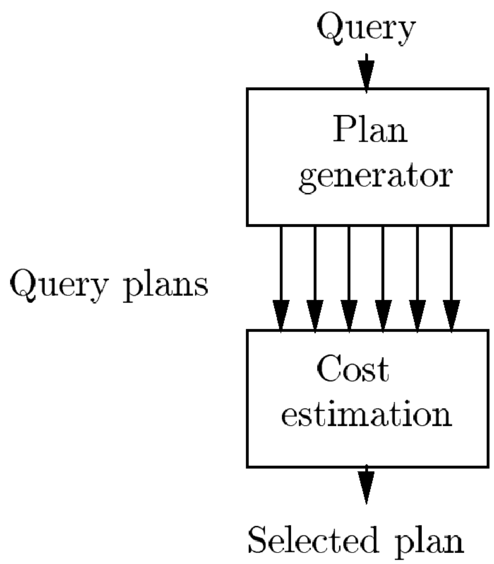
Query Processing

Overview Query Processing



Query Plans

Choose operations, e.g., σ, $\bowtie$

Order operations.

Detailed strategy of operations, e.g.:

- Join method.
- Pipelining: consume result of one operation by another, to avoid temporary storage on disk.
- Use of indexes?
- Sort intermediate results?

We focus on relational systems

Example

Select B,D
From R,S
Where R.A = “c” AND S.E = 2 AND R.C=S.C

R	A	B	C
	a	1	10
	b	1	20
	c	2	10
	d	2	35
	e	3	45

S	C	D	E
	10	x	2
	20	y	2
	30	z	2
	40	x	1
	50	y	3

Answer

B	D
2	x

Implementation of DBMS

How Do We Execute the Query?

One idea

- Do Cartesian product
- Select tuples
- Do projection

WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Implementation of DBMS

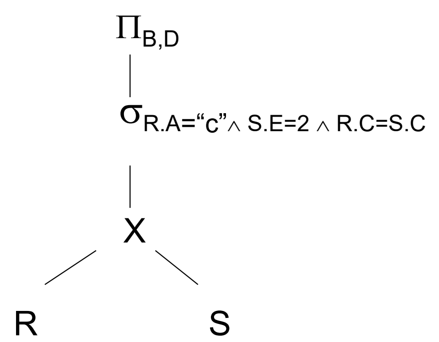
RXS	R.A	R.B	R.C	S.C	S.D	S.E
	a	1	10	10	x	2
	a	1	10	20	y	2
	.					
	.					
Bingo! Got one...	C	2	10	10	x	2
	.					
	.					

WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Implementation of DBMS

Relational Algebra to Describe Plans

Ex: Plan I

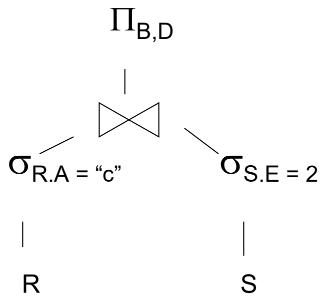


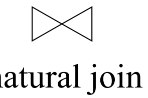
OR: $\Pi_{B,D} [\sigma_{R.A=\text{“c”} \wedge S.E=2 \wedge R.C = S.C} (RxS)]$

WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Another Plan

Plan II





R

A	B	C
a	1	10
b	1	20
c	2	10
d	2	35
e	3	45

σ (R)

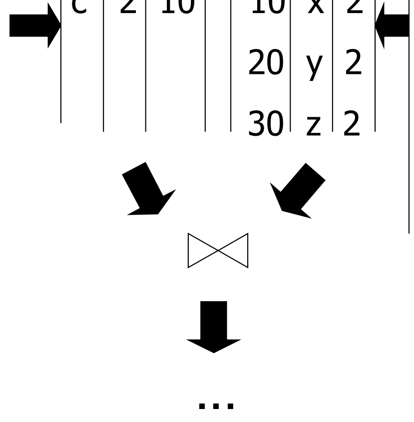
A	B	C
c	2	10

σ(S)

C	D	E
10	x	2
20	y	2
30	z	2

S

C	D	E
10	x	2
20	y	2
30	z	2
40	x	1
50	y	3



Plan III

Use R.A and S.C Indexes

(1) Use R.A index to select R tuples with R.A = "c"

