

# 15-826: Multimedia Databases and Data Mining

Christos Faloutsos
CMU

www.cs.cmu.edu/~christos





#### **Problem**

Given a large collection of (multimedia) records, find similar/interesting things, ie:

- · Allow fast, approximate queries, and
- Find rules/patterns

15-826

Copyright: C. Faloutsos (2005)



#### Sample queries

- Similarity search
  - Find pairs of branches with similar sales patterns
  - find medical cases similar to Smith's
  - Find pairs of sensor series that move in sync
  - Find shapes like a spark-plug
  - (nn: 'case based reasoning')

-826 Copyright: C. Faloutsos (2005)



#### CMU SC

# Sample queries -cont'd

- Rule discovery
  - Clusters (of branches; of sensor data; ...)
  - Forecasting (total sales for next year?)
  - Outliers (eg., unexpected part failures; fraud detection)

15-826

Copyright: C. Faloutsos (2005)



#### **Outline**

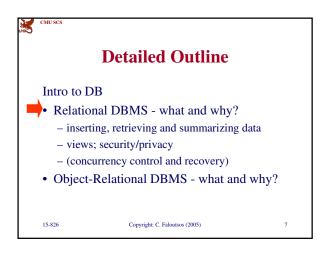
Goal: 'Find similar / interesting things'

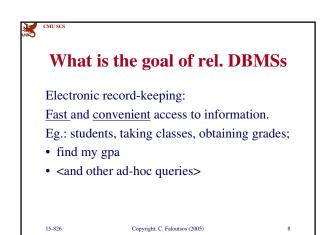
Intro to DB

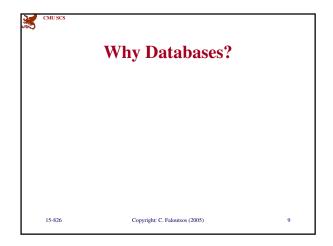
- Indexing similarity search
- Data Mining

15-826

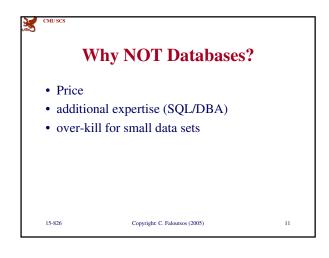
Copyright: C. Faloutsos (2005)

