TEMA 2

26/10/16

Problema 1:

$S = \sigma_{a=10 \vee b<20}(R)$

→ $S1 = \sigma_{a=10}(R)$

$T(S1) = \dfrac{1000\cancel{0}}{5\cancel{0}} = \boxed{200}$

→ $S2 = \sigma_{b<20}(R)$

$T(S2) = \dfrac{10000}{3} = \boxed{3333,33}$

¡cuando es desigualdad divido entre 2 ó 3!

EJERCICIO INVENTADO EN CLASE (Rangos)

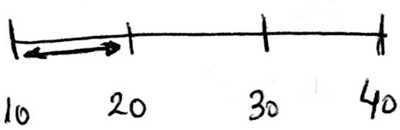
$T(R) = 10000$

$S = \sigma_{b<20}(R)$

$V(R,b) = 100$

$MIN = 10$

$MAX = 40$



$\dfrac{20 - \min}{(\max - \min) + 1} = \dfrac{10}{31} = 0'32 = f$

$f \times V(R,b) = 0,32 \times 100 = 32$

$\dfrac{T(R)}{V(R,b)} = \dfrac{100\cancel{00}}{1\cancel{00}} = 100$

$T(S) = 32 \times 100 = 3200$

Formula: $\frac{T(R)}{V(R,a)}$ ↳ por cada tabla

PROBLEMA 2

$R(a,b)$

$T(R) = 10000$

$V(R,a) = 50$

$S = \sigma_{a=10 \wedge a>20}(R)$

¿CUAL SERA EL NUMERO ESTIMADO DE TUPLAS DE S?

→ Cero, ya que no hay ninguna fila que cumpla esta condición.

PROBLEMA 3

$X = R1(A,B) \bowtie R3(C,D)$

$T(X) = T(R1) \times T(R3) = 3000000$

$V(X,A) = 50$

$V(X,B) = 100$

$V(X,C) = 90$

$V(X,D) = 500$

$Z = X \bowtie R2(B,C)$

$$(T_Z) = \frac{T(R2) \times T(X)}{V(R2,B) \times V(R2,C)} = \frac{2000 \times 3\cancel{000000}}{2\cancel{00} \times 3\cancel{00}} = \frac{60000}{6} = \boxed{10000}$$

$V(Z,A) = 50$

$V(Z,B) = 100$

$V(Z,C) = 90$

$V(Z,D) = 500$

PROBLEMA 5

$R(a,b)$; $V(R,a)=25$

$T(R) = 2500$;

$S = \sigma_{a=10 \vee a \leq 10}(R)$

NOTA: $a=10$ esta dentro de $a \leq 10$, con lo cual voy a resolver $\underline{a \leq 10}$

¡Genérico de rango dividido entre 2 ó 3!

$$T(S) = \frac{2500}{2} = 1250$$

$$V(S,a) = \frac{25}{2} \simeq 12$$

¡explicar resultado!

PROBLEMA 6

$R(a,b)$

$T(R) = 2500$

$V(R,a) = 25$

$S = \sigma_{a=10 \vee a \leq 10 \vee a \geq 10}(R)$

¡Tomamos todos los valores, ya que las condiciones no simplifican ninguna fila!

$T(S) = T(R)$

PROBLEMA 7 $T(R) = 1000$.

$R(a,b,c)$

$S(d,b,f)$

$$X = \underbrace{\left(\sigma_{a=10 \wedge b \leq 10}(R)\right)}_{y} \bowtie S$$

$$y = \sigma_{a=10 \wedge b \leq 10}(R) = \underbrace{\sigma_{=10}\left(\underbrace{\sigma_{\leq 10}(R)}_{y_1}\right)}_{y_2}$$

$$T(y_1) = \frac{T(R)}{2} = \frac{1000}{2} = \boxed{500}$$

$$T(y_2) = \frac{T(y_1)}{V(y_1,a)} = \frac{500}{20} = \boxed{25} = T_y$$

sigue

$T(y) = 25$

¡ numero de valores diferentes de la columna a, b, c !

$V(y,a) = 1$

$V(y,b) = 25$.

$V(y,c) = 12$ pongo la media.

