Relational Model and Algebra

P.J. McBrien

Imperial College London

Example Relational Schema

branch		
sortcode	bname	cash
56	'Wimbledon'	94340.45
34	'Goodge St'	8900.67
67	'Strand'	34005.00

movement			
mid	no	amount	tdate
1000	100	2300.00	5/1/1999
1001	101	4000.00	5/1/1999
1002	100	-223.45	8/1/1999
1004	107	-100.00	11/1/1999
1005	103	145.50	12/1/1999
1006	100	10.23	15/1/1999
1007	107	345.56	15/1/1999
1008	101	1230.00	15/1/1999
1009	119	5600.00	18/1/1999

account				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

key branch(sortcode)
key branch(bname)
key movement(mid)
key account(no)

movement(no) $\stackrel{fk}{\Rightarrow}$ account(no)

account(sortcode) $\stackrel{fk}{\Rightarrow}$ branch(sortcode)

Relational Algebra: A Query Language for the Relational Model

- Five primitive operators
  - Unary operators π and σ
  - Binary operators $\times$, $\cup$ and $-$
- All operators produce one relation as their output
- Other (useful) operators may be defined in terms of the five primitve operators

Relational Algebra: Project

account				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

$\pi_{no,type}$ account

no	type
100	'current'
101	'deposit'
103	'current'
107	'current'
119	'deposit'
125	'current'

$\pi_{sortcode}$ account

sortcode
67
34
56

Relational Algebra: Select

account				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

$\sigma_{\text{rate}>0}$account				
no	type	cname	rate	sortcode
101	'deposit'	'McBrien, P.'	5.25	67
119	'deposit'	'Poulovassilis, A.'	5.50	56

Relational Algebra: Product

branch		
sortcode	bname	cash
56	'Wimbledon'	94340.45
34	'Goodge St'	8900.67
67	'Strand'	34005.00

$\sigma_{rate>0}$account				
no	type	cname	rate	sortcode
101	'deposit'	'McBrien, P.'	5.25	67
119	'deposit'	'Poulovassilis, A.'	5.50	56

branch $\times$ $\sigma_{rate>0}$account							
sortcode	bname	cash	no	type	cname	rate	sortcode
56	'Wimbledon'	94340.45	101	'deposit'	'McBrien, P.'	5.25	67
56	'Wimbledon'	94340.45	119	'deposit'	'Poulovassilis, A.'	5.50	56
34	'Goodge St'	8900.67	101	'deposit'	'McBrien, P.'	5.25	67
34	'Goodge St'	8900.67	119	'deposit'	'Poulovassilis, A.'	5.50	56
67	'Strand'	34005.00	101	'deposit'	'McBrien, P.'	5.25	67
67	'Strand'	34005.00	119	'deposit'	'Poulovassilis, A.'	5.50	56

Quiz 1: RA Queries

branch		
sortcode	bname	cash
56	'Wimbledon'	94340.45
34	'Goodge St'	8900.67
67	'Strand'	34005.00

account				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

Which RA query lists the name of branches that have deposit accounts?

A

$\pi_{\text{sortcode}}\, \sigma_{\text{type}=\text{'deposit'}}\, \text{account}$

B

π_{bname}

$\sigma_{\text{account.sortcode}=\text{branch.sortcode} \wedge \text{type}=\text{'deposit'}}$

$(\text{account} \times \text{branch})$

C

$\pi_{\text{bname}}(\text{branch} \times \sigma_{\text{type}=\text{'deposit'}}\, \text{account})$

D

$\pi_{\text{bname}}\, \sigma_{\text{type}=\text{'deposit'}}(\text{account} \times \text{branch})$

SPJ Queries

Select Project Join (SPJ) queries

If a product of tables is formed, where a selection is then done that compares the attributes of those tables, we say that a **join** has been performed.
Normally not all columns of the product are returned, and therefore a project is also required.

