1-9 关系及其基本性质 (II)

魏恒峰

hfwei@nju.edu.cn

2017 年 12 月 18 日

了呢？如果要详细地回答这

此只能举例一二，点到为止.

现在计算机专业的大学一、二年级学生，普遍不愿意学习逻辑演算与集合论课程，认为相关内容与计算机专业没有什么用. 那么我们的教育管理部门和相关专业人士又是如何认知的呢？据我所知，南京大学早年不仅要给计算机专业本科生开设这两门课程，而且还要开设递归论和模型论课程. 然而随着思维模式的不断转移，不仅递归论和模型论早已停开，而且逻辑演算与集合论课程的学时数也在逐步缩减. 现在国内坚持开设这两门课的高校已经很少了，大部分高校只在离散数学课程中，给学生讲很少一点逻辑演算与集合论知识. 其实，相关知识对于培养计算机专业的高科技人才来说是至关重要的，即使不谈这是最起码的专业文化素养，难道不明白我们所学之程序设计语言是靠逻辑设计出来的？而且柯特(E. P. Codd)博士创立关系数据库，以及许华兹(J. T. Schwartz)教授开发的集合论程序设计语言 SETL，可谓全都依靠数理逻辑与集合论知识的积累. 但却很少有专业教师能从历史的角度并依此为例去教育学生，甚至还有极个别的专家教授，竟然主张把“计算机科学理论”这门硕士研究生学位课取消，认为这门课相对于毕业后去公司就业的学生太空洞，这真是令人瞠目结舌. 特别是对于那些初涉高等学府的学子来说，其严重性

The Relational Data Model

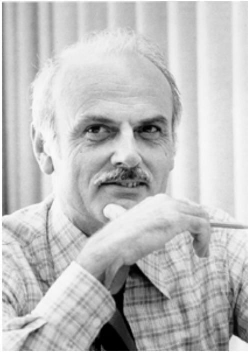
A Relational Model of Data for Large Shared Data Banks

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

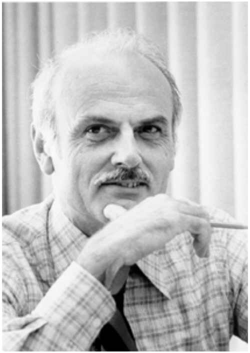
— 如何靠“关系”赢得图灵奖?

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
quotefancy

1

A Relational View of Data

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

$$R \subseteq A \times B$$

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

$$R \subseteq A \times B$$

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

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

$$R \subseteq A \times B$$

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

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

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

$$R \subseteq A \times B$$

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

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

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

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

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

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

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

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

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

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

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

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

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

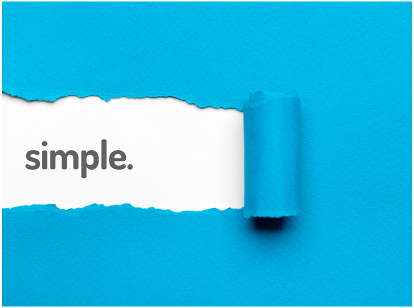
Natural Join ($R \bowtie S$)

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

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.

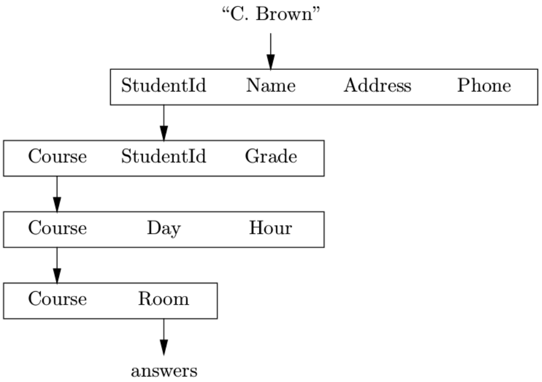
Course	StudentId	Grade
CS101	12345	A
CS101	67890	B
EE200	12345	C
EE200	22222	B+
CS101	33333	A−
PH100	67890	C+

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

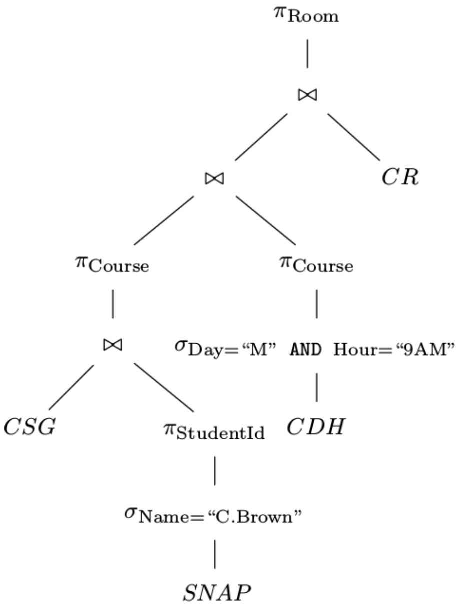
Course	Day	Hour
CS101	M	9AM
CS101	W	9AM
CS101	F	9AM
EE200	Tu	10AM
EE200	W	1PM
EE200	Th	10AM