$x = y \bowtie S$

$$T(x) = \frac{\cancel{5000} \times 25}{\cancel{1000}} = \boxed{125}$$

$V(x,a) = 1$ no tiene nada en comun asi que dejo el original.

$V(x,b) = 25$ ¡ tiene en comun !

$V(x,c) = 12$

$V(x,d) = 2$

$V(x,f) = 7$

~~xn~~

PROBLEMA 8 $\quad T(R) = 1000$ $\quad$ ¡¡ $\boxed{\text{si es } \wedge \text{ si es } \lor}$

$R(a,b,c)$

$S(d,b,f,e,g)$

$$X = \underbrace{\sigma_{a=10 \lor b<=10}(R)}_{y} \bowtie \underbrace{(\Pi_{d,b}(S))}_{w}$$

$y_1 = \sigma_{a=10}(R)$

$T(y_1) = \dfrac{100\cancel{0}}{2\cancel{0}} = \boxed{50}$ $\quad +$

$y_2 = \sigma_{b \le 10}(R) = \dfrac{1000}{2} = \boxed{500}$

$T(y) = T(y_1) + T(y_2) = \boxed{550}$

$V(y,a) = 20/2 = 10$ $\quad$ ¡por la condicion!

$V(y,b) = 550$

$V(y,c) = 150$

$T(w) = T(S) = 5000$

$V(w,d) = 2$

$V(w,b) = 1000$

$X = y \bowtie w$

$$T(x) = \frac{T(y) \times T(w)}{\max(V(y,b)\ V(w,b))} = \frac{550 \times 5\cancel{000}}{1\cancel{000}} = \boxed{2750}$$

$V(x,a) = 10$

$V(x,b) = 550$

$V(x,c) = 150$

$V(x,d) = 2$

$T(R) \rightarrow$ # tuplas en R (nº filas de la tabla R)

$S(R) \rightarrow$ # de bytes en cada tupla de R.

$B(R) \rightarrow$ # de bloques para guardar todas las tuplas de R

$V(R.A) \rightarrow$ # valores diferentes de A en R.

$\wedge \rightarrow y$
$\vee \rightarrow o$

tocb

$\boxed{\text{ESTIMACION DEL TAMAÑO}}$ — $\boxed{W = R_1 \times R_2}$

$$T(W) = T(R_1) * T(R_2)$$

$$S(W) = S(R_1) + S(R_2)$$

$\boxed{W = \sigma_{A=a}(R)}$

$$= \frac{T(R)}{V(R.a)}$$

des $\frac{T\ R}{2'3}$

$$S(W) = S(R)$$

$$T(W) = \underbrace{\frac{T(R)}{2}}_{\text{Media.}} \quad \text{ó} \quad T_W = \underbrace{\frac{T(R)}{3}}_{\text{Trozos pequeños}}$$

$\boxed{\text{RANGO}}$

Para el rango vemos un ejemplo para entenderlo mejor.

$Q = [1, 20]$

min max

max

min

intervalo.

$T(R) = 16$
MIN $= 1$
MAX $= 20$
$V(R.Z) = 10$
$W = \sigma_{Z \geq 15}(R)$

$$f = \frac{20 - 15 + 1}{20}$$

$\boxed{W = R_1 \bowtie R_2}$

$X \cap Y = \emptyset \rightarrow$ si no hay columnas en comun es un producto Cartesiano

$X \cap Y = A \rightarrow$ si tenemos columnas comunes

PROBLEMA 1

¿ Numero de tuplas estimado de S?

$R(a,b) \longrightarrow$ RELACION

$T(R) = 10000 \longrightarrow$ Nº tuplas (filas de R)

$V(R,a) = 50 \longrightarrow$ Valores distintos de a en R

$S = \sigma_{a=10 \vee b<20}(R) \longrightarrow$ Condiciones.

$\longrightarrow S_1 = \sigma_{a=10} \longrightarrow$ AL SER $(=)$ $\frac{TR}{V(Ra)} = \frac{10000}{50} = \boxed{200}$

$\longrightarrow S_2 = \sigma_{b<20} \longrightarrow$ AL SER (desigualdad $<$) $\frac{TR}{3} = \boxed{3333}$

! cuando hay desigualdad <> divido entre 2 ó 3 depende del resultado de cada uno elijo. !

