



Eigenschaften von Unternehmensanwendungen

Stefan Halfpap, Ralf Teusner, Dr. Werner Sinzig

27. April 2020

- **Einführung zu Unternehmensanwendungen**
- Grundlagen des IT-gestützten Rechnungswesens und der Planung
- Einführung zu relationalen Datenbanken und Anfrageverarbeitung
- Grundlagen von (spaltenorientierten) Hauptspeicherdatenbanken
- Trends in Hauptspeicherdatenbanken
- Klausur

- Unternehmensanwendungen
- Beispiele für Unternehmensanwendungen
- Datenbankanfrageeigenschaften von Unternehmensanwendungen
 - OLTP vs. OLAP
 - Trennung in eigenständige Systeme
 - Optimierung eines Systems für beide Anfragearten
- Eigenschaften von Unternehmensdaten
- Zusammenfassung
- Übungsblatt 1

- „Enterprise applications are about the **display, manipulation, and storage of large amounts of often complex data** and the **support or automation of business processes with that data.**“ Martin Fowler „Patterns of Enterprise Application Architecture Patterns“ (2002)
 - Große komplexe Datenmengen, die in der realen Welt existierende Entitäten abbilden
 - Unterstützung/Automatisierung von Geschäftsprozessen auf Basis dieser Daten
 - Nutzen üblicherweise (relationale) Datenbanken

- Klassische Domäne
 - ERP, SRM, CRM (siehe 01_Unternehmensanwendungen)
- Erweiterte Domäne

Neue Technologien haben Einfluss auf Unternehmensanwendungen.

Informationsintegration aus unterschiedlichen Quellen bietet neue Möglichkeiten zur Anreicherung und dadurch Optimierung der klassischen Unternehmensprozesse.

- Sensordaten in Produktionsüberwachung
- Unstrukturierte Daten wie medizinische Behandlungsberichte und Röntgenbilder
- Posts im Web und auf sozialen Netzwerken zum Einfluss von Produkten

Beispiele für Unternehmensanwendungen

Klassische Domäne: Abwicklung eines Kaufauftrags

Basierend auf Frank Körsgen „SAP R/3 Arbeitsbuch“ (2005)

- Prozesse im Vertrieb werden durch Vertriebs-/Verkaufsbelege abgebildet
 - Kundenanfrage
 - Angebot
 - Auftrag
 - Lieferung
 - Warenausgang + Buchhaltungsbeleg
 - Ausgangsrechnung + Buchhaltungsbeleg
 - Bezahlung + Buchhaltungsbeleg

- (Details in der Vorlesung zum Rechnungswesen)

Beispiele für Unternehmensanwendungen

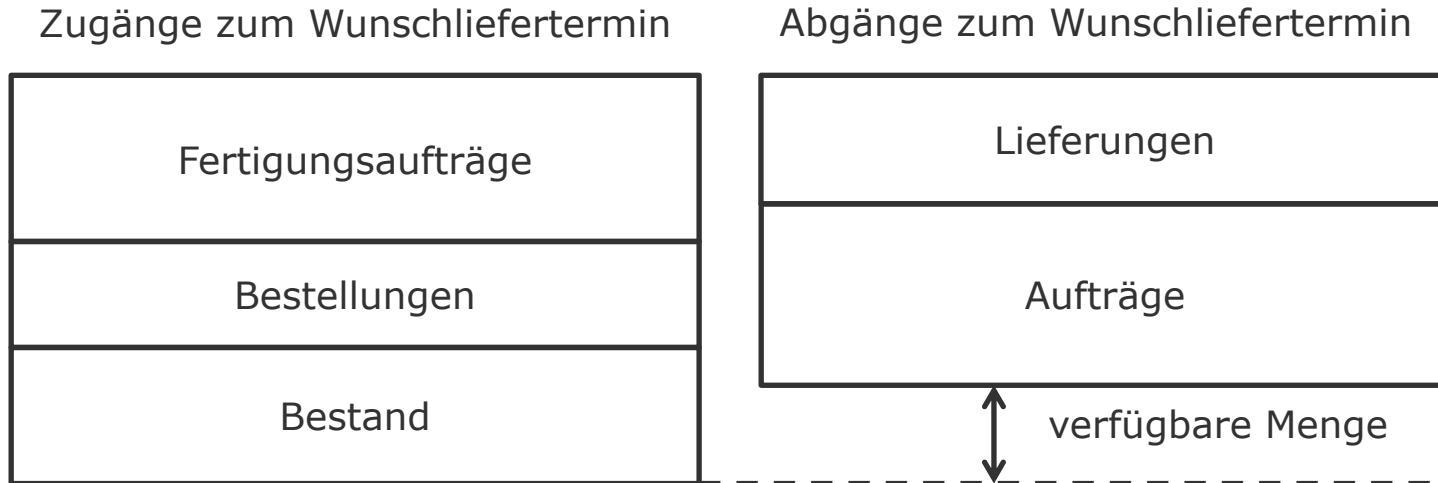
Klassische Domäne: Abwicklung eines Kaufauftrags

