Note Title 2/13/2014

Relational Algebra

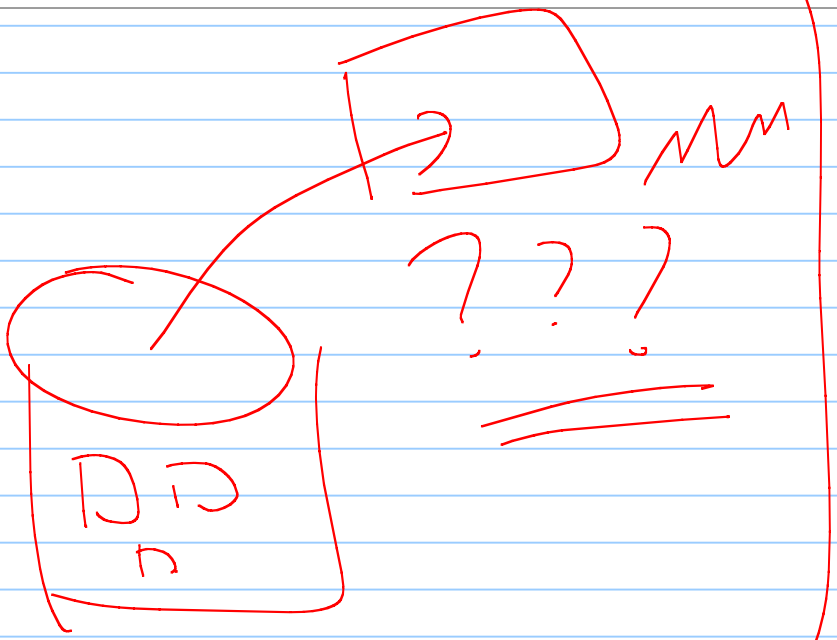
R ∝ S

GS_A (R)

Level of physical plan

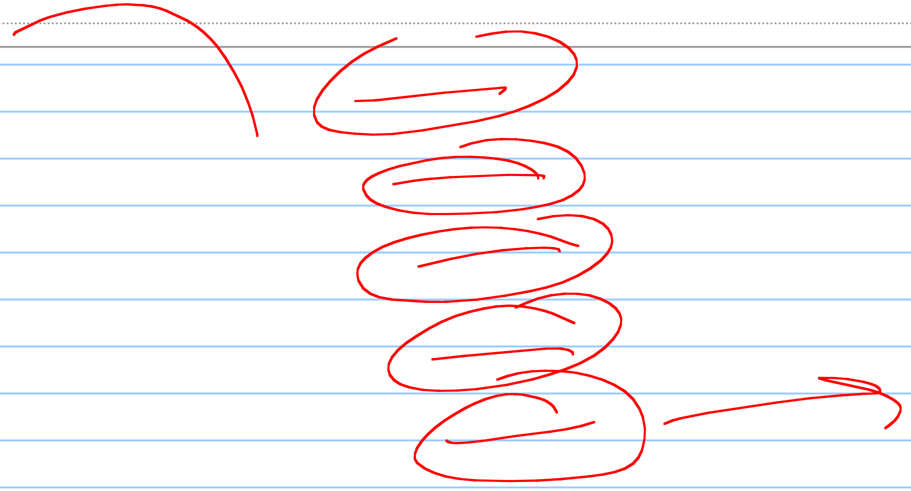
Query

Strategy



Note Title 2/13/2014

SQL



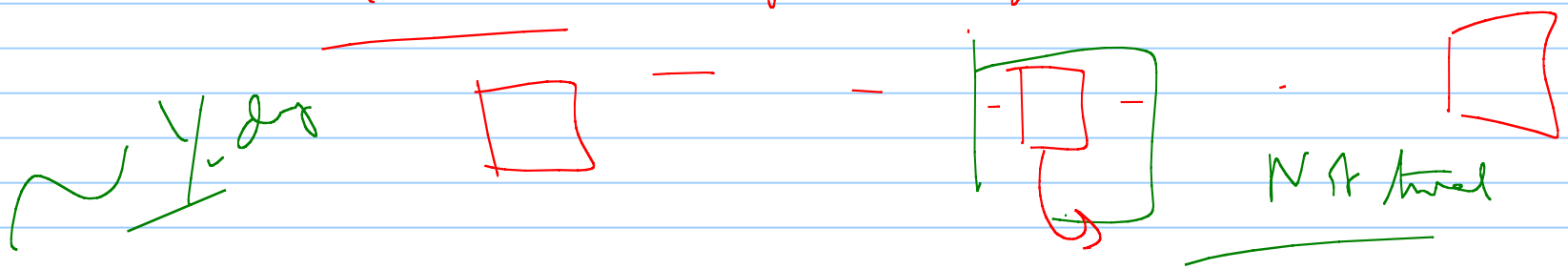
set of program codes.

Evolution is getting the worlds.

Note Title 2/13/2014

Problem n° ~ 10

10 ! ways in which you can be done



Go $\pi \left(\sigma_{C_1 \wedge C_2 \ldots} (R_1 \times R_2 \times \ldots \times R_{10}) \right)$

$\pi \left[\sigma_{C_1}(R_1) \bowtie \sigma_{C_2}(R_2) \bowtie \ldots \bowtie \sigma_{C_{10}}(R_{10}) \right]$

Go A plus

$\Downarrow$

$\sigma_{C_i}(R_i)$ $\left[\sigma_{C_i}(R_i) \bowtie \sigma_{C_j}(R_j) \right]$

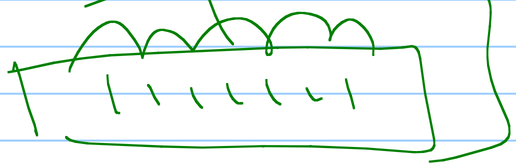
~~(R_i and R_j)~~

Study greedy

a plan

Signature

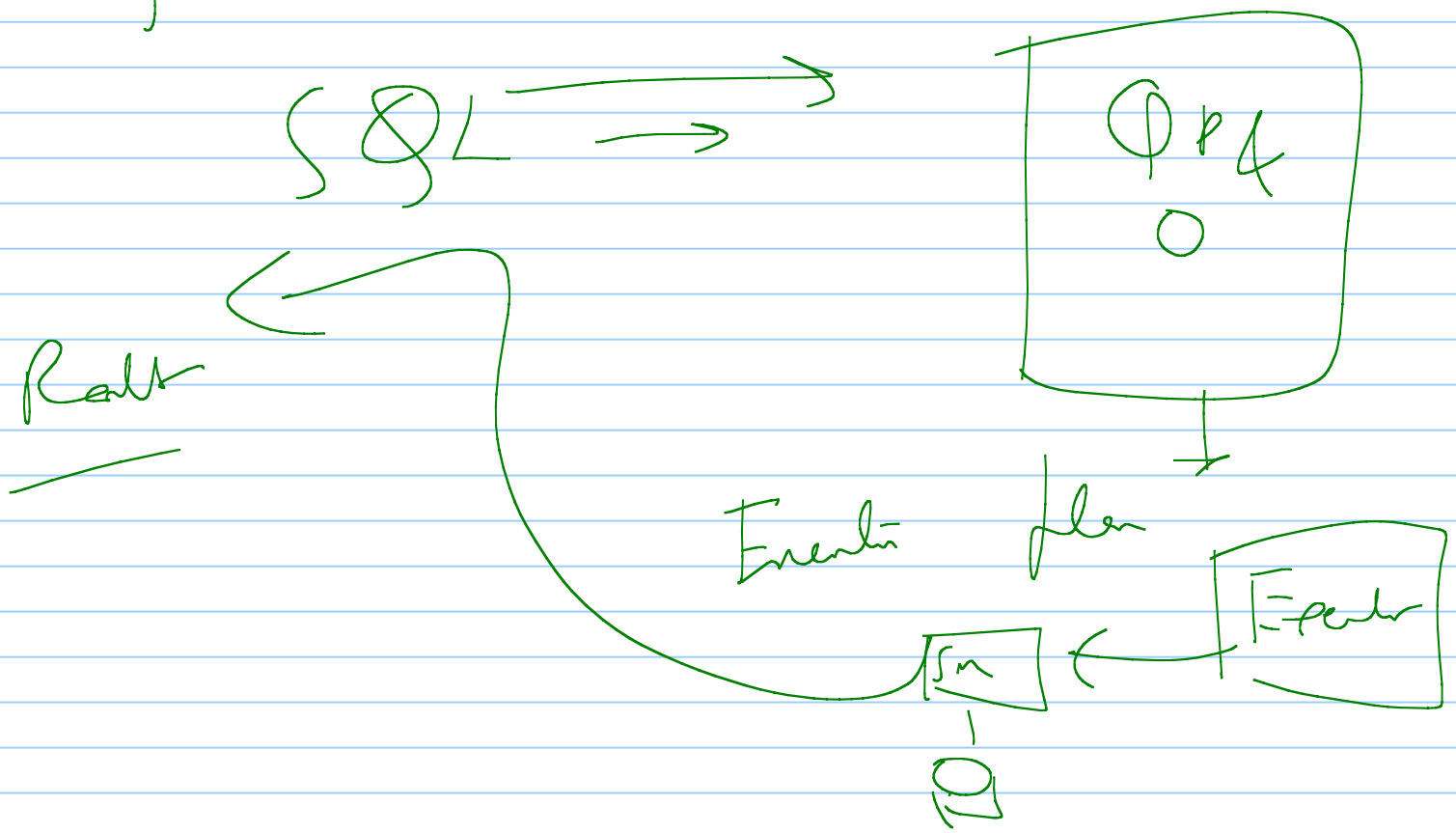
incremental | DW | Dynamic



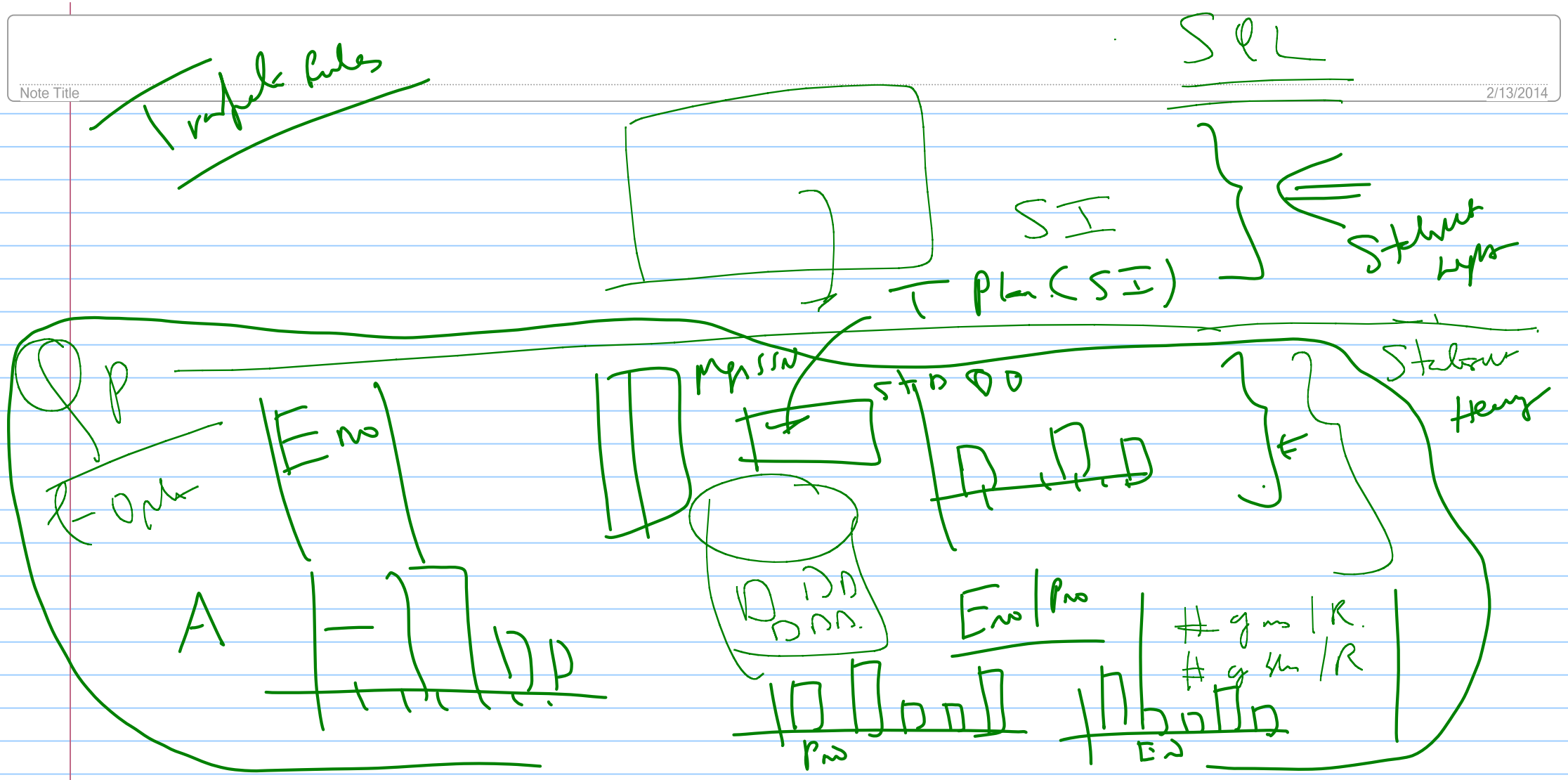
reusable | optimal |

Note Title 2/13/2014

How do I learn this?



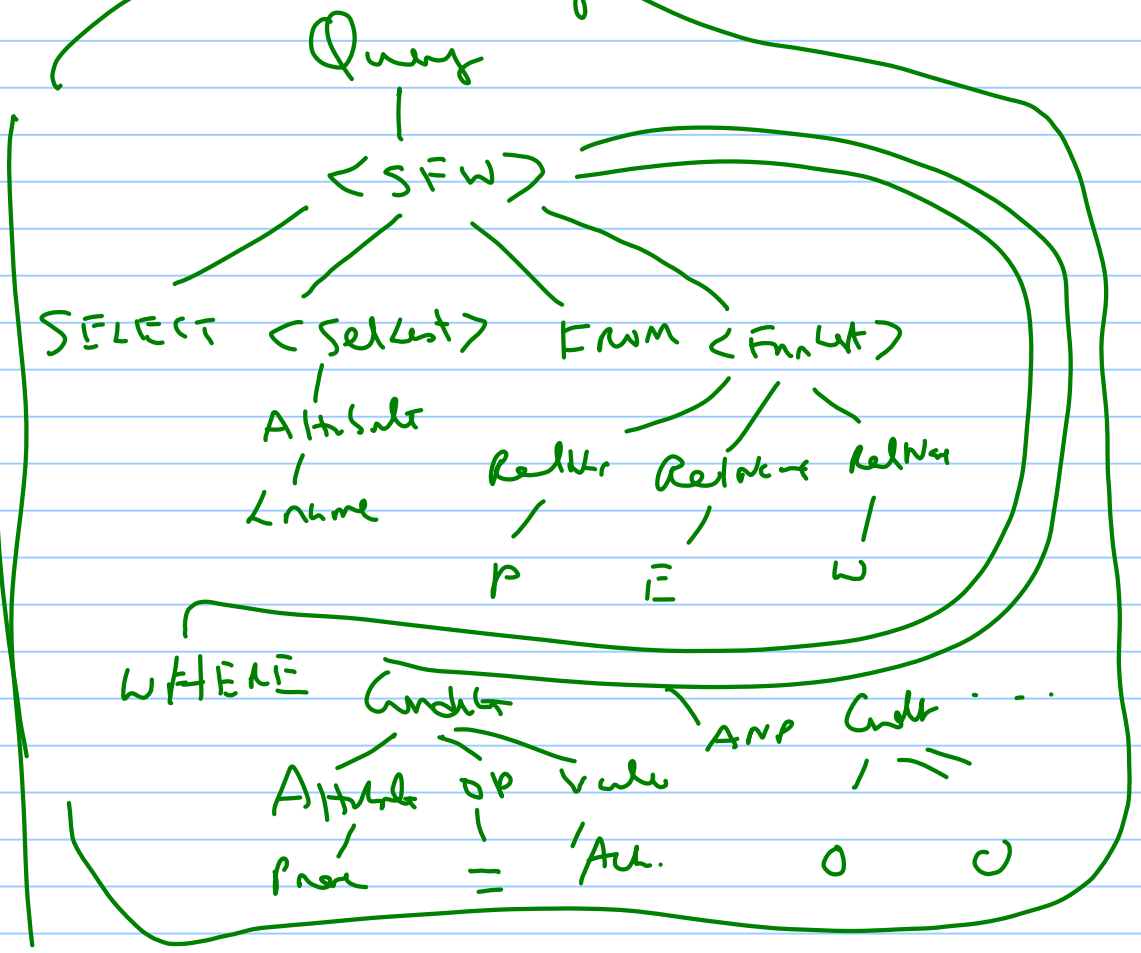
Note Title 2/13/2014



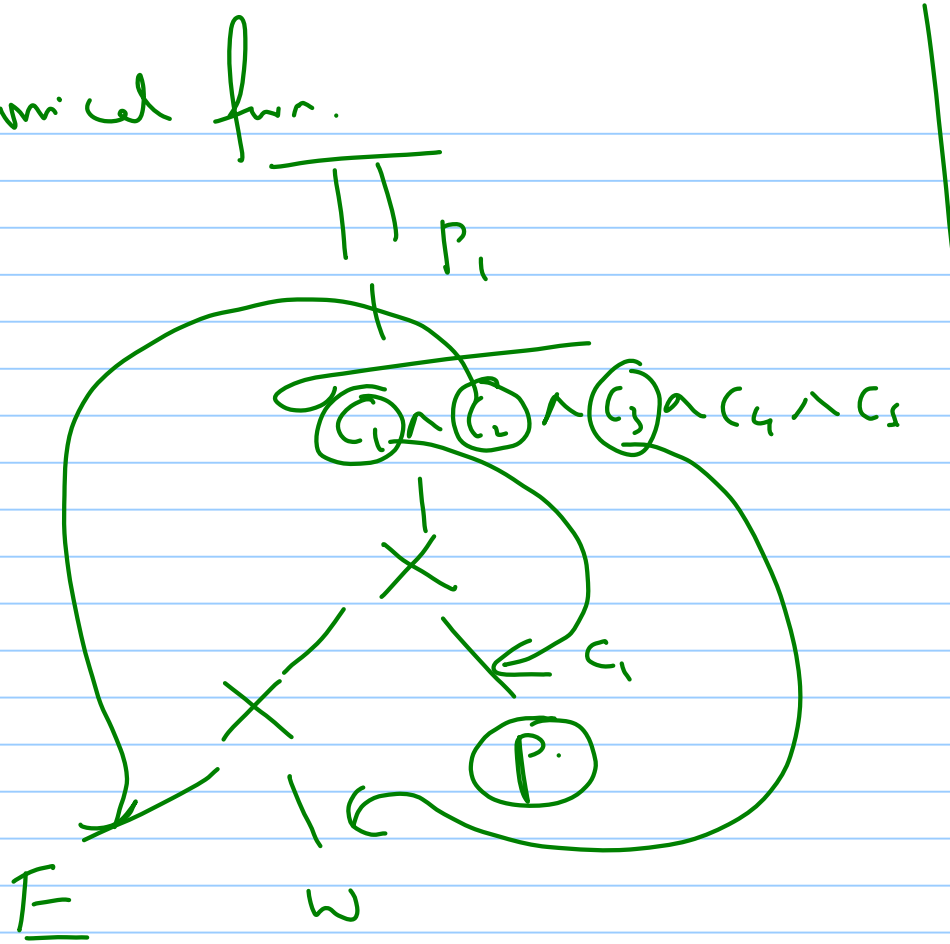
E(SSN, BDate, LName, ...) WorksON(ESSN, PNO, Hours) Project(PNumber, PName, ...)

```
SELECT LNAME
FROM   E, W, P
WHERE  PName = 'Aquarius'
  AND  BDate > 'Dec-31-1997'
  AND  Hours >= 40
  AND  ESSN = SSN
  AND  PNO = PNUMBER;
```

Syntactically correct



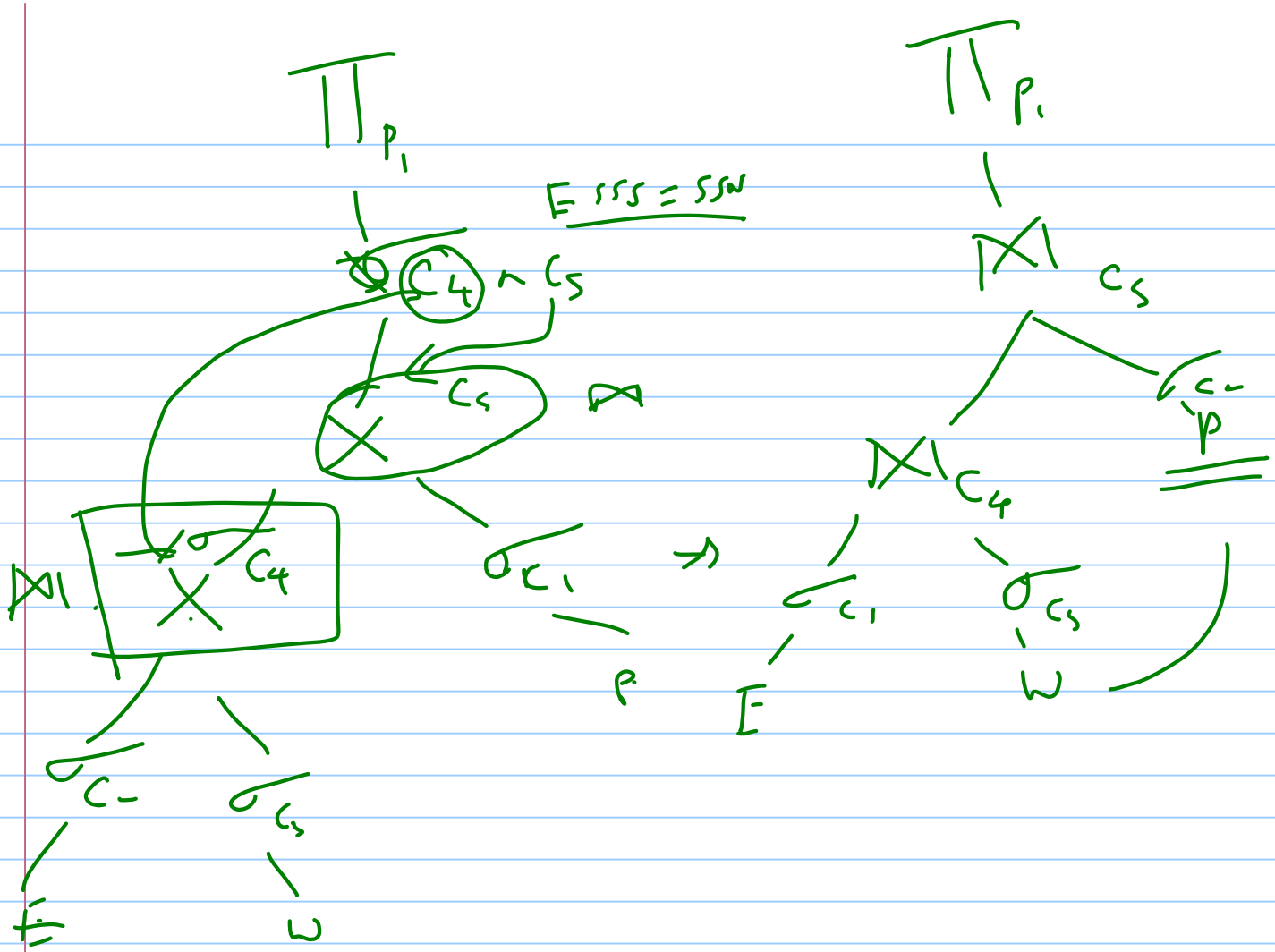
Canonical form.



