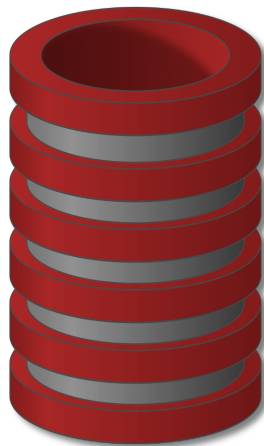


Announcements

- Assignments:
 - HW#3: due today
 - here, homework box
 - SmartSite ☹
- Project #2
 - Focus is XML to relational mapping
 - (... possible ECS-199 topics ...)
- Midterm
 - Wed 2/19
 - Recursion (Datalog, PostgreSQL)
 - XML: DTDs, XPath, XQuery
 - Relation Design Theory (BCNF, 4NF)
 - OLAP

ECS-165B

1



On-Line Analytical Processing (OLAP)

Introduction

Two broad types of database activity

■ OLTP – Online Transaction Processing

- Short transactions
- Simple queries
- Touch small portions of data
- Frequent updates

traditional

■ OLAP – Online Analytical Processing

- Long transactions
- Complex queries
- Touch large portions of the data
- Infrequent updates

decision support

More terminology

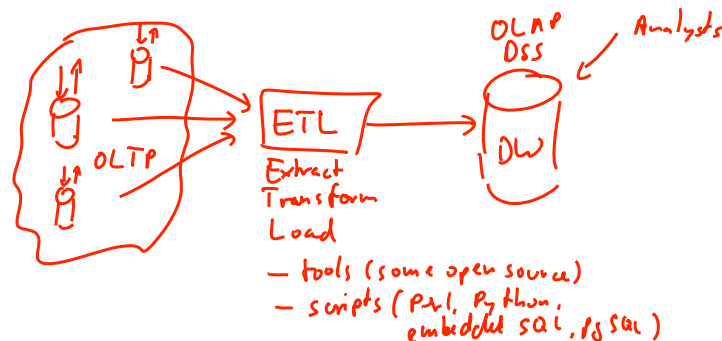
*195 : "OLAP-SQL" → SQL
195 : Skyline-Query →
199 : Schemas → BCNF/4NF*

■ Data warehousing

Bring data from operational (OLTP) sources into a single "warehouse" for (OLAP) analysis

■ Decision support system (DSS)

Infrastructure for data analysis
E.g., data warehouse tuned for OLAP



“Star Schema”