- Vorkommende teils aufwändige Prozesse
 - Preisfindung
 - **Verfügbarkeitsprüfung**
 - Bedarfsübergabe (an Einkauf und/oder Fertigung)
 - Versandterminierung
 - Versandstellen- und Routenermittlung
 - Kreditlimitprüfung

Beispiele für Unternehmensanwendungen

Klassische Domäne: Abwicklung eines Kaufauftrags

■ Verfügbarkeitsprüfung



Unternehmensanwendungen

Datenbankanfrageeigenschaften

Clark D. French „'One Size Fits All' Database Architectures Do Not Work for DDS" (1995)

■ Online Transaction Processing (OLTP)

- Beispiele: Auftrag anlegen/suchen, Rechnung erstellen, Kontenbuchung, Anlegen und Anzeigen von Stammdaten
- Einfache Anfragen
- Oft festgelegte Anfragen
→ Index-Unterstützung
- Geringe Datengrundlage (stark-selektive Anfragen)
- Daten(tupel)eingabe und –abruf

■ Online Analytical Processing (OLAP)

- Beispiele: Mahnlauf, Verfügbarkeitsprüfung, Cross-Selling, Operationales Reporting (Auflistung offener Rechnungen)
- Komplexe Anfragen
- Ad-hoc-Anfragen
- „Decision Support“
- Große Datengrundlage (schwach-selektive Anfragen)
- Analyseanfragen (Aggregation und Gruppierung weniger Attribute)

