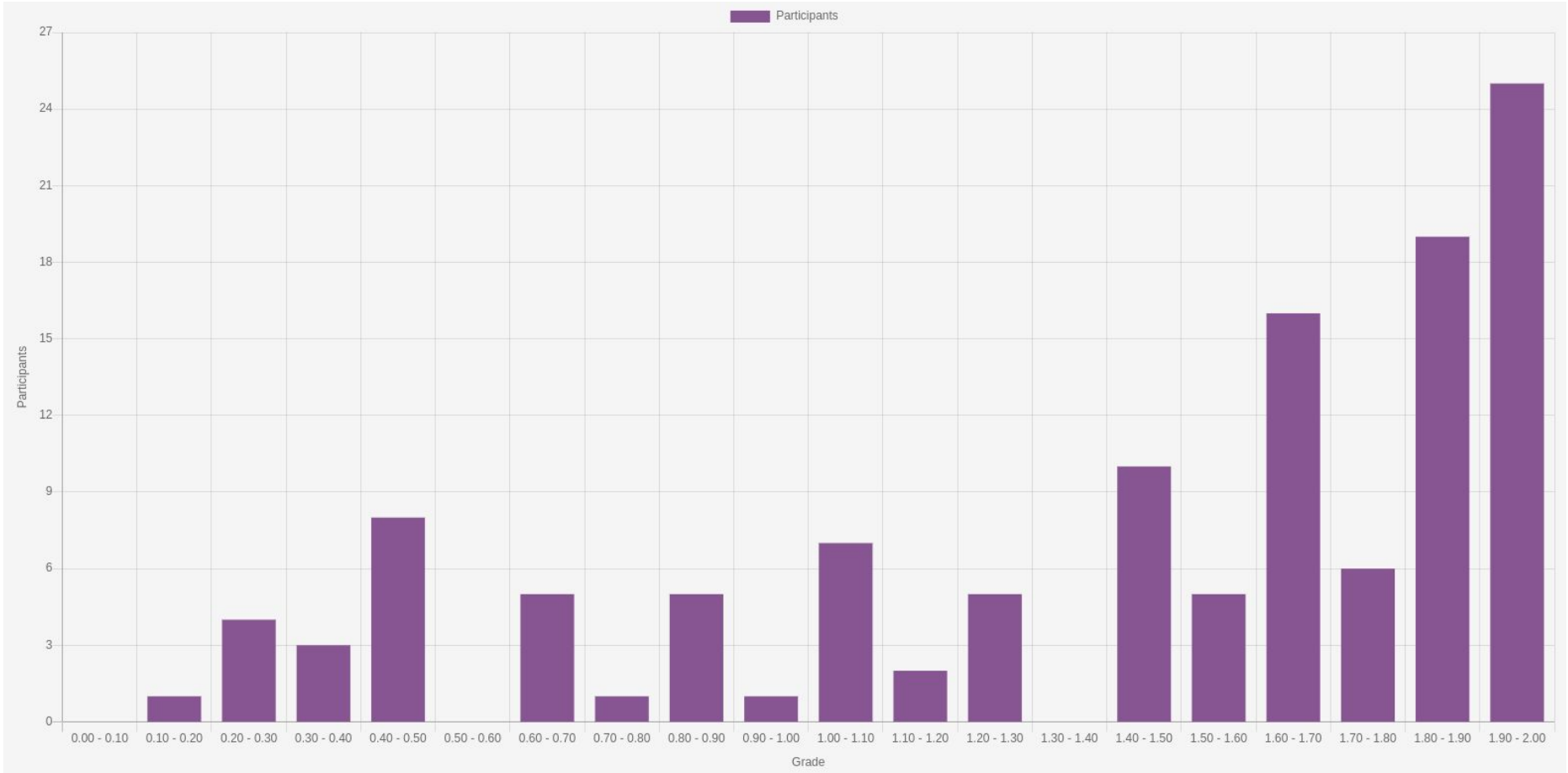
Database Technology

Exercise 7: Review


Participants
27
24
21
18
15
12
9
6
3
0
Participants
0.00 - 0.10
0.10 - 0.20
0.20 - 0.30
0.30 - 0.40
0.40 - 0.50
0.50 - 0.60
0.60 - 0.70
0.70 - 0.80
0.80 - 0.90
0.90 - 1.00
1.00 - 1.10
1.10 - 1.20
1.20 - 1.30
1.30 - 1.40
1.40 - 1.50
1.50 - 1.60
1.60 - 1.70
1.70 - 1.80
1.80 - 1.90
1.90 - 2.00
Grade

Q1: Which of the following is a valid condition rule in the grammar for simple subset of SQL?

Select all that apply. Wrong answers subtract partial points. The minimum grade is 0.

