

大数据开发工程师求职简历



个人信息 (Personal Info)

姓名:	魏孝伟	性别:	男
籍贯:	福建省福清市	年龄:	25
电话/微信:	13276047202	电子邮箱:	weixiaowei1010@foxmail.com
工作经验:	2 年	是否在职:	是



教育背景 (Education)

2014.09—2018.07	郑州大学	安全工程	本科
-----------------	------	------	----



专业技能 (Professional skills)

- 1、掌握 Hive、HBase、Flume、Kafka 等 Hadoop 生态技术圈的相关技术的使用;理解 Spark 工作机制, 熟练掌握 Spark core、Spark sql 的开发、熟悉 Spark Streaming 的开发
2. 熟悉 Spark 大数据项目的性能调优、线上 troubleshooting 与数据倾斜的解决方案
3. 熟悉 MySQL、Hive、Oracle、PostgreSQL 数据库的使用, 能熟练编写 sql 语句进行复杂场景的指标计算
- 4、熟悉 java、scala 语言开发, 掌握 Linux 常用命令的使用



工作经验 (Work Experience)

2018.7—2019.5	河南金明源技术有限公司	大数据开发工程师
2019.7—至今	深圳文思海辉信息技术有限公司	大数据开发工程师



项目经验 (Project experience)

项目名称: 平安寿险保全运营监控平台

软件架构: Oracle+Kafka+SparkStreaming+Hive+Spark Sql+PostgreSQL+Zookeeper

所属公司: 深圳文思海辉信息技术有限公司

项目描述: 为了提高中国平安寿险的服务质量和优化客户的体验。中国平安寿险启动了针对保全业务的大数据支持业务。该项目包含离线和实时。离线部分使用数仓的模式进行开发。离线数仓分为五层架构进行开发。实时的部分利用 Spark 进行开发。实时部分是从 oracle 获取业务数据, 通过 Ogg 传入到 kafka 中, 使用 spark streaming 对接 kafka 进行数据处理, 最后将结果数据写入到 PG 数据库。

责任描述: 我主要负责的是项目离线处理和部分实时开发。针对 BDI 从 Oracle 导过来的数据进行建模离线处理。实时部分负责 Spark Sql 对 Spark Streaming 处理后存入 Hive 的数据进行业务逻辑编写。

项目名称: Api 日志信息统计分析

软件架构: MongoDB+Hive+Spark

所属公司: 深圳文思海辉信息技术有限公司

项目描述: api 系统的请求日志保存在 MongoDB 中, 由于数据量大 (每天的请求日志大概有 1 亿条), 需导入到大数据平台进行统计分析, 并将统计结果写回 MongoDB, 以保证能顺利获取每天的 api 数据, 为趋势扩容做参考。

责任描述:

- 1、抽数模块, 使用 spark 从 MongoDB 中抽取每天写入的请求日志
- 2、计算模块, 使用 spark sql 对抽取的数据进行统计分析
- 3、写入模块, 使用 spark 将计算后的结果写回 MongoDB

项目名称: 平安寿险品质监测项目

软件架构: Hive+Zookeeper+Spark+PG+Oracle

所属公司: 深圳文思海辉信息技术有限公司

项目描述: 平安寿险有大量的业务代理人员, 代理人可以通过售出保险获取佣金。为了预防代理人员在获取佣金以后, 客户进行退保, 造成平安公司的损失。平安寿险利用 Python 设计机器学习模型, 我组负责向模型输入正确的业务数据。整个项目分为数据采集, 各种指标 (场景、服务、投保等) 的计算, 指标结果导出, 我主要负责投保场景指标的计算。项目初期使用 hive 来进行开发工作, 但由于 hive 效率较低, 我提出使用 spark 进行开发, 改为 spark 后效率提高将近 4 倍。

责任描述:

- 1、投保指标的 hql 开发以及结果数据的正确性验证
- 2、Spark 代码开发, 由 hive 转为 spark 开发后, 为增强代码健壮性与拓展性, 避免不同开发人员一人一套代码, 设计了通过读取 xml 的方式, 只需简单配置 xml 文件, 即可让不同开发人员共用一套代码, 提高开发效率

项目名称: APP 日志统计分析系统

软件架构: Flume+Spark+Kafka+Hive+HBase

所属公司: 河南金明源技术有限公司

项目描述: APP 日志统计分析系统 是收集用户使用公司内部 APP 的日志记录, 并进行统计分析, 通过统计 APP 的新增用户、活跃用户、沉默用户等, 了解 APP 的详细使用情况; 统计分析 APP 不同版本的分布, 了解用户对不同版本 APP 的接受程度; 统计 APP 的使用时长, 来了解用户对于 APP 的依赖程度; 通过留存用户分析、新鲜度分析, 帮助公司了解不同用户的使用情况。

项目描述:

- 1、使用双 Flume 拓扑架构以及 Flume 拦截器对服务器中的日志文件进行监控、采集, 并发送到 Kafka 集群;
- 2、消费 Kafka 中的数据, 并写入 HDFS, 编写 shell 脚本将数据导入到 Hive 分区表中;
- 3、通过 Hive SQL 以及自定义聚合函数对离线数据进行分析;
- 4、使用 SparkStreaming 处理实时数据流, 实时统计每个城市的用户数量, 并将结果写入 HBase。