1-9 关系及其基本性质 (II)

魏恒峰

hfwei@nju.edu.cn

2017 年 12 月 11 日


康托的无穷的
数学和哲学
(美)周·道本○著
郑毓信 刘晓力○编译
03
(珍藏版)
数学科学文化理念传播丛书(第一辑)
Georg Cantor: His Mathematics and Philosophy of The Infinite
大连理工大学出版社

The Relational Data Model

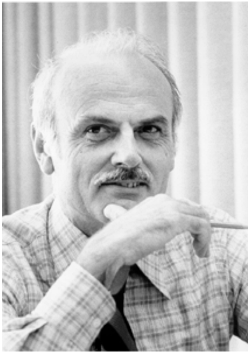
A Relational Model of Data for Large Shared Data Banks

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.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on *n*-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

Codd@CACM'1970



Edgar F. Codd (1923 – 2003)

The Relational Data Model

— 如何靠“关系”赢得图灵奖?

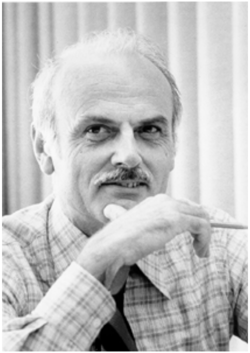
A Relational Model of Data for Large Shared Data Banks

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.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on *n*-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

Codd@CACM'1970



Edgar F. Codd (1923 – 2003)


1

2

3


Big ideas are usually simple ideas.
David Ogilvy

1

A Relational View of Data

$$(a, b) = \{\{a\}, \{a, b\}\}$$

$$R \subseteq A \times B$$

$$(a, b) = \{\{a\}, \{a, b\}\}$$

$$R \subseteq A \times B$$

$$(a, b, c) = ((a, b), c)$$

$$(a, b) = \{\{a\}, \{a, b\}\}$$

$$R \subseteq A \times B$$

$$(a, b, c) = ((a, b), c)$$

$$(x_1, x_2, \cdots, x_n) = ((x_1, x_2, \cdots, x_{n-1}), x_n)$$

$$(a, b) = \{\{a\}, \{a, b\}\}$$

$$R \subseteq A \times B$$

$$(a, b, c) = ((a, b), c)$$

$$(x_1, x_2, \cdots, x_n) = ((x_1, x_2, \cdots, x_{n-1}), x_n)$$

$$R \subseteq X_1 \times X_2 \times \cdots \times X_n$$

$$\begin{aligned}\textsf{Course} &= \{\mathrm{CS101}, \mathrm{EE200}, \mathrm{PH100}\} \\ \textsf{StudentId} &= \{12345, 67890, 22222, 33333\} \\ \textsf{Grade} &= \{A, B, C, D, A^-, B^+, C^+\}\end{aligned}$$

$$\text{Course} = \{\text{CS}101, \text{EE}200, \text{PH}100\}$$
$$\text{StudentId} = \{12345, 67890, 22222, 33333\}$$
$$\text{Grade} = \{A, B, C, D, A^-, B^+, C^+\}$$

$$\text{CSG} \subseteq \text{Course} \times \text{StudentId} \times \text{Grade}$$

$$\begin{aligned}\text{Course} &= \{\text{CS}101, \text{EE}200, \text{PH}100\}\\ \text{StudentId} &= \{12345, 67890, 22222, 33333\}\\ \text{Grade} &= \{A, B, C, D, A^-, B^+, C^+\}\end{aligned}$$

$$\text{CSG} \subseteq \text{Course} \times \text{StudentId} \times \text{Grade}$$

$$\begin{aligned}\text{CSG} = \{&(\text{CS}101, 12345, A),\\ &(\text{CS}101, 67890, B),\\ &(\text{CS}101, 33333, C),\\ &(\text{EE}200, 12345, B^+),\\ &(\text{EE}200, 22222, A^-),\\ &(\text{PH}100, 67890, C^+)\}\end{aligned}$$

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

Row : tuple

Column : attribute/component

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

StudentId	Name	Address	Phone
12345	C. Brown	12 Apple St.	555-1234
67890	L. Van Pelt	34 Pear Ave.	555-5678
22222	P. Patty	56 Grape Blvd.	555-9999

Course	Day	Hour
CS101	M	9AM
CS101	W	9AM
CS101	F	9AM
EE200	Tu	10AM
EE200	W	1PM
EE200	Th	10AM

Course	Room
CS101	Turing Aud.
EE200	25 Ohm Hall
PH100	Newton Lab.


