



教育经历

北京邮电大学

2020 - 2024

本科 计算机学院(国家示范性软件学院) 计算机科学与技术

专业排名: 36/425(8%)

英语四级成绩644, 六级成绩580

中国人民大学

2024 - 至今

博士研究生 信息学院 计算机科学与技术

研究方向: AI4DB 智能查询优化技术

荣誉奖项

校级三好学生

2020-2021 学年

校级二等奖学金

2020-2021 学年

校级三等奖学金

2021-2022 学年, 2022-2023 学年

第13届全国大学生数学竞赛三等奖, 北京市二等奖

2021年6月

博士研究生学业奖学金一等奖

2024-2025 学年, 2025-2026 学年

博士研究生学习优秀奖学金三等奖

2025-2026 学年

项目经历

本科毕业设计: 基于深度学习的分布式查询优化器系统的设计与实现

2023年9月 - 2024年06月

- 基于优化规则开关的“提示”来辅助和指导Spark SQL的查询优化器。
- 实现了一种基于贪心算法的提示集搜索算法和两种基于深度学习的代价模型。
- 在TPC-DS, TPC-H和JOB数据集上取得了1.05-1.19的加速比。

华为横向项目: 大数据SQL优化器技术合作项目

项目负责人

2024年3月 - 2025年3月

- 在毕业设计的“提示”优化基础上, 在Spark SQL中进一步集成了查询改写, 基数估计等多种优化技术。
- 在TPC-DS SF3000数据集上以华为提供的一组基线配置作为参照取得了20%的加速效果。

研究论文

- 何家豪, 王嘉辰, 王晓, 张喜盈, 李翠平, 陈红.**智能查询优化算法研究综述. 软件学报, 2026, 37(1): 279-300 (CCF-A 中文期刊)
- Jiahao He, Yutao Cui, Cuiping Li, Jikang Jiang, Yuheng Hou, Hong Chen.** LQRS: Learned Query Re-optimization Framework for Spark SQL. (VLDB 在投)
动机: 现有学习型查询优化器遵循“先优化、后执行”的范式, 无法利用执行过程中产生的真实运行时信息, 限制了进一步的性能提升。
设计: 本文提出 LQRS, 一个基于 Spark SQL 的学习型查询重优化框架, 将强化学习与 Spark AQE 技术结合, 支持在执行前与执行中基于运行时观测动态调整执行计划。
结果: 在多个基准上, LQRS 相比现有学习型优化器和重优化方法可将端到端查询执行时间最多降低约 90%。
- Jiahao He, Cuiping Li, Hong Chen.** Calibra: An Unified Learning Framework for End-to-End Query Optimization in Spark SQL. (在研)
动机: 现有学习型查询优化方法往往针对单一优化阶段独立设计与训练, 导致跨阶段信息无法协同利用, 且模型设计与训练存在明显冗余。
设计: 本文提出 Calibra, 一个统一的端到端学习框架, 通过统一计划编码、Tree Linear RNN 计划表示模型以及多级反馈训练机制, 在不扩展原生搜索空间的前提下, 对 Spark SQL 各优化阶段的决策进行协同校准。
- Jikang Jiang, Jiahao He, Zihang Qiu, Cuiping Li, Yuheng Hou, Yutao Cui.** LEGO: Learning Graph For Operator-Level Cost Estimation. (SIGMOD 在投)
本文提出 LEGO, 通过将查询计划建模为算子级计算图并利用图神经网络学习算子之间的结构与依赖关系, 实现精细化的算子级执行代价估计。