CSE 4125: Distributed Database Systems
Chapter – 6

Optimization of Access Strategies.

(part – C)

Outline

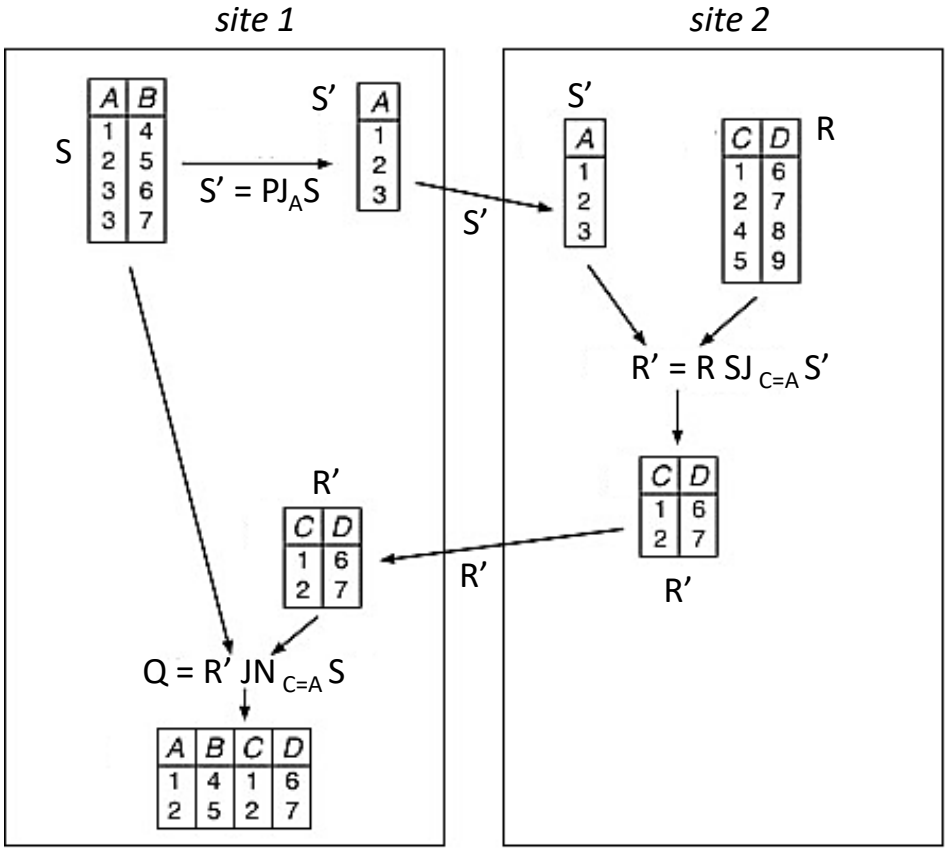
- Semi-join programs.
- Full reducer.

Semi-join Programs

$$\boldsymbol{R} \text{ JN}_{C=A} \boldsymbol{S} \leftrightarrow (\boldsymbol{R} \text{ SJ}_{C=A} \text{PJ}_{A} \boldsymbol{S}) \text{ JN}_{C=A} \boldsymbol{S}$$

Semi-join Programs (contd.)

$$\boldsymbol{R}\ \mathrm{JN}_{C=A}\ \boldsymbol{S} \leftrightarrow (\boldsymbol{R}\ \mathrm{SJ}_{C=A}\ \mathrm{PJ}_{A}\ \boldsymbol{S})\ \mathrm{JN}_{C=A}\ \boldsymbol{S}$$