Too simple?

2

Relational Algebra

Relational Algebra

Relational Algebra

Relational Algebra

A set of relations

Operations on relations

Laws of operations

$$\cup \quad \cap \quad \setminus$$

Selection ($\sigma_C(R)$)

$\sigma_{\text{Course} = \text{"CS101"}}(\text{CSG})$

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
CS101	33333	A−

Projection ($\pi_L(R)$)

$$\pi_{\text{StudentId}}\left(\sigma_{\text{Course} = \text{“CS101”}}(\text{CSG})\right)$$

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
CS101	33333	A−

StudentId
12345
67890
33333

Join $(R \underset{R.A_i = S.B_j}{\bowtie} S)$

$$R \subseteq A_1 \times \cdots \times A_n$$

$$S \subseteq B_1 \times \cdots \times B_m$$

$$R \underset{R.A_i = S.B_j}{\bowtie} S \subseteq A_1 \times \cdots \times A_n \times B_1 \times \cdots B_{j-1} \times B_{j+1} \times \cdots \times B_m$$

Join $(R \underset{R.A_i=S.B_j}{\bowtie} S)$

$$R \subseteq A_1 \times \cdots \times A_n$$

$$S \subseteq B_1 \times \cdots \times B_m$$

$$R \underset{R.A_i=S.B_j}{\bowtie} S \subseteq A_1 \times \cdots \times A_n \times B_1 \times \cdots B_{j-1} \times B_{j+1} \times \cdots \times B_m$$

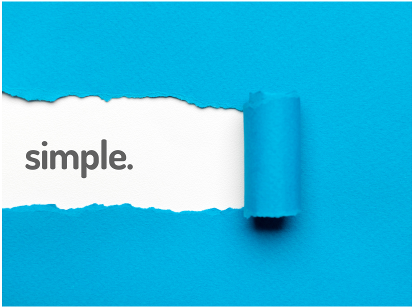
Natural Join $(R \bowtie S)$

$$CR \underset{CR.Course = CDH.Course}{\bowtie} CDH \qquad (CR \bowtie CDH)$$

Course	Room
CS101	Turing Aud.
EE200	25 Ohm Hall
PH100	Newton Lab.

Course	Day	Hour
CS101	M	9AM
CS101	W	9AM
CS101	F	9AM
EE200	Tu	10AM
EE200	W	1PM
EE200	Th	10AM

Course	Room	Day	Hour
CS101	Turing Aud.	M	9AM
CS101	Turing Aud.	W	9AM
CS101	Turing Aud.	F	9AM
EE200	25 Ohm Hall	Tu	10AM
EE200	25 Ohm Hall	W	1PM
EE200	25 Ohm Hall	Th	10AM


simple.

Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

StudentId	Name	Address	Phone
12345	C. Brown	12 Apple St.	555-1234
67890	L. Van Pelt	34 Pear Ave.	555-5678
22222	P. Patty	56 Grape Blvd.	555-9999

Course	Day	Hour
CS101	M	9AM
CS101	W	9AM
CS101	F	9AM
EE200	Tu	10AM
EE200	W	1PM
EE200	Th	10AM

Course	Room
CS101	Turing Aud.
EE200	25 Ohm Hall
PH100	Newton Lab.

