Data Warehouse

1. Definition

Ein Data Warehouse ist eine zentrale Sammlung von gleich strukturierten Daten, deren Inhalt sich aus verschiedenen, unterschiedlich strukturierten Quellen zusammensetzt. Diese Daten werden dann vor allem zwecks Zusammenfassung, übergreifender Analyse und betriebswirtschaftlicher Entscheidungshilfe dauerhaft gespeichert.   
Es fungiert häufig als Grundelement einer unternehmensweiten Strategie für Performance Management (bedeutet dass Prozesse so optimiert werden, das die höchst möglichste Leistung erzielt wird, wobei auch auf die Entwicklung der Mitarbeiter eingegangen wird) oder Business Intelligence Lösungen (bedeutet dass anhand der Verzahnung von Daten die Zusammenhänge zwischen allen Unternehmensbereichen erkannt werden können und sich daraus Rückschlüsse ziehen lassen). Außerdem ermöglicht es einen leichteren Zugriff auf die Daten.

Quelle: <https://www.pmone.com/wiki/data-warehouse/>

1. 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.

Quelle: <http://datawarehouse4u.info/OLTP-vs-OLAP.html>

1. 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.

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**

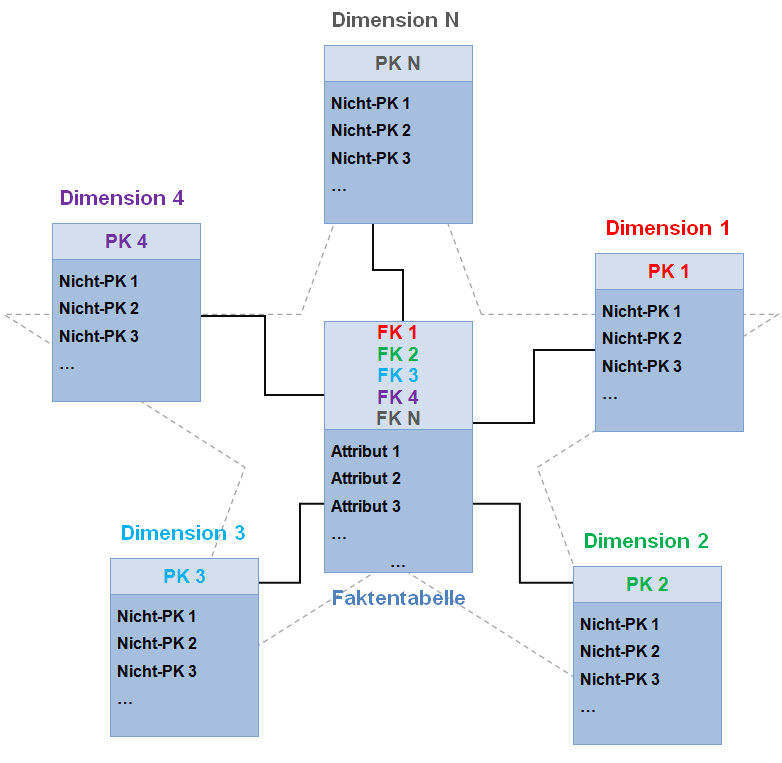
Dabei werden die Daten nicht verändert, sondern es wird lediglich die Perspektive des Würfels verändert d.h. die Dimensionen werden neu angeordnet.

Quelle: <http://www.wirtschafts-lehre.de/online-analytical-processing.html>

1. Sternschema

Das Sternschema ist eine besondere Form eines Datenmodells, dessen Ziel nicht die Normalisierung ist, sondern eine Optimierung auf effiziente Leseoperationen. Sie soll eine multidimensionale Sicht auf die relationale Datenbank ermöglichen.

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.



Quelle Bild: Von Dhbw dbi - Eigenes Werk, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=6969631