

15-826: Multimedia Databases and Data Mining

Lecture#1: Introduction

Christos Faloutsos

CMU

www.cs.cmu.edu/~christos



Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing similarity search
- Data Mining

15-826

Copyright: C. Faloutsos (2013)



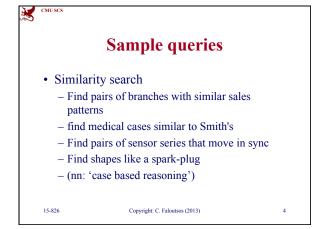
Problem

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

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

15-826

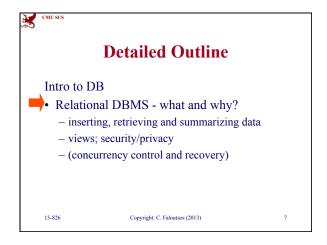
Copyright: C. Faloutsos (2013)

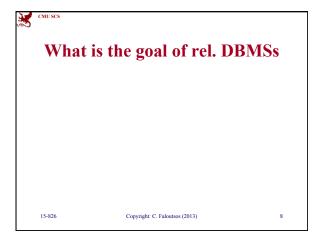


Sample queries —cont'd

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







What is the goal of rel. DBMSs

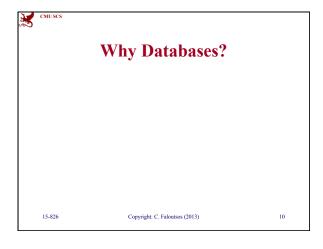
Electronic record-keeping:

Fast and convenient access to information.

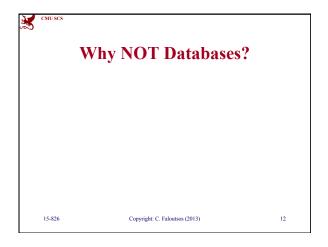
Eg.: students, taking classes, obtaining grades;

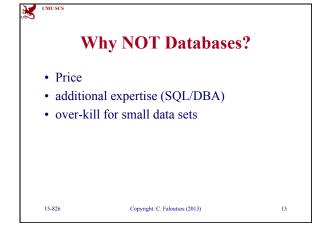
• find my gpa

• <and other ad-hoc queries>

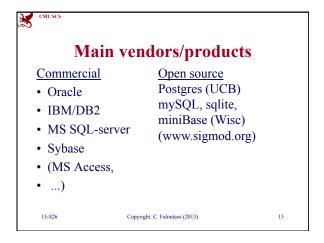


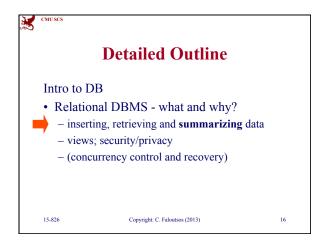
• Flexib	pility
• data in attribu	ndependence (can add new tables; new utes)
• data s	haring/concurrency control
• recov	ery

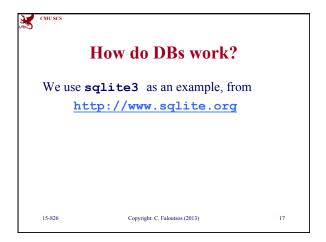


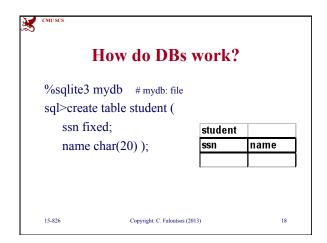


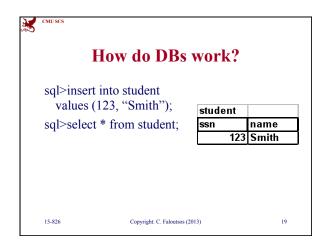


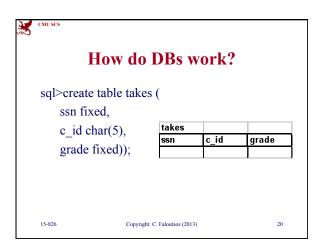




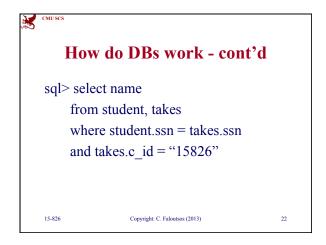


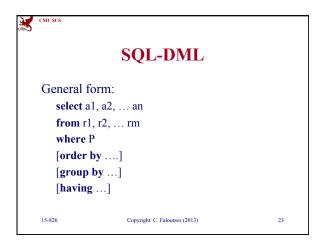


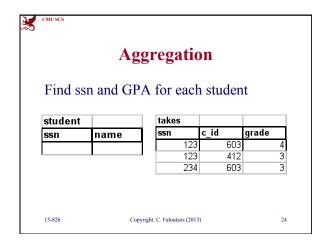


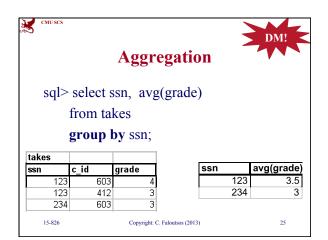


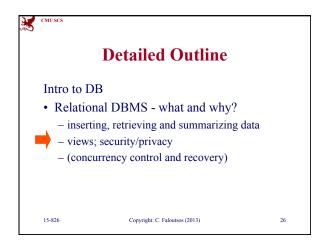
110	w do DE	3s work	- con	t'd
More th	an one tab	oles - join	S	
	ter (names	•		6
Lg., 103	ter (marries	only) lo		O
Lg., 103	ter (numes	omy) io	. 10 02	O
student	Ter (name)	takes	10 02	
	name		c_id	grade