Algebra

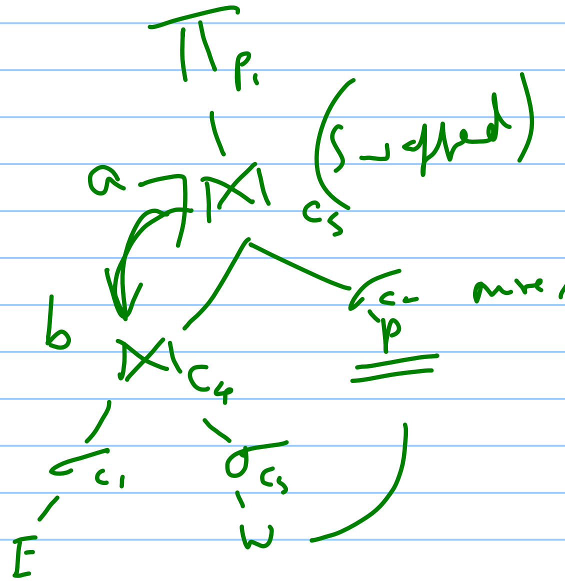
Starting Point

SELECT LNAME (P1)
FROM E, W, P.
C1 WHERE PNAME = 'Aquarius'
C2 AND BDATE > 'DEC-31-1957'
C3 AND HOURS > 40
C4 AND E.SSN = SSN
C5 AND PNO = PNUMBER.



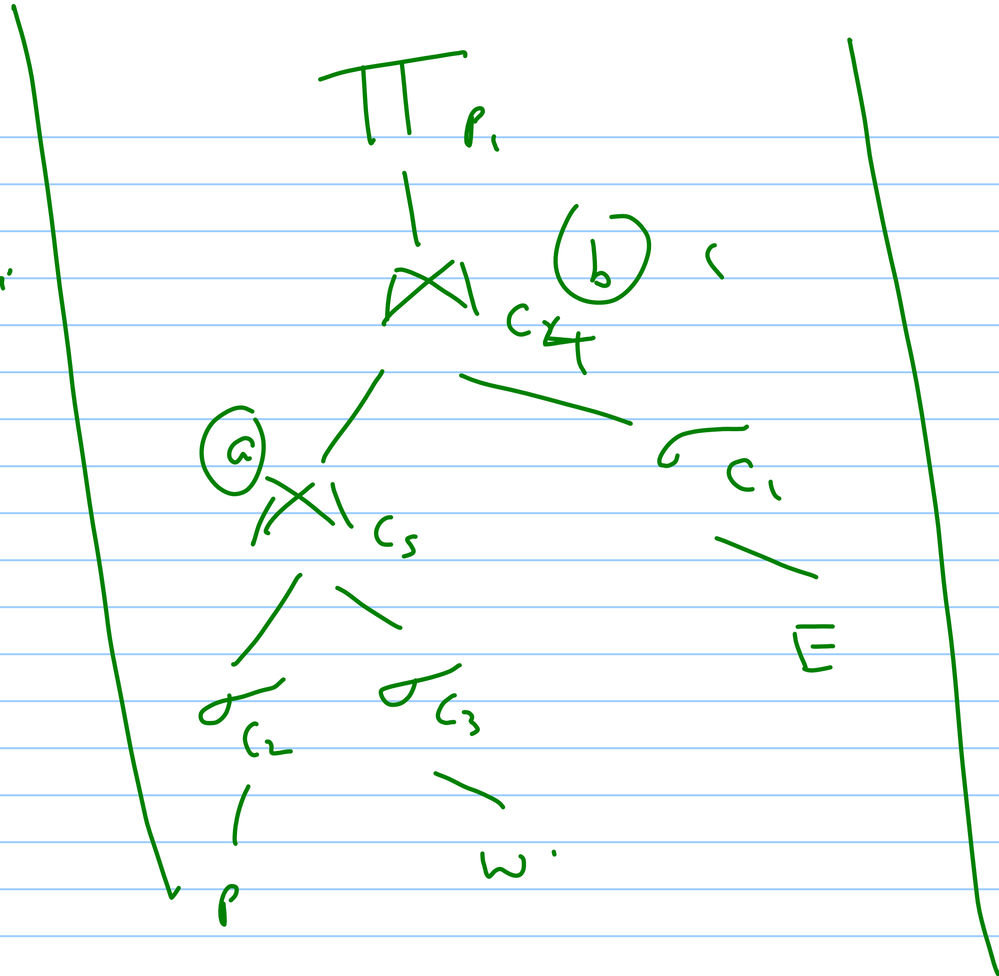

```
SELECT L.NAME P.
FROM E, W, P.
C. WHERE  P.NAME = 'ACHUMA'
C-  AND   B.DATE > 'DEC-31-1997'
C5  AND   HOURS  > 40
C4  AND   E.SSN = SSN
C1  AND   PNO = PNUMBER.
```

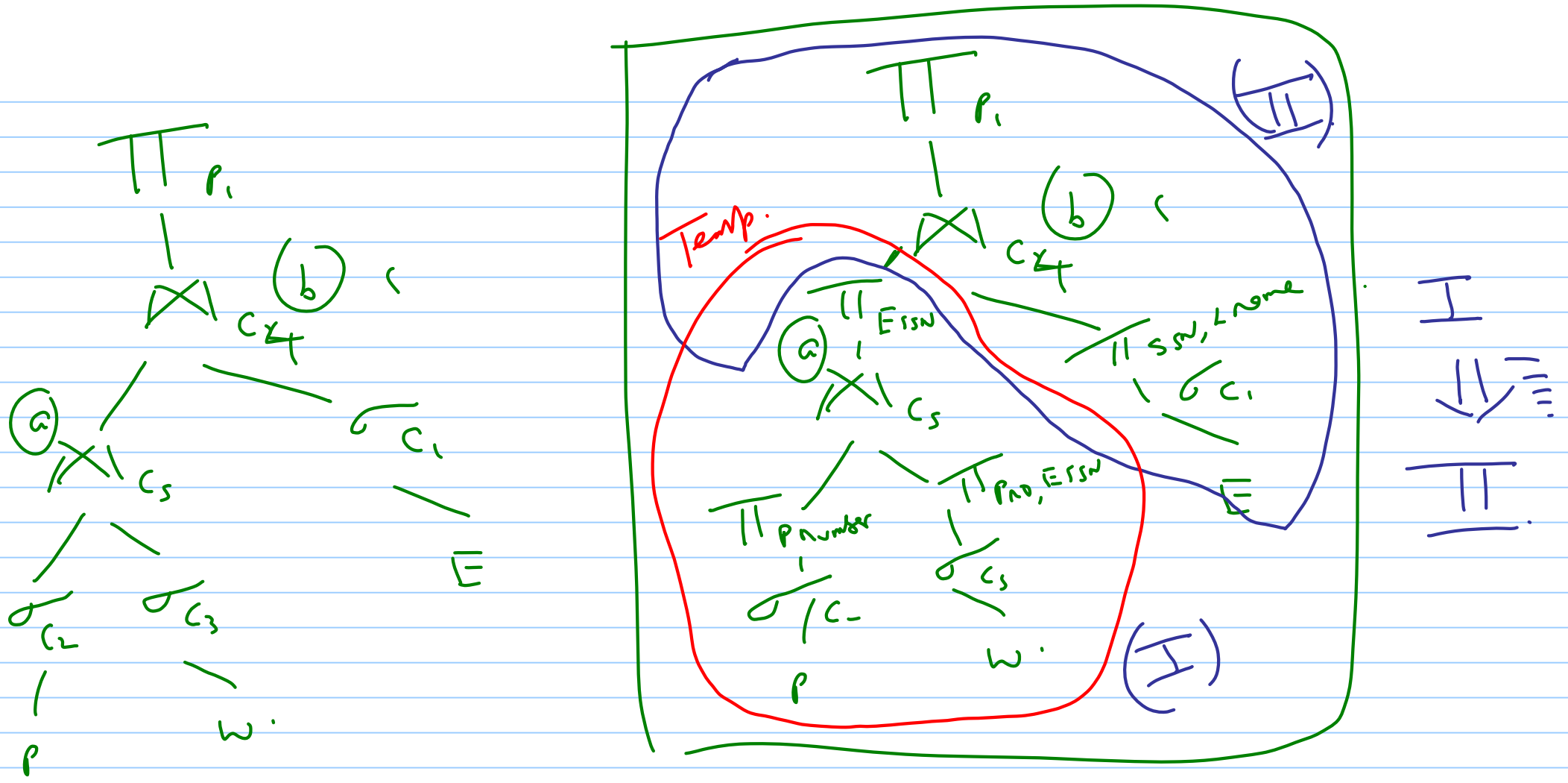
more restrictive

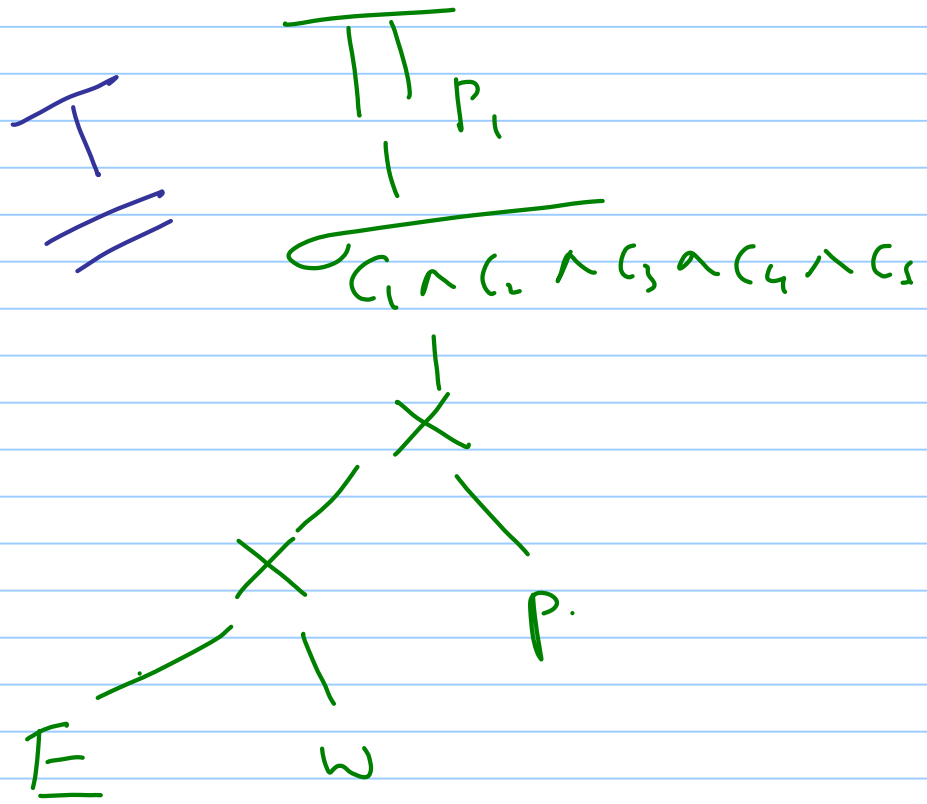


more restrictive

SELECT LNAME P1
FROM E, W, P.
C1 WHERE Pname = 'Aquarius'
C2 AND BDate > 'Dec-31-1957'
C3 AND Hours > 40
C4 AND E.SSN = SSN
C5 AND PNO = PNUMBER.




Temp.
I
II



while (no more transf.) {

transf^(rule, T, T')

T = T'

}

$\Pi(R \bowtie S) \supseteq \Pi(S \bowtie R)$

$\subseteq$

Rules

rule1 $R \bowtie S = S \bowtie R$ [C]

rule2 $(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$ [A]

$(E \bowtie W) \bowtie P = E \bowtie (W \bowtie P)$

$\Rightarrow$ rule3 $R \times S = S \times R$

rule4 $(R \times S) \times T = R \times (S \times T)$

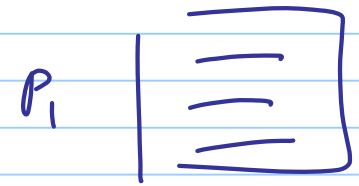
$R \cup S = S \cup R$

$(R \cup S) \cup T = R \cup (S \cup T)$

1st bunch of rules

$$\sigma_{p_1 \wedge p_2}(R) \equiv \sigma_{p_1}(\sigma_{p_2}(R))$$

$p_1 \wedge p_2 \wedge \dots \wedge p_n$



$$\sigma_{p_1 \vee p_2}(R) \equiv$$

$$\sigma_{p_1}(R) \cup \sigma_{p_2}(R)$$

$p_1 \vee p_2 \vee \dots \vee p_n$

$p_1 \vee (p_2 \wedge p_3) \vee (p_4 \wedge p_5)$

$R = \{ \underset{(a\ b\ c\ d\ e\ f)}{a, a, b, b, b, c} \}$ | Rule sets 4 | Bags

$S = \{ \underset{(g,h,i,j,k)}{b, b, c, c, d} \}$

$R \cup S = \{ \underset{(a,1)(a,2)(b,1)(b,1)(b,3)}{a, a, b, b, b, b, c, c, c, d} \}$ SUM

$= \{ a, a, b, b, b, c, c, d \}$ MAX

$R \subseteq$ (2, 3, 2, 1)

Select Distinct

Rules — $X \subseteq Attr(R)$

— $Y \subseteq Attr(R)$

$XY = X \cup Y$.

$\Pi_{XY}(R) = \Pi_X[\Pi_Y(R)]$

$\Pi_{\emptyset}(R)$

$\Pi_{(Emp, Emp, Pno)}(E) = \Pi_{Emp}\Pi_{Emp}\Pi_{Pno}(E)$

Note Title 2/13/2014

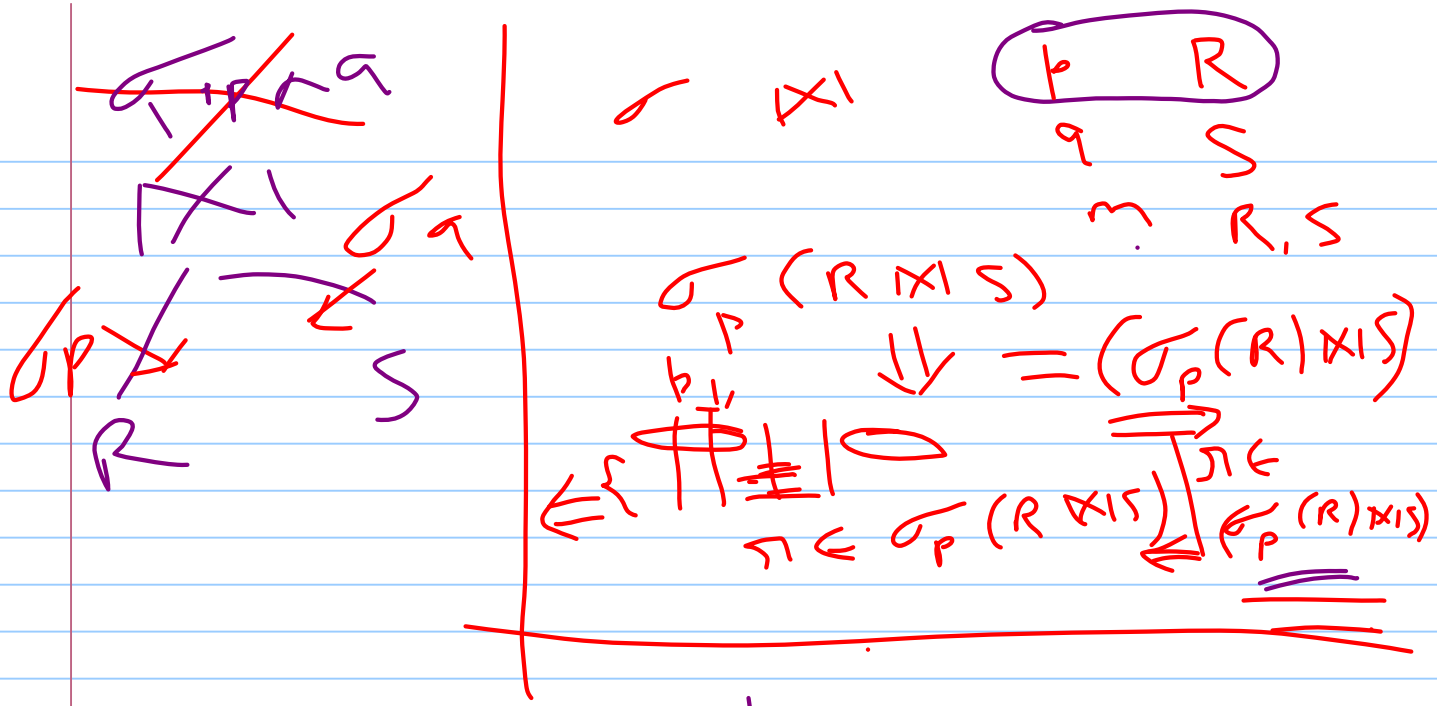
σ | π | ⋈ | GB | SnA | operators

E

HJ

(C I)

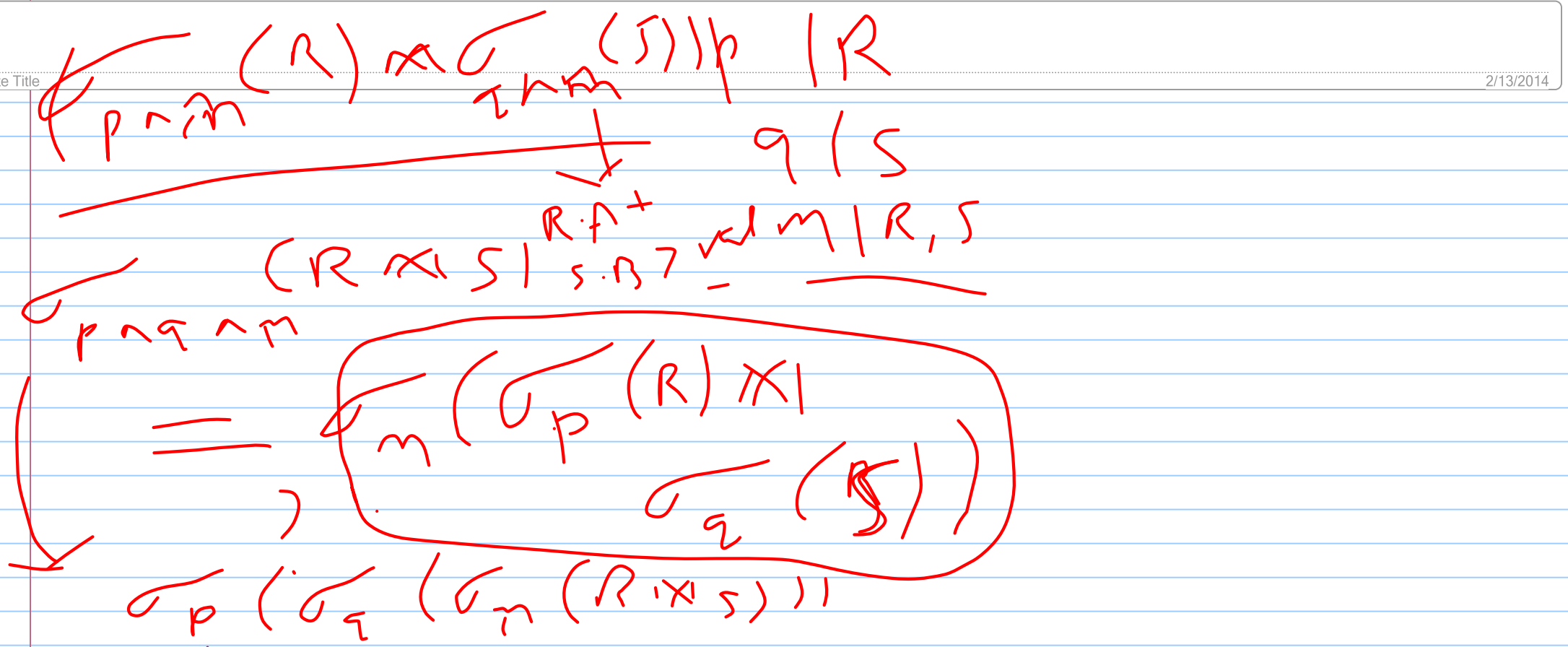
Exec (pred, R)



$$\sigma_{p \wedge q}(R \bowtie S) = \sigma_p(R) \bowtie \sigma_q(S)$$

Note Title 2/13/2014

Note Title 2/13/2014



Π, σ

$X \quad R$

$2 \quad p \mid R$

$\Pi_x(\sigma_p(R))$

$= \Pi_u(\sigma_p(\Pi_{x2}(R)))$ | $x \mid R$, $y \mid S$, $2 \mid R \cap S$?

