

Song Hao 求职意向： JAVA 高级软件工程师

自我评价

- 📁 **工作背景：**7年多 Java 开发工作经验，精通 web 后端和大数据开发，熟悉数据全生命周期 ETL 开发管理，具备较好的编程功底、代码知识，熟悉分布式系统及主流开发框架的原理和适用场景，对项目技术架构选型有良好的判断。业务理解能力强，具有 TB 级数据处理与成功落地经验。
- 📁 **专业能力：**
 - ✍ 精通 Spring, SpringMVC, SpringBoot, SpringCloud, MyBatis 等后端开发框架。
 - ✍ 熟练掌握 Hadoop 生态圈，独立搭建 Hadoop 集群，应用 Flume, Kafka, HDFS, Hive, MapReduce, Hbase, Zookeeper, 解决实际的问题。
 - ✍ 熟练掌握 Spark 生态圈，对 Spark 的批处理, SparkSQL, SparkStreaming 有良好的认识。
 - ✍ 掌握全文检索 Elasticsearch, 并在实际项目中运用。
 - ✍ 对 Java 和 Python 语言有较好的认识和实践。
- 📁 **学习能力：**喜欢沉淀，爱学习，并善于从日常工作中总结复盘，持续提升工作效率。具备良好的代码习惯，重视codereview，总结并记录。关注学习行业技术方面相关前沿知识，目前在备考高级系统架构师。
- 📁 **性格特质：**喜欢钻研技术，个性沉稳，工作认真负责，较强的执行能力和抗压能力。爱好阅读，台球。

工作经历

- | | | |
|------------------------|---|----------------|
| 2020.04-至今 | 达而观信息科技(上海)有限公司 | 资深技术专家 |
| 📁 API 开发： | 负责知识图库系统 0-1 的API 开发设计，主要有图谱管理，数据源管理，图谱构建，图谱探索，词典管理，权限管理模块，包括需求讨论会方案的梳理、需求文档的编写，与开发团队共同设定项目计划并确保完成，运行后持续推进产品的迭代发布和优化升级。 | |
| 📁 需求分析： | 参与项目开发工作和部署，包括系统核心模块的框架选型、需求分析，确保方案可行性和扩展性。 | |
| 📁 数据库设计开发： | 使用 Elasticsearch、Mysql、Janusgraph/ArangoDB 数据库，设计系统核心模块数据库表，以及对整个平台系统代码评审和数据表设计审核，确保系统数据的利用率和一致性，便于程序的后续开发、维护和延展。 | |
| 📁 功能优化： | 依据客户需求，优化产品功能模块结构和流程逻辑，确保系统使用及时满足客户要求。 | |
| 📁 新人带教： | 带领一名新员工，进行技术指导，帮助新人成长，克服技术困难，保证项目质量以及进度和时间。 | |
| 2019.03-2020.03 | 二三四五科技有限公司 | 高级后端工程师 |
| 📁 数据采集： | 对平台 5 个 APP 的商业数据进行数据采集，传送到中台仓库，进行 ETL 清洗，提取分析，提炼相关内容制作商业项目报表、分类统计，补漏，展示在“趣看看”内容 APP 上，日处理达上亿条数据，输出合并报表达 50 张左右/日。 | |

-
- 📁 **ETL 脚本**: 负责数据仓库的数据抽取及存储脚本开发, 使用 Hive 和 SparkSQL 编写, 数据迁移脚本, 配置调度信息实现数据萃取逻辑。通过优化 ETL 脚本实现效率提升和降低系统对服务器资源的需求。
 - 📁 **数仓设计开发**: 数仓数据模型设计及编写, 存储过程实现 BUFFER 层, ODS 层, DW 层, DWS 层的业务开发, 完成数据同步, 归纳同步。通过 SQL 脚本根据需求和主题完成图层报表输出, 帆软 BI 实现报表可视化展示。

