

CS5226 Lecture 3

Query Tuning II

Query Tuning II

- ▶ Reordering groupby & select
- ▶ Reordering groupby & join

Reordering group by & select

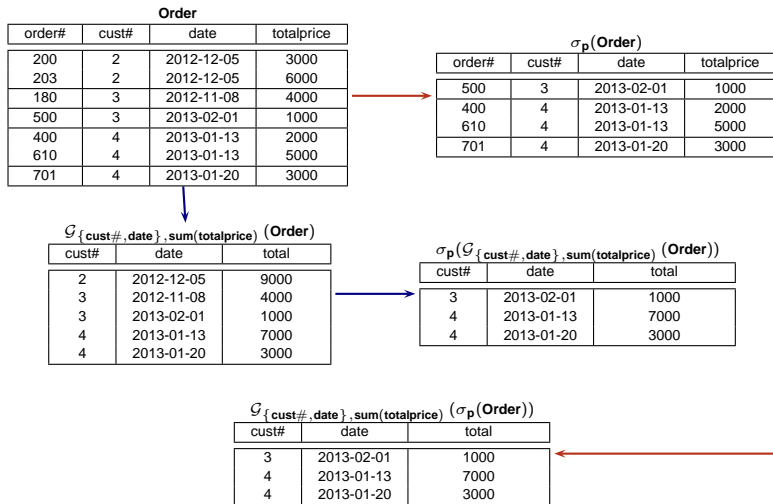
- ▶ Customer (cust#, cname, country)
- ▶ Order (order#, cust#, date, totalprice)

```
select cust#, date, total
from    (select cust#, date, sum(totalprice) as total
          from    Order
          group by cust#, date)
where   date >= '2013-01-01'
```

```
select cust#, date, sum(totalprice) as total
from    Order
where   date >= '2013-01-01'
group by cust#, date
```

Reordering group by & select

- Let p denote the predicate “ $\text{date} \geq 2013-01-01$ ”



Reordering group by & select

$$\begin{aligned} \sigma_p(\mathcal{G}_{A,F}(R)) &= \mathcal{G}_{A,F}(\sigma_p(R)) \\ &\text{iff} \\ A &\rightarrow \text{columns}(p) \end{aligned}$$

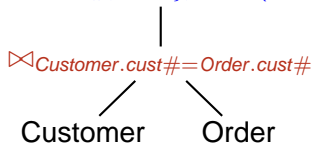
Pushing group by below join

```
select    c.cust#, cname, sum(o.totalprice) as T
from      Customer c join Order o
           on c.cust# = o.cust#
group by c.cust#, cname
```

```
select c.cust#, cname, T
from   customer c,
        (select    cust#, sum(totalprice) as T
         from      Order
         group by cust#
        ) as o
where  c.cust# = o.cust#
```

