RELATIONAL DATABASE MANAGEMENT SYSTEM (RDBMS)

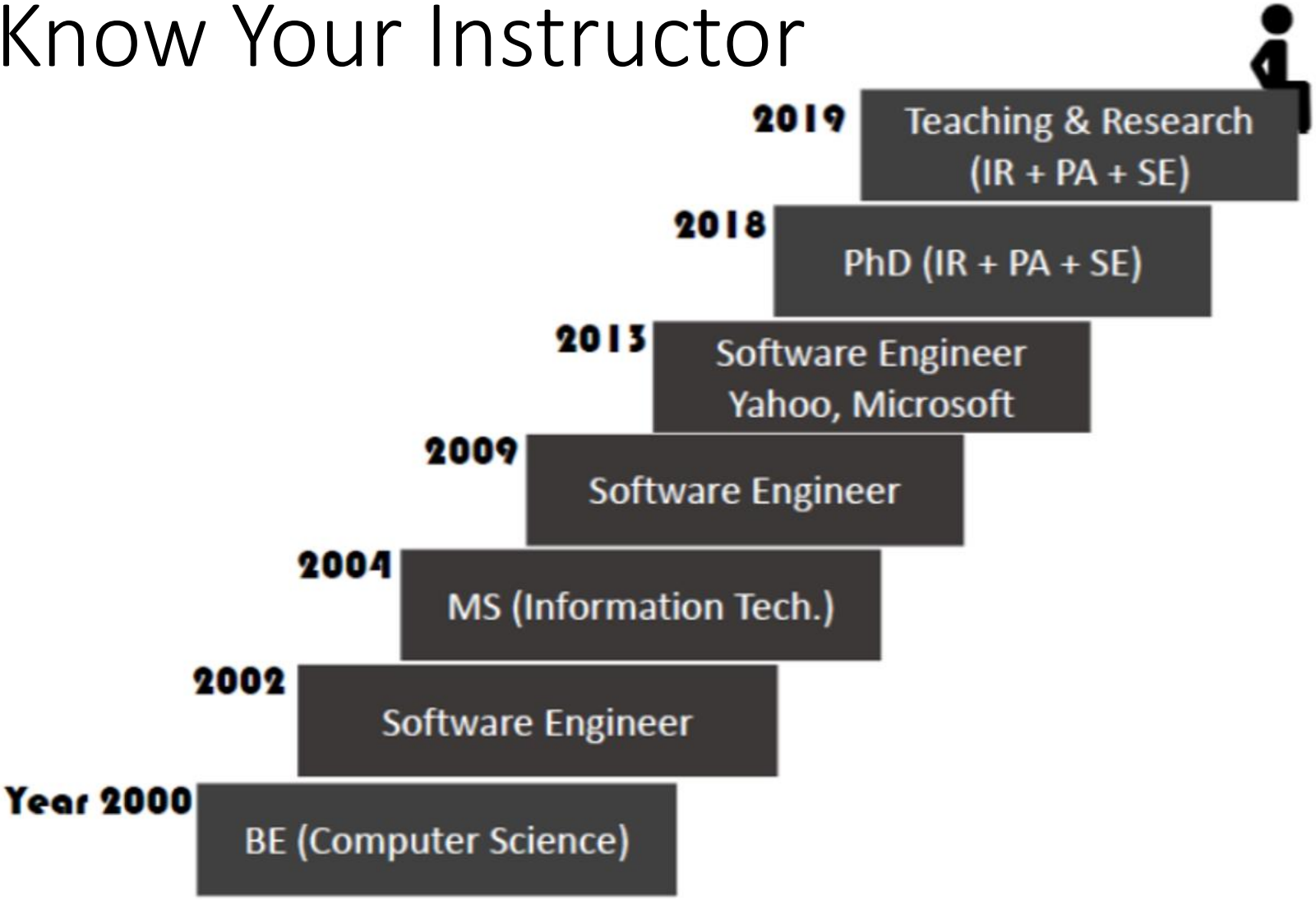
Venkatesh Vinayakarao
venkateshv@cmi.ac.in
http://vvtesh.co.in

Chennai Mathematical Institute

I was vehemently against acquisitions. Now let's buy everything in sight.

– Larry Ellison.

Know Your Instructor



Larry Ellison

Oracle Corporation

- “Good for nothing” – said his Father.
- Dropped out of college twice.
- Never took a CS class

  … “DBMS” …
- Became World’s sixth richest!

A database-management system (DBMS) is a collection of:

1. interrelated data and
2. a set of programs to access those data.

The collection of data is usually referred to as database.

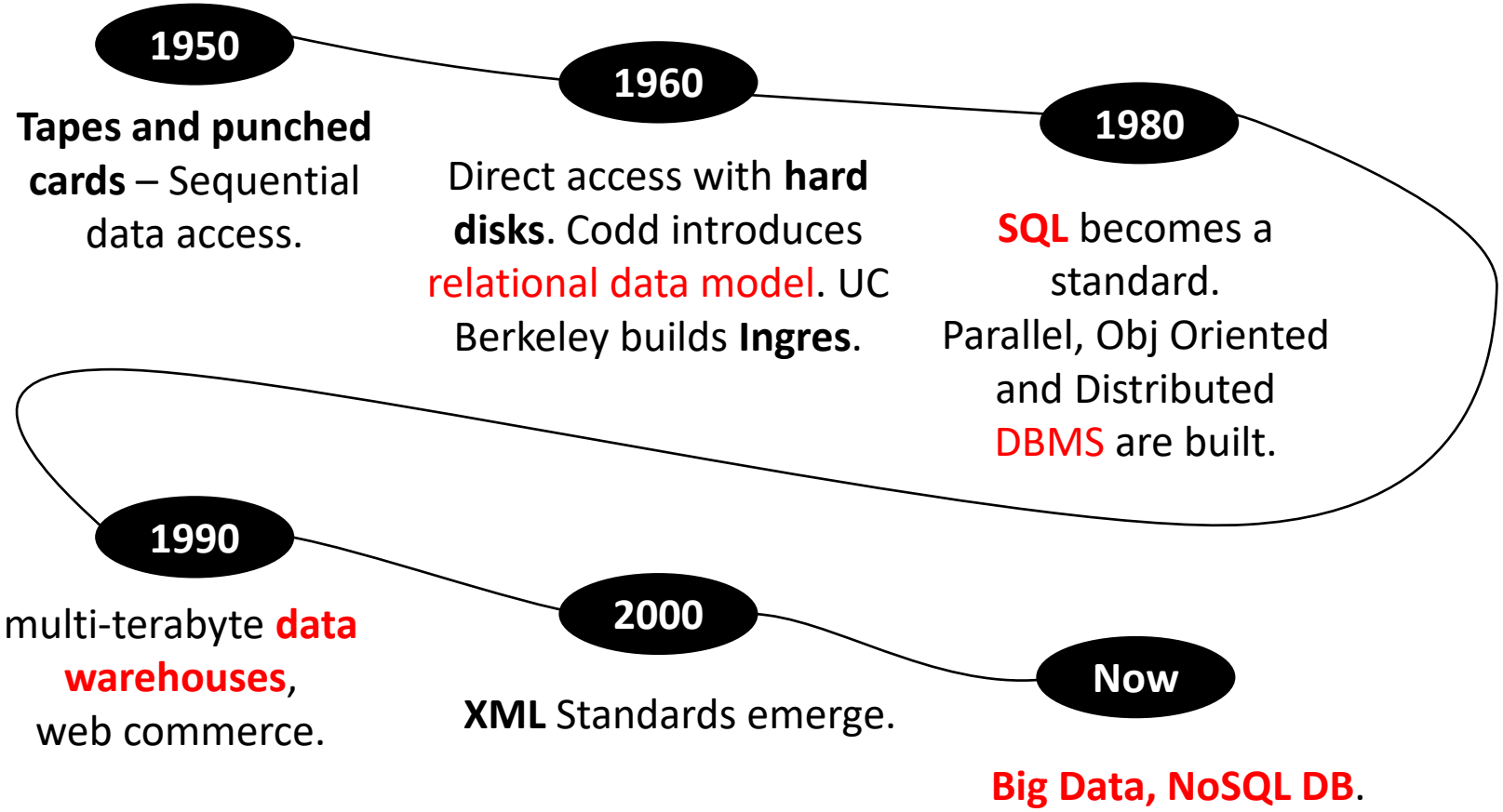
Why files are insufficient to store data?

1. Redundancy
2. Inconsistency
3. Difficulty in accessing data
4. Too many file formats… pdf, ppt, jpg, gif, txt, …

Why files are insufficient to store data?

5. Backups
6. Collaboration (for editing)
7. Security
8. Integrity

History


1950
Tapes and punched cards – Sequential data access.
1960
Direct access with hard disks. Codd introduces relational data model. UC Berkeley builds Ingres.
1980
SQL becomes a standard. Parallel, Obj Oriented and Distributed DBMS are built.
1990
multi-terabyte data warehouses, web commerce.
2000
XML Standards emerge.
Now
Big Data, NoSQL DB.

Assume you are in 1960s… We only have files… We do not have any DBMS (like SQL Server, MySQL, Oracle, etc)

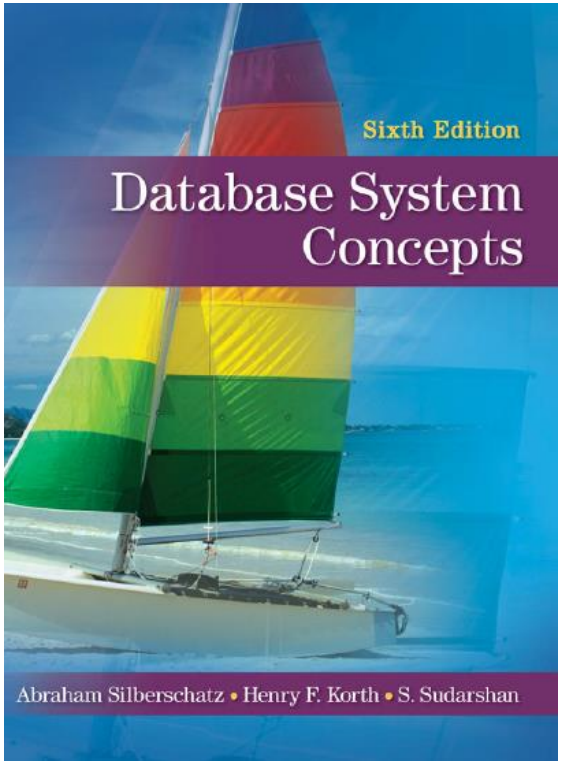
Objective-1
How to build a DBMS?

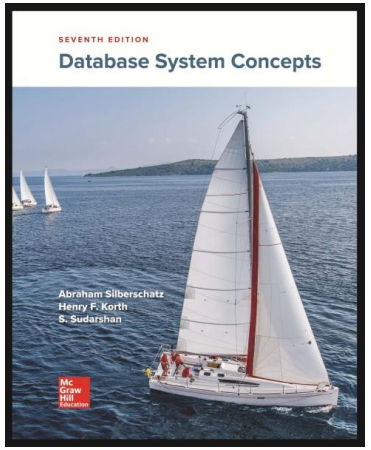
Now, you are in 2020

Objective-2
How to use a DBMS?

Course Text

We will follow the…





7th Edition Covers
Big Data, Block Chain,
Distributed Comptuing…

https://www.db-book.com/db6/index.html

Acknowledgment

- (Some) contents are borrowed from the official website of the course text. For the authors' original version of slides, visit:
  - https://www.db-book.com/db6/slide-dir/index.html

Let Us Build a DBMS

Venkatesh Vinayakarao
venkateshv@cmi.ac.in
http://vvtesh.co.in

Chennai Mathematical Institute

Slide contents are borrowed from the official website of the course text. For the authors' original version of slides, visit: https://www.db-book.com/db6/slide-dir/index.html.

What is Relational about the Relational Database Management System?

Venkatesh Vinayakarao
venkateshv@cmi.ac.in
http://vvtesh.co.in

Chennai Mathematical Institute

The Text Book Definition

A database-management system (DBMS) is a collection of:

1. inter**related** data and
2. a set of programs to access those data.

The collection of data is usually referred to as database.

Remember…

A database-management system (DBMS) is a collection of:

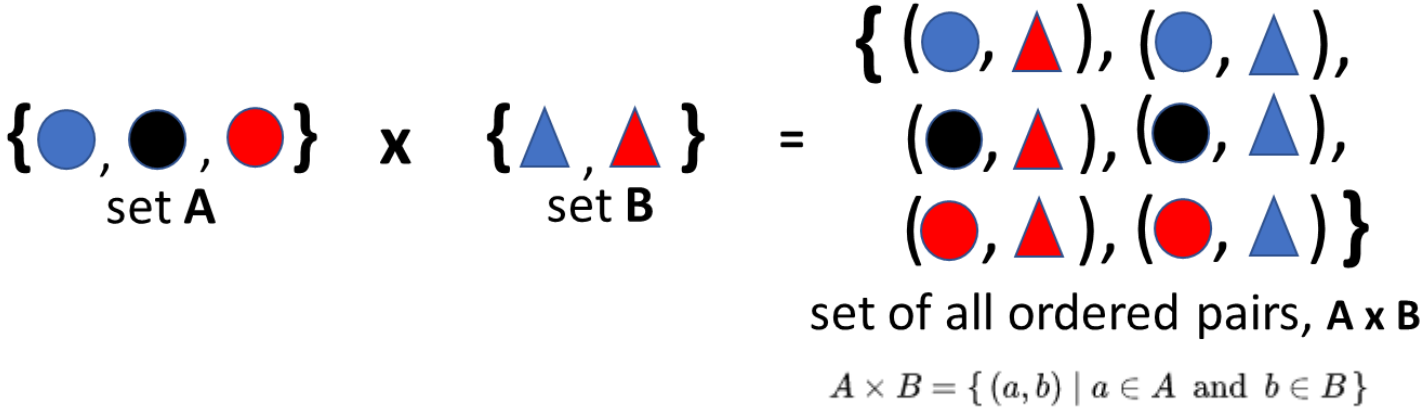
1. inter**related** data and
2. a set of programs to access those data.

The collection of data is usually referred to as database.

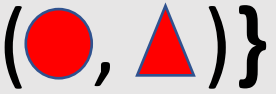
The idea of “**Relations**”

Quiz

- A relation R from a set A to set B is a subset of the cartesian product A x B. True/False?



Let's say a relation exists between the reds:

Relation R = 

A Relation

- Let the set, id = {1,2,3}
- Let the set, name = {vv, sd}
- What is id x name?
- We have a relation if we assign a sequential id to each name.

id	name
1	sd
2	vv

id	name
1	sd
1	vv
2	sd
2	vv
3	sd
3	vv

Interrelated “Data” as a Relation

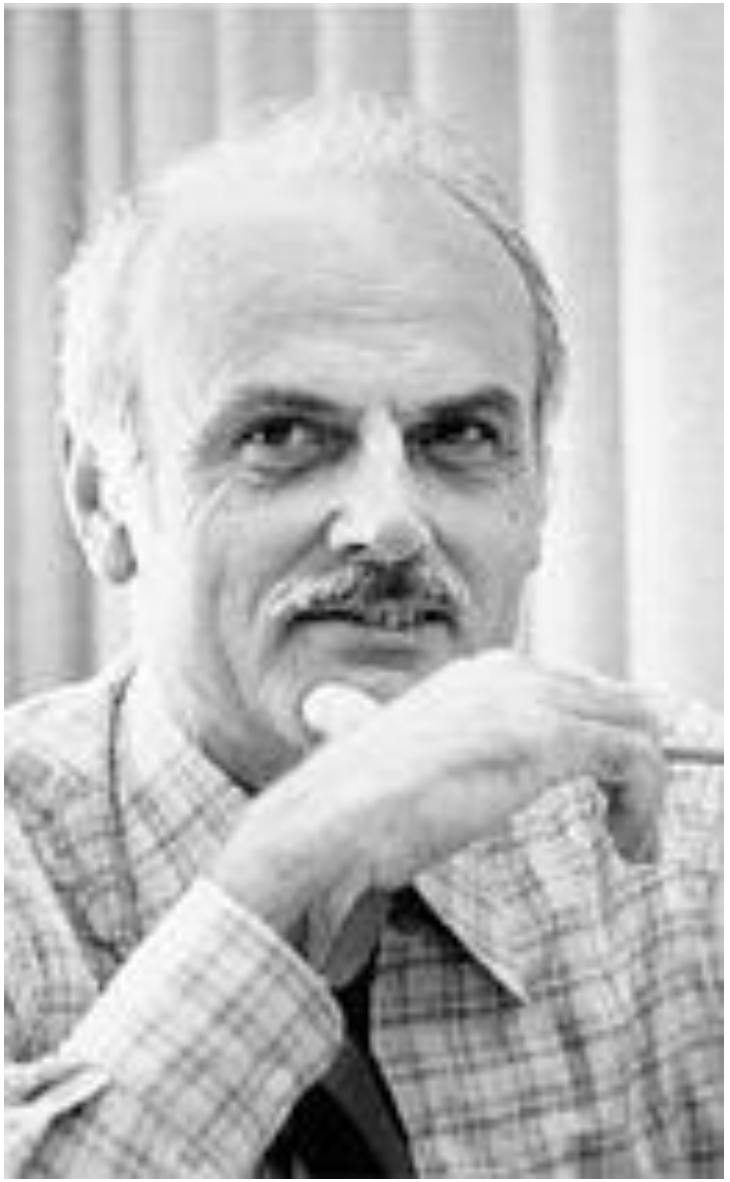
The Relational Model

Edgar F. Codd

PhD in Computer Science

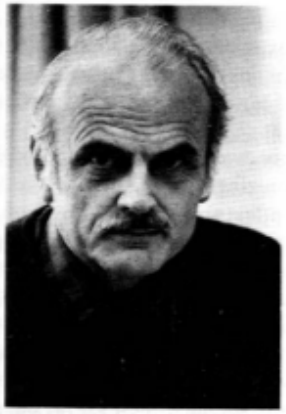
Winner of the Turing Award

He made other valuable contributions to computer science, but the relational model, a very influential general theory of data management, remains his most mentioned, analyzed and celebrated achievement. –Wikipedia.



The 1981 ACM Turing Award Lecture

Delivered at ACM '81, Los Angeles, California, November 9, 1981



The 1981 ACM Turing Award was presented to Edgar F. Codd, an IBM Fellow of the San Jose Research Laboratory, by President Peter Denning on November 9, 1981 at the ACM Annual Conference in Los Angeles, California. It is the Association's foremost award for technical contributions to the computing community.

Codd was selected by the ACM General Technical Achievement Award Committee for his "fundamental and continuing contributions to the theory and practice of database management systems." The originator of the relational model for databases, Codd has made further important contributions in the development of relational algebra, relational calculus, and normalization of relations.

Edgar F. Codd joined IBM in 1949 to prepare programs for the Selective Sequence Electronic Calculator. Since then, his work in computing has encompassed logical design of computers (IBM 701 and Stretch), managing a computer center in Canada, heading the development of one of the first operating systems with a general multiprogramming capability, contributing to the logic of self-reproducing automata, developing high level techniques for software specification, creating and extending the relational approach to database management, and developing an English analyzing and synthesizing subsystem for casual users of relational databases. He is also the author of *Cellular Automata*, an early volume in the ACM Monograph Series.

Codd received his B.A. and M.A. in Mathematics from Oxford University in England, and his M.Sc. and Ph.D. in Computer and Communication Sciences from the University of Michigan. He is a Member of the National Academy of Engineering (USA) and a Fellow of the British Computer Society.

The ACM Turing Award is presented each year in commemoration of A. M. Turing, the English mathematician who made major contributions to the computing sciences.

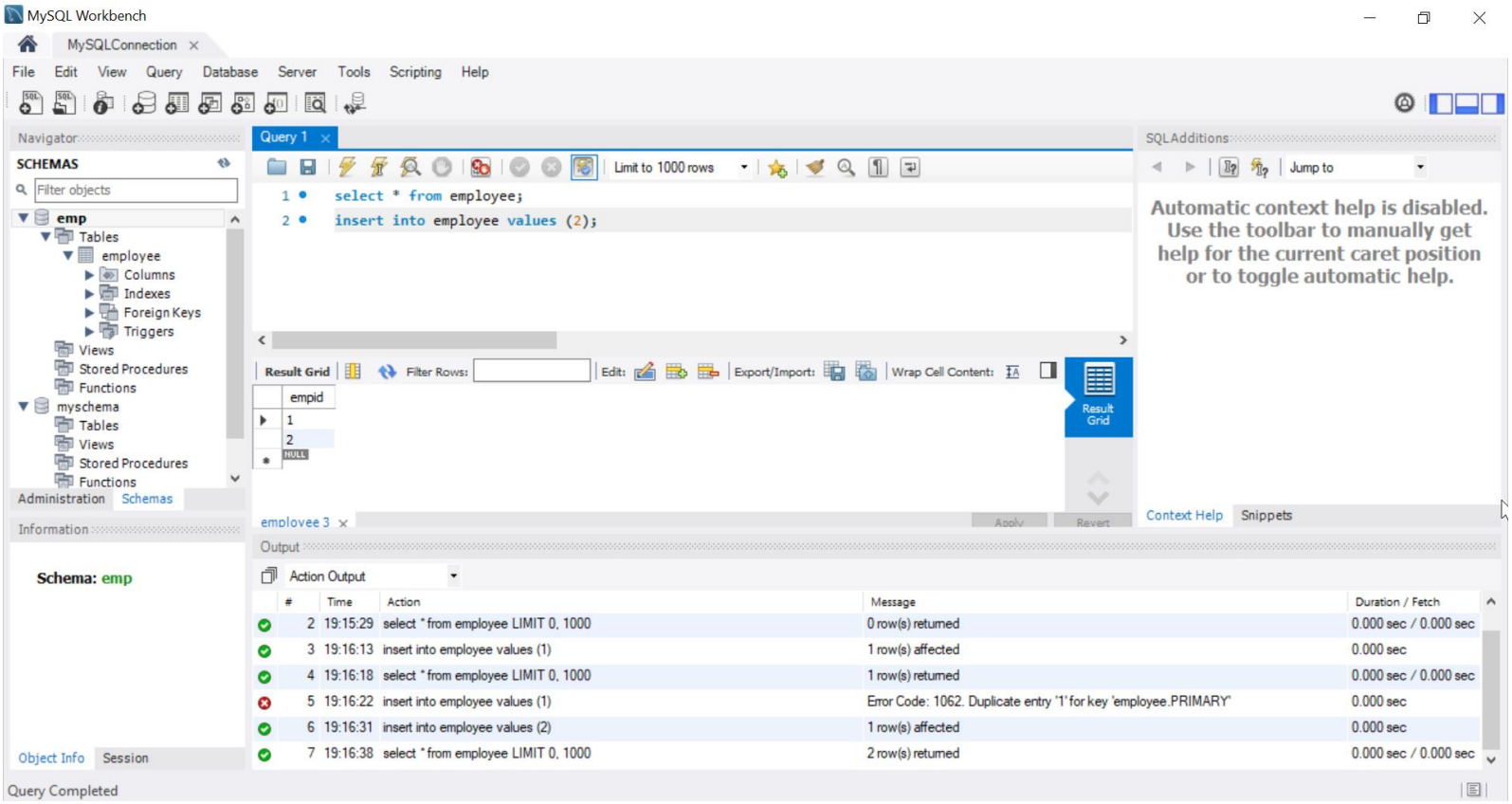
Relational Database: A Practical Foundation for Productivity

E. F. Codd
IBM San Jose Research Laboratory

https://dl.acm.org/doi/pdf/10.1145/1283920.1283937?download=true

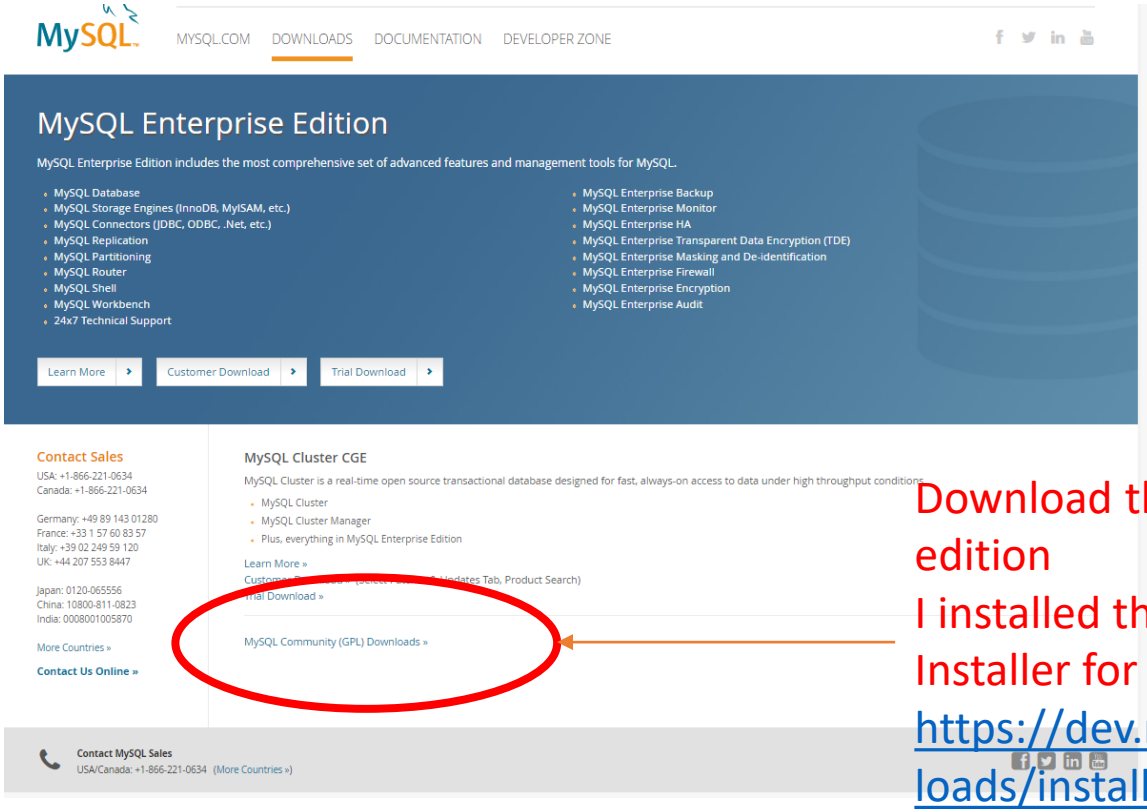
A “**DBMS**” named MySQL

MySQL - Tutorial

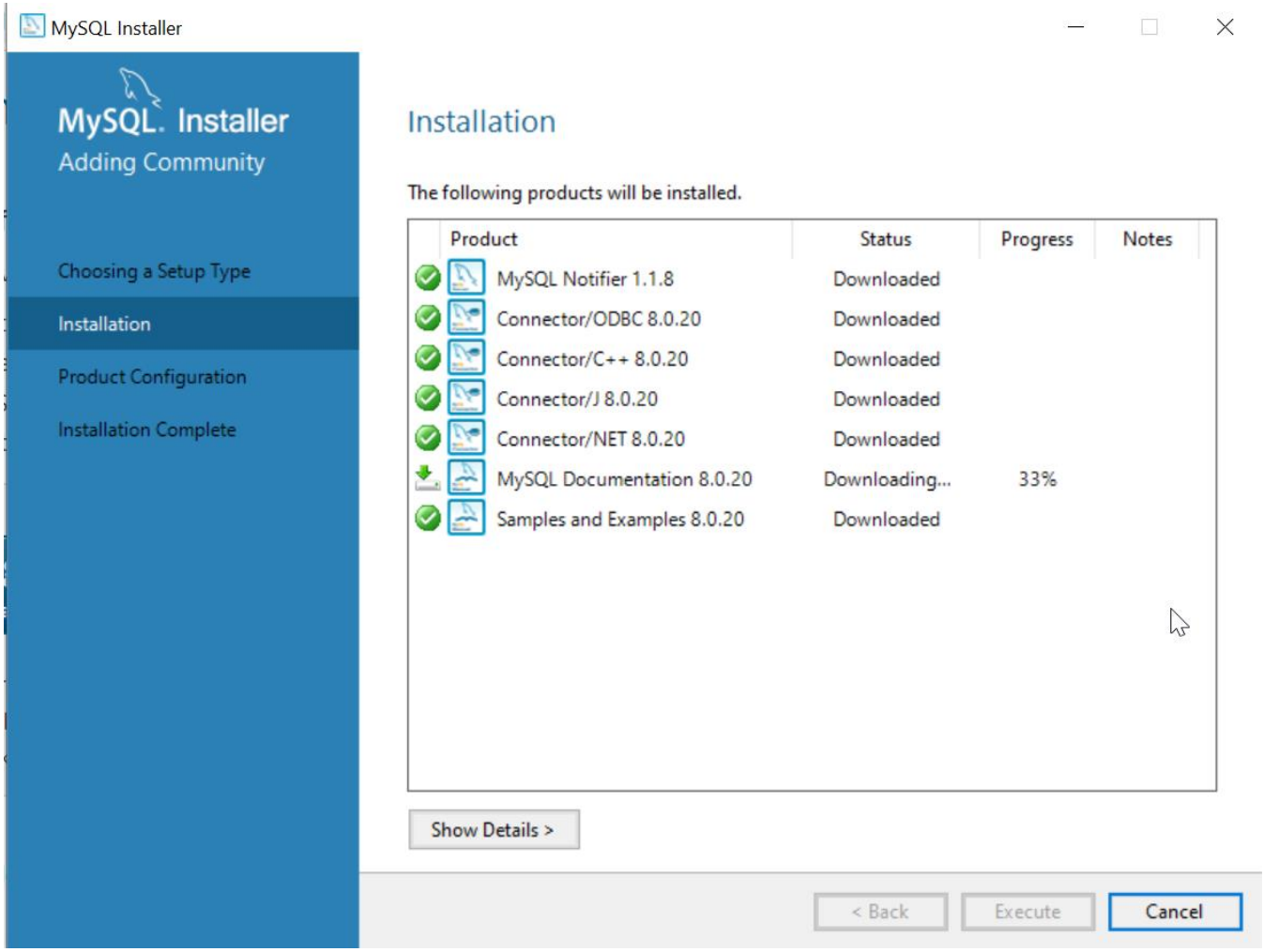


Install MySQL

- https://www.mysql.com/downloads/



Download the community edition
I installed the MySQL Installer for Windows from https://dev.mysql.com/downloads/installer/



You may need to install pre-requisites such as Visual C++ Redistributable for Visual Studio 2015. Visit https://www.microsoft.com/en-in/download/details.aspx?id=48145

Relations

Let R be the set of all binary relations on the set {1, 2, 3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is reflexive (round off to 3 decimal places) is _____ ?

[GATE 2020]

Relations

Let R be the set of all binary relations on the set {1, 2, 3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is reflexive (round off to 3 decimal places) is _____ ?

What is a reflexive relation?

- If A = {1,2,3}, then 1R1, 2R2, and 3R3.
  - Then R = {(1,1),(2,2),(3,3),(1,2)} is a reflexive relation on A.
  - R = {(1,1),(2,2),(3,3),(1,2),(3,1)} is also a RR.

Relations

Let R be the set of all binary relations on the set {1, 2, 3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is reflexive (round off to 3 decimal places) is ____ ?

Probability = #reflexive relations / #relations

- #reflexive relations = #relations that can be formed with {(1,1),(2,2),(3,3),…} = 2^6
- #relations = #relations that can be formed with {(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)} = 2^9

Therefore, Probability = $\frac{2^6}{2^9} = \frac{1}{2^3}$ = 0.125.

Relations

Let R be the set of all binary relations on the set {1, 2, 3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is symmetric (round off to 3 decimal places) is _____ ?

Relations

Let R be the set of all binary relations on the set {1, 2, 3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is symmetric (round off to 3 decimal places) is _____ ?

What is a symmetric relation?

- If A = {1,2,3}. Then R = {(1,1),(2,1),(3,3),(1,2)} is a symmetric relation on A.
- R = {(1,1),(2,2),(3,3),(1,2),(2,1)} is also an SR.
- In general, if $(a,b) \in R \Longrightarrow (b,a) \in R$

Relations

Let R be the set of all binary relations on the set {1, 2,3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is symmetric (round off to 3 decimal places) is ____ ?

Probability = # symmetric relations / #relations

- # symmetric relations = $2^{\frac{n(n+1)}{2}} = 2^6$
- #relations = #relations that can be formed with {(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)} = 2^9

Therefore, Probability = $\frac{2^6}{2^9} = \frac{1}{2^3} = 0.125.$

Notes

No. of reflexive relations $= 2^{n(n-1)}$

No. of irreflexive relations $= 2^{n(n-1)}$

No. of symmetric relations $= 2^{n(n+1)/2}$

No. of asymmetric relations $= 3^{n(n-1)/2}$

No. of Anti Symmetric Relations $= 2^{n} * 3^{n(n-1)/2}$

Antisymmetry is different from asymmetry: a relation is asymmetric if, and only if, it is antisymmetric and irreflexive.

We don't know how to generalize for transitivity.

See https://gateoverflow.in/120762/how-many-transitive-relations-are-there-set-with-elements%C2%A0if

Story So Far…

What is a Relation?

{●, ●, ●} x {▲, ▲} = { (●, ▲), (●, ▲), (●, ▲), (●, ▲), (●, ▲), (●, ▲) }

set A set B

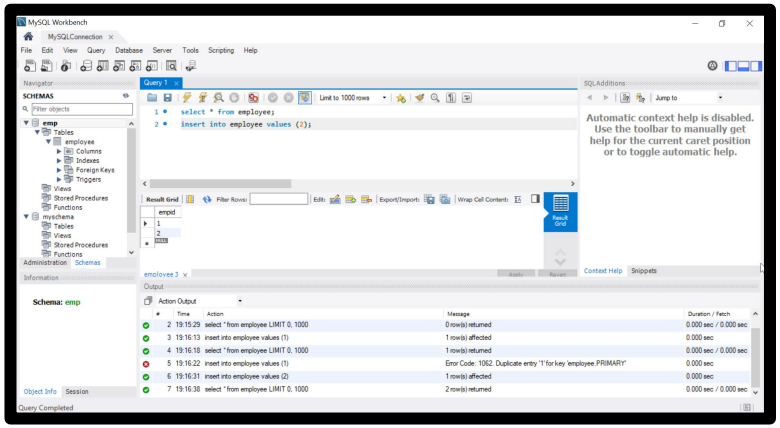
set of all ordered pairs, **A x B**

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

Let's say a relation exists between the reds:

Relation R = {(●, ▲)}

MySQL – An RDBMS



Relational Data Model

id	name
1	sd
1	vv
2	sd
2	vv
3	sd
3	vv

Vs

id	name
1	sd
2	vv

Attendance **relation**
{(1,1), (2,1)}
is same as the table

studentid	sessionid
1	1
2	1

Some problems to understand relations better…

No. of reflexive relations $= 2^{n(n-1)}$

No. of irreflexive relations $= 2^{n(n-1)}$

No. of symmetric relations $= 2^{n(n+1)/2}$

No. of asymmetric relations $= 3^{n(n-1)/2}$

No. of Anti Symmetric Relations $= 2^{n} * 3^{n(n-1)/2}$