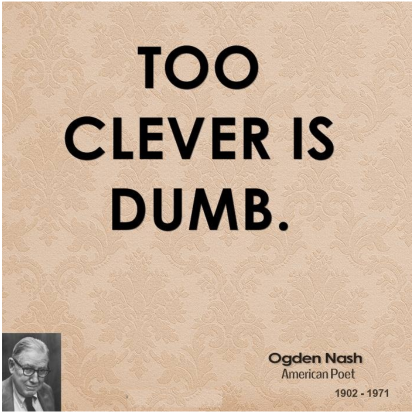
Course	Room
CS101	Turing Aud.
EE200	25 Ohm Hall
PH100	Newton Lab.

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

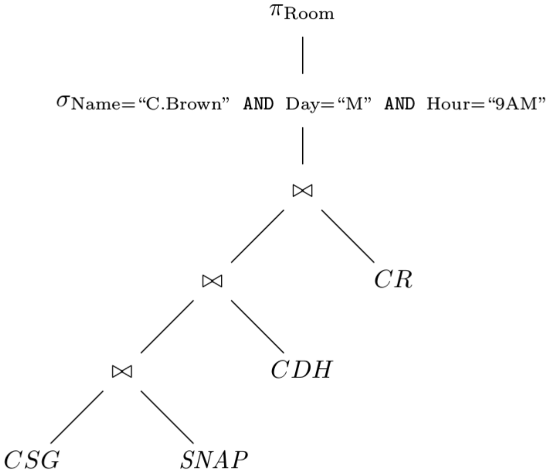

"C. Brown"
StudentId Name Address Phone
Course StudentId Grade
Course Day Hour
Course Room
answers

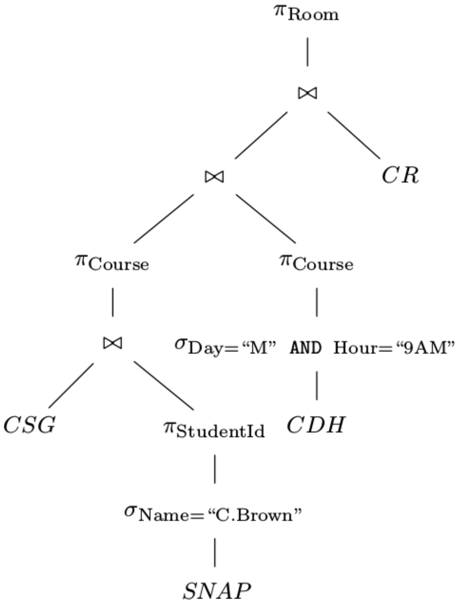
```
π Room (
  π Course (
    π StudentId (
      σ Name = 'C. Brown' (SNAP)
    ) ⋈ (CSG)
  ) ⋈ (
    π Course (
      σ Day = 'M' ∧ Hour = '9AM' (CDH)
    )
  ) ⋈ CR
)
```


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


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

"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

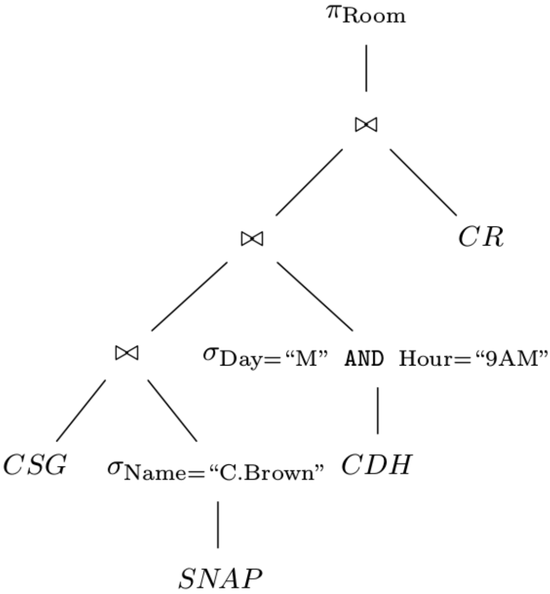
$$\sigma_{C \wedge C'}(R) \equiv \sigma_C(\sigma_{C'}(R))$$

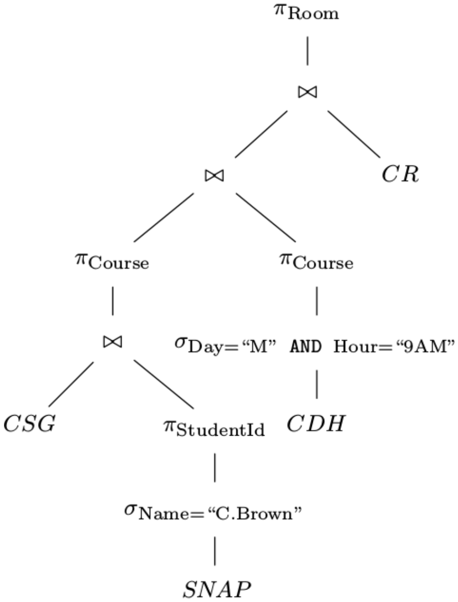
Commutativity

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

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
$SNAP$


π Room
⋈
⋈
CR
π Course
π Course
⋈
σ Day="M" AND Hour="9AM"
CSG
π StudentId
CDH
σ Name="C.Brown"
SNAP

