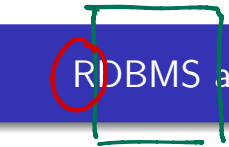
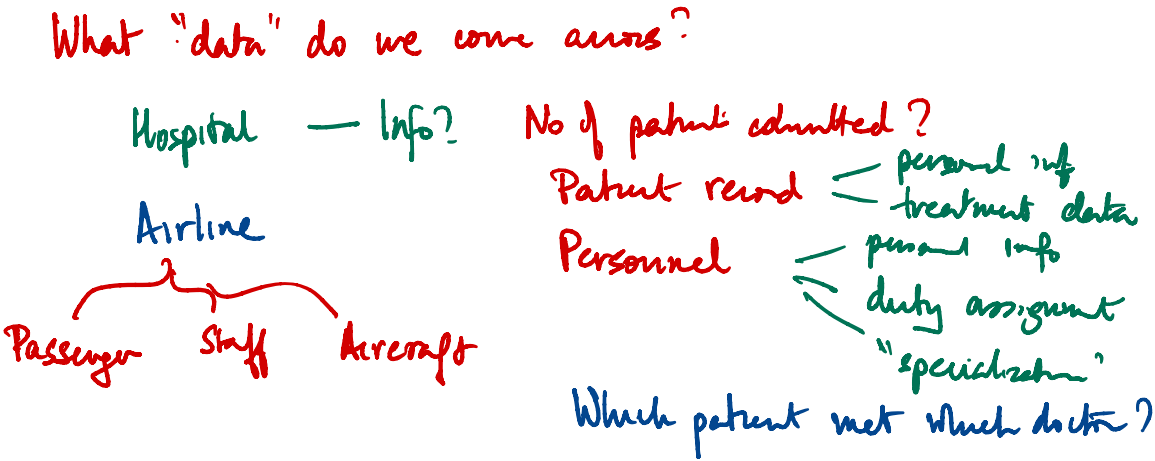


Madhavan Mukund

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

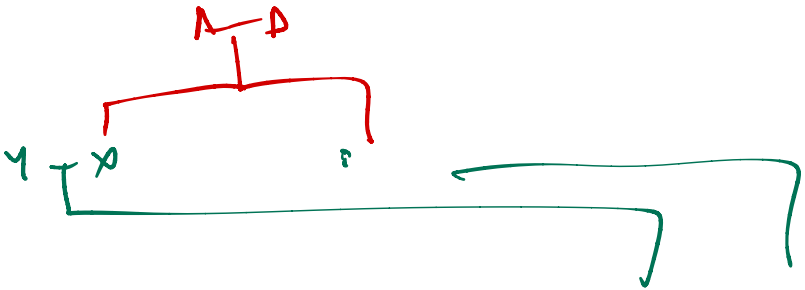
Lecture 1, 19 August 2026

Why DBMS?


What "data" do we come across?
Hospital — Info?
No of patients admitted?
Patient record
personal inf
treatment data
Personnel
personal info
duty assignment
"specialization"
Which patient met which doctor?
Airline
Passenger
Staff
Aircraft

Representing information

- Personnel data
- Courses, instructors, grades, enrollment] Running example in textbook
- Airline schedules
- Family trees



Representing information

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

1960's

Customized solution

Internet banking

General framework

- Flexible
- Efficient - Unknown requirements

The relational model

A Relational Model of Data for Large Shared Data Banks

E. F. Codd
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

≈ 1970
Edgar Codd
Turing Award
1970's
System R
↓
SQL

The relational model

- All information is in tables (relations)

S, T (s, t)

R ⊆ S × T

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

The relational model

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

ID	*course_id*	*sec_id*	*semester*	*year*
10101	CS-101	1	Fall	2017
10101	CS-315	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	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

The relational model

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

course_id	*prereq_id*
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

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

Storing information

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

By author name

By subject
↓
classification
|
numerical

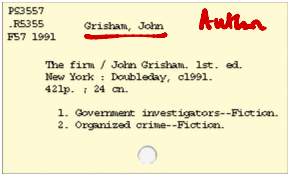
Dewey Decimal Number

Indexing information

- Finding a book in the library

Indexing information

- Finding a book in the library



Indexing information

- Finding a book in the library

PG3557
.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.

PG3557
.R5355 The firm Title
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

PG3557
.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.

PG3557
.R5355 The firm
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.

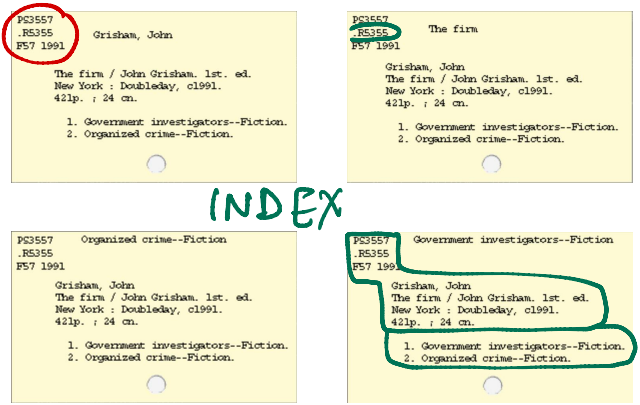
PG3557 Organized crime--Fiction
.R5355 Topic
F57 1991

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

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

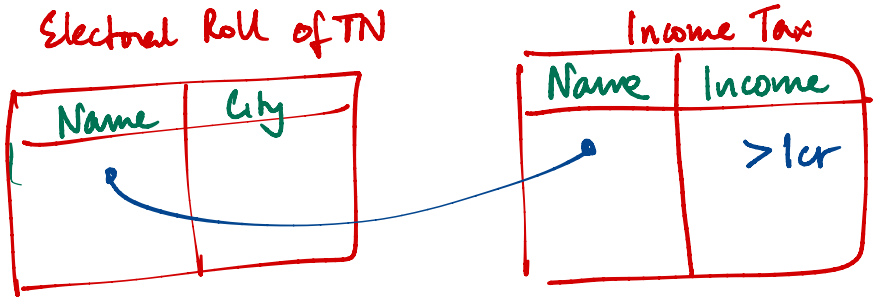
Indexing information

- Finding a book in the library



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

for voter in TN

for taxpayer in IT

If name matches,
city is Chennai,
income > 1cr

Output name

Querying information efficiently

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

How long?

Taxpayers? 8-9 cr

> 1cr? ≈ 4L

Voters in TN? ~5 cr

Chennai ~30L

for each voter 10^7

for each taxpayer 10^7

10^{14} comparisons

10^7 comparisons/sec in Python

10^7 sec

$60 \times 60 \times 24 = 90000$

10^4 1 day

10^3 days

Querying information efficiently

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

Extract TN voters in Chennai $10^7 \longrightarrow 30L$

Extract IT returns > 1c $+ 10^8 \longrightarrow 5L$

30×10^5

$3 \times 10^6 \times 5 \times 10^5 = 10^{10}$

Querying information efficiently

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

DBMS keeps "metadata" — How many for each city
How many for each income slab

Transactions and concurrency

- Choosing your seat on a flight

Updates

Sequence of actions / updates

Transactions and concurrency

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

Atomic
Consistent
Isolated
Durable