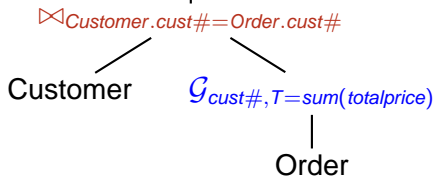
Example 1

$\mathcal{G}_{\{Order.cust\#,cname\},T=sum(totalprice)}$



(a) Group by after join

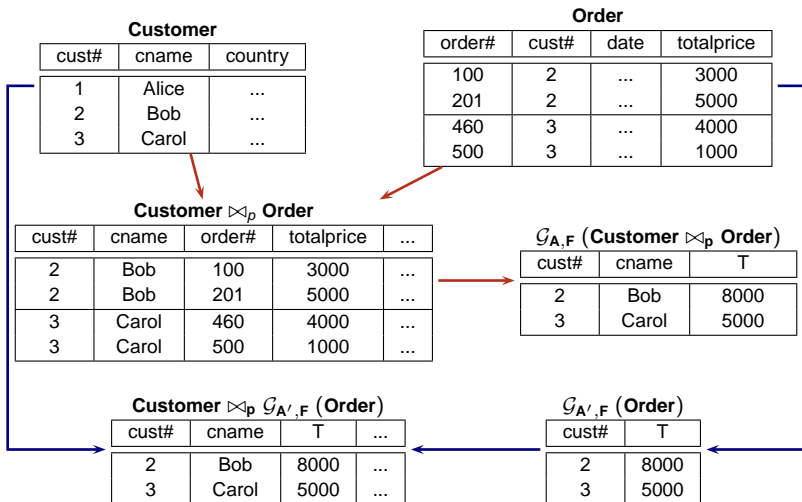
$\pi_{cust\#,cname,T}$



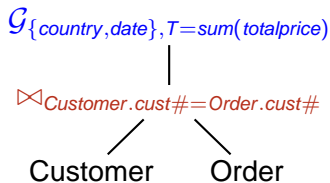
(b) Group by before join

Example 1

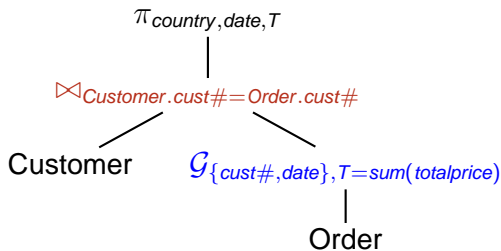
$$A = \{cust\#, cname\}, A' = \{cust\# \}, columns(p) = \{cust\#\}$$



Example 2



(a) Group by after join



(b) Group by before join

Example 2

$A = \{country, date\}$, $A' = \{cust\#, date\}$, $columns(p) = \{cust\#\}$

Customer

cust#	cname	country
1	Alice	UK
2	Bob	US
3	Carol	US

Order

order#	cust#	date	totalprice
100	2	2012-12-10	3000
201	2	2013-01-04	5000
460	3	2012-12-10	4000
500	3	2013-01-04	1000

Customer \bowtie_p Order

cust#	country	date	totalprice	...
2	US	2012-12-10	3000	...
3	US	2012-12-10	4000	...
2	US	2013-01-04	5000	...
3	US	2013-01-04	1000	...

$\mathcal{G}_{A,F}(\text{Customer} \bowtie_p \text{Order})$

country	date	T
US	2012-12-10	7000
US	2013-01-04	6000

Customer $\bowtie_p \mathcal{G}_{A',F}(\text{Order})$

country	date	T	...
US	2012-12-10	3000	...
US	2013-01-04	5000	...
US	2012-12-10	4000	...
US	2013-01-04	1000	...

$\mathcal{G}_{A',F}(\text{Order})$

cust#	date	T
2	2012-12-10	3000
2	2013-01-04	5000
3	2012-12-10	4000
3	2013-01-04	1000

Pushing group by below join

$\mathcal{G}_{A,F} (S \bowtie_p R)$ can be rewritten as

$$\pi_{A,F}(S \bowtie_p (\mathcal{G}_{A',F} R))$$

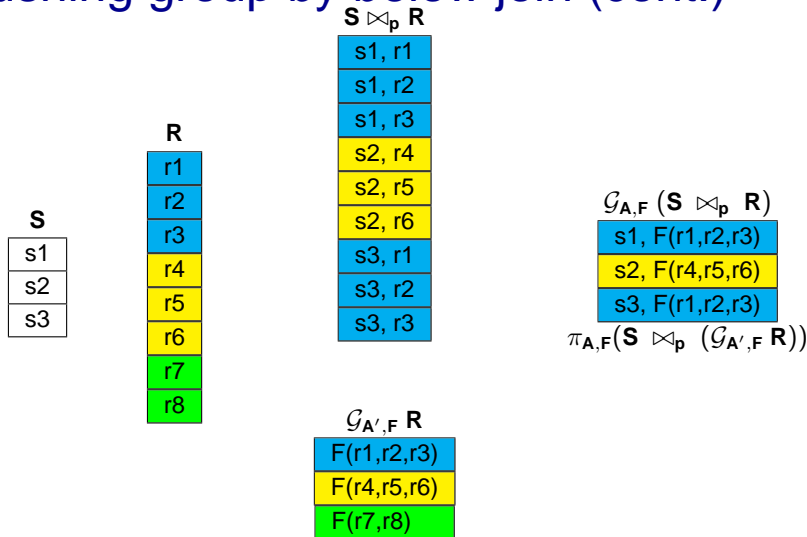
where

$$A' = A \cup \text{columns}(p) - \text{columns}(S)$$

iff the following conditions hold:

1. $\text{columns}(p) \cap \text{columns}(R) \subseteq A$,
2. $\text{key}(S) \subseteq A$, and
3. $\text{columns}(F) \subseteq \text{columns}(R)$

Pushing group by below join (cont.)