Unternehmensanwendungen

Historischer Hintergrund

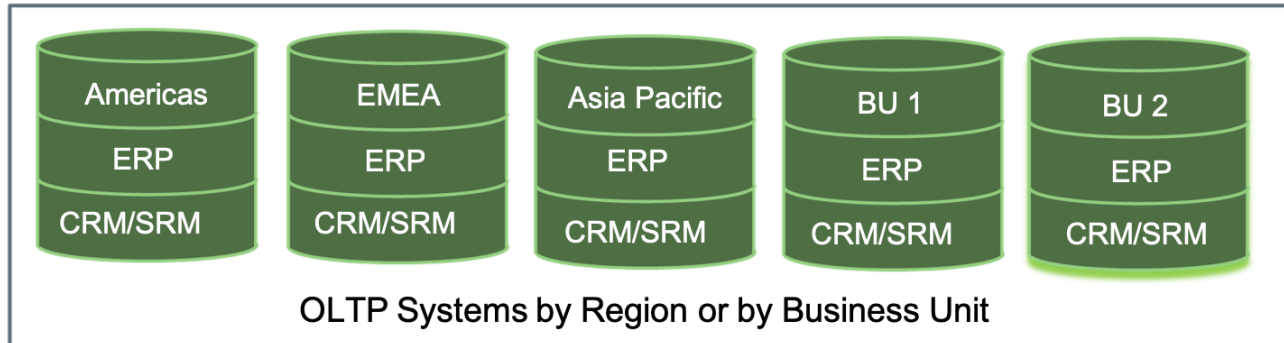
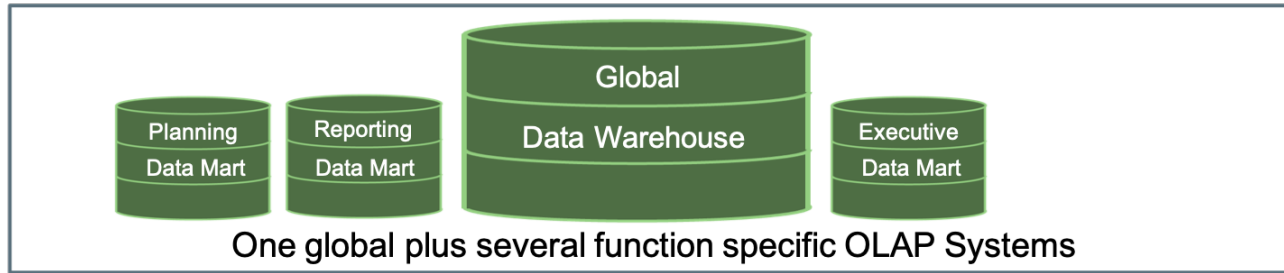
Michael Stonebraker „One Size Fits All: An Idea Whose Time Has Come and Gone“ (2005)

- Relationale DBMSs entstanden in den 1970er Jahren als Forschungs-Prototypen
 - System R
 - INGRES
- Genealogy der relationalen Datenbanken
- https://hpi.de/fileadmin/user_upload/fachgebiete/naumann/projekte/RDBMSGenealogy/RDBMS_Genealogy_V6.pdf
- Beide Systeme wurden für die Verarbeitung von Unternehmensdaten (OLTP) geschaffen
 - Seit Anfang der 1980er: „one size fits all“ Strategie
Gründe: Kosten, Kompatibilität, Verkauf, Marketing
 - In der 1990er Jahren: Unternehmen wollen Daten aus verschiedenen operationalen DBs in ein Data-Warehouse für Business Intelligence vereinen

“Although most warehouse projects were dramatically over budget and ended up delivering only a subset of promised functionality, they still delivered a reasonable return on investment. In fact, it is widely acknowledged that historical warehouses of retail transactions pay for themselves within a year [...]”

Unternehmensanwendungen

Typische Systemlandschaft



Unternehmensanwendungen

Data-Warehouse und OLAP Technologien

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

- Gründe für Entstehung: Performanz und Datenintegration
 - Relationale Datenbanken waren historisch für OLTP optimiert: Angst vor Performanzeinbruch für Transaktionen durch komplexe Ad-hoc-Anfragen
 - Ältere Daten für historische Tendenzen + zusätzliche Informationen z.B. Aktienpreise