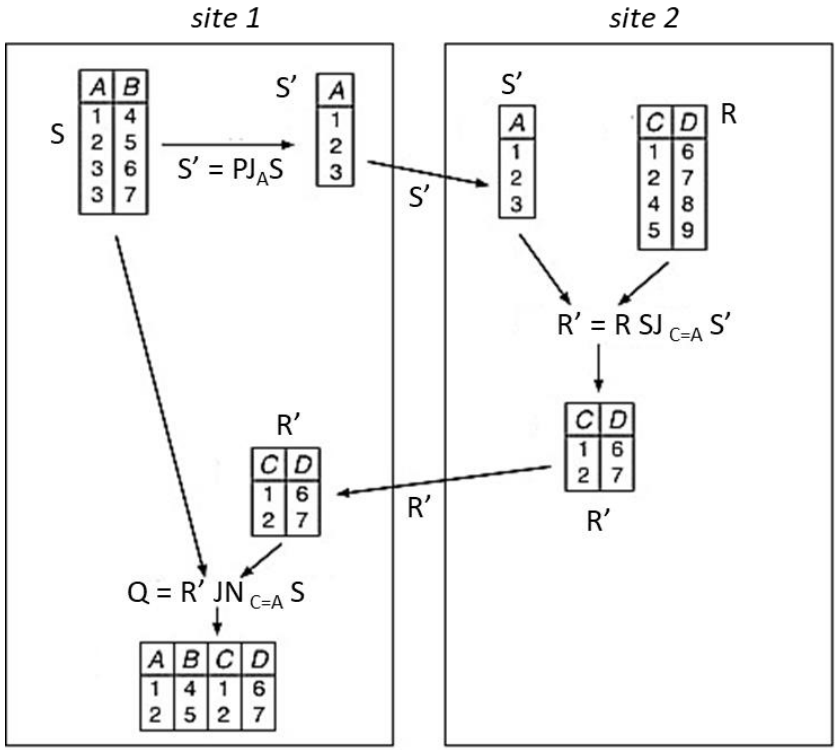


Semi-join Programs (contd.)

Cost of Semi-join program

- Cost of sending S' to site – 2:

  TC_1 = ?

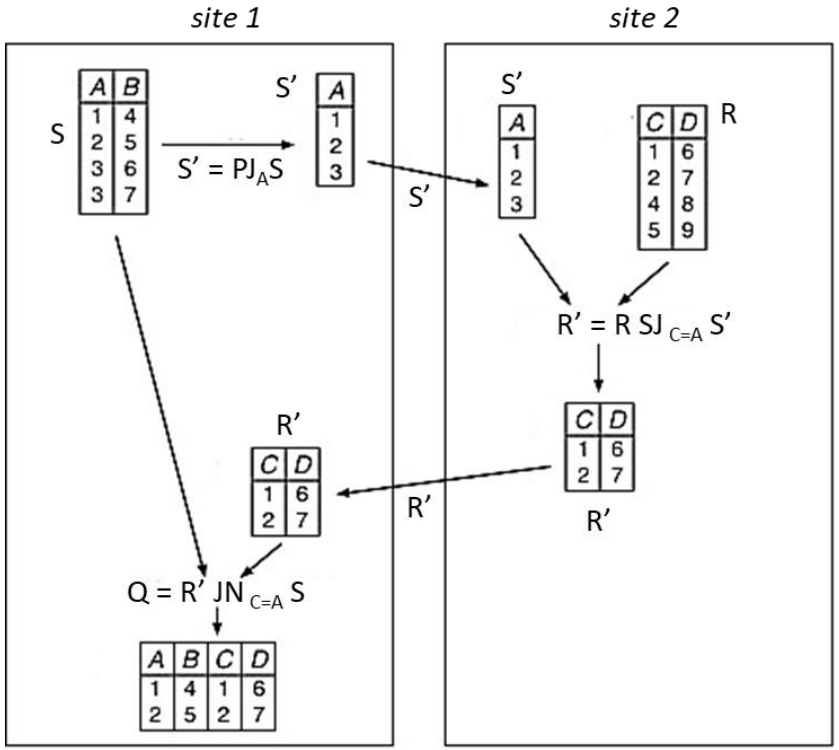


Semi-join Programs (contd.)

Cost of Semi-join program

- Cost of sending S' to site – 2:

  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$



Semi-join Programs (contd.)

