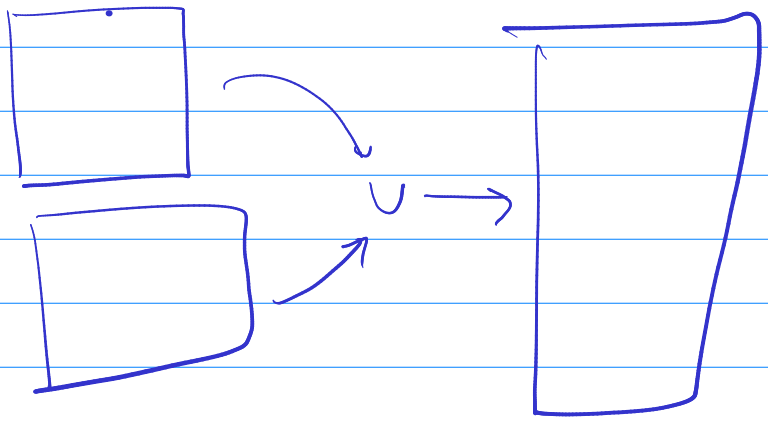
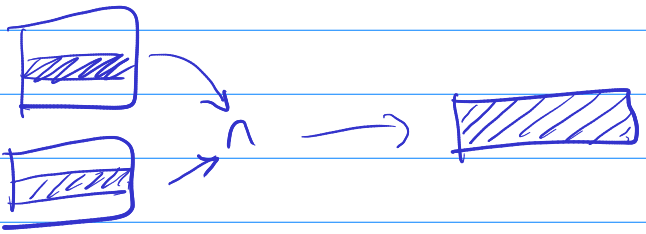
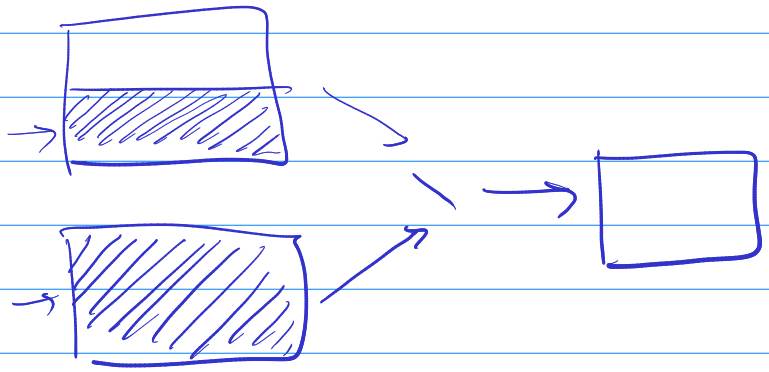
Spreadsheets | Relational → Databases

Relational Algebra



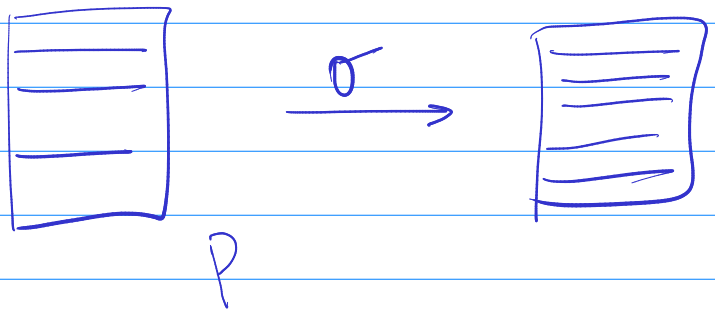


Unions ∪ Intersections ∩ set differences \



Practical people Academic researchers

where / filter / selection

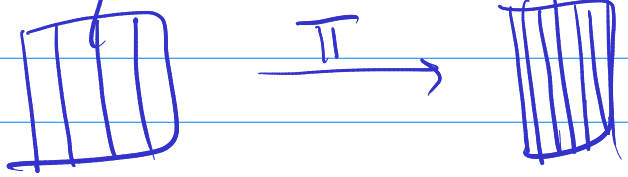


$\sigma_P T$

All rows in T which satisfy P

Practical Acad.