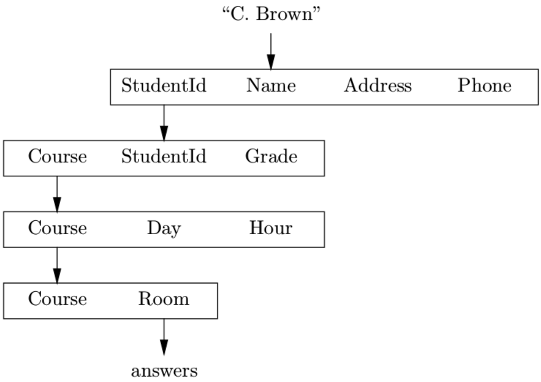
Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

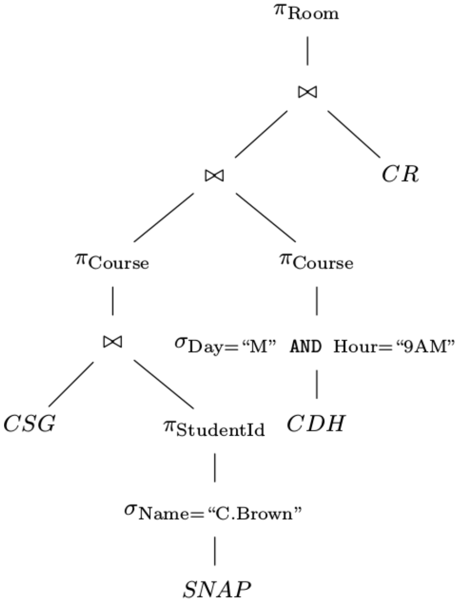
StudentId	Name	Address	Phone
12345	C. Brown	12 Apple St.	555-1234
67890	L. Van Pelt	34 Pear Ave.	555-5678
22222	P. Patty	56 Grape Blvd.	555-9999

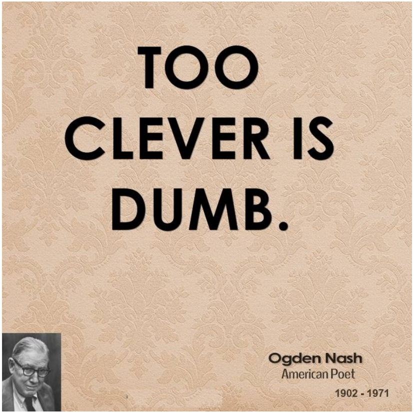
Course	Day	Hour
CS101	M	9AM
CS101	W	9AM
CS101	F	9AM
EE200	Tu	10AM
EE200	W	1PM
EE200	Th	10AM

Course	Room
CS101	Turing Aud.
EE200	25 Ohm Hall
PH100	Newton Lab.

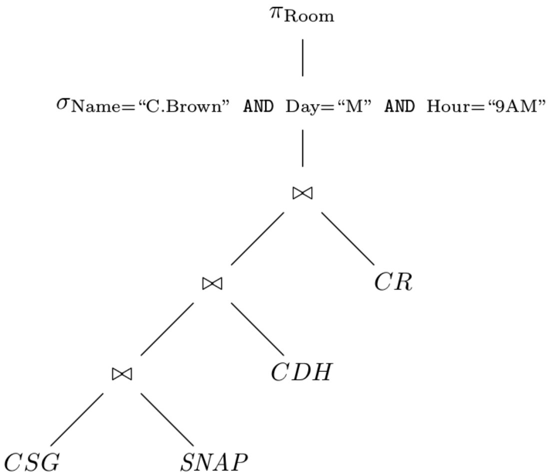
Q : Where *is* "C. Brown" 9 AM *on* Mondays?

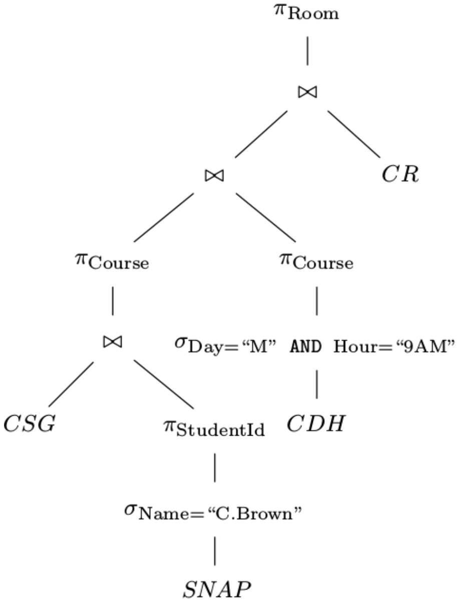

"C. Brown"
StudentId Name Address Phone
Course StudentId Grade
Course Day Hour
Course Room
answers