- [ ] A. <Condition> ::= <Attribute> AVAILABLE <Pattern>
- [x] B. <Condition> ::= <Condition> AND <Condition>
- [x] C. <Condition> ::= <Attribute> IN (<Query>)
- [ ] D. <Condition> ::= <Attribute> NOT FOUND IN <Pattern>

Q2: Is the following a valid algebraic law?

$\sigma_{C1 \text{ AND } C2}(R) = \sigma_{C2}(\sigma_{C1}(R))$

Assume that the relation R is a set.

True

Q3: How many tuples are selected by the selection $\sigma_{a\neq75 \text{ AND } b=20}(R)$?

Given is the table $R(a, b)$ with the following tuple and attribute domain cardinalities:

- $T(R) = 172800$
- $V(R, a) = 80$
- $V(R, b) = 40$

Assume that:

- 1 / (3 * $V(R, f)$) tuples satisfy the comparison when an attribute is equated to a constant (e.g., $\sigma_{f=15}$)
- Inequality (e.g., $\sigma_{f<90}$) tends to retrieve one-fourth of the tuples
- For not equals (e.g., $\sigma_{f\neq90}$), a fraction of 4 / $V(R, f)$ number of the tuples will fail to meet the condition

Total tuples = Ceil(T(R)*(1-4/V(R,$\sigma_{a\neq75}$))*(1/3*V(R,$\sigma_{b=20}$))
= Ceil(172800*(1-(4/80))*(1/(3*40))) = 1368

Q4: How many tuples are selected by the selection $\sigma_{a\neq 19 \text{ OR } b=28}(R)$?

Given is the table $R(a, b)$ with the following tuple and attribute domain cardinalities:

- $T(R) = 37800$
- $V(R, a) = 70$
- $V(R, b) = 30$

Assume that:

- $1 / V(R, f)$ tuples satisfy the comparison when an attribute is equated to a constant (e.g., $\sigma_{f=15}$)
- Inequality (e.g., $\sigma_{f<90}$) tends to retrieve half of the tuples
- For not equals (e.g., $\sigma_{f\neq 90}$), a fraction of $6 / V(R, f)$ number of the tuples will fail to meet the condition
- The conditions are independent

m1($\sigma_{a\neq 19}$) = Ceil(37800*(1-6/70)) = 34560

m2($\sigma_{b=28}$) = Ceil(37800*1/30) = 1260

Total tuples = Ceil(n*(1-(1-m1/n)*(1-m2/n)))
= Ceil(37800*(1-(1-34560/37800)*(1-1260/37800)))
= 34668

Q5: How many tuples are selected by the join $R \bowtie \sigma_{s2=51}(S)$?

Given are the tables $R(\underline{r_1}, s_1)$ and $S(\underline{s_1}, s_2)$. $R.\,s_1$ is a foreign key to $S.\,s_1$. The tables have the following tuple and attribute domain cardinalities:

- $T(R) = 600$
- $T(S) = 400$
- $V(S, s_2) = 30$

Assume that 1 / $V(R, f)$ tuples satisfy the comparison when an attribute is equated to a constant (e.g., $\sigma_{f=15}$)

R Join S = T(R) = 600 due to foreign key relationship

$T(R)*1/V(S,\sigma_{s2=51}) = 600*1/30 = 20$

Q6: What is the estimated size of the natural join?

Consider the natural join $R(a,d) \bowtie S(d,e) \bowtie U(a,c,e)$, and the following important statistics:

R(a, d)	S(d, e)	U(a, c, e)
T(R) = 1200	T(S) = 2400	T(U) = 18000
$V(R,a) = 20$		$V(U,a) = 450$
		$V(U,c) = 450$
$V(R,d) = 40$	$V(S,d) = 20$	
	$V(S,e) = 80$	$V(U,e) = 450$

Size of R(a,d) ⋈ S(d,e) = T(R)*T(S)/max[V(R,d),V(S,d)] as d is the join key
= 1200*2400/40 = 7200

Size of (R(a,d) ⋈ S(d,e)) ⋈ U(a,c,e) = T(R⋈ S)*T(U)/((max[V(R⋈S,a),V(U,a)])*(max[V(R⋈S,a),V(U,a)]))
as “a” and “e” are the join keys
= 7200*18000/(450*450) = 6400

Q7: What is the cost of the following join order ((T⋈U)⋈R)⋈S?

