

Data Warehousing und Data Mining

Sommersemester 2007

Ulf Leser

Wissensmanagement in der
Bioinformatik



Große Datenbanken

- „... Der typische Walmart Kaufagent verwendet täglich mächtige Data Mining Werkzeuge, um die Daten der **300 Terabyte Datenbank** zu erforschen“
[Jim Gray, Computer Zeitung 17/2003]
 - Was verkaufen wir wo am häufigsten?
 - Gibt es einen Zusammenhang zwischen der Regalposition und den Umsatzzahlen von Produkten?
 - Wie lange müssen wir Produkt X durchschnittlich lagern und wo?

Data Warehousing und Data Mining

- Halbkurs
 - Di, 11.00 – 13.00 Uhr, RUD 26, 1'306
 - Do, 11.00 – 13.00 Uhr, RUD 26, 1'306
- Übung
 - Do, 13.00 – 15.00 Uhr, RUD 26, 1'306
 - Erster Termin: Donnerstag, 26.4.2007
 - Teilnahme ist Voraussetzung für Prüfung
 - Anmeldung in GOYA erforderlich
 - [Überlaufliste](#) benutzen!
- Bisher als Spezialvorlesung, erstmals als Halbkurs
 - Data Warehousing [und Data Mining](#)
- Sprechstunde
 - Nach Vereinbarung
 - IV.105, (030) 2093 – 3901, [leser\(...\)informatik.hu-berlin.de](mailto:leser(...)informatik.hu-berlin.de)

Einbettung

- Voraussetzungen
 - DBS-I
 - Relationalenmodell, ER-Modell
 - Joins, Anfrageübersetzung
 - Grundzüge de Anfrageoptimierung
 - SQL
 - (Logik)
- Keine Voraussetzung
 - DBS-II

Inhaltsübersicht –1-

- Einleitung & Motivation
 - Was ist ein DWH?
- Architektur & Prozesse
 - Konzepte
 - Komponenten eines DWH
- Modellierung von DWH
 - Multidimensionale Modellierung
 - OLAP-Operationen
- Umsetzung des multidimensionalen Datenmodells
 - Relationale Abbildung
 - SQL-basierte OLAP-Operatoren
 - MDX

Inhaltsübersicht –2-

- Extraction, Transformation & Load (ETL)
 - Integrationsprobleme und -lösungen
 - Updatestrategien
 - Datenqualität, Duplikaterkennung, Data Cleansing
- Indexstrukturen für DWH
 - Bitmap-Index & Join-Index
 - Multidimensionale Indexstrukturen: Grid-File, KD Trees
- Logische Optimierung
 - Star-Join
 - Partitionierung
 - Optimierung von OLAP Operationen
- Materialisierte Sichten
 - Auswahl, Aktualisierung, Verwendung
 - Ableitbarkeit von Anfragen, Query Containment

Inhaltsübersicht –3-

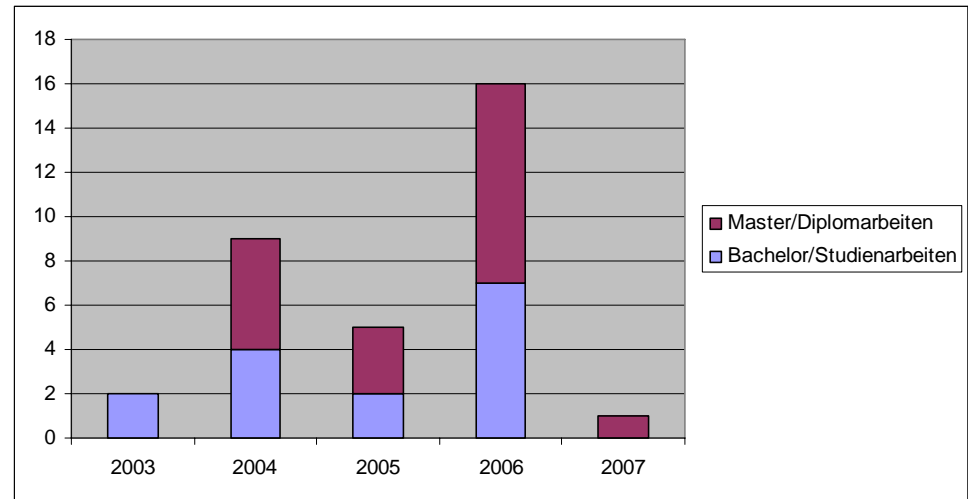
- Datenaufbereitung für das Data Mining
- Clustering
 - Klassiker: K-Means, nearest neighbor
 - DBSCAN und BIRCH
- Klassifikation
 - Klassiker: Decision Trees, Naive Bayes
 - SPRINT, RainForest
- Assoziationsregeln
 - A-Priori Algorithmus
 - Closed Pattern und quantitative ARM
- Ableitung von Kaufempfehlungen
 - Collaborative Filtering

Literatur

- Primär
 - Lehner: „Datenbanktechnologie für Data Warehouse Systeme“, dpunkt.Verlag, 2003, ca. 40.- Euro
 - Han/Kamber: „Data Mining“, Morgan Kaufmann, 2006. ca. 40.- Euro
- Weitere
 - Bauer/Günzel: „Data Warehouse Systeme“, dpunkt.Verlag, 2004
 - Dunham: „Data Mining“, Prentice Hall, 2002
 - Ester/Sander: „Knowledge Discovery in Databases“, Springer
 - Oehler: „OLAP: Grundlagen, Modellierung und betriebswirtschaftliche Grundlagen“, Hanser Verlag, 2000
- Übersichtsartikel
 - Chaudhuri, Dayal: „An Overview of Data Warehousing and OLAP Technology“, SIGMOD Record, 1997
 - Widom: „Research Problems in Data Warehousing“, CIKM, 1994
 - Ganti, Gherke, Ramakrishnan: „Mining Very Large Databases“, IEEE, 1999
 - Fayyad: „From Data Mining to Knowledge Discovery in Databases“, Ai Magazine, 1996

Wissensmanagement in der Bioinformatik

- **Schwerpunkte**
 - Graphen und Datenbanken, Semantic Web
 - Data Mining und Text Mining in der Bioinformatik
 - Datenintegration
 - Sensor/Geodaten
- **Laufend möglich**
 - Studienarbeiten
 - Diplomarbeiten
 - ... oft in Kooperation mit (Bioinformatik-)Firmen
 - ... oft in Kooperation mit (Bioinformatik-)Instituten



Weitere Termine

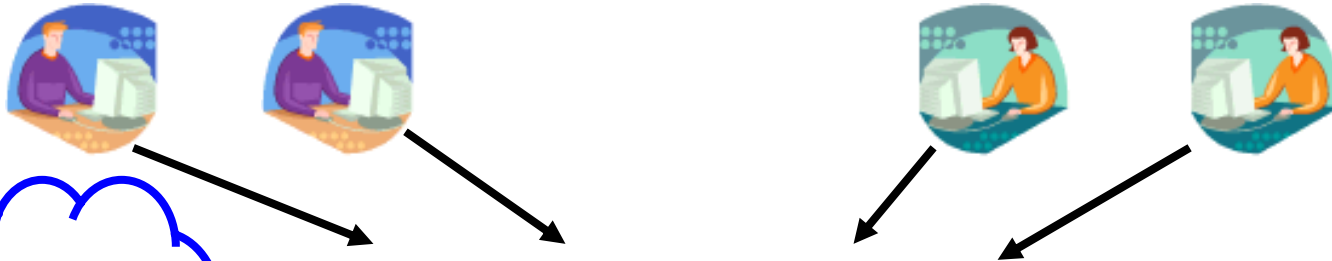
- Feiertage: 1.5., 17.5.
- Gastvortrag: 26.6.2007, Herr Eberl-Koschny, PSI AG:
"DWH Projekte in der Praxis"

Fragen ?

Fragen

- Diplominformatiker?
- Semester?
- DBS-I?
- Prüfung?

