

RDBMS and SQL

Madhavan Mukund

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

Lecture 1, 14 August 2025

Why DBMS?

What? Data Base Management System

Manage lots of data?

Data? Nature of information

Ticket information

Route information — Equipment info

Staff information

Information is not static

- Storing & Updating Information
- Retrieve information — flexible

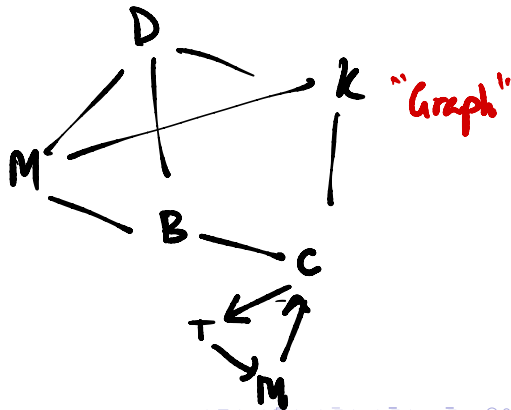
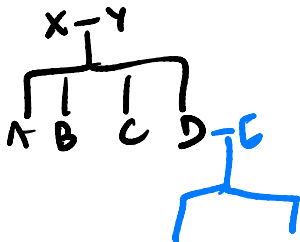
Representing information

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

— Running example



"Tree"



Why DBMS?

- Store info
- Query info
- Generic solution

Historically (1960s)

Many proposed models

X

A Relational Model of Data for Large Shared Data Banks

E. F. Codd

IBM Research Laboratory, San Jose, California

TURING AWARD

~1970



**IBM research
project**

System R



SQL

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.

The relational model

- All information is in tables (relations)

Silberschatz, Korth, Sudarshan
7th edition



<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



<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

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-315	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
55565	Katz	Comp. Sci.	75000
58583	Califleri	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



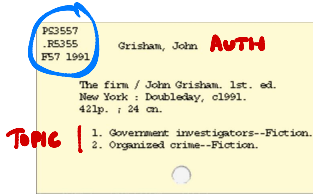
Indexing information

- Finding a book in the library



Indexing information

■ Finding a book in the library



PS3557
.R5355
F57 1991

Grisham, John **AUTH**

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

Topic | 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
.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.

Indexing information

■ Finding a book in the library



PS3557
.R5355 Grisham, John
F57 1991

The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
42lp. ; 24 cm.

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

PS3557
.R5355 The firm
F57 1991

Grisham, John
The firm / John Grisham. 1st. ed.
New York : Doubleday, c1991.
42lp. ; 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.
42lp. ; 24 cm.

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

Indexing information

■ Finding a book in the library



PS3557
.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
.R5355
F57 1991

The firm

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

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

PS3557
.R5355
F57 1991

Organized crime--Fiction

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

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

PS3557
.R5355
F57 1991

Government investigators--Fiction

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

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

"Indexes"

Querying information efficiently

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

Elected Roll
for TN

Name	District

ITR. data

Name	Income

Match



List
↓

List
↓

for each r1 in TN Voter list

for each r2 in ITR list

if r1.name == r2.name &
r1.district == Chennai

&
r2.income > lcr

add r1+r2 to output

Voters in TN? ≈ 6 cr

ITR filed? ≈ 7 cr

$$\begin{array}{r} 6 \times 10^7 \\ \times 5 \times 10^7 \\ \hline \end{array}$$

10^{15} comparisons

$\leq 10^9$ comparisons/sec

$$\Rightarrow \frac{10^6 \text{ secs}}{9 \times 10^4} \approx 10 \text{ days}$$

$$\begin{aligned} 60 \times 60 \times 24 &= 3600 \times 24 \\ &\approx 900 \times 100 \end{aligned}$$

$$\frac{100,000}{90,000} \approx 1.1$$

Create temporary tables

TN voter list $\rightarrow$ Chennai Voter list ≈ 50 lakhs
6 cr

ITR list $\rightarrow$ ITR > 1cr list ≈ 2 lakhs?
7 cr

$$\begin{array}{c} 6 \times 10^7 \\ \downarrow \\ 5 \times 10^6 \end{array}$$

$$\begin{array}{c} 7 \times 10^7 \\ \downarrow \\ 2 \times 10^5 \end{array}$$

DBMS maintain extra info

Size of each table

Multiplicity of each value



"Metadata"

Transactions and concurrency

- Choosing your seat on a flight

Transactions and concurrency

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

Atomic

Consistent

Isolated

Durable