Consider four relations $R(c,d)$, $S(a,d)$, $T(b,c)$, and $U(a,b)$. Each relation has 1000 tuples. The relations have the following attribute domain cardinalities:

R(c, d)	S(a, d)	T(b, c)	U(a, b)
	$V(S,a) = 20$		$V(U,a) = 500$
		$V(T,b) = 40$	$V(U,b) = 20$
$V(R,c) = 50$		$V(T,c) = 100$	
$V(R,d) = 10$	$V(S,d) = 50$		

Cost of (((T(b,c) ⋈ U(a,b)) ⋈ R(c,d)) ⋈ S(a,d)) = Cost of ((T⋈ U)⋈R)) + Cost of (S)

= Cost of ((T⋈ U)⋈R)) + 0

= Cost of intermediate (T⋈U) + Size of ((T⋈ U)⋈R))

= Size of (T⋈U)+Cost(T)+Cost(U)+Size of ((T⋈ U)⋈R))

= 25000 + 0 + 0 +250000 = 275000

Size of T(b,c) ⋈ U(a,b) = 1000*1000/max[V(T,b),V(U,b)] as b is the join key = 1000*1000/40 = 25000

Size of (T(b,c) ⋈ U(a,b)) ⋈ R(c,d) = T(T⋈ U)*T(R)/max[V(T⋈U,c),V(R,c)] as “c” is the join key

= 25000*1000/100 = 250000

Q8: What is the size of the following join order ((S⋈R)⋈U)?

Consider four relations $R(c,d)$, $S(a,b)$, $T(a,d)$, and $U(b,c)$. Each relation has 1000 tuples. The relations have the following attribute domain cardinalities:

R(c, d)	S(a, b)	T(a, d)	U(b, c)
	$V(S,a) = 10$	$V(T,a) = 50$	
	$V(S,b) = 250$		$V(U,b) = 50$
$V(R,c) = 100$			$V(U,c) = 200$
$V(R,d) = 100$		$V(T,d) = 10$	

Size of S(a,b) ⋈ R(c,d) = T(S)*T(R) as no join key
= 1000000

Size of (S(a,b) ⋈ R(c,d)) ⋈ U(b,c) = T(S⋈R)*T(U)/((max[V(R⋈S,b),V(U,b)])*(max[V(R⋈S,c),V(U,c)]))
as “a” and “e” are the join keys
= 1000000*1000/(250*200) = 20000

Q9: When determining the optimal join order, which join orders can be ignored because they contain a cross product?

Consider four relations $R(c, d)$, $S(a, b)$, $T(b, c)$, and $U(a, d)$. Each relation has 1000 tuples. The relations have the following attribute domain cardinalities:

R(c, d)	S(a, b)	T(b, c)	U(a, d)
	$V(S, a) = 250$		$V(U, a) = 20$
	$V(S, b) = 250$	$V(T, b) = 250$	
$V(R, c) = 400$		$V(T, c) = 40$	
$V(R, d) = 50$			$V(U, d) = 200$

- [ ] A. $(R \bowtie S) \bowtie U$
- [ ] B. $(R \bowtie S) \bowtie T$
- [ ] C. $(R \bowtie T) \bowtie S$
- [ ] D. $(R \bowtie U) \bowtie S$
- [ ] E. $(R \bowtie T) \bowtie U$
- [ ] F. $(S \bowtie R) \bowtie T$

Q10: From the options below, select the best join order (i.e., the one with the lowest cost):

Consider four relations $R(a,d)$, $S(b,c)$, $T(a,b)$, and $U(c,d)$. Each relation has 1000 tuples. The relations have the following attribute domain cardinalities:

R(a, d)	S(b, c)	T(a, b)	U(c, d)
$V(R,a) = 1000$		$V(T,a) = 10$	
	$V(S,b) = 200$	$V(T,b) = 200$	
	$V(S,c) = 40$		$V(U,c) = 200$
$V(R,d) = 100$			$V(U,d) = 250$

a. $((S \bowtie U) \bowtie T) \bowtie R$

b. $((U \bowtie T) \bowtie S) \bowtie R$

c. $((R \bowtie U) \bowtie S) \bowtie T$

d. $((T \bowtie R) \bowtie U) \bowtie S$

Cost of Cost of (((T⋈R)⋈U)⋈S) = Cost of ((T⋈R)⋈U) + Cost of (S)
= Cost of ((T⋈R)⋈U) + 0
= Cost of intermediate (T⋈R) + Size of ((T⋈R)⋈U))
= Size of (T⋈R)+Cost(T)+Cost(R)+Size of ((T⋈R)⋈U))
= 1000 + 0 + 0 +4000 = 5000

Similarly,
Cost of (((S⋈U)⋈T)⋈R) = 30000
Cost of (((U⋈T)⋈S)⋈R) = 1050000
Cost of (((R⋈U)⋈S)⋈T) = 24000

Size of T(a,b) ⋈ R(a,d) = 1000*1000/max[V(T,a),V(R,a)] as a is the join key
= 1000*1000/1000 = 1000
Size of (T(a,b) ⋈ R(a,d))⋈U(c,d) = T(T⋈R)*T(U)/max[V(T⋈R,d),V(U,d)] as “d” is the join key
= 1000*1000/250 = 4000