On-Line Analytical Processing (OLAP)

Introduction

Jennifer Widom

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
  - Infrequent updates

Jennifer Widom

More terminology

- ## 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"

🔲 **Fact table**

Updated frequently, often append-only, very large

Sales transactions, course enrollments, page views

🔲 **Dimension tables**

Updated infrequently, not as large

- Stores, items, customers
- Students, courses
- Web pages, users, advertisers

Star Schema – fact table references dimension tables

(F)
(D)

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

Dimension attributes

Dependent attributes

Store

SID	...

Sales

SID	iID	CID	§	▯

Customer

CID	...

Item

iID	...

Jennifer Widom
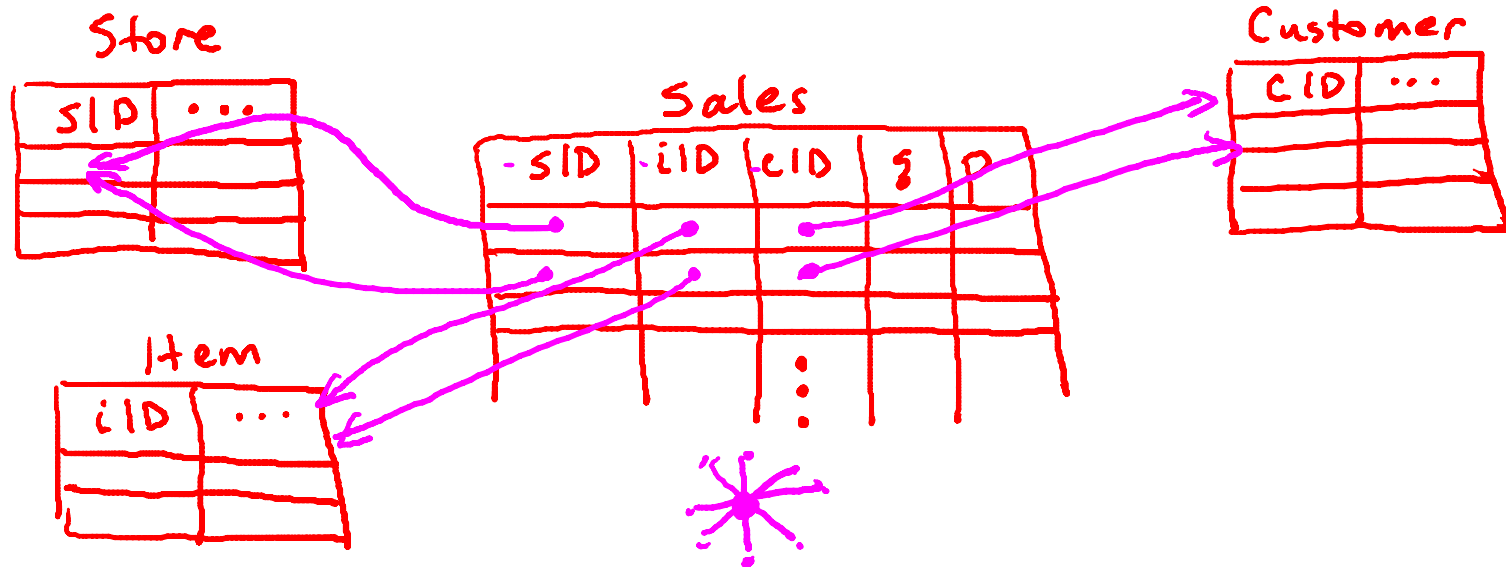
OLAP queries

```
Sales(storeID, itemID, custID, qty, price)
Store(storeID, city, state)
Item(itemID, category, brand, color, size)
Customer(custID, name, address)
```
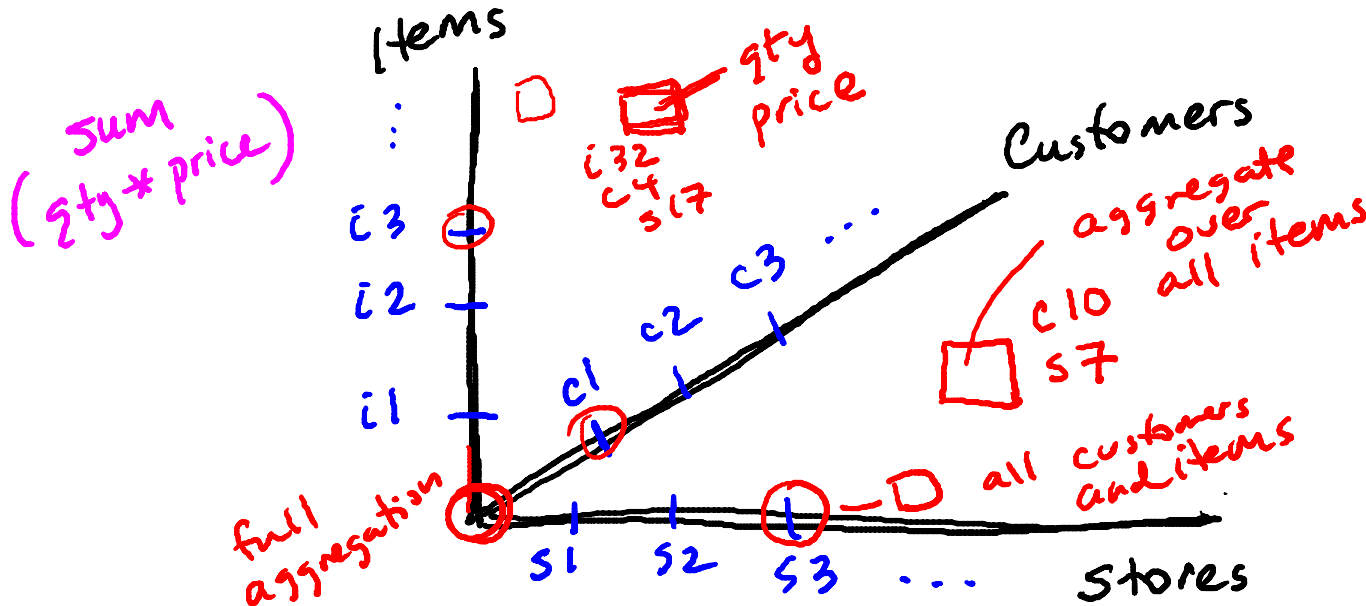
Join → Filter → Group → Aggregate

Performance

– Inherently very slow:
   special indexes, query processing techniques

– Extensive use of materialized views

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

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

Jennifer Widom

Fact table uniqueness for data cube

date

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

- If dimension attributes not key, must aggregate

- **Date** can be used to create key

  Dimension or dependent?

Drill-down and Roll-up

Drill-down and Roll-up

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

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
```

SQL Constructs

With Cube **and** With Rollup

Select **dimension-attrs**, **aggregates**
From **tables**
Where **conditions**
Group By **dimension-attrs** With Cube

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`

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

  - Star schemas
  - Data cubes
  - `With Cube` and `With Rollup`
  - Special indexes and query processing techniques

Jennifer Widom