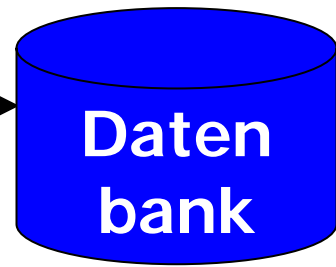
Bücher im Internet bestellen



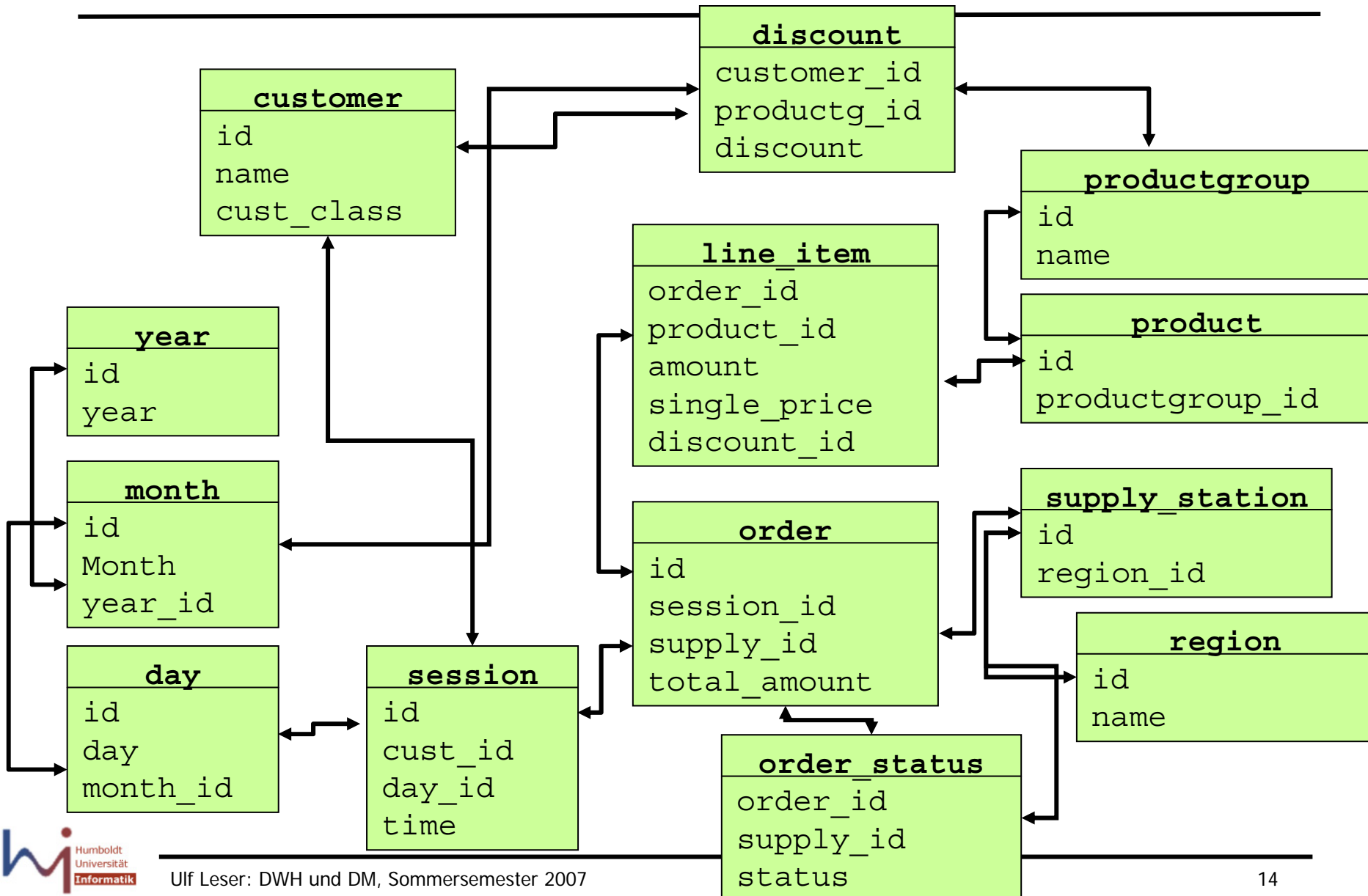
Backup
Durchsatz
Load-
balancing



Portfolio
Umsatz
Werbung

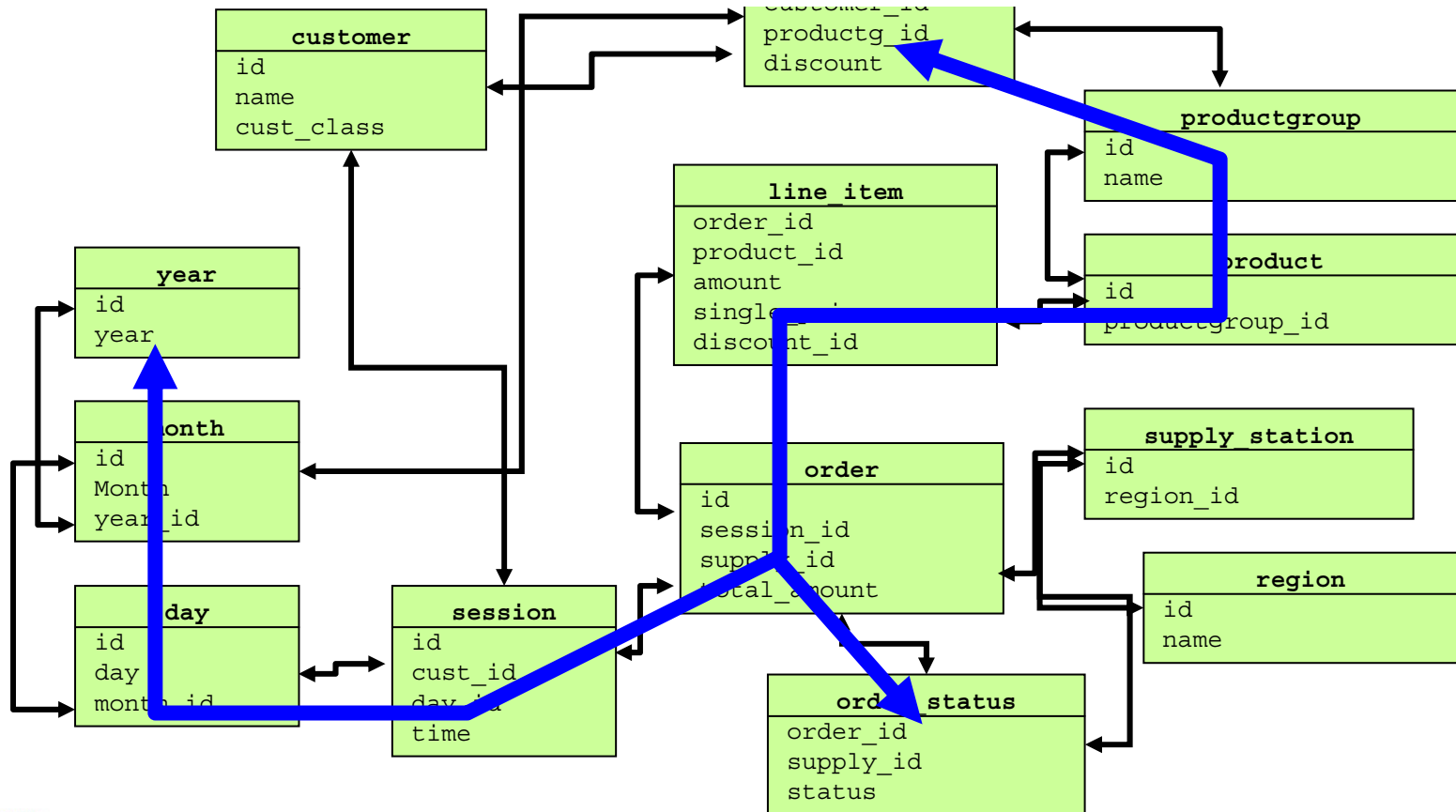


Die Datenbank dazu



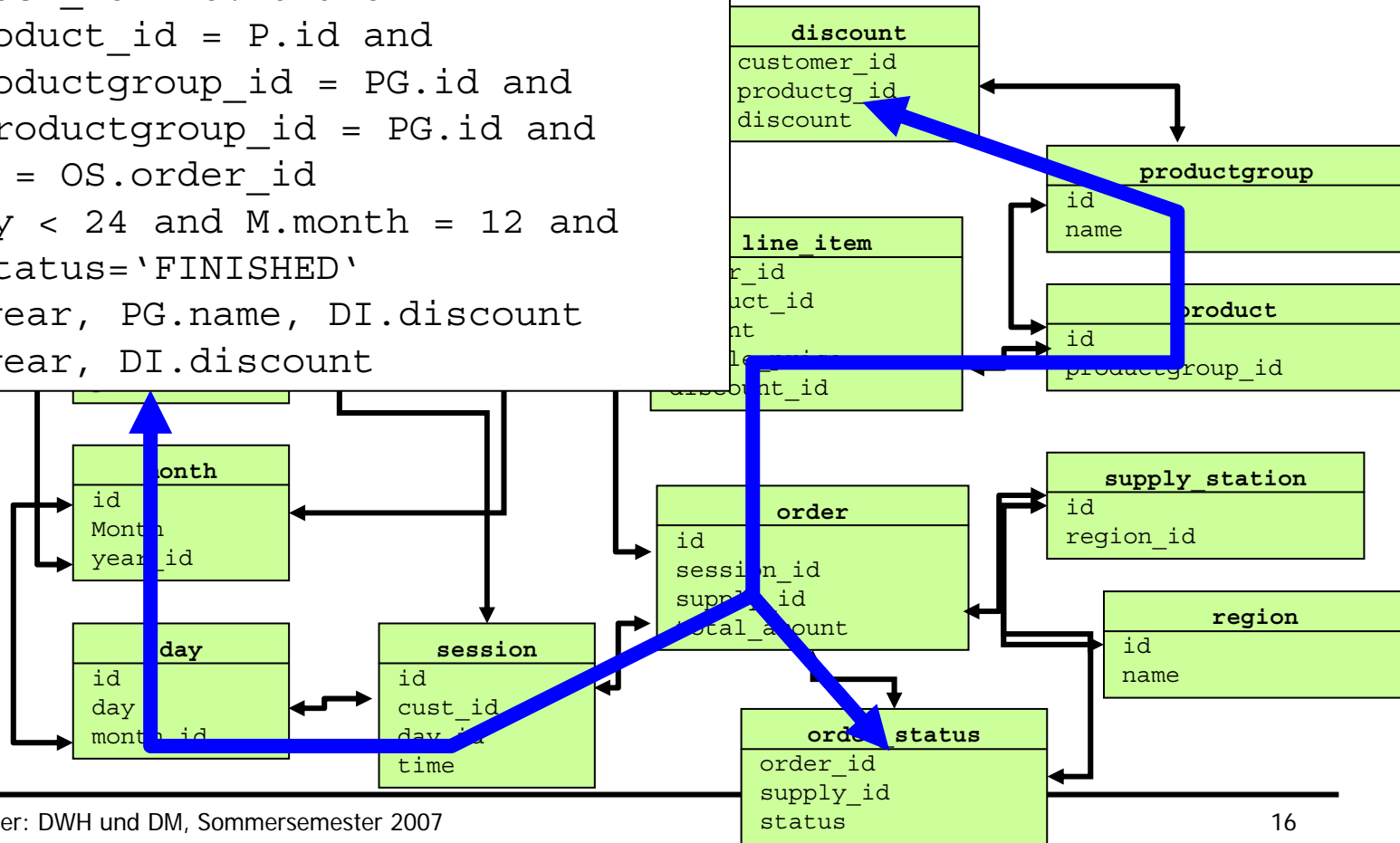
Fragen eines Marketingleiters

Wie viele abgeschlossene Bestellungen haben wir jeweils im Monat vor Weihnachten, aufgeschlüsselt nach Produktgruppen und Promotion?



Technisch

```
SELECT Y.year, PG.name, DI.discount, count(*)
FROM year Y, month M, day D, session S,
line_item I, order O, product P,
productgroup PG, discount DI,
order_status OS
WHERE M.year_id = Y.id and
D.month_id = M.id and
S.day_id = D.id and
O.session_id = S.id and
I.order_id = O.id and
I.product_id = P.id and
P.productgroup_id = PG.id and
DI.productgroup_id = PG.id and
O.id = OS.order_id
