

程松

求职意向：大数据开发工程师

现居住地：北京

出生年月：1997/8

性 别：男

学 历：本科

电 话：18395665241

邮 箱：idforwork@qq.com

教育经历

2015.09-2019.06 河北建筑工程学院 物联网工程（全日制）

工作经历

2018.10-2021.09 杭州时趣信息技术有限公司 | 大数据开发工程师

专业技能

- 熟悉Flink的基本API，可以根据计算逻辑进行代码实现，了解底层运行原理和工作流程
- 熟练使用SparkCore和SparkSql进行数据处理
- 熟悉Hadoop生态圈，HDFS的读写流程以及常用的操作命令，熟悉hdfs工作原理
- 熟悉Linux命令，熟悉常见的shell脚本
- 熟悉yarn集群搭建以及yarn的资源调度原理
- 熟悉mysql数据库，熟练SQL语法
- 熟悉MapReduce工作原理和离线计算,对MR各组件关系有一定理解
- 熟练使用java、scala编程语言
- 熟悉Hive工作原理，熟练hql脚本开发、熟练使用hive窗口函数，有一定的hive调优经验
- 了解ClickHouse的基本架构和运行原理，Zookeeper工作机制

项目经历

1 | 基于事件驱动的实时分析系统

项目功能：根据用户的日常行为，利用flink的实时处理优势，为其提供多方位实时的服务，安全-实时监测用户异常行为，防止用户利用漏洞获利；营销-结合用户画像特征，使广告更精确的投放到每个用户。

工具介绍：Flink、Nginx、Canal、Drools、MySQL、Kafka、Clickhouse、Hbase、Redis

大致流程

- 1) 通过规则管理平台定义规则
- 2) 通过规则注入模块（规则数据广播流以及对应的处理工具）,加载（删除，停用，启用）定义好的规则
- 3) 从kafka中消费用户行为日志数据
- 4) 根据在线规则的个类分组需求，对日志数据进行复制和标记分组key，实现动态keyBy的分组
- 5) 数据进入到keyedProcessFunction
- 6) 在keyedProcessFunction中获取到所有的在线规则
- 7) 遍历所有的符合分组字段的规则，对输入事件进行规则的逻辑计算
- 8) 输出计算的结果到上层应用系统所对接的kafka消息缓存中参与设计和优化系统，实现用户行为日志采集，规则注入，规则计算和查询缓存模块

主要职责：

1. 参与设计和开发了该系统的部位内容
2. 对数据清洗过滤、转化、脱敏
3. 实现用户行为日志采集，规则注入，规则计算和查询缓存模块
4. 负责部分的 flink 代码

2 | 离线数仓综合系统

项目功能：该项目基于用户行为数据以及用户产生的业务数据，将这些行为数据和业务数据导入到hive数仓中，在数仓中进行分层,对数据进行分析,以及统计各类指标,利用这些信息分析运营状况，从而为公司提供更好的市场支持，对运营提供决策。

工具介绍：Flume、Sqoop、Hive、Spark、HDFS、Zookeeper、Kylin、Azkaban、Atlas、Hbase

大致流程

数据来源：APP/小程序/Web端的埋点，通过Flume写入到HDFS，每日一个分区

仓库层级介绍：

ODS层：建表，将HDFS中对应分区的数据导入Hive，生成原始表；同时也将一部分业务数据通过Sqoop采集过来

DWD层：用Spark对原始表进行过滤、规范化、session分割、地理信息集成，以及全局唯一ID的生成，之后与业务表Join，可生成宽表、拉链表等

DWS层：该层是对报表需求做一个轻度聚合，包含但不止流量分析报表、广告转化分析、新用户留存分析、归因分析等

ADS层：此层是最终报表的计算存储层，针对业务人员的需求

画像系统：Hbase存储，基本标签、行为标签、消费喜好标签等

Azkaban：进行每天定点的任务脚本调度

Atlas：元数据管理系统，便于血缘追踪 | Kylin：OLAP多维分析平台

主要职责：

1. 参与数仓的维度建模和设计，将数仓分层
2. 完成原始数据的ETL工作
3. 数据入库后的报表开发
4. 拉链表每日更新的Shell脚本编写

自我评价

本人喜欢挑战自我，性格认真负责，对待工作沉稳严谨，热心助人，待人真诚，有良好的团队合作精神与责任心。有上进心，乐于学习，希望能未来的日子里在公司领导与前辈的带领下，努力工作，稳定地提高自己的工作能力，与公司同步发展。