π_{Room}
$\bowtie$
$\bowtie$
CR
π_{Course}
π_{Course}
$\bowtie$
$\sigma_{\text{Day}=\text{"M"} \texttt{ AND } \text{Hour}=\text{"9AM"}}$
CSG
$\pi_{\text{StudentId}}$
CDH
$\sigma_{\text{Name}=\text{"C.Brown"}}$
$SNAP$


TOO
CLEVER IS
DUMB.
Ogden Nash
American Poet
1902 - 1971

Algebraic Laws for Relations


π_{Room}
$\sigma_{\text{Name=“C.Brown” AND Day=“M” AND Hour=“9AM”}}$
$\bowtie$
$\bowtie$
CR
$\bowtie$
CDH
CSG
$SNAP$


π_{Room}
$\bowtie$
$\bowtie$
CR
π_{Course}
π_{Course}
$\bowtie$
$\sigma_{\text{Day=“M” AND Hour=“9AM”}}$
CSG
$\pi_{\text{StudentId}}$
CDH
$\sigma_{\text{Name=“C.Brown”}}$
$SNAP$

"In fact, the improvement in efficiency made possible by transforming expressions of relational algebra is arguably the most striking example of the power of algebra that we find in computer science."

Laws Involving Selection ($\sigma_C(R)$)

Selection Splitting

$$\sigma_{C \wedge C'}(R) \equiv \sigma_C(\sigma_{C'}(R))$$

Commutativity

$$\sigma_C(\sigma_{C'}(R)) \equiv \sigma_{C'}(\sigma_C(R))$$

Laws Involving Selection ($\sigma_C(R)$)

Selection Splitting

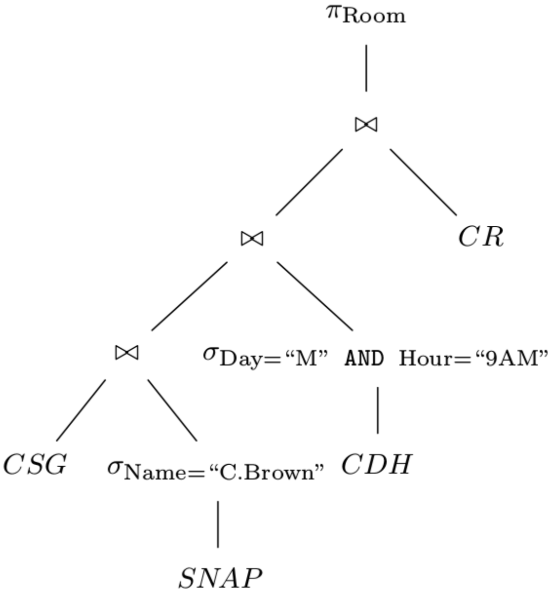
$$\sigma_{C \wedge C'}(R) \equiv \sigma_C(\sigma_{C'}(R))$$

Commutativity

$$\sigma_C(\sigma_{C'}(R)) \equiv \sigma_{C'}(\sigma_C(R))$$

Selection Pushing

$$(\sigma_C(R \bowtie S)) \equiv (\sigma_C(R) \bowtie S)$$
$$(\sigma_C(R \bowtie S)) \equiv (R \bowtie \sigma_C(S))$$

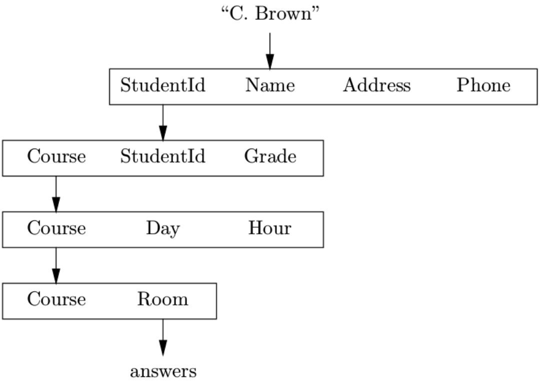

π_{Room}
$\bowtie$
$\bowtie$
CR
$\bowtie$
$\sigma_{\text{Day}=\text{“M”}\ \texttt{AND}\ \text{Hour}=\text{“9AM”}}$
CSG
$\sigma_{\text{Name}=\text{“C.Brown”}}$
CDH
$SNAP$

