

Join minimization examples

Let R be a relation over attributes ABC .

(i) Simplify the following conjunctive SQL query, knowing that it is applied only to relations R satisfying the set of FDs $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$ (use pattern minimization and the chase):

```

select  $t_1.A, t_2.B, t_4.C$ 
from  $R\ t_1, R\ t_2, R\ t_3, R\ t_4$ 
where  $t_3.A = t_4.A$  and  $t_2.B = t_3.B$  and  $t_1.C = t_2.C$  and  $t_3.B = t_4.C$ 

```

(ii) Redo (i) for the query:

```

select  $t_1.A, t_2.B, t_4.C$ 
from  $R\ t_1, R\ t_2, R\ t_3, R\ t_4$ 
where  $t_2.C = 5$  and  $t_3.A = t_4.A$  and  $t_2.B = t_3.B$  and  $t_1.C = t_2.C$  and  $t_4.A = 8$ 

```

Solution

(i) The pattern P corresponding to the query is

R	A	B	C		A	B	C
	a	—	c_1	answer	a	b	b
	—	b	c_1				
	a_1	b	—				
	a_1	—	b				

We now chase the pattern with $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$. The steps are as follows:

1. we first consider $AC \rightarrow B$. However, there is no violation of this FD at this point.
2. we consider $B \rightarrow C$, which is violated by rows (2) and (3). Chasing leads us to identify the — in (3) with c_1 , yielding:

R	A	B	C		answer	A	B	C
	a	$-$	c_1			a	b	b
	$-$	b	c_1					
	a_1	b	c_1					
	a_1	$-$	b					

3. we consider $C \rightarrow A$, which is violated by rows (1),(2),(3). Chasing leads us to identify a , the $-$ in (2), and a_1 , which all become equal to a . Note that this includes the a_1 in row (4), which is the same a_1 as in row (3). This yields:

R	A	B	C		answer	A	B	C
	a	$-$	c_1			a	b	b
	a	b	c_1					
	a	b	c_1					
	a	$-$	b					

4. we are not yet done, because now $AC \rightarrow B$ is violated by rows (1) and (2). Chasing turns the $-$ in (1) into b , yielding:

R	A	B	C		answer	A	B	C
	a	b	c_1			a	b	b
	a	b	c_1					
	a	b	c_1					
	a	$-$	b					

The above pattern satisfies F , so the chase is done.

Eliminating duplicate rows from $CHASE_F(P)$ yields the following pattern, which is minimal:

R	A	B	C		answer	A	B	C
	a	b	c_1			a	b	b
	a	$-$	b					

and the minimal SQL query is:

select $t_1.A, t_1.B, t_2.C$
from $R\ t_1, R\ t_2$
where $t_1.A = t_2.A$ and $t_1.B = t_2.C$

(ii) The pattern P corresponding to the query is

R	A	B	C				
	a	$-$	5				
	$-$	b	5				
	8	b	$-$				
	8	$-$	c				

answer	A	B	C
	a	b	c

Chasing with respect to $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$ yields (after eliminating duplicate rows):

$CHASE_F(P) =$	R	A	B	C				
		8	b	5				
		8	$-$	c				

answer	A	B	C
	8	b	c

This pattern is minimal, and a corresponding SQL query with minimum number of joins is

select $t_1.A, t_1.B, t_2.C$
from $R\ t_1, R\ t_2$
where $t_1.A = 8$ and $t_2.A = 8$ and $t_1.C = 5$