


CMU SCS

Carnegie Mellon Univ.
School of Computer Science
15-415/615 - DB Applications

C. Faloutsos & A. Pavlo
Lecture #4: *Relational Algebra*




CMU SCS

Overview

- history
- concepts
- Formal query languages
 - relational algebra
 - rel. tuple calculus
 - rel. domain calculus

Faloutsos - Pavlo CMU SCS 15-415/615 #2




CMU SCS

History

- before: records, pointers, sets etc
- introduced by E.F. Codd in 1970
- revolutionary!
- first systems: 1977-8 (System R; Ingres)
- Turing award in 1981

Faloutsos - Pavlo CMU SCS 15-415/615 #3



CMU SCS

Concepts - reminder

- Database: a set of relations (= tables)
- rows: tuples
- columns: attributes (or keys)
- superkey, candidate key, primary key

Faloutsos - Pavlo CMU SCS 15-415/615 #4

Example

Database:

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

SSN	c-id	grade
123	15-413	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #5

Example: cont'd

Database:

k-th attribute
(D_k domain)

↓

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

rel. schema (attr+domains)

← tuple

Faloutsos - Pavlo CMU SCS 15-415/615 #6

Example: cont'd

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

rel. schema (attr+domains)

↕ instance

Faloutsos - Pavlo CMU SCS 15-415/615 #7

Example: cont'd


- D_i: the domain of the i-th attribute (eg., char(10))

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

rel. schema (attr+domains)

↕ instance

Faloutsos - Pavlo CMU SCS 15-415/615 #8




CMU SCS

Overview

- history
- concepts
- **Formal query languages**
 - relational algebra
 - rel. tuple calculus
 - rel. domain calculus

Faloutsos - Pavlo CMU SCS 15-415/615 #9




CMU SCS

Formal query languages

- How do we collect information?
- Eg., find ssn's of people in 415
- (recall: everything is a set!)
- One solution: Rel. algebra, ie., set operators
- Q1: Which ones??
- Q2: what is a minimal set of operators?

Faloutsos - Pavlo CMU SCS 15-415/615 #10




CMU SCS

Relational operators

- .
- .
- .
- set union \cup
- set difference \setminus

Faloutsos - Pavlo CMU SCS 15-415/615 #11



CMU SCS

Example:

- Q: find all students (part or full time)
- A: PT-STUDENT union FT-STUDENT

FT-STUDENT		
Ssn	Name	
129	peters	main str
239	lee	5th ave

PT-STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

Faloutsos - Pavlo CMU SCS 15-415/615 #12

CMU SCS

Observations:


- two tables are ‘union compatible’ if they have the same attributes (‘domains’)
- Q: how about intersection \cap

Faloutsos - Pavlo CMU SCS 15-415/615 #13

CMU SCS

Observations:

- A: redundant:
- STUDENT intersection STAFF =

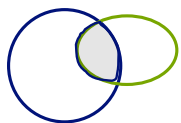


Faloutsos - Pavlo CMU SCS 15-415/615 #14

CMU SCS

Observations:

- A: redundant:
- STUDENT intersection STAFF =

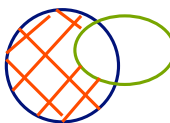


Faloutsos - Pavlo CMU SCS 15-415/615 #15

CMU SCS

Observations:

- A: redundant:
- STUDENT intersection STAFF = STUDENT - (STUDENT - STAFF)



Faloutsos - Pavlo CMU SCS 15-415/615 #16

CMU SCS

Observations:

- A: redundant:
- $STUDENT \cap STAFF = STUDENT - (STUDENT - STAFF)$

Double negation:
We'll see it again, later...

Faloutsos - Pavlo CMU SCS 15-415/615 #17

CMU SCS

Relational operators

- .
- .
- .
- set union \cup
- set difference $'-'$

Faloutsos - Pavlo CMU SCS 15-415/615 #18

CMU SCS

Other operators?

- eg, find all students on 'Main street'
- A: 'selection'

$$\sigma_{address='main str'} (STUDENT)$$

Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

Faloutsos - Pavlo CMU SCS 15-415/615 #19

CMU SCS

Other operators?

- Notice: selection (and rest of operators) expect tables, and produce tables (-> can be cascaded!!)
- For selection, in general:

$$\sigma_{condition} (RELATION)$$

Faloutsos - Pavlo CMU SCS 15-415/615 #20

CMU SCS

Selection - examples

- Find all 'Smiths' on 'Forbes Ave'

$$\sigma_{name='Smith' \wedge address='Forbes ave'}(STUDENT)$$

'condition' can be any boolean combination of '=', '>', '>=', ...

