

陈思翰

电话: 18175686352 | [邮箱](#) | [GitHub](#) | [Website](#)

教育背景

北京航空航天大学，计算机科学与技术专业

2019.9-2023.6

- 综合绩点: **3.84/4.0**; 加权平均分: 92/100; 专业绩点:**3.88/4**; 专业平均分: 93/100
- 主修课程: 离散数学 98, 概率统计 99, 数学分析 96, 高等代数 100;形式语言与自动机 96、数据结构 90、操作系统 93、计算机组成 94、面向对象设计与构造 97、编译技术 92、数据库系统原理 95、计算机网络 88、X86 汇编程序设计 100
- 主要技能: 熟悉 Java, 了解 Springboot, Maven, Mybatis; 熟悉 MySQL, Redis, Git; 了解 RocketMQ
熟悉 Python, 了解 Go, C++
- 英语水平: 雅思总分 7.5/9; GRE 323/340 (Verbal: 153/170,Quantative: 169/170)
- 荣誉奖项: 蓝桥杯 PythonA 组国家级三等奖, 北京市一等奖
北京航空航天大学三好学生
学科竞赛奖学金: 一等奖, 2020, 2021; 特等奖, 2022
学习优秀奖学金: 二等奖, 2020, 2021; 一等奖, 2022
社会工作奖学金: 二等奖, 2020

Imperial College London, Advanced Computing MSc(offered)

2023.Fall

实习经历

字节跳动 | 飞书套件主数据部门后端实习生-Go

2022.7-2022.9

主要负责人力资源数据的管理 (查询、存储、迁移), 主要项目如下:

ETL 模块消息去重

- 对 ETL 模块中消息队列里的消息进行去重, 减少 Vault 服务查询接口被调用的次数
- 基于 RocketMQ 的 PartitionKey 进行分布式去重, 替代基于 Redis 进行集中式去重的方式, 避免了繁琐低效的锁操作和设计
- 设计容灾策略, 保证宕机之后的数据一致

MySQL 数据表迭代

- 由于之前设计上的缺陷, 将已经投入使用的 manage_relations 表换为 job_data 数据表
- 采用双写策略, 让两个数据表同时在线上环境工作, 防止数据表替换过程中产生脏数据或者数据丢失、不一致
- 利用 Go 脚本基于增量数据对 job_data 数据表相关代码的逻辑正确性进行了检查
- 逐步增加线上流量到新数据表中, 使换表过程中的线上事故可控

Vault 服务 MultiGeo 功能改造

- 应法律法规要求将原来存储在中国的新加坡员工的敏感数据迁移到新加坡
- 使用软删除和硬删除结合的方式, 避免了迁移过程中的数据丢失
- 通过 EGW 增加了 Vault 服务跨 Geo 存储和查询的功能

京东科技 | 科技产品创新部门后端实习生-Java

2022.11-2023.2

主要负责 Ku+平台 (基于 AutoML 和 MLOps 的机器学习平台) 功能的开发与维护, 主要项目如下:

OSS 数据迁移

- 利用 Java 脚本遍历 OSS 服务中的文件树, 将图片、文本等不同形式的文件做相应的处理, 最后存入 MySQL 数据库
- 通过事务机制保证了数据迁移过程中的数据一致性

部门树生成器

- 生成器以定时任务的形式, 每天自动根据平台用户信息生成用户部门树

- 采用增量处理的策略，保证了定时任务的性能；考虑了各种部门异动，包括：取消，新建，更名和移动

异常 Pod 报警器和看板

- 报警器定时轮询 Pod 的健康状况，将异常的 Pod 信息发送给 Pod 拥有者和系统管理员（邮件、钉钉 ME 等形式）；创建了一个异常任务面板以展示异常信息，包括：K8s 中 Pod 的 status 和 description 等信息
- 预先考虑了报警器动态配置报警条件的需求，支持在服务运行过程中动态配置报警条件
- 使用了 cache 加速后端接口的查询速度

项目经历

基于 SysY 的编译器 -Java lines:8000+

项目地址: <https://github.com/sihan-chen-yes/Compiler-BUAA-2021>

2021.9-2021.12

- 使用 Java 完成编译器，可对 SysY(C 语言的一个子集)进行从源码到 MIPS 汇编代码的转换
- 系统组成部分：词法分析，语法分析，语义分析生成抽象语法树(AST)，中间代码生成，MIPS 汇编代码生成，代码优化
- 利用自动机理论手动进行词法分析，利用递归下降进行语法分析,同时进行错误处理，然后扁平化 AST 生成中间代码(四元式)，最后完成从中间代码到目标代码(MIPS 汇编代码)的转换
- 对中间代码、目标代码进行优化:数据流分析、划分基本块，利用图着色分配寄存器，执行常量传播和复写传播之后进行死代码、无用代码删除，除此之外还进行了窥孔优化、条件短路处理、常量折叠等细节优化
- 综合 rank:24/250 (对不同类型的汇编指令进行加权统计)

多功能电影网站 -Python lines:1000+

项目地址: <https://github.com/sihan-chen-yes/MovieWeb>

2021.9-2021.12

- 支持基本功能的电影网站，包括：在线观看预告片，进行影评等，已经部署在个人云服务器上，可正常使用 (<http://movie.ito.vin/>) 游客账号:123 密码:123
- 前端采用 Vue 框架，后端采用 Django 框架，数据库采用 MySQL
- 主要负责后端工作，包括所有逻辑操作的实现，数据库的交互(基本操作、触发器、存储过程的设计)，保障安全性(防注入、密码的明文存储)以及网站在服务器上的部署工作

基于 ARC 的页面置换 -C++ lines:400+

项目地址: <https://github.com/sihan-chen-yes/pageReplaceOS-BUAA-2021>

2021.5-2021.6

- 对操作系统缓存中的页面置换的过程进行模拟，设计算法最大程度避免页面缺失和置换，同时尽量减少计算成本
- 实现了 2003 年的 Adaptive Replacement Cache 算法(具体见 github 文档)，该算法主要结合 LRU 和 LRU 的优点，并改进了二者的缺点，具有自适应性，个人在实现时对数据结构、C++ 的代码进行了优化，此外，还进行了算法参数的调优
- 综合 rank:1/250 (对 Linux 下的 perf 指令的各项结果赋予不同的权重进行排名)

研究经历

VLN 模型优化-Python

项目地址: <https://github.com/sihan-chen-yes/VLN-Model-Optimization>

2022.5-2023.2

- 提出了 Structured state-Evolution(SEvol)模块，解决了基于一维状态向量的导航智能体模型对于物体级别的时空信息过度压缩的问题，提高了此类模型在 R2R, R4R 和 REVERIE 数据集上的精度
- 基于强化学习的 A2C 算法，提出了 Reinforced Layout clues Miner(RLM)生成物体子图
- 利用 Dynamic Graph Neural Network(DGNN)模型对物体的时空特征进行了聚合
- 提出了矩阵版本的 GRU 模型，在每个时间步对 DGNN 中的权重矩阵进行了更新

互动式机器狗-Python

项目地址: https://github.com/sihan-chen-yes/my_dog

2021.10-2022.5

- 跟随实验室机器人团队开发了带有语音交互和导盲功能的机器狗
- 目标检测和语音识别模块分别采用 YOLOv5 算法和 CMUSphinx 开源库实现
- 基于机器狗 SDK，并结合目标检测和语音识别模块产生的信息实现了主控逻辑