|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **基本信息** | | | | |
| 姓名： | 施丽源 | 工作经验： | 4年 |  |
| 手机： | 13395039962 | 学历水平： | 本科 |
| 邮箱： | shiliyuan\_li@163.com | 毕业年限： | 2016 |
| 性别： | 女 | 专 业： | 软件工程 |
| 英语： | 4级 |  |  |
| **求职意向** | | | | |
| |  | | --- | | 职位：大数据分析师 薪资：面议 目前状况：在离职，可随时上岗 | |  | | | | | |
| **专业技能** | | | | |
| |  | | --- | | * 编程语言  1. 能使用java进行MR、Spark的业务开发，掌握JVM垃圾回收机制及垃圾回收算法； 2. 能使用Scala进行Spark的业务开发，掌握伴生对象、RDD的属性等。 | | * Hadoop体系  1. 能够使用MR进行业务开发，掌握MR运行原理、Shuffle过程； 2. 掌握HDFS的读写原理； 3. 能够描述Hive架构原理，熟练使用HQL，掌握Hive的优化； 4. 掌握Hbase的架构及各个组件的工作机制。  * Spark框架  1. 掌握RDD的属性，特征，能够使用RDD算子进行Spark Core的业务开发； 2. 能够使用Spark Streaming进行数据的流式处理； | | 1. 能够使用Spark SQL进行数据的分析及处理； 2. 掌握Spark任务的启动流程及Shuffle过程； 3. 能够根据实际业务情况对Spark优化，能够对数据倾斜、OOM问题进行优化。  * 其他类  1. 熟练idea开发工具，git版本控制工具； 2. 能够使用Flume对数据进行采集和传输； 3. 掌握Kafka文件存储机制及其原理 4. 能够使用Mysql/Hbase/Redis对数据进行存储，熟练使用SQL； 5. 掌握ZooKeeper的架构和工作原理； | | | | | |
| **工作经历** | | | | |
| |  |  |  | | --- | --- | --- | | **2019.03-2020.01** | **软通动力有限公司** | **大数据分析师** | | **所属部门：**大数据分析部 **工作职责：** 1、负责对用户行为及产品运行等多维数据的分析和挖掘，为决策和优化提供数据支撑； 2、负责对分析结果可视的处理，提供报表等可视化视图。 | | | | | | | |
| |  |  |  | | --- | --- | --- | | **2016.09-2018.12** | **行吟信息科技(上海)有限公司** | **大数据开发工程师** | | **所属部门：**大数据分析部 **工作职责：** 1、负责数据仓库和大数据模块的架构设计和开发，针对业务层逻辑需求进行业务建模，满足业务需求； 2、负责对用户行为及产品运行等多维数据的分析和挖掘，为决策和优化提供数据支撑；  3、**负责公司内部同事大数据知识培训。** | | | | **2015.10-2016.08** | **沪江教育科技(上海)股份有限公司** | **大数据开发工程师** | | **所属部门：**大数据中心部  **工作职责：**  1、负责数据数据模型设计、ETL实施、性能优化；  2、根据需求使用大数据开发技术进行数据处理、查询、统计、分析等工作。 | | | | | | | |
| **项目经历** | | | | |
| |  |  |  | | --- | --- | --- | | **2019.03-2020.01** | **QQ浏览器分析** |  |   **项目架构：**Flume + Kafka + Spark Core + HDFS+ Hive + Sqoop + Mysql  **项目描述：**  QQ浏览器是目前市面上用户活跃度最高的浏览器之一。本项目主要是对该浏览器的用户行为进行分析，输出可靠结果，为其他部门提供决策。  **项目职责：**   1. 用户活跃度分析，包括日活，周活，月活等活跃分析； 2. 用户留存分析，包括次留，三留等常见的留存分析； 3. 用户行为分析，包括不同年龄段男女使用浏览器的情况，新增用户行为分析等； 4. 把分析结果可视化，如需要长期观察分析的则做出日常报表输出。  |  |  |  | | --- | --- | --- | | **2018.7-2018.12** | **小红薯社区内容数据支撑系统** |  | | **项目架构：**Flume + Kafka + Spark Core + HDFS+ Hive + Sqoop + Mysql  **项目描述：**  小红薯社区是小红书APP中的一款分享兴趣爱好的社区。社区种类有21种，包括视频、护肤、彩妆、时尚、美食、旅行、明星、婚礼、母婴等各类女性感兴趣的热门话题。该内容运营系统旨在分析热门社区话题、活跃在社区的用户等，分析忠诚的用户，计算用户粘度，为其他部门提供数据支持。  **项目流程：**   1. 使用Flume将数据读取到Kafka中，不同的数据分类采集到不同的topic中； 2. 通过Flume采集Kafka中的数据到HDFS中，通过Hive清洗数据后再次放到HDFS中，并与Hive建立映射； 3. 使用Spark Core进行离线统计分析各话题、用户等； 4. 将各项指标分析结果保存到Mysql中。   **项目职责：**   1. 将数据进行ETL清洗，并落地到Hive建立ODS层信息宽表； 2. 抽取相对应字段DW层建立事实表、维度表，并设计DWD层的中间表； 3. 分析DAU、MAU指标，计算用户粘度、黏性发展指数； 4. 分析忠诚用户占比、统计忠诚用户特征等； 5. 内容详情页转化漏斗分析。   **项目排障：**  Spark的优化 | | | |  | | | | | | | |