D.day < 24 and M.month = 12 and
OS.status='FINISHED'
GROUP BY Y.year, PG.name, DI.discount
ORDER BY Y.year, DI.discount
```



Ergebnis

```
SELECT Y.year, PG.name, DI.disc, count(*)
FROM year Y, month M, day D, session S,
line_item I, order O, product P,
productgroup PG, discount DI,
order_status OS
WHERE M.year_id = Y.id and
D.month_id = M.id and
S.day_id = D.id and
O.session_id = S.id and
I.order_id = O.id and
I.product_id = P.id and
P.productgroup_id = PG.id and
DI.productgroup_id = PG.id and
O.id = OS.order_id
D.day < 24 and M.month = 12 and
S = 'FINISHED'
PG.name, DI.discount
DI.discount
```

9 Joins

- year: 10 Records
- month: 120 Records
- day: 3650 Records
- session: 36.000.000
- order: 37.000.000
- line_item: 72.000.000
- order_status: 37.000.000
- product: 200.000
- productgroup: 100
- discount: 50

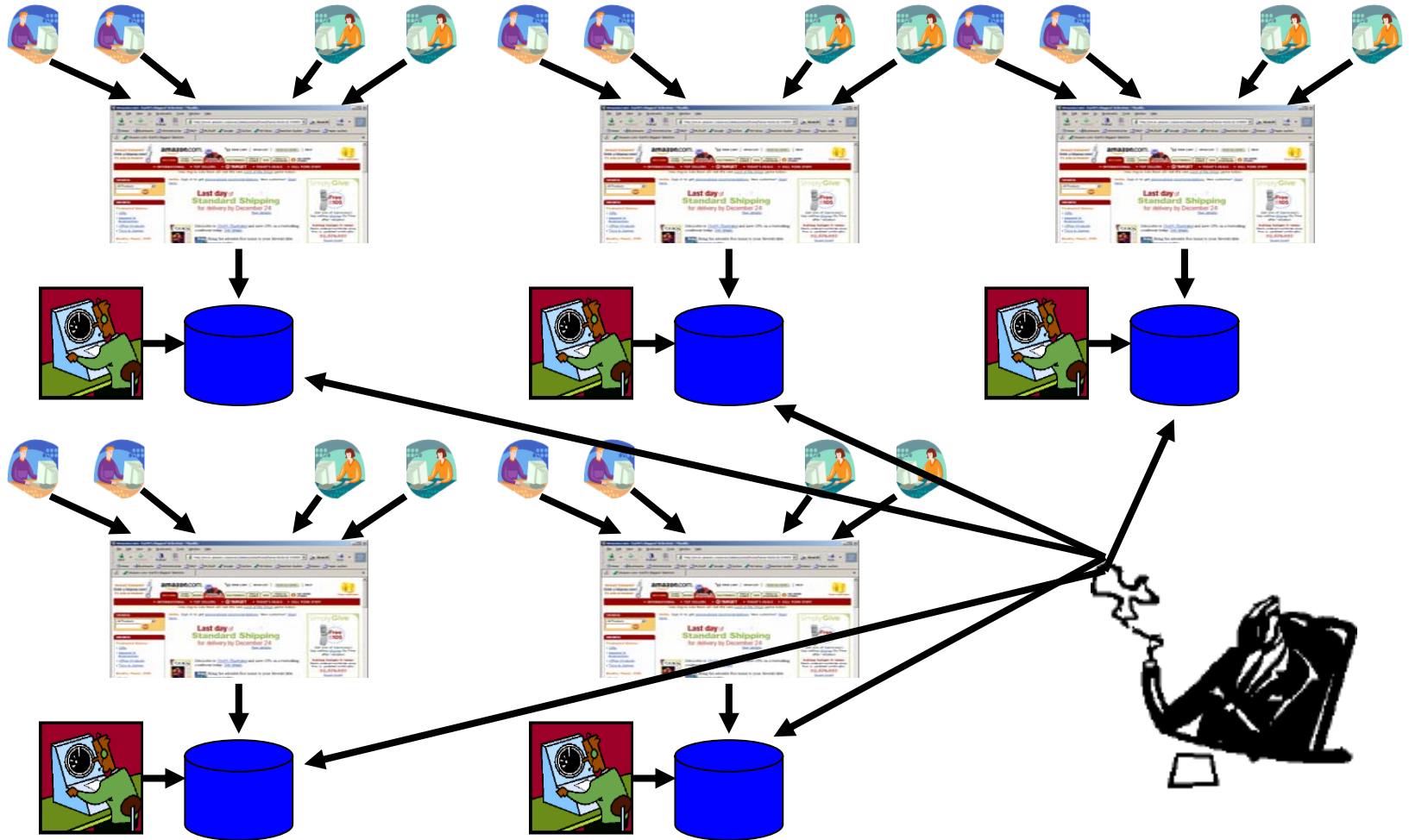
Problem!