$\Pi_{xy}(R \times S) =$

$\Pi_{xy}(\Pi_{x_2}(R) \times \Pi_{y_2}(S))$

$$\Pi_{xy}(\sigma_{P}(R \bowtie S)) = \Pi_{xy}\left[\sigma_{P}\left(\Pi_{x,z'}(R) \bowtie \Pi_{y,z'}(S)\right)\right]$$

x	R
y	S
z	R ∩ S

$z' = z \cup$ (attributes in P)

$$\sigma_p(R \cup S) = \sigma_p(R) \cup \sigma_p(S)$$

$$\sigma_p(R - S) = \sigma_p(R) - \sigma_p(S)$$

R(A, B, C, D, E) $X = \{E\}$

p A = 3 $\wedge$ B = "cat"

$\Pi_X(\sigma_p(R))$ vs. (I)

$\Pi_E(\sigma_p(\Pi_{ABE}(R)))$ (II)

(I) or (II)

Transformation

homomorphism

$$\left[\bigcup_{A=1} (R) \times S\right] = \bigcup_{A=1} \{(R) \times S\}.$$

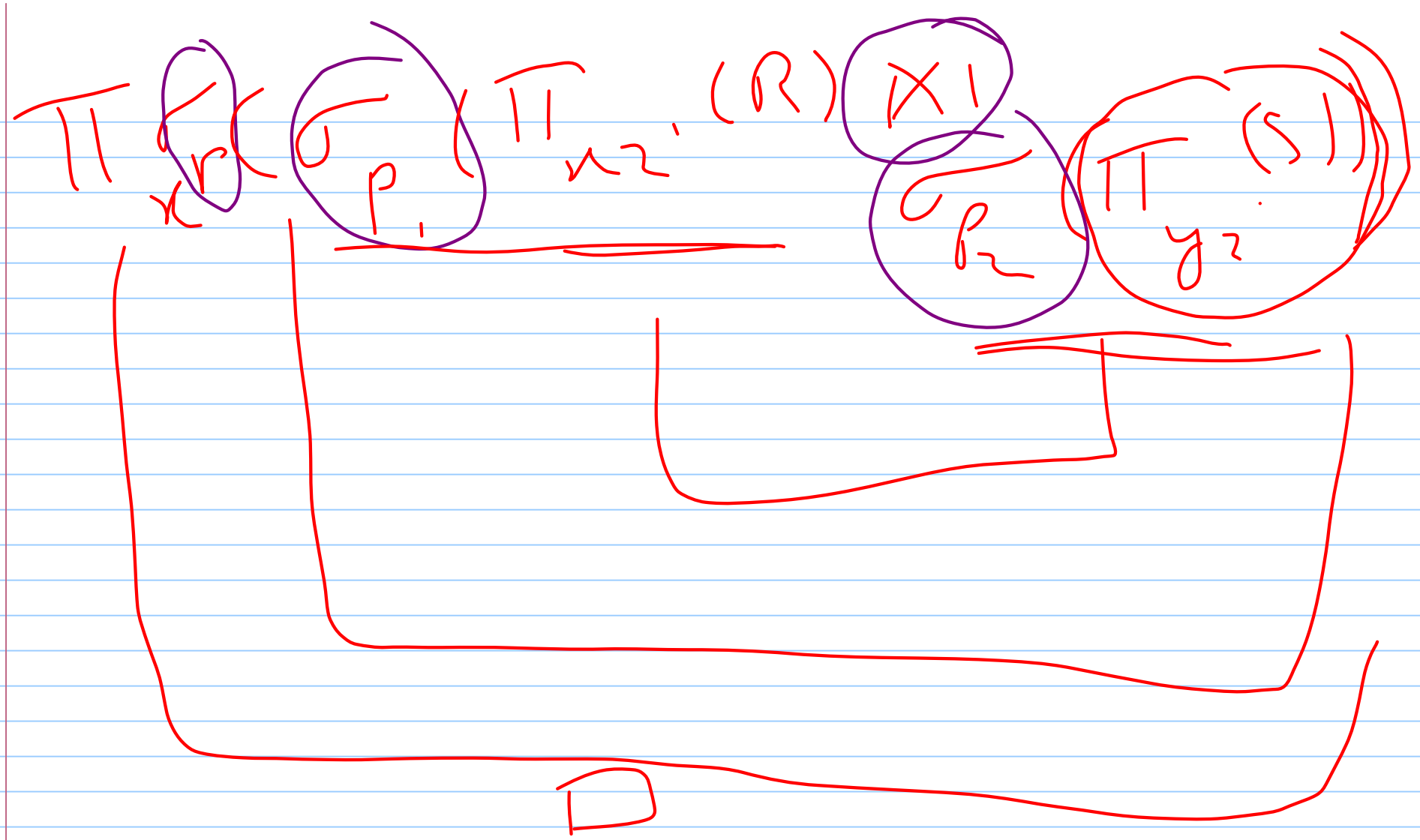
Triple is only half

$$\mathbb{Q} \quad \mathbb{O}$$

$$\Downarrow$$

Exact Consts $(\mathbb{R} \times \mathbb{S})_{\mathbb{H}}$

$$\mathbb{R} \times (\mathbb{S} \cdot \times \mathbb{T})$$


(R)
(S)

Estimating sizes of Results

$T(R)$ = # of tuples in R

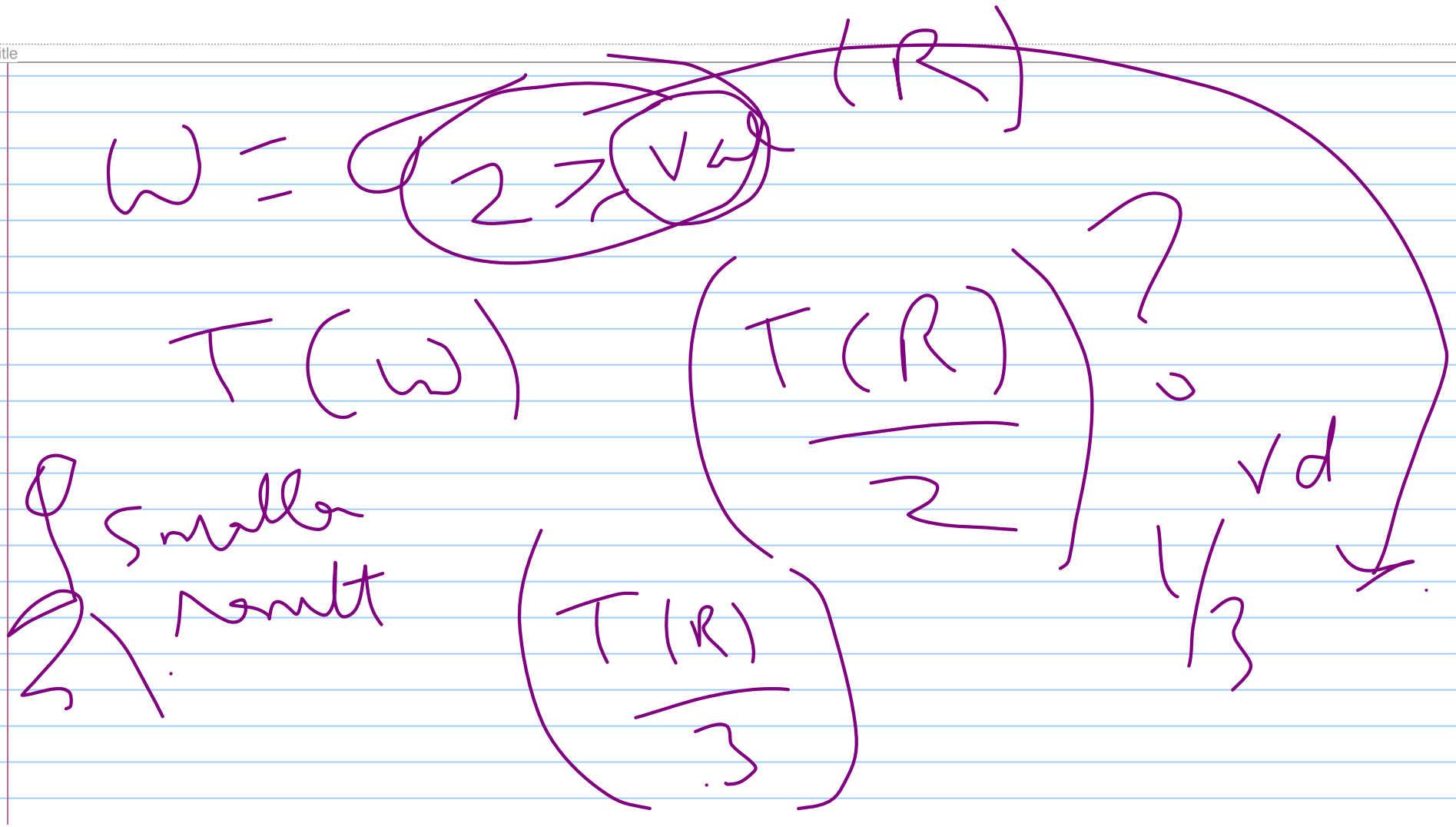
$S(R)$ = # bytes

$B(R)$ = blocks

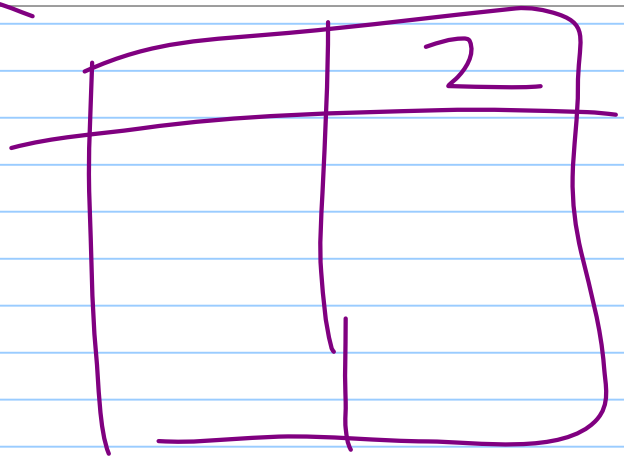
$V(R, A)$ = # of distinct values of A in R

(Select distinct A From R)

Note Title 2/13/2014



R



$M_{1,1} = 1$ $\quad V(R,2) = 10$

$M_{4,2} = 20$

$W = \sigma_{27,15}(R)$ $\quad T(w) = f \cdot T(R)$

$f = \frac{20-15+1}{20-1+1} = \frac{6}{20} = \frac{3}{10}\,(T(R))$

Note Title 2/13/2014

$$\prod^{A} V(z,n) \left(\frac{f}{+}\right) V(z,R)$$

$$T(w) = f \times V(z,R) \times \overline{T(R)}$$

$$= f \times \overline{T(R)} \quad \because V(z,R)$$

$W = R1 \bowtie R2 = R1 \times R2$

$X \quad R1 \quad Y \quad R2$

$X \cap Y = \phi$

$T(W) = T(R1) \times T(R2)$

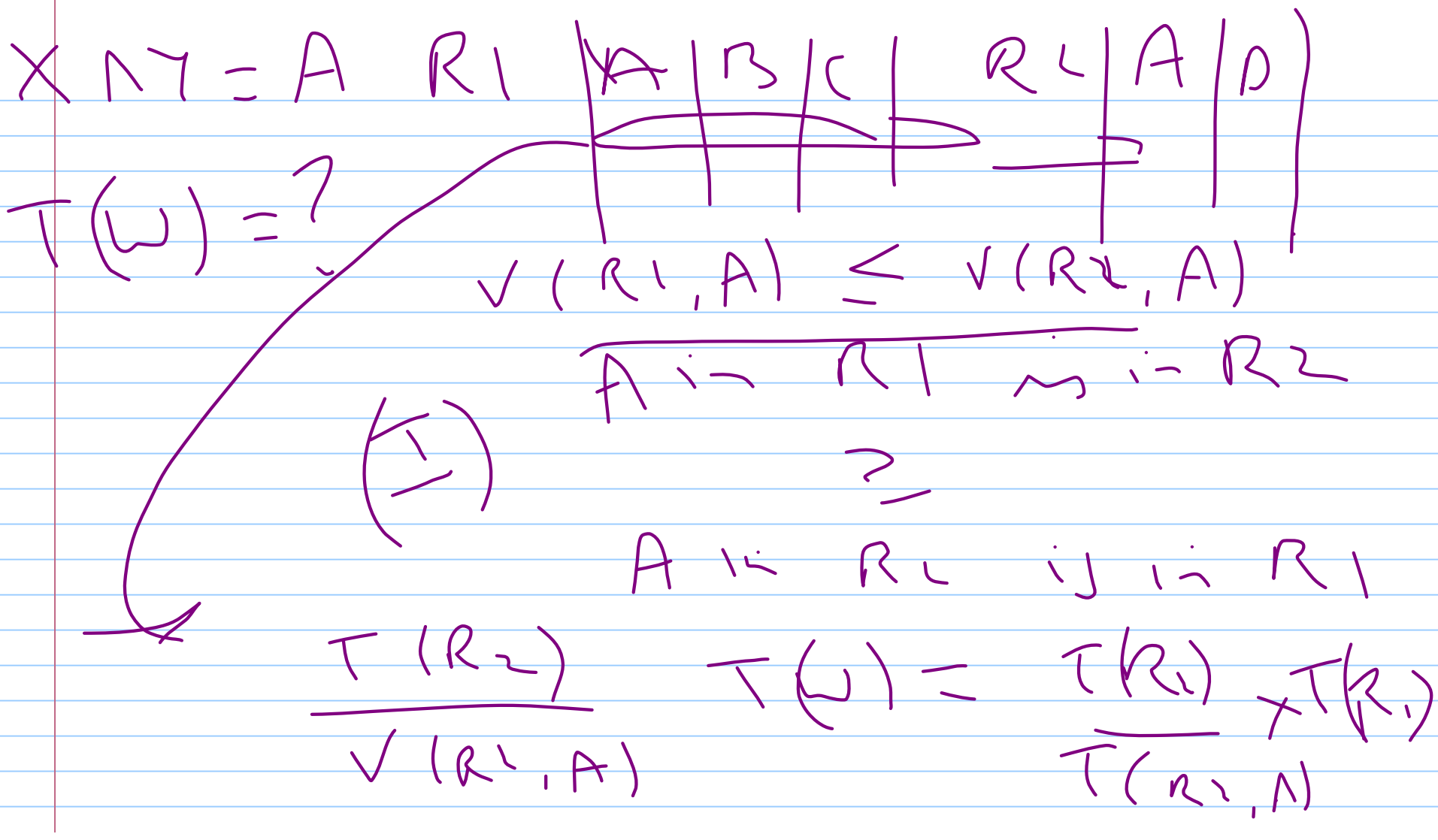
$X \cap Y = A$

R1	A	B	C

R2	A	D

$T(W) = ?$

$V(R1, A)$ $V(R2, A)$


X ∩ Y = A
R1
A B C
R2 A D
T(ω) = ?
V(R1, A) ≤ V(R2, A)
A in R1 ω in R2
(1)
≥
A in R2 ij in R1
T(R2) / V(R1, A)
T(ω) = T(R1) / T(R1, A) × T(R1)

$$T(w) = T(R_1) \times T(R_2)$$

$$T(w) = \frac{T(R_1) \cdot T(R_2)}{\max(V(R_1, A), V(R_2, A))}$$

$$T(w) = \frac{T(R_1) \cdot T(R_2)}{\max(D_m(R_1, A), D_m(R_2, A))}$$

$$S(R_1) + S(R_2) - S(A)$$

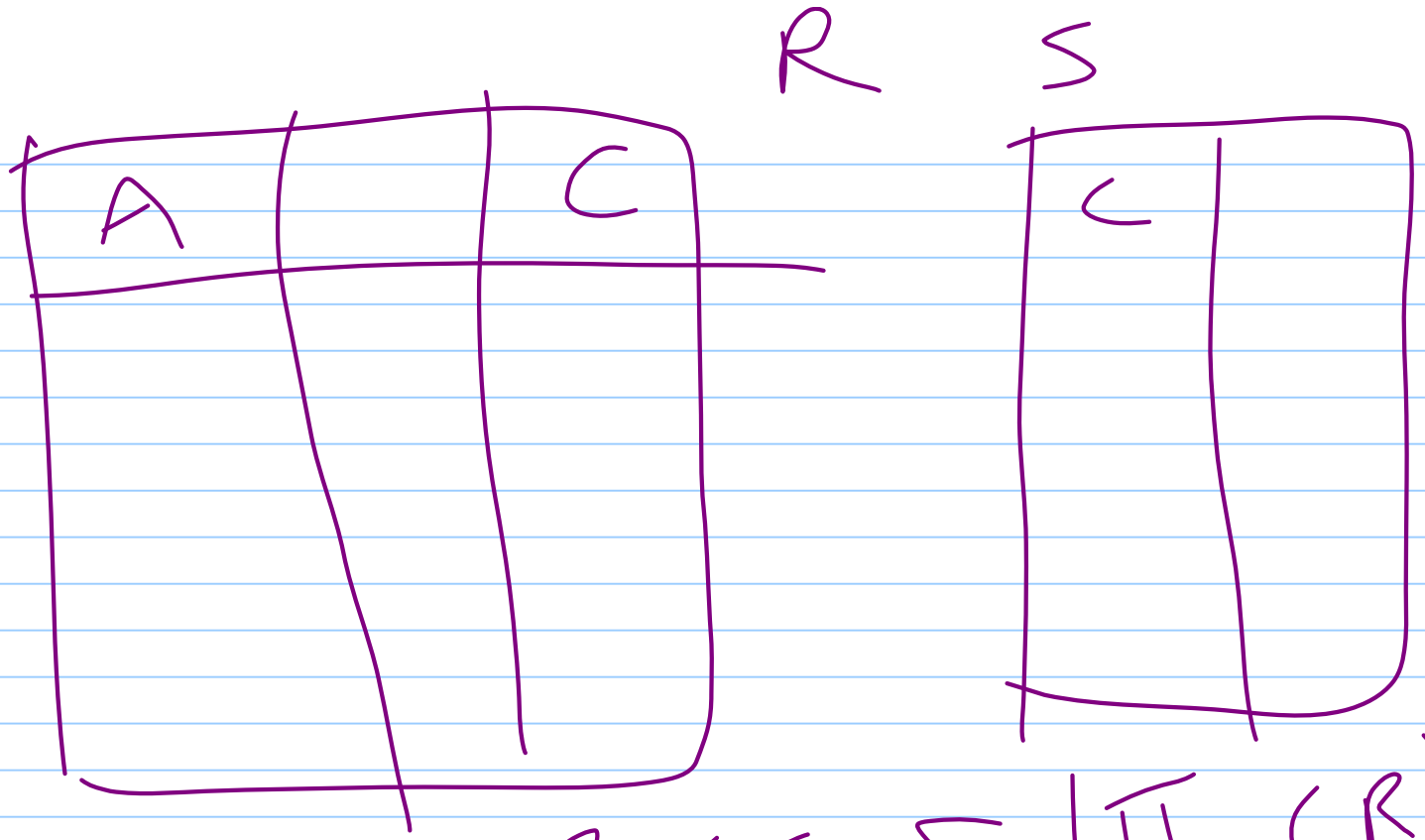
$$V(R,Y) \subseteq V(S,Y)$$

R S Y

Fix

$\subseteq$ [pic.