(2) For each R.C value found, use S.C index to find matching tuples

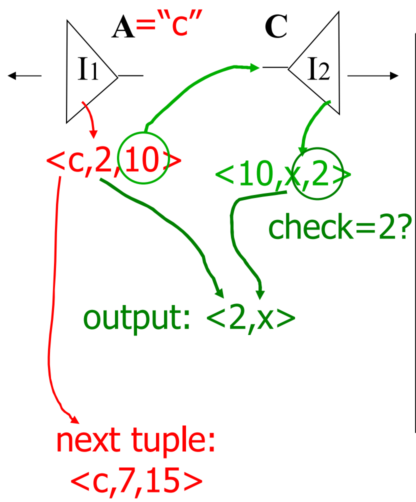
(3) Eliminate S tuples with S.E ≠ 2

(4) Join matching R,S tuples,

(5) Project B,D attributes and place in result

R

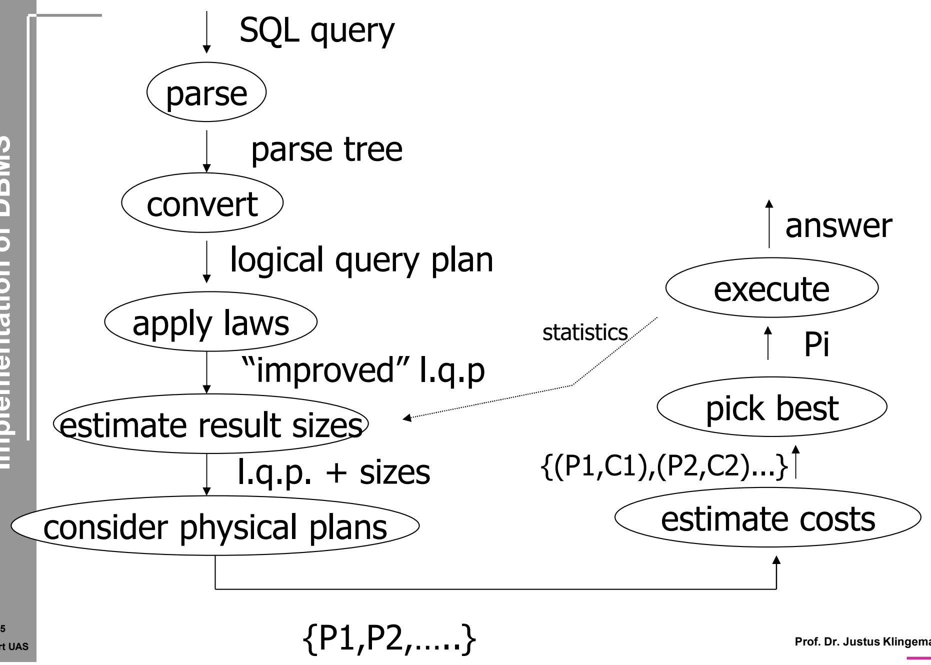
A	B	C
a	1	10
b	1	20
c	2	10
d	2	35
e	3	45