- Schwierig zu optimieren (Join-Order)
- Ja nach Ausführungsplan riesige Zwischenergebnisse
- Ähnliche Anfragen – ähnlich riesige Zwischenergebnisse

In Wahrheit ...

- Es gibt noch:
 - Amazon.de
 - Amazon.fr
 - Amazon.it
 - ...
- Verteilte Ausführung
 - Count über Union mehrerer gleicher Anfragen in unterschiedlichen Datenbanken?

In Wahrheit ...



Technisch

```
CREATE VIEW christmas AS
```

```
SELECT      Y.year, PG.name, DI.disc, count(*) AS o_count
FROM        FR.year Y, FR.month M, FR.day D, FR.session S, ...
WHERE       M.year = Y.id and
...
ORDER BY    Y.year, DI.discount
```

```
UNION ALL
```

```
SELECT      Y.year, PG.name, DI.disc, count(*) AS o_count
FROM        EN.year Y, EN.month M, EN.day D, EN.session S, ...
WHERE       M.year = Y.id and
...
ORDER BY    Y.year, DI.discount
```

```
...
```

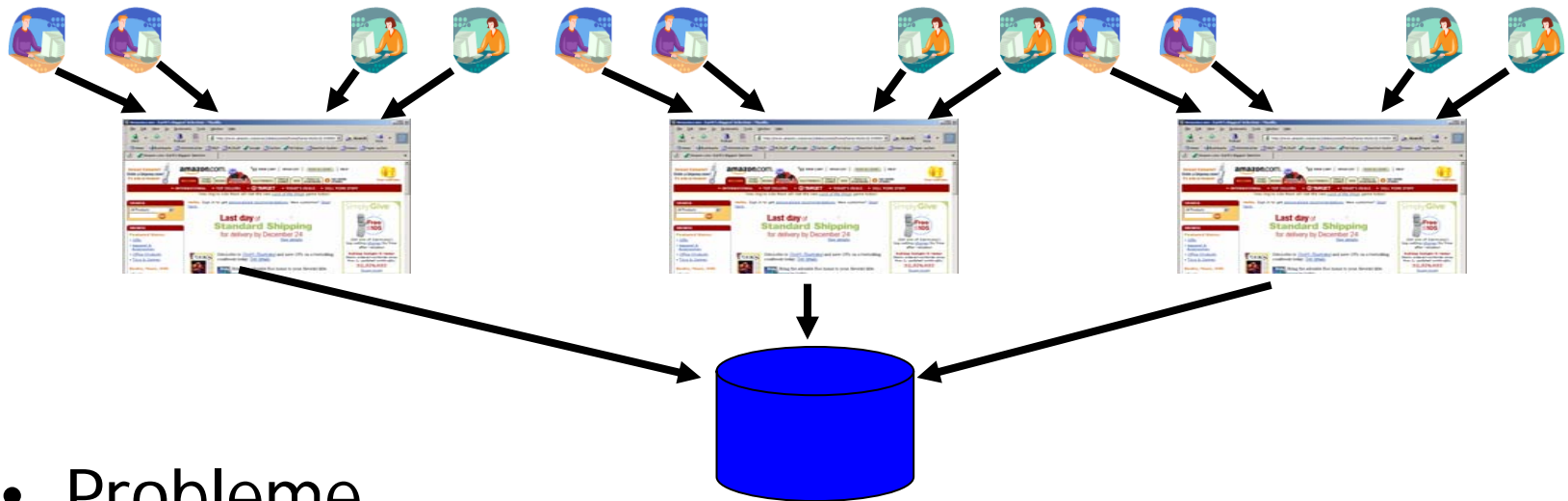
```
SELECT      year, name, disc, sum(o_count)
FROM        christmas
GROUP BY    year, name, disc
ORDER BY    year, disc
```

Probleme

- Count über Union über verteilte Datenbanken?
 - **Heterogenitätsproblem**
 - Quellen werden Schemata verändern
 - Länderspezifischer Eigenheiten (MWST, Versandkosten, Sonderaktionen, ...)
 - Oftmals verborgene Änderungen in der Semantik der Daten
- Berechnung der Zwischenergebnisse bei jeder Anfrage?
 - **Datenmengenproblem**
 - Erfordert Transport großer Datenmengen durchs Netz
 - Historische Sicht - **Datenmengen wachsen immer weiter**
 - Operative Systeme brauchen die historischen Daten nicht
 - Ziel: Frühes löschen (abgeschlossene Bestellungen)
 - Manager brauchen viele der operativen Daten nicht
 - Ziel: Alles aufheben

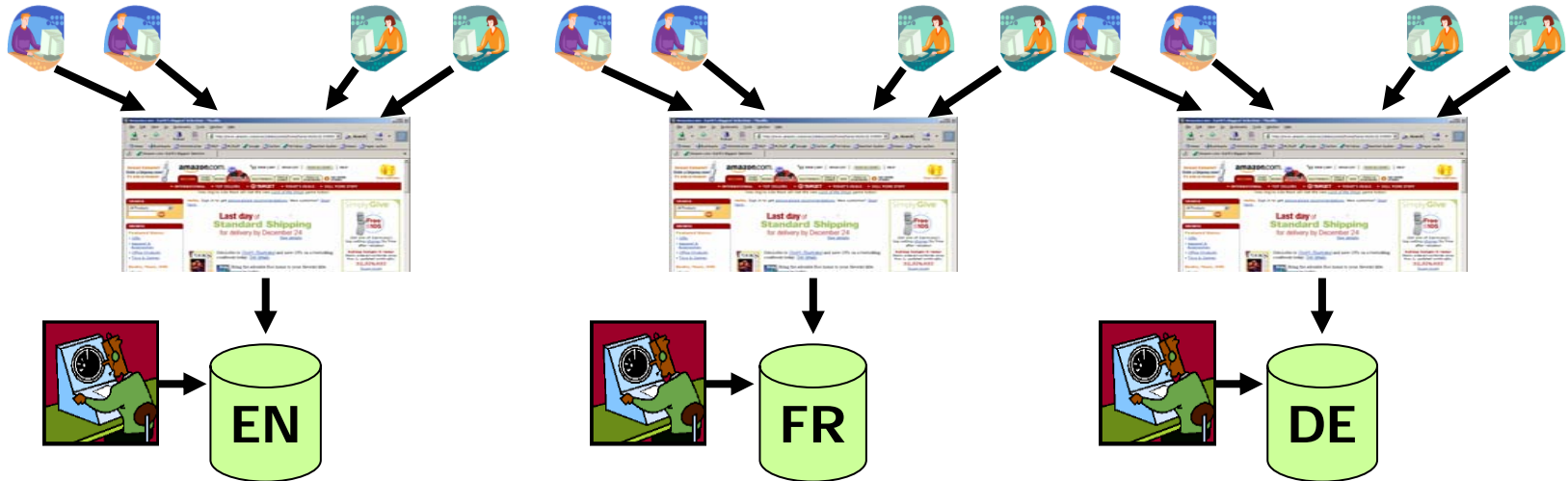
Lösung Heterogenitätsproblem?