Tuples with same color have same $columns(R) \cap columns(p)$ values

Revisiting Example 1

- ▶ $\mathcal{G}_{A, T=\text{sum}(\text{totalprice})} (\text{Customer} \bowtie_{\text{cust}\#} \text{Order})$
where $A = \{\text{cust}\#, \text{cname}\}$
- ▶ $\pi_{\text{cust}\#, \text{cname}, T} (\text{Customer} \bowtie_{\text{cust}\#} (\mathcal{G}_{A', T=\text{sum}(\text{totalprice})} \text{Order}))$
where $A' = \{\text{cust}\#\}$
- ▶ $S = \text{Customer}$
- ▶ $\text{key}(S) = \text{cust}\#$
- ▶ $R = \text{Order}$
- ▶ $\text{columns}(p) = \{\text{Customer.cust}\#, \text{Order.cust}\#\}$
- ▶ $\text{columns}(F) = \{\text{Order.totalprice}\}$

Revisiting Example 2

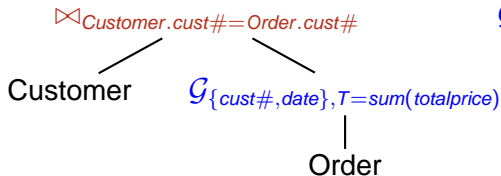
- ▶ $\mathcal{G}_{A, T=\text{sum}(\text{totalprice})} (\text{Customer} \bowtie_{\text{cust}\#} \text{Order})$
where $A = \{\text{country}, \text{date}\}$
- ▶ $\pi_{\text{country}, \text{date}, T} (\text{Customer} \bowtie_{\text{cust}\#} (\mathcal{G}_{A', T=\text{sum}(\text{totalprice})} \text{Order}))$
where $A' = \{\text{date}, \text{cust}\#\}$
- ▶ $S = \text{Customer}$
- ▶ $\text{key}(S) = \text{cust}\#$
- ▶ $R = \text{Order}$
- ▶ $\text{columns}(p) = \{\text{Customer.cust}\#, \text{Order.cust}\#\}$
- ▶ $\text{columns}(F) = \{\text{Order.totalprice}\}$

Pulling group by above join

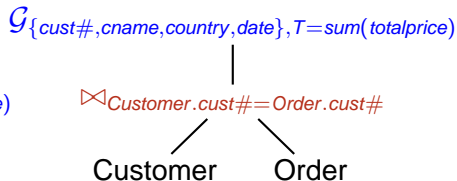
```
select c.cust#, cname, country, date, T
from    customer c,
        ( select    cust#, date, sum(totalprice) as T
          from      Order
          group by cust#, date
        ) as o
where   c.cust# = o.cust#
```

```
select    c.cust#, cname, country, date,
          sum(o.totalprice) as T
from      Customer c join Order o
          on c.cust# = o.cust#
group by c.cust#, cname, country, date
```

Example 3



(a) Group by before join



(b) Group by after join

Example 3

$$A = \{cust\#, date\}, A' = \{cust\#, cname, country, date\}$$

Customer

cust#	cname	country
1	Alice	UK
2	Bob	US
3	Carol	US

Order

order#	cust#	date	totalprice
100	2	2012-12-10	3000
201	2	2013-01-04	5000
460	3	2012-12-10	4000
500	3	2013-01-04	1000

$$\mathcal{G}_{A',F}(\text{Customer} \bowtie_p \text{Order})$$

cust#	cname	country	date	totalprice
2	Bob	US	2012-12-10	3000
2	Bob	US	2013-01-04	5000
3	Carol	US	2012-12-10	4000
3	Carol	US	2013-01-04	1000

$$\text{Customer} \bowtie_p \mathcal{G}_{A,F}(\text{Order})$$

$$\mathcal{G}_{A,F}(\text{Order})$$

cust#	date	T
2	2012-12-10	3000
2	2013-01-04	5000
3	2012-12-10	4000
3	2013-01-04	1000

