

RDBMS and SQL

Madhavan Mukund

<https://www.cmi.ac.in/~madhavan>

Lecture 1, 8 August 2024

Why DBMS?

Payroll

Leave status

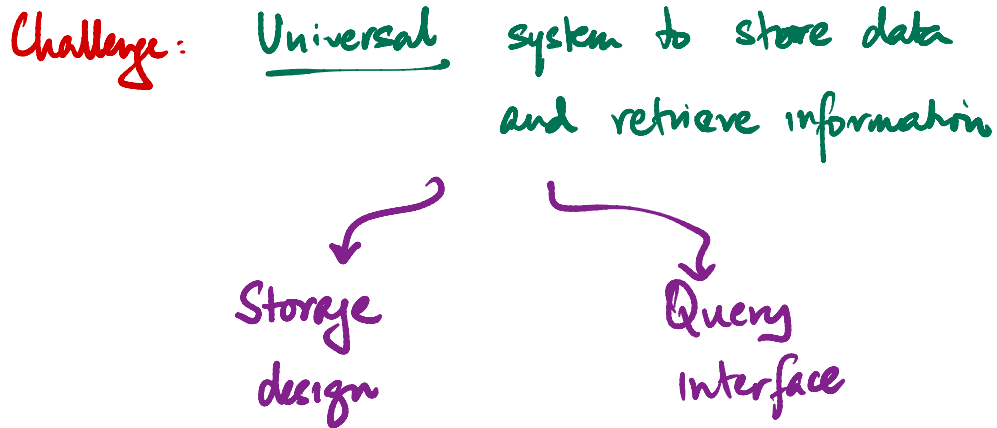
Length of service

On time percentage

Crew availability

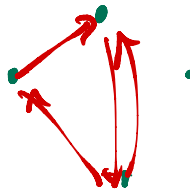
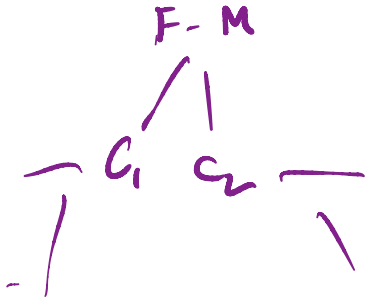
⋮

New requirements come up



Representing information

- Personnel data
- Courses, instructors, grades, enrollment
- Airline schedules
- Family trees



The relational model

- All information is in tables (relations)

~ 1970 Edgar Codd

↳ data representation + query interface



ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor



course_id	title	dept_name	credits
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

Course

Silberchatz, Korth & Sudarshan

7th edition

The relational model

- All information is in tables (relations)
 - How do we represent interrelationships — teaching allocation, prerequisites, timetable conflicts ...
- Avoid redundancy*

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2017
10101	CS-319	1	Spring	2018
10101	CS-347	1	Fall	2017
12121	FIN-201	1	Spring	2018
15151	MU-199	1	Spring	2018
22222	PHY-101	1	Fall	2017
32343	HIS-351	1	Spring	2018
45565	CS-101	1	Spring	2018
45565	CS-319	1	Spring	2018
76766	BIO-101	1	Summer	2017
76766	BIO-301	1	Summer	2018
83821	CS-190	1	Spring	2017
83821	CS-190	2	Spring	2017
83821	CS-319	2	Spring	2018
98345	EE-181	1	Spring	2017

Teaches

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califeri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor

course_id	title	dept_name	credits
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

Course

The relational model

- All information is in tables (relations)
- How do we represent interrelationships — teaching allocation, prerequisites, timetable conflicts ...

<i>course_id</i>	<i>prereq_id</i>
BIO-301	BIO-101
BIO-399	BIO-101
CS-190	CS-101
CS-315	CS-101
CS-319	CS-101
CS-347	CS-101
EE-181	PHY-101

Prerequisites

<i>course_id</i>	<i>title</i>	<i>dept_name</i>	<i>credits</i>
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

Course

Storing information

- How is information organized physically? Example: books in a library



Indexing information

- Finding a book in the library



■ Finding a book in the library



```
PS3557
.R5355      Grisham, John
F57 1991

The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.
```

Indexing information

■ Finding a book in the library



PS3557
.R5355 Grisham, John
F57 1991

The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

PS3557 The firm
.R5355
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

Indexing information

■ Finding a book in the library



PS3557
.R5355 Grisham, John
F57 1991

The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

PS3557 The firm
.R5355
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

PS3557 Organized crime--Fiction
.R5355
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

Indexing information

■ Finding a book in the library



PS3557
.R5355 Grisham, John
F57 1991

The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

PS3557 The firm
.R5355
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

PS3557 Organized crime--Fiction
.R5355
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

PS3557 Government investigators--Fiction
.R5355
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
421p. ; 24 cm.