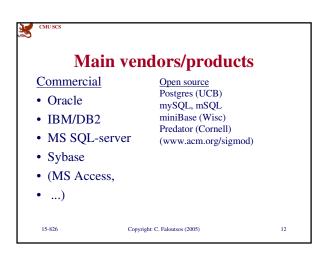


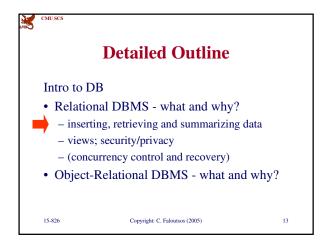


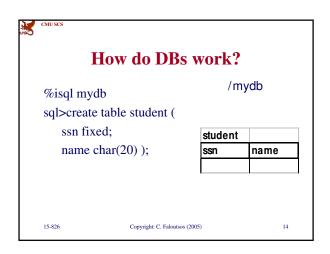


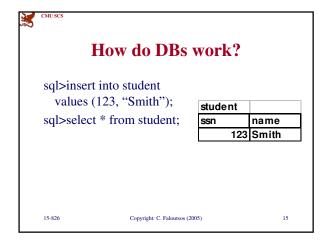


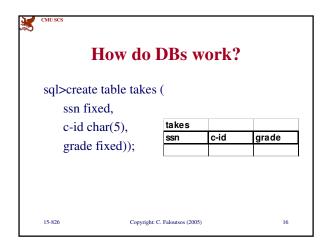


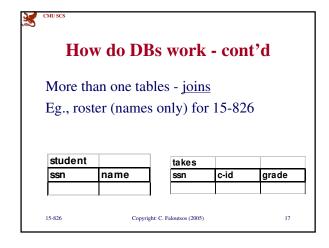


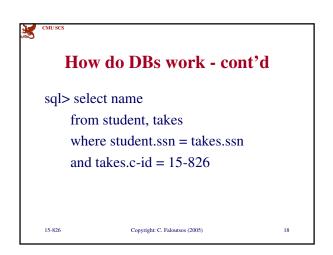


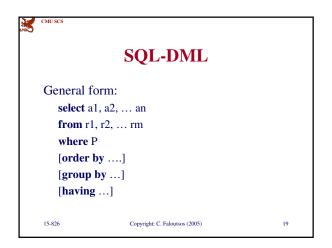


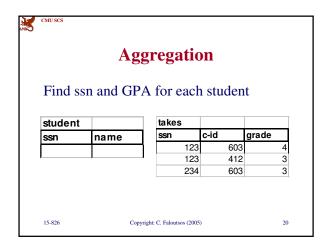


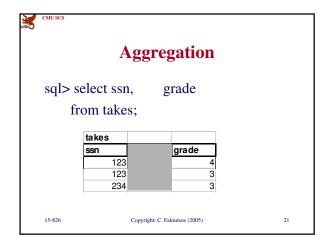


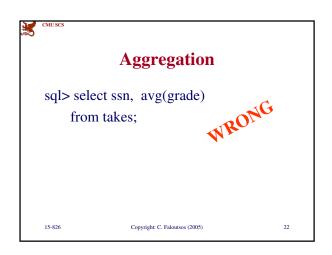


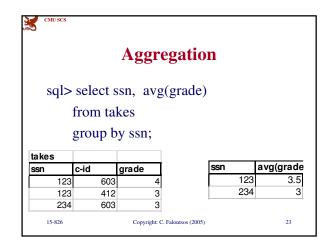


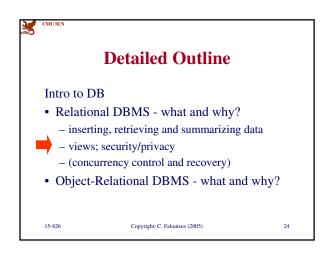


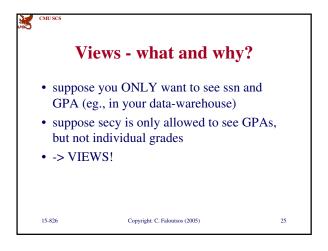


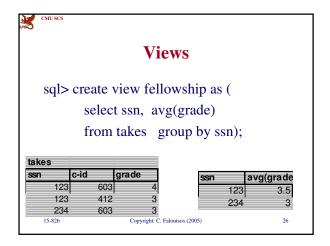


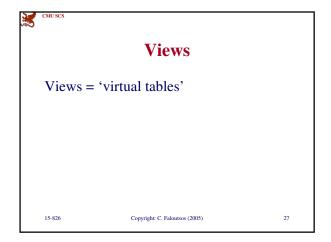


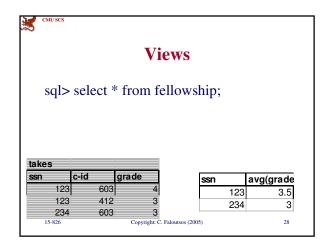


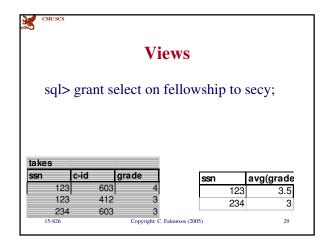


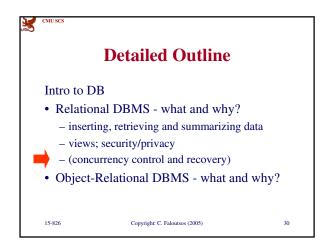


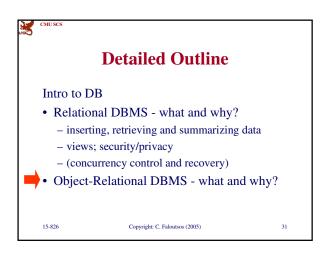


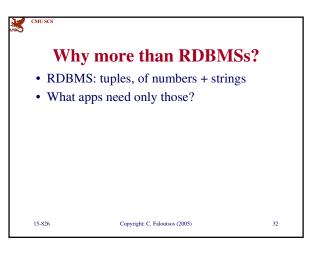


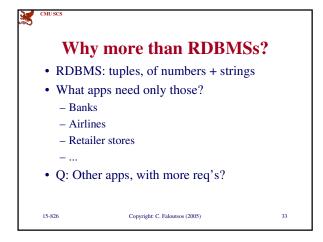


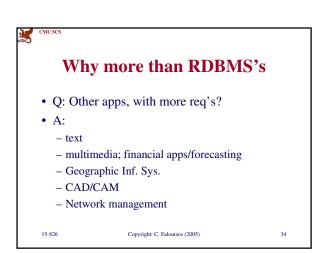


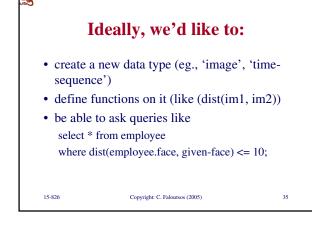


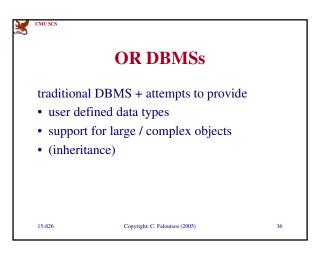


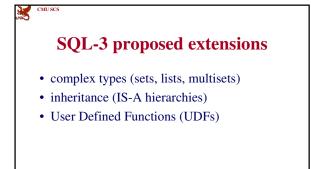












37

15-826 Copyright: C. Faloutsos (2005)

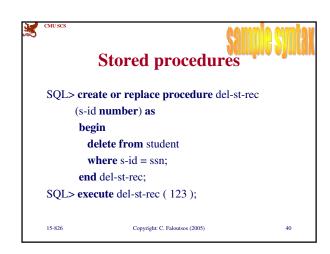
```
eg,
create type MyDate (
day decimal(2),
month char(3),
year decimal (4)
);

15-826 Copyright C. Faloutsos (2005) 38
```

BLObs etc:
Large objects, eg., video, images, 3d-MRI scans
new data types: LOB (=Large OBject)

BLOB: (up to 4Gb; binary: jpeg, mpeg, ...)
CLOB: (up to 2Gb; character: english text)
NCLOB:(.....; multi-byte characters)

Copyright: C. Faloutsos (2005)

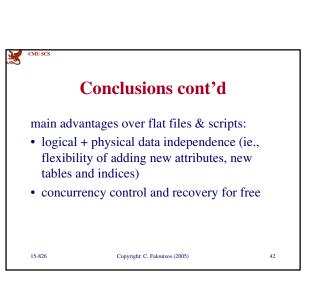


Conclusions

• (relational) DBMSs: electronic record keepers

• customize them with create table commands

• ask SQL queries to retrieve info





## Conclusions cont'd

• OR-DBMS: user-defined data types (eg., images), and U.D. functions.

15-826

Copyright: C. Faloutsos (2005)



## For more info:

- Microsoft Access: available on ANDREW clusters (PC)
- Ramakrishna + Gehrke, 3rd edition
- 15-415 web page, eg,
  - (http://www.cs.cmu.edu/~natassa/courses/15-415/current)

15-826

43

Copyright: C. Faloutsos (2005)