

# An Overview of Query Optimization in Relational Systems

Presenter: Dutch Meyer  
Discussion: Mike "Debo" DiBernardo

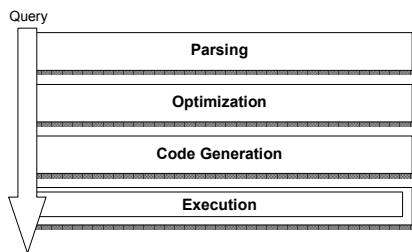
Somewhat based on slides from Albert Wong, modified by Rachel Pottinger.

## Key Points

- Query optimization in relation to RDBMS
- Query optimization strategy
- Statistical modeling considerations

Goal: Sampling of query optimization as of 1998

## Query handling



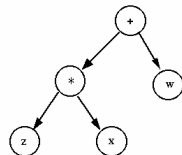
Remember: Our Query is declarative!

## Having a Plan: Optimization Strategy

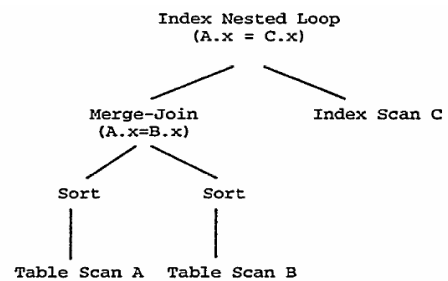
- A space of plans (search space)
- An enumeration algorithm
- A cost estimation technique

## Operator Trees

- General Concept:
  - Operat-ors are nodes
  - Operat-ees are leaves

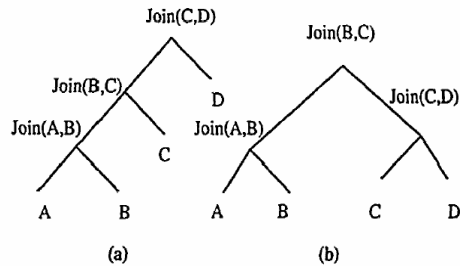


## Applied to Queries



## Join Sequences

### ■ Linear Vs Bushy



Statistics!

## Statistics and Cost

- Statistics vs. Cost
- Resources at our disposal
  - CPU, IO, memory, bandwidth...
- Summarize the data
  - Tuple counts
  - Physical Pages
  - Column information
- Statistic collecting is an interesting problem

## Determining cost

- Considering an operator and its input
- Histograms
  - Information on column values for predicate matching
- Information Propagation

## Cost Computation

- Translating into CPU
- Translating into I/O
- Other resources
  - Bandwidth/Communication
  - Buffer Utilization?

## Key Contributions

- Elements of a search strategy
- Operator Trees
- Cost computations and statistics