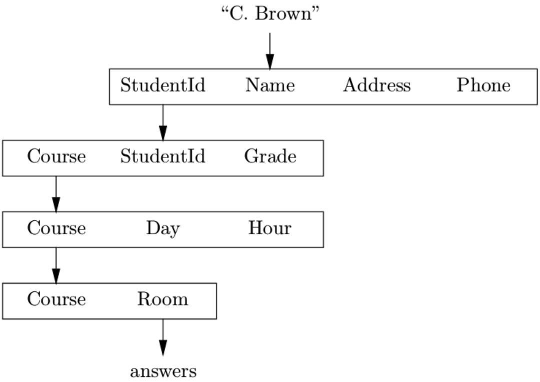
Laws Involving Projection ($\pi_L(R)$)

Projection Pushing

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

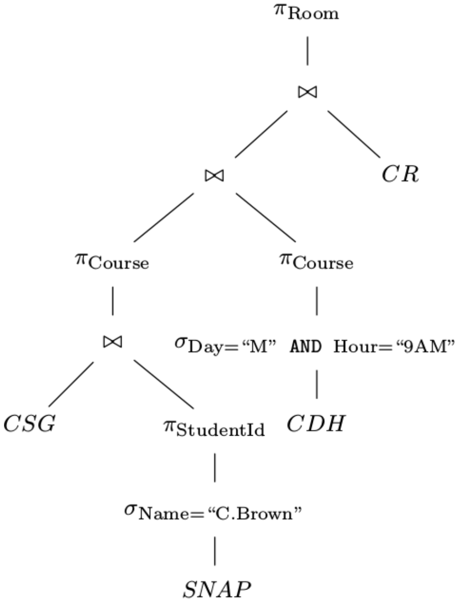
$$M = \Big(L \cup \{A_i\}\Big) \cap \{A_i\}_i$$

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



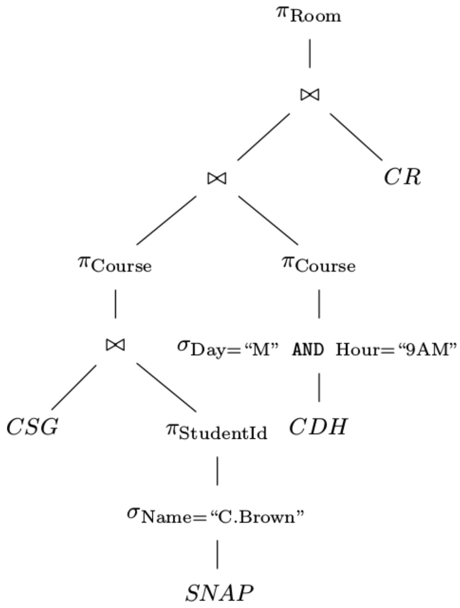
$$\pi_{\text{Course}}\left(\text{CSG} \underset{\text{StudentId}}{\bowtie} \text{SNAP}\right)$$

$$\equiv \pi_{\text{Course}}\left(\pi_{\text{Course,StudentId}}(\text{CSG}) \underset{\text{StudentId}}{\bowtie} \pi_{\text{StudentId}}(\text{SNAP})\right)$$


π Room
⋈
⋈
CR
π Course
π Course
⋈
σ Day="M" AND Hour="9AM"
CSG
π StudentId
CDH
σ Name="C.Brown"
SNAP

3

Algebra as Query Language


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

CourseManage...

CSG
- Course string
- StudentId number
- Grade string

SNAP
- StudentId number
- Name string
- Address string
- Phone string

CDH
- Course string
- Day string
- Hour string

CR

Relational Algebra | SQL

π σ ρ ← τ γ ∧ ∨ ¬ = ≠ ≥ ≤ ∩ ∪ ÷ - × ⋈

```
π Room (
  π Course (
    π StudentId (
      σ Name = 'C. Brown' (SNAP)
    ) ⋈ (CSG)
  ) ⋈ (
    π Course (
      σ Day = 'M' ∧ Hour = '9AM' (CDH)
    )
  ) ⋈ CR
)
```

execute query

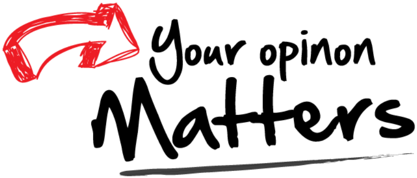
Structured Query Language

Structured Query Language

Declarative Programming


1
2
3


Thank
You!



Office 302

Mailbox: H016

hfwei@nju.edu.cn