

# CSE 562 Database Systems

## Query Processing: Overview

Some slides are based or modified from originals by  
*Database Systems: The Complete Book,*  
Pearson Prentice Hall 2<sup>nd</sup> Edition  
©2008 Garcia-Molina, Ullman, and Widom

*cse@buffalo*

## Outline – Query Optimization

- **Overview**
- Relational algebra level
  - Algebraic Transformations
- Detailed query plan level
  - Estimate Costs
    - Estimating size of results
    - Estimating # of IOs
  - Generate and compare plans

## Query Processing

- The query processor turns user queries and data modification commands into a query plan – a sequence of operations (or algorithms) on the database
  - from high level queries to low level commands
- Decisions taken by the query processor:
  - Which of the algebraically equivalent forms of a query will lead to the most efficient algorithm?
  - For each algebraic operator, what algorithm should we use to run the operator?
  - How should the operators pass data from one to the other? (e.g., main memory buffers, disk buffers)

## Example

```
SELECT R.B, S.D  
FROM R, S  
WHERE R.A = 'c' AND S.E = 2 AND R.C = S.C
```

## Example (cont.)

R	A	B	C	S	C	D	E
a	1	10	10	x	2		
b	1	20	20	y	2		
c	2	10	30	z	2		
d	2	35	40	x	1		
e	3	45	50	y	3		

**Answer**

B	D
2	x

UB CSE 562 Spring 2009

5

## Query Processing

How do we execute a query eventually?

One idea

- Scan relations
- Do Cartesian product
- Select tuples
- Do projection

UB CSE 562 Spring 2009

6

## Example (cont.)

R x S	R.A	R.B	R.C	S.C	S.D	S.E
a	1	10	10	x	2	
a	1	10	20	y	2	
		⋮				
c	2	10	10	x	2	
		⋮				

Bingo! → Got one...

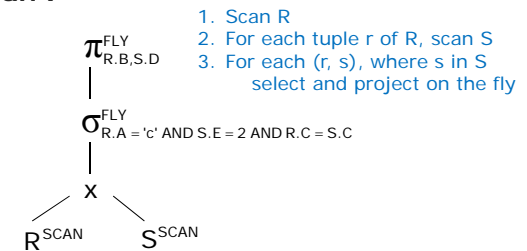
UB CSE 562 Spring 2009

7

## Example (cont.)

Relational algebra can be enhanced to describe plans

Example: **Plan I**



OR:  $\pi_{B,D}^{FLY} \left[ \sigma_{R.A='c' \text{ AND } S.E=2 \text{ AND } R.C=S.C}^{FLY} (R^{SCAN} \times S^{SCAN}) \right]$

