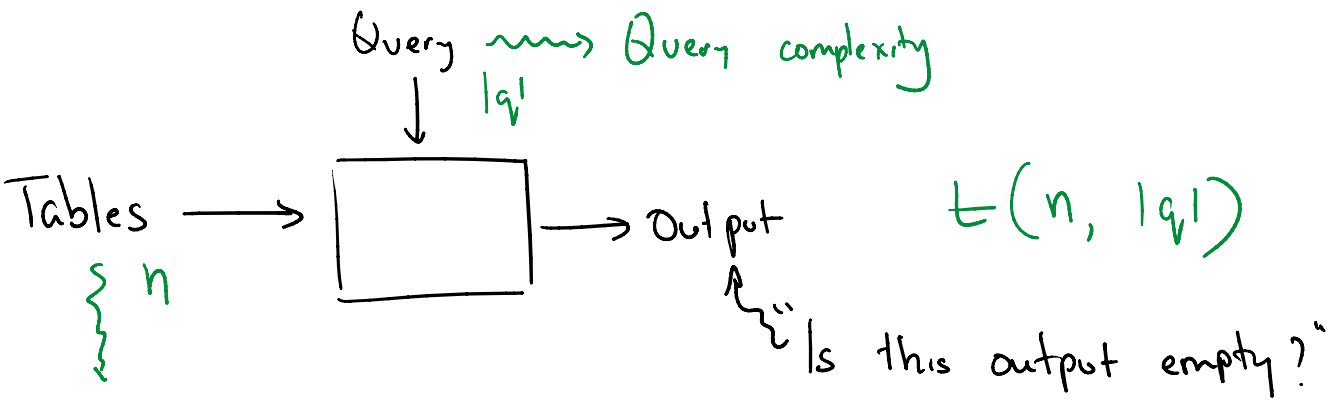
Lecture 24

Tuesday, November 16, 2021 11:57 AM

– How to evaluate relational queries?

(Decision version)



$t(n, |q|)$

Combined complexity

$T_1 \bowtie_{T_1 \cdot c_1 = T_2 \cdot c_2} T_2 \longrightarrow$

for t_1 in T_1:

for t_2 in T_2:

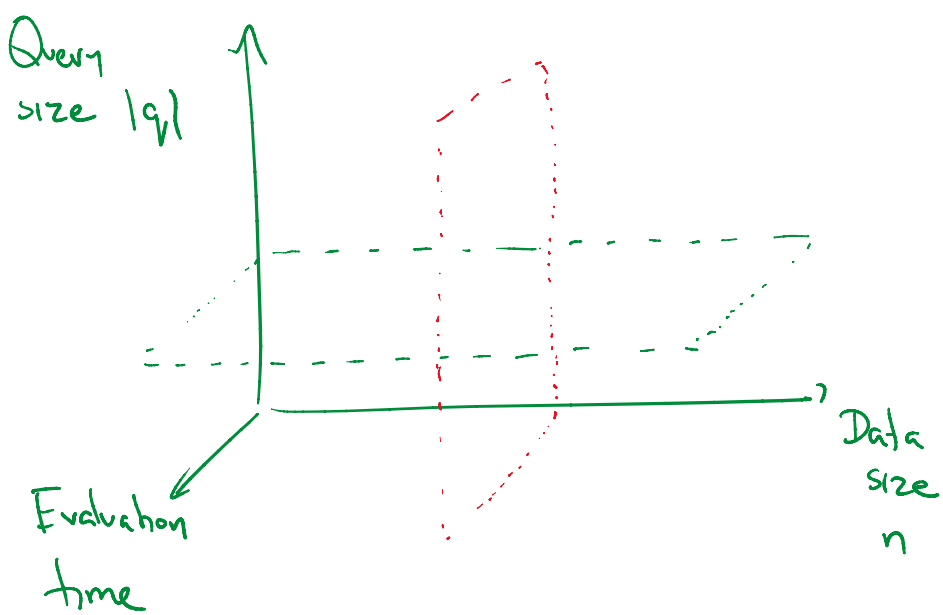
if $t_1 \cdot c_1 = t_2 \cdot c_2$:

then emit $(t_1 \cdot t_2)$

$\boxed{\text{Data complexity of relational algebra} \in P}$

T_1 ; $T_1 \underset{c_1=c_2}{\bowtie} T_2$; $\left(T_1 \underset{c_1=c_2}{\bowtie} T_2\right) \underset{c_3=c_4}{\bowtie} T_3$; . . .