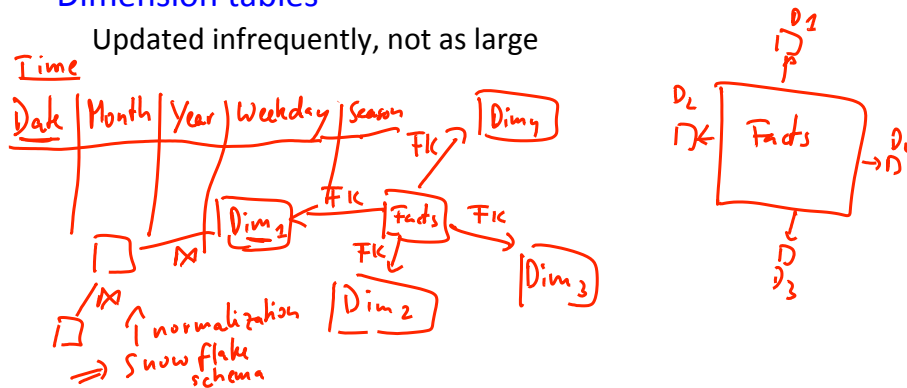
OLAP: Intro

Fact table

Updated frequently, often append-only, very large

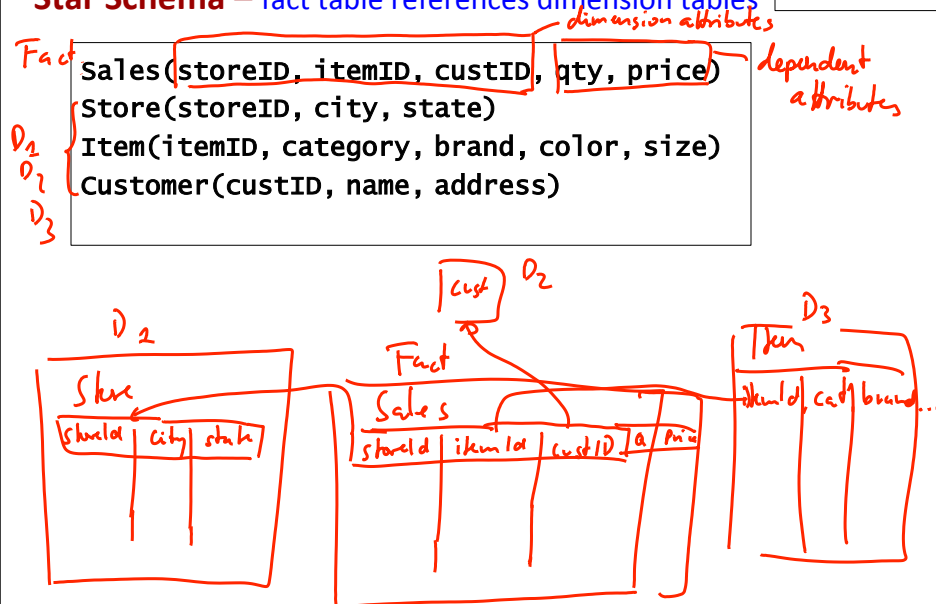
Dimension tables

Updated infrequently, not as large



Star Schema – fact table references dimension tables

OLAP: Intro



OLAP queries

OLAP: Intro

```

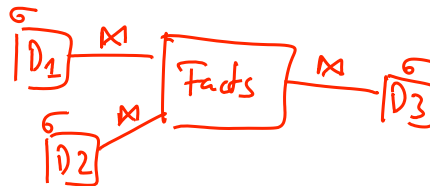
Sales(storeID, itemID, custID, qty, price)
Store(storeID, city, state)
Item(itemID, category, brand, color, size)
Customer(custID, name, address)

```

select Join → Filter → Group → Aggregate

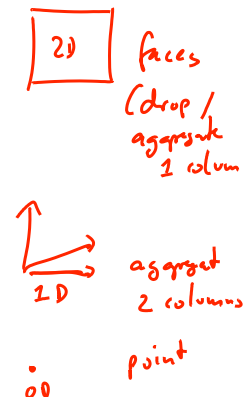
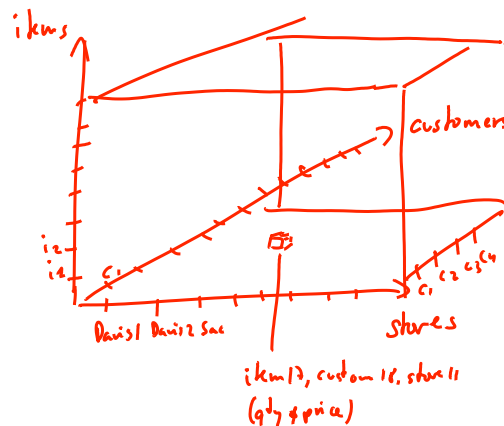
Performance

