Asymptotically Better Query Optimization Using Indexed Algebra

Philipp Fent, Guido Moerkotte, Thomas Neumann
fent@in.tum.de

Motivation

Complex queries on small workloads are common. E.g., BigQuery/Tableau: 90% of queries process < 100MB

Optimization time is super-linear with algebra depth!

Algebra trees can have $O(n)$ Depth. Especially before Optimization.

E.g., TPC-DS Q64 →

Query Optimization

Analyze algebra to find optimization opportunities

Operator-centric $O(n^2)$

Path-centric $O(n \log(n))$ with Indexed Algebra

Indexed Algebra

Index of paths through the algebra

$O(n \log(n))$ index height

Balanced binary index of the path from B⁶ to the root

Represented algebra plan

Link/cut trees make indexing efficient

Evaluation

Asymptotic behaviour

Expression evaluation

Cardinality estimation

Outer-Join nullability

Query decorrelation

Implementation

Indexed Algebra

Path Traversal

Column Sets

Optimization time