$\boxed{\text{Query complexity of R.A} \in NP}$



Two Practical Optimizations

Ex 1: Two tables, T_1 T_2

$$\sigma_{\substack{c_1 = 50 \\ c_2 = 80K}} \left(\underbrace{T_1}_{\text{Employees}} \bowtie_{c_1 = c_2} \underbrace{T_2}_{\text{Compensation}} \right)$$

```
select age, salary
from Employees join Compensation
where age = 50 and salary = 80K
```

$$\underbrace{\sigma_{\substack{c_3 = 50 \\ c_4 = 80K}}}_{} \underbrace{\left(\underbrace{T_1}_{2000 \text{ tuples}} \bowtie_{c_1 = c_2} \underbrace{T_2}_{2000 \text{ tuples}} \right)}_{30000 \text{ tuples}}$$

Computing this requires 30000 units of time

Computing this requires 2000 log 2000 units of time

$(30000 + 2000 \log 2000)$

$(30000 + 2000 \log 2000)$ time units

$$\sigma_{\substack{c_3=50 \\ c_4=80K}} \left(T_1 \bowtie_{c_1=c_2} T_2\right) = \underbrace{\left(\sigma_{c_3=50} \underset{2000}{T_1}\right)}_{40} \bowtie_{c_1=c_2} \underbrace{\left(\sigma_{c_4=80K} \underset{2000}{T_2}\right)}_{30}$$

$\boxed{\text{Selection pushdown}}$

1200 time units to join

Old query

Total time $= 30000 + 2000 \log 2000$ time units

New Query

Total time $= 2000 + 2000 + 1200$

$= 5200$ time units

Selection pushdown = Distributivity of σ over $\bowtie$.

$$\sigma_f (T_1 \bowtie T_2) = (\sigma\ T_1) \bowtie (\sigma\ T_2)$$

$$\sigma_f(T_1 \bowtie T_2) = (\sigma_{\boxed{f_1}} T_1) \bowtie (\sigma_{\boxed{f_2}} T_2)$$

