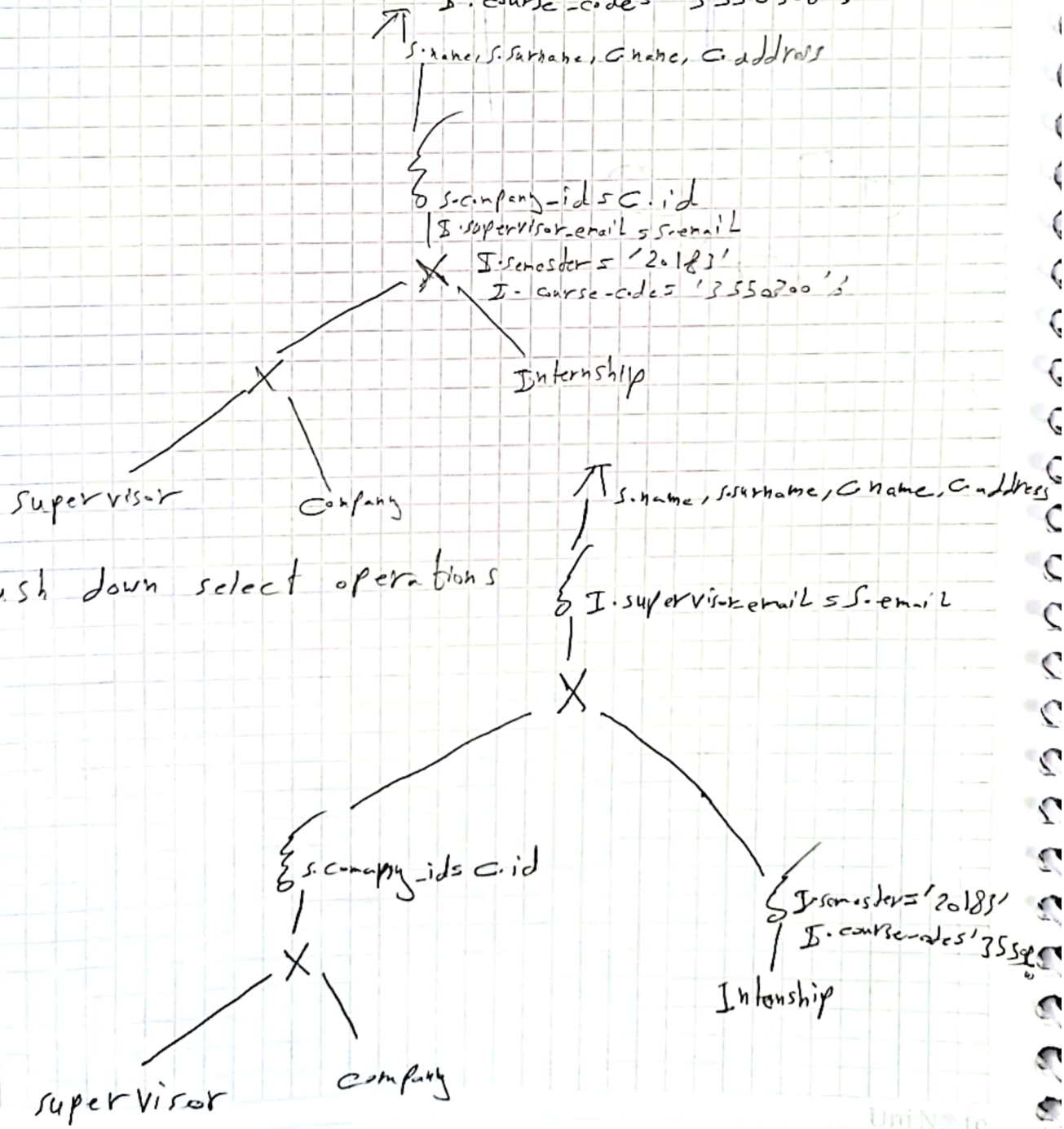
Assignment 1

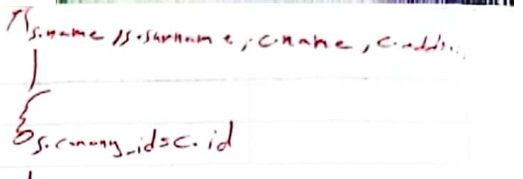
1) relational algebra:

$\Pi_{S.name, S.surname, C.name, C.address}\left(\sigma_{S.company\_id = C.id \wedge I.supervisor\_email = S.email \wedge I.semester = '2018' \wedge I.course\_code = '3550300'}\left((Supervisor \times Company) \times Internship\right)\right)$

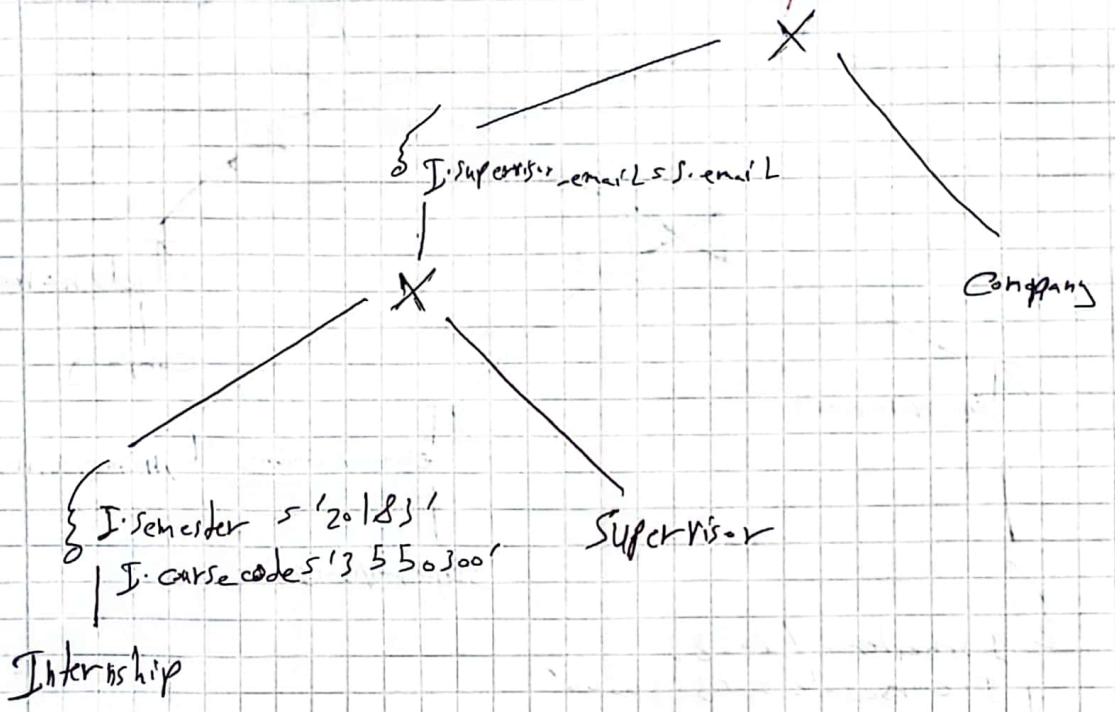
12AT!