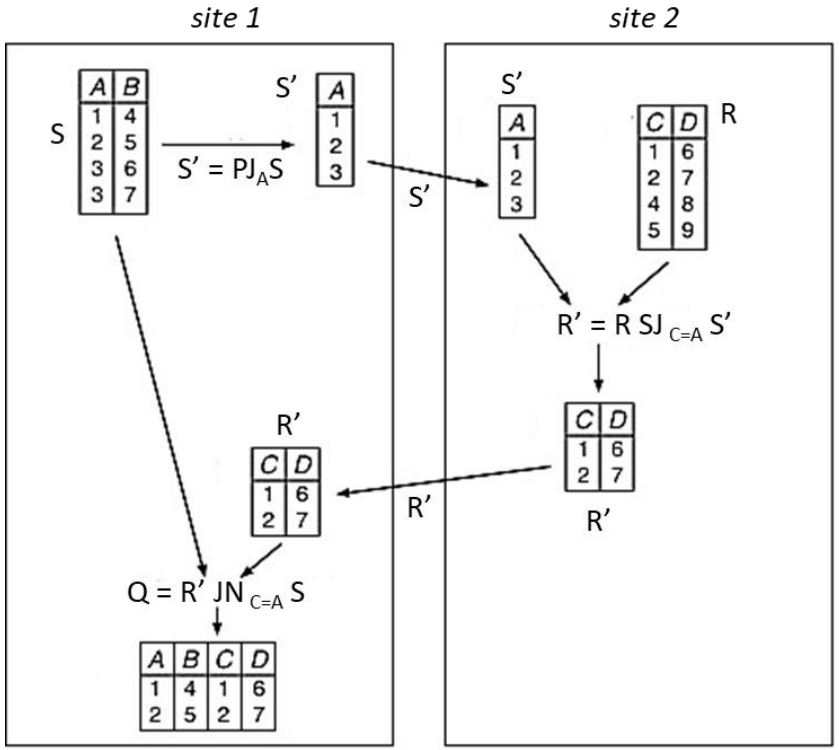
Cost of Semi-join program

- Cost of sending S' to site – 2:

  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$

- Cost of computing R' at site – 2:

  $TC_2 = ?$



Semi-join Programs (contd.)

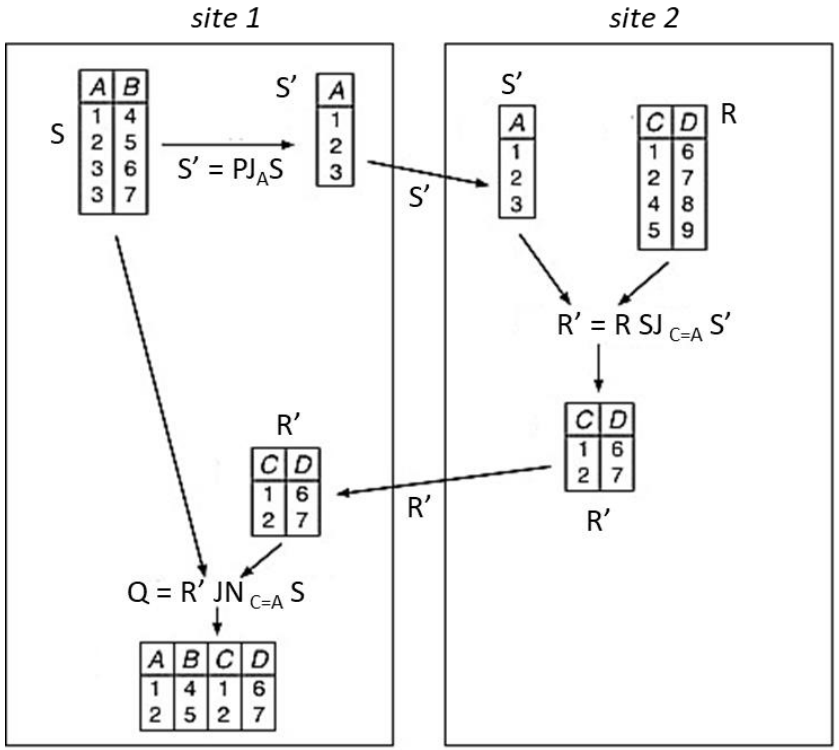
Cost of Semi-join program

- Cost of sending S' to site – 2:

  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$

- Cost of computing R' at site – 2:

  $TC_2 = 0$



Semi-join Programs (contd.)