Faloutsos - Pavlo CMU SCS 15-415/615 #21

CMU SCS

Relational operators

- selection $\sigma_{condition} (R)$
- .
- .
- set union $R \cup S$
- set difference $R - S$

Faloutsos - Pavlo CMU SCS 15-415/615 #22

CMU SCS

Relational operators

- selection picks rows - how about columns?
- A: 'projection' - eg.: $\pi_{ssn}(STUDENT)$

finds all the 'ssn' - removing duplicates

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

Faloutsos - Pavlo CMU SCS 15-415/615 #23

CMU SCS

Relational operators

Cascading: 'find ssn of students on 'forbes ave'

$$\pi_{ssn}(\sigma_{address='forbes ave'}(STUDENT))$$

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

Faloutsos - Pavlo CMU SCS 15-415/615 #24

CMU SCS

Relational operators

- selection $\sigma_{condition}(R)$
- projection $\pi_{att-list}(R)$
- .
- set union $R \cup S$
- set difference $R - S$

Faloutsos - Pavlo CMU SCS 15-415/615 #25

CMU SCS

Relational operators

Are we done yet?
Q: Give a query we can **not** answer yet!

Faloutsos - Pavlo CMU SCS 15-415/615 #26

CMU SCS

Relational operators

A: any query across **two** or more tables,
eg., 'find names of students in 15-415'
Q: what extra operator do we need??

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

TAKES		
SSN	c-id	grade
123	15-413	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #27

CMU SCS

Relational operators

A: any query across **two** or more tables,
eg., 'find names of students in 15-415'
Q: what extra operator do we need??
A: surprisingly, cartesian product is enough!

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

TAKES		
SSN	c-id	grade
123	15-413	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #28

Cartesian product

- eg., dog-breeding: MALE x FEMALE
- gives all possible couples

MALE
name
spike
spot

x

FEMALE
name
lassie
shiba

=

M.name	F.name
spike	lassie
spike	shiba
spot	lassie
spot	shiba

Faloutsos - Pavlo CMU SCS 15-415/615 #29

so what?

- Eg., how do we find names of students taking 415?

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

SSN	c-id	grade
123	15-415	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #30

Cartesian product

- A: $\sigma_{STUDENT.ssn=TAKES.ssn}(STUDENT \times TAKES)$

Ssn	Name	Address	ssn	cid	grade
123	smith	main str	123	15-415	A
234	jones	forbes ave	123	15-415	A
123	smith	main str	234	15-413	B
234	jones	forbes ave	234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #31

Cartesian product

- $\sigma_{cid=15-415}(\sigma_{STUDENT.ssn=TAKES.ssn}(STUDENT \times TAKES))$

Ssn	Name	Address	ssn	cid	grade
123	smith	main str	123	15-415	A
234	jones	forbes ave	123	15-415	A
123	smith	main str	234	15-413	B
234	jones	forbes ave	234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #32

CMU SCS

$$\pi_{\text{name}}(\sigma_{\text{cid}=15-415}(\sigma_{\text{STUDENT.ssn}=\text{TAKES.ssn}}(\text{STUDENT} \times \text{TAKES})))$$

Ssn	Name	Address	ssn	cid	grade
123	smith	main str	123	15-415	A
234	jones	forbes ave	123	15-415	A
123	smith	main str	234	15-413	B
234	jones	forbes ave	234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #33

CMU SCS

FUNDAMENTAL Relational operators

• selection	$\sigma_{\text{condition}}(R)$
• projection	$\pi_{\text{att-list}}(R)$
• cartesian product	MALE \times FEMALE
• set union	$R \cup S$
• set difference	$R - S$

Faloutsos - Pavlo CMU SCS 15-415/615 #34

CMU SCS

Relational ops

- Surprisingly, they are enough, to help us answer almost any query we want!!
- derived/convenience operators:
 - set intersection
 - join** (theta join, equi-join, natural join) \bowtie
 - 'rename' operator $\rho_{R'}(R)$
 - division $R \div S$

Faloutsos - Pavlo CMU SCS 15-415/615 #35

CMU SCS

Joins

- Equijoin: $R \bowtie_{R.a=S.b} S = \sigma_{R.a=S.b}(R \times S)$

Faloutsos - Pavlo CMU SCS 15-415/615 #36

CMU SCS

Cartesian product

- A: $\sigma_{STUDENT.ssn=TAKES.ssn} (STUDENT \times TAKES)$

Ssn	Name	Address	ssn	cid	grade
123	smith	main str	123	15-415	A
234	jones	forbes ave	123	15-415	A
123	smith	main str	234	15-413	B
234	jones	forbes ave	234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #37

CMU SCS

Joins

- Equijoin: $R \bowtie_{R.a=S.b} S = \sigma_{R.a=S.b} (R \times S)$
- theta-joins: $R \bowtie_{\theta} S$
generalization of equi-join - any condition θ