select / projection

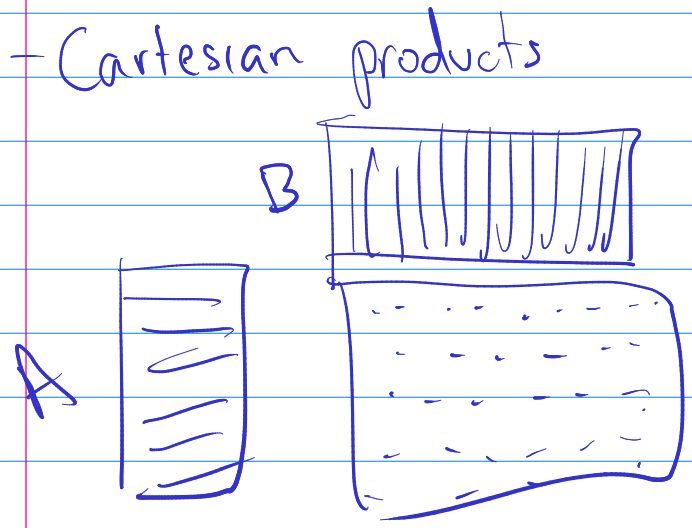


$\Pi_{C_1 C_2 C_3} T$

Just the columns C_1, C_2, C_3 in T.

Joins

- Cartesian products



A × B

Every row in A combined with every row in B

$\sigma_{T_1.c_1 = T_2.c_2}(A \times B)$

Θ

Θ-Join, Natural join, equijoin, Semijoin

SPJ calculus ($\cup$ $\cap$ $\setminus$)

select project join $\bowtie$

Query Evaluation for Dummies

$\cup$ $\cap$ $\setminus$

Union(A B)

For every row r in A: emit r

For every row r in B: emit r

Intersection (A B)

Assuming A is smaller

- For every row r in A
  - If r in B: emit r.

Set difference (A B)

- ∀ row r ∈ A
  - if r ∉ B: emit r

Select (A p)

- ∀ r ∈ A if p(r) then emit(r)

Project (A c_1 c_2)

[For $r \in A$: output respective columns

Join (A B $A.c_1 = B.c_2$)

[For every row r in A

If row $r' \in B$ where

$r'.c_2 = r.c_1$

Output $(r \cdot r')$

Fancy database indices

- How to perform complicated joins?

$(T_1 \bowtie T_2) \bowtie T_3$ for $r_1 \in T_1$

for $r_2 \in T_2$

$T_1 \bowtie (T_2 \bowtie T_3)$ for $r_3 \in T_3$

...

Given: Database T

SPJ query q

Calculate $q(T)$.

Question 1: Fix q, vary T

How difficult is the problem?

Data complexity $\in$ Polynomial

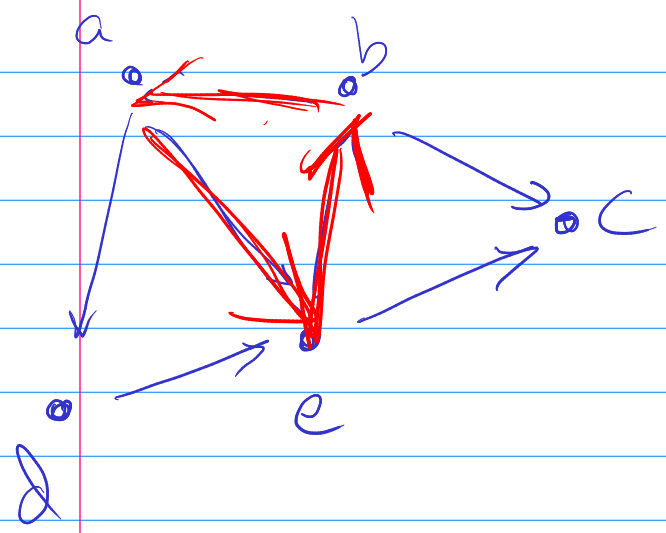
Question 2: Fix T, vary q

How difficult? q_1 q_2 q_3

Query complexity $\in$ Exptime n^1 n^2 n^3 ... n^n

Q3: Vary both T & q

Combined complexity



Edge	
From	To
a	d
a	e
b	a
b	c
d	e
e	b
e	c

Question: Given a graph, list all triangles.

$$\bigcup_{c_2=c_3} \left(E \times E \times E\right)$$

$c_4 = c_5$

$c_6 = c_1$