



## *Inhalt:*

- *Konzepte*
- *OLAP*
- *ROLAP und MOLAP*
- *Lebenszyklus*
- *Vergleich OLTP und OLAP*
- *ETL*



## Konzepte:

- *Batch-Verarbeitung*
- *OLTP = Online Transaction Processing*
- *OLAP = Online Analytical Processing*

*OLAP-Systeme sind unverzichtbare Instrumente betrieblicher Analyse der umfangreichen und mehrdimensionalen Daten. Diese Systeme gewähren anwendungsspezifische Sichten auf Daten. In erster Linie werden sie von Managern unterschiedlicher Ebenen verwendet.*



## Gründe für OLAP:

- *Trennung von Tagesgeschäft und Auswertungen.*
- *Historisierte Daten mit Zeitraum-Bezug.*
- *Große Mengen von Nur-Lese-Daten (Permanenz).*
- *Multidimensionale Datenmodelle.*
- *Gezielte Denormalisierung des ganzen Modells.*



## Eigenschaften von OLAP:

- *Intuitive und interaktive Analyse der Daten.*
- *Flexible Darstellung der Daten aus unterschiedlichen Perspektiven.*
- *Basis der OLAP-Analyse ist der Hypercube (kartesisches Produkt).*
- *Besondere Operationen sind definiert – Rotation, Slice, Dice, Drill-Through, Drill-Across, Roll-Up, Drill-Down.*
- *Als Clients werden spezielle Programme oder bekannte Tabellenkalkulationstools wie Excel verwendet.*



*Data Warehouse ist eine OLAP-Datenbank zwecks:*

- *Unterstützung von Entscheidungen für strategische Unternehmenssteuerung.*
- *Analyse von Tendenzen und Mustern mittels großer Datenmengen über große Zeiträume hinweg.*
- *Bessere Entscheidungen durch bessere Informationen.*
- *Flexiblere Analysemöglichkeiten.*
- *Verlagerung der Analyse-Möglichkeiten in die Fachabteilungen.*
- *Geringere Berichterstellungskosten.*
- *Gemeinsame Informationsbasis im Unternehmen.*



*Die OLAP-Datenbanken werden physikalisch organisiert als:*

- *ROLAP – Relationales OLAP. Solche Systeme basieren auf den relationalen Datenbanken wie Oracle, DB2. Zum Einsatz kommen Star-Schema (Daten sind in Fakten- und Dimensionstabellen zugeordnet, 3NF ist bei Dimensionstabellen verletzt) und Snowflake-Schema (Tabellen sind normalisiert). ROLAP-Systeme orientieren sich an Anwendungsbereichen mit hohem Datenvorkommen und großen Nutzerzahlen und zeigen sich in diesen Bereichen sicher und stabil.*
- *MOLAP – Multidimensionales OLAP. Als Basis liegen hier herstellerspezifische Datenbanken vor. Sie sind für hohe Performance in multidimensionalen Datenstrukturen optimiert und gewährleisten u.a. schnelle Zusammenfassungen von Daten (Aggregationen).*
- *HOLAP – Hybrides OLAP ist eine Variante aus ROLAP und MOLAP.*



## Vorteile von ROLAP:

- *Abfragen, Verwaltung, Speicherung, Recovery, Archivierung mit bewährten relationalen Technologien.*
- *Sperrmechanismen, Transaktionen etc. werden nicht benötigt.*

## Vorteile von MOLAP:

- *Hohe Performance.*
- *Speicherung am multidimensionalen Modell ausgerichtet.*



## Nachteile von ROLAP:

- *Umfangreiche Zugriffe auf Daten erfordern hohes Maß an JOINS, Indizes, Table Scans.*
- *Umfangreiche Aggregationen und Berechnungen.*

## Nachteile von MOLAP:

- *Hoher Schulungsaufwand wegen geringer Verbreitung.*
- *Proprietäre Verwaltung.*
- *Oft fehlende Unterstützung von Standardschnittstellen.*





## Planung:

- *Komplette Analyse von Architektur und Infrastruktur.*
- *Definition der erforderlichen Ressourcen.*
- *Definition von Zeitvorgaben.*
- *Archivierungsstrategien.*
- *Verbindungsmöglichkeiten zur Datenbank.*
- *Ladeprogramme, um Daten in die Datenbank zu kopieren.*



## *Spezifikation von Datenanforderungen und Modellierung der Datenbank:*

- *Ermittlung der notwendigen Entitäten und deren Attribute aus dem Anwendungsbereich.*
- *Geschäftsprozesse und Geschäftsanwendungsfälle identifizieren und beschreiben.*
- *Ein- und Ausgabedaten definieren.*
- *Detailierungsgrad der Daten festlegen.*