$S(R) = S_1 + S_2 \longrightarrow S(R) = 200 + 3333 = \boxed{3533}$

* Si el calculo hubiera sido superior al T(R) deberíamos hacerlo de otra forma

PROBLEMA 2

¿ Número de tuplas estimado de S?

$R(a,b) \longrightarrow$ Relación

$T(R) = 10000 \longrightarrow$ nº tuplas de R (filas)

$V(R,a) = 50 \longrightarrow$ Valores distintos de a en R

$S = \sigma_{a=10 \wedge a>20}(R) \longrightarrow$ Condiciones.

* No podemos estimar ya que ~~son~~ las dos condiciones no se pueden cumplir.

PROBLEMA 3 Calcule $T(Z)$, teniendo en cuenta el primer $\bowtie$ es $R_1 \bowtie R_3$.

$$Z = R_1(A,B) \bowtie R_2(B,C) \bowtie R_3(C,D)$$

$T(R_1) = 1000$	$V(R_1,A) = 50$	$V(R_1,B) = 100$
$T(R_2) = 2000$	$V(R_2,B) = 200$	$V(R_2,C) = 300$
$T(R_3) = 3000$	$V(R_3,C) = 90$	$V(R_3,D) = 500$

$U = R_1 \bowtie R_3 = \emptyset$ (no tienen columnas en comun, por lo tanto es un producto cartesiano)

$T(U) = 1000 \times 3000 = 3000000$

$V(U,A) = 50$

$V(U,B) = 100$

$V(U,C) = 90$

$V(U,D) = 500$

$$\boxed{\frac{T(R)}{V(R,-)}}$$ por cada tabla

$Z = U \bowtie R_2(BC) \longrightarrow T(Z) = \dfrac{T(U) \times T(R_2)}{V(R_2,B) \times V(R_2,C)} = \longrightarrow$

$\longrightarrow T(Z) = \dfrac{30000\cancel{00} \times 20\cancel{00}}{2\cancel{00} \times 3\cancel{00}} = \dfrac{60000}{6} = 10000$

$V(Z,A) = 50$

$V(Z,B) = 100$

$V(Z,C) = 90$

$V(Z,D) = 500$

¿Porque vuelve a tomar los valores de R1, R3?

PROBLEMA 4 Estadisticas para las relaciones E, F y G

E (a,b,c)	F (a,b,d)	G (a,c,d)
T(E) = 1000	T(F) = 2000	T(G) = 3000
V(E,a) = 500	V(F,a) = 50	V(G,a) = 500
V(E,b) = 100	V(F,b) = 200	V(G,c) = 300
V(E,c) = 20	V(F,d) = 100	V(G,d) = 100

Estadisticas finales $(G \bowtie (E \bowtie F))$

1° paso

$U = E \bowtie F$ → donde tienen (a,b) en comun.

$$T(U) = \frac{\dfrac{T(E) \times T(F)}{\max(V(E,a), V(F,a))}}{\max(V(E,b), V(F,b))} = \frac{\dfrac{1000 \times 2000}{500}}{200} = \frac{1000 \times 2000}{200 \times 500} = \boxed{20}$$

$V(U,a) = 50$

$V(U,b) = 100$

$V(U,c) = 20$

$V(U,d) = 100$

Valores minimos ¿?

2° Paso

$Z = U \bowtie G$ → donde los atributos comunes son (a,c,d)

$$T(Z) = \frac{\dfrac{\dfrac{T(U) \times T(G)}{\max(V(U,a), V(G,a))}}{\max(V(U,c), V(G,c))}}{\max(V(U,d), V(G,d))} = \frac{\dfrac{\dfrac{20 \times 3000}{500}}{300}}{100} = \frac{20 \times 3000}{500 \times 300 \times 100} = \frac{60000}{15.000.000} = \boxed{0}$$

$V(Z,a) = 0$

$V(Z,b) = 0$

$V(Z,c) = 0$

$V(Z,d) = 0$

$\boxed{\text{PROBLEMA 5}}$

$R(a,b) \longrightarrow$ Relación

$T(R) = 2500 \longrightarrow$ numero tuplas de R (filas)

$V(Ra) = 25 \longrightarrow$ Valores distintos de a en R

$S = \sigma_{a=10 \vee a \leq 10}(R) \longrightarrow$ Condiciones.