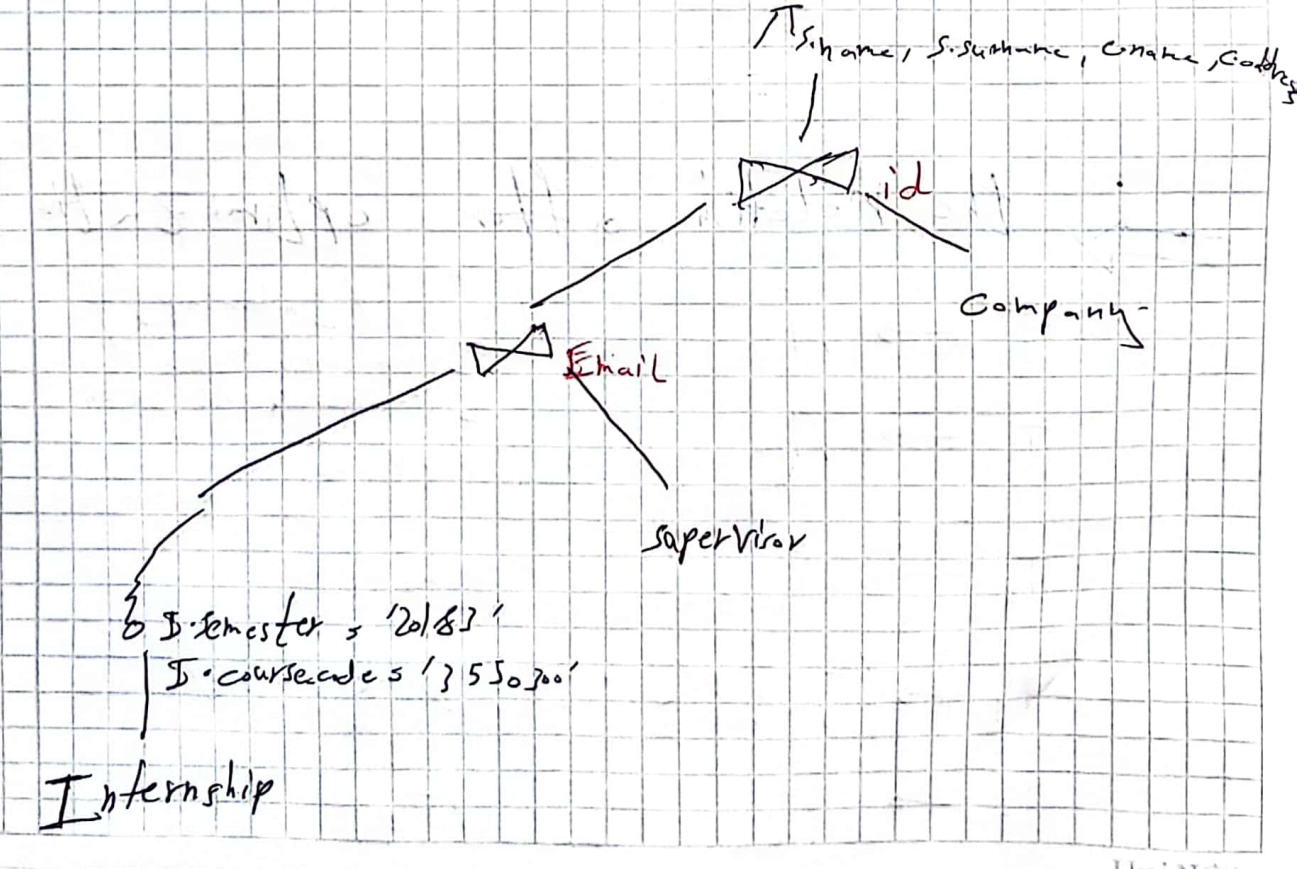
1) push down select operations



2) more restrictive goes first (left aligned tree)



