Adaptive Recursive Query Optimization

Anna Herlihy, Guillaume Martres, Anastasia Ailamaki, Martin Odersky

Adaptive Meta-programming in Datalog

Datalog is a logic-based programming language that has been used in Java program analysis, TensorFlow, Rust compilation, Ethereum VM, AWS network analysis, and other performance-critical applications.

Adaptive Meta-programming uses Multi-Stage Programming to continuously reoptimize Datalog via phases of compile and runtime code generation, so the optimizer can adapt to new information as it becomes available.

Advanced program analysis requires solving constraint systems with complex, recursive interdependencies.

Running Example: Graspan’s Context-Sensitive Pointer Analysis (CSPA) on Apache httpd

The Carac compiler uses Adaptive Meta-programming to partially evaluate the input Datalog program and continuously regenerate specialized and parallelized imperative programs.

Carac compiles Datalog to an IR using a Futamura Projection and the bottom-up Semi-Naive algorithm. The IR is lowered to various targets, including:

(1) Scala 3 Quotes and Splices
(2) JVM bytecode
(3) Higher-order lambda functions

Where possible, JIT compilation is combined with macros to push expensive optimizations offline.

Up to 5000x speedup over unoptimized queries and 6x over hand-optimized queries using adaptive metaprogramming.