Zentrale Datenbank



- Probleme
 - Zweigstellen schreiben übers Netz
 - **Lange Antwortzeiten** im operativen Betrieb
 - Datenmengenproblem bleibt

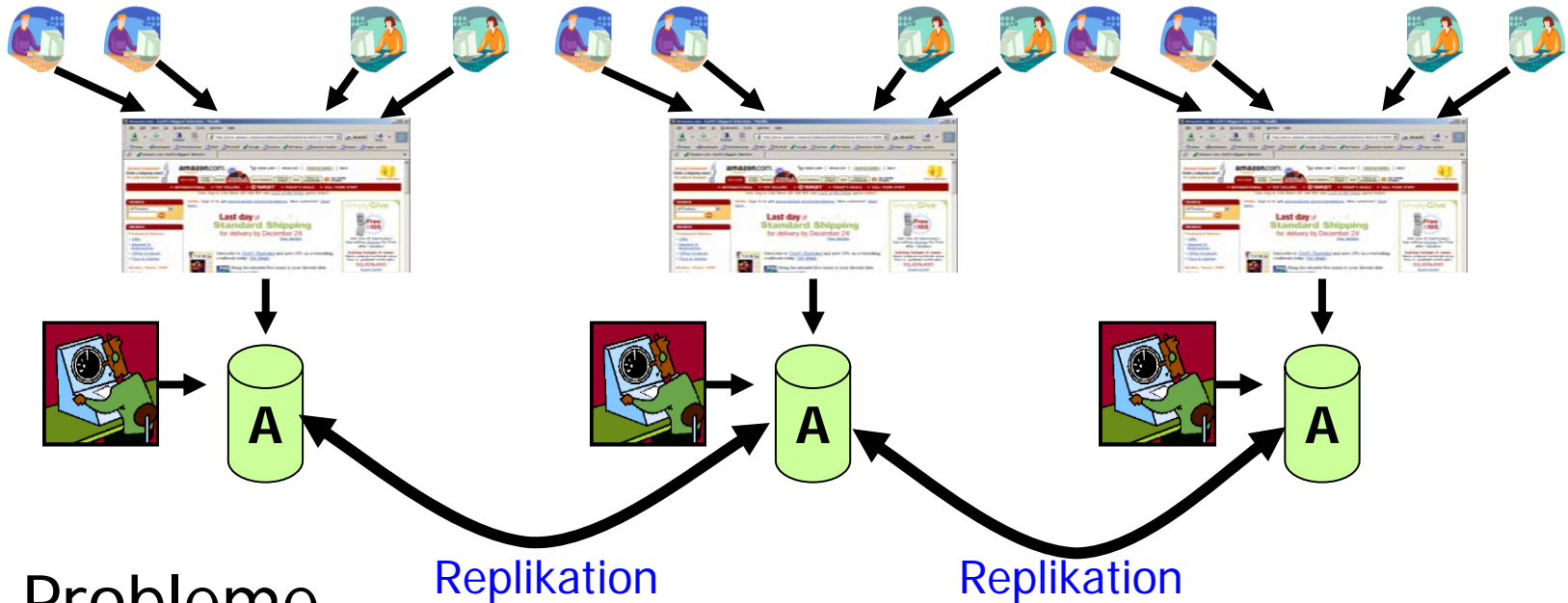
Lösung Anfragezeiten?



- Probleme

- Schnelle lokale Anfragen
- **Lange Antwortzeiten** für strategische Anfragen
- Heterogenitätsproblem bleibt

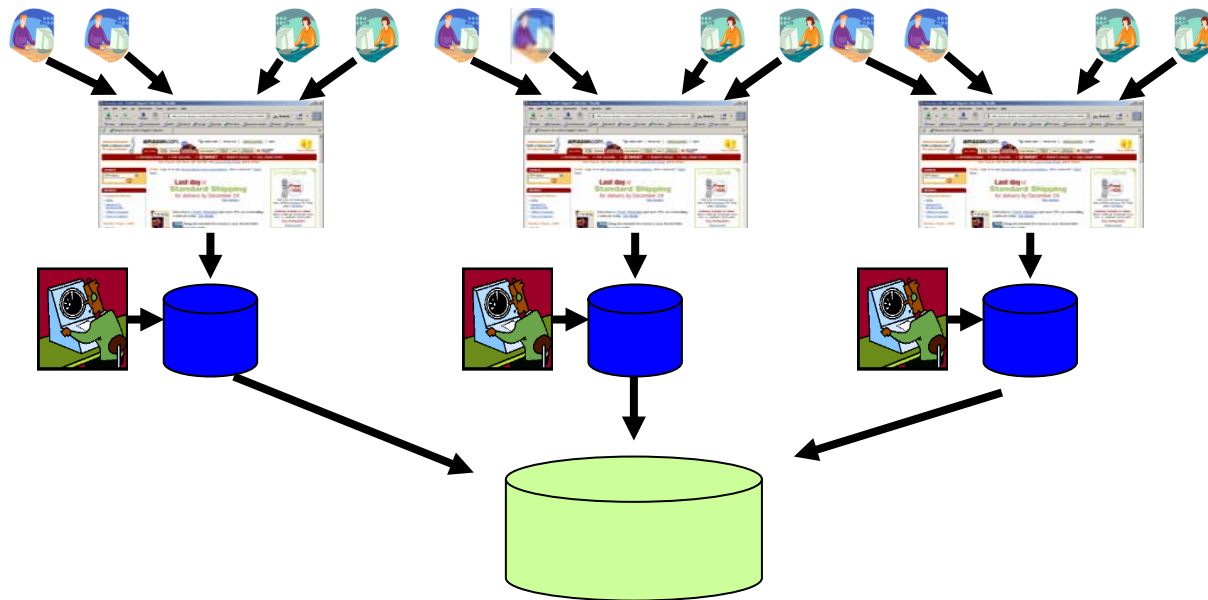
Lösung Datenmengenproblem?



- Probleme

- Lokale Anfrage arbeiten auf riesigen Tabellen
- Verzögerung im operativen Betrieb
- Lange Antwortzeiten für strategische Anfragen

Besser: Data Warehouse



- Redundante Datenhaltung
- Spezielle Modellierung
- Transformierte und selektierte Daten
- Asynchrone Aktualisierung

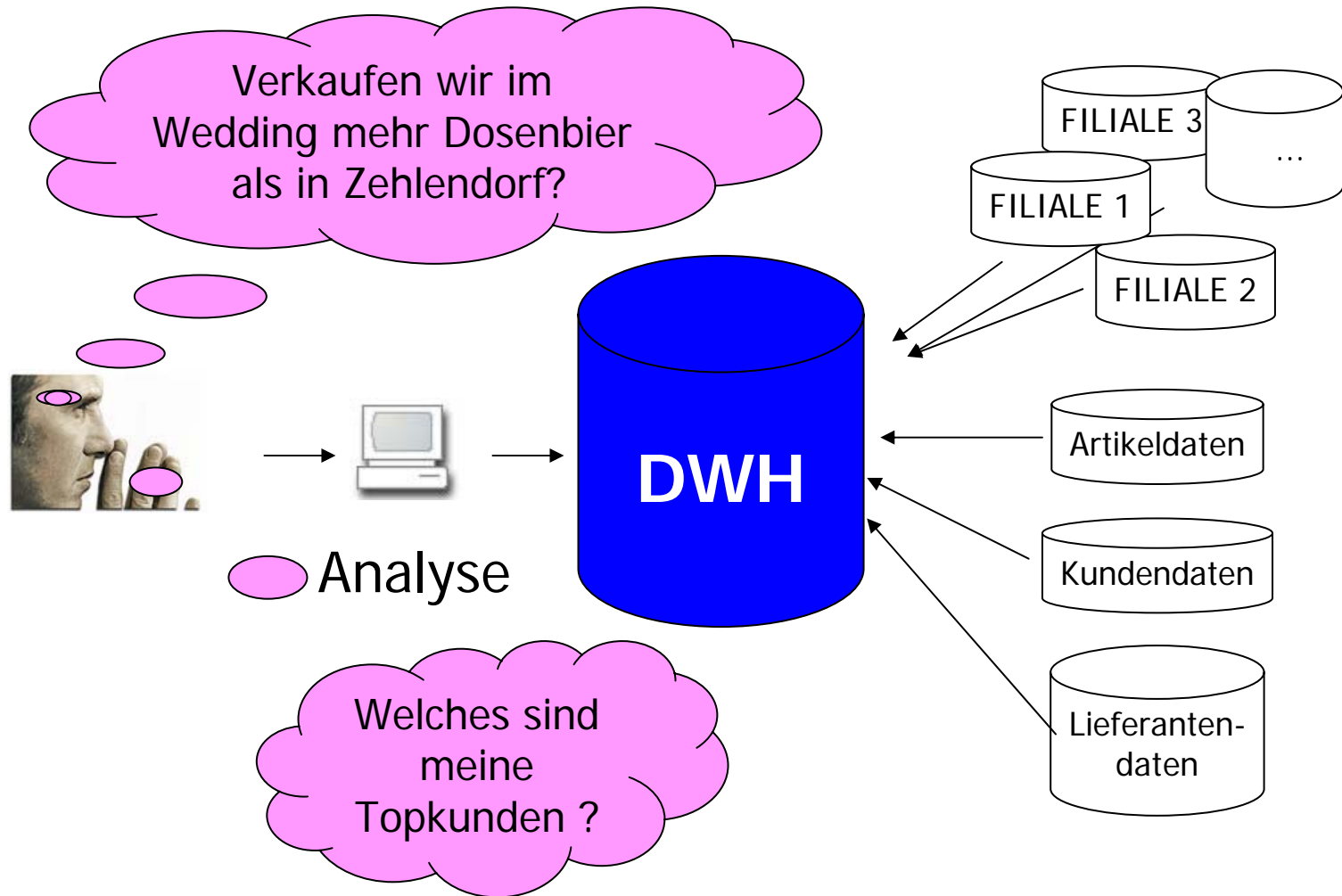
Inhalt dieser Vorlesung

- Szenario
- Definition & Geschichte & Abgrenzung
- Sichtweisen
- Typische Anwendungsgebiete
- Grosse Datenmengen und TPC-H

