichtiges Merkmal von OLTP-Systemen ist deren Fähigkeit, gleichzeitige Zugriffe zu unterstützen. Um Single Points of Failure zu vermeiden, sind OLTP-Systeme häufig dezentral organisiert.

OLAP-CUBE

Der OLAP-Cube dient zur Grundlage für multidimensionale Analysen. Die Dimensionen werden durch die Kanten des Würfels repräsentiert. Die Länge der Kanten wird durch die Anzahl der Merkmale bestimmt.

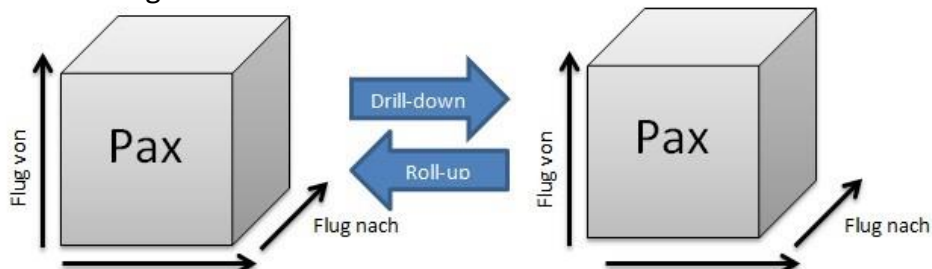


Dieser OLAP-Cube zeigt ein einfaches Beispiel einer Airline. Die Dimensionen werden durch die Kanten repräsentiert. In diesem Fall existieren drei Dimensionen: Flug von, Flug nach und Zeitraum. Innerhalb der Würfelzellen werden ein oder mehrere Fakten hinterlegt z.B. Passagierzahlen.

In einem OLAP-Cube können beliebig viele Operationen durchgeführt werden. Diese OLAP-Operationen ermöglichen verschiedene Sichten auf den OLAP-Würfel, wie zum Beispiel eine Drill-Down-/Roll-Up-, Slice-/Dice- und Rotate-Funktion.

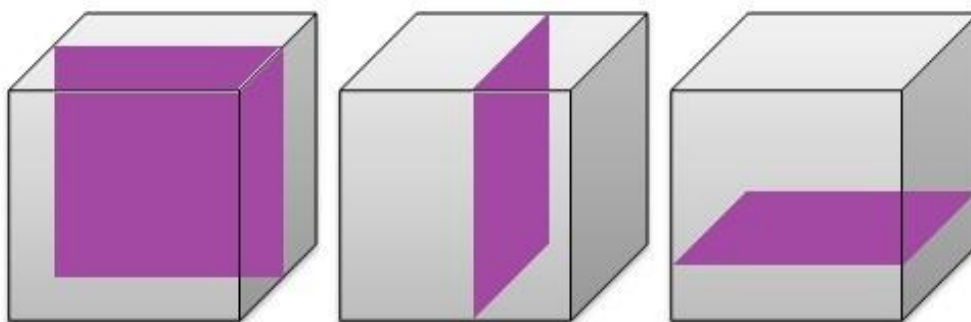
Die OLAP-Operationen Drill-Down und Roll-Up ermöglichen das Durchlaufen der Verdichtungsstufen, entlang der auf einer Dimension zugeordneten Merkmale in einem Würfel.

Mit der Funktion Drill-Down steigt man von einem aggregierten Merkmal auf ein Merkmal mit der nächsttieferen detaillierten Verdichtungsstufe ab. Mit Hilfe der Operation Roll-Up aggregiert man von einer tieferen detaillierten Verdichtungsstufe zu einer nächsthöheren verallgemeinerten Verdichtungsstufe.