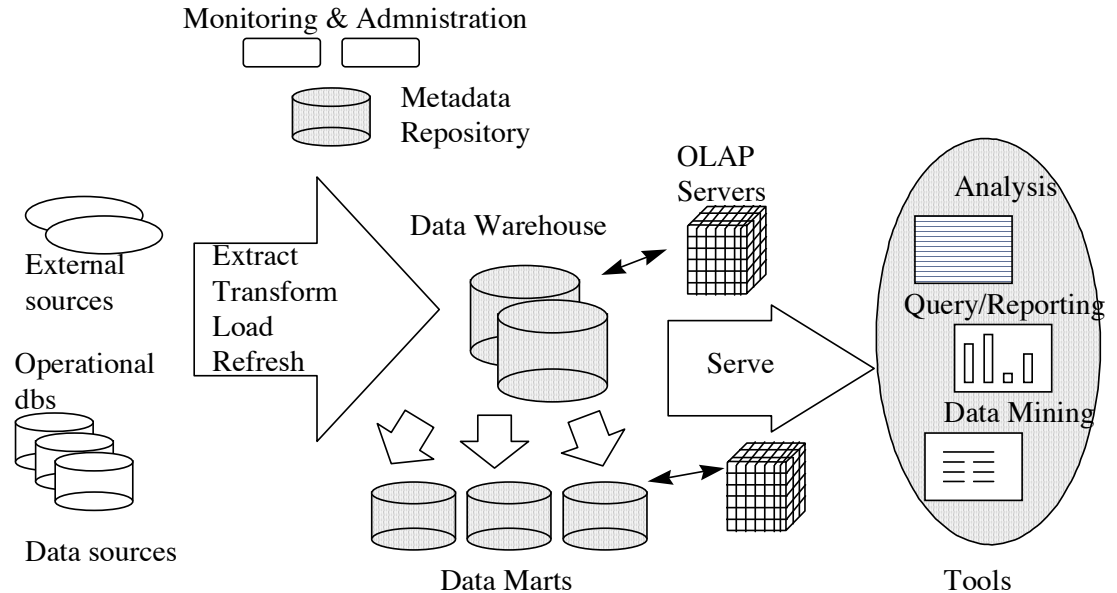
- Aufbau eines Data-Warehouses für Unternehmen ist langwieriger und komplexer Prozess
- Erfordert umfangreiche Geschäftsprozessmodellierung (Abdeckung des gesamten Unternehmens)
- Data-Marts: integrieren Datenuntermenge, die sich auf einen einzelnen Organisationsbereich fokussieren z.B. Marketing Data-Mart

Unternehmensanwendungen

Data-Warehouse Architektur

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

- Datenintegration
- Datenmodellierung
- Operationen
- Datenspeicherung
- Hilfsstrukturen



Unternehmensanwendungen

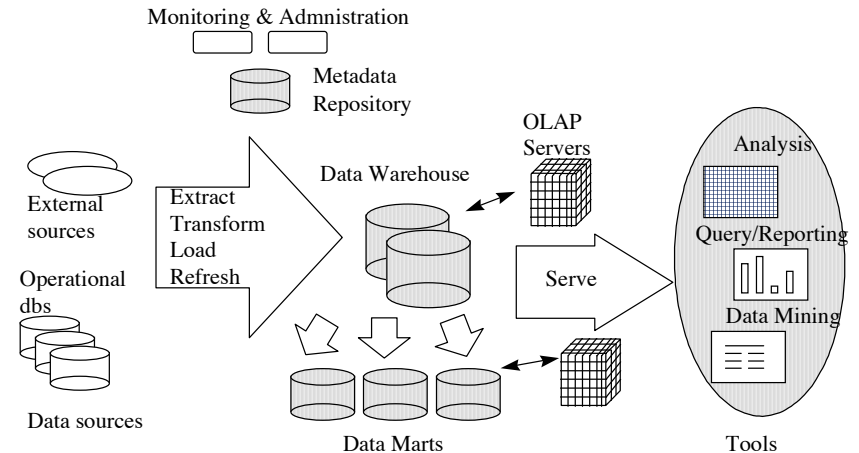
Data-Warehouse: Datenintegration

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

- Aufwendiger ETL-Prozess
 - Extraktion der relevanten Daten aus operationalen DBs und externen Quellen
 - Transformation der Daten in das Warehouse-Schema
 - Laden der Daten in das Warehouse

- Metadatenmanagement

- Beschreibung der Datenquellen
- Beschreibung des Warehouse-Schemas (siehe Folgefolien)
- Beschreibung der Datenextraktions- und Datentransformationsregeln



Unternehmensanwendungen

Data-Warehouse: Datenmodellierung

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

- OLTP Schema-Design basiert auf Normalisierung (redundanzfreie Speicherung)
- Für OLAP sind effiziente Anfragen und effizientes Laden wichtig