f_1 = All parts of f which are relevant to T_1

f_2 = All parts of f which are relevant to T_2

Employee

N	Z	O
John	90089	LA
Mark	12345	NY

Office

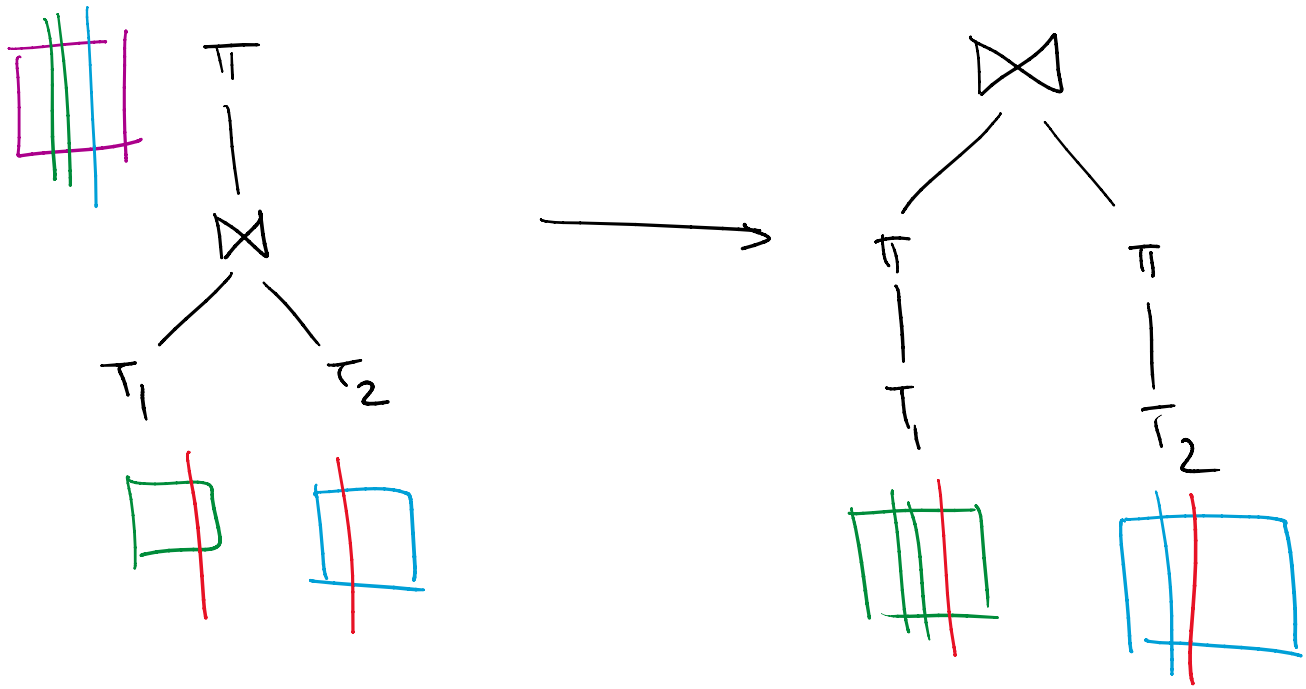
O	Z
LA	90015
NY	12345

$$\sigma_{E.Z \neq O.Z}\left(\text{Employee} \bowtie_{E.O = O.O} \text{Office}\right)$$

Selection pushdown only applicable when predicate independently filters

predicate independently filters the tables

Optimization 2: Projection Pushdowns



Ex2:

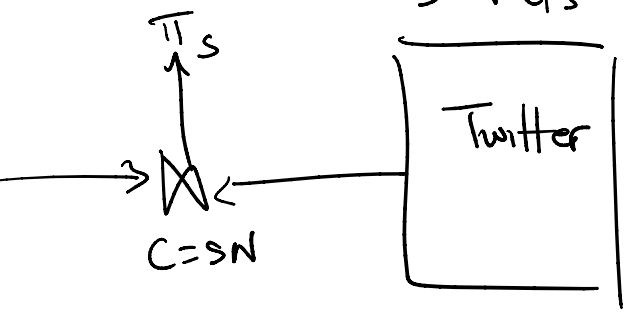
T_1

Company	State
Apple	CA
Microsoft	WA
Oracle	CA
Twitter	CA

T_2

SocNets

Twitter



All states which have the HQ of a social network

C	S
Apple	CA
Microsoft	WA
Oracle	CA
Twitter	CA

$\xrightarrow{\pi_S}$

S
CA
WA

Projections can cause reduction in # of tuples

But, in this case the C column is necessary

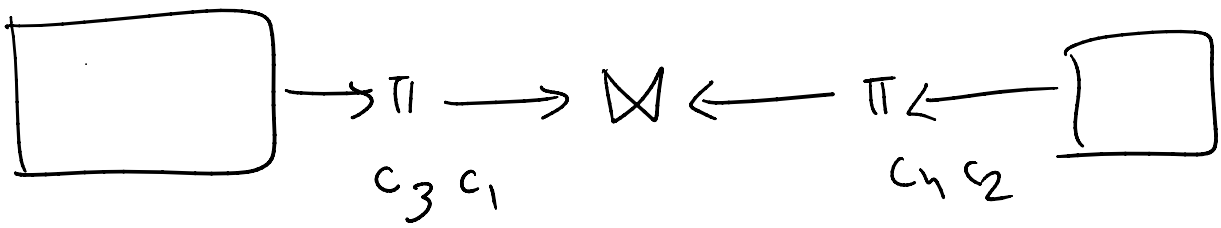
C	O	S
Apple	SF	CA
Microsoft	Redmond	WA
Oracle	SJ	CA
Twitter	PA	CA

$\pi_{c_3\, c_4}$

$\bowtie_{c_1 = c_2}$

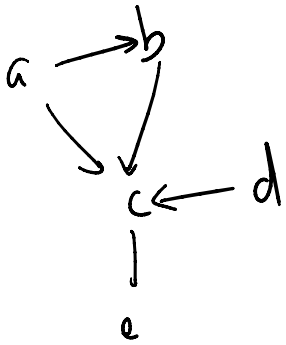
SN

Twitter



Beyond Relational Databases

Question: How do you determine reachability in a graph?



Edge

a	b
a	c
b	c
d	c
c	e

Reachability

a	b
a	c
a	e
b	c
b	e
d	c
d	e
c	e

- All pairs reachability
- Transitive closure

$$\sigma_{14}\left(\left(\sigma_{14}\left(\underbrace{\underbrace{E}_{\text{Paths of length 1}} \bowtie_{E_2 = E_1'} E}_{\text{Paths of length 2}}\right)\right) \bowtie_{2=1'} E\right)$$

Paths of length 3

$\vdots$