Faloutsos - Pavlo CMU SCS 15-415/615 #38

CMU SCS

Joins

- very** popular: natural join: $R \bowtie S$
- like equi-join, but it drops duplicate columns:
STUDENT (ssn, name, address)
TAKES (ssn, cid, grade)

Faloutsos - Pavlo CMU SCS 15-415/615 #39

CMU SCS

Joins

- nat. join** has 5 attributes $STUDENT \bowtie TAKES$

Ssn	Name	Address	ssn	cid	grade
123	smith	main str	123	15-415	A
234	jones	forbes ave	123	15-415	A
123	smith	main str	234	15-413	B
234	jones	forbes ave	234	15-413	B

$\text{equi-join: } 6 \text{ } STUDENT \bowtie_{STUDENT.ssn=TAKES.ssn} TAKES$

Faloutsos - Pavlo CMU SCS 15-415/615 #40

CMU SCS

Natural Joins - nit-picking

- if no attributes in common between R, S:
nat. join \rightarrow cartesian product

Faloutsos - Pavlo CMU SCS 15-415/615 #41

CMU SCS

Overview - rel. algebra

- fundamental operators
- derived operators
 - joins etc
 - rename**
 - division
- examples

Faloutsos - Pavlo CMU SCS 15-415/615 #42

CMU SCS

Rename op.

- Q: why? $\rho_{AFTER}(BEFORE)$
- A: shorthand; self-joins; ...
- for example, find the grand-parents of 'Tom', given PC (parent-id, child-id)

Faloutsos - Pavlo CMU SCS 15-415/615 #43

CMU SCS

Rename op.

- PC (parent-id, child-id) $PC \bowtie PC$

PC		PC	
p-id	c-id	p-id	c-id
Mary	Tom	Mary	Tom
Peter	Mary	Peter	Mary
John	Tom	John	Tom

Faloutsos - Pavlo CMU SCS 15-415/615 #44

CMU SCS

Rename op.

- first, WRONG attempt:

$$PC \bowtie PC$$
- (why? how many columns?)
- Second WRONG attempt:

$$PC \bowtie_{PC.c=id=PC.p=id} PC$$

Faloutsos - Pavlo CMU SCS 15-415/615 #45

CMU SCS

Rename op.

- we clearly need two different names for the same table - hence, the 'rename' op.

$$\rho_{PC1}(PC) \bowtie_{PC1.c=id=PC.p=id} PC$$

Faloutsos - Pavlo CMU SCS 15-415/615 #46

CMU SCS

Overview - rel. algebra

- fundamental operators
- derived operators
 - joins etc
 - rename
 - **division**
- examples

Faloutsos - Pavlo CMU SCS 15-415/615 #47

CMU SCS

Division

- Rarely used, but powerful.
- Example: find suspicious suppliers, ie., suppliers that supplied **all** the parts in A_BOMB

Faloutsos - Pavlo CMU SCS 15-415/615 #48

Division

SHIPMENT	
s#	p#
s1	p1
s2	p1
s1	p2
s3	p1
s5	p3

÷

ABOMB	
p#	
p1	
p2	

=

BAD_S	
s#	
s1	

Faloutsos - Pavlo CMU SCS 15-415/615 #49

Division

- Observations: ~reverse of cartesian product
- It can be derived from the 5 fundamental operators (!!)
- How?

Faloutsos - Pavlo CMU SCS 15-415/615 #50

Division

- Answer:

$$r \div S = \pi_{(R-S)}(r) - \pi_{(R-S)}[(\pi_{(R-S)}(r) \times S) - r]$$

- Observation: find ‘good’ suppliers, and subtract! (**double negation**)

Faloutsos - Pavlo CMU SCS 15-415/615 #51

Division

SHIPMENT	
s#	p#
s1	p1
s2	p1
s1	p2
s3	p1
s5	p3

÷

ABOMB	
p#	
p1	
p2	

=

BAD_S	
s#	
s1	

- Answer:

$$r \div S = \pi_{(R-S)}(r) - \pi_{(R-S)}[(\pi_{(R-S)}(r) \times S) - r]$$

- Observation: find ‘good’ suppliers, and subtract! (**double negation**)

Faloutsos - Pavlo CMU SCS 15-415/615 #52

CMU SCS

Division

• Answer:

SHIPMENT	
s#	p#
s1	p1
s2	p1
s1	p2
s3	p1
s5	p3

ABOMB
p#
p1
p2

BAD_S
s#
s1

$$r \div S = \pi_{(R-S)}(r) - \pi_{(R-S)}[(\pi_{(R-S)}(r) \times S) - r]$$

←→ ←→
All suppliers
All bad parts

Faloutsos - Pavlo CMU SCS 15-415/615 #53

CMU SCS

Division

• Answer:

SHIPMENT	
s#	p#
s1	p1
s2	p1
s1	p2
s3	p1
s5	p3

ABOMB
p#
p1
p2

BAD_S
s#
s1

$$r \div S = \pi_{(R-S)}(r) - \pi_{(R-S)}[(\pi_{(R-S)}(r) \times S) - r]$$

←→ ←→
←→
all possible suspicious shipments

Faloutsos - Pavlo CMU SCS 15-415/615 #54

CMU SCS

Division

• Answer:

SHIPMENT	
s#	p#
s1	p1
s2	p1
s1	p2
s3	p1
s5	p3

ABOMB
p#
p1
p2

BAD_S
s#
s1

$$r \div S = \pi_{(R-S)}(r) - \pi_{(R-S)}[(\pi_{(R-S)}(r) \times S) - r]$$

←→ ←→
←→
←→
all possible suspicious shipments that didn't happen

Faloutsos - Pavlo CMU SCS 15-415/615 #55

CMU SCS

Division

• Answer:

SHIPMENT	
s#	p#
s1	p1
s2	p1
s1	p2
s3	p1
s5	p3

ABOMB
p#
p1
p2

BAD_S
s#
s1

$$r \div S = \pi_{(R-S)}(r) - \pi_{(R-S)}[(\pi_{(R-S)}(r) \times S) - r]$$

←→ ←→
←→
←→
←→
all suppliers who missed at least one suspicious shipment, i.e.: 'good' suppliers

Faloutsos - Pavlo CMU SCS 15-415/615 #56

CMU SCS

Overview - rel. algebra

- fundamental operators
- derived operators
 - joins etc
 - rename
 - division
- **examples**

Faloutsos - Pavlo CMU SCS 15-415/615 #57

CMU SCS

Sample schema

find names of students that take 15-415

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

CLASS		
c-id	c-name	units
15-413	s.e.	2
15-412	o.s.	2

TAKES		
SSN	c-id	grade
123	15-413	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #58

CMU SCS

Examples

- find names of students that take 15-415

Faloutsos - Pavlo CMU SCS 15-415/615 #59

CMU SCS

Examples

- find names of students that take 15-415

$$\pi_{name} [\sigma_{c-id=15-415} (STUDENT \bowtie TAKES)]$$

Faloutsos - Pavlo CMU SCS 15-415/615 #60

CMU SCS

Sample schema

find course names of 'smith'

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

CLASS		
c-id	c-name	units
15-413	s.e.	2
15-412	o.s.	2

TAKES		
SSN	c-id	grade
123	15-413	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #61

CMU SCS

Examples

- find course names of 'smith'

$$\pi_{c-name} [\sigma_{name='smith'} ($$

STUDENT ⋈ *TAKES* ⋈ *CLASS*

$$)]$$

Faloutsos - Pavlo CMU SCS 15-415/615 #62

CMU SCS

Examples

- find ssn of 'overworked' students, ie., that take 412, 413, 415

Faloutsos - Pavlo CMU SCS 15-415/615 #63

CMU SCS

Examples

- find ssn of 'overworked' students, ie., that take 412, 413, 415: almost correct answer:

$$\sigma_{c-name=412} (TAKES) \cap$$

$$\sigma_{c-name=413} (TAKES) \cap$$

$$\sigma_{c-name=415} (TAKES)$$

Faloutsos - Pavlo CMU SCS 15-415/615 #64

CMU SCS

Examples

- find ssn of ‘overworked’ students, ie., that take 412, 413, 415 - Correct answer:

$$\pi_{ssn}[\sigma_{c-name=412}(TAKES)] \cap$$

$$\pi_{ssn}[\sigma_{c-name=413}(TAKES)] \cap$$

$$\pi_{ssn}[\sigma_{c-name=415}(TAKES)]$$

Faloutsos - Pavlo CMU SCS 15-415/615 #65

CMU SCS

Examples

- find ssn of students that work at least as hard as ssn=123, ie., they take all the courses of ssn=123, and maybe more

Faloutsos - Pavlo CMU SCS 15-415/615 #66

CMU SCS

Sample schema

STUDENT		
Ssn	Name	Address
123	smith	main str
234	jones	forbes ave

CLASS		
c-id	c-name	units
15-413	s.e.	2
15-412	o.s.	2

TAKES		
SSN	c-id	grade
123	15-413	A
234	15-413	B

Faloutsos - Pavlo CMU SCS 15-415/615 #67


CMU SCS

Examples

- find ssn of students that work at least as hard as ssn=123 (ie., they take all the courses of ssn=123, and maybe more

$$[\pi_{ssn,c-id}(TAKES)] \div \pi_{c-id}[\sigma_{ssn=123}(TAKES)]$$

Faloutsos - Pavlo CMU SCS 15-415/615 #68



CMU SCS

Conclusions

- Relational model: only tables ('relations')
- relational algebra: powerful, minimal: 5 operators can handle almost any query!

Faloutsos - Pavlo CMU SCS 15-415/615 #69