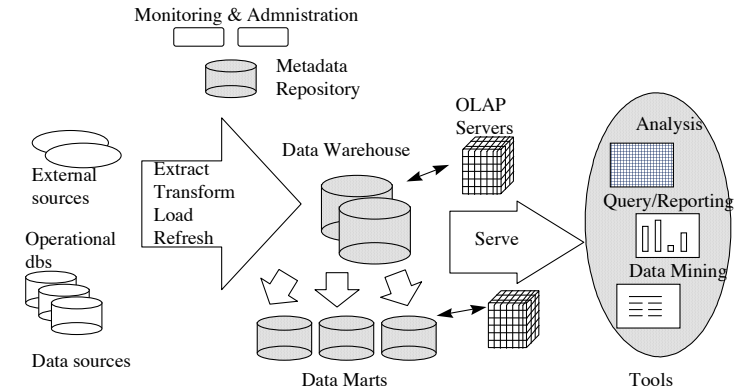
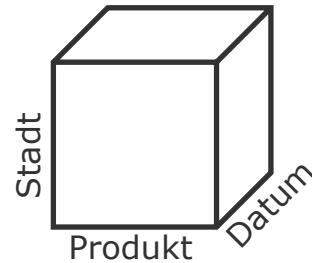
- Mehrdimensionale Datensicht
Datenwürfel (OLAP-Würfel)

- Numerische Kennzahlen werden analysiert z.B. Umsätze oder Bestände

- Jede Kennzahl hängt von einer Menge an Dimensionen ab

- Dimensionen sind oft hierarchisch

- Tag - Monat - Quartal - Jahr
- Produkt - Kategorie - Industrie
- Stadt - Bezirk - Land

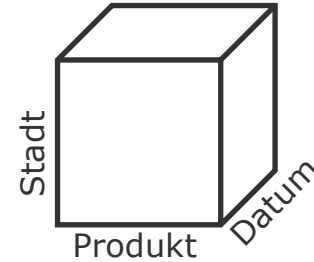


Unternehmensanwendungen

Data-Warehouse: Operationen

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

- Slicing und Dicing: Selektion/Auswahl bestimmter Informationen
- Drill-Down: zunehmende Datendetails
- Drill-Up/Roll-Up: zunehmende Aggregation/Verdichtung der Daten
- Pivoting/Rotation: Umorganisation der Datenansicht



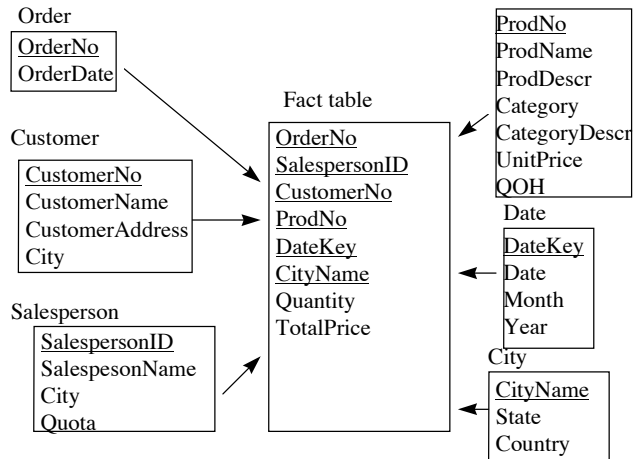
Unternehmensanwendungen

Data-Warehouse: Datenspeicherung - ROLAP

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

■ R(elational)OLAP

- Daten sind in Tabellen (relational) gespeichert
- Umsetzung im Sternschema mit Denormalisierungen aus Performanzgründen: eine Faktentabellen + eine Tabelle pro Dimension



Unternehmensanwendungen

Data-Warehouse: Datenspeicherung - ROLAP

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

■ R(elational)OLAP

- Schneeflockenschema als Alternative: Dimensionshierarchie wird explizit durch Normalisierung dargestellt