Laws Involving Projection ($\pi_L(R)$)

$$\Big(\pi_L(R \underset{A_i = B_j}{\bowtie} S)\Big) \equiv \Big(\pi_L\big(\pi_M(R) \underset{A_i = B_j}{\bowtie} \pi_N(S)\big)\Big)$$

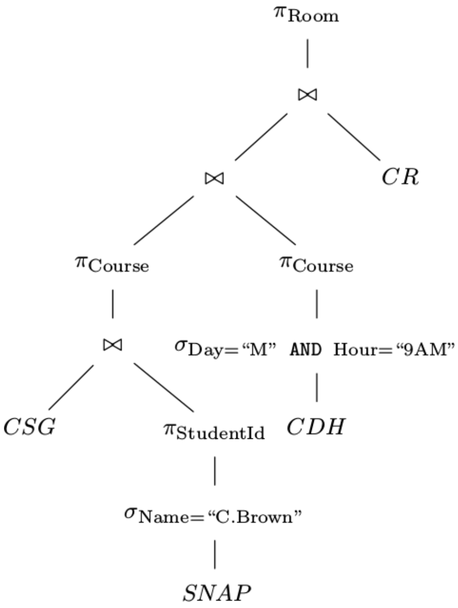
$$M = \Big(L \cup \{A_i\}\Big) \cap \{A_i\}_i$$

$$N = \Big(L \cup \{B_j\}\Big) \cap \{B_j\}_j$$

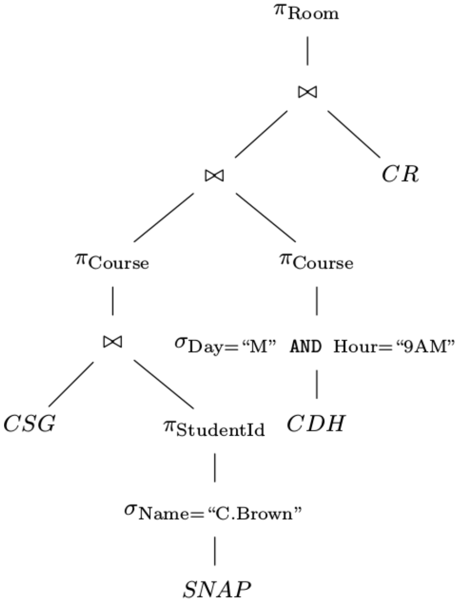


$$\pi_{\text{Course}}\Big(\text{CSG} \bowtie_{\text{StudentId}} \text{SNAP}\Big)$$

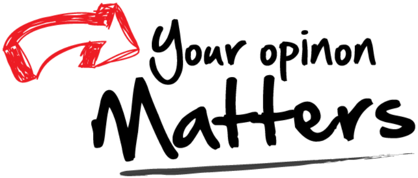
$$\equiv \pi_{\text{Course}}\Big(\pi_{\text{Course,StudentId}}(\text{CSG}) \bowtie_{\text{StudentId}} \pi_{\text{StudentId}}(\text{SNAP})\Big)$$


π_{Room}
$\bowtie$
$\bowtie$
CR
π_{Course}
π_{Course}
$\bowtie$
$\sigma_{\text{Day}=\text{"M"} \ \texttt{AND} \ \text{Hour}=\text{"9AM"}}$
CSG
$\pi_{\text{StudentId}}$
CDH
$\sigma_{\text{Name}=\text{"C.Brown"}}$
$SNAP$

Algebra as Language


π_{Room}
$\bowtie$
$\bowtie$
CR
π_{Course}
π_{Course}
$\bowtie$
$\sigma_{\text{Day=“M” AND Hour=“9AM”}}$
CSG
$\pi_{\text{StudentId}}$
CDH
$\sigma_{\text{Name=“C.Brown”}}$
$SNAP$


Thank
You!



Office 302

Mailbox: H016

hfwei@nju.edu.cn