(content



$A \subseteq$ $R \bowtie S$ $\Pi_A(R)$

$\Pi_A(R \bowtie S)$ $\subseteq$

$$W = \sigma_{A=a}^{(R1) \bowtie R_L} B$$

$$\left(\frac{T(R1)}{V(R,A)} \right) \times \frac{T(R_2)}{\max(V(R,D), V(R_L, B))}$$

R(a, b)

T(R) = 1000

V(R, b) = 26

S(b, c)

T(S) = 2000

V(S, b) = 50

V(S, c) = 100

U(c, d)

T(U) = 500

V(U, d) = 50

(RMS) ⋈ U

$= \frac{1000 \times 2000}{40000}$

40000

$\frac{40000 \times 8000}{800}$

400,000

R ⋈ (S ⋈ U)

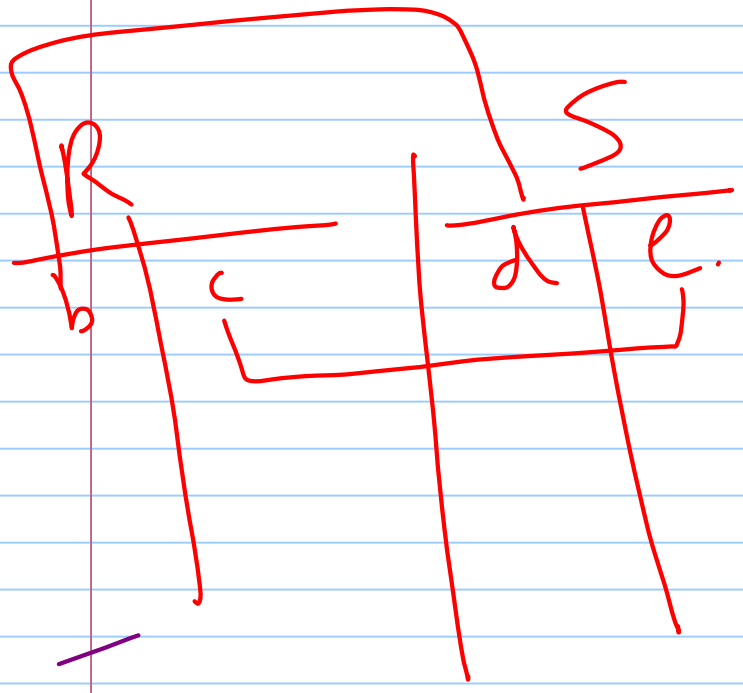
R(a, b, c) S(d, e, f)

T(R) = 1000 | T(S) = 2000

V(R, b) = 20 | V(S, d) = 50

V(R, c) = 100 | V(S, e) = 50

$T(R) \times T(S) / \max(V(R,b), V(S,d)) \cdot \min(V(R,c), V(S,e))$



W = R(a, b, c) ⋈ S(d, e, f)

I(W)

$R.b = S.d$ and $R.c = S.e$

$= \frac{1000 \times 2000}{50 \times 100} = 400$

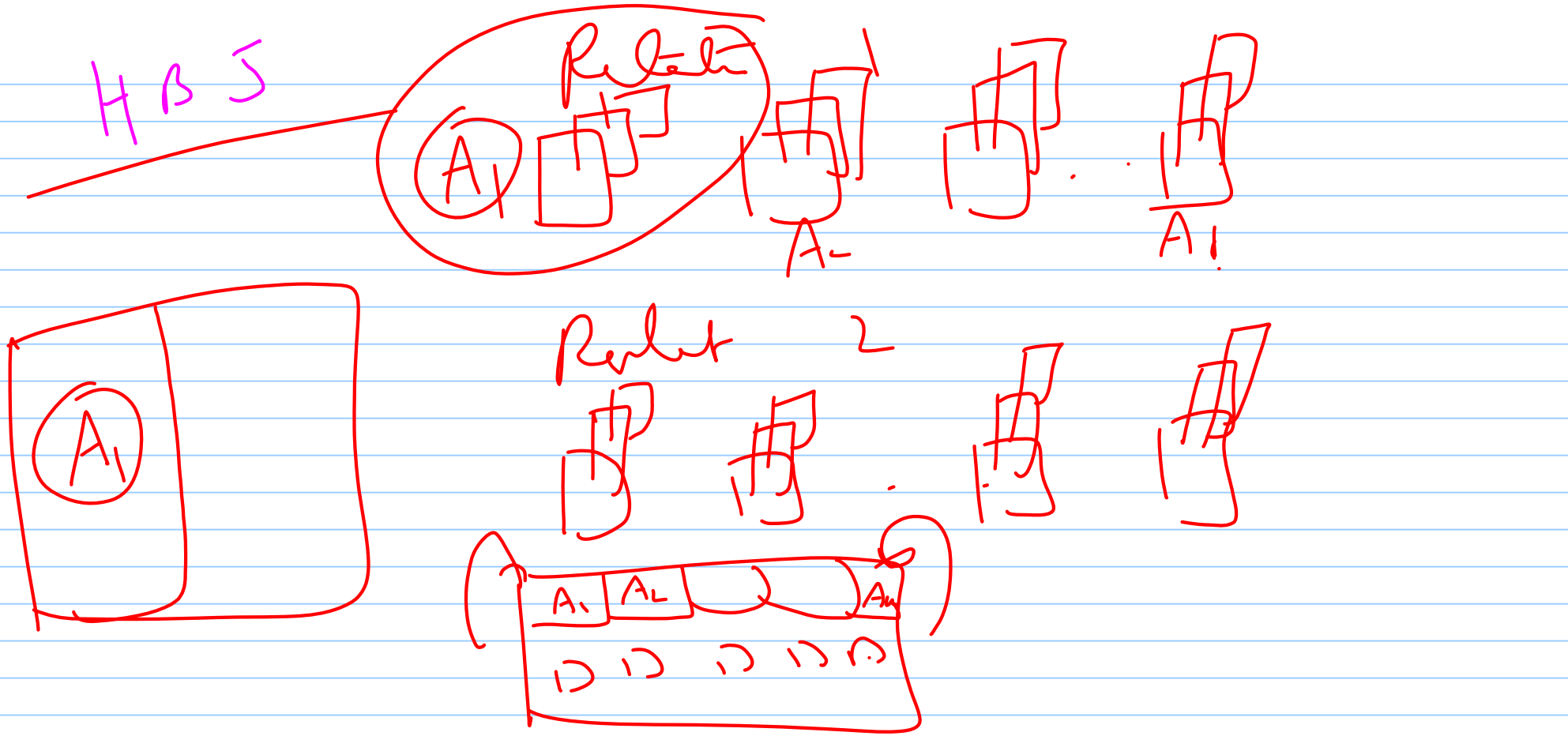
	EMPTY			
b = c	Y	Y	N	N
d = e	Y	N	Y	N

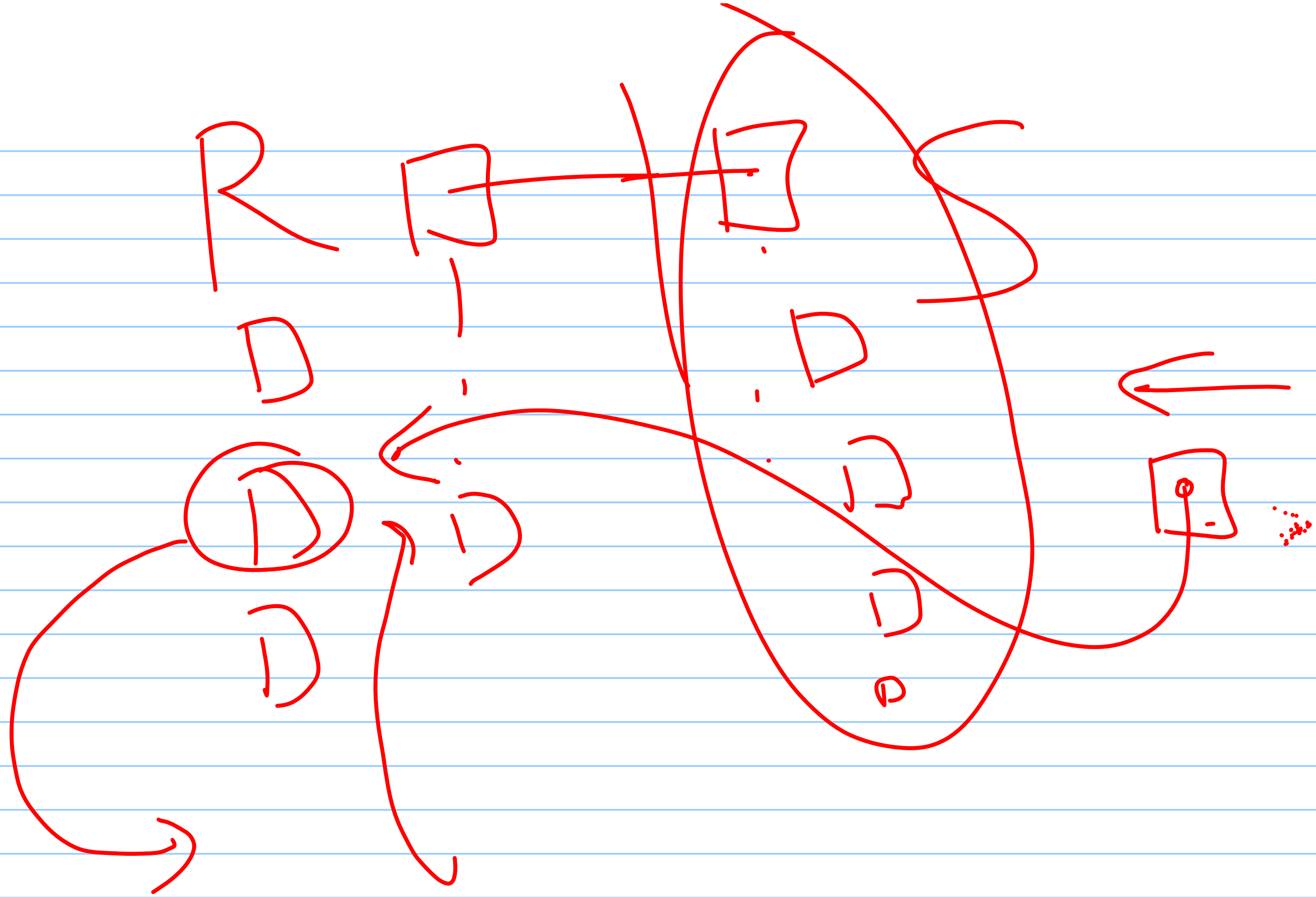
Saving Disk I/O in HBJ

How optimal is our Operating choice?

Statistics / Relation / Attributes (...

→ Best plan / Operation / ...