Numero de tuplas estimado de S.

Explicacion : $\left.\begin{matrix} a = 10 \\ a \leq 10 \end{matrix}\right\}$ a=10 esta dentro de $a \leq 10$.

por lo tanto resuelvo $a \leq 10$.

$$\left.\begin{matrix} T(S) = \dfrac{2500}{2} = 1250 \text{ Filas} \\ \text{ó} \\ T(S) = \dfrac{2500}{3} = 834 \text{ Filas.} \end{matrix}\right\}$$ cualquiera de las dos valdría ya que siguen siendo menor que T(R)

$$V(S,a) = \frac{25}{2} \simeq 12$$

ó

$$V(S,a) = \frac{25}{3} \simeq 8$$

$\boxed{\text{PROBLEMA 6}}$

$R(a,b)$

$T(R) = 2500$

$V(R,a) = 25$

$S = \sigma_{a=10 \vee a \leq 10 \vee a \geq 10}$

Número de tuplas estimado de S.

* En este Caso $T(S) = T(R) = 2500$ ya que las condiciones engloban todas las filas de la tabla.

PROBLEMA 7 $R(a,b,c)$ y $S(d,b,f)$

$T(R) = 1000$	$T(S) = 5000$
$V(R,a) = 20$	$V(S,d) = 2$
$V(R,b) = 1000$	$V(S,b) = 1000$
$V(R,c) = 150$	$V(S,f) = 7$

Numero estimado de tuplas de x.

$$X = \left(\sigma_{a=10 \wedge b \leq 10}(R)\right) \bowtie (S)$$

1° PASO estimar n° tuplas y valores estimados de V para $\boxed{\sigma_{a=10 \wedge b \leq 10}(R)}$ a la que llamare (R')

$$T(R') \longrightarrow R' = \left(\sigma_{a=10 \wedge b \leq 10}(R)\right) = \Big(\underbrace{\sigma_{a=10}}_{2°}\big(\underbrace{\sigma_{b \leq 10}(R)}_{1°}\big)\Big)$$

1° $T\left(\sigma_{b \leq 10}(R)\right) = \dfrac{1000}{2} = 500$ $\boxed{\text{Tambien podriamos haber dividido } /3}$

2° $T\left(\left(\underbrace{\sigma_{a=10}}_{V(\ )}\big(\underbrace{\sigma_{b \leq 10}(R)}\big)\right)\right) = \dfrac{500}{20} = \boxed{25 = T(R')}$

$T(R') = 25$ ¿? número minimo de valores diferentes.

$V(R',a) = 1$

$V(R',b) = 25$

$V(R,c) = 25/2 \simeq 12$ ¡la media escojo.!

2° PASO Estimamos las estadísticas de $\boxed{X = R' \bowtie S}$

$T(R') = 25$ $\bowtie$ $T(S) = 5000$

$V(R,a) = 1$ $V(S,d) = 2$

$V(R,b) = 25$ $V(S,b) = 1000$

$V(R,c) = 12$ $V(S,f) = 7$

3° PASO Solo tienen b en comun

$$T(X) = \frac{T(R') \times T(S)}{\max V(R,b), (S,b)} = \frac{5000 \times 25}{1000} = \boxed{125}$$

$V(X,a) = 1$

$V(X,b) = 25$

$V(X,c) = 12$

$V(X,d) = 2$ ¡Importante!

$V(X,f) = 7$

PROBLEMA 8 $R(a,b,c)$ y $S(d,b,f,e,g)$

$T(R) = 1000$	$T(S) = 5000$
$V(R,a) = 20$	$V(S,d) = 2$
$V(R,b) = 1000$	$V(S,b) = 1000$
$V(R,c) = 150$	$V(S,f) = 7$
	$V(S,e) = 200$
	$V(S,g) = 127$

Numero estimado tuplas de X

$X = (\sigma_{=10} \vee b \leq 10(R)) \bowtie (\Pi\, d,b(S))$

1° Paso Valores estimadas V para $(\sigma = 10 \vee b \leq 10(R)) = R'$

1° $(\sigma\, a = 10(R) \rightarrow T(R)/V(R,a) = \frac{1000}{20} = 50$

2° $(\sigma\, b \leq 10(R)) \rightarrow T(R)/2 = \frac{1000}{2} = 500$

- No podemos una OR con dos condiciones, por lo tanto nos quedemos con el maximo $\boxed{500}$