S

C	D	E
10	x	2
20	y	2
30	z	2
40	x	1
50	y	3

Overview of Query Optimization

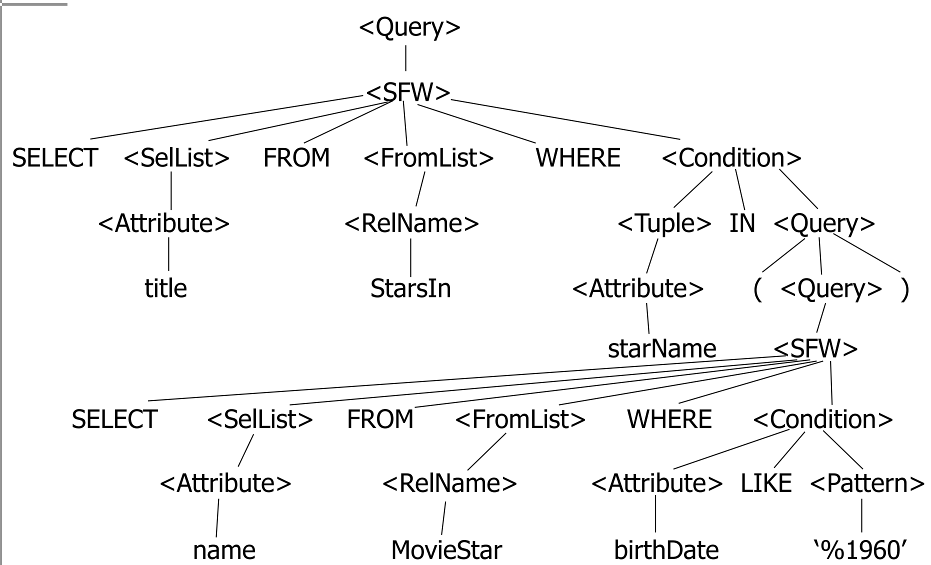


Example: SQL Query

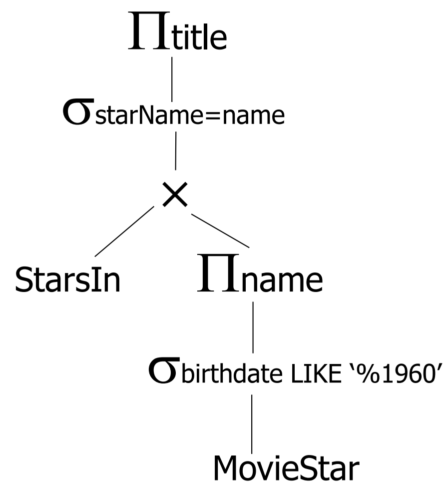
```
SELECT title
FROM StarsIn
WHERE starName IN (
        SELECT name
        FROM MovieStar
        WHERE birthdate LIKE '%1960'
);
```

(Find the movies with stars born in 1960)

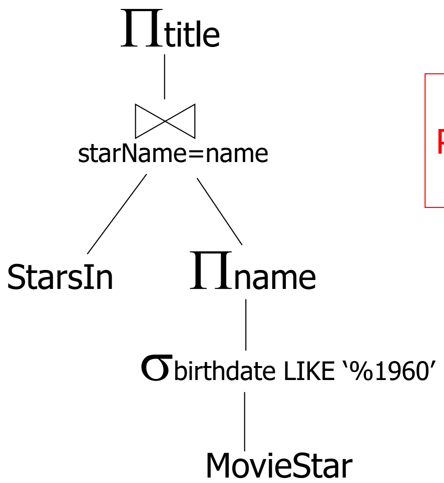
Example: Parse Tree



Example: Logical Query Plan



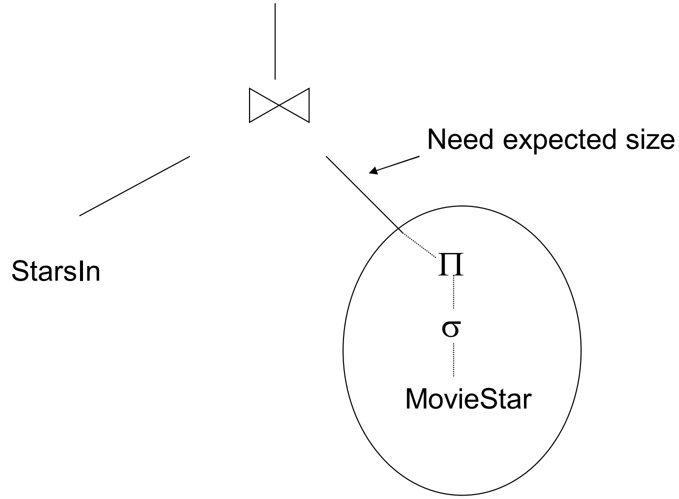
Example: Improved Locical Query Plan



Question:
Push project to
StarsIn?