HBJ
A1
A2
A1
A1
A2
Am


R
S

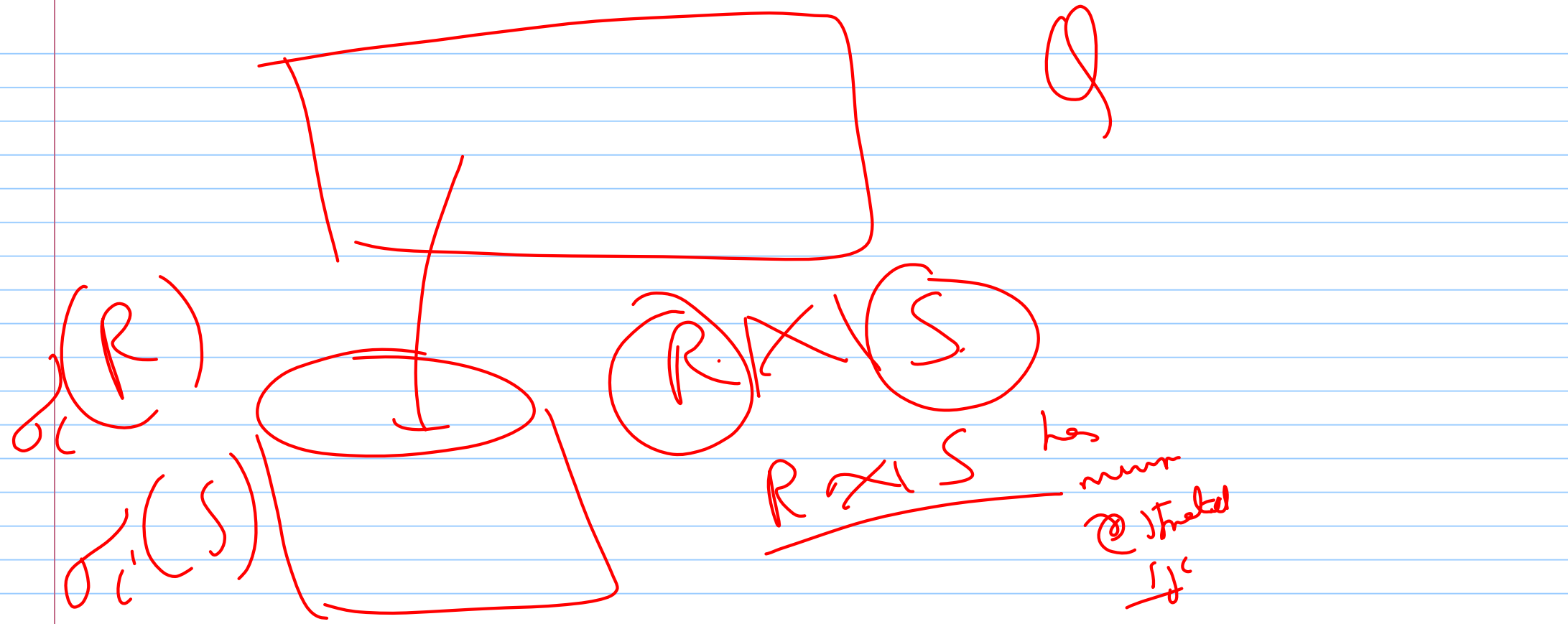
How upload is youtube

↓

Quality → ID IDD

→

→ buffering effect


Q
R × S
R × S

R | A | B | C | A | K

S | D | E | F |

$\frac{1000}{1}$ | 500 | R

$\frac{1000}{\sqrt{2}}$ | 1000

| 999 | 0

$\frac{1000}{1.414}$ | 750 | 15

| 800 | 16

| 100 | 4

| 0

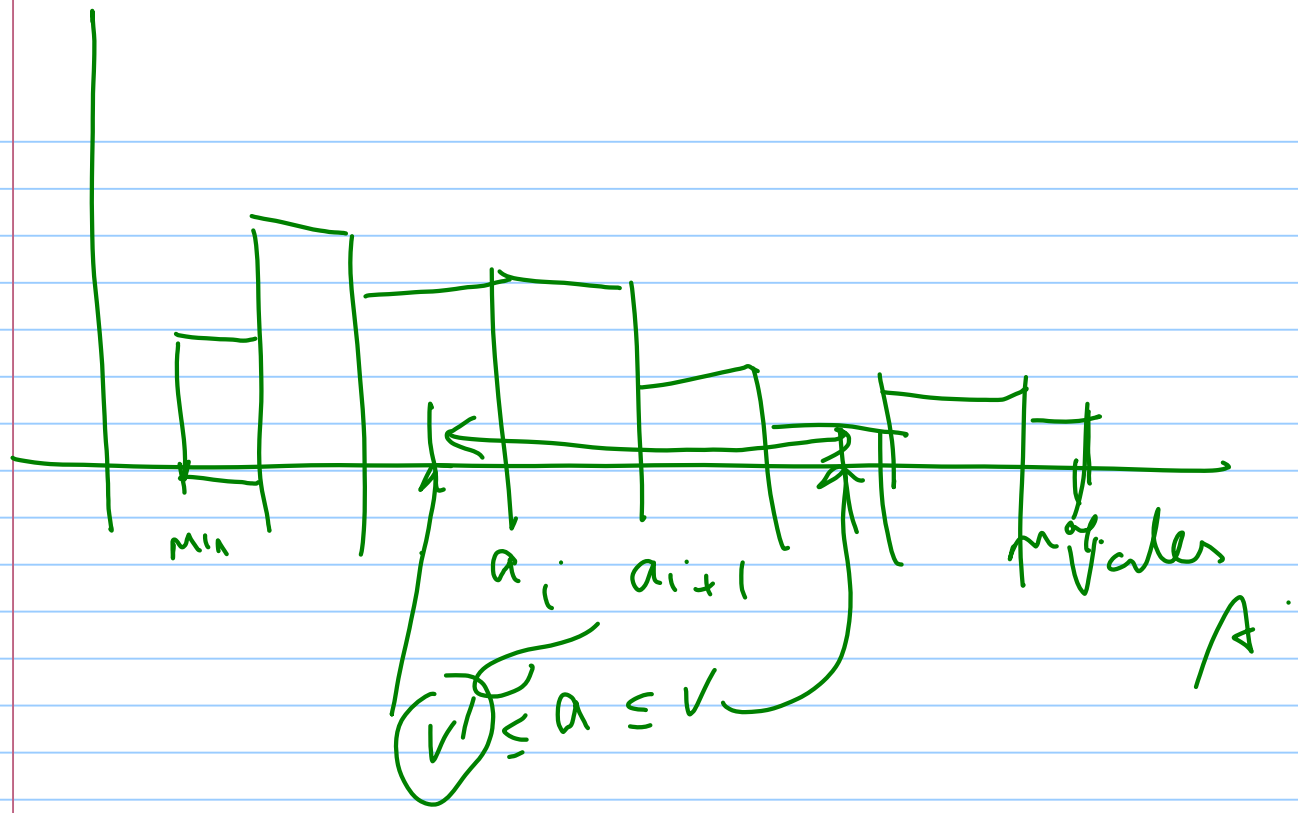
~700

$\frac{1000}{\sqrt{3}}$ 500

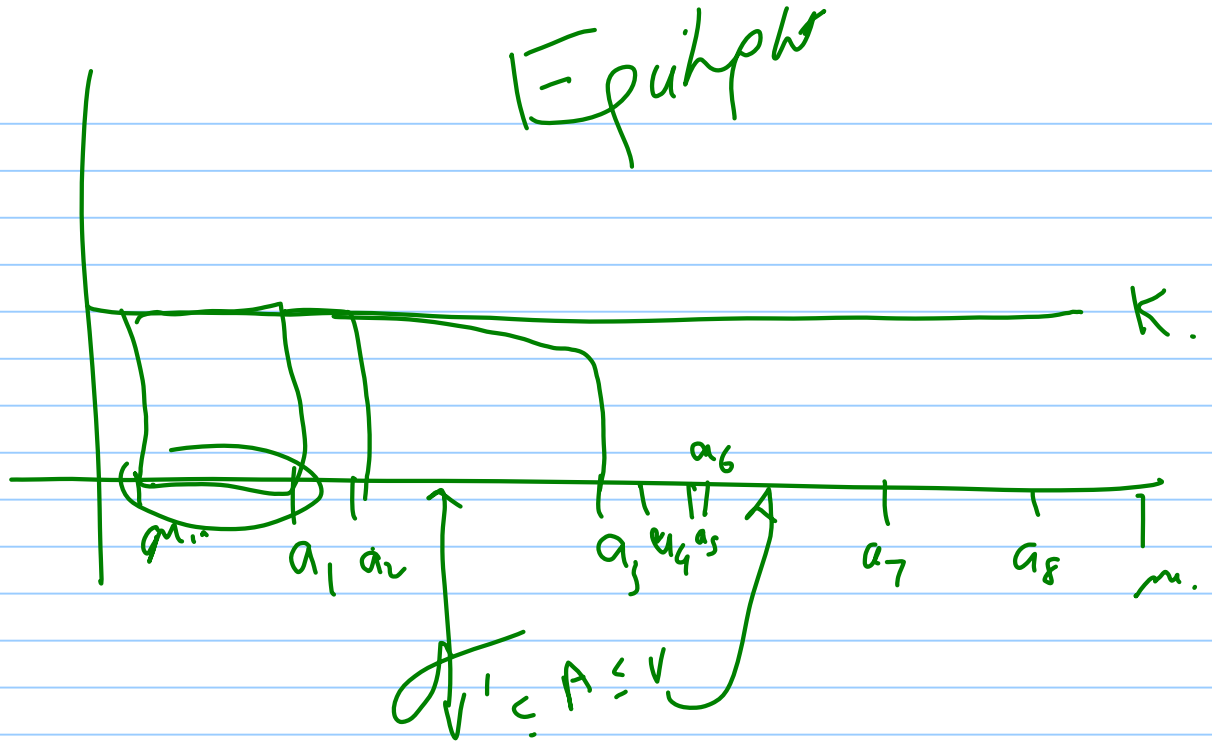
E → D

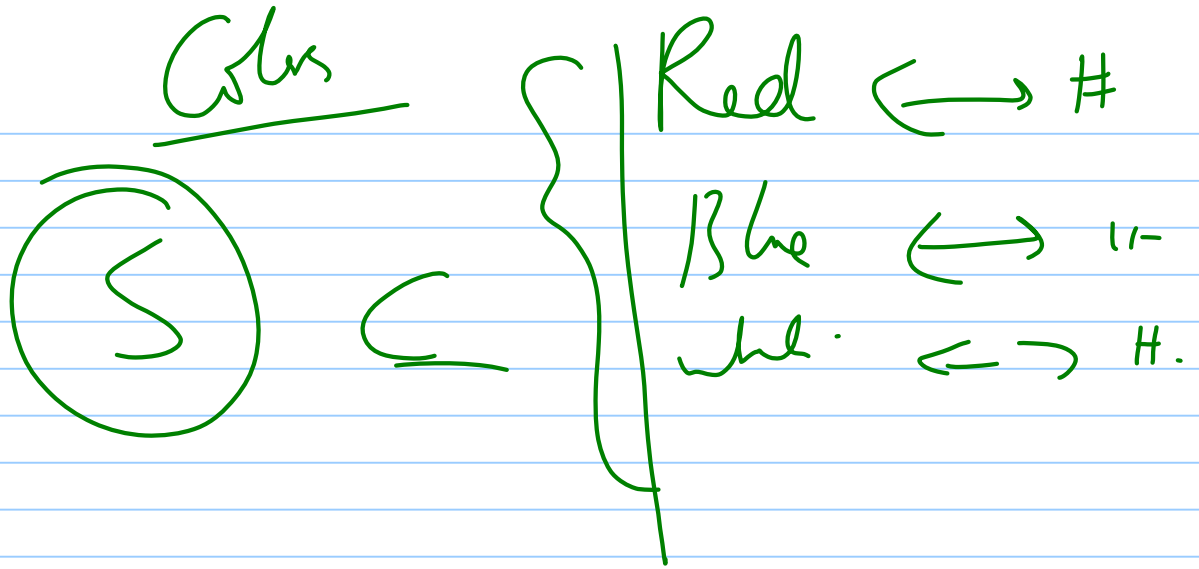
Страна

Zipf


min
a_i
a_{i+1}
$v \leq a \leq v$

Equilibre




Colors
S
Red ⟷ #
Blue ⟷ 11-
Wht. ⟷ H.

Selectivity factor products

f·n f·T(R)

Predicate	Uniform Domain	Uniform Actual	Zipf	Histogram
A = val	$f = \frac{1}{V(Dom(A))}$	$\frac{1}{V(R,A)}$	$\frac{n}{\sqrt{n}}$	$\frac{T(\sigma_{A=val}(R))}{T(R)} \cdot n$
A > val ≤ val	1/2	1/3	$A \le val$: $\sum_{j \le i} \frac{T(\sigma_{A=v_j}(R))}{T(R)}$	(Actual)
$C_1 \wedge C_2$	$\sigma_{C_1}(R) \times \sigma_{C_2}(R) \Rightarrow$		$sf_{C_1} \times$	$sf_{C_2} \times T(R)$
$C_1 \vee C_2$	$\sigma_{C_1}(R) \cup \sigma_{C_2}(R)$		$= (sf_1 + sf_2 - sf_1 \times sf_2)$	$T(R)$

T

Containment principle.

$R \subseteq_Y S$

$V(R, Y) \subseteq V(S, Y)$ ∎

predicate symbols A

k $\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{bmatrix}$

$(R \bowtie_C S)$

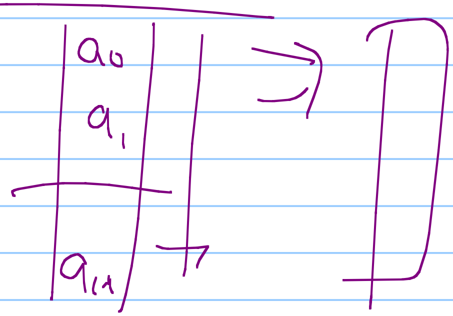
$\Downarrow$

$(a_0, a_1, \ldots a_{k-1})$

(2) X PK - FK

in the screen

$A \leq a_p$



$$(R \bowtie_A S) = \frac{T(R) \times T(S)}{\max(V(R,A), V(S,A))}$$

$R \bowtie S \Rightarrow T(R) \times T(S)$

$<< T(R) \times T(S)$ — drastic

$$\text{JSF} = \frac{1}{\max(V(R,A), V(S,A))}$$

⋮

$R_1 \bowtie R_2 \bowtie \cdots \bowtie R_n$

$R_i \bowtie R_j$ JSF$_{(R_i, R_j)}$

JSF $R_p \bowtie R_q$ · ≤

no comm attribute

$\sigma_{C_1}(R_1) \bowtie \sigma_{C_2}(R_2) \bowtie \cdots \bowtie \sigma_{C_K}(R_K)$

$\sigma_{C_i}(R_i) \bowtie \sigma_{C_j}(R_j)$

$= T(\sigma_{C_i}(R_i) \times T(\sigma_{C_j}(R_j))$

$\max(V(\sigma_{C_i}(R),A), V(\sigma_{C_j}(R_j),A))$

$= \frac{S_{C_i} \times T(R_i) \quad S_{C_j} \times T(R_j)}{\max(V(S_{C_i} \times T(R_i),A), V(S_{C_j} \times T(R_j),A))}$

$C_i = (B = val)$

$C_j = (C = val)$

$= \frac{1}{T(R_i \bowtie B)} \times (T(R_i)) \times \frac{1}{T(R_j,C)} \times (T(R_j))$

$\max\left(V\left(\frac{1}{T(R_i,B)} \times T(R_i), A\right), V\left(\frac{1}{T(R_j,C)} \times T(R_j), A\right)\right)$

$\frac{1}{V(R_i,B)} (T(R_i) \times \frac{1}{V(R_j,C)} \times (T(R_j))$

$\frac{1}{V(R_i,B) \times V(R_j,C) \times \max(V(R_i,A), V(R_j,C))} \max((V(R_i,A), V(R_j,A))$

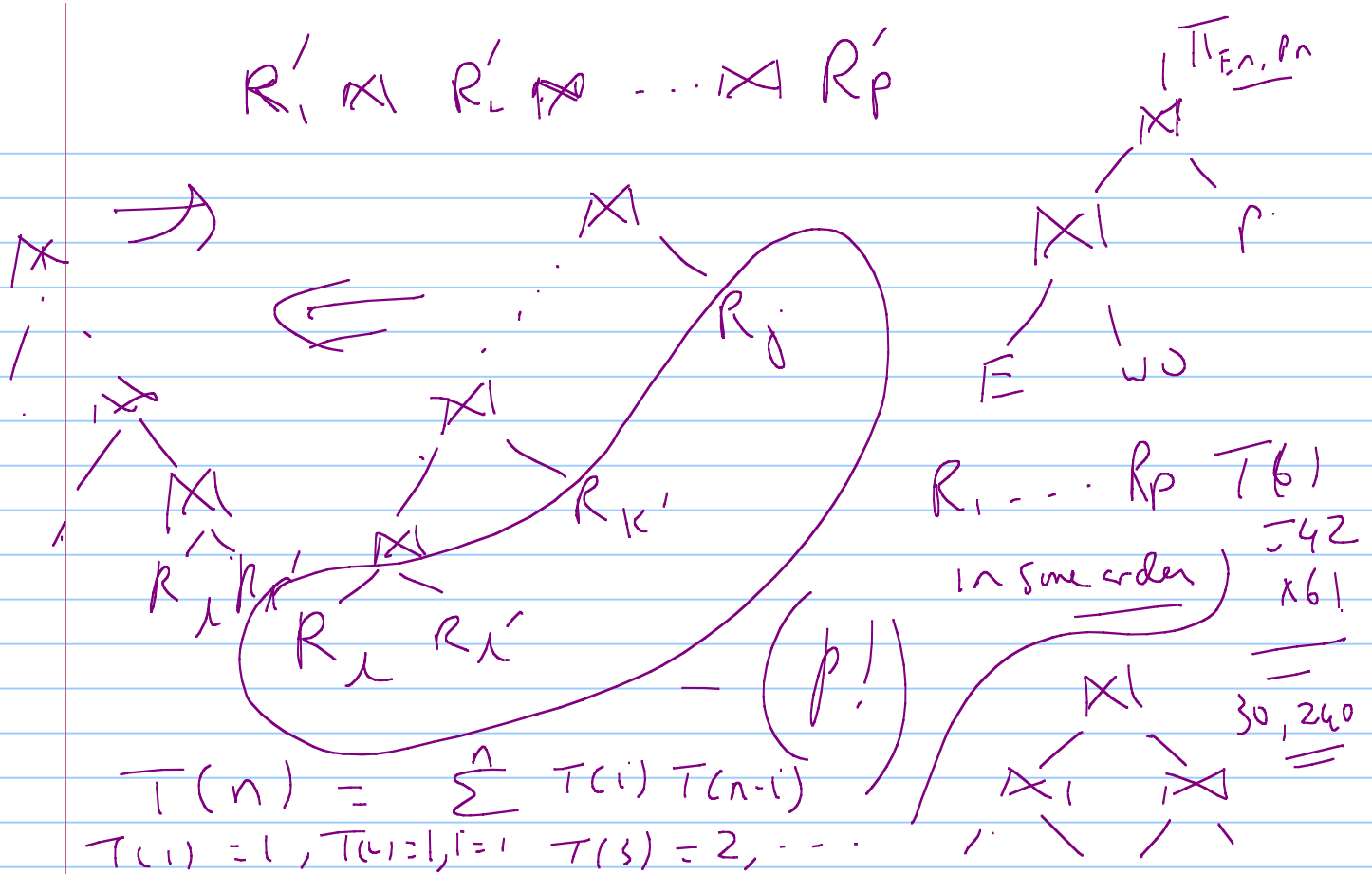
Select $A_1 \cdots A_k$

From $R_1, R_2 \cdots R_p$ $p \geq 5$

Where $c_1 \cdots c_l$

$k \geq p-4$

(Pseudo the order of joins)



$$T(n) = \sum^{n} T(i)\,T(n-i)$$

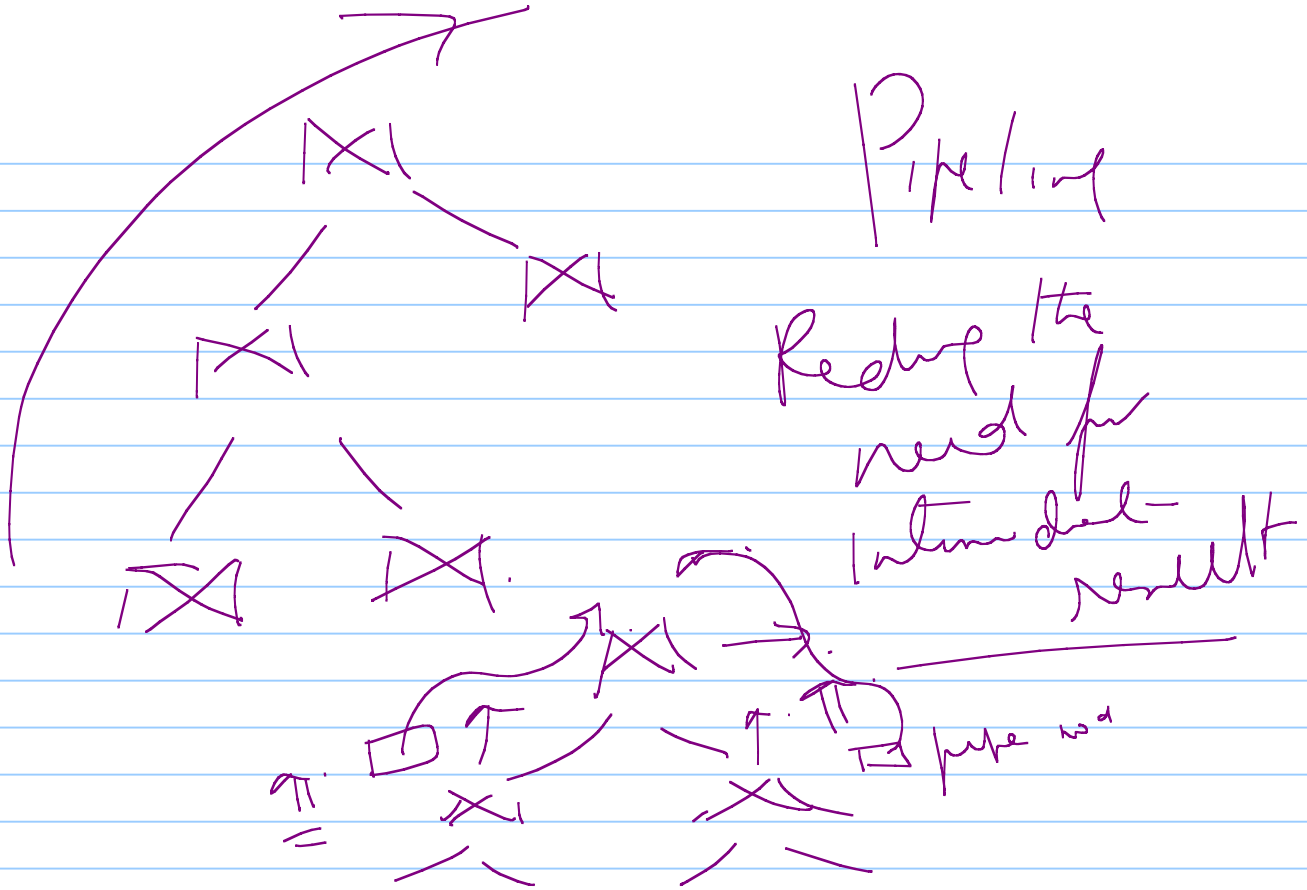
$T(1) = 1,\ T(2) = 1,\ T(3) = 2, \cdots$

$R_1 \cdots R_p$ $T(6)$

in some order

$= 42 \times 6!$

$= 30,240$


Pipeline
Reduce the need for intermediate result

E Emp (Name, DNO, JOB, Sal)

D Dept (DNO, Name, Loc)

J Job (Job, JTitle)

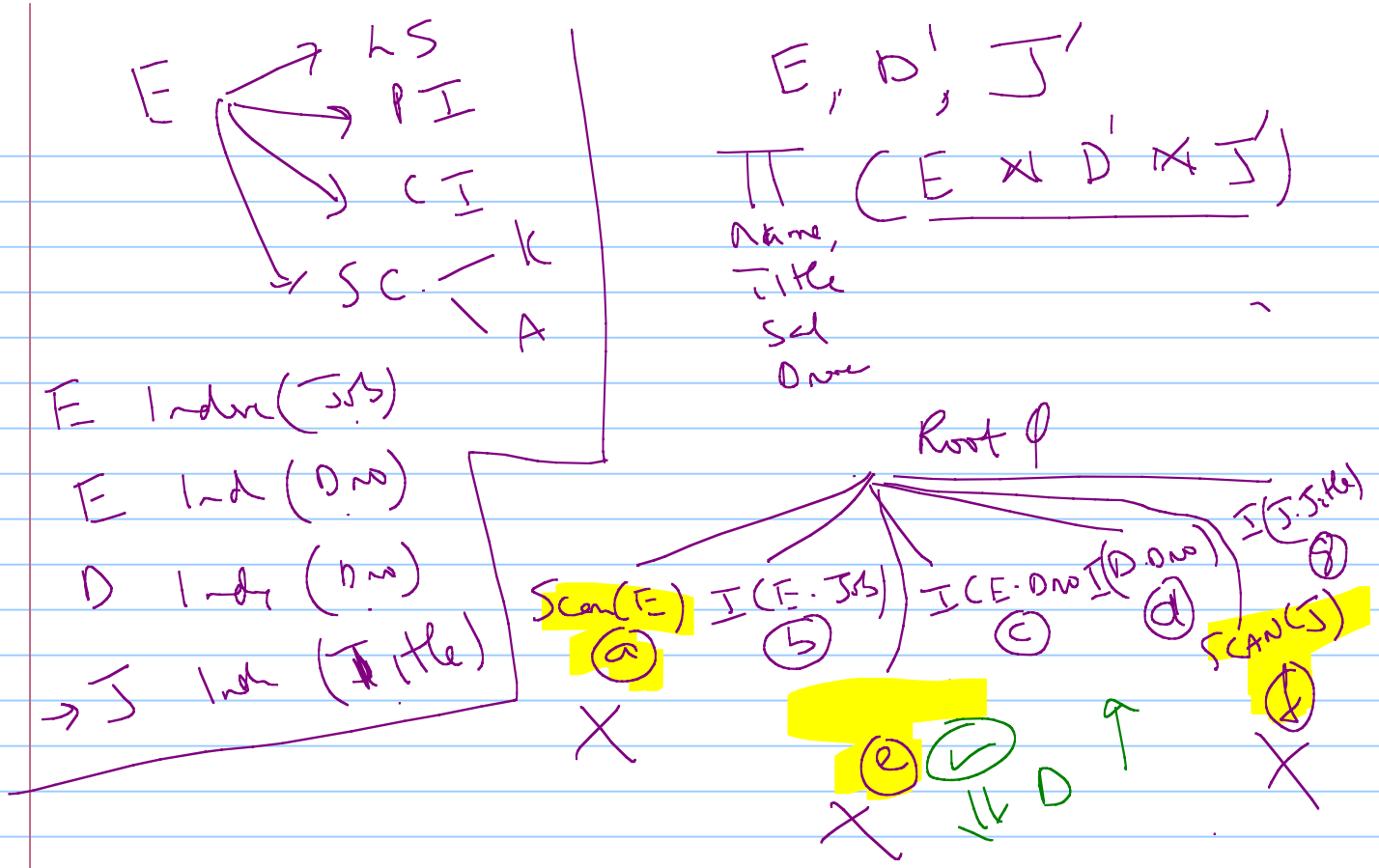
Select Name, Title, Sal, Dname.
From E, D, J
where Title = 'Admin'
AND Loc = 'Denver'
AND E.DNO = D.DNO
AND E.JOB = J.Job;

