Motivation

- Database Systems handle concurrent execution of queries
  - But, query optimization is based on individual cost evaluation
- Query interactions impact the performance of the system significantly.
  - Negative Interactions
  - Positive Interactions
- Considering query interactions improves performance and scheduling options
Report Generation

QShuffler

Supports Concurrent Execution of Queries!

Arrival Queue

Cost Model

Scheduled Query Mix

DATA

DBMS

Multiple Queries

Supports Concurrent Execution of Queries!

DATA

DBMS
Cost Metric

Normalized Runtime Overhead (NRO)

\[ NRO_i = \frac{1}{M^2} \sum_{j=1}^{T} N_{ij} \frac{A_{ij}}{t_j} \]

QShuffler Challenges

- Multiprogramming Level (MPL)
  - Static Vs Dynamic
- Regression Model Selection
  - Linear Vs Quadratic Features
- Regression Model Training
  - Off-line Vs On-line Training
QShuffler Challenges

- Multiprogramming Level (MPL)
  - Static Vs Dynamic

- Regression Model Selection
  - Linear Vs Quadratic Features

- Regression Model Training
  - Off-line Vs On-line Training

Static vs. Dynamic MPL

**Static:**
- Too High: data contention thrashing
- Too Low: under-utilization of resources
- Training sampling over single MPL

**Dynamic:**
- Avoids System Overload
- Ensures smart utilization of resources
- Training sampling over multiple MPLs
Algorithm Modification

Arrival Queue

QShuffler

Active Queries

Algorithm Modification

Arrival Queue

QShuffler

Active Queries
Completion Times

QShuffler Challenges

- Multiprogramming Level (MPL)
  - Static Vs Dynamic

- Regression Model Selection
  - Linear Vs Quadratic Features

- Regression Model Training
  - Off-line Vs On-line Training
Linear Vs Quadratic Model

**Linear Model**
- Less Accurate
- Less Overhead
- Less Training Sampling

\[ \overline{NRO}_i = \sum_{j=1}^{T} \beta_j N_{ij} \]

**Quadratic Model**
- More Accurate
- More Overhead
- More Training Sampling

\[ \overline{NRO}_i = w_0 + \sum_{j=1}^{T} w_j N_{ij} + \sum_{j=1}^{T} v_j N_{ij}^2 \]

Variable MPL – Linear Fit

![Graph showing the linear fit for variable MPL](image)
Variable MPL – Quadratic Fit

Completion Times

Workload Size/Type
Linear Model
Quadratic Model
QShuffler Challenges

- Multiprogramming Level (MPL)
  - Static Vs Dynamic
- Regression Model Selection
  - Linear Vs Quadratic Features
- Regression Model Training
  - Off-line Vs On-line Training

Gradient Descent

- On-line Calculation of $NRO$ for mix $I$
  - “Average” mix running over a period of time
- Minimize the Error:
  \[ E_i = \left( NRO_i - \overline{NRO}_i \right)^2 \]
- Update to Weights:
  \[ w_j \leftarrow w_j + \alpha N_{ij} \left( NRO_i - \overline{NRO}_i \right) \]
Completion Times

![Graph showing completion times for different workloads and training methods.]

Future Investigation

- Selecting Load Threshold
  - Static Vs Dynamic
- Examining On-line Update Model
  - Exploration Vs Exploitation
- Integrating Scheduler into a Database System
  - Black Box → White Box