Cost of Semi-join program

- Cost of sending S' to site – 2:

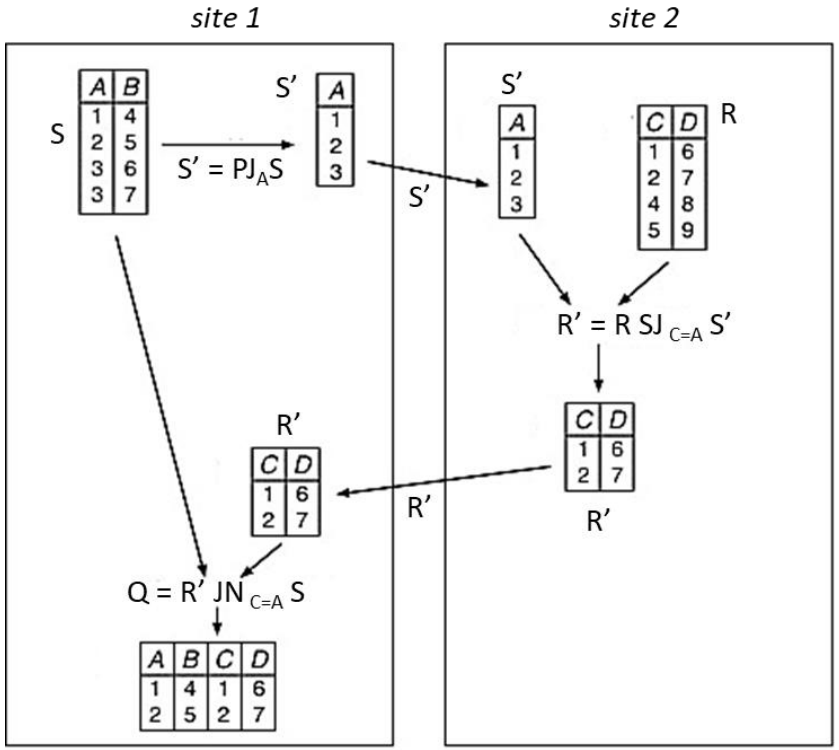
  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$

- Cost of computing R' at site – 2:

  $TC_2 = 0$

- Cost of sending R' to site – 1:

  $TC_3 = ?$



Semi-join Programs (contd.)

Cost of Semi-join program

- Cost of sending S' to site – 2:

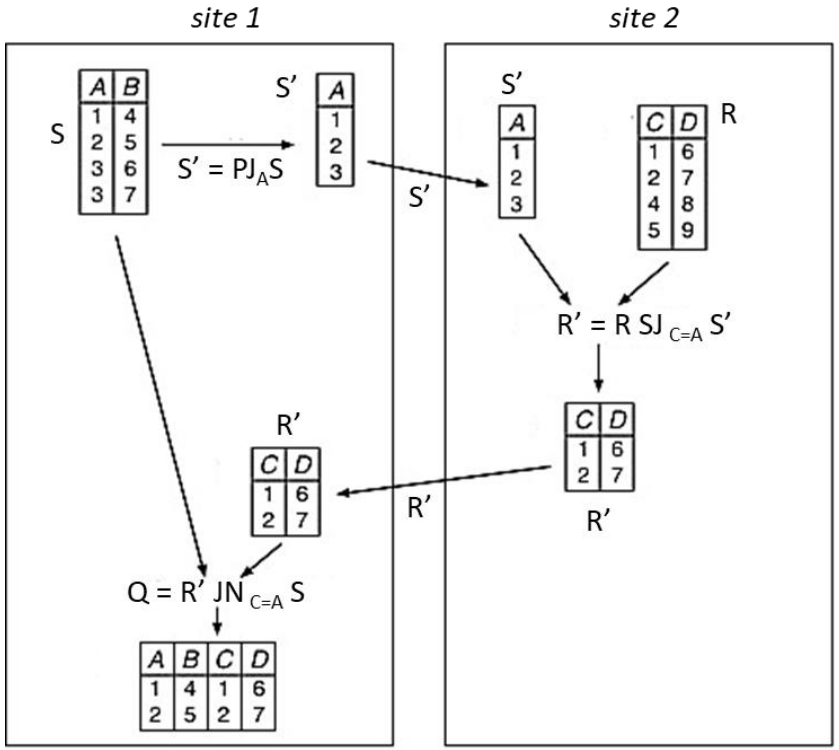
  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$