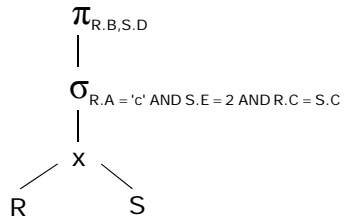
UB CSE 562 Spring 2009

8

## Example (cont.)

FLY and SCAN are the defaults

Example: **Plan I**



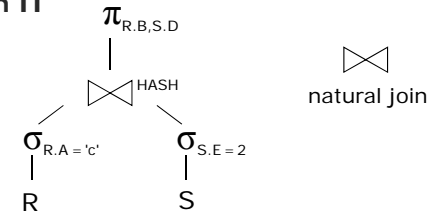
UB CSE 562 Spring 2009

9

## Example (cont.)

Another Idea

Example: **Plan II**

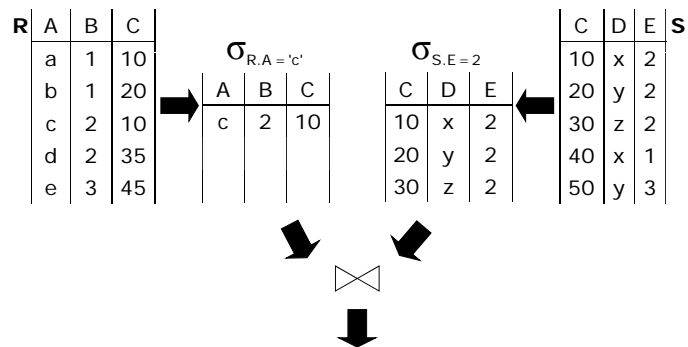


1. Scan R and S
2. Perform on the fly selections
3. Do hash join
4. Project

UB CSE 562 Spring 2009

10

## Example (cont.)



UB CSE 562 Spring 2009

11

## Example (cont.)

Example: **Plan III**

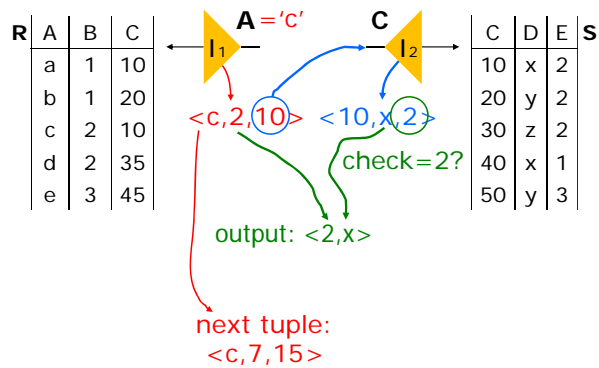
Use R.A and S.C Indexes

- (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 join tuples
- (3) Eliminate join tuples for which S.E  $\neq$  2
- (4) Project B,D attributes

UB CSE 562 Spring 2009

12

## Example (cont.)

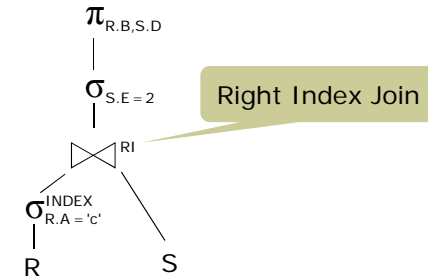


UB CSE 562 Spring 2009

13

## Example (cont.)

### Algebraic Form of Plan



UB CSE 562 Spring 2009

14

## From Query To Optimal Plan

- Complex process
- Algebra-based logical and physical plans
- Transformations
- Evaluation of multiple alternatives

UB CSE 562 Spring 2009

15

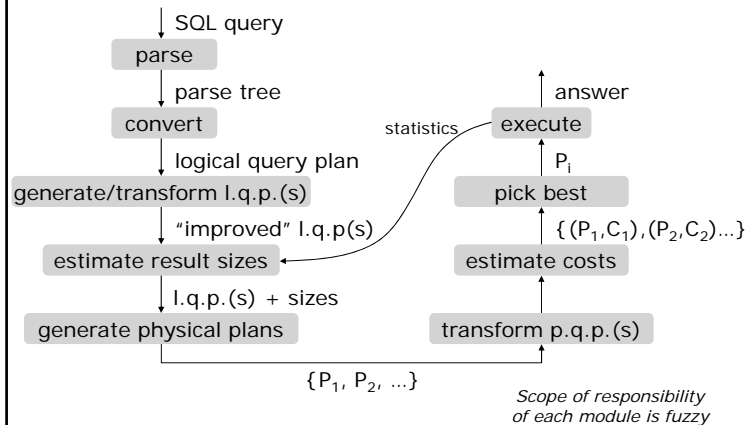
## Issues in Processing and Optimization

- Generate Plans
  - Employ efficient execution primitives for computing relational algebra operations
  - Systematically transform expressions to achieve more efficient combinations of operators
- Estimate Cost of Generated Plans
  - Statistics
- “Smart” Search of the Space of Possible Plans
  - always do the “good” transformations (relational algebra optimization)
  - prune the space (e.g., System R)
- Often the above steps are mixed