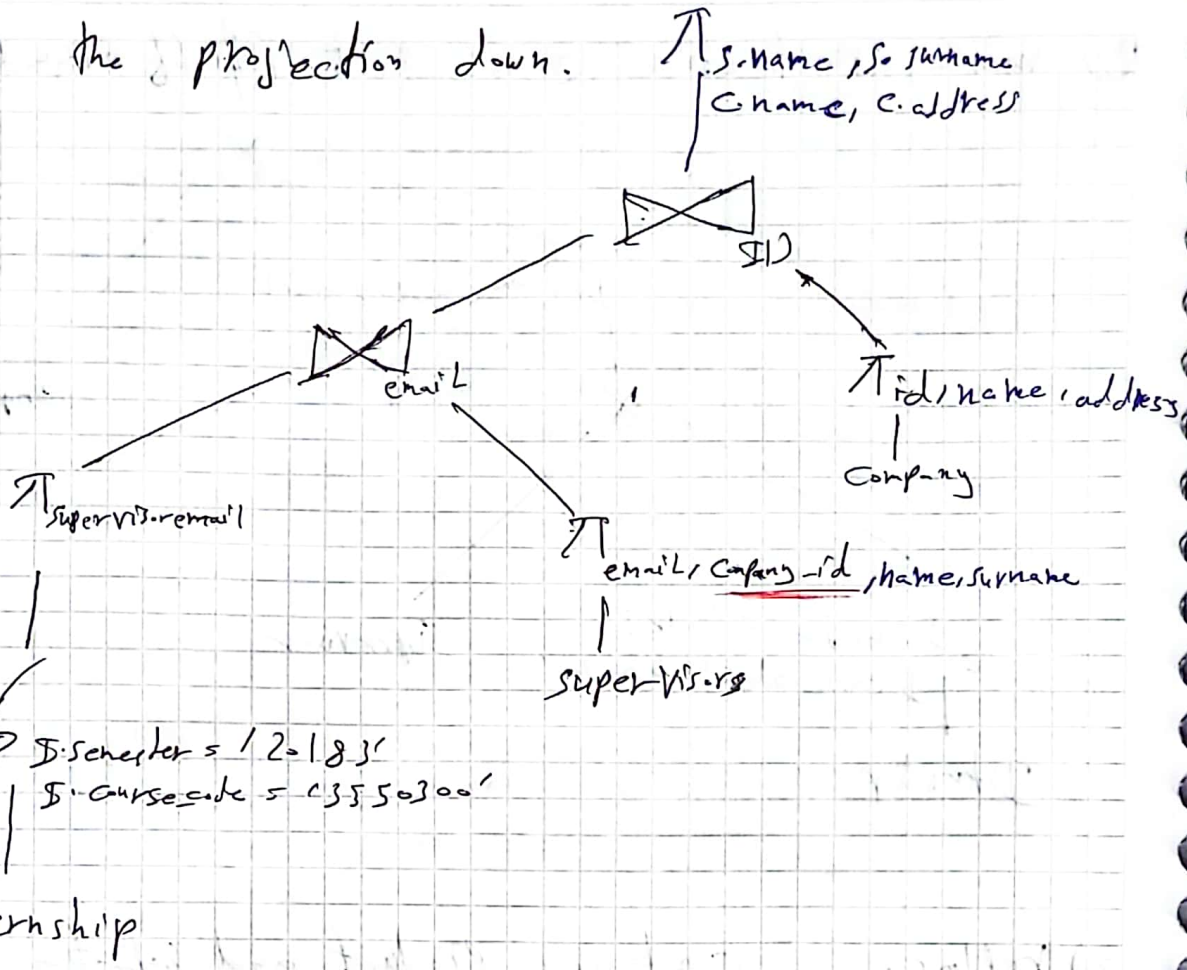
3) replace cartesian product and selection with Join



UniNotes

CamScanner

4) push the projection down.



⇒ the relation after optimization

$T_1 \rightarrow \Pi_{supervis.email}(\sigma_{I.semester='2-1837' \wedge I.course\_code='C355030'}(Internship))$