- Cost of computing R' at site – 2:

  $TC_2 = 0$

- Cost of sending R' to site – 1:

  $TC_3 = C_0 + C_1 * size(R) * card(R')$



Semi-join Programs (contd.)

Cost of Semi-join program

- Cost of sending S' to site – 2:

  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$

- Cost of computing R' at site – 2:

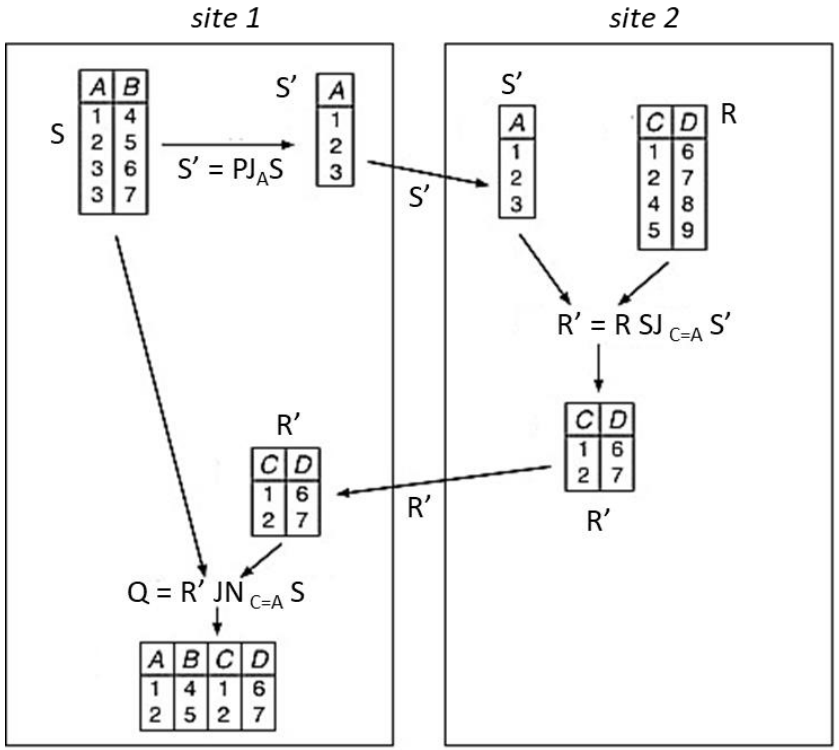
  $TC_2 = 0$

- Cost of sending R' to site – 1:

  $TC_3 = C_0 + C_1 * size(R) * card(R')$

- Cost of computing Q at site – 1:

  $TC_4 = ?$



Semi-join Programs (contd.)