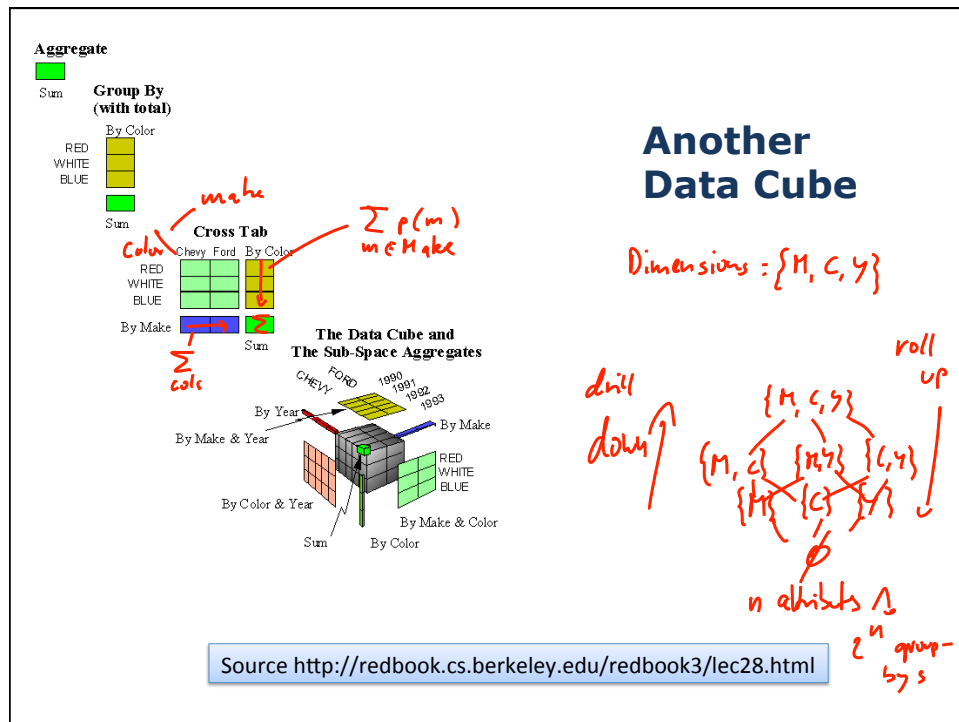
- Inherently very slow:
 special indexes, query processing techniques
- Extensive use of materialized views

**Data Cube (a.k.a. multidimensional OLAP)**

OLAP: Intro

- Dimension data forms axes of “cube”
- Fact (dependent) data in cells
- Aggregated data on sides, edges, corner





Fact table uniqueness for data cube

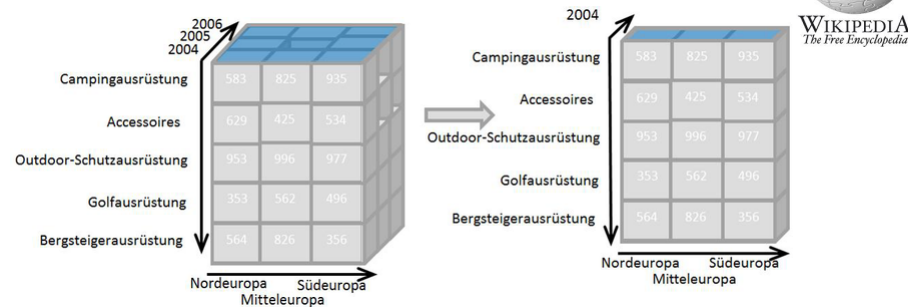
OLAP: Intro

```
Sales(storeID, itemID, custID, qty, price)
```

- If dimension attributes not key, must aggregate
- **Date** can be used to create key
Dimension or dependent?

OLAP Slicing

*Select ... from ...
when Year = 2004*

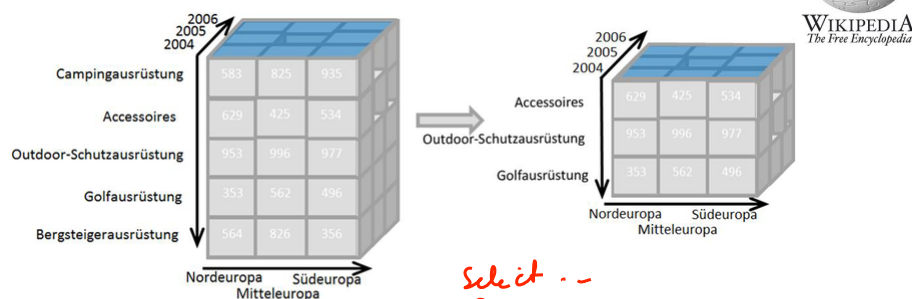


Slicing is the act of picking a rectangular subset of a cube by **choosing a single value for one of its dimensions**, creating a new cube with one fewer dimension. Here: The sales figures of all sales regions and all product categories of the company in the year 2004 are "sliced" out of the data cube.

