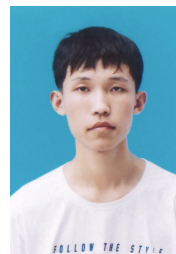


# 朱奎 JAMES

✉ zhukui.1998@qq.com · ☎ (+86) 131-2671-6166

🏠 山东潍坊 · 💻 后端开发、数据工程



## 🎓 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

- 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）集团控股有限公司

2023 年 6 月 – 2023 年 9 月

后端开发实习生 飞猪 · 住宿行业基础平台-国际 EBK · 商品

国际 EBK 商品直连系统部门建立独立的海外 EBK 直连系统，负责 Java 后端开发工作，涉及 **Spring-Boot**、**MyBatis**、**Redis**、**MySQL**、**MetaQ (RocketMQ)** 等技术栈。

- 负责国际 EBK 直连系统的账号和权限模块开发，主要负责用户账号鉴权、联系人消息通知等模块的开发，通过主子账号、functionBit、角色授权等设计，实现了用户账号的**权限分配**、**鉴权**、**登录**、**子账号创建**以及基于 **MetaQ** 消息中间件实现了联系人的消息通知等功能。
- 了解**供应商数据中心 MOS** 需求。MOS 为国际 EBK 提供经营能力，帮助供应商感知 N 漏斗全过程提升整体酒店/房型匹配率。

微软（亚洲）互联网工程院

2022 年 11 月 – 2023 年 5 月

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

1. **Microsoft Creator Copilot**（微软内容创作者助手）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。主要负责**后端系统开发**。涉及 **Scope SQL**、**微软数据平台 Cosmos**、**离线数仓**、**ABTest 分析**、**Redis**、**线程池**、**Azure 数据流**等技术和方法。

- 采用**代理模式**设计**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生成质量，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**：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，并生成更合理的类人任务解决轨迹，提高 LLM 新闻创作的真实性和创造性。

2. **Bing 用户增长数据工程 & 搜索平台工具**

- 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 Hive SQL）对 Bing 搜索引擎实时日志数据进行 ETL，产出关于增长、日活、群体、地域、热搜话题等图表，结合可视化用户增长平台进行监控，并进行分析；改进本土化搜索的业务策略和功能、分析修复线上 bug，提高 DAU、DSQ 等用户增长指标，并通过 ABTest 平台进行分析，最终提升了搜索用户体验。
- 维护 **Bing Search** 的平台工具，负责 **Bing 中国区首页新闻热点数据的更新业务功能**。首先 ETL 产出的热搜数据实时入库，通过调度框架执行 Azure function 实现从 KV 存储中读取热搜数据进行提取、加工、转换。完善 Bing HotSearch 热搜、Bing 新标签页股票指数更新等功能的链路。

**B 站大数据平台 · 平台工具 · 元数据**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。
- 在部门降本增效计划中，实现无效数据表定期下线的功能，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
- 提供表使用热度的统计功能，参与了覆盖全、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。

## 👥 项目

### 电力市场数据实时分析平台

2022 年 9 月 - 至今

实验室项目 数据工程、后端开发

通过用电情况和电网公司内部的电力系统数据、日志数据、业务数据，建立实时数据仓库，以供电力公司管理人员及时了解用电时的运作情况和用电负荷。为用户提供与电力市场业务相关的报表、业务预测，以辅助各管理层有效决策。

项目架构：Hadoop + Flink + Kafka + MySQL + HBase + Redis + ClickHouse + SpringBoot

- 采用 FlinkCDC2.0 对 MySQL 中业务数据进行采集，从 ODS 层读取数据利用 MySQL + 广播流实现动态分流。
- 将数据分别写入 Kafka、HBase 构建 DWD 层和 DIM 层。
- 通过 Flink 将 DWD 层的数据进行实时聚合、过滤、分析，形成各种指标数据，并写入 ClickHouse 和 Redis。ClickHouse 作为 OLAP 引擎，支持多维度的实时查询和分析，Redis 作为缓存数据库，提供快速的数据访问和更新。使用时间序列预测算法实现实时电价预测。
- 使用 SpringBoot 实现可视化大屏系统后端系统搭建。

### 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

并发编程项目模块 Java 并发编程

设计并实现了一个可线性化的高效列车售票系统，并能在并发环境下保证高性能运行。

- 利用乘车区间二进制编码的涉及，将锁的粒度从区间级别降低到车座粒度，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。

## ♡ 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省二等奖
- 全国大学生数学竞赛三等奖

## ⚙️ 爱好与技能

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