Branches with current accounts

$\pi_{\text{bname,no}}\, \sigma_{\text{branch.sortcode=account.sortcode} \wedge \text{account.type='current'}} (\text{branch} \times \text{account})$

bname	no
'Goodge St'	103
'Wimbledon'	107
'Wimbledon'	125
'Strand'	100

Relational Algebra: Union

$\pi_{\text{sortcode as id}}$account
id
67
34
56

$\pi_{\text{no as id}}$account
id
100
101
103
107
119
125

$\pi_{\text{sortcode as id}}$account $\cup$ $\pi_{\text{no as id}}$account
id
67
34
56
100
101
103
107
119
125

- relations must be **union compatible**

Relational Algebra: Difference

π_{no}account
no
100
101
103
107
119
125

π_{no}movement
no
100
101
103
107
119

π_{no}account − π_{no}movement
no
125

Rules for Combining Operators

Since all operators produce a relation as output, *any* operator may produce one of the inputs to any other operator.

well formed RA query

- the output of the nested operator must contain the attributes required by an outer π or σ
- the two inputs to a $\cup$ or $-$ must contain the same number of attributes

Derived Relational Algebra: Natural Join

branch ⋈ account						
sortcode	bname	cash	no	type	cname	rate
34	'Goodge St'	8900.67	103	'current'	'Boyd, M.'	NULL
56	'Wimbledon'	94340.45	107	'current'	'Poulovassilis, A.'	NULL
56	'Wimbledon'	94340.45	119	'deposit'	'Poulovassilis, A.'	5.50
56	'Wimbledon'	94340.45	125	'current'	'Bailey, J.'	NULL
67	'Strand'	34005.00	100	'current'	'McBrien, P.'	NULL
67	'Strand'	34005.00	101	'deposit'	'McBrien, P.'	5.25

$$R \bowtie S = \sigma_{R.A_1=S.A_1 \wedge \ldots \wedge R.A_m=S.A_m} R \times S$$

$$\text{branch} \bowtie \text{account} = \sigma_{\text{branch.sortcode}=\text{account.sortcode}} \text{branch} \times \text{account}$$

Quiz 2: Natural Join

What is the result of $\pi_{no}(account \bowtie movement)$?

A

$\pi_{no}(account \bowtie movement)$
no
100
101
103
107
119
125

B

$\pi_{no}(account \bowtie movement)$
no
100
101
103
107
119

C

$\pi_{no}(account \bowtie movement)$
no
125

D

$\pi_{no}(account \bowtie movement)$
no

Derived Relational Algebra: Semi Join

account ⋉ movement				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56

$$R \ltimes S = R \bowtie \pi_{Attr(R) \cap Attr(S)}(S)$$

Derived Relational Algebra: Semi Join

account ⋉ movement				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56

$$R \ltimes S = R \bowtie \pi_{Attr(R) \cap Attr(S)}(S)$$

$$\text{account} \ltimes \text{movement} = \text{account} \bowtie \pi_{\text{no}}(\text{movement})$$

Derived Relational Algebra: Joins

Natural Join

$$R \bowtie S = \sigma_{R.A_1=S.A_1 \wedge \ldots \wedge R.A_m=S.A_m} R \times S$$

Equi Join

$$R \overset{A=B}{\bowtie} S = \sigma_{R.A=S.B} R \times S$$

Semi Join

$$R \ltimes S = R \bowtie \pi_{Attr(R) \cap Attr(S)}(S)$$

Theta Join

$$R \overset{\theta}{\bowtie} S = \sigma_{\theta} R \times S$$

Quiz 3: Understanding join operators

branch		
sortcode	bname	cash
56	'Wimbledon'	94340.45
34	'Goodge St'	8900.67
67	'Strand'	34005.00

account				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

Which RA query produces the most tuples?

A

$$\text{branch} \overset{\text{branch.sortcode} < \text{account.sortcode}}{\bowtie} \text{account}$$

B

$\text{branch} \bowtie \text{account}$

C

$\text{branch} \ltimes \text{account}$

D

$$\text{branch} \overset{\text{branch.sortcode} = \text{account.sortcode}}{\bowtie} \text{account}$$

Quiz 4: Foreign Keys and Natural Joins (1)

Suppose R and S only share attribute A, and there is a foreign key $S(A) \overset{fk}{\Rightarrow} R(A)$.