Pulling group by above join

Pulling group by above join

$S \bowtie_p (\mathcal{G}_{A,F} R)$ can be rewritten as

$$\mathcal{G}_{A',F} (S \bowtie_p R)$$

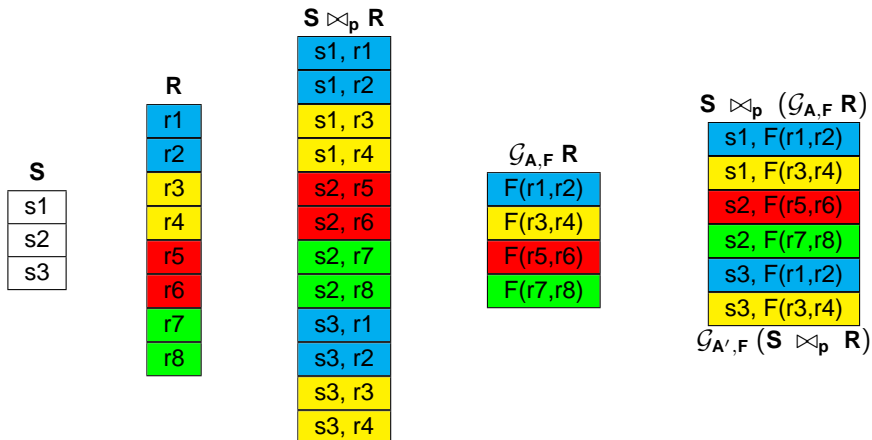
where

$$A' = A \cup \text{columns}(S)$$

iff the following conditions hold:

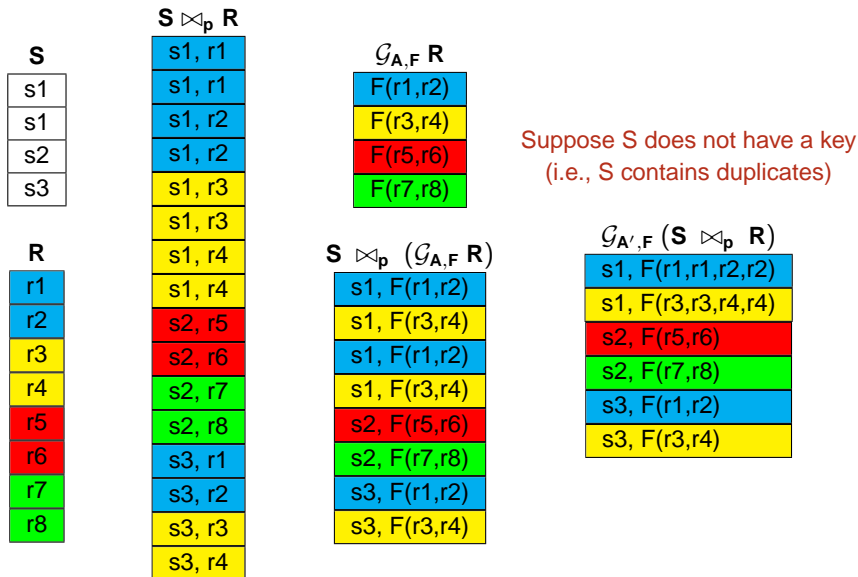
1. S has a key, and
2. p does not refer to aggregated columns in F ;
i.e., $\text{columns}(p) \cap \text{columns}(R) \subseteq A$

Pulling group by above join

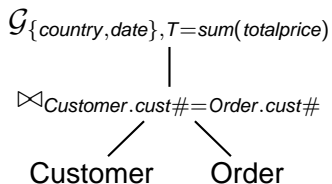


Tuples with same color have same A values

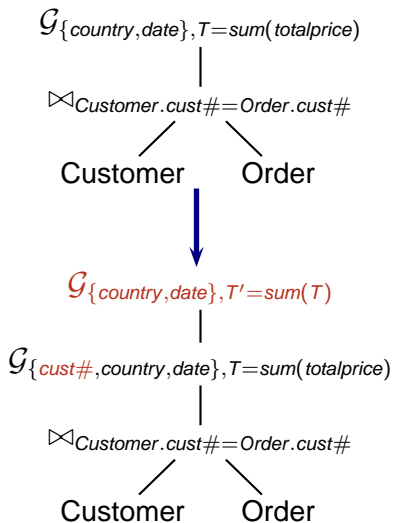
Pulling group by above join (cont.)