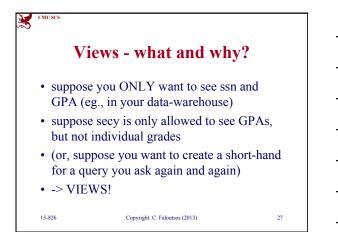


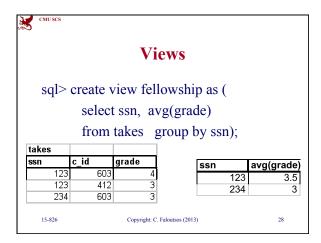


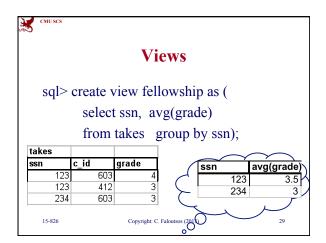


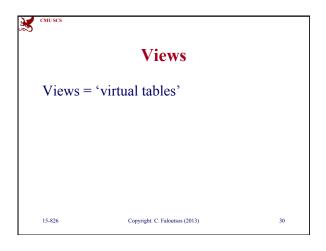


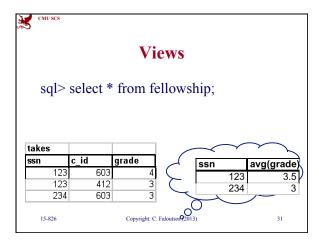


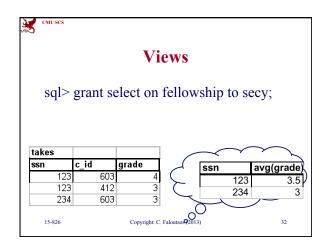


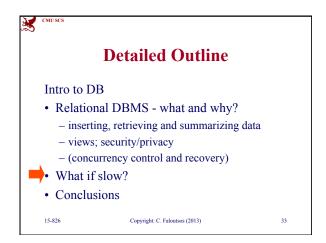


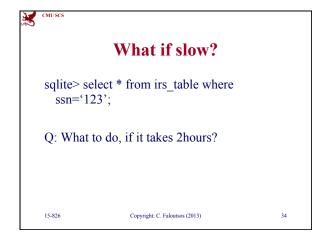


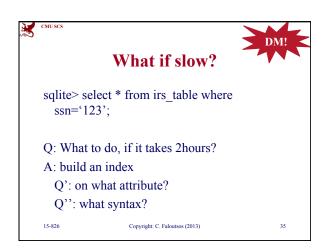


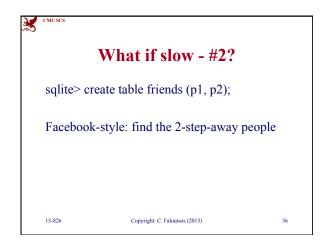


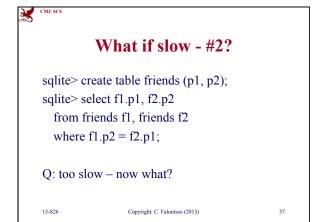












what if slow - #2?

sqlite> create table friends (p1, p2);
sqlite> select f1.p1, f2.p2
from friends f1, friends f2
where f1.p2 = f2.p1;

Q: too slow - now what?
A: 'explain': sqlite> explain select

15-826

Copyright: C. Faloutos (2013)

38



Long term answer:

• Check the query optimizer (see, say, Ramakrishnan + Gehrke 3rd edition, chapter15)

15-826

Copyright: C. Faloutsos (2013)



Conclusions

• (relational) DBMSs: electronic record keepers

- customize them with **create table** commands
- ask SQL queries to retrieve info

15-826

Copyright: C. Faloutsos (2013)



Conclusions cont'd

main advantages over flat files & scripts:

- logical + physical data independence (ie., flexibility of adding new attributes, new tables and indices)
- (concurrency control and recovery for free)

15-826

Copyright: C. Faloutsos (2013)

,

41

42



Conclusions cont'd

D.M. practitioner's guide

- Data mining: group by + aggregates
- If a query runs slow:
 - -explain select to see what happens-create index often speeds up queries

15-826

Copyright: C. Faloutsos (2013)

CMU SCS		
	For more info:	
• Sqlite3:	www.sqlite.org - @ linux.and	lrew
• postgres	S: v.postgresql.org/docs/	
Microso clusters	oft Access: available on andrew (PC)	
Ramakr	rishnan + Gehrke, 3rd edition	
• 15-415	web page, eg,	
- http://v	www.cs.cmu.edu/~christos/courses/dbms.S1	3/
15 926	Commishts C. Felantose (2012)	42