| |  |  |  | | --- | --- | --- | | **2017.08-2018.03** | **小红书流媒体高潜精准营销分析系统** |  | | **项目架构：**Flume + Kafka + HDFS + Hive +Sprark Streaming + Spark Core + Spark SQL + Redis  **项目介绍：**  小红书社区内广告的投放来源于社区达人分享+明星种草，针对用户点击行为数据进行实时计算，分析明星或达人对于社区内商品的转化率，据此统计各个广告的引流量情况和热门广告榜，为运营部门提供广告投放策略的数据支持。  **项目流程：**   1. 使用Flume收集用户日志，sink到Kafka的topic中； 2. Spark Streaming采用Director方式消费Kafka数据； 3. 通过Spark SQL + Spark Streaming分析各个指标； 4. 将最终结果缓存到Redis中，供前台缓存。   **项目职责：**   1. 实时点击量分析； 2. 广告转化分析，如个广告在同一内容模块的流量占比等； 3. 第三方广告分析，如第三方广告引流量分析； 4. 优质广告分析，如统计广告红黑榜等； 5. 渠道推广后的付费转化率。   **项目排障：**  Spark优化、Kafka消息的防止重复消费和避免数据丢失处理 | | | | | | | |
| |  |  |  | | --- | --- | --- | | **2016.12-2017.05** | **小红书福利社智能数据运支撑系统** |  | | **项目架构：**Flume + HDFS + Hive + MapReduce + Hbase  **项目介绍：**  小红书福利社是一个社交型跨境电商平台，通过用户在小红书网页端、APP端产生的海量点击数据进行大数据离线分析，并进行数据建模，生成主题报表为其他部门提供数据支持。  **项目流程：**   1. Flume采集数据到HDFS上，通过Hive以及编写UDF对数据进行ETL清洗； 2. 将清洗后的数据根据业务需求进行数据建模； 3. 使用HQL语句分析各项指标并建立ODS、DW、DWD、DM层，将最后的结果按主题展示，并保存到Hbase。   **项目职责：**   1. 分析用户规模KPI，如分析DNU，WNU，MNU、用户环比增长等； 2. 分析用户活跃KPI，如天/周/月登陆峰值，分析周/月用户流失率等； 3. 分析用户行为KPI，如页面单挑转化率，次日留存、七日留存率等 4. 分析订单规模KPI，如日、周订单额的环比、同比增长，每日、每月订单额峰值分布等。   **项目排障：**  Hive数据倾斜，MR的优化 | | | | | | | |
| |  |  |  | | --- | --- | --- | | **2016.02-2016.08** | **CCtalk-LIS(一期)** |  | | **项目架构：**Flume + HDFS + MapReduce + Hive + Sqoop + MySQL  **项目介绍：**  CCtalk-LIS，通过大数据分析技术，多维度，细致化分析学生多次考试中的进退步信息，让学生精准查缺补漏，提示成绩。  **项目流程：**   1. 使用多个Flume将原始数据采集到一个Flume上，其主要目的是将Flume产生的多个小文件进行合并，最后将合并的数据sink到HDFS上； 2. 使用MR对数据进行清洗，包括数据预处理，JSON解析等，最后将数据格式化输出到HDFS上并映射成Hive表； 3. 使用Hive按照指标对数据进行统计和分析 4. 使用Sqoop将统计结果和分析结果导出到MySQL中用于前端展示。   **项目职责：**   1. 将数据仓库建立为四层，ODS操作数据层、DW数据仓库层、DWD中间结果集、DM数据集市层； 2. 使用HQL进行相关的数据转换，包括自定义UDF书写、窗口函数的使用等； 3. 使用HQL进行相关指标的实现   **项目排障：**  Flume的高可用和故障转移，Hive的优化 | | | | | | | |
| |  | | --- | |  | | | | | |
| **教育背景** | | | | |
| |  |  |  | | --- | --- | --- | | 2011.9-2016.7 | 厦门理工学院 | 本科-软件专业 | |  |  |  | | | | | |
| **个人评价** | | | | |
| |  | | --- | | 1、参与过多个项目开发，熟悉常见的业务处理，能快速融入团队。 2、遇到问题善于分析、总结，对新技术有着强烈的好奇心。 3、乐于与用户、同事、领导沟通，以便快速解决项目中遇到的问题。 4、抗压能力强，抗压适应高度工作环境，对未来一直保持乐观的心态 | | | | | |