

Publication Review - Bao: Making Learned Query Optimization Practical

Authors:

- Ryan Marcus, MIT & Intel Labs ryanmarcus@csail.mit.edu
- Nesime Tatbul, MIT & Intel Labs tatbul@csail.mit.edu
- Parimarjan Negi, MIT pnegi@csail.mit.edu
- Mohammad Alizadeh, MIT alizadeh@csail.mit.edu
- Hongzi Mao, MIT hongzi@csail.mit.edu
- Tim Kraska, MIT kraska@csail.mit.edu

Presented at [SIGMOD '21](#), June 20–25, 2021, Virtual - Event, China

Reviewed for ITEC-6220 by: Firoz Kabir

Before We Get Started

- Practitioner's Lens
- Traditional Process
- Credit for Diagrams

Traditional Optimization Process

- Identify slow sql
- Run explain plan
- Add hints to sql code

Learned Query Optimization

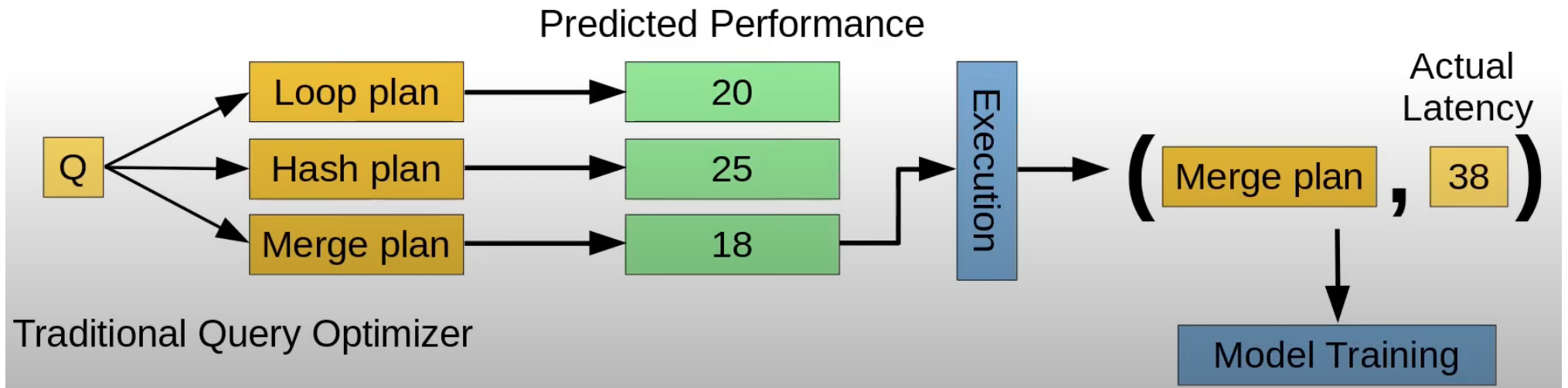
- Predict Performance
- Observe Actual
- Retrain

Problem Description

- Long training time
- Inability to adjust to data and workload changes
- Tail catastrophe
- Black-box decisions
- Integration cost

Deeper look at the problem

- Predict Performance
 - Inductive Bias
 - Tree Convolution
- Training
 - Multi-Arm Bandit Problem (Exploration vs Exploitation)
 - Thompson Sampling



The Bao difference:

- Short training time
- Robustness to schema, data, and workload changes
- Better tail latency
- Interpretability and easier debugging
- Low integration cost
- Extensibility

Bao System Model:

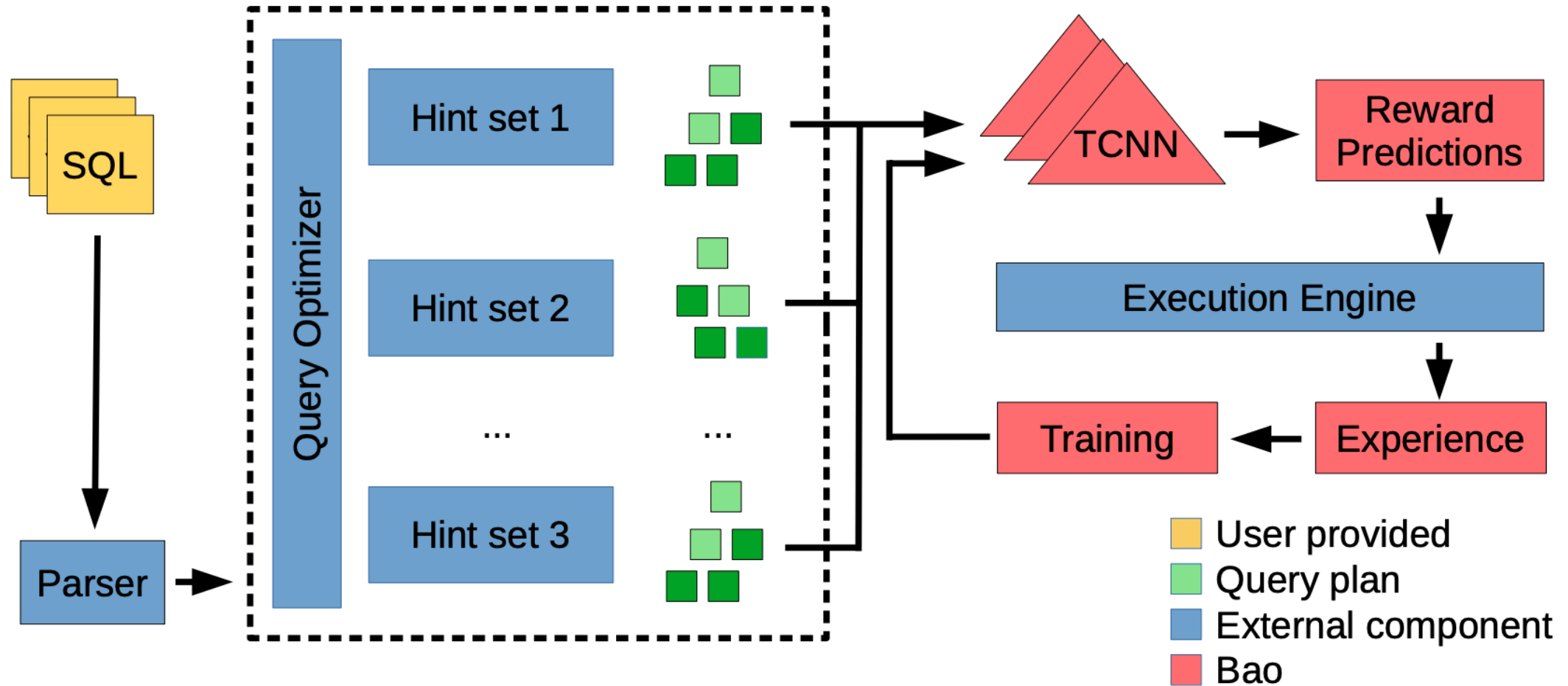


Figure 2: Bao system model

Bao Prediction Model:

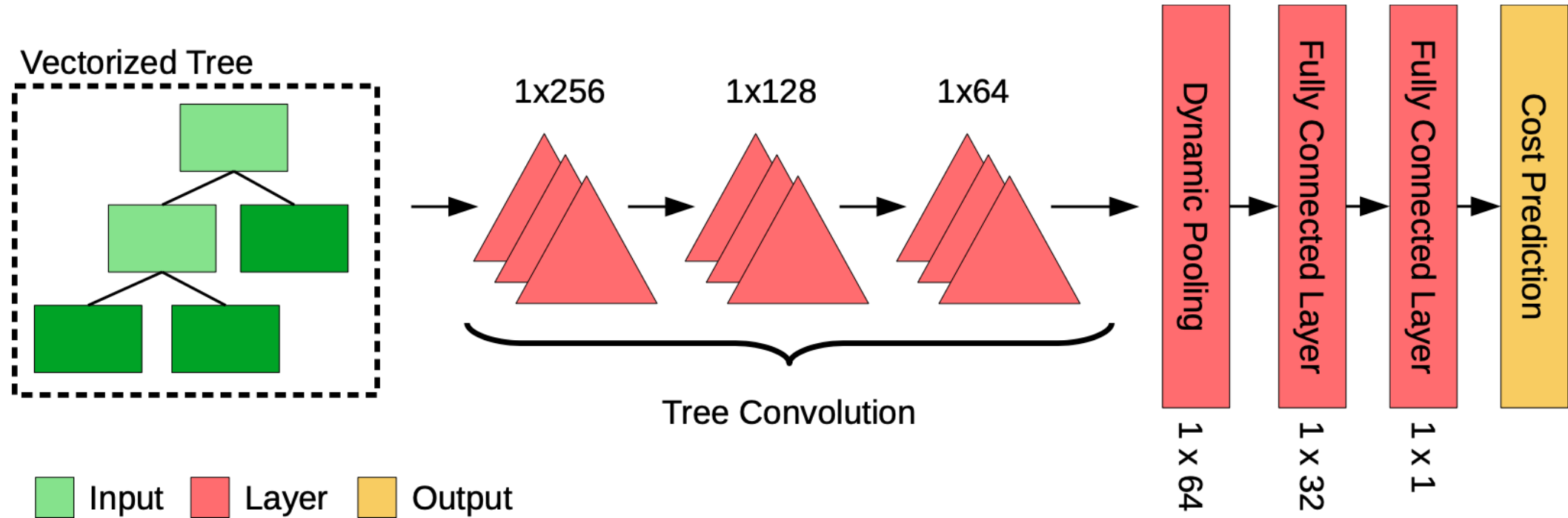
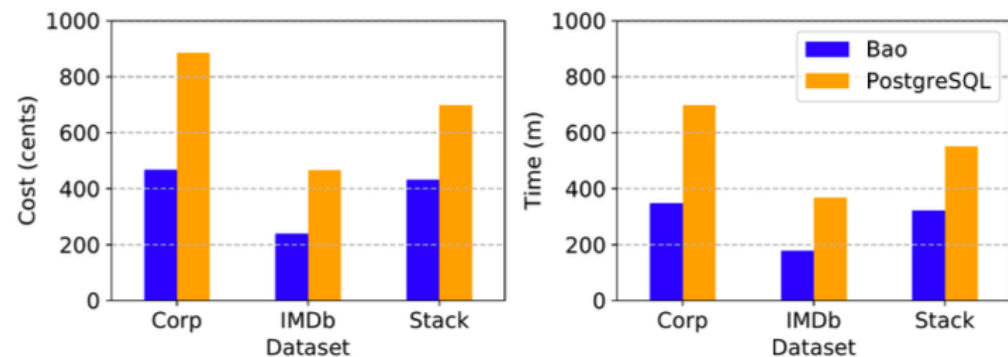
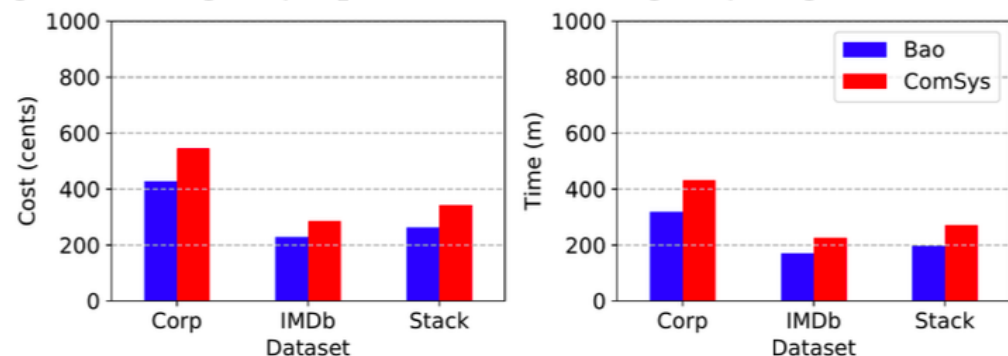