If $|R| = 100$ and $|S| = 1,000$, what is $|R \bowtie S|$?

A

100

B

1,000

C

100,000

D

900

Note that $|R|$ returns the number of tuples in the current extent of R

Quiz 5: Foreign Keys and Natural Joins (2)

Suppose R and S only share attribute A, and there is a foreign key $R(A) \overset{fk}{\Rightarrow} S(A)$.

If $|R| = 100$ and $|S| = 1,000$, what is $|R \bowtie S|$?

A

100

B

1,000

C

100,000

D

900

Derived Relational Algebra: Intersection

Intersection

$$R \cap S = R - (R - S)$$

π_{no}account $\cap$ π_{no}movement

π_{no}account
no
100
101
103
107
119
125

π_{no}account $-$ π_{no}movement
no
125

π_{no}account $\cap$ π_{no}movement
no
100
101
103
107
119

Quiz 6: Intersection

email	
name	address
'McBrien, P.'	p.mcbrien@imperial.ac.uk
'Poulovassilis, A.'	ap@dcs.bbk.ac.uk
'Pietzuch, P.'	prp@doc.ic.ac.uk

What is the result of π_{cname} account $\cap$ π_{name} email?

A

cname
'McBrien, P.'
'Boyd, M.'
'Poulovassilis, A.'
'Bailey, J.'
'Pietzuch, P.'

B

cname
'McBrien, P.'
'Boyd, M.'
'Poulovassilis, A.'
'Bailey, J.'

C

cname
'McBrien, P.'
'Poulovassilis, A.'
'Pietzuch, P.'

D

cname
'McBrien, P.'
'Poulovassilis, A.'

Worksheet: Relational Algebra Operators

branch		
sortcode	bname	cash
56	'Wimbledon'	94340.45
34	'Goodge St'	8900.67
67	'Strand'	34005.00

movement			
mid	no	amount	tdate
1000	100	2300.00	5/1/1999
1001	101	4000.00	5/1/1999
1002	100	-223.45	8/1/1999
1004	107	-100.00	11/1/1999
1005	103	145.50	12/1/1999
1006	100	10.23	15/1/1999
1007	107	345.56	15/1/1999
1008	101	1230.00	15/1/1999
1009	119	5600.00	18/1/1999

account				
no	type	cname	rate	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

key branch(sortcode)
key branch(bname)
key movement(mid)
key account(no)

movement(no) $\stackrel{fk}{\Rightarrow}$ account(no)

account(sortcode) $\stackrel{fk}{\Rightarrow}$ branch(sortcode)

Equivalences Involving Project

Project and Project

$\pi_{\vec{X}}\, \pi_{\vec{Y}}\, \mathsf{R} \equiv \pi_{\vec{X}}\, \mathsf{R}$

You can eleminate any inner project (note that to be well formed $\vec{X} \subseteq \vec{Y}$)

Project and Select

$\pi_{\vec{X}}\, \sigma_{\mathsf{P}(\vec{Y})}\, \mathsf{R} \equiv \sigma_{\mathsf{P}(\vec{Y})}\, \pi_{\vec{X}}\, \mathsf{R}$

You can move a project of attributes $\vec{X}$ inside a select, provided the select predicate can be answered from those attributes, *i.e.* $\vec{Y} \subseteq \vec{X}$

Project and Product

$\pi_{\vec{X}}(\mathsf{R} \times \mathsf{S}) \equiv \pi_{\vec{X} \cap \mathsf{Atts}(\mathsf{R})}\, \mathsf{R} \times \pi_{\vec{X} \cap \mathsf{Atts}(\mathsf{S})}\, \mathsf{S}$

Project and Union

$\pi_{\vec{X}}(\mathsf{R} \cup \mathsf{S}) \equiv \pi_{\vec{X}}\, \mathsf{R} \cup \pi_{\vec{X}}\, \mathsf{S}$