1. Government investigators--Fiction.
2. Organized crime--Fiction.

Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 1 cr



Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 1 cr

Electoral Rolls for TN

IT Data

m
rows

Name	Distnct

matches
uniquely

Name	Income

n
rows

Querying information efficiently

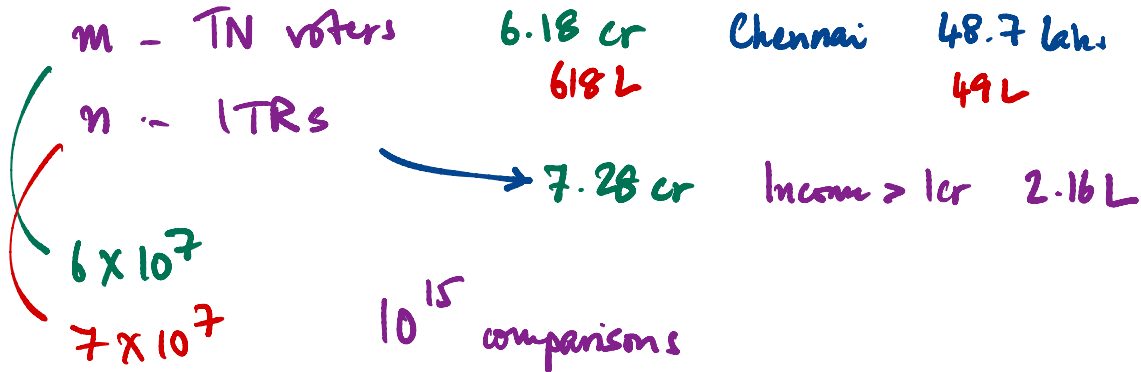
- Find all taxpayers from Chennai with annual income over Rs 1 cr

for each row e in Electoral rolls m times
for each t in Tax table n times

if e & t have same name, $m \cdot n$
 e is from Chennai
 t income $>$ 1cr ✓
Checks

Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 1 cr



Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 1 cr

10^{15} comparisons? $\Rightarrow 10^6$ secs

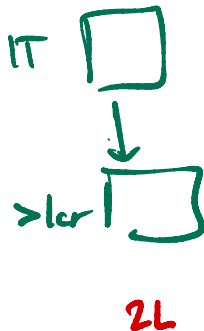
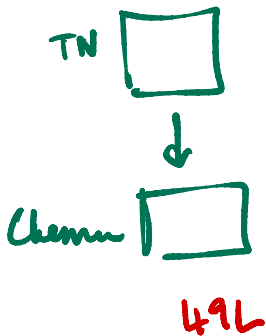
10^9 ops/sec

$60 \times 60 \times 24 \approx 90,000$

36×10^6

Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 1 cr



$$49 \times 10^5 \\ \times 2 \times 10^5$$

$$10^{12}$$

~ 1000 sec \approx 20 mins

Transactions and concurrency

- Choosing your seat on a flight

Transaction - Database update that involves multiple steps

Atomic
Consistent
Isolation
Durable

Transactions and concurrency

- Choosing your seat on a flight
- Transactions — ACID properties

Mathematically?

Mapping?

$$f: S \rightarrow T$$

$$x \mapsto \sqrt{x}$$

$$x \mapsto x^3$$

↑

Domain?

Binary relation

$$R \subseteq S \times T$$

$$S \times T = \{ (s, t) \mid \begin{array}{l} s \in S, \\ t \in T \end{array} \}$$

↑
Cartesian Product

S = Courses

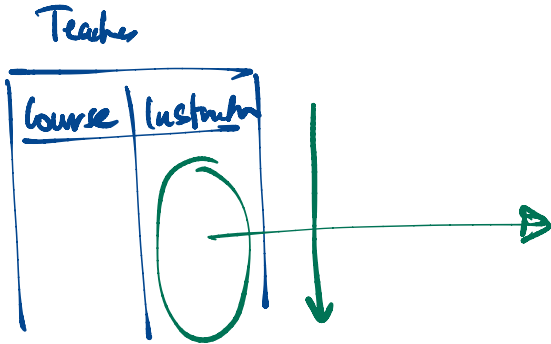
T = Instructors

Teaches \subseteq Courses \times Instructors

Relations are over sets — No duplicates
Order is unimportant

Tables are over sequences — Duplicates
— Order is important

Who is teaching this semester?



Relation: Each name appears once

Table - As many times as in original table

Relations

- Tables have fixed structure — **relation schema**

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor

Instructor			
ID	name	dept_name	salary
Text	Text	Text	Number

Relations

- Tables have fixed structure — relation schema

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor

Instructor			
ID	name	dept_name	salary
Text	Text	Text	Number
Format?		Fixed set?	Range?

CS
Fin
Mus
Phy
Hc
Bio
EE