UB CSE 562 Spring 2009

16

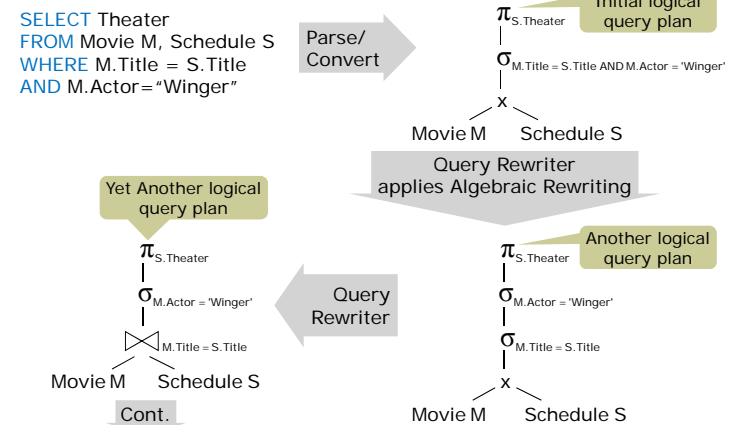
## Architecture



UB CSE 562 Spring 2009

17

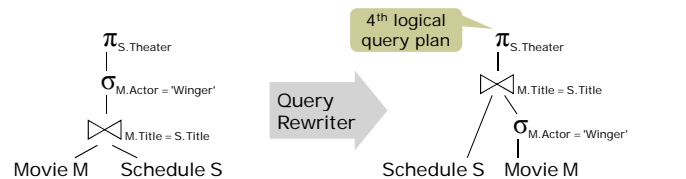
## Example Journey of a Query



UB CSE 562 Spring 2009

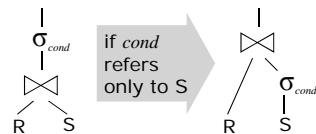
18

## Example Journey of a Query (cont.)



### Summary of Logical Plan Generator

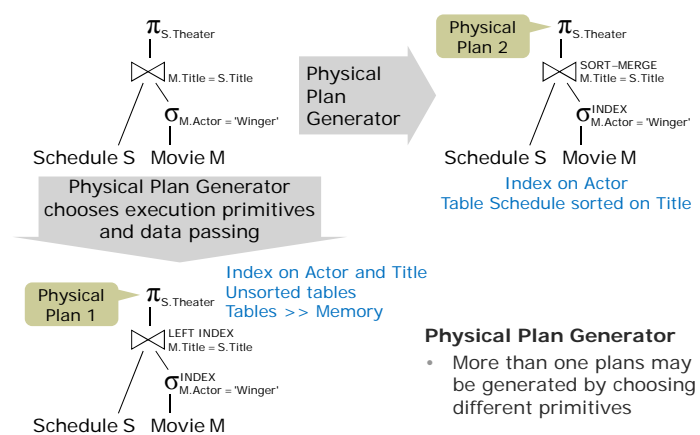
- 4 logical query plans created
- algebraic rewritings were used for producing the candidate logical query plans
- the last one is the winner (at least, cannot be a big loser)
- in general, multiple logical plans may "win" eventually