Sternschema

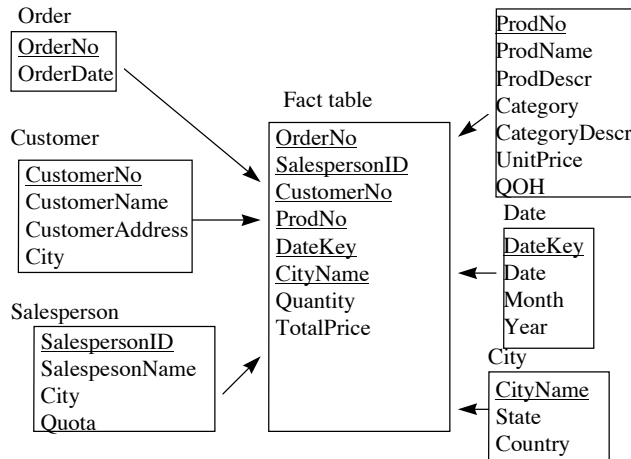
vs.

Schneeflockenschema

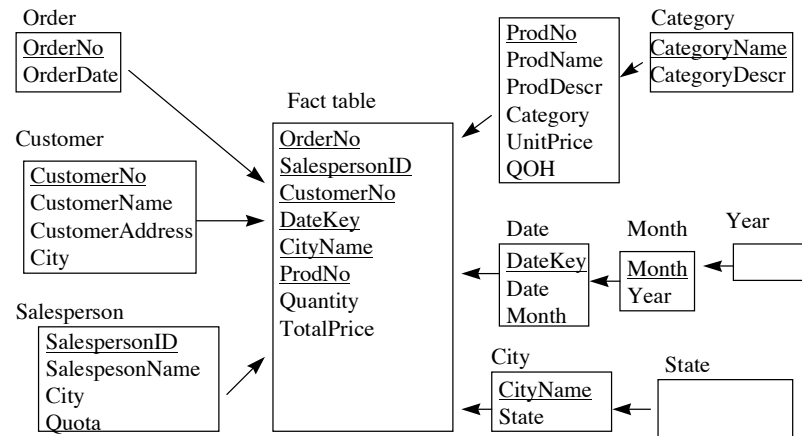
Einfacheres Schema (Anfragen)

vs.

Weniger Redundanz



VS.



Unternehmensanwendungen

Data-Warehouse: Datenspeicherung - MOLAP

S. Chaudhuri et U. Dayal „An Overview of Data Warehousing and OLAP Technology “ (1997)

- M(ultidimensional)OLAP
 - Mehrdimensionale Daten sind in speziellen Datenstrukturen gespeichert z.B. mehrdimensionale Arrays
 - Nutzen meist zwei Dimensionen (+ Kompressionsverfahren) um dünn besetzte Datenwürfel („sparse data“) effizient zu unterstützen

- ROLAP- und MOLAP-Systeme nutzen Hilfsstrukturen für effiziente Anfragebearbeitung
 - Voraggregierte Daten (nach verschiedenen Dimensionen)
 - Materialisierte Sichten
 - Indexstrukturen
 - Bit-Map-Index als Alternative zu Pointer/ID-Listen
 - Effiziente Operationen: Index-Intersection, Index-Union
 - Join-Indizes: Dimensionstabelle → Faktentabelle
 - Indizes für Textsuche
- Weitere Effizienzsteigerung durch Partitionierung/Sortierung möglich

- Unternehmensanwendungen beinhalten sowohl transaktionale als auch analytische Anfragen
- Aber: Anforderungen an OLAP- und OLTP-Datenbanksysteme sind widersprüchlich

(historischer) Ansatz 1:

Trennung in eigenständige Systeme (Tuning der Datenspeicherung und des Schemas)

- Data Warehouse: separate optimierte OLAP-Datenbank
 - Effizientere Bearbeitung komplexer analytischer Anfragen
 - Daten vieler Datenbanken können integriert werden

(moderner) Ansatz 2:

Optimierung eines Systems für beide Anfragearten (mit Kompromissen)

Unternehmensanwendungen

Nachteile der Systemtrennung

- Datenredundanz
- Datenkonsistenz schwieriger
- Synchronisierung der Systeme durch Extract, Transform, Load (ETL)
 - Kostenintensiver Prozess
 - ETL verursacht eine Zeitverzögerung
- Unterschiedliche Datenschemas erzeugen eine höhere Komplexität für Anwendungen, die beide Systeme nutzen

Unternehmensanwendungen

Kombination von OLTP und OLAP in einem System

- Eine Datenbasis als „Wahrheitsquelle“ („Single source of truth“)
 - Ziel sind ad-hoc-Analysen auf dem transaktionalen Schema ohne materialisierte Sichten (z.B. vorberechnete Aggregate) – „Mixed Workloads“ auf einem System
- Vereinfachte Anwendungen (nur eine DB mit allen Datenpunkten anfragen) und Datenbankstrukturen (weniger Indizes, keine materialisierte Sichten)
- Ermöglicht durch modernere/schnellere Hardware
 - Dennoch: für spezielle Charakteristiken optimierte Datenbanksysteme sind „One size fits all“ Systemen überlegen

Michael Stonebraker „One Size Fits All: An Idea Whose Time Has Come and Gone“ (2005)

Unternehmensanwendungen

Vielseitige Anfrageeigenschaften („Mixed Workloads“)

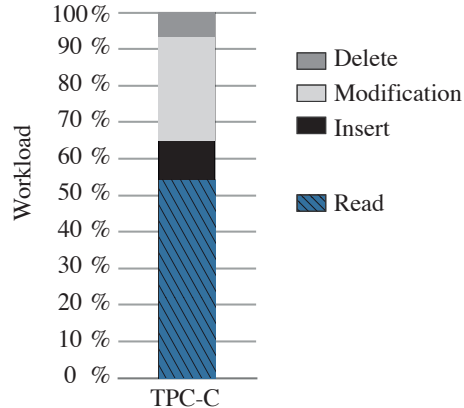
- Auch Hybrid Transactional/Analytical Processing (HTAP) oder OLxP
- Vereinen Eigenschaften von OLTP **und** OLAP
 - Einfache **und** komplexe Abfragen
 - Vordefinierte **und** ad-hoc-Abfragen
 - „Zeilenoperationen“ **und** Analysen (Aggregationen, Gruppierungen, Joins) über komplette Spalten
 - INSERTS, UPDATES **und** viele SELECTS

Unternehmensanwendungen

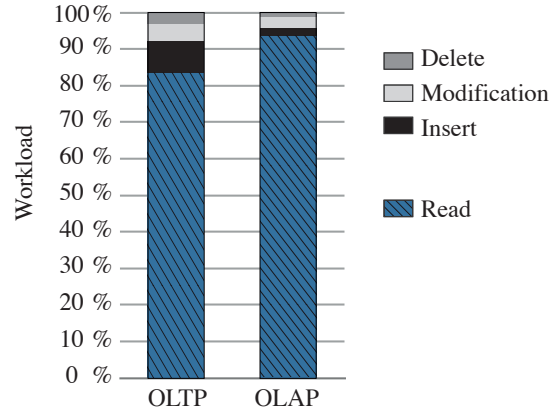
Anfrageeigenschaften: OLTP Datenzugriffsmethoden

Hasso Plattner „The Impact of Columnar In-Memory Databases on Enterprise Systems“ (2014)