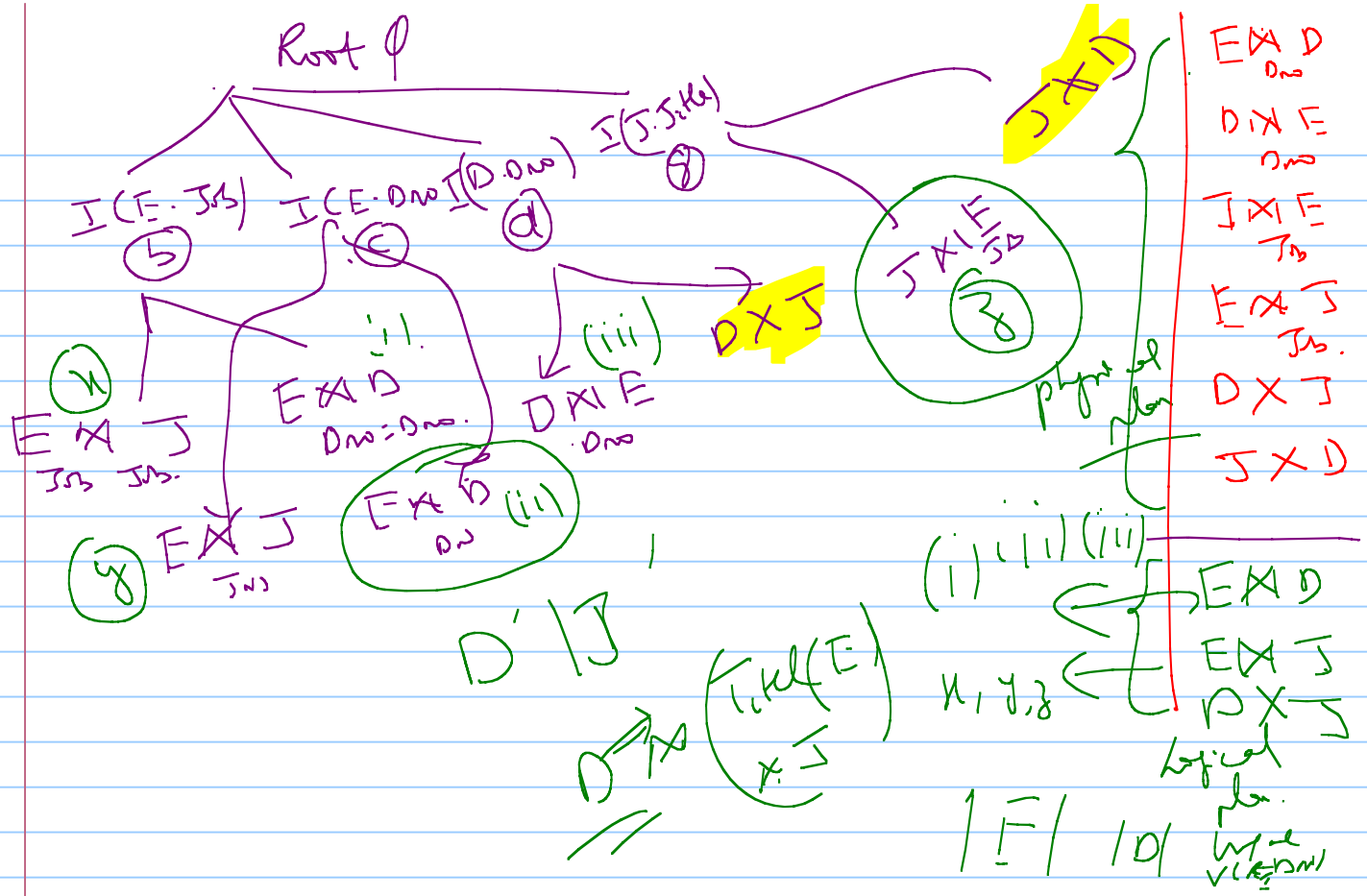
$J' = \sigma_{Title='Admin'}(J)$

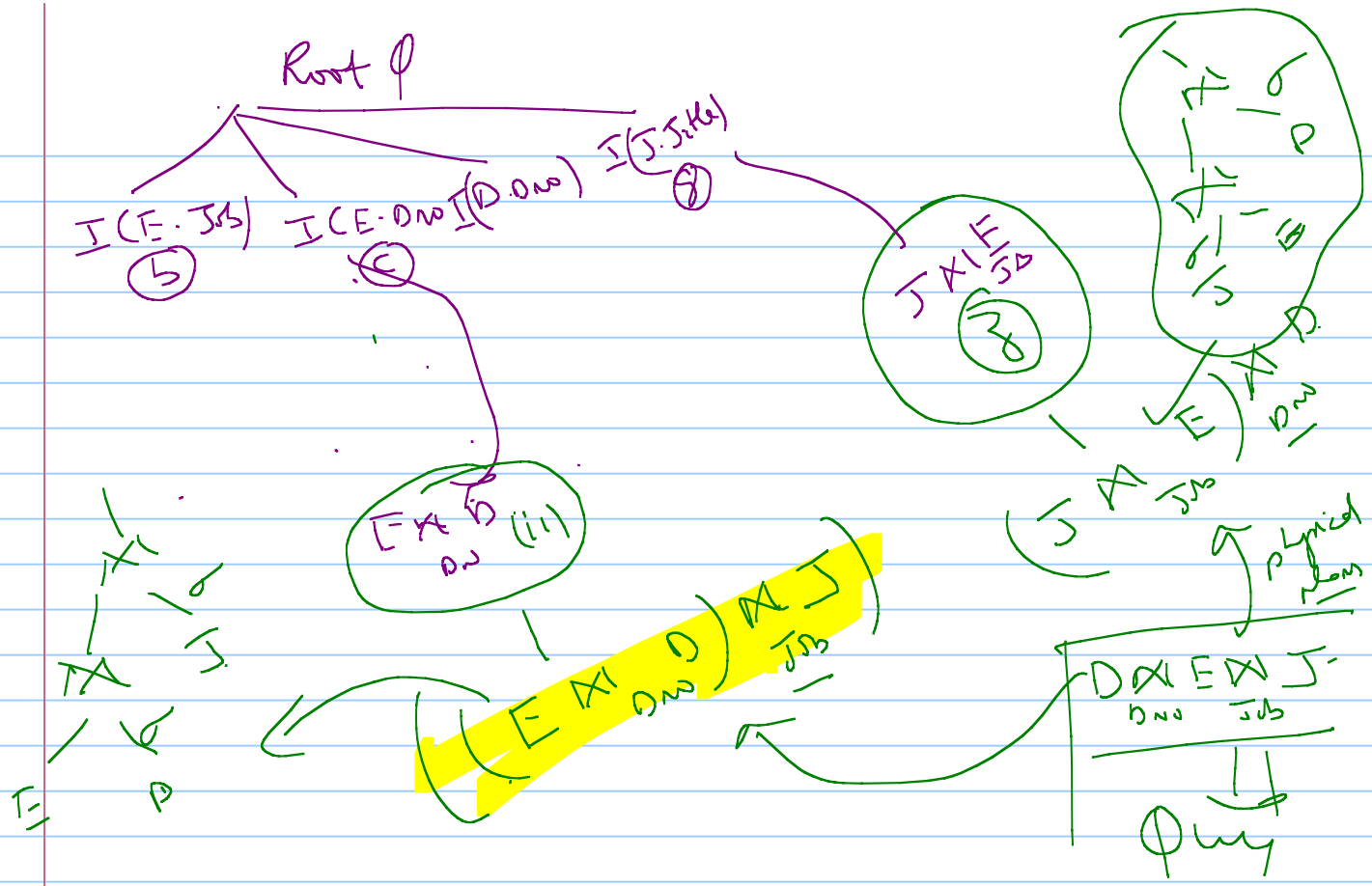
$D' = \sigma_{Loc='Denver'}(D)$

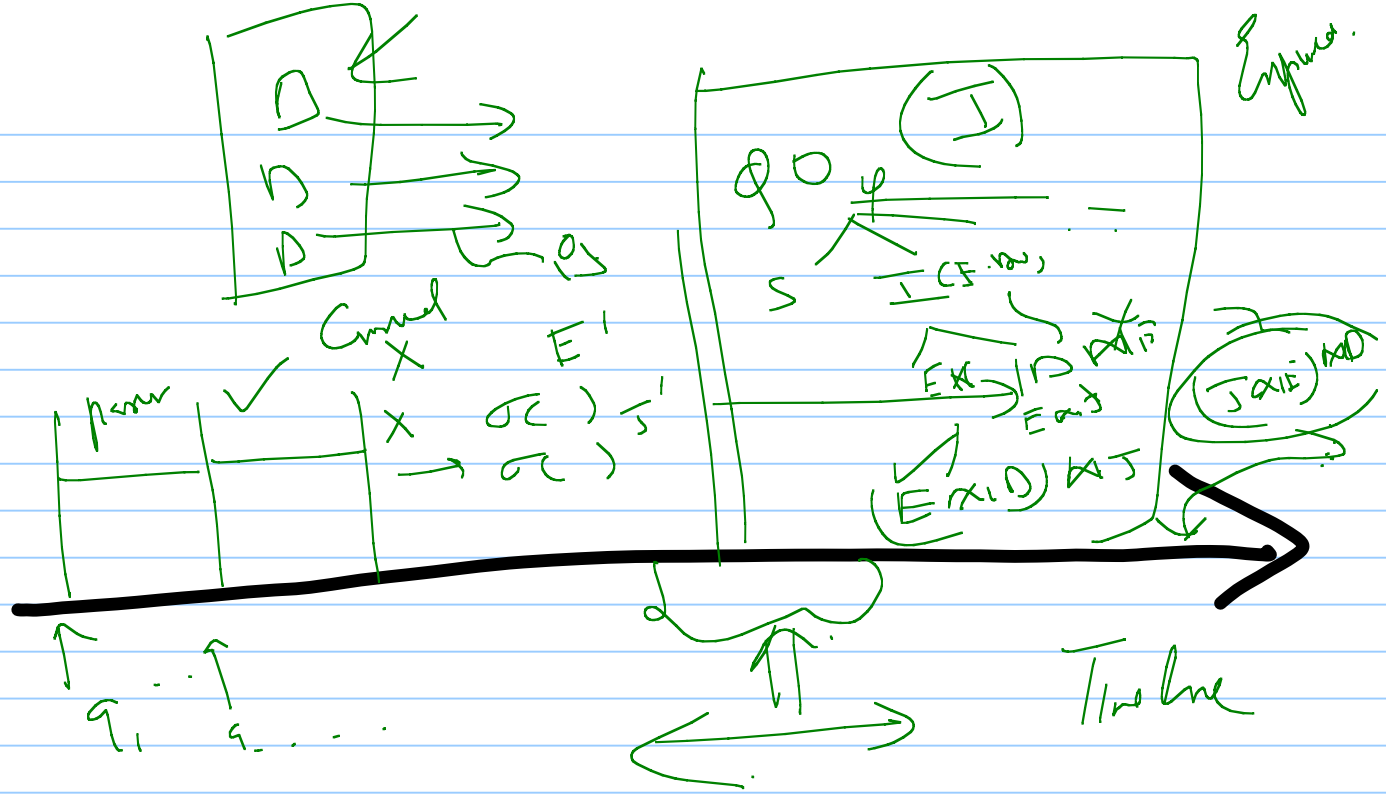
E, D', J'

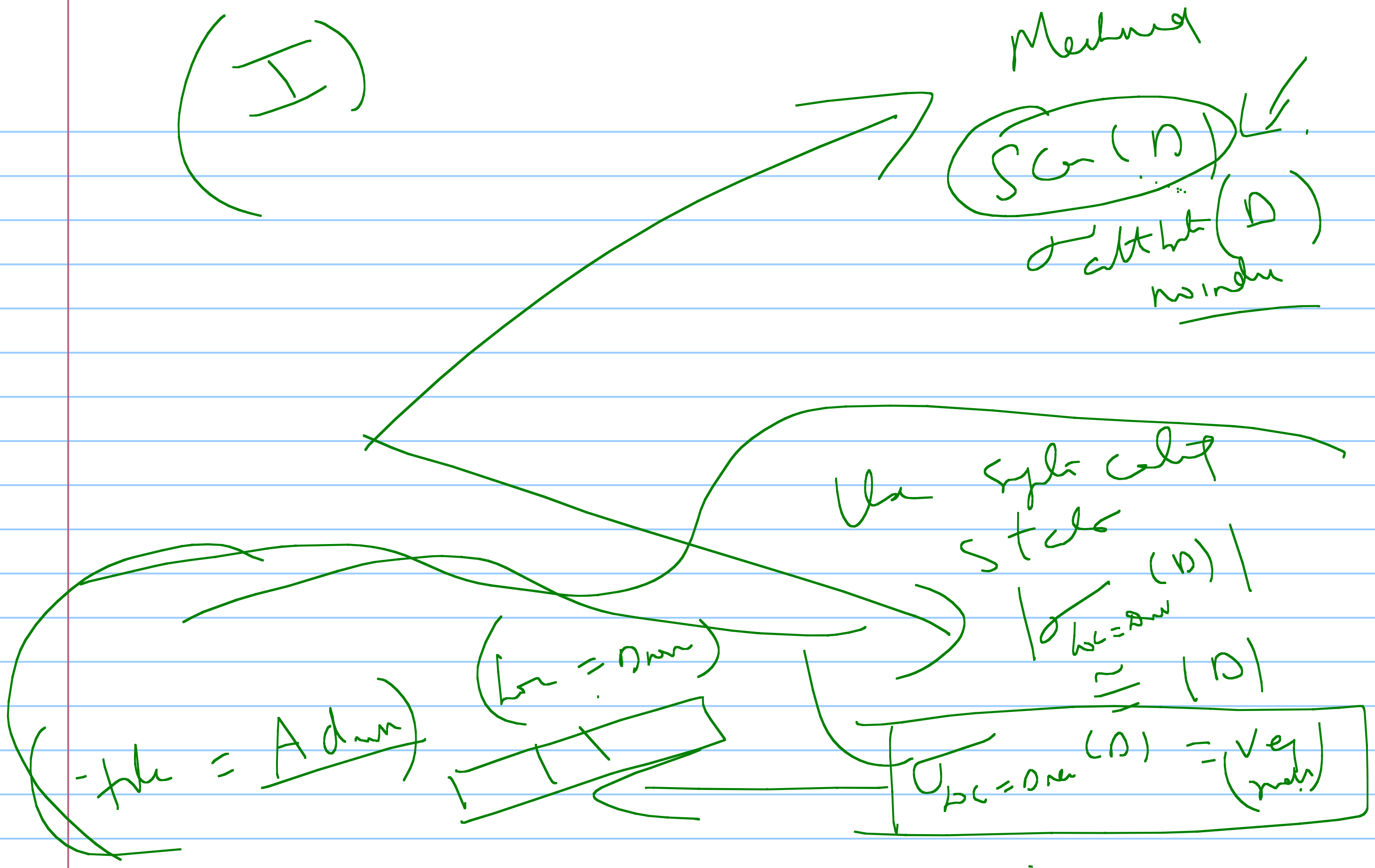
$\Pi_{Name, Title, Sal, DName}(E \bowtie D' \bowtie J')$

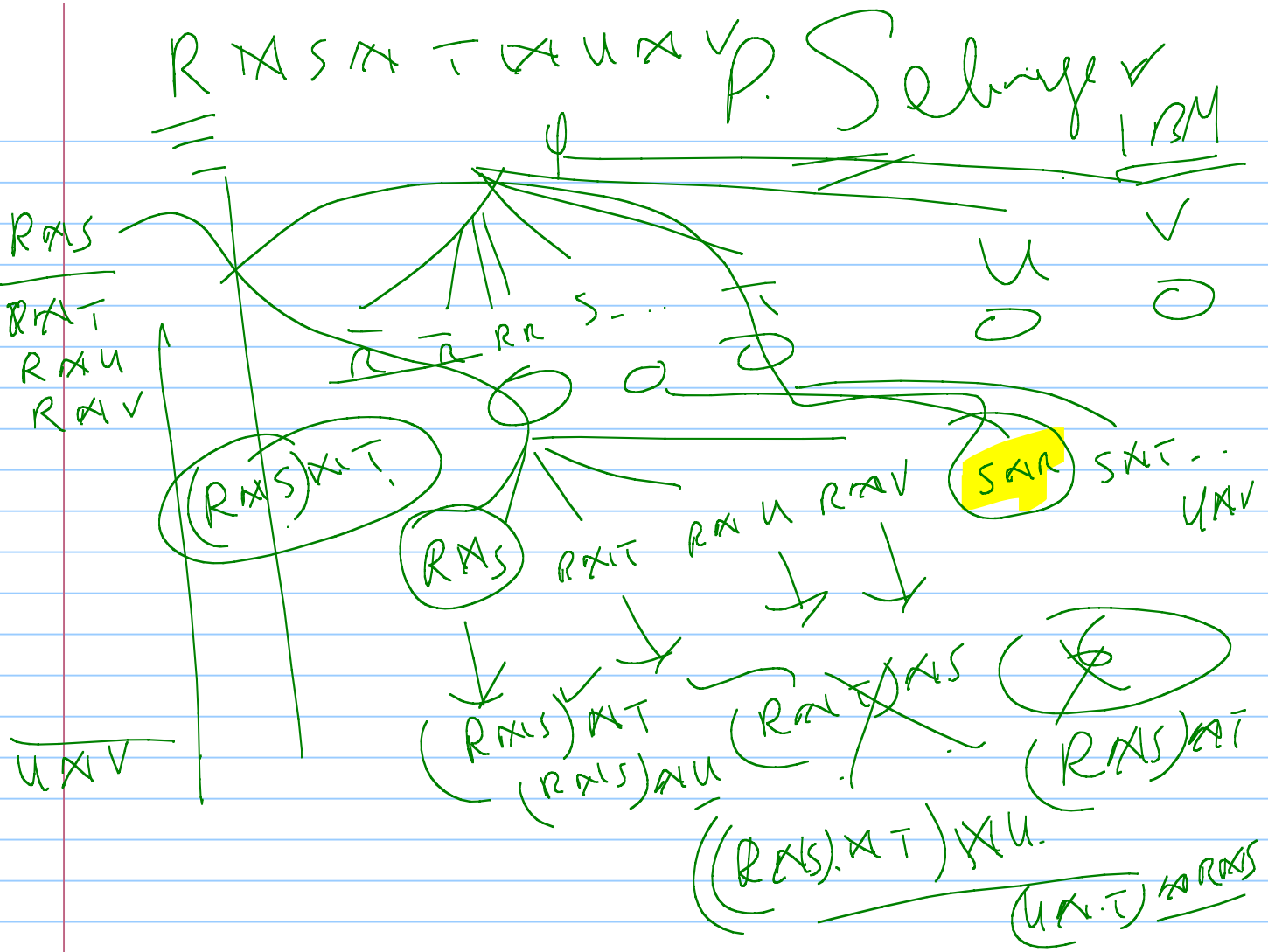

E
LS
PI
CI
SC
K
A
E Index(SSN)
E Ind(DNO)
D Ind(DNO)
J Ind(Title)
E, D', J'
π (E ⋈ D' ⋈ J)
Name,
Title
Sal
DNO
Root ϕ
Scan(E)
I(E.SSN)
I(E.DNO)
I(D.DNO)
I(J.Title)
SCAN(J)
a
b
c
d
e
f
g
X
X
X
D