Cost of Semi-join program

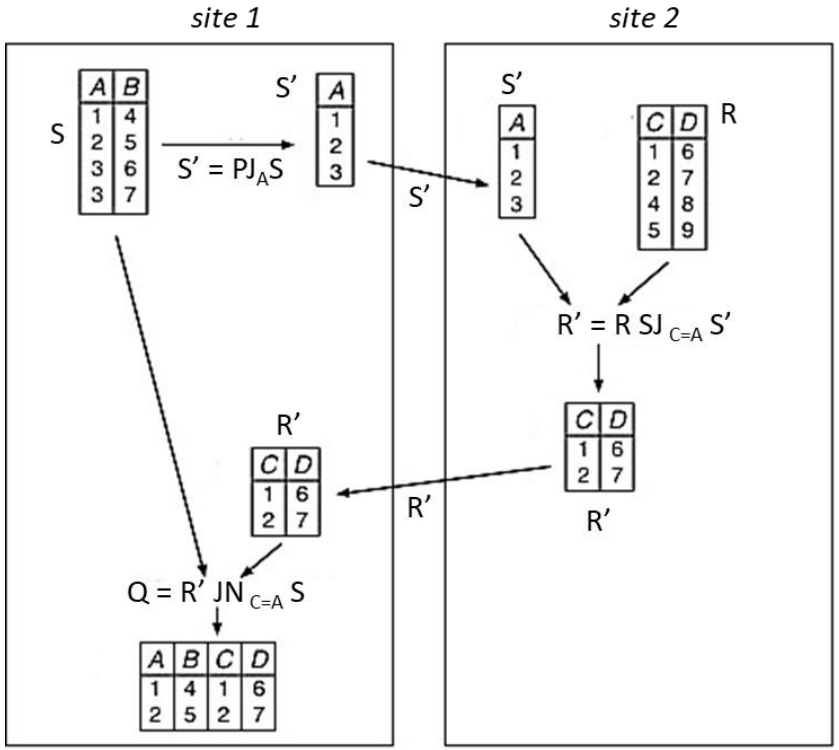
- Cost of sending S' to site – 2:

  $TC_1 = C_0 + C_1 * size(A) * val(A[S])$
- Cost of computing R' at site – 2:

  $TC_2 = 0$
- Cost of sending R' to site – 1:

  $TC_3 = C_0 + C_1 * size(R) * card(R')$
- Cost of computing Q at site – 1:

  $TC_4 = 0$



Semi-join Programs (contd.)

Cost of Semi-join program

- Total cost $TC_{SJ} = TC_1 + TC_2 + TC_3 + TC_4$
- If $TC_{SJ} < TC_{JN}$ then semi-join program is profitable.
  - Here TC_{JN} is the cost of performing join without semi-join program.

Other Applications of Semi-join Programs

- Semi-join programs can be used as fragment reducers (operations that can reduce cardinality of a relation).
  - Similarly to unary operations.
- Full reducer:
  - Chain of semi-joins.

Full Reducer

R

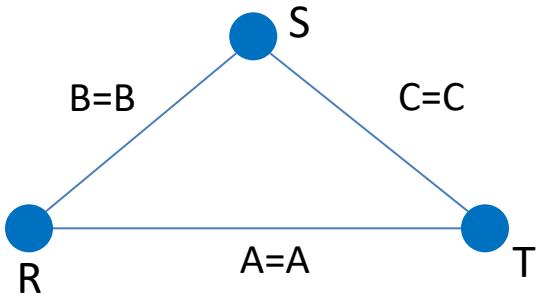
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



Full Reducer (contd.)

R

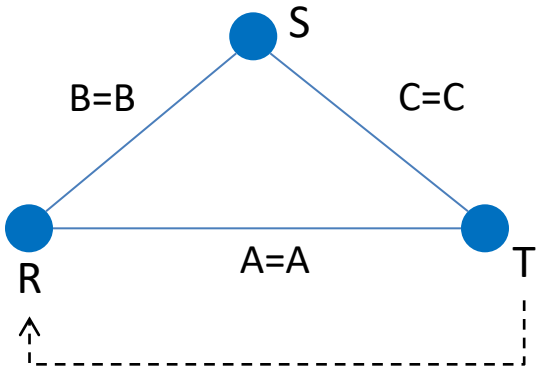
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



$R' = R \; SJ_{A=A} \; T$

A	B
2	b
3	c

Full Reducer (contd.)