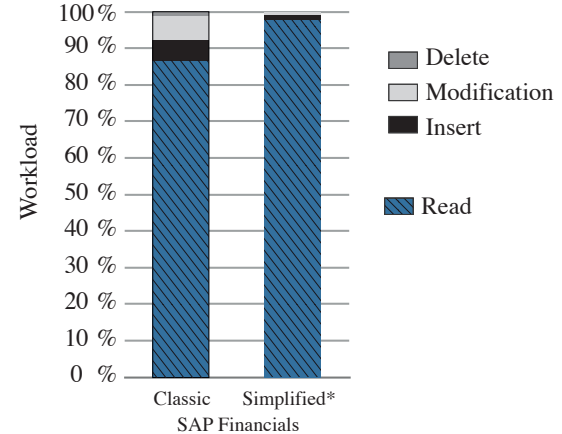
- Workload Analysen von Unternehmensanwendungen zeigen: OLTP und OLAP Workloads unterscheiden sich NICHT wesentlich bezüglich ihres Schreibanteil



(a) TPC-C Benchmark



(b) Workload Analysis
Krueger et al. VLDB'11



(c) SAP Financials
Workload Analysis

Unternehmensanwendungen

Eigenschaften von Unternehmensdaten

- Viele Tabellen haben hunderte von Attributen
- Viele Attribute werden in der Regel NICHT benutzt
- Für viele Attribute dominieren NULL oder DEFAULT-Werte
- Viele Attribute haben eine geringe Kardinalität (Anzahl verschiedener Attributwerte)
- Attributwerte sind häufig ungleichmäßig verteilt
- Attributwerte verschiedener Attribute hängen oft voneinander ab

- Tabellen sind breit und spärlich gefüllt → erlaubt effiziente Komprimierung

- Detaillierte Informationen in

Jens Krüger „Fast Updates on Read-Optimized Databases Using Multi-Core CPUs“ (2011)

http://www.vldb.org/pvldb/vol5/p061_jenskrueger_vldb2012.pdf

- Unternehmensanwendungen beinhalten sowohl transaktionale als auch analytische Anfragen; Insbesondere analytische Anforderungen sind mit der Zeit gewachsen
- Tabellen von Unternehmensanwendungen sind breit und spärlich gefüllt
- Zwei Ansätze: Separat optimierte Systeme vs. **Vereinigung in einem System**
 - Vereinigung überwindet viele Nachteile der Systemtrennung
 - Entsprechende Datenbankarchitekturen beinhalten viele Optimierungen, die den Anfrage- und Datencharakteristiken gerecht werden (kommende Vorlesungen)

Unternehmensanwendungen

Übungsblatt 1

Unternehmensanwendungen
Sommersemester 2020: Übung 1

Enterprise Platform and
Integration Concepts
Fachgebiet | Hasso-Plattner-Institut
Universität Potsdam



Aufgabe 1 (32 Punkte)

Lesen Sie Ausschnitte aus Hasso Plattners Veröffentlichungen

- „A Common Database Approach for OLTP and OLAP Using an In-Memory Column Database“
Abschnitt 1 – 3
<http://www.sigmod09.org/images/sigmod1ktp-plattner.pdf>
- „The Impact of Columnar In-Memory Databases on Enterprise Systems“
Abschnitt 1 – 5.2
<http://www.vldb.org/pvldb/vol7/p1722-plattner.pdf>

und fassen Sie die Inhalte unter Beantwortung folgender Fragen zusammen.

- a) Was sind Gründe für die historische Trennung von OLTP und OLAP in verschiedene Systeme? (6 Punkte)
- analytische Anfragen verlangsamen das transaktionale System
 - verschiedene Datenbankschemata notwendig
 - inkompatible SQL-Dialekte
 - kein Bedarf an Vereinigung
 - performante und flexible analytische Anfragen möglich
 - unterschiedliche Workload-Eigenschaften
- b) Warum und wie ist es heutzutage möglich OLTP- und OLAP-Anfragen in einem System effizient zu bearbeiten? (7 Punkte)
- Schnellere Hardware (Moore's Law)
 - Entwicklung spaltenbasierter Hauptspeicher-DBs
 - Entwicklung von NV-RAM
 - Entwicklung von Streaming-Systemen
 - größerer und günstigerer Hauptspeicher
 - Entwicklung der Blockchain