Beispielszenario

- Ein beliebiges Handelshaus : Spar, Extra, Kaufhof, ...
- Physikalische Datenverteilung
 - Viele Niederlassungen (bis zu mehrere tausend)
 - Noch mehr Registerkassen
- Aber: **Zentrale Planung, Beschaffung, Verteilung**
 - Strategische Unternehmensentscheidungen
 - Was wird wo und wie oft verkauft?
 - Was muss wann wohin **geliefert** werden?
 - Bedenke: Verderbliche Waren
 - Was werden wir in Zukunft verkaufen?
- Erfordert
 - **Zentrale Übersicht über Umsätze, Kunden, Produkte, ...**
 - Integration von Verkaufsdaten mit Lieferanten-, Kunden-, Produkt-, ... daten

Handels - DWH



DWH Datenquellen

- Lieferantendatenbanken
 - Produktinformationen: Packungsgrößen, Farben, ...
 - Lieferbedingungen, Rabatte, Lieferzeiten, ...
- Personaldatenbank
 - Zuordnung Kassenbuchung auf Mitarbeiter
 - Stundenabrechnung, Prämien
- Kundendatenbank
 - Kundenklassen; Premium, Normal, soziale Brennpunkte, ...
 - Persönliche Vorlieben & Historie
 - Kundenkarten, TESCO
- Weitere Vertriebswege (Multi-Channel)
 - Internet, Katalogbestellung, Verkaufclubs, ...
- ...

Definition DWH

- *A DWH is a subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management's decision* [Inm96]
 - Subj-orient.: Verkäufe, Personen, Produkte, etc.
 - Integrated: Erstellt aus vielen Quellen
 - Non-Volatile: Hält Daten unverändert über die Zeit
 - Time-Variant: Vergleich von Daten über die Zeit
 - Decisions: Wichtige Daten rein, unwichtige raus

Geschichte von DWH

- Managementinformationssysteme (MIS),
Decision Support Systeme (DSS),
Executive Information Systeme (EIS)
 - Seit den 60er Jahren
 - Feste, programmierte Reports
 - Redundante Datenhaltung sehr teuer – dadurch langsam, oft Downtimes für operative Systeme
 - Datenzugriff sehr schwierig: fehlende Vernetzung, Schnittstellen, Anwendungsneutralität nicht gegeben
 - Analysemethoden ungenügend (Data Mining)
 - Schwierig zu bedienen
- Teuer und unflexibel
- Schattendasein

Erfolg von DWH

- Top-Thema seit Mitte der 90er Jahre
- Voraussetzungen
 - Extreme Verbilligung von Plattenspeicherplatz
 - Relationale Modellierung: Anwendungsneutral
 - Graphische Benutzeroberflächen und Terminals
 - IT in allen Unternehmensbereichen (SAP R/3)
 - Vernetzung und DB Standardisierung (SQL)
- Aber
 - Vision der vollständigen Integration scheitert bisher (immer wieder aufs neue)
 - Soziale versus technische Aspekte

Abgrenzung

- Parallele Datenbanken
 - Blendet Heterogenität aus
 - Ausfallsicherheit und Performanceverbesserung
 - Technik zur Realisierung eines DWH
- Verteilte Datenbanken
 - Gewollte Verteilung als Mittel zur Lastverteilung
 - Keine physische Integration der Daten (Materialisierung)
- Föderierte Datenbanken
 - Höhere Autonomie und Heterogenität
 - Verteilung bleibt erhalten

Inhalt dieser Vorlesung

- Szenario
- Definition & Geschichte & Abgrenzung
- Sichtweisen
 - Betriebswirtschaftliche Sichtweise
 - Informatiker Sichtweise: OLAP versus OLTP
- Typische Anwendungsgebiete
- Grosse Datenmengen und TPC-H

Betriebswirtschaftliche Sicht

- Ein DWH
 - Ermöglicht **viele neue Fragen**
 - Verbessert viele Antworten erheblich
- ... durch ...
 - Zugriff auf **integrierte Daten**
 - Ohne DWH nur manuell (Programmierung) möglich
 - Übergreifende Analysen
 - Inhaltliche Verknüpfung der Daten
 - **Bessere Datenqualität**
 - Fehlerminimierung, Ergänzung, Plausibilitätschecks
 - Entfernung von Fehlern ist hier vertretbar, aber nicht in den operativen Systemen
 - Anreicherung mit **externen Daten**
 - Externe Kundenprofile, geographische Daten

Informatische Sicht

- Operative Systeme
 - Viele Benutzer
 - Kurze Transaktionen, **einfache Queries**
 - Echtzeitanforderungen
 - Kurzes Gedächtnis
 - Beispiel: Kassensysteme, Bankautomaten
 - **OLTP** (Online **Transaction** Processing)
- DWH
 - Wenige(r) Benutzer
 - **Komplexe Queries** mit langer Laufzeit, nur lesend
 - Zeitlich eher unkritisch
 - Historische Daten
 - Beispiel: Sortimentplanung, Kapazitätsplanung
 - **OLAP** (Online **Analytical** Processing)

OLTP Beispiel

Login

```
SELECT pw FROM kunde WHERE login=„...“  
UPDATE kunde SET last_acc=date, tries=0 WHERE
```

COMMIT

Willkommen

```
SELECT k_id, name FROM kunde WHERE login=„...“  
SELECT last_pur FROM purchase WHERE k_id=...
```

COMMIT

Bestellung

```
SELECT av_qty FROM stock WHERE p_id=...  
UPDATE stock SET av_qty=av_qty-1 where ...  
INSERT INTO shop_cart VALUES( o_id, k_id, ...
```

COMMIT

Best. löschen

```
DELETE FROM shop_cart WHERE o_id=...  
UPDATE stock SET av_qty=av_qty+1 where ...
```