*Logisches Datenmodell entsteht in diesem Schritt.*



## *Physischer Datenbankentwurf:*

- *Entwerfen des physischen Modells der Datenbank: Star-Schema (Fakten- und Dimensionstabellen), Snowflake-Schema.*
- *Aufheben der Normalisierung.*
- *Festlegen von Schlüsseln.*
- *Indizierungsstrategien festlegen.*
- *Partitionierung.*



### *Befüllen des Data Warehouse:*

- *Definition der Quellsysteme.*
- *Umformungsspezifikationen.*
- *Aktualisierungszyklus festlegen.*
- *Definition der Prozeduren für ETL-Prozess (Extract, Transform, Load).*
- *Test der ETL-Prozeduren.*
- *Automatisierung der Ladevorgänge, Backup- und Recovery-Prozeduren, Test.*
- *Anwendungsentwicklung (Berichte entwerfen, Dokumentation, Test).*



*Test und Überprüfung der Daten.*

*Schulung (Verwendung des Frontend-Tools).*

*Produktabnahme.*

*Wartung.*

*Verbesserungen.*

*Weiterentwicklung des Systems (Berichte, Prozeduren).*

*Performance-Untersuchungen.*



## OLTP-Datenbanken:

- *Vorhersehbare Abfragen, meist Abfragen einzelner Datensätze.*
- *Ständige Änderungen.*
- *Nur notwendige Redundanz (normalisiertes Modell).*
- *Hohe Transaktionsrate.*

## OLAP-Datenbanken:

- *Komplexe unvorhersehbare Abfragen.*
- *Statische, unveränderbare Daten.*
- *Verständliche Datenstruktur (denormalisiertes Modell).*
- *Aggregation – Zusammenfassung von vielen Fakten zu einem einzelnen Fakt.*



## *Extraktion:*

- *Periodischer, ereignisgesteuerter oder anfragegesteuerter Abzug. Komplette oder Delta-Übertragungen.*
- *Protokollierung der Änderungen und Übertragungen.*

*Transformation wird im Arbeitsbereich durchgeführt:*

- *Datentypkonvertierung.*
- *Wertumsetzung.*
- *Schlüsselvergabe, -anpassung oder -bereinigung.*
- *Zeitstempelvergabe.*
- *Datenverdichtung, -bereinigung.*

*Laden bedeutet Übertragung der Daten aus dem Arbeitsbereich in das Data-Warehouse.*



*Grundlage der Datenverarbeitung in DWH ist der mehrdimensionale Hypercube (kartesisches Produkt), der solche Dimensionen hat, wie z.B. Zeitperioden, Produkte, Abteilungen, und Werte, wie z.B. Absatzvolumen von bestimmten Produkten in bestimmten Zeitperioden in Abteilungen.*

*Folgende Operationen erlauben Navigation und Auswertungen im Würfel:*

- Rotation. Das ist Auswahl von zwei konkreten Dimensionen, die für aktuelle Auswertung notwendig ist. Geometrisch sieht es wie eine Drehung des Würfels um eine eigene Achse aus.*
- Slice. Das ist ein voller zweidimensionaler Ausschnitt aus dem ganzen Würfel.*
- Dice. Das ist ein mehrdimensionaler Ausschnitt aus dem ganzen Würfel, praktisch eine Untermenge des Würfels, ein kleiner Würfel.*
- Drill-Across. Das ist Verbindung von mehreren Würfeln gleicher Dimension zu einer Kette, um die Auswertungen horizontal zu erweitern.*





*Einige Dimensionen des Würfels können noch hierarchisch gegliedert werden. Folgende Operationen erlauben Navigation in solchen Hierarchien:*

- Drill-Down. Das ist Navigieren von der oberen Ebene zu den tieferen Ebenen der Hierarchie.*
- Roll-Up. Das ist Navigieren von den tieferen Ebenen zu der oberen Ebene der Hierarchie.*
- Drill-Through. Ist die Operation Drill-Down nicht mehr möglich, dann kann durch diese Operation eine neue Datenquelle (Würfel) angeschlossen werden und Navigieren nach unten fortgesetzt.*



*Die Einsätze von Data Warehouse können sich variieren:*

- *Data Marts. Ein Data Mart wird in einem sehr viel begrenzten Anwendungsbereich eingesetzt, z.B. in einer Abteilung einer Firma. Gibt es in einer Firma mehrere Data Marts, so entsteht ein Problem, Konsistenz aller Daten über mehrere Data Marts zu pflegen. Andererseits Einrichtung der Data Marts ist selbstverständlich einfacher als DWH.*
- *Operation Data Stores. Diese DWH sind für aktuelle (tägliche) Auswertungen gedacht, somit unterstützen sie so gut wie keine Abfragen, die die Informationen über längere Zeit erfassen.*

