

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 (2011)



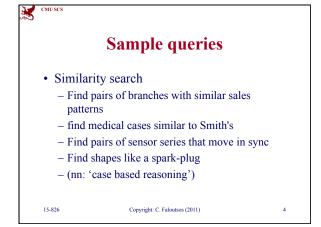
Problem

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

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

15-826

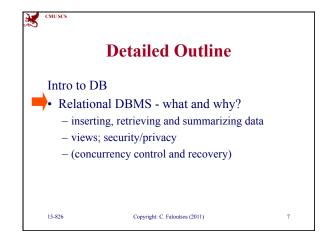
Copyright: C. Faloutsos (2011)

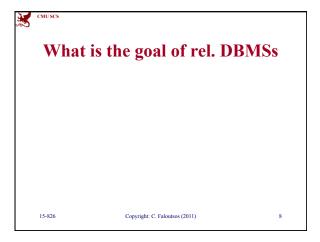


Sample queries —cont'd

• 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

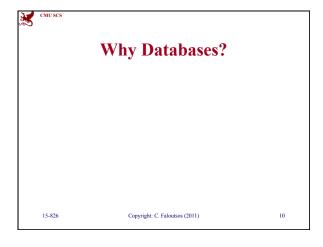
Electronic record-keeping:

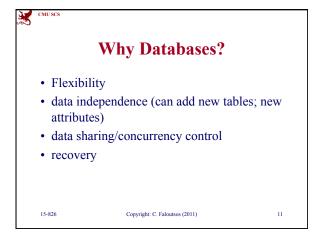
Fast and convenient access to information.

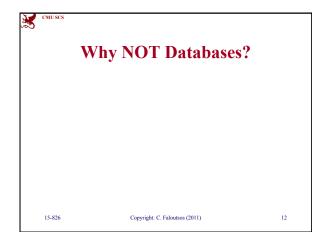
Eg.: students, taking classes, obtaining grades;

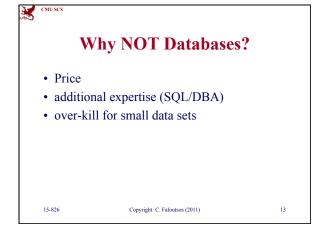
• find my gpa

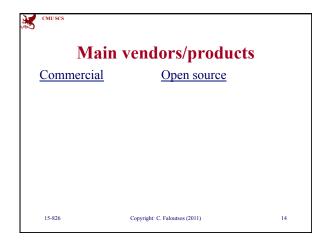
• <and other ad-hoc queries>

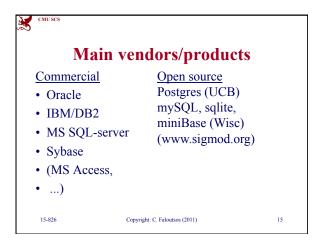


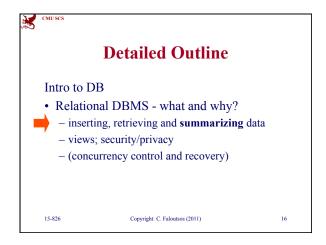




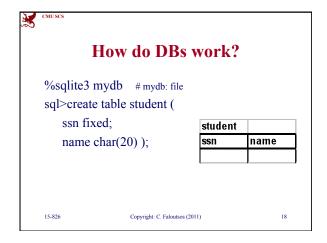


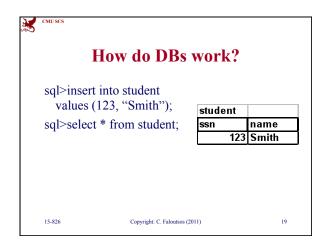


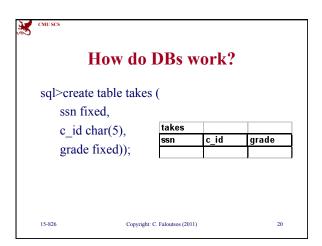




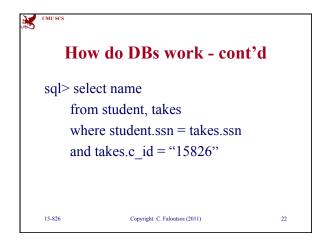


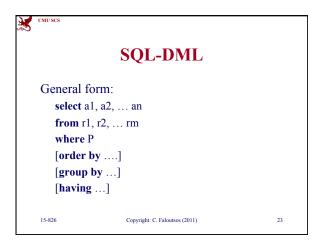


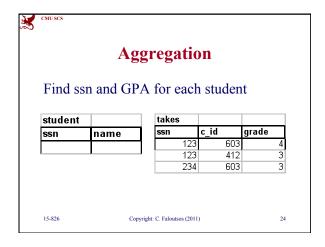


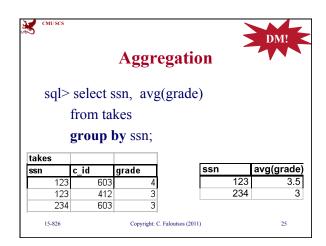


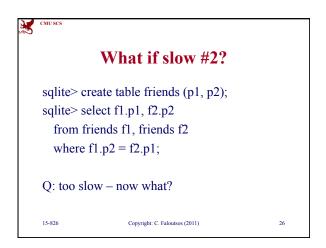
Hoy	w do DE	k work	- con	t'd
110	" do Di	25 WUIN	Con	· u
More th	an one tab	alec ioin	c	
wiore un	ian one tat	7168 - <u>Join</u>	.5	
Eg., ros	ter (names	s only) for	r 15-82	6
28., 100	(1101110)	01115) 10.		
student		takes		
student ssn	name	takes ssn	c_id	grade