$T(R') = 500$

$V(R'a) = 1$

$V(R'b) = 500$

$V(R'c) =$ min $= 1$ y max $= 150$, así que nos quedamos con la media 75

2° Paso $S' = (\Pi\, d,b,(S))$, hemos tenido suerte ya que el numero de tuplas coincide con el de S y ademas se eliminen las columnas f, e y g.

$T(S') = 5000$

$V(S,d) = 2$

$V(S,b) = 1000$

3° Paso $X = R' \bowtie S'$

$T(R') = 500$

$V(R'a) = 1$

$V(R'b) = 500$

$V(R'c) = 75$

$T(S') = 5000$

$V(S',d) = 2$

$V(S',b) = 1000$

4° Paso atributo b en comun.

$T(X) = 500 \times 5000 / 1000 = 2500$

$V(X\,a) = 1$

$V(X\,b) = 500$

$V(X\,c) = 75$

$V(X\,d) = 2$

a) $(Z \bowtie X) \bowtie W$

$\wedge \xrightarrow{I} y$ (alcanzan)
$\vee \longrightarrow o$ (suman)

$Z \cap X = \emptyset$ No tienen columnas en comun, por lo tanto es un producto cartesiano.

$U = Z \bowtie X \longrightarrow T(U) = 100 \times 300 = 30000$

$V(U,d) = 10$

$V(U,e) = 50$

$V(U,f) = 100$

$V(U,b) = 60$

$V(U,c) = 100$

$\dfrac{T(\ )}{maxV(\_.\_)}$ por cada tabla

$$R = U \bowtie W\ (a,b) \longrightarrow T(R) = \frac{T(U) \times T(W)}{max(V(W,b)\,V(X,b))} = \frac{30000 \times 4\cancel{00}}{6\cancel{0}} = \frac{120000\cancel{0}}{6\cancel{0}}$$

$T(R) = 200.000$

$V(T.a) = 50$

$V(T.b) = 40$

$V(T.c) = 50$

$V(T.d) = 10$

$V(T,e) = 50$

$V(T,f) = 100$

Tomamos los valores minimos en caso de dos columnas iguales.

* f es primary Key por eso nos sale T(R) muy elevado.

(b) $\sigma_{a=1} \vee b > 22 (w)$ Me centro en la tabla (w)

$$w' = \underbrace{\sigma_{a=1}}_{R_1} \vee \underbrace{b > 22 (w)}_{R_2}$$

$R_1 \rightarrow$ al ser una igualdad $\rightarrow \frac{T(w)}{V(w,a)} = \frac{400}{50} = 8$

$R_2 \rightarrow$ al ser una desigualdad $\rightarrow \frac{T(w)}{2'3} = \frac{400}{3} \approx 133$

$w' = 8 + 133 \approx 201$

* Al no disponer de un criterio para estimar operaciones OR de dos condiciones en este caso tomaremos el max.

$T(w') = 133$

$V(w', a) = 50$

$V(w', b) = 40$

$T(w', c) =$

c) $\sigma_{Z.d = Y.d}\left(\left(\underbrace{(W \times Y)}\right) \times Z\right) = R$; $W \times Y$ es producto carteriano xq no tienen nd en com..

$R = \left(\underbrace{\underbrace{(W \times Y)} \times Z}_{R_1}\right) \rightarrow R_1 = T(W) \times T(Y) = 400 \times 200 = 80000$

$T(R_1) = 80000$

$V(R_1, a) = 50$

$V(R_1, b) = 40$

$V(R_1, c) = 50$

$V(R_1, d) = 20$

$V(R_1, f) = 100$

$R = R_1 \times Z \rightarrow R_1 \bowtie Z\ (b)$; ya que tienen d en comun.

