## 0050001

CS5226 Lecture 3

**Query Tuning II** 

#### **Query Tuning II**

- Reordering groupby & select
- ► Reordering groupby & join

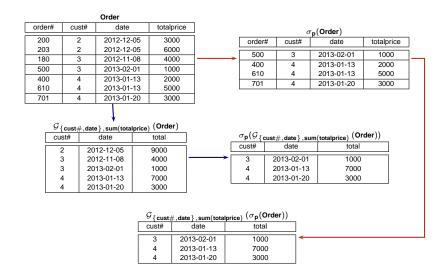
#### Reordering group by & select

- Customer (<u>cust#</u>, cname, country)
- ▶ Order (order#, cust#, date, totalprice)

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

#### Reordering group by & select

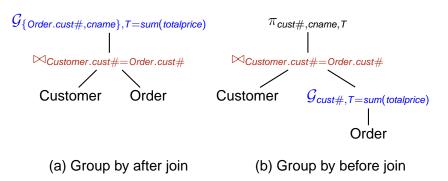
Let p denote the predicate "date≥ 2013-01-01"



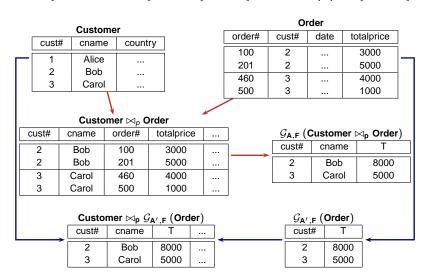
#### Reordering group by & select

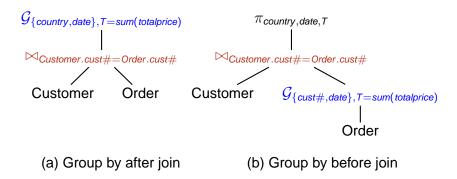
$$\sigma_p(\mathcal{G}_{A,F}\left(R
ight)) = \mathcal{G}_{A,F}\left(\sigma_p(R)
ight)$$
iff
 $A o columns(p)$ 

#### Pushing group by below join



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





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

Customer							Order								
	cust#	cname	count	rv		(	order#	С	ust#		date	tot	alprice		
	1	Alice	UK	_			100		2	201	12-12-10	;	3000		
	2	Bob	US				201		2	2013-01-04		5000			
	3	Carol	US				460		3	201	12-12-10		4000		
		Caro	03				500		3	201	13-01-04		1000		
		Cu	stomer D	$\bowtie_p$ Ord	er										
L	cust#	cust#         country         date         totalprid           2         US         2012-12-10         3000		ice			$\mathcal{G}_{A,F}$ (Customer $\bowtie_p$ Order)								
	2			3000	)			country		date		Т			
	3	US	US 2012-12-10 4000			US		2012-12-10		7000	,				
Ī	2	US	2013-	01-04	5000	)			US	;	2013-01		6000	į	
	3	US 2013-01-04 1000 .								•					
Customer $\bowtie_{p} \mathcal{G}_{A',F}$ (Order)										$\mathcal{G}$	$G_{A',F}$ (Orde	er)			
	count	ry d	ate	Т					cust#	‡	date		Т		
	US	2012	-12-10	3000	T ]				2		2012-12-1	0	3000	l	
US US		2013	-01-04	5000					2		2013-01-0	4	5000	l	
		2012	-12-10	4000					3		2012-12-1	0	4000	١	
	US	2013	-01-04	1000					3	İ	2013-01-0	4	1000	ĺ	

### Pushing group by below join

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

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

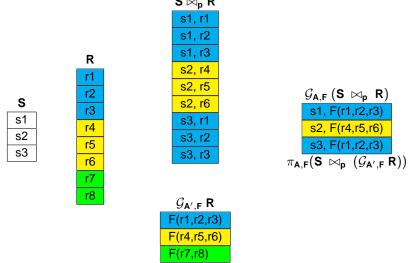
where

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

iff the following conditions hold:

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

# Pushing group by below join (cont.)

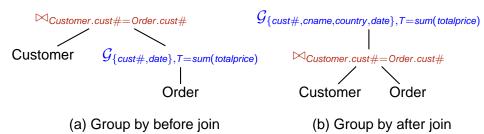


Tuples with same color have same  $columns(R) \cap columns(p)$  values

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

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

#### Pulling group by above join



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

		Custo	mer		Order								
	cust#			ntry		order# cust#		date	totalprice				
				<u> </u>		100	2	2012-12-1	10 3000				
	1 2	Alice Bob				201	2	2013-01-0	5000				
	3	Carol US				460	3	2012-12-1	10 4000				
		J Calor 03		500	3	2013-01-0	1000						
		Г	1			mer ⋈ <sub>p</sub> (			1				
		cust# cname country		date		totalprice							
		2		Bob	US	2012	-12-10	3000					
_		2 Bob US 2013-01-		-01-04	5000								
		3 Carol US 2012-12-10		-12-10	4000								
		3 Carol US 2013-01-0		-01-04	1000								
Customer $\bowtie_{p} \mathcal{G}_{A,F}$ (Order)													
$\mathcal{G}_{A,F}$ (Order)													
		cust#		‡   C	date T								
					2-12-10 3000		<del></del>						
				2		3-01-04	5000						
				2012	2-12-10	4000							
				3	2013	3-01-04	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  $\bowtie_p R$ )