R

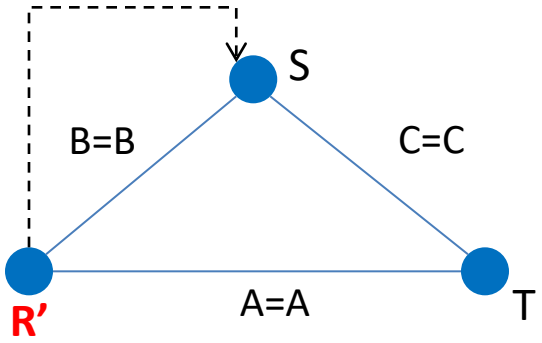
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



R' = R SJ $_{A=A}$ T

A	B
2	b
3	c

S' = S SJ $_{B=B}$ **R'**

B	C
b	y
c	z

Full Reducer (contd.)

R

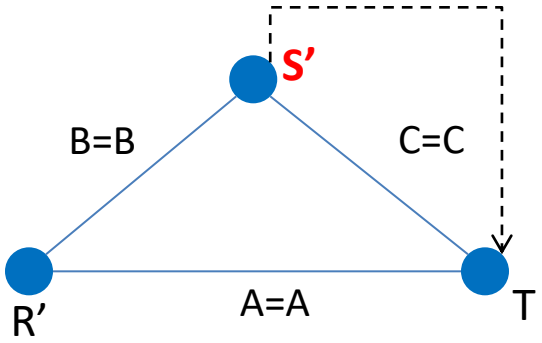
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



$R' = R \text{ SJ}_{A=A} T$

A	B
2	b
3	c

$S' = S \text{ SJ}_{B=B} R'$

B	C
b	y
c	z

$T' = T \text{ SJ}_{C=C} S'$

C	A
y	3
z	4

Full Reducer (contd.)

R

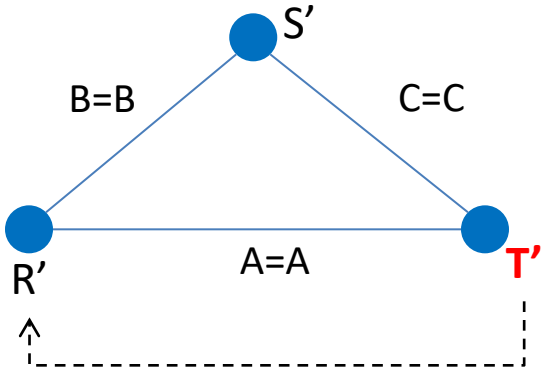
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



$R' = R \; SJ_{A=A} \; T$

A	B
2	b
3	c

$S' = S \; SJ_{B=B} \; R'$

B	C
b	y
c	z

$T' = T \; SJ_{C=C} \; S'$

C	A
y	3
z	4

$R'' = R' \; SJ_{A=A} \; T'$

A	B
3	c

Full Reducer (contd.)

R

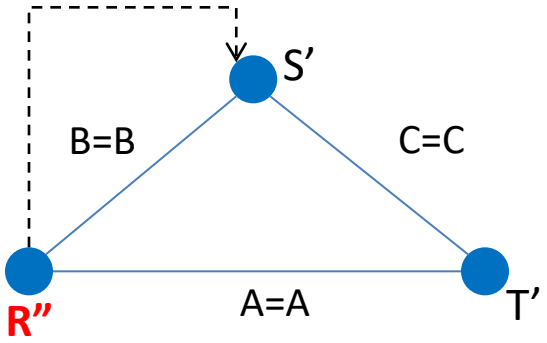
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



$R' = R \; SJ_{A=A} \; T$

A	B
2	b
3	c

$S' = S \; SJ_{B=B} \; R'$

B	C
b	y
c	z

$T' = T \; SJ_{C=C} \; S'$

C	A
y	3
z	4

$R'' = R' \; SJ_{A=A} \; T'$

A	B
3	c

$S'' = S' \; SJ_{B=B} \; R''$

B	C
c	z

Full Reducer (contd.)