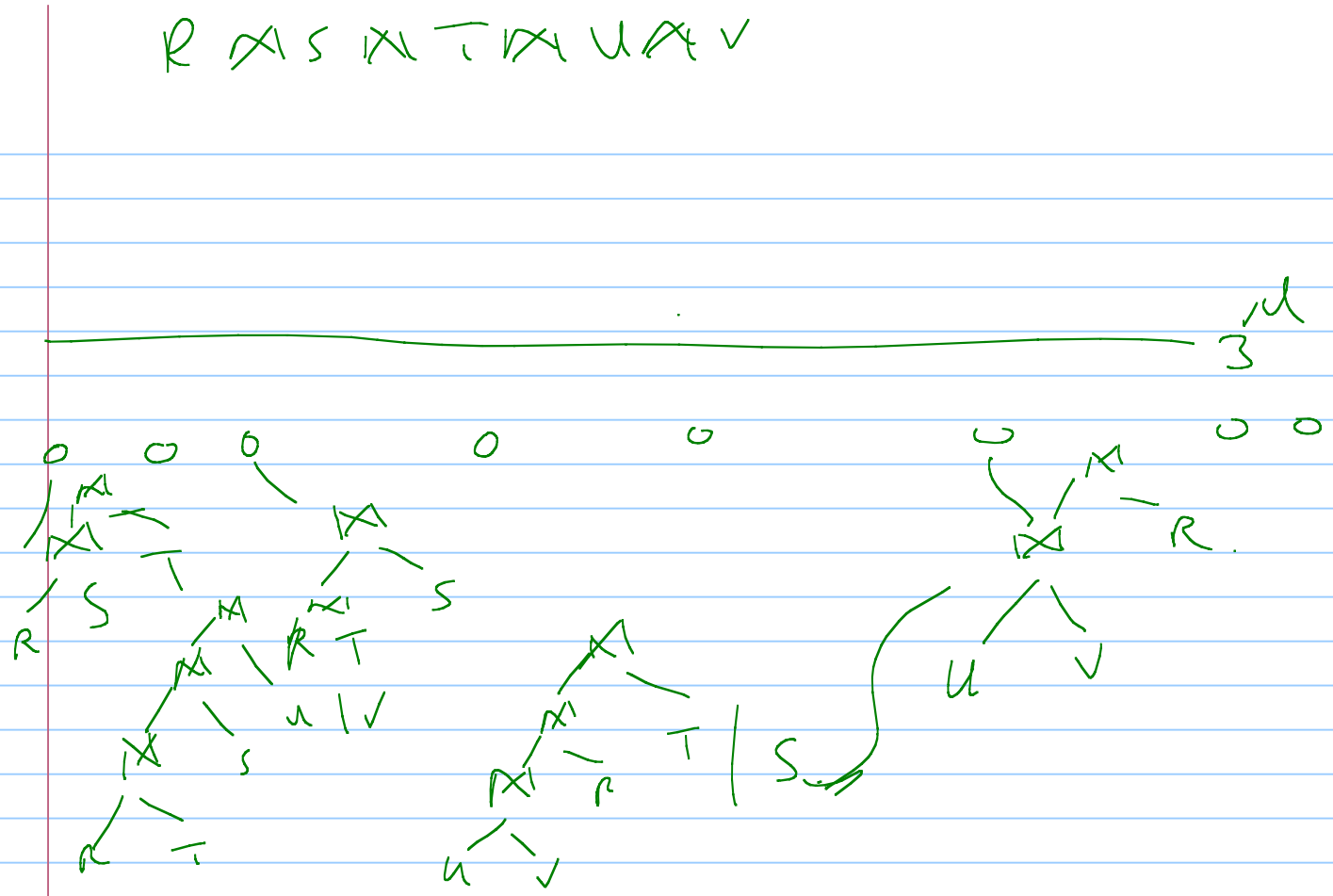

Root
physical plan
logical plan

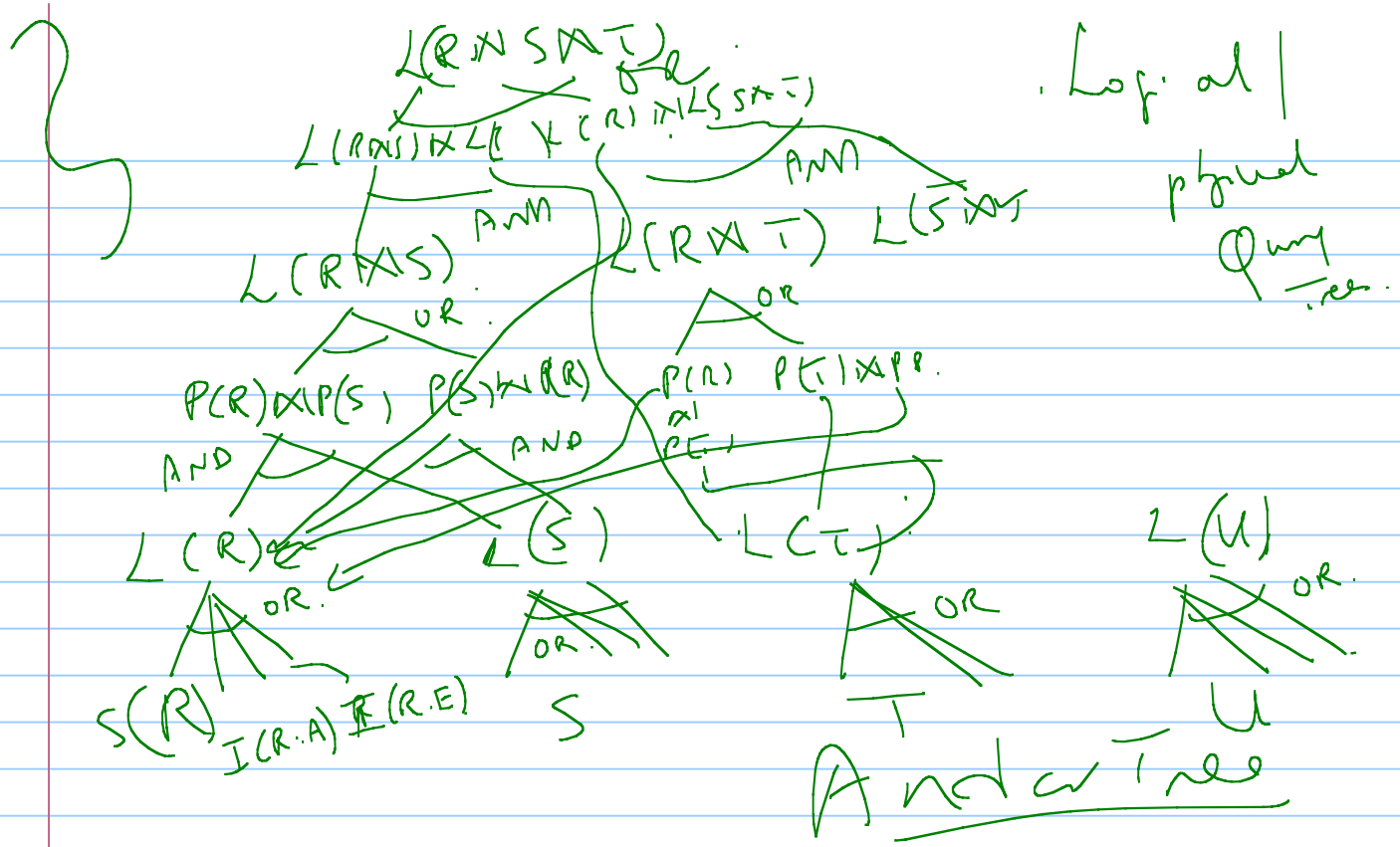

Root
5
6
8


Timeline


(I)



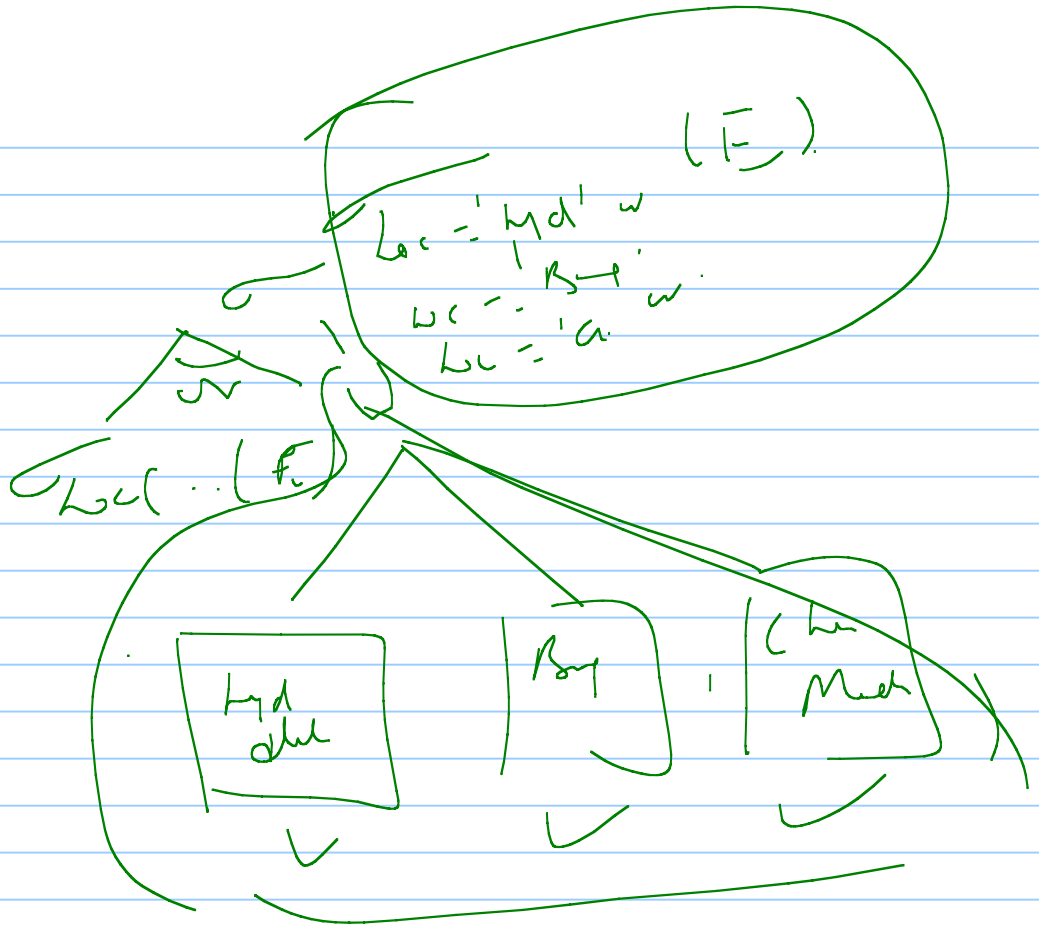




Logical plan

Query tree

Andor tree



And / OR

Implicity based selection

$\sigma_{b_i = bd}(E \bowtie D)$

=

OR

AND / OR

$\sigma_{Loc = Ld}(E \bowtie D)$ $E \bowtie (\sigma^{(D)}_{Loc = Ld})$

I ce premptu of the uk

OR

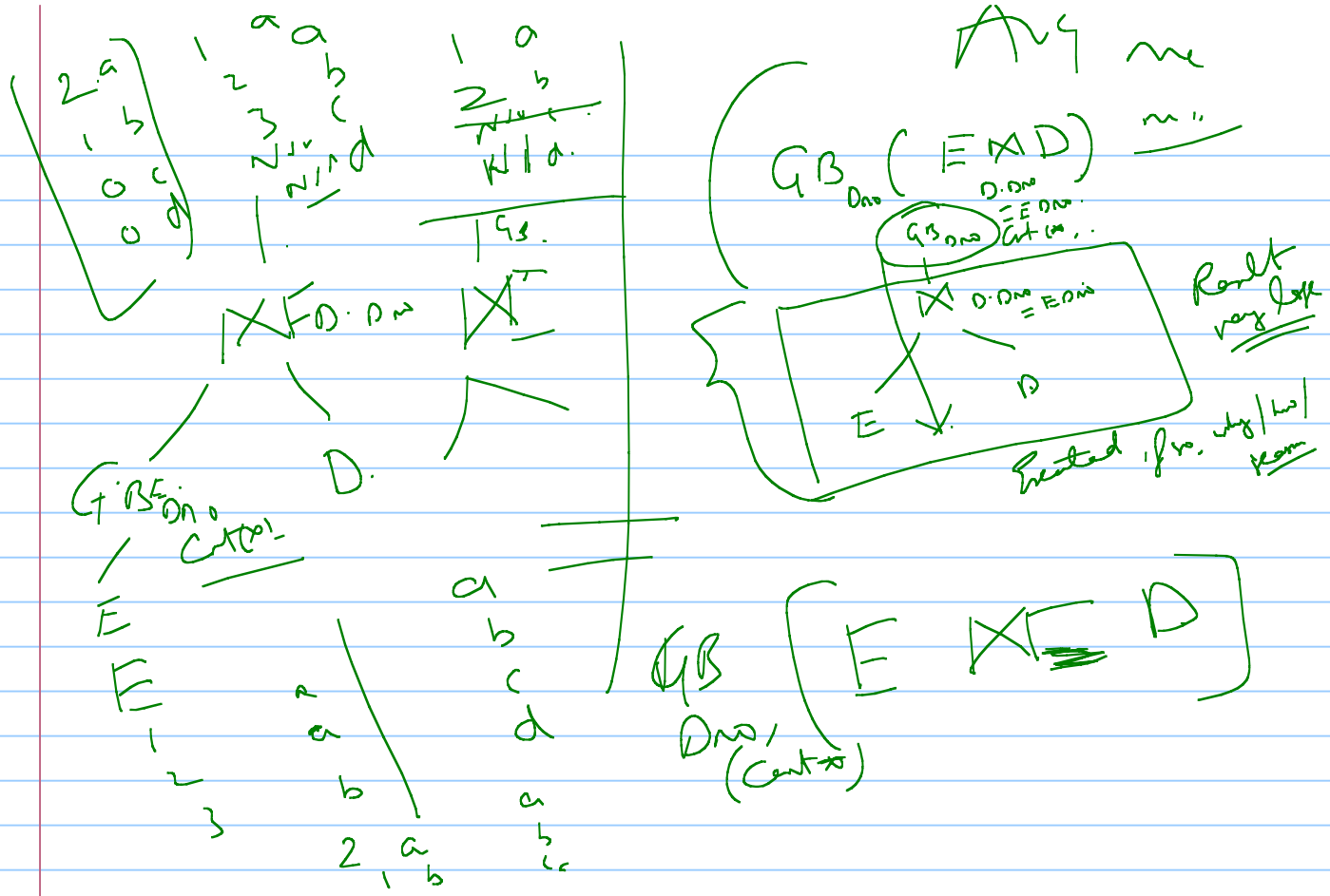
(A)

Eqular

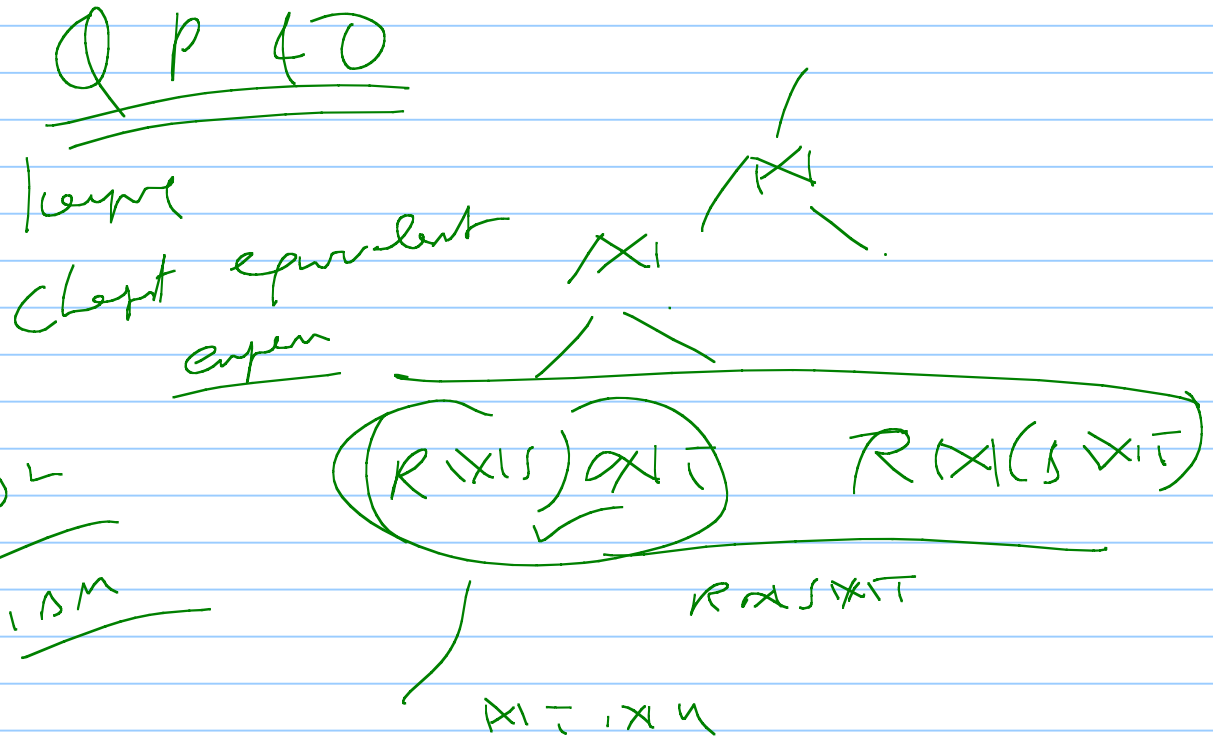
Expmn

Second

Az val.



Q P 40



Emp (Eno, Ename)
ASG (Eno, Pno, . .)
PROJ (Pno, Pname)

q SELECT Emp. Ename
From Emp, Asg, Prj.
Where Emp. Eno = Asg. Eno
And Asg. Pno = Prj . Pno
AND PROJ. Pname = 'CAD/CAM'.

q_1 SELECT Prj . Pno into JVAR
From Prj
Where Prj . Pname = 'CAD/CAM'.

q_2 Select Emp . Eno
From Emp, Asg, JVAR.
Where Emp. Eno = Asg . Eno
And Asg . PNO = JVAR . PNO.

Premise is
Only
Select A
From R
Where Cond
R . . . A

Matter

Prj

1	DB
2	CAD/CA
3	CAD/CA
4	DB

JVAR

2	CAD/CA
3	CAD/CA

$q \leftarrow q_2 \leftarrow q_1$
Result

Select Emp.Ename
From Emp, (ASG ⋈ JVAR)
Where E.ENO = A.ENO AND A.PNO = JVAR.PNO

ⓐ ASG.ENO
A.PNO = JVAR.PNO

q_{21} Select ASG.ENO into GVAR
From ASG, JVAR
Where ASG.PNO = JVAR.PNO

q_{22} Select Emp.Ename
From EMP, GVAR
Where EMP.ENO = GVAR.ENO

$q \Leftarrow q_{12} \Leftarrow q_{21} \Leftarrow q_2 — q_1$
result GVAR JVAR

JVAR

2	<ADD/CAR
3	<ADD/CAR

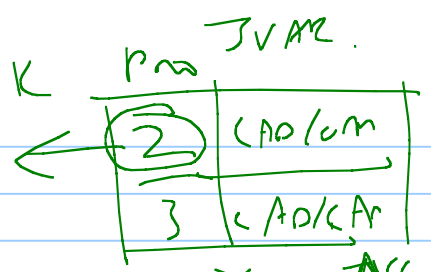
ASG

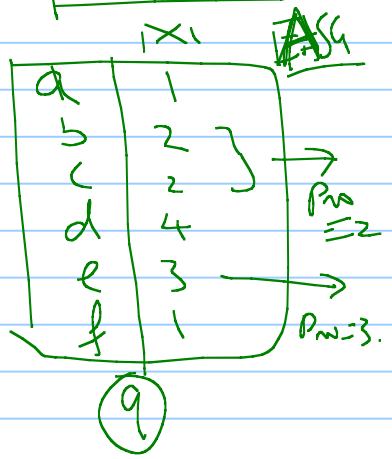
a	1
b	2
c	2
d	4
e	3
f	1

ⓠ

q_{21} Select ASG.ENO into GVAR
From ASG, JVAR
Where ASG.PNO = JVAR.PNO

q_{22} Select EMP.ENO
From EMP, GVAR
Where EMP.ENO = GVAR.ENO

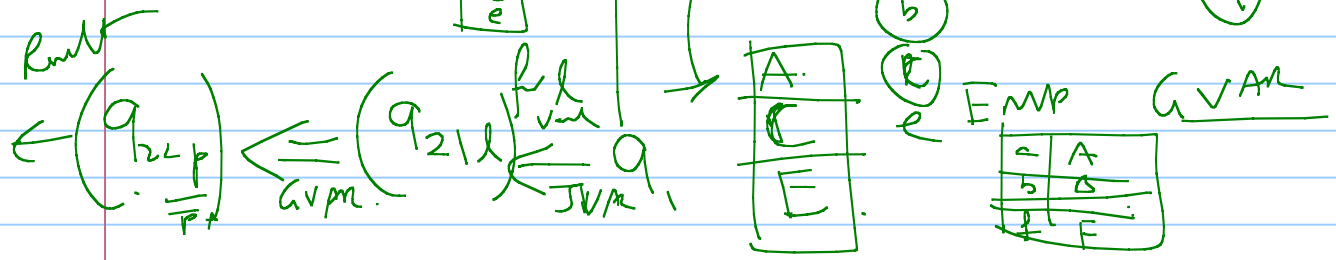


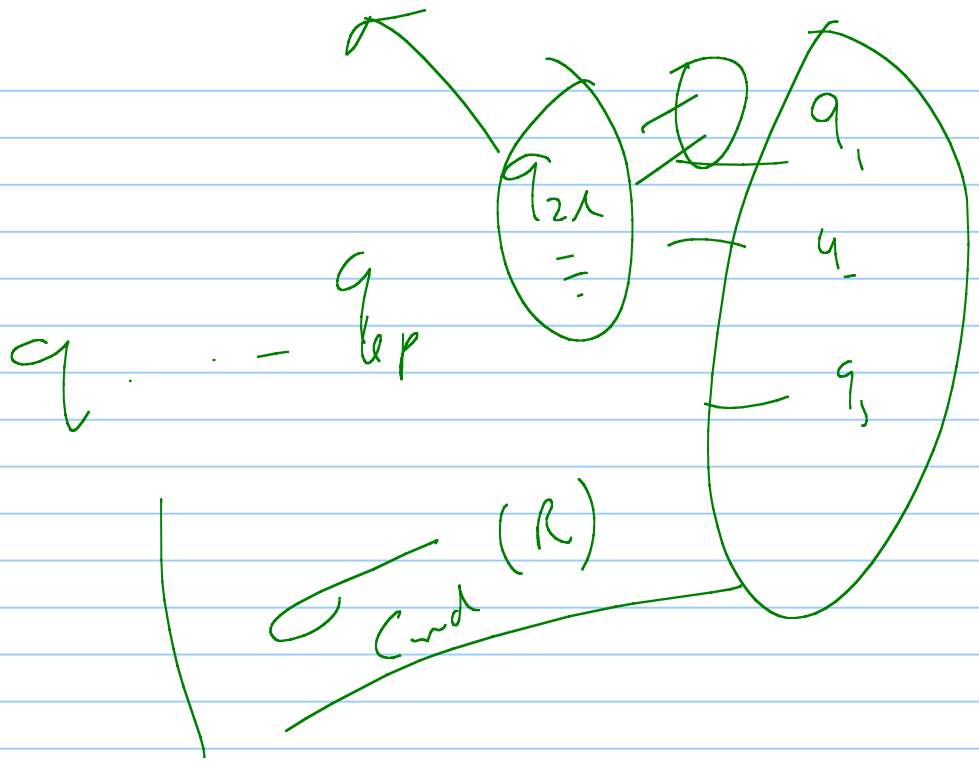


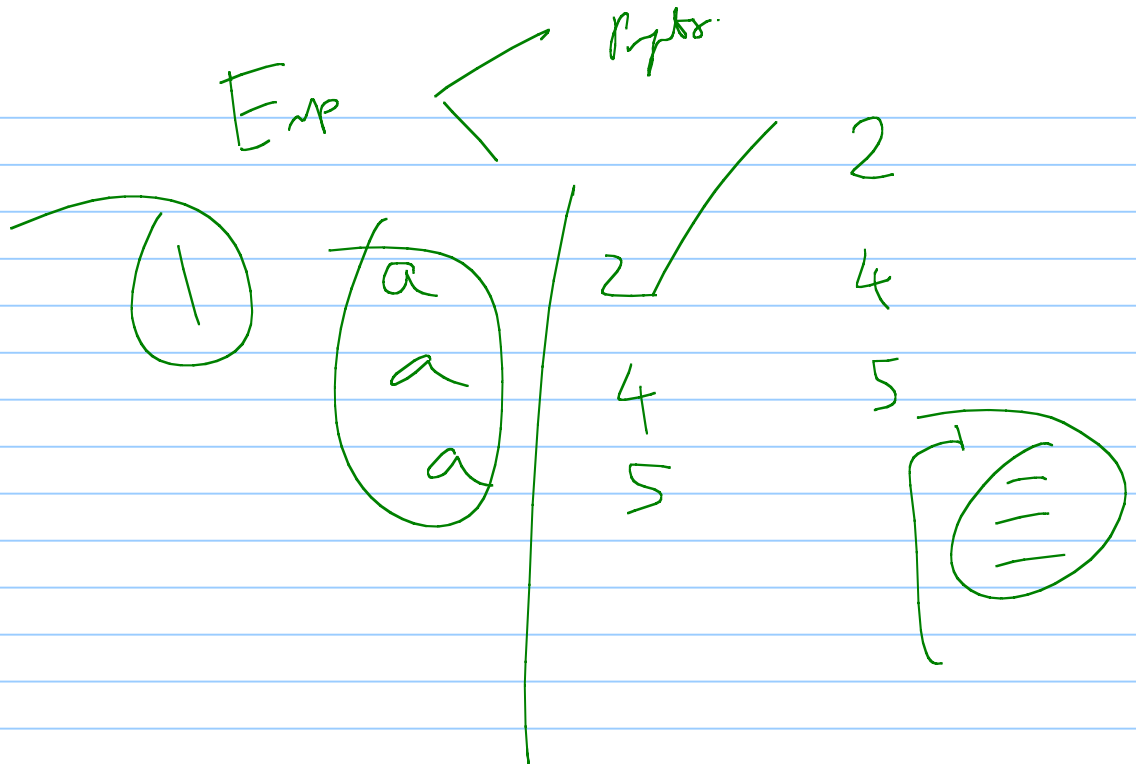
$q_{21}(l=1$ Select ASG.ENO into GVAR. (or cursor)
From ASG
Where ASG.PNO = JVAR.PNO value
$l=2$
$l=k$

