

RDBMS and SQL

Madhavan Mukund

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

Lecture 1, 8 September 2023

Data Base Management System



Maintain information — organizational

- Simple "universal" format

- Flexible ("programmable") query mechanism

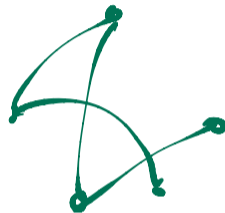
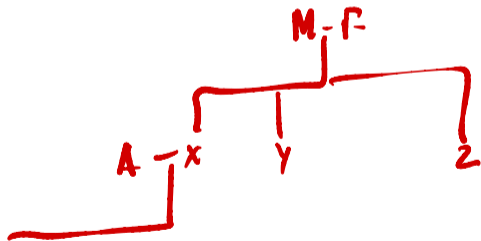
Backend support

- How to physically store data
- How to "compute" results

Representing information

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

} "Graphical"



The relational model

- All information is in tables (relations)

E.F. Codd ~ 1970 Turing Award
IBM - System R \Rightarrow SQL

<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 ...

<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

Storing information

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



Anna
Centenary
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

all taxpayers with income > 5cr who are from Chennai

- Find all taxpayers from Chennai with annual income over Rs 5 cr

TN voters

Name	Dist

~ 6cr 10⁷

approx 10% in Chennai
~ 60L

Income Tax

Name	Income

? ~ 8cr 10⁷

Income > 5cr?
~ 25,000

For simplicity,
names are
unique

Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 5 cr

for taxpayer in IT:

if income > 5cr:

for voter in TN:

if district == Chennai

report

Worst case

$$10^7 \times 10^7$$

$$= 10^{14}$$

Querying information efficiently

- Find all taxpayers from Chennai with annual income over Rs 5 cr

Better

Extract relevant rows

25,000 income
> 5cr

60 L voters
from Chennai

10⁷
+
10⁷

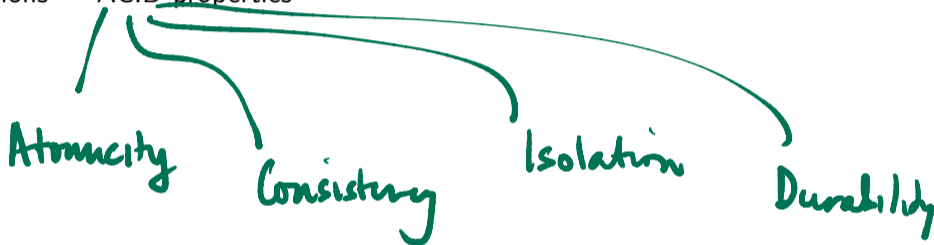
"Queries are declarative" → What I want

How to compute? - "Imperative"

- Choosing your seat on a flight

Transactions and concurrency

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



Sets S_1, S_2, \dots, S_n

$$S_1 \times S_2 = \{(s_1, s_2) \mid s_1 \in S_1, s_2 \in S_2\}$$

$$S_1 \times S_2 \dots \times S_n = \{(s_1, s_2, \dots, s_n) \mid s_1 \in S_1, \dots, s_n \in S_n\}$$

Relation $r \subseteq S_1 \times \dots \times S_n$

$$S_1 = \mathbb{Z} \quad S_2 = \mathbb{R}$$

$$\text{SquareRoot} \subseteq \mathbb{Z} \times \mathbb{R} = \{(n, r) \mid r = \sqrt{n}\}$$

$$\text{Pyt} \subseteq \mathbb{N} \times \mathbb{N} \times \mathbb{N} = \{(p, q, r) \mid p^2 + q^2 = r^2\}$$

Relations

- Tables have fixed structure — relation schema

$r \subseteq \mathcal{D}_1 \times \mathcal{D}_2 \dots \times \mathcal{D}_k$

<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?