Figure 5: Bao prediction model architecture

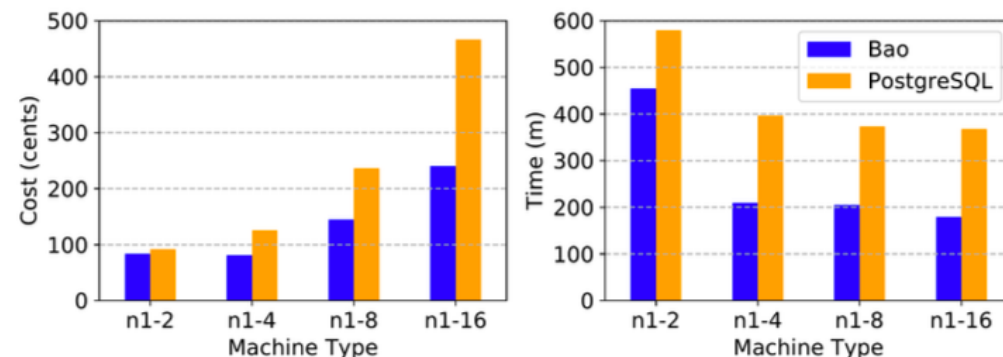


(a) Across our three evaluation datasets, Bao on the PostgreSQL engine vs. PostgreSQL optimizer on the PostgreSQL engine.

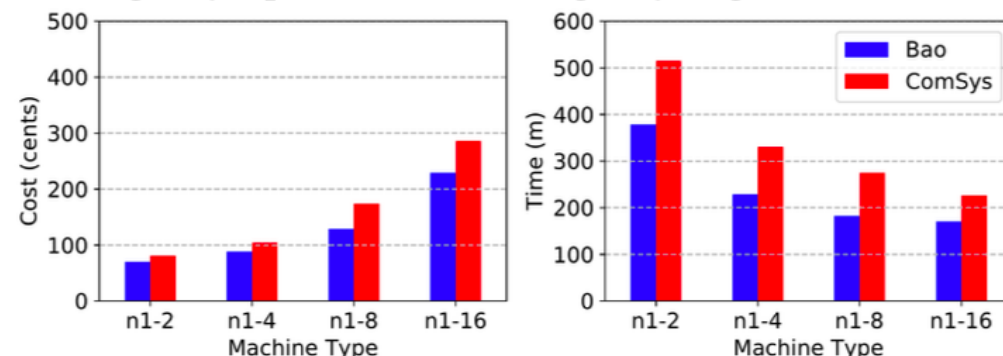


(b) Across our three evaluation datasets, Bao on the ComSys engine vs. ComSys optimizer on the ComSys engine.

Figure 7: Cost (left) and workload latency (right) for Bao and two traditional query optimizers across three different workloads on a N1-16 Google Cloud VM.



(a) Across four different VM types, Bao on the PostgreSQL engine vs. PostgreSQL optimizer on the PostgreSQL engine.



(b) Across four different VM types, Bao on the ComSys engine vs. ComSys optimizer on the ComSys engine.

Figure 8: Cost (left) and workload latency (right) for Bao and two traditional query optimizers across four different Google Cloud Platform VM sizes for the IMDb workload.

Questions ?