Plan Generation

- selection of physical access path and [[Plan Operators]]
- selection of execution order
- convert logical query plan into optimal physical query plan
- cost query optimization must be less than actual improvements

Cost Models

- Relies on statistics (cardinalities, selectivities via histograms + estimators)
- Operator-specific and general-purpose cost models

$$C_{\text{out}}(T) = \begin{cases} 0 & \text{if } T \text{ is a single relation} \\ |T| + C_{\text{out}}(T_1) + C_{\text{out}}(T_2) & \text{if } T = T_1 \bowtie T_2 \end{cases}$$

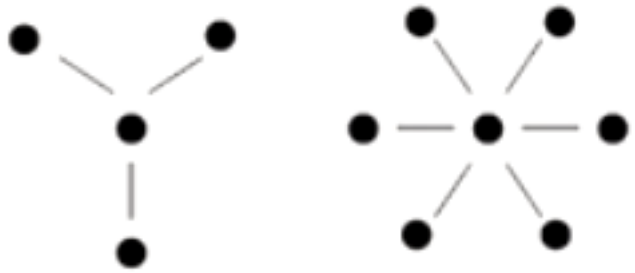
- **I/O costs** (number of read pages, tuples)
- **Computation costs** (CPU costs, path lengths)
- **Memory** (temporary memory requirements)
- **Beware assumptions of optimizers** (no skew, independence, no correlation)

  - do not consider skew and [[Correlation]]

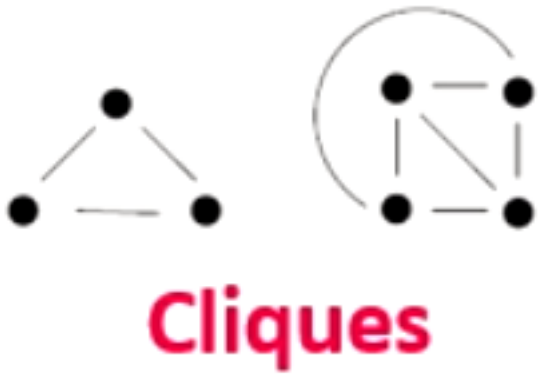
Query Types

- nodes: tables
- edges: join conditions
- hardness depends on structure
- types:
  - chains
    - few join orders

    -
  - stars
    - central table
    - outer tables add information
    - almost no alternative join orders



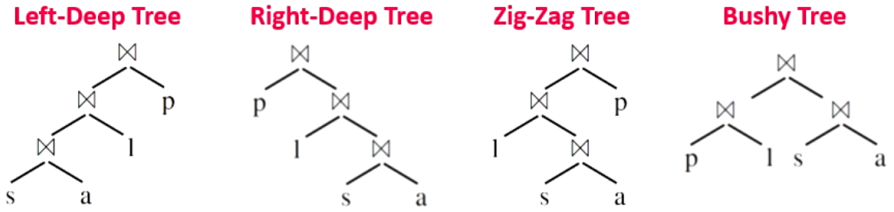
  - *
- cliques
  - * lots of different join orders
  - * difficult to calculate



  - *

Join Tree/Plan Types

- data flow graph of tables and joins
- edges data dependencies
- types:



  -