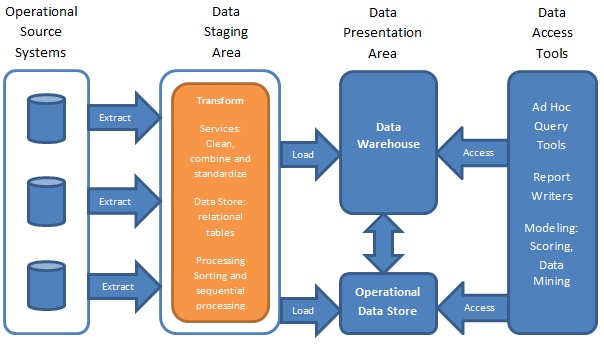
Data Warehouse – Was ist ein Data Warehouse?

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.

Die Daten werden unabhängig von den operativen Datenbanken und Anwendungssystemen bereitgestellt und zu analytischen Zwecken aufbereitet.

Data Warehouse Aufbau



Das Data Warehouse-System wird von operativen Systemen mit Informationen befüllt, die in der Data Staging Area konsolidiert werden. Mittels ETL-Prozesse 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 vs. OLTP Systeme

Online Analytical Processing versteht man die Verarbeitung und Auswertung von Daten mit der Hilfe von Programmen und IT-Systemen. Da diese Daten meist mehrdimensional sind, werden spezielle OLAP Programme benutzt. Die Daten kommen meist aus einem Data-Warehouse.

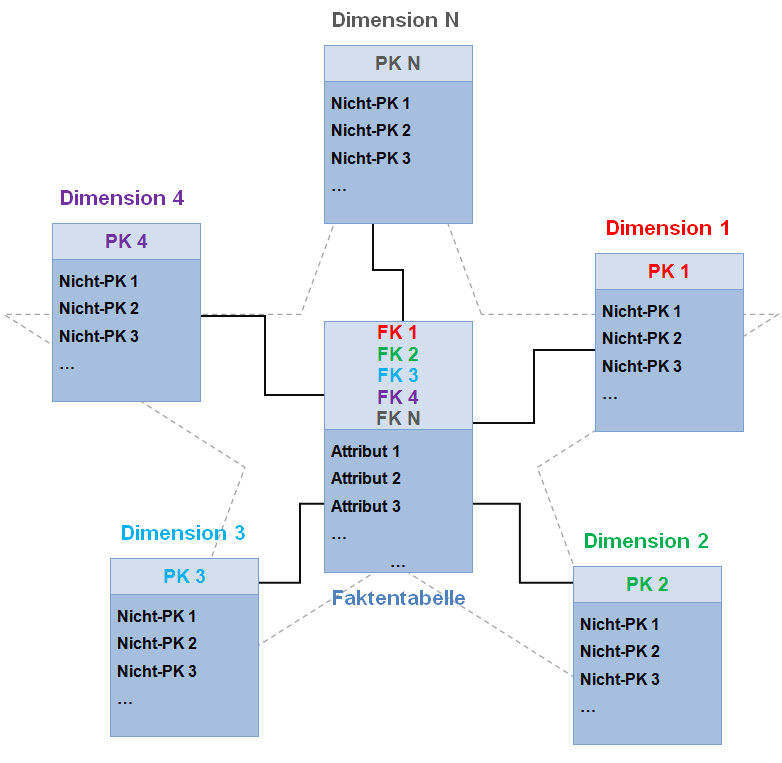
OLAP hat eine kleine Anzahl an Transaktionen. Die Abfragen sind dabei komplex und brauchen dabei eine gewissen Abfragezeit, welche von der Komplexität abhängt.

OLTP hat eine große Anzahl an Transaktion pro Sekunde. Die Abfragen sind dabei einfach, damit die Daten so schnell wie möglich abgefragt werden. OLTP spezialisiert sich also auf die Schnelligkeit.

Sternshema

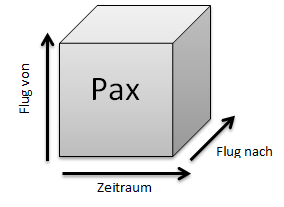
Das Sternschema ist ein Datenbankdesign bzw. eine Modellierungstechnik welches eine multidimensionale Sicht auf eine relationale Datenbank ermöglicht. Ziel dieses Datenbankdesigns ist es nicht eine Normalisierung zu schaffen, sondern eine Optimierung auf effiziente Leseoperationen, z.B. im Fall einer Datenbankabfrage.

Die Daten werden Leseoptimierung in zwei Tabellengruppen eingeteilt: Faktentabellen und Dimensionstabellen.   
Üblicherweise gibt es eine Faktentabelle und mehrere Dimensionstabellen.  
Faktentabellen benötigt man zum Analysieren. Fakten sind Ereignisse die gezählt oder gemessen werden können. Im Data Warehouse beinhaltet die Faktentabelle Fremdschlüssel auf die Primärschlüssel der Dimensionstabellen. So ergibt die Gesamtanzahl der Fremdschlüssel den Primärschlüssel der Faktentabelle.   
Dimensionstabellen, auch Nebentabellen genannt, sind die Tabellen mit zusätzlichen beschreibenden Informationen.



Cube

Mehrdimensionale Daten werden in einem sogenannten OLAP-Würfel abgebildet. Dieser dient dazu um Mehrdimensionale Daten schnell und einfach abzufragen. Der Würfel bietet dabei verschieden Operationen an um auf die Daten zuzugreifen.



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.

Operationen

**Roll-Up**:

Bei der Roll-Up Operation werden Dimensionen zu einer höheren Hierarchiestufe zusammengefasst.

z.B.: Januar, Februar und März werden zu 1.Quartal zusammengefasst.

**Drill-Down:**

Beim Drill-Down werden Daten in kleinere Hierarchiestufe aufgespalten.

z.B.: Lehrer werden zu POS\_Lehrer, DBI\_Lehrer, AM\_Lehrer

**Dice**

Dabei werden bestimmte Daten gefiltert d.h. es wird ein Teilwürfel aus dem Gesamtwürfel generiert.

**Slice**

Dabei wird ein Teil der Gesamtwürfels herausgeschnitten.

**Pivotierung**

Dabei werden die Daten nicht verändert, sondern es wird lediglich die Perspektive des Würfels verändert d.h. die Dimensionen werden neu angeordnet.