R

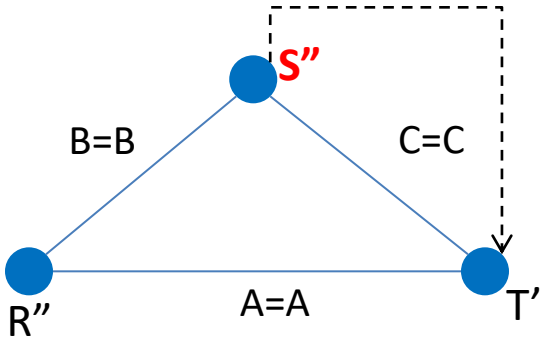
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



R’ = R SJ $_{A=A}$ T

A	B
2	b
3	c

S’ = S SJ $_{B=B}$ R’

B	C
b	y
c	z

T’ = T SJ $_{C=C}$ S’

C	A
y	3
z	4

R” = R’ SJ $_{A=A}$ T’

A	B
3	c

S” = S’ SJ $_{B=B}$ R”

B	C
c	z

T” = T’ SJ $_{C=C}$ **S”**

C	A
z	4

Full Reducer (contd.)

R

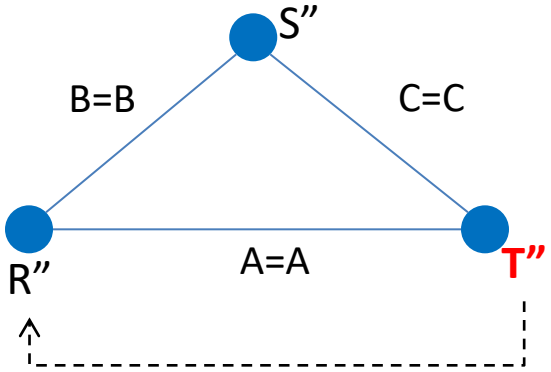
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



R’ = R SJ $_{A=A}$ T

A	B
2	b
3	c

S’ = S SJ $_{B=B}$ R’

B	C
b	y
c	z

T’ = T SJ $_{C=C}$ S’

C	A
y	3
z	4

R” = R’ SJ $_{A=A}$ T’

A	B
3	c

S” = S’ SJ $_{B=B}$ R”

B	C
c	z

T” = T’ SJ $_{C=C}$ S”

C	A
z	4

R”’ = R” SJ $_{A=A}$ **T”**

Φ

Full Reducer (contd.)

R

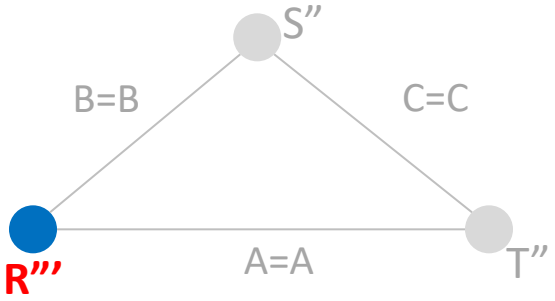
A	B
1	a
2	b
3	c

S

B	C
a	x
b	y
c	z

T

C	A
x	2
y	3
z	4



R’ = R SJ $_{A=A}$ T

A	B
2	b
3	c

S’ = S SJ $_{B=B}$ R’

B	C
b	y
c	z

T’ = T SJ $_{C=C}$ S’

C	A
y	3
z	4

R” = R’ SJ $_{A=A}$ T’

A	B
3	c

S” = S’ SJ $_{B=B}$ R”

B	C
c	z

T” = T’ SJ $_{C=C}$ S”

C	A
z	4

R”’ = R” SJ $_{A=A}$ T”

Φ

Additional Reading

- Length of full reducer.
- Tree queries.
- “Best” reducer.

Practice Problems/ Questions

1. With an example, prove that the semi-join is not symmetric. *[hint: page. 142]*
2. Textbook exercise: 6.1, 6.4