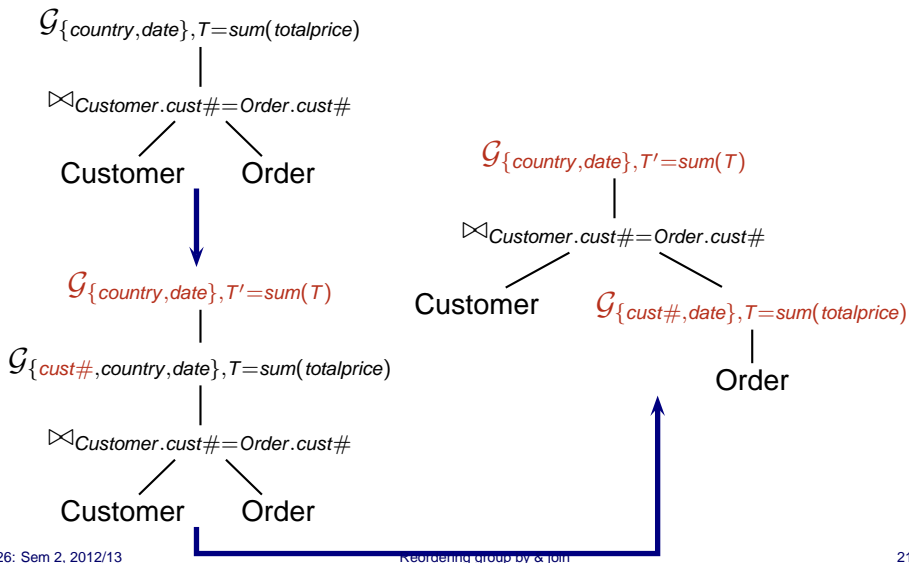
Revisiting Example 2



Revisiting Example 2



Revisiting Example 2



Revisiting Example 2

$$A = \{country, date\}, F = \{T = \text{sum}(\text{totalprice})\},$$

$$A' = \{cust\#, date\}, F' = \{T' = \text{sum}(T)\}$$

Customer

cust#	cname	country
1	Alice	UK
2	Bob	US
3	Carol	US

Order

order#	cust#	date	totalprice
100	2	2012-12-10	3000
201	2	2013-01-04	5000
460	3	2012-12-10	4000
500	3	2013-01-04	1000

Customer \bowtie_p Order

cust#	country	date	totalprice	...
2	US	2012-12-10	3000	...
3	US	2012-12-10	4000	...
2	US	2013-01-04	5000	...
3	US	2013-01-04	1000	...

$\mathcal{G}_{A,F}(\text{Customer} \bowtie_p \text{Order})$

country	date	T
US	2012-12-10	7000
US	2013-01-04	6000

$\mathcal{G}_{A,F'}(X)$

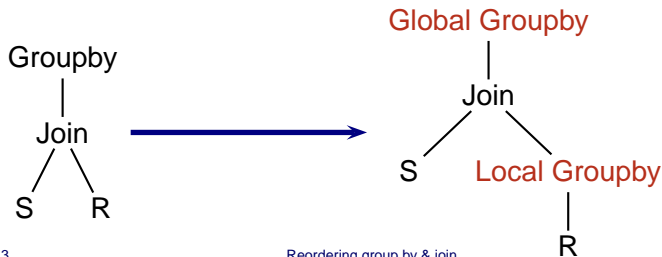
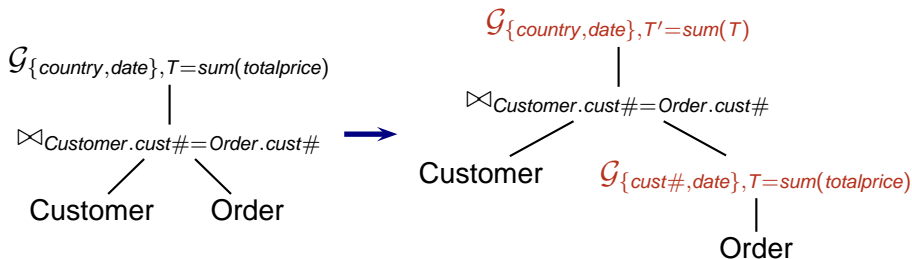
$X = \text{Customer} \bowtie_p \mathcal{G}_{A',F'}(\text{Order})$

country	date	T	...
US	2012-12-10	3000	...
US	2013-01-04	5000	...
US	2012-12-10	4000	...
US	2013-01-04	1000	...

$\mathcal{G}_{A',F'}(\text{Order})$

cust#	date	T
2	2012-12-10	3000
2	2013-01-04	5000
3	2012-12-10	4000
3	2013-01-04	1000

Local/Global group by



Local/Global group by (cont.)

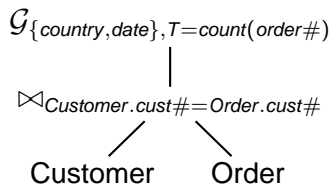
- ▶ Consider an aggregation function f on a relation R
- ▶ Let R be partitioned into n groups
 - ▶ $R = R_1 \cup \dots \cup R_n$,
 - ▶ $R_i \cap R_j = \emptyset, \forall i, j \in [1, n], i \neq j$
- ▶ f_ℓ = a local aggregate function
- ▶ f_g = a global aggregate function

$$\begin{aligned} f(R) &= f(R_1 \cup \dots \cup R_n) \\ &= f_g(f_\ell(R_1) \cup \dots \cup f_\ell(R_n)) \end{aligned}$$