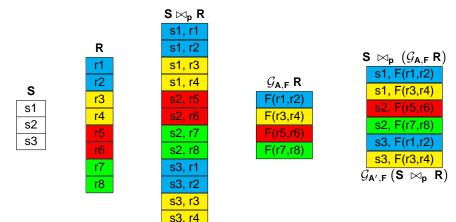
where

$$A' = A \cup columns(S)$$

iff the following conditions hold:

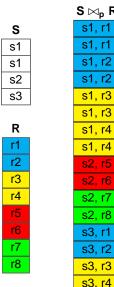
- 1. S has a key, and
- 2. p does not refer to aggregated columns in F; i.e.,  $columns(p) \cap columns(R) \subseteq A$

#### Pulling group by above join



Tuples with same color have same A values

### Pulling group by above join (cont.)



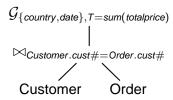
```
S ⋈<sub>D</sub> R
 s1, r2
 s1, r2
 s1, r3
 s1, r3
 s1, r4
 s1, r4
 s2, r5
 s2, r6
 s2, r7
 s2. r8
 s3. r2
 s3. r3
```

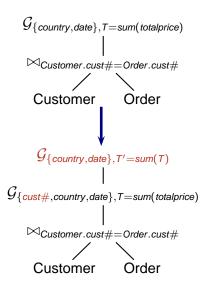
```
\mathcal{G}_{AF}R
  F(r1,r2)
  F(r3,r4)
  F(r5,r6)
  F(r7,r8)
s1, F(r1,r2)
s1, F(r3,r4)
s1, F(r1,r2)
s1, F(r3,r4)
s2, F(r5,r6)
```

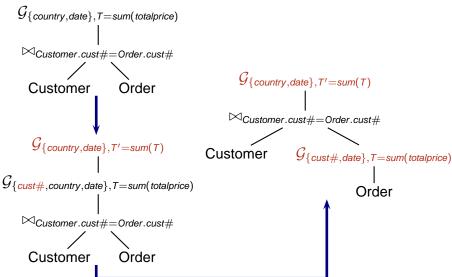
 $S\bowtie_{D} (\mathcal{G}_{A,F}R)$ s2, F(r5,r6) s2, F(r7,r8) s3, F(r3,r4) s2, F(r7,r8) s3, F(r1,r2) s3, F(r3,r4)

Suppose S does not have a key (i.e., S contains duplicates)

```
\mathcal{G}_{\mathsf{A}',\mathsf{F}}(\mathsf{S}\bowtie_{\mathsf{p}}\mathsf{R})
s1, F(r1,r1,r2,r2)
s1, F(r3,r3,r4,r4)
s3, F(r1,r2)
```