ECS-165B

11

OLAP Dicing



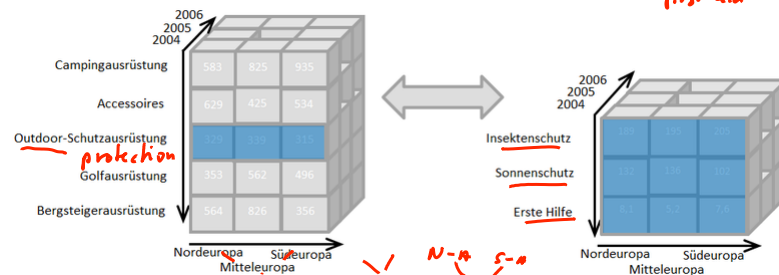
*Select ...
from ...
where ...*

Dice: The dice operation produces a subcube by allowing the analyst to **pick specific values of multiple dimensions**. Here, the new cube shows the sales figures of a limited number of product categories (the time and region dimensions cover the same range as before.)

ECS-165B

12

OLAP Roll-Up and Drill-Down



Drill Down allows the user to navigate among levels of data ranging from the most summarized (up) to the most detailed (down).

Here: The analyst moves from the summary category "Outdoor-Schutzausrüstung" to see the sales figures for the individual products.

Roll-up: A roll-up involves summarizing the data along a dimension. The summarization rule might be computing totals along a hierarchy or applying a set of formulas such as "profit = sales - expenses".

ECS-165B

13

Drill-down and Roll-up "zoom-in"

OLAP: Intro

Examining summary data, break out by dimension attribute

```

select state, brand, Sum(qty*price)
From Sales F, Store S, Item I
where F.storeID = S.storeID And F.itemID =
       I.itemID
Group By state, brand
  
```

category

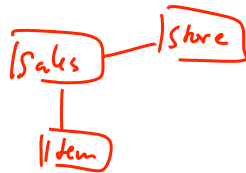
OLAP: Intro

Drill-down and **Roll-up**Examining data, summarize by dimension attribute

```

select state, brand, Sum(qty*price)
From Sales F, Store S, Item I
where F.storeID = S.storeID And F.itemID =
      I.itemID
Group By state, brand

```



OLAP: Intro

SQL Constructs

with Cube **and** with Rollup

```

select dimension-attrs, aggregates
From tables
Where conditions
Group By dimension-attrs with cube

```

$S = \{A, B, C\}$
 $2^S = P(S) = \{\phi, \{A\}, \{B\}, \{C\}, \{A, B\}, \{B, C\}, \{A, C\}, \{A, B, C\}\}$
 ϕ
 $\begin{matrix} A & B & C \\ A & B & A & C & B & C \end{matrix}$
 $A & B & C$

\downarrow drill down \uparrow roll-up

Add to result: faces, edges, and corner of cube using NULL values

SQL Constructs

with cube **and** with rollup

```
select dimension-attrs, aggregates
From tables
where conditions
Group By dimension-attrs with rollup
```

For hierarchical dimensions, portion of **with cube**

*Instead of 2^n subqueries (with cube) for
n dimensions, only $n+1$ are used.*

Two broad types of database activity

- OLTP – Online Transaction Processing

- Short transactions
- Simple queries
- Touch small portions of data
- Frequent updates

- OLAP – Online Analytical Processing

- Long transactions
- Complex queries
- Touch large portions of the data

- Star schemas
- Data cubes
- **with cube and with rollup**
- Special indexes and query processing techniques