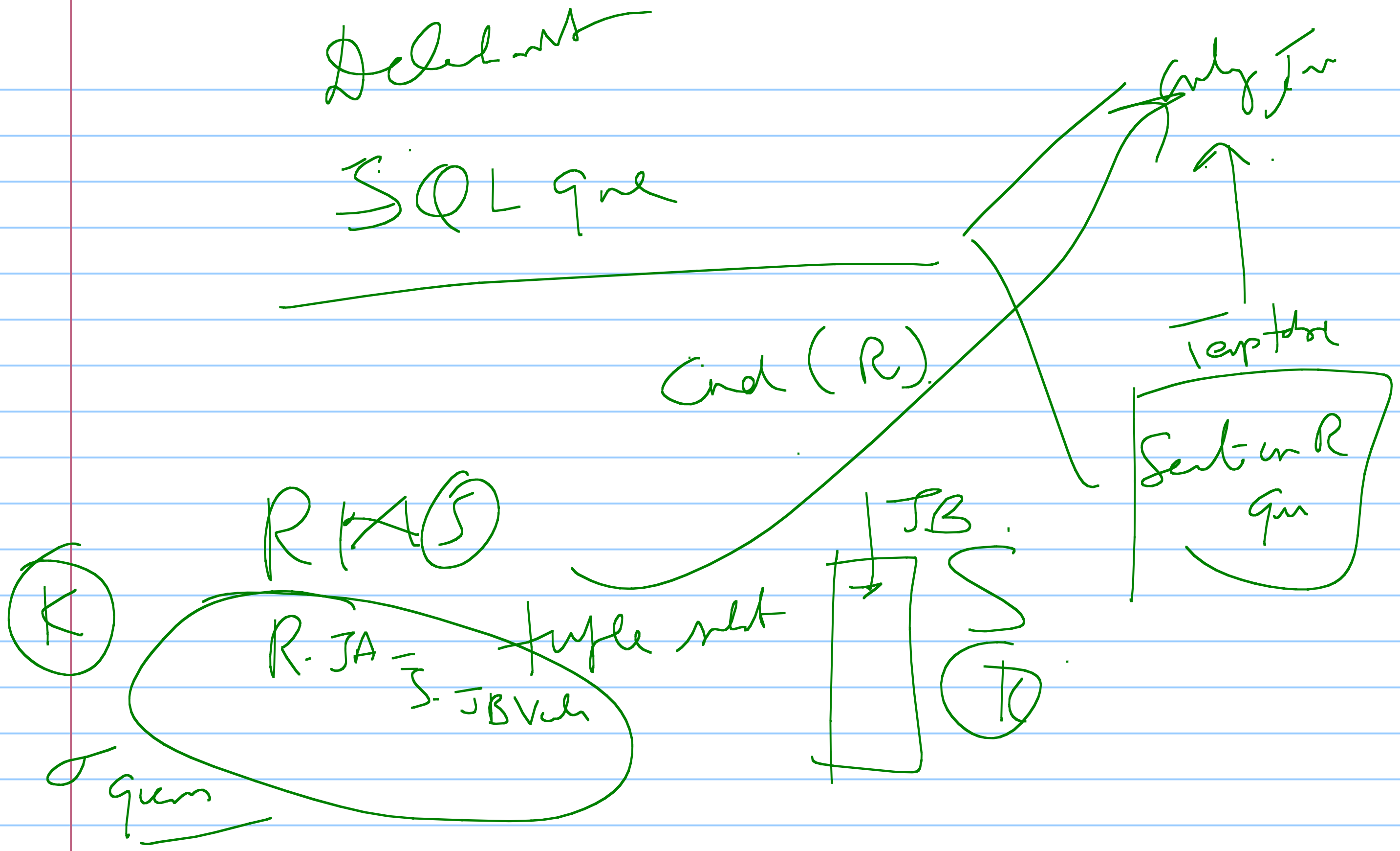
GVAR: b, c, e

$q_{22}(p)$ Select EMP.ENO
From EMP
Where EMP.ENO = GVAR.tuple.val

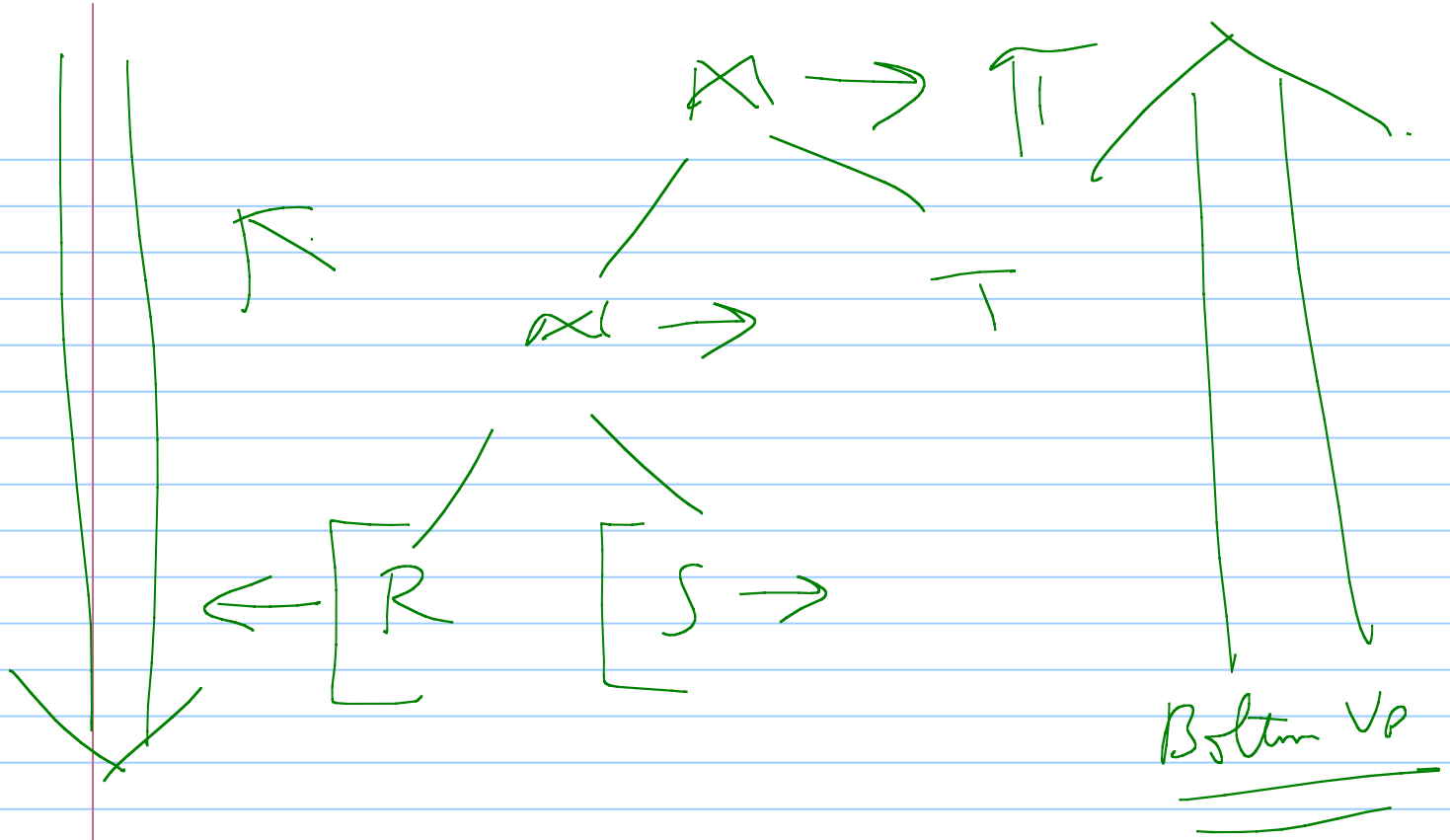


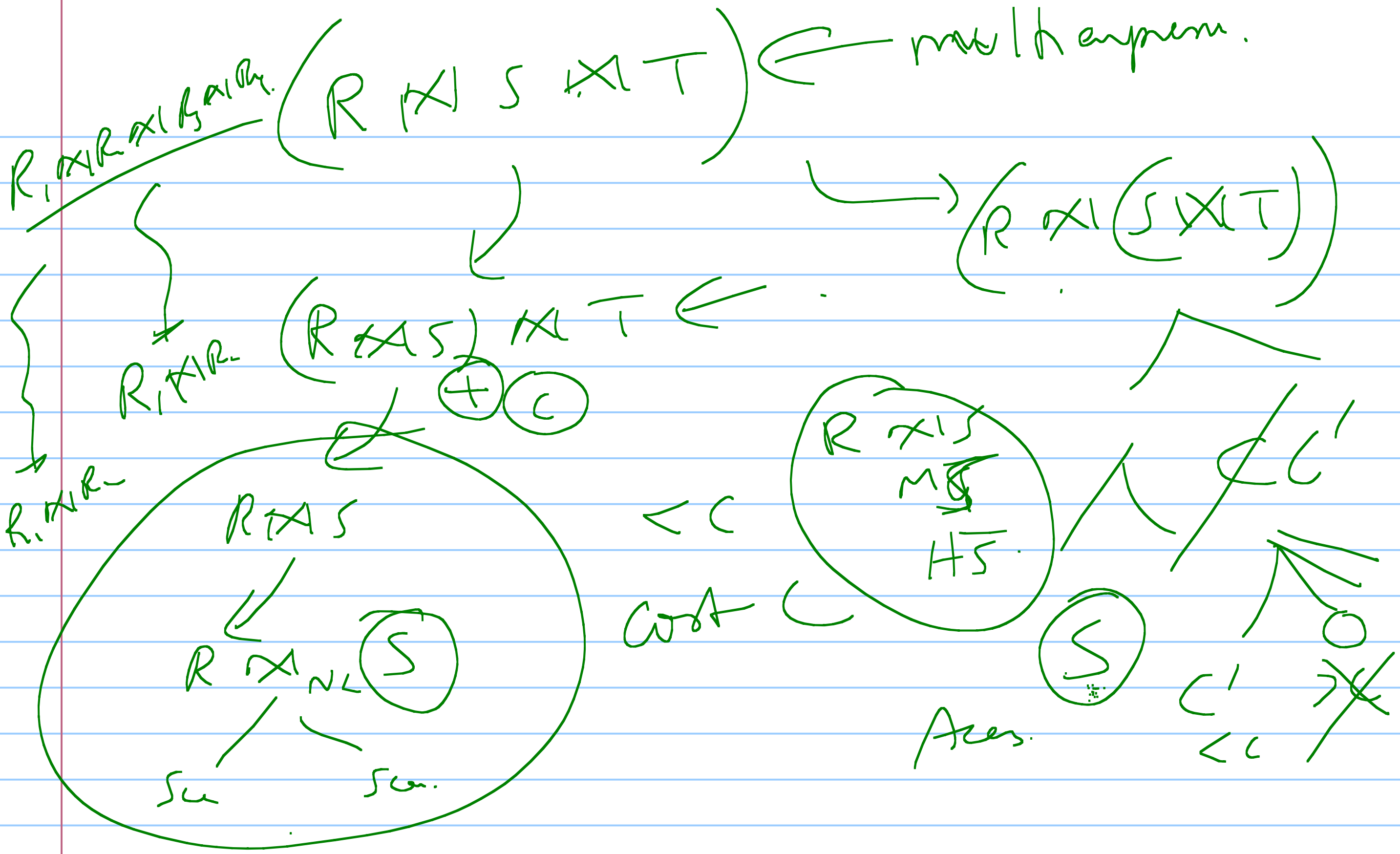

$q_{2\Lambda}$
q_1
q_-
q_3
(R)
Cond


Emp
Pupils
1
a
a
a
2
4
5
2
4
5

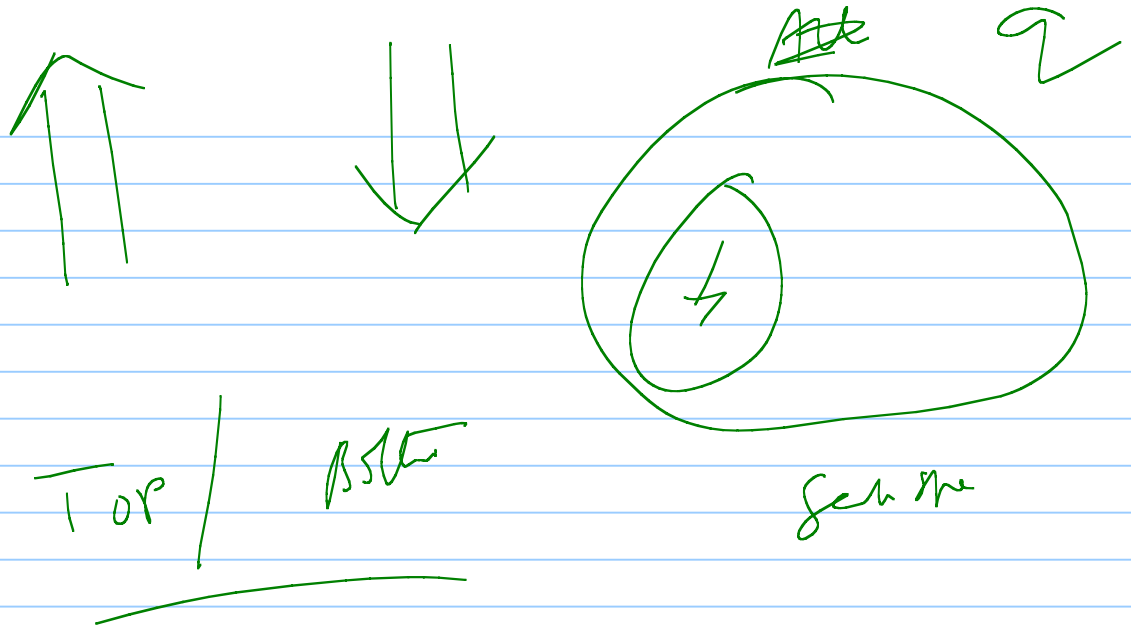


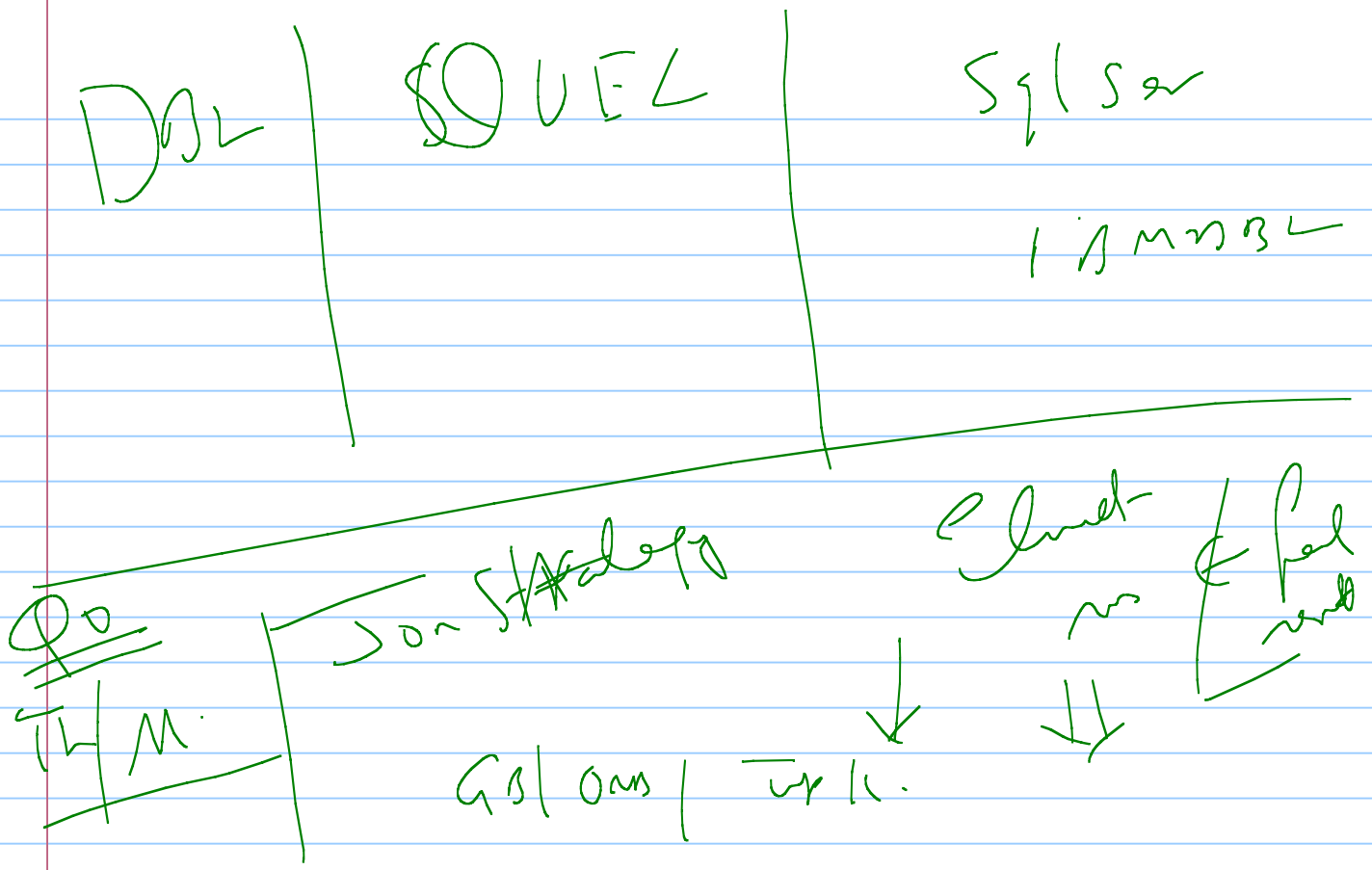
SQL Query
Cond (R)
R ⋈ S
R.JA = S.JBValue
tuple select
JB
S
Temp table
Select on R Query
K
Query


R
S
T
Bottom Up



(R ⋈ S ⋈ T)
(R ⋈ (S ⋈ T))
(R ⋈ S) ⋈ T
R ⋈ S
R ⋈NL S
< C
< C'





Note Title 2/13/2014

Note Title 2/13/2014

Note Title 2/13/2014

Note Title 2/13/2014