

联系方式

- 手机：13175121637
- Email：chenzebedee@gmail.com
- QQ/微信号：331471867

个人信息

- 陈绍迪/男/1995
- 本科/杭州电子科技大学信息工程学院
- 工作年限：3年
- 期望职位：大数据工程师
- 期望城市：杭州

技能清单

1. 熟练使用 linux 的命令行，能熟练编写 shell 脚本，以及熟练使用 Nginx
2. 熟练使用Git与SVN版本管理工具
3. 熟练掌握 Hadoop 生态系统，熟练 Hive、HBase、Kafka、Sqoop、Flume等
4. 熟悉 MySQL，Oracle 等数据库，熟练使用SQL进行开发，有MySQL运维优化经验
5. 熟悉使用 JAVA 语言，有良好的编码规范，熟悉使用SpringBoot框架进行开发
6. 熟练使用 OGG，Canal 数据库同步工具
7. 熟悉 Docker 命令操作，以及 DockerFile 的编写
8. 熟悉 Flink、Spark等实时分析工具，了解RDD模型
9. 熟悉使用 scala、Python语言
10. 熟悉 Ambari 与 Cloudera Manager 大数据管理系统

工作经历

医惠科技有限公司（2019年9月 ~ 至今）

工作职责

1. 大数据环境搭建，维护，优化
2. 大数据应用开发
3. 数据转换应用开发

杭州玛瑙湾科技有限公司（2018年6月 ~ 2019年9月）

工作职责

1. 大数据环境搭建，维护，优化
2. 离线数据同步

3. 数据 etl 清理
4. 实时同步数据处理
5. 数据应用
6. 数据平台架构搭建

北京思特奇信息技术股份有限公司（实习）（2017年9月 ~ 2018年6月）

工作职责

1. 报表开发
2. 离线etl开发

项目经历

医院大数据平台搭建

项目描述

在医院信息化建设中，数据中心建设是最重要的核心部分。但是目前国家还没有出台相关标准、各厂商建设数据中心的目标也不统一，导致医院数据中心建设的效果良莠不齐、为医院带来的价值有限，离医院对数据中心的期望目标还有不少的差距。还存在以下问题：1. 数据中心建设需求不明确、目标感不强。2. 数据条理性、结构性不强，数据质量无法验证。3. 数据中心缺乏管理工具，可视化程度不强。4. 大数据存储，数据实时性等技术参数较难真正达到

而我们为了建设体系化、层次化、模型化、标准化、集中化、合理化的数据中心。解决上诉问题，为医院进行的大数据应用平台的搭建与开

技术应用： oracle + ogg + Spring Boot + Kafka + HBase + Ambari

主要工作职责

1. 构建ods数据库,cdr数据中心,与Ambari大数据环境
2. 构建源端数据库到达ods的ogg数据同步链路， ods 数据实时同步到 cdr 数据中心与 HBase 的数据链路
3. ods,cdr,Ambari一键脚本开发
4. ogg推送Kafka延迟堆积等问题解决
5. HBase 预分区处
6. 大数据组件应用开发

数据中心交互平台建设

项目描述

在金融行业中，为出于安全考虑，让出借人与借款人信息，存在通信壁垒，但是由于不方便宏观分析，于是需要一个线下安全的统一的数据中心，并且拥有分析能力的数据应用中心。

通过同类数据库进行数据初次同步，通过etl手段对数据进行清洗，同步至hive与HBase中，并通过

Canal同步到Kafka消息队列中, 实现实时数据分析, 由于用户数量大量增长,传统型数据库应对不了大数据量分析,并且需要特殊数据类型分析,需要构建一个对外使用的数据中心,用于运营与风控等数据分析。对数据进行挖掘分析, 根据时间段判断活跃时间, 进行广告推送活动, 进行推广活动时, 会发生无限使用新用户获取活动金的薅羊毛行为, 对于此类行为, 我们进行了联系人对比, 以及机型监控的方法进行风控。对于骗贷案例, 我们通过数据分析, 做好第三道风控。

技术应用：MySQL + Canal + Kafka + HBase + Spark + Docker + Spring Boot

1. Docker 封装HBase、Spark、Canal容器, 各自独立
2. Canal配置按表同步数据到Kafka中, 开发Kafka消费者代码, 数据存入HBase中
3. 日活数据分析, 通过HSQL分析日活数据
4. 联系人碰撞分析, 开发HBase与Spark计算代码
5. 评分卡模型, 开发MapReduce程序
6. 数据实时同步上报网信办

业务经理人行为轨迹分析及万号手机版动态日报平台

项目描述

根据 10000 号大屏制作方便各地业务员查看的手机版动态日报平台, 以及对业务经理的行为轨迹分析。主要任务如下:

1. 10000 当月话务情况、人工话务走势、各地市话务情况(日均)、人工话务热点。
2. 全省及区域性故障实时预警、宽带故障申告走势、宽带故障原因分类