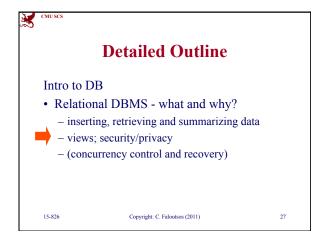


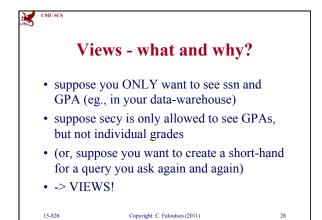




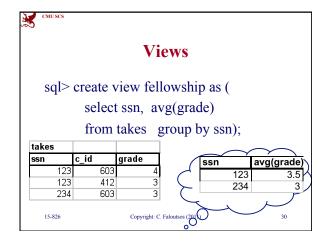


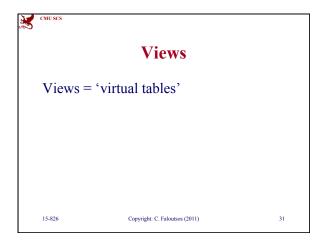


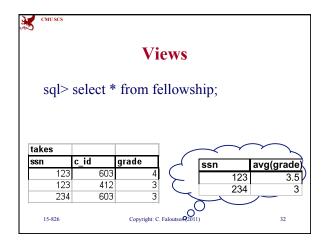


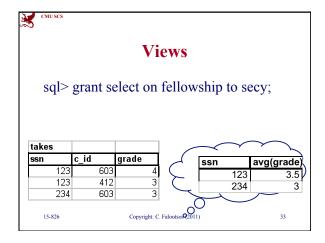


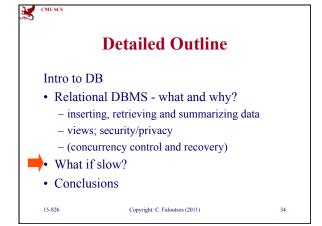
Views sql> create view fellowship as (select ssn, avg(grade) from takes group by ssn); takes c_id grade ssn ssn avg(grade) 603 123 3.5 123 412 234 234 603 15-826 Copyright: C. Faloutsos (2011)

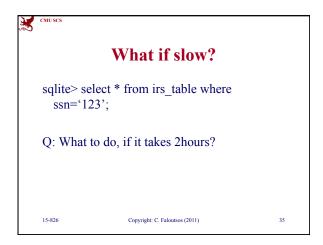


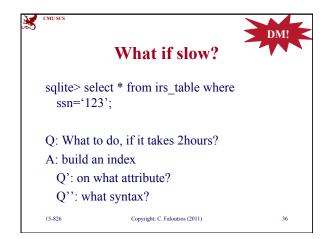


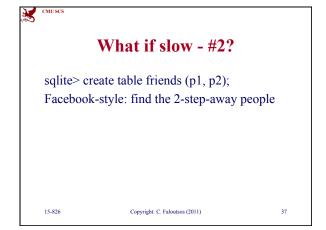












What if slow - #2?

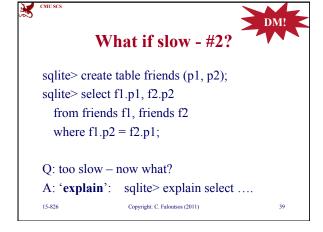
sqlite> create table friends (p1, p2);
sqlite> select f1.p1, f2.p2
from friends f1, friends f2
where f1.p2 = f2.p1;

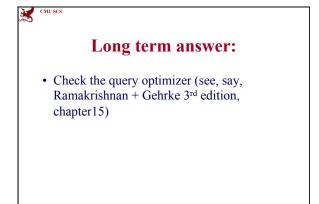
Q: too slow - now what?

15-826

Copyright: C. Faloutsos (2011)

38





Copyright: C. Faloutsos (2011)

S CMU SCS

Conclusions

- (relational) DBMSs: electronic record keepers
- customize them with **create table** commands
- ask SQL queries to retrieve info

15-826

Copyright: C. Faloutsos (2011)

41

42

CMU SCS

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 (2011)

CMU SCS		
	For more info:	
Microso clusters	oft Access: available on ANDR (PC)	EW
	www.sqlite.org	
• postgres	S: _postgresql.org/docs/	
 Ramakr 	rishna + Gehrke, 3rd edition	
	web page, eg,	T00/
- http://w	vww.cs.cmu.edu/~christos/courses/dbms-	·F09/
15-826	Copyright: C. Faloutsos (2011)	43