Implementation of DBMS

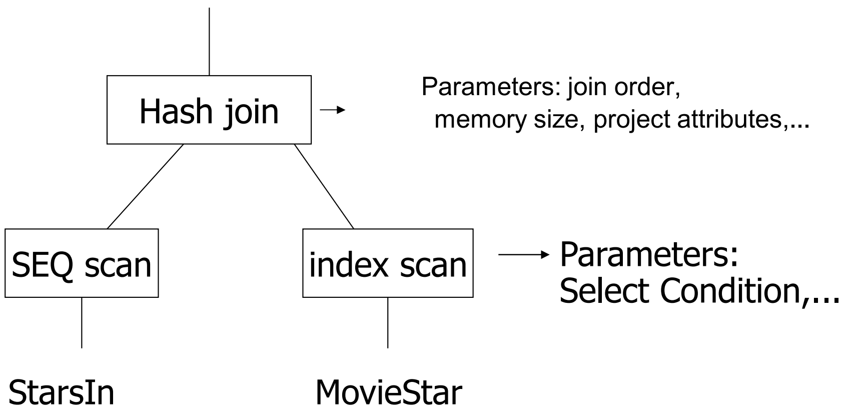
Example: Estimate Result Sizes



WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Implementation of DBMS

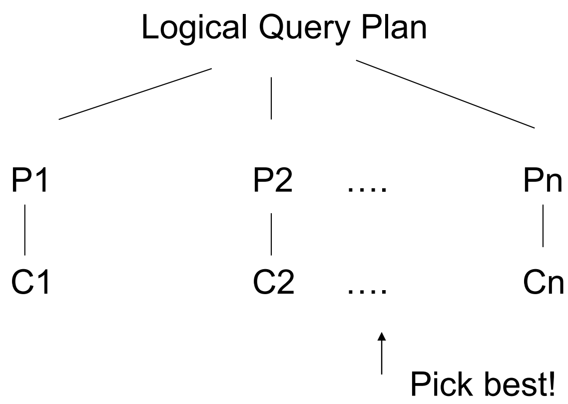
Example: One Physical Plan



WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Implementation of DBMS

Example: Estimate Costs



WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Algebraic Transformations

Generating Plans

- Start with query definition.
  - A plan, but usually a terrible one.
- Apply algebraic transformations to find other plans.
- Relational algebra is a good start, but we need also to consider: GROUP BY, duplicate elimination, HAVING, ORDER BY.

Algebraic Transformations

- Rules give equivalent expressions. meaning that whatever relations are substituted for variables, the results are the same.

Rules: Natural joins & cross products & union

$R \bowtie S \quad = \quad S \bowtie R$

$(R \bowtie S) \bowtie T \quad = R \bowtie (S \bowtie T)$

$R \times S = S \times R$

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

$R \cup S = S \cup R$

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

But beware of thetajoin (join condition different from =)

- associative law does not hold.

Rules: Selects

$\sigma_{p1 \wedge p2}(R) \quad = \quad \sigma_{p1}\ [\sigma_{p2}\ (R)]$

$\sigma_{p1 \vee p2}(R) \quad = \quad [\sigma_{p1}\ (R)] \cup [\sigma_{p2}\ (R)]$

Rules: Project

Let: X = set of attributes

Y = set of attributes

XY = X U Y

$\pi_{xy}\ (R) \quad = \quad \cancel{\pi_x\ [\pi_y\ (R)]}$

Rules: $\sigma + \bowtie$ combined

Let p = predicate with only R attributes
q = predicate with only S attributes
m = predicate with only R,S attributes

$$\sigma_p (R \bowtie S) = [\sigma_p (R)] \bowtie S$$

$$\sigma_q (R \bowtie S) = R \bowtie [\sigma_q (S)]$$

Implementation of DBMS

Rules: $\sigma + \bowtie$ combined

Some rules can be derived:

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

$$\sigma_{p \wedge q \wedge m} (R \bowtie S) = \sigma_m \left[(\sigma_p R) \bowtie (\sigma_q S)\right]$$

$$\sigma_{p \vee q} (R \bowtie S) = \left[(\sigma_p R) \bowtie S\right] \cup \left[R \bowtie (\sigma_q S)\right]$$

WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann

Implementation of DBMS

Rules: π, σ combined

Let x = subset of R attributes
z = attributes in predicate P
(subset of R attributes)

$$\pi_x[\sigma_p (R)] = \pi_x \{\sigma_p [\overset{\pi_{xz}}{\cancel{\pi_x}} (R)]\}$$

WS 24/25
Frankfurt UAS
Prof. Dr. Justus Klingemann