Project and Difference

$\pi_{\vec{X}}(\mathsf{R} - \mathsf{S}) \equiv \pi_{\vec{X}}\, \mathsf{R} - \pi_{\vec{X}}\, \mathsf{S}$

Equivalences Involving Select

Select and Project

$$\sigma_{P(\vec{X})}\, \pi_{\vec{X}}\, R \equiv \pi_{\vec{X}}\, \sigma_{P(\vec{X})}\, R$$

Select and Select

$$\sigma_{P_x(\vec{X})}\, \sigma_{P_y(\vec{Y})}\, R \equiv \sigma_{P_x(\vec{X}) \wedge P_y(\vec{Y})}\, R$$

Select and Product

$$\sigma_{P(\vec{X})}(R \times S) \equiv \sigma_{P(\vec{X})}\, R \times S \iff \vec{X} - \mathsf{Atts}(S) = \vec{X}$$

You can move a select predicate $P(\vec{X})$ onto one of the relations inside a product provided all the attributes of the select belong to that relation.

Select and Union

$$\sigma_{P(\vec{X})}(R \cup S) \equiv \sigma_{P(\vec{X})}\, R \cup \sigma_{P(\vec{X})}\, S$$

Select and Difference

$$\sigma_{P(\vec{X})}(R - S) \equiv \sigma_{P(\vec{X})}\, R - S$$

Quiz 7: Equivalent RA Expressions

Which RA expression is not equivalent to the other three?

A

$\pi_{\text{no}}\ \sigma_{\text{type}=\text{'current'}}\ \text{account}$

B

$\pi_{\text{no}}\ \sigma_{\text{type}=\text{'current'}}\ \pi_{\text{no,type,cname}}\ \text{account}$

C

$\pi_{\text{no}}\ \sigma_{\text{type}<>\text{'deposit'}}\ \pi_{\text{no,type,cname}}\ \text{account}$

D

$\pi_{\text{no}}\ \sigma_{\text{type}=\text{'current'}}\ \sigma_{\text{type}<>\text{'deposit'}}\ \text{account}$

Equivalences Involving Binary Operators

Product and Union

$R \times (S \cup T) \equiv (R \times S) \cup (R \times T)$

Product and Difference

$R \times (S - T) \equiv (R \times S) - (R \times T)$

Union and Product

$R \cup (S \times T)$ unable to move $\cup$ inside $\times$

Union and Difference

$R \cup (S - T)$ unable to move $\cup$ inside $-$

Difference and Product

$R - (S \times T)$ unable to move $-$ inside $\times$

Difference and Union

$R - (S \cup T) \equiv (R - S) - T$

Worksheet: Equivalences Between RA Expressions

1. $(R \times S) \times T \equiv R \times (S \times T)$
   $(R - S) - T \equiv R - (S - T)$
   $(R \cup S) \cup T \equiv R \cup (S \cup T)$
   $(R \cap S) \cap T \equiv R \cap (S \cap T)$
2. $\pi_{\text{no,type}}\, \sigma_{\text{sortcode}=56}\, \pi_{\text{no,type,sortcode}}\, \sigma_{\text{type}=\text{'deposit'}}\, \text{account}$
3. $\sigma_{\text{account.no}=\text{movement.no}}(\pi_{\text{no,cname}}\, \text{account} \times \pi_{\text{mid,no}}\, \sigma_{\text{amount}>1000}\, \text{movement})$
4. $\sigma_{\text{account.no}=\text{movement.no}}(\pi_{\text{no,cname,rate}}\, \text{account} \times (\sigma_{\text{amount}>1000}\, \pi_{\text{mid,no}}\, \text{movement} \cup \sigma_{\text{amount}<100}\, \pi_{\text{mid,no}}\, \text{movement}))$
5. $\pi_{\text{no,cname,tdate}}\, \sigma_{\text{amount}<0 \wedge \text{account.no}=\text{movement.no}}\, \text{branch} \times \text{movement}$