

# 朱奎 JAMES

✉ zhukui.1998@qq.com · ☎ (+86) 131-2671-6166 · 🌐 <https://github.com/zkJames>

🏠 山东潍坊 · 💻 后端开发、数据工程 · ♂ 男

## 🎓 教育背景

中国科学院大学, 计算机软件与理论

2021.9 - 预计 2024.6 毕业

- 学术硕士在读 (保研)
- **GPA:** 3.5 / 4.0 **排名:** 4 / 63 (前 10%)

河南科技大学, 计算机科学与技术

2017.9 - 2021.6

- 本科
- **GPA:** 4.5 / 5.0 **排名:** 3 / 144 (前 3%)

## 👨‍💻 实习经历

阿里巴巴（中国）有限公司

2023 年 6 月 – 至今

直连交易平台 Java 后端开发实习生 飞猪 · 住宿行业研发中心 · 行业基础平台

学习直连交易平台、商品技术框架相关知识，负责交易和商品搜索相关的 Java 后端开发工作

- 端午节线上问题排查处理，通过**分析线上日志**，定位到问题所在。通过修复程序 bug，处理标准商品交易链路中的错误码和外部接口返回解析不一致的问题，解决了线上问题。
- 梳理并优化商品框架的**直连订单批量检索功能**，研究如何优化商品和订单搜索。
- 梳理商品信息抓取链路，从供应商接口获取酒店信息，分析酒店信息的数据结构，实现酒店图片的批量保存和房型描述信息的解析和存储。

微软（亚洲）互联网工程院

2022 年 11 月 – 2023 年 6 月

Bing 数据/软件开发实习生 (Data/Software Develop Engineer Intern) STCA Beijing Bing

### 1. Microsoft Creator Copilot

基于 Bing 搜索引擎和 ChatGPT 模型的 **AIGC 自动创作工具**，提供段落实时改写、热搜话题新闻生成以及自动配图等功能。

- 后端开发：我主要负责设计和实现写作工具的**后端架构**，采用**代理模式**包括搜索、LLM 模型的**代理层**和**业务层**。代理层负责与**搜索引擎**和**LLM 模型服务接口**进行交互，实现搜索、模型的调用和结果返回。业务层负责处理用户请求和响应，实现写作工具的核心功能。为了提高框架的性能和内容生成体验，我在代理层采用了**接口响应优化**、**HTTP 连接池**、**线程池**、**Redis 缓存**等技术，有效地提升了并发请求的吞吐量和响应速度。
- Prompt 工程：采用 LLM ReAct(Reasoning and Acting) 范式，结合**推理动作**和**搜索动作**功能使 ChatGPT 模型可以与 Bing 搜索引擎交互，获取**实时热搜话题**和**Bing 搜索结果**作为知识库，克服了思维链推理中普遍存在的妄想和错误传播问题，并生成更合理的类人任务解决轨迹，提高大语言模型 AI 新闻创作的真实性和创造性。

### 2. Bing 用户增长数据分析、搜索平台工具研发

- 负责 Bing Search 中**用户搜索相关的数据开发**和**数据分析**，使用微软 Cosmos 数据平台 Scope 语言（类 Spark SQL）对搜索日志实时数仓相关数据进行查询、整合，计算 New Bing 相关的用户增长指标进行分析。通过改进本土化搜索功能提高 DAU、DSQ 等指标，并通过 ABTest 进行功能分析。
- 维护 Bing Search 的平台工具，负责**Bing 中国区首页新闻热点数据的实时更新功能**。从对象存储中读取数据，通过 Scope 查询语言对热搜数据的提取、加工、转换、实现了 Bing HotSearch 热搜、Bing 新标签页股票指数更新等功能，推动了 Bing 首页界面显示的优化。

Berserker 是提供数据的存储、查询、数据开发、数据质量分析、分布式调度的大数据平台。实习中负责工具侧**元数据**、**数据运营**、**数据管理**等方向，专注**元数据采集**、**治理工具**等功能的研发。

- 设计了 Hive MetaStore 中**分区信息数据的解耦方案**。通过**全量拷贝**、**Binlog 增量变更**、**数据一致性检测**、**分布式调度同步**的方法落入 **TiDB** 的方案加速查询，优化了 Hive 表分区信息查询速度和资源占用，**接口性能提升 60%**
- 在部门降本增效计划中，参与了无效数据表集中下线的功能设计，为业务提供打标，批量定时删除并在删除后及时告知表所有人的功能，通过**分布式定时调度框架**设计实现了高效的数据表安全删除逻辑功能
- 为了提高数据访问的性能，需要区分**冷热数据**。同时，为了支持各类数据应用和数据治理的需求，需要提供表使用热度的统计和查询功能，在数据库层面统计各个表的使用情况，设计了覆盖全、数据准、粒度细的**表使用热度**功能服务，为各类数据应用和数据治理提供支持。

## 👤 个人项目

### 用于列车售票的可线性化并发数据结构

2021 年 9 月 - 2021 年 12 月

课程设计 中国科学院大学《并发数据结构与多核编程》，林惠民院士

在本项目中，我使用 Java 语言设计并实现了一个**基于 lock-free 算法的高效列车售票系统**，该系统支持查票、购票、退票等操作，并能在并发环境下保证每个操作都能**线性化**地执行，即按照某个全局顺序执行，不会出现不一致或冲突的情况。具体来说，我采用了以下技术和优化：

- 利用乘车区间**二进制编码运算**的技术，将每个区间用一个二进制位表示，从而将锁的粒度从区间级别降低到**位级别**，大大减少了锁的竞争和开销。
- 基于 **CAS 原语**（比较并交换），实现了 **lock-free**（无锁）的并发方案，避免了死锁、饥饿等问题，提高了并发效率。
- 引入余票表**缓存**，在购票、退票后子线程异步刷新余票表，使得查票操作可以直接从缓存中读取数据，提高了查询速度；通过随机占座等**负载均衡**的优化，使得不同线程尽量访问不同的资源，避免了资源竞争，提高了并发性能。
- 使用日志系统进行**容灾处理**，即每次执行操作时都会记录操作的类型、参数、结果等信息到一个日志文件中。这样可以保证在断电后可以根据日志文件重放操作，并且恢复到最新的状态。
- 按照 70% 查票、20% 购票、10% 退票的概率进行压力测试，利用**吞吐量**、**操作时延**等性能评价指标，对算法性能进行评估，最终性能评测分数排名达到所有方案的**前 5%**。并且使用可线性化验证工具对并发数据结构进行可线性化分析验证，证明了算法的正确性。

## ♡ 获奖情况

- 中国科学院大学三好学生
- 国家励志奖学金
- 中国科学院大学计算机科学与技术学院优秀学生
- 全国大学生数学建模竞赛二等奖
- 全国大学生数学竞赛三等奖

## ⚙️ 额外技能

- 英语：CET 4 566、CET 6 479
- 爱好：乒乓球、羽毛球
- 文档：Markdown、 $\LaTeX$