$T(R) = \dfrac{T(R_1) \times T(Z)}{\max(V(R_1 d)\ V(Z, d))} = \dfrac{80000 \times 10\cancel{0}}{2\cancel{0}} = \dfrac{800000}{2} = 40000$

$T(R) = 4000$

$V(R a) = 50$

$V(R b) = 40$

$V(R c) = 50$

$V(R d) = 10$

d) $\sigma_{b \neq 10 \wedge b = 10}(x) = x'$

No podemos estimar nada, ya que las dos condiciones son contradictorias, no puede haber una tupla en la que se cumplan las dos condiciones a la vez.

$\boxed{T(x') = 0}$

(e) $\sigma_{c\, != 25}(X) = R$

Este apartado podemos calcularlo de dos formas distintas.

(1) $\sigma_{c < 25 \vee c > 25}$

(2) $\sigma_{c = 25}$

(2) $\sigma_{c = 25}$

$$R = \frac{T(X)}{V(X,c)} = \frac{300}{100} = 3 \text{ filas iguales a } 25$$

Por lo tanto $300 - 3 = 297$ filas $!= 25$

$T(R) = 297$

$V(R,c) =$ min 1 max 297 ó 97 ya que le -3.

$V(R,b) = 60$

(f) $\sigma\ d = 10 \wedge e = 7 \wedge f < 15\ (Z)$

En este apartado solo puedo resolver $f < 15$ ya que

$$\left.\begin{array}{l} d = 10 \ \checkmark \\ e = 7 \ \times \end{array}\right] \quad Min(f) = 0 \leftrightarrow Max(f) = 20$$

$R = \sigma\ f<15\ (Z)$

tomo la mitad

10

$$\boxed{f = 0{,}71} = \frac{15 - min}{(max - min) + 1} = \frac{15 - 0}{(20-0)+1} = \frac{15}{21}$$

$$f \times V(Z\ f) = 0{,}71 \times 100 = 71$$

$$\frac{T(Z)}{V(Z f)} = \frac{100}{100} = 1 \quad \text{Primary Key}$$

$$T(R) = 71 \times 1 = \boxed{71 \text{ tuplas}}$$

$V(R\ d) = 10$

$V(R\ e) = 50$

$V(R\ f) = 71$ al ser primary Key

9) $\sigma \underbrace{(e=10 \wedge e>20)}_{\text{no se cumple.}} \vee d>3 \; (Z)$

el and no se puede estimar ya que $e=10$ no puede ocurrir nunca por lo tanto nos centramos en $d>3 \; (Z)$

$$T_{(R)} = \frac{T_Z}{V(Z\,d)} = \frac{100}{2} = 50$$

$T_{(R')} = 50$

$V(R\,d) = 10$

$V(R\,e) = 50$

$V(R\,f) = 10$ por la media.

(H) $\underbrace{W \bowtie X}_{R_1} \bowtie \underbrace{Y \bowtie Z}_{R_2} = T(R)$

$T(R) = R_1 \bowtie R_2 \rightarrow T(R_1) = \dfrac{T(W) \times T(X)}{\max(V(W,b)\,V(X,b))} =$

$= \dfrac{400 \times 300}{60} = \dfrac{1200\cancel{0}}{6\cancel{0}} = 2000$

$T(R_1) = 2000$

$V(R_1 a) = 50 \rightarrow \min\ 10 - \max\ 55$

$V(R_1 b) = 40$

$V(R_1 c) = 100$

$T(R_2) = \dfrac{T(Y) \times T(Z)}{\max(V(Yd))(V(Zd))} = \dfrac{200 \times 100}{20} = \dfrac{2000\cancel{0}}{2\cancel{0}} = 1000$

$T(R_2) = 1000$

$V(R_2, c) = 50$

$V(R_2, d) = 10$

$V(R_2, e) = 50$

$V(R_2, f) = 100 \quad \min 0 \ \max 20$

$T(R) = \dfrac{T(R_1) \times T(R_2)}{\max(V(R_1 c))(V R_2, c)} = \dfrac{2000 \times 1000}{50} = \dfrac{200000\cancel{0}}{5\cancel{0}} = 40000$

$T(R) = 40000$

$V(Ra) = 50 \quad (10, 55]$	$V(Rd) = 10$
$V(Rb) = 40$	$V(Re) = 50$
$V(Rc) = 50$	$V(Rf) = 100 \quad \max(0, 20]$