

徐智涛

联系电话: 15910985025

电子邮件: littlexzt@126.com

微信: littlexzt

应聘岗位 机器学习算法工程师

工作及项目经历

2017.7-至今 软件开发工程师 北京谷数科技有限公司

- ◆ 工作期间共参与公司项目 4 个, 在项目中主要作为开发人员提供技术支持
- ◆ 阅读分布式大数据 SQL 引擎 presto 源代码, 理解其架构并形成文档贡献在公司文档库中。
- ◆ 参与公司大规模分布式并行处理数据库研发, 并在某客户单位数据库性能测试中性能排名第一
- ◆ 开发某小文件定期管理项目的部分程序代码编写, 为项目系统稳定运行提供支持
- ◆ 数据库的 GPU 计算能力的拓展研发, 目前仍在进行中

工作项目

2017.10-至今 核心研发人员 数据库 GPU 计算能力拓展

- ◆ 研究数据库从 SQL 语句到查询计划树的源代码, 通过在数据库的查询计划树的相应节点下插入自定义扫描节点, 在不改变数据库原本架构下, 为 GPU 计算的插入提供插入点;
- ◆ 将相关计算任务规划到自定义扫描节点, 并在自定义扫描节点初始化时动态生成 cuda 代码, 执行时执行 cuda 代码并将结果返回给 CPU
- ◆ 在复杂计算 SQL (如聚合、join 等) 中, 数据库查询性能提升 10 倍以上;
- ◆ 通过保留执行完 SQL 的数据以及 GPUContext, 缓存数据, 当再次执行的 SQL 的数据已被缓存, 进一步提高数据库查询性能

2017.9-2017.11 核心研发人员 某部队单位数据库性能测试

- ◆ 对大规模并行处理 (MPP) 数据库集群环境进行搭建;
- ◆ 通过对 CPU 和磁盘 IO 性能的分析, 修改数据传输所启动的进程数目以及 segment 的数目, 极大地提高了数据并行导入数据库的性能, 最终能达到一千亿条数据导入时间为 117ms;
- ◆ 在最终测试中, 和多家竞争厂商的数据库性能对比中排名第一。

2017.7-2017.9 核心研发人员 某部队单位数据跨域访问项目

- ◆ 参与测试方案的编写与评估, 以及测试现场环境的搭建与调试;
- ◆ 阅读分布式大数据 SQL 查询引擎 presto 的源代码, 理解其架构及实现原理, 并形成文档贡献到公司文档库中;
- ◆ 实现用于访问分布在上海, 崇明, 陵水三地数据库的接口。

2016.6-2016.12

实习软件开发工程师

中国科学院信息工程研究所

- ◆ 实习工作期间共参与公司项目 1 个，在项目中主要作为开发人员提供技术支持；
- ◆ 参与研发大数据实时处理分析平台，每天对上亿条日志数据实时清洗、处理、分析，从中发掘有用信息。

2016.6-2016.12

研发人员

大数据处理平台研发

- ◆ 调用了 Spark Streaming 的 API，从 kafka 消息队列中消费实时的流数据，对海量数据实时清洗处理。最后将结果储存到 hive 大数据仓库中。
- ◆ 调用 Spark MLlib 的 API，实现利用决策树、朴素贝叶斯算法和随机森林的机器学习程序，使用标记数据对算法模型进行训练，并比较两个算法的分类效果
- ◆ 大数据处理平台每天对上亿条日志数据实时清洗、处理、分析，很好地满足项目要求

教育经历

2013.9-2017.6

哈尔滨工业大学

计算机科学与技术学院

计算机科学与技术专业

技能

专业技能

- ◆ 熟练掌握 JAVA，了解 c、python；
- ◆ 熟悉 Spark、hadoop、hive、presto 等常用的大数据开源项目；
- ◆ 熟悉大规模分布式并行处理数据库的基本架构及原理，对分布式存储有一定了解；
- ◆ 有一定的脚本语言编写能力，能通过 shell 编写脚本，实现文本处理或其他功能；
- ◆ 熟悉 greenplum 数据库（基于 postgresql）的源码，对其基本实现原理和框架较为清楚；
- ◆ 熟悉常用的机器学习算法（如：决策树、朴素贝叶斯、k-means 等）的原理及适用场景（本人对机器学习非常感兴趣，希望能够从事机器学习或数据挖掘相关工作）；

英语能力

- ◆ 英语达到 CET-6 水平。