UB CSE 562 Spring 2009

19

## Example Journey of a Query (cont.)



### Physical Plan Generator

- More than one plans may be generated by choosing different primitives

UB CSE 562 Spring 2009

20

## Example: Nested SQL Query

```

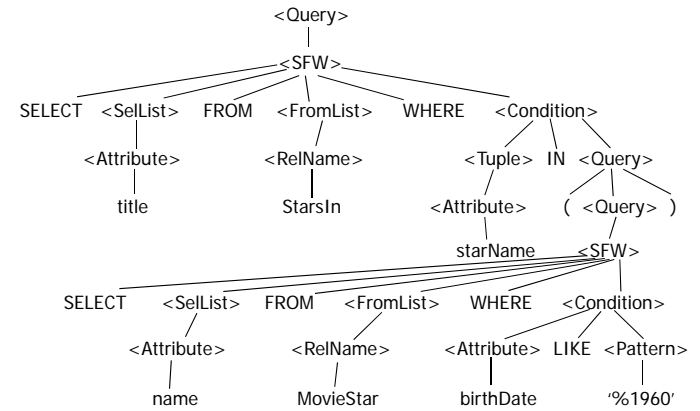
SELECT title
FROM StarsIn
WHERE starName IN
  (SELECT name
   FROM MovieStar
   WHERE birthdate LIKE '%1960')
    
```

Find the movies with stars born in 1960

UB CSE 562 Spring 2009

21

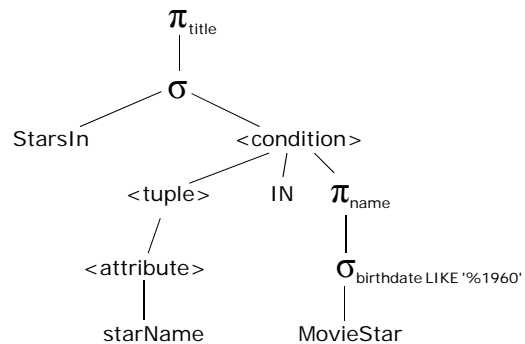
## Example: Parse Tree



UB CSE 562 Spring 2009

22

## Example: Generating Relational Algebra

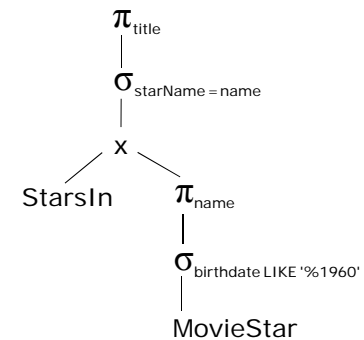


- An expression using a two-argument  $\sigma$ , midway between a parse tree and relational algebra

UB CSE 562 Spring 2009

23

## Example: Logical Query Plan

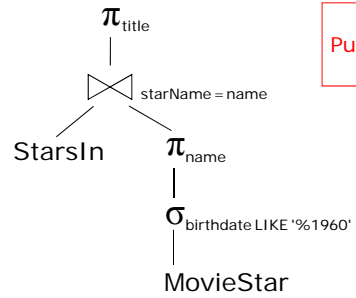


- May consider "IN" elimination as a transformation in the query rewriter or may consider it a task of the converter

UB CSE 562 Spring 2009

24

## Example: Improved Logical Query Plan



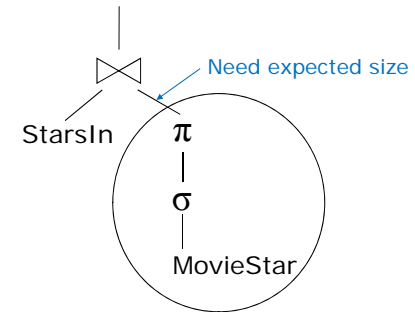
**Question**  
Push project to  
StarsIn?

UB CSE 562 Spring 2009

25

## Example: Estimate Result Sizes

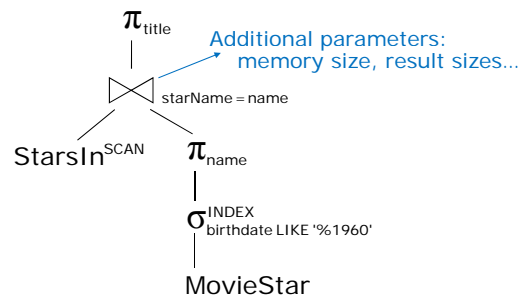
Result sizes are important for selecting physical plans



UB CSE 562 Spring 2009

26

## Example: One Physical Plan



UB CSE 562 Spring 2009

27