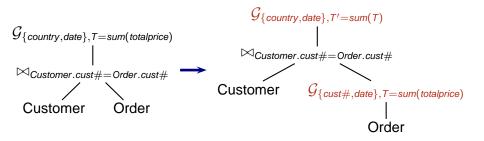


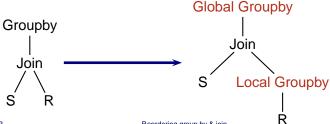
CS5226: Sem 2, 2012/13

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

Custom							Ord	er			
	-	$\neg$		(	order#	С	ust#		date	total	orice
		4			100		2	201	2-12-10	30	00
	_				201		2	201	3-01-04	50	00
Carol	US				460 500		3	2012-12-10		4000 1000	
					300		3	201	3-01-04	10	00
Customer ⋈ <sub>p</sub> Order							GA I	- (C	ustomer ⊳	ın Ore	der)
country	date		totalpı	rice				_		√p Οι (	T
US	2012-12	2-10	300	0						40	
US	2012-12	2-10	400	0		$\overline{}$		- 1			7000
US	2013-01	-04	500	0			08				
US	2013-01	-04	1000		L	~	$\mathcal{G}_{\mathbf{A},\mathbf{F}'}\left(\mathbf{X}\right)$				
$X = Customer \bowtie_{D} \mathcal{G}_{A',F}$ (Order)				/				$\mathcal{G}_{I}$	<sub>A',F</sub> (Orde	r)	
ntry d	ate	Т					cust#		date		Т
S 2012	-12-10	3000	T				2	1	2012-12-10	) 3	000
S 2013	-01-04	5000			ď		2	1	2013-01-04	5	000
S 2012	-12-10	4000					3	2	2012-12-10	) 4	000
\$ 2013	-01-04	1000					3	1	2013-01-04	L   1	000
======================================	#   cname   Alice   Bob   Carol	Alice UK Bob US Carol US  Customer ⋈ US 2012-12 US 2013-01 04 US 2013-01-04 US 2013-01-04 US 2013-01-04	# cname   country   Alice   UK   Bob   US   Carol   US    Customer $\bowtie_p$ Order   US   2012-12-10   US   2013-01-04   US   2012-12-10   US   2012-12-10   US   2013-01-04   US	# cname country  Alice UK Bob US Carol US  Customer $\bowtie_p$ Order    Country   date   totalprication	# cname   country    Alice   UK   Bob   US   Carol   US    Customer $\bowtie_p$ Order      country   date   totalprice     US   2012-12-10   4000     US   2013-01-04   5000     US   2013-01-04   1000    = Customer $\bowtie_p \mathcal{G}_{A',F}$ (Order)     intry   date   T       IS   2012-12-10   3000       IS   2013-01-04   5000       IS   2013-01-04   5000       IS   2012-12-10   3000       IS   2012-12-10   4000       IS   2012-12-10   4000	# cname   country   100   201   460   500   500	# cname   country   100   201   460   500    Customer $\bowtie_p$ Order     country   date   totalprice      US   2012-12-10   3000     4000      US   2013-01-04   5000      Example 2   2012-12-10   3000      US   2013-01-04   5000      Example 3   2012-12-10   3000      Example 4   3   3   3   3   3    Example 5   2012-12-10   3000      Example 6   3   3   3   3    Example 7   3   3   3    Example 7   3   3   3    Example 8   3   3    Example 9   3   3    Example 9   3   3    Example 9   3   3    Example 9   3    Exa	Customer       # cname   country     order#   cust#         Alice   UK   Bob   US   Carol   US     201   2   2460   3   5000   3         Customer $\bowtie_p$ Order $460$   3   5000   3           Country   date   totalprice     US   2012-12-10   3000     US   2012-12-10   4000         US   US   2013-01-04   5000     US   2013-01-04   1000             US   2013-01-04   5000     Intry   date   T     Intry   date   T     Intry   Cust#         Cust#   Cust#   Cust#   Cust#   2012-12-10   3000     2012-12-10   3000     2012-12-10   3000     300	Customer           #         cname         country           Alice         UK           Bob         US           Carol         US           Carol         US           Customer $\bowtie_p$ Order $G_{A,F}$ (Country           US         2012-12-10           US         2012-12-10           US         2012-12-10           US         2013-01-04           US         2013-01-04           US         2013-01-04           US         2013-01-04           US         2012-12-10           US         2013-01-04           US         2012-12-10           US         20           US	# cname   country   date   100   2   2012-12-10   201   2   2013-01-04   460   3   2012-12-10   500   3   2013-01-04   201   2   2013-01-04   201   2   2013-01-04   201   2   2013-01-04	Customer           # cname   country         order#   cust#   date   total            100   2   2012-12-10   30           201   2   2013-01-04   50           201   2   2013-01-04   50           460   3   2012-12-10   40           500   3   2013-01-04   10           Customer $\bowtie_p$ Order           US   2012-12-10   3000             US   2012-12-10   4000             US   2013-01-04   5000             US   2013-01-04   5000             US   2013-01-04   5000             US   2012-12-10   3000             US   2013-01-04   5000             US   2012-12-10   3000             US   2012-12-10   3000

#### Local/Global group by





CS5226: Sem 2, 2012/13

Reordering group by & join

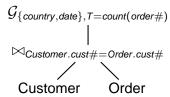
#### Local/Global group by (cont.)

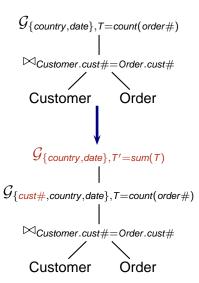
- Consider an aggregation function f on a relation R
- Let R be partitioned into n groups
  - ►  $R = R_1 \cup \cdots \cup R_n$ , ►  $R_i \cap R_i = \emptyset$ ,  $\forall i, j \in [1, n], i \neq j$
- $f_{\ell}$  = a local aggregate function
- $f_q$  = a global aggregate function

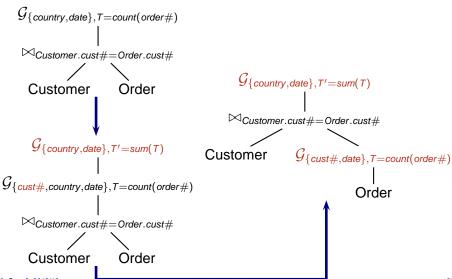
$$f(R) = f(R_1 \cup \cdots \cup R_n)$$
  
=  $f_g(f_\ell(R_1) \cup \cdots \cup f_\ell(R_n))$ 

### Local/Global group by (cont.)

f	$f_\ell$	f <sub>g</sub>
SUM	SUM	SUM
MIN	MIN	MIN
MAX	MAX	MAX
COUNT	COUNT	SUM



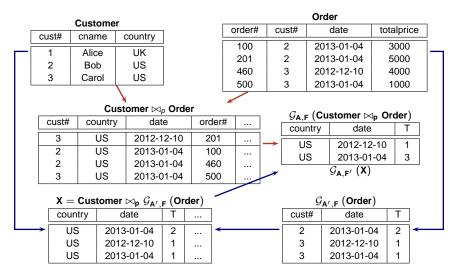




CS5226: Sem 2, 2012/13

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

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



#### 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