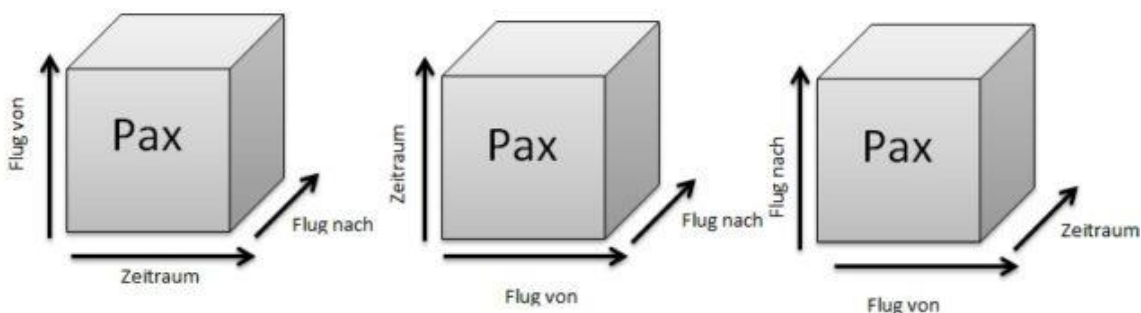


Die OLAP-Operation Slice gleicht dem Herausschneiden einer Scheibe aus einem dreidimensionalen Würfel. Das Ergebnis wird in einer zweidimensionalen Tabelle zurückgeliefert.

Die OLAP-Operation Dice schneidet einen Teilwürfel aus einem OLAP-Cube heraus. Auf Basis des Teilwürfels können weitere Operationen, wie z.B. dem weiteren Herausschneiden (Slice) durchgeführt werden.



Die OLAP-Operation Rotate dreht einen OLAP-Cube um seine eigenen Achsen. Durch die Drehung erhält der Anwender unterschiedliche Sichten auf einen Datenwürfel.



Sternschema

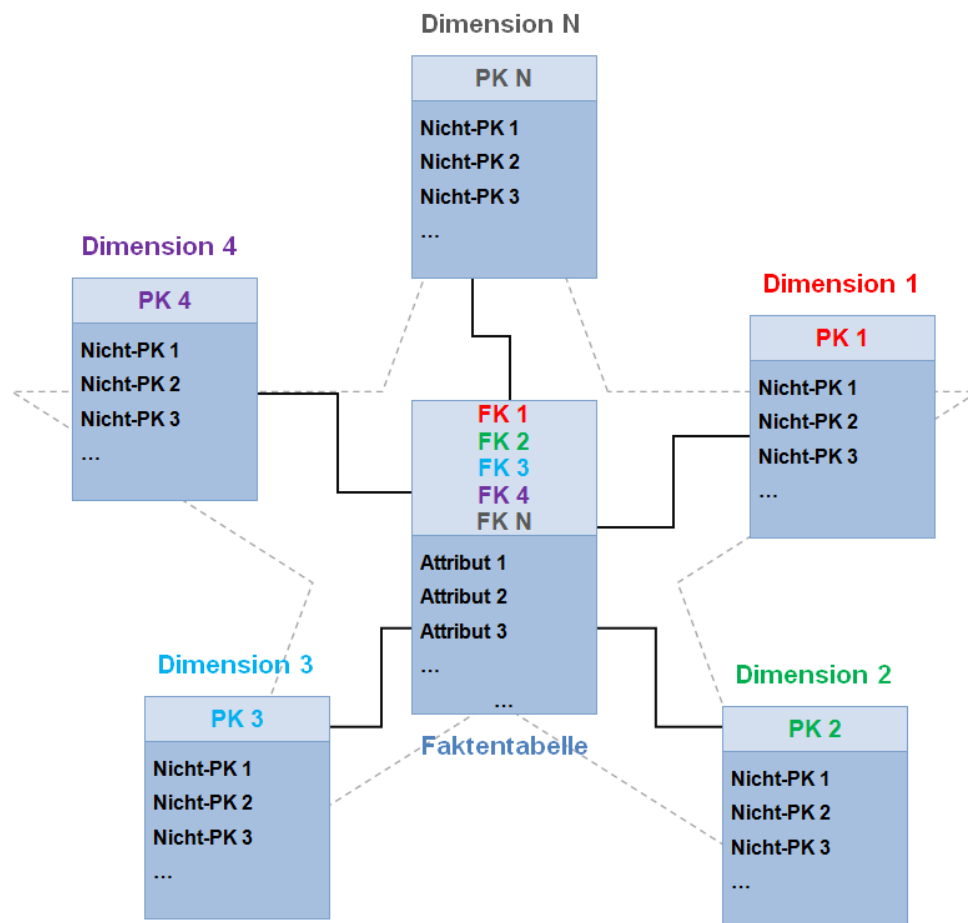
Das Sternschema ist sternförmig aufgebaut und auch für analytische Anwendungen im Data Warehouse-Umfeld geeignet.

Die Fakten in einem Sternschema 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 in einem Sternschema 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, 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.



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

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