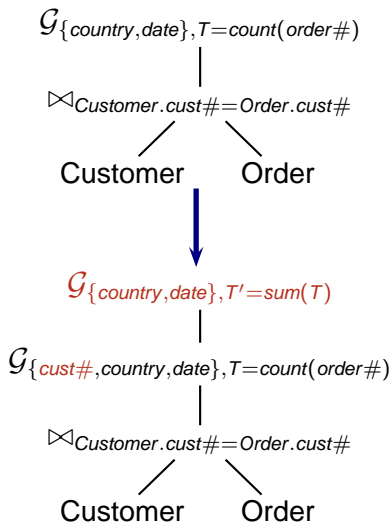
Local/Global group by (cont.)

\mathbf{f}	\mathbf{f}_ℓ	\mathbf{f}_g
SUM	SUM	SUM
MIN	MIN	MIN
MAX	MAX	MAX
COUNT	COUNT	SUM

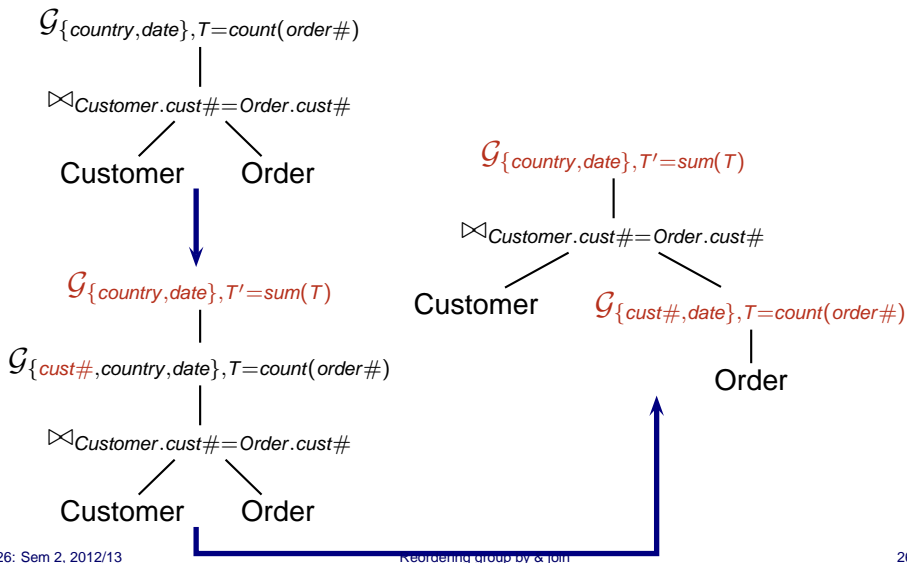
Example 4



Example 4



Example 4



Example 4

$$A = \{country, date\}, F = \{T = count(order\#)\},$$

$$A' = \{cust\#, date\}, F' = \{T' = sum(T)\}$$

Customer		
cust#	cname	country
1	Alice	UK
2	Bob	US
3	Carol	US

Order			
order#	cust#	date	totalprice
100	2	2013-01-04	3000
201	2	2013-01-04	5000
460	3	2012-12-10	4000
500	3	2013-01-04	1000

Customer \bowtie_p Order				
cust#	country	date	order#	...
3	US	2012-12-10	201	...
2	US	2013-01-04	100	...
2	US	2013-01-04	460	...
3	US	2013-01-04	500	...

$\mathcal{G}_{A,F}(\text{Customer} \bowtie_p \text{Order})$		
country	date	T
US	2012-12-10	1
US	2013-01-04	3

$\mathcal{G}_{A,F'}(X)$

$X = \text{Customer} \bowtie_p \mathcal{G}_{A',F'}(\text{Order})$

country	date	T	...
US	2013-01-04	2	...
US	2012-12-10	1	...
US	2013-01-04	1	...

$\mathcal{G}_{A',F'}(\text{Order})$

cust#	date	T
2	2013-01-04	2
3	2012-12-10	1
3	2013-01-04	1

References

Required Readings

- ▶ C.A. Galindo-Legaria, M.M. Joshi, *Orthogonal optimization of subqueries and aggregation*, SIGMOD 2001

Additional Readings

- ▶ W.P. Yan, P.A. Larson, *Performing group-by before join*, ICDE 1994
- ▶ W.P. Yan, P.A. Larson, *Eager aggregation and lazy aggregation*, VLDB 1995