大数据工程师

姓名：蒋琼香 毕业院校：河南科技大学

毕业时间：2018-06 学历：本科

邮箱：jop2793752687@163.com 专业：计算机科学与技术

电话：17521162143 工作经验：3年

基本信息

岗位：大数据工程师（Hadoop / Spark）

目前状况：在职状态

求职意向

**1、数据仓库**

(1) 熟悉Hive和Hbase的系统架构；能够自定义UDF以及UDFT函数，编写运行hql的shell脚本，熟悉Hive相关调优方法；根据实际需求设计Hive表和Hbase优化。

（2）熟练运用flume，运用flume的选择器与自定义拦截器对日志数据进行过滤与分发到不同的sinks中

**2、spark计算框架**

（1）熟练使用SparkCore、SparkSQL、SparkString相关的API进行离线和实时应用的开发，理解Spark集群的启动过程相关类源码(Master/Worker)和任务提交流程。了解spark相关源码开发。

（2）熟悉spark streaming原理和相关的编程API，能使用Kafka+Sparking开发实时流处理应用程序。

**3、开发语言**

（1）熟练使用Java，scala等语言开发进行数据处理

（2）熟练使用Spark框架进行离线或者实时数据分析

（3）掌握Linux常用操作命令

**4、数据库相关**

1. 熟练运用MySQL数据库增删改查，索引，视图，用户授权，建库建表等相关操作。
2. 熟悉Redis数据库的架构，Redis 的主从架构，分布式集群 和API操作。

(3) 熟悉Hive，HBASE 大数据存储，熟练运用hql 语句对Hive数据进行分析

专业技能

**5、开发工具**

（1）熟悉Kafka消息系统的架构、原理和文件存储机制以及运用spark streaming提供的receiver以及 Direct 消费Kafka 消息。

（2）熟悉Zookeeper的架构，为其他分布式集群提供一致性服务

（3）能使用Flume采集不同应用场景的数据

（4）能使用Sqoop，logstash进行数据迁移，使用elasticSearch 存储以及查询数据

(5) 掌握Hadoop,Hive,Hbase等大数据分析，存储工具

(6) 掌握git/svn 项目管理工具

专业技能

2020.3-2020.12 上海华钦信息科技股份有限公司 大数据开发工程师

2018.12-2019.7 沿线教育 开发工程师

2017.11-2018.8 中和世纪投资公司 开发工程师

工作经历

项目名称 EUROCARS

项目日期 2020-3 到 至今

项目描述 EUROCARS 是一个风险控制项目。通过对报表数据的分析，进行风险评估。

项目架构 Hadoop +Spark+ Hive + Oracle + Sybase +Autosys +DTS

数据来源 旧sybase 平台数据迁移

责任描述 1 负责使用spark 技术进行sybase项目的迁移，将store procedure 迁移到大数据平台。

2 负责编写shell脚本，生成feed 文件，提供项目初始数据。

3 根据用户需求，生成对应的报表数据。

4 对比新老平台的数据，对于不匹配的数据找出原因并给出合理解释

5 使用jill file 部署相应的autosys job ，确保项目能在production 平台上平稳运行。

项目经验一

项目名称 易客数据管理平台

项目日期 2019-1 到2019-6

项目描述 易客数据管理平台，是一款用于分析用户行为数据的平台，通过对用户行为的分析，为所有的用户打上标签，用于供给需求方平台，精准的受众数据，使广告主可以精准的进行广告投放

项目架构 Hadoop+Flume+Kafka+spark+Hbase+Redis+ElasticSearch+Mysql

数据来源 本公司网站日志数据 + 第三方网站数据购买 + AdX 对接数据

项目实现 1 通过flume采集工具，将实时竞价数据从各个日志服务器采集到Kafka消息引擎中

2 通过Flume采集工具将用户行为日志存储到HDFS中，然后通过Spark对HDFS进行处理，然后将处理后的结果存储到Hbase中

3 从Kafka消息引擎中读取数据，进行实时报表（例如 展示量、点击量、点击率、参与竞价数、竞价成功数、千人成本等）数据展示，分析后的结果存入Mysql 数据库中

责任描述 1 将flume从各个日志服务器采集到的用户数据进行数据清洗，过滤掉不符合条

的数据后，以Parquet 的文件格式存储到hdfs上

2 根据用户行为数据，分别统计各个地域、各个媒体、各个渠道、操作系统的广告点击数，有效点击数，广告位竞价数，实时降价数，竞价成功率，广告成本费等相关数据，然后将分析后的结果存入MYSQL 数据库中。

3 过滤数据，为每日的数据打上各个标签，例如App标签、广告位标签、渠道标签、

设备标签、关键词标签、地域以及商圈标签，其中商圈标签是通过高德地图提供的逆向地图编码获取，这里首先会根据经纬度求出GeoHash值，然后通过GeoHash值从Redis服务器中查找商圈标签，如果找不到，则通过高德地图提供的APi获取，并添加进入Redis 数据库

4 将用户数据与上下文数据进行合并，得到当天的用户标签数据，并将数据存储到ElasticSearch 当中

项目经验二

项目名称 算力盒子用户分析系统

项目日期 2017-11 到2018-8

项目描述 算力盒子是一个b/s架构的web应用，算力盒子主要是做线上线下蚂蚁矿机租赁以及BTC,ETC 等虚拟货币交易的一个平台。算力盒子用户分析系统主要是针对算力盒子这个web应用进行用户行为，矿机租赁，虚拟货币交易等数据的分析，用于给公司的上层领导进行相关决策分析的数据分析系统。

项目架构 Flume + Hadoop + Hive + Kafka + Spark + MySQL

数据来源 主要是js埋点的日志数据，以及后台数据库提供的sql数据。

项目实现 1 通过flume 将日志数据采集到Kafka里，然后通过SparkStreaming 消费数据

2 使用SparkStream初步处理数据之后，将数据存入到hive的ods 层,作为源数据

3 根据项目指标，将hive表中的原始数据分离出维度层与集市层，最后将数据存入Mysql数据库，供给开发人员调用。

责任描述 1 负责flume日志数据采集的相关操作

2 负责用户数据，交易数据的初步清理

3 根据交易记录 分析各个币种每日最高成交价、最低成交价、平均成交价、每日成交数、每日求购广告数、每日售出广告数等各项指标

4 通过矿机产品数据，分析各个矿机产品的点击数，线上/线下交易总数，交易人，交易失败数，交易成功数等相关数据的各项指标

项目经验三

本人性格开朗大方，待人热情、真诚，能很好地融入到团队工作中，能吃苦耐劳，按时按量完成上级领导分配的工作任务，爱好打羽毛球和转呼啦圈。

自我评价