$T_2 \rightarrow \Pi_{email, Company\_ID, name, surname}(Supervisors)$

$T_3 \rightarrow \Pi_{id, name, address}(Company)$

$T_4 \rightarrow ((T_1 \bowtie_{email} T_2) \bowtie_{ID} T_3)$

$T_5 \rightarrow \Pi_{s.name, s.surname, c.name, c.address}(T_4)$

UniNote

CamScanner

2) company (id, name, address, tel) — Hash

supervisor (email, company_id, name, surname, tel) — Hash, B+tree

internship (internship_id, start_date, course_code, semester, supervisor_email, student_id) — Hash, clustering

nTuples(supervisor) = 8271 ⟹ nB(supervisor) = 207

bfactor(supervisor) = 40

nTuples(internship) = 28271 ⟹ nB(internship) = 1414

bfactor(internship) = 20

nDistinct$_{courscode}$(internship) = 15

nlevels$_{email}$(I) = 2, nlevels$_{name}$(I) = 2, nlevels$_{courscode}$(I) = 2

a) $\sigma_{name = 'Nikat'}$(supervisor)

number of blocks

Linear-search = $\lceil \frac{nTuples(supervisor)}{bfactor(supervisor)} \rceil = \lceil \frac{8271}{40} \rceil = 207$

Equality - non clustering (B+ tree):

Assumption:

nDistinct$_{name}$(supervisor) = 1000

⟹ SC$_{name}$(supervisor) = $\lceil \frac{8271}{1000} \rceil = \lceil 8.28 \rceil = 9$

⟹ cost = nlevels$_{name}$ + SC(supervisor) = 2 + 9 = 11

with hash

b) $\sigma_{courseCode=1355400}$ (Internship)

Linear: $\left\lceil \frac{nTuples(internship)}{bfactor(internship)} \right\rceil = \left\lceil \frac{28271}{20} \right\rceil = 1414$

Equality - clustering:

$SC_{courseCode}(internship) = \left\lceil \frac{28271}{15} \right\rceil = 1885$

$\Rightarrow$ cost = $nlevels_{courseCode}(I) + \left\lceil \frac{SC_{courseCode}(internship)}{bfactor(internship)} \right\rceil$

$= 2 + \left\lceil \frac{1885}{20} \right\rceil$

$= 2 + 95 = \boxed{97}$ → minimum.

c) Internship $\bowtie_{email}$ Supervisor (Assume nBuffer=100)

① Block nested loop Join: nBlocks(I2) + (nB(I2) * nB(S))

$\Rightarrow$ 1414 + (1414 * 207) = 294,112

② Buffer: nBlocks(I2) + [nBlocks(S) * (nB(I2)/nBuffer - 2)]

= 1414 + ⌈207 * ⌈1414/100 - 2⌉⌉ = $\boxed{4,401}$

③ Indexed nested loop Join: nBlocks(I2) + nTuples(I2) * (hLevel + 1)

= 1414 + 28271 * (2+1) = 86,227

④ sort_Merge: nBlocks(I2) + nBlocks(S) + sortCost(I2)

(5) Hash Join: $3 * (nBlocks(R) + nBlocks(S))$

$= 3 * (1414 + 207)$

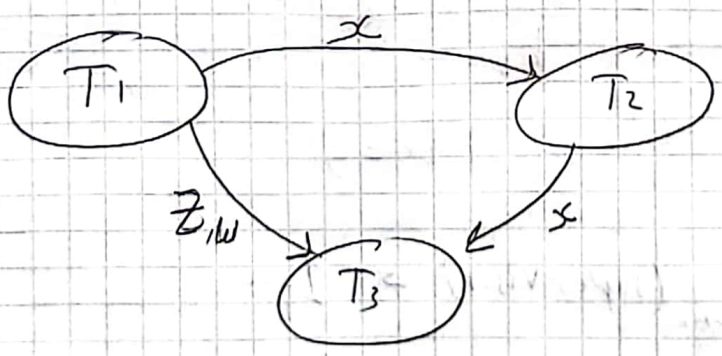
$= 4863$

3)

T_1	T_2	T_3
	$R(y)$	
	$W(y)$	
$W(x)$		
$W(z)$		
		$R(z)$
	$R(x)$	
$R(w)$		
$W(w)$		
	$W(x)$	
		$R(w)$
		$W(w)$
		$W(x)$

(precedence graph)

a)



no cycle $\Rightarrow$ conflict serializable

b) $< T_1, T_2, T_3 >$

UniNote

CamScanner