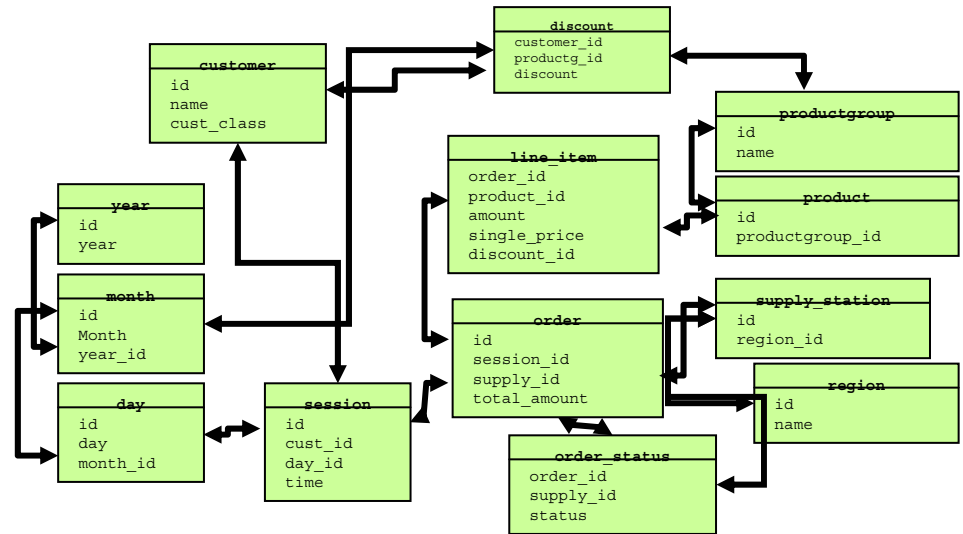
COMMIT

OLAP Beispiel

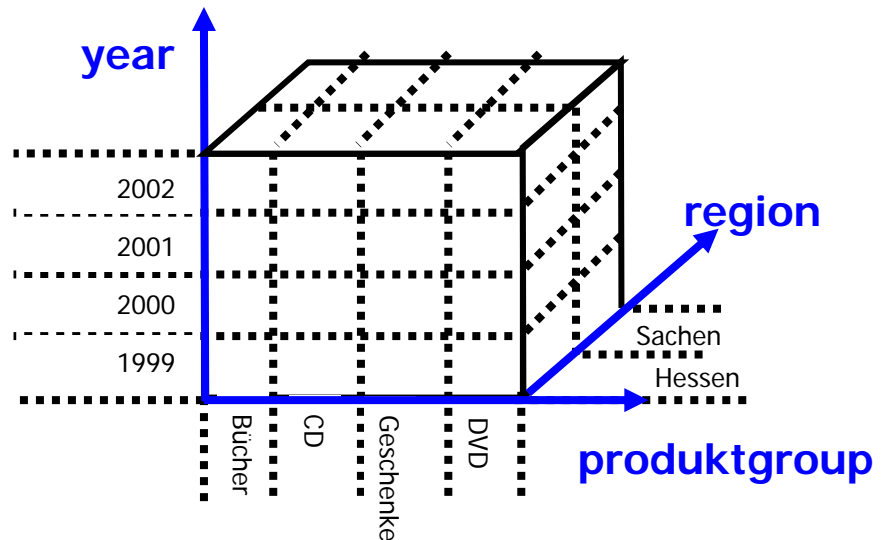
- Welche Produkte hatten im letzten Jahr im Bereich Bamberg einen Umsatzrückgang um mehr als 10%?
 - Welche Produktgruppen sind davon betroffen?
 - Welche Lieferanten haben diese Produkte?
- Welche Kunden haben über die letzten 5 Jahre eine Bestellung über 50 Euro innerhalb von 4 Wochen nach einem persönlichen Anschreiben aufgegeben?
 - Wie hoch waren die Bestellungen im Schnitt?
 - Wie hoch waren die Bestellungen im Vergleich zu den durchschnittlich. Bestellungen des jew. Kunden in einem vergleichbaren Zeitraum?
 - Lohnen sich Mailing-Aktionen?
- Haben Zweigstellen einen höheren Umsatz, die gemeinsam gekaufte Produkte zusammen stellen ?
 - Welche Produkte werden überhaupt zusammen gekauft – und wo?

Modellierung im DWH

- Modellierung in operativen Systemen:
Normalisierung



- Modellierung in DWH:
Dimensionen und Fakten



Ausfall kostet Millionen.
 Hochverfügbarkeit:
 24x7x52

OLAP versus OLTP

Ausfall ärgerlich

	OLTP	OLAP
Typische Operationen	Insert, Update, Delete, Select	Select Bulk-Inserts
Transaktionen	Viele und kurz	Lesetransaktionen
Typische Anfragen	Einfache Queries, Primärschlüsselzugriff, Schnelle Abfolgen von Selects/inserts/updates/deletes	Komplexe Queries: Aggregate, Gruppierung, Subselects, etc. Bereichsanfragen über mehrere Attribute
Daten pro Operation	Wenige Tupel	Mega-/ Gigabyte
Datenmenge in DB	Gigabyte	Terabyte
Eigenschaften der Daten	Rohdaten, häufige Änderungen	Abgeleitete Daten, historisch & stabil
Erwartete Antwortzeiten	Echtzeit bis wenige Sekunden	Minuten
Modellierung	Anwendungsorientiert	Themenorientiert
Typische Benutzer	Sachbearbeiter	Management

Inhalt dieser Vorlesung

- Szenario
- Definition & Geschichte & Abgrenzung
- Sichtweisen
- **Typische Anwendungsgebiete**
- Grosse Datenmengen und TPC-H

Wal Mart

- Unternehmensweites Data Warehouse
- Größe
 - 100.000.000 Kunden pro Woche, 3.600 Geschäfte
 - „Wal-Mart ... will expand its data-warehouse system, which handles 50,000 queries a week, **from 7.5 to 24 terabytes.**“ [Info. Week, 1997]
 - „**At 70 terabytes and growing ...**“ [Wes, 2000]
 - „... **300 Terabyte** Datenbank“ [Computerzeitung, 16/2003]
 - Ca. **480 Terabyte** (11/04, New York Times)
 - Datenbank: Teradata, NCR
- Täglich bis zu 20.000 DW-Anfragen
 - Die meisten Analysen laufen auf Shop-, nicht auf Kundenebene
 - Überprüfung des Warensortiments zur Erkennung von Ladenhütern oder Verkaufsschlägern
 - Standortanalyse, Rentabilität von Niederlassungen
 - Untersuchung der Wirksamkeit von Marketing-Aktionen

SBC Communications

- Amerikanisches Telekommunikationsunternehmen
- 57.000.000 Telefon / DSL Kunden
- Teradata Datenbank
- „Derzeit größtes DWH der Welt“ (4/2004)
 - Ca. 360 Terabyte
 - 12.000 Tabellen (?)
 - 300.000 Logins pro Tag
 - Massiv parallele 500 Prozessor-Maschine
- Teradata
 - Erste Terabyte Warehouse: 1990
 - Testsysteme laufen heute (5/2004) im Petabyte Bereich
 - Erwartung für den kommerziellen Betrieb: 2006

CRM: Customer Relationship Management

- Vertriebswege, Vertriebsorganisationen
- Kundenkäufe, Präferenzen, Kontaktdaten, Reklamationen, Bonität
- Wer sind Premiumkunden -> Sonderbehandlung
- Beratungssysteme, Personalisierung, Mailings

Amazon

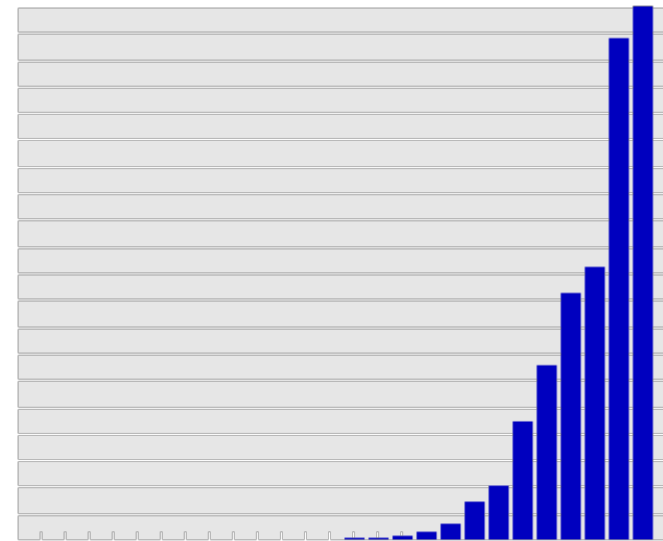
- Customers who bought X also bought Y
- Personalisierte Empfehlungen
- Neuerscheinungen

Weitere Anwendungsgebiete

- Controlling & Accounting
 - Kostenstellen, Kostenträger, Kostenarten
 - Bilanzierung
 - Plan – Ist Zahlen, Szenarios
 - Kennzahlensysteme
- Logistik
 - Flottenmanagement
 - Disposition
 - Tracking
- Finanzen
 - Kreditkartenanalyse
 - Risikoanalysen
 - Fraud Detection
- Health Care
 - Studienüberwachung
 - Wirkstoffanalyse

Was sind große Datenbanken?

- Wirklich groß: Astronomie, Teilchenphysik, CIA (?)
- Beispiel Google
 - [ct, 1/2003, Heise Verlag]
10 Terabyte Rohdaten
15.000 PC, 1 Petabyte (Redundanz)
 - [Technology Review, 4/2004, Heise Verlag]
>100.000 PC, **4 Petabyte**
- EMBL (Humanes Genom)
 - Release 104, 2004, als Flatfile:
Ca. 100 Gigabyte
 - Seitdem ...











EMBL, Stand 1 März 2007

TPC.ORG: Benchmarks

- Vergleich der Leistungsfähigkeit von Datenbanken
 - TPC-C OLTP Benchmark
 - TPC-H Ad-Hoc Decision Support (variable Anfragen)
 - TPC-R Reporting Decision Support (feste Anfragen)
 - TPC-W eCommerce Transaktionsverarbeitung
 - [TPC-D Abgelöst durch H und R]
 - TPC-E Ganz frischer OLTP Benchmark
- TPC-H
 - Vorgegebene Schemata (Lieferwesen)
 - Schema-, Anfrage- und Datengeneratoren
 - Unterschiedliche DB-Größen
 - 100 GB - 300 GB - 1 TB - 3 TB

TPC-H Ergebnisse 100 GB

100 GB Results									
Rank	Company	System	QphH	Price / QphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		HP AlphaServer ES45 Model 68/1000	5,578	404 US \$	07/15/02	Oracle 9iR2 w/Real Application Cluste	HP Tru64 Unix V5.1A/IPK	10/09/02	Y
2		IBM eServer x350 with DB2 UDB	2,960	336 US \$	06/20/02	IBM DB2 UDB 7.2	Turbolinux 7 Servers	02/01/02	Y
3		SGI 1450 Server with DB2 UDB EEE v7.2	2,733	347 US \$	10/31/01	IBM DB2 UDB EEE 7.2	Linux 2.4.3	05/11/01	Y
4		HP ProLiant DL760 X900	1,933	89 US \$	12/31/02	Microsoft SQL Server 2000 Enterprise Edition	Microsoft Windows .NET Enterprise Server	07/31/02	N
5		ProLiant 8000-X700-8P	1,699	161 US \$	08/01/00	Microsoft SQL 2000	Microsoft Windows 2000	07/21/00	N
6		HP ProLiant DL580 G2	1,695	82 US \$	06/26/02	Microsoft SQL Server 2000 Enterprise Edition	Microsoft Windows 2000 Advanced Server	06/26/02	N
7		e-@ction Enterprise Server ES5085R	1,669	169 US \$	01/31/01	Microsoft SQL Server 2000	Microsoft Windows 2000	12/22/00	N
						Microsoft SQL	Microsoft		











Document: Done (2.353 secs)

Quelle: <http://www.tpc.org>, Januar 2003

TPC-H Ergebnisse 100 GB

All Results
 Clustered Results
 Non-Clustered Results
 Currency: All

100 GB Results







Rank	Company	System	QphH	Price/QphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		HP ProLiant DL585G2 4P	19,323	10.67 US \$	01/16/07	Microsoft SQL Server 2005 x64 Enterprise Edt. SP1	Microsoft Windows Server 2003 Enterprise x64 Edition SP1	09/25/06	N
2		PowerEdge 6950/2.8GHz/2MB	17,179	10.02 US \$	12/04/06	Microsoft SQL Server 2005 Enterprise x64 Edition	Microsoft Windows Server 2003 Enterprise x64 Edition	10/23/06	N
3		HP ProLiant DL580G4 4P	17,120	7.91 US \$	11/22/06	Microsoft SQL Server 2005 x64 Enterprise Edt. SP1	Microsoft Windows Server 2003 Enterprise x64 Edition SP1	09/05/06	N
4		PowerEdge 6800/3.4GHz/16MB	16,320	13.40 US \$	08/28/06	Microsoft SQL Server 2005 Enterprise Edt (x64)	Microsoft Windows Server 2003 Enterprise x64 Edition	08/01/06	N
5		PowerEdge 2900/2.66GHz/8MB	15,723	9.61 US \$	12/31/06	Microsoft SQL Server 2005 x64 Enterprise Edt. SP1	Microsoft Windows Server 2003 Enterprise x64 Edition SP1	11/27/06	N
6		PowerEdge 6950/2.8GHz/2MB	14,923	6.04 US \$	12/04/06	Microsoft SQL Server 2005 Enterprise x64 Edition	Microsoft Windows Server 2003 Enterprise x64 Edition	10/23/06	N
7		HP ProLiant ML570G4 4P	14,242	10.88 US \$	11/22/06	Microsoft SQL Server 2005 Enterprise x64 Edition	Microsoft Windows Server 2003 Enterprise x64 Edition SP1	05/22/06	N
8		HP ProLiant DL585 G1 4P	12,600	7.67 US \$	11/07/05	Microsoft SQL Server 2005 Enterprise x64 Edition	Microsoft Windows Server 2003 Enterprise x64 Edition	11/04/05	N
9		IBM eServer 325	12,216	70.68 US \$	11/08/03	IBM DB2 UDB 8.1	Suse Linux Enterprise Server 8	07/29/03	Y
10		PowerEdge PE6800/3.0/64GB	11,529	11.18 US \$	05/01/06	Microsoft SQL Server 2005 Enterprise x64 Edition	Microsoft Windows Server 2003 Enterprise x64 Edition	05/01/06	N

300 GB Results

Quelle: <http://www.tpc.org>, März 2007

TPC-H Ergebnisse 3000 GB












3,000 GB Results

Rank	Company	System	QphH	Price / QphH	System Availability	Database	Operating System	Date Submitted	Cluster
1	 HP invent	HP 9000 Superdome Enterprise Server	27,094	240 US \$	10/30/02	Oracle 9i Database Enterprise Edition v9.2.0.2.0	HP UX 11.i 64-bit	10/04/02	N
2	 Sun microsystems	Sun Fire[™] 15K Server with Oracle9i R2	23,813	237 US \$	10/30/02	Oracle 9i R2 Enterprise Edition	Sun Solaris 9	06/26/02	N
3	 HP invent	Compaq ProLiant DL760 X900-128P	21,053	291 US \$	06/20/02	IBM DB2 UDB 7.2	Microsoft Windows 2000 Advanced Server	02/06/02	Y
4	 Teradata a division of  NCR	WorldMark 5250	18,803	989 US \$	07/27/01	Teradata V2R4.1	MP-RAS 3.02.00	10/09/01	Y
5	 HP invent	HP 9000 Superdome Enterprise Server	17,908	569 US \$	05/15/02	Oracle 9i Database Enterprise Edition	HP UX 11.i 64-bit	01/28/02	N

Quelle: <http://www.tpc.org>, Januar 2003

TPC-H Ergebnisse 3000 GB

3,000 GB Results

Rank	Company	System	QphH	Price/QphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		HP BladeSystem ProLiant BL25p cluster 64p DC	110,576	37.80 US \$	06/08/06	Oracle Database 10g R2 Enterprise Edt w/Partitioning	Red Hat Enterprise Linux 4 ES	06/08/06	Y
2		Sun Fire[TM] E25K server	105,430	54.87 US \$	01/27/06	Oracle Database 10g R2 Enterprise Edt w/Partitioning	Sun Solaris 10	01/27/06	N
3		IBM eServer p5 595	100,512	53.00 US \$	03/01/06	Oracle 10g Enterprise Ed R2 w/ Partitioning	IBM AIX 5L V5.3	09/19/05	N
4		HP Integrity Superdome - Itanium2/1.6 GHz-64p/64c	71,847	56.00 US \$	01/18/06	Oracle Database 10g R2 Enterprise Edt w/Partitioning	HP UX 11.i V2 64 bit	07/18/05	N
5		Sun Fire(TM) E25K Server	59,435	100.66 US \$	07/27/05	Oracle Database 10g Enterprise Edition	Sun Solaris 10	01/27/05	N
6		IBM eServer xSeries 346	54,465	32.34 US \$	08/15/05	IBM DB2 UDB 8.2	Suse Linux Enterprise Server 9	05/18/05	Y
7		HP Integrity Superdome - Itanium2/1.5 GHz-64p/64c	45,247	108.78 US \$	03/25/04	Oracle Database 10g Enterprise Edition	HP UX 11.i 64-bit	09/25/03	N
***		PRIMEPOWER 2500	34,345	146.71 US \$	02/22/04	Oracle Database 10g Enterprise Edition	Sun Solaris 9	08/26/03	N
8		HP Integrity Superdome - Itanium2/1.6 GHz-32p/32c	30,956	75.16 US \$	05/05/06	Microsoft SQL Server 2005 Enterprise Edt SP1	Microsoft Windows Server 2003 Datacenter Ed.(64-bit)SP1	11/07/05	N
9		Unisys ES7000/one Enterprise Server (16P)	30,013	37.83 US \$	09/08/06	Microsoft SQL Server 2005 Enterprise Itanium Ed.	Microsoft Windows Server 2003 Datacenter Itanium Ed SP1	07/18/06	N
10		Sun Fire[TM] 15K server	28,948	184.32 US \$	04/30/03	Oracle 9i R2 Enterprise Edition	Sun Solaris 9	04/07/03	N

Quelle: <http://www.tpc.org>, März 2007