2016.05-2019.04

数说故事科技有限公司

ETL 数据工程师

- 📁 **Java 开发**: 参与项目的需求调研和需求分析, 对项目进行可行性分析。设计和搭建系统架构、平台、数据库、接口和应用架构等, 完成系统框架和核心代码的实现。期间开发了新希望地产项目和自助查询系统。
- 📁 **ETL&爬虫**: 使用 ETL 工具对 Weibo、电商购物平台、论坛等抽取商家价格舆情信息、质量分析及竞品价格质量商业信息, 完成数据清洗和抽取分析归类。对爬虫系统的技术缺陷, 对策略架构做出合理地调整和改正。
- 📁 **数仓开发**: 参与数据仓库系统的开发工作, 对接上下游需求, 制定产品需求及产品上线方案, 完成数据模型设计、编码、部署设计, 实现业务逻辑。挖掘数据价值, 定期编撰数据分析报告, 为屈臣氏等多家零售商提供商业舆情监控报表, 构建商业智能。

2014.11-2016.04

数字思新技术有限公司

Java 后端工程师

- 📁 **Java 开发**: 负责平台游戏产品及外包项目的 Java 代码开发。根据业务需求编写代码, 接口文档, 及功能模块的自测, 提测。期间开发了网站日志分析系统, 优联租房系统。
- 📁 **系统优化**: 系统线上环境的维护, 对 BUG 及时做出处理及数据修复。

项目经历

2020.04-至今

源海知识图谱为系统

资深技术专家

- 📁 **项目描述**: 项目为一站式知识图谱应用平台, 可应用于金融, 政府, 军工, 教育等行业, 包含数据源管理, 图谱构建, 图谱管理, 图谱应用等, 根据搜索图谱的实体, 并根据设定的深度, 对图节点进行深度分析。
- 📁 **项目职责**: 维护整个产品的开发与迭代, 需求评审, 接口开发, 数据库设计, 确保系统稳定运行。图谱后端的 SpringBoot 程序开发, 数据处理模块 Spark ELT 程序的开发与维护, 大数据集群的搭建与维护, 数据库 Janusgraph 的维护, 推进项目的验收与交付。
- 📁 **技术环境**: SpringBoot + Kafka + Zookeeper + Spark + HBase + Elasticsearch + Janusgraph + ArangoDB
- 📁 **项目业绩**: 推进项目的按时验收与交付, 完成千万级别的实体, 亿级别关系图谱数据构建。产品迄今应用于军工, 医疗, 政务, 金融等十多家企事业单位。

2019.04-2020.03

移动端数据仓库建设

高级后端工程师

- 📁 **项目描述**: 为了对公司的移动端产品进行数据分析, 创建一个高效的采集, 分析, 展示的大数据平台, 平台包含 ETL 模块, 元数据管理模块, 数据质量模块和调度系统四个大块。
- 📁 **项目职责**: 参与数据仓库架构的设计及开发, 对接移动端所有产品的业务和日志数据到中台的数据仓库, 抽取成各种数据需求, ETL 流程优化及解决 ETL 相关技术问题。
- 📁 **技术环境**: Hive + EasyScheduler + DataX + Spark + Clickhouse + 帆软 BI
- 📁 **项目业绩**: 项目可处理日数据量 TB 级, 开发通用性采集脚本, 让数据对接通过配置方式接入, 人力成本节省50%; ETL 脚本可每天监控指标的异常情况, 实现按时按标准的计算指标报表数据; 可根据任务的数据量计算相应的服务

器资源，缩减达 1/3。

2015.09-2016.03

电商竞品分析系统

后端工程师

- 📁 **项目描述:** 通过爬虫爬取舆情与商品信息数据，将采集的数据存储到 HBase 中，通过编写程序对采集到的数据进行 ETL 处理，最终将数据存储到 Elasticsearch。然后通过 java web 后端程序对需求分析师整理的相关指标进行统计与展示处理。
- 📁 **项目职责:** 负责 0-1 核心研发工作，采集，分析，接口开发等。编写 Spark 程序对原始日志数据进行清洗；根据项目需求对统指标进行分析和统计。
- 📁 **技术环境:** Java + SpringMVC + Spark + HBase + Elasticsearch
- 📁 **项目业绩:** 在极短的时间内，将数据的各个链路进行打通，使客户能快速浏览各个指标数据，指导运营决策。对 Elasticsearch 查询进行不断优化处理，提升客户体验，使项目顺利交付。

教育背景

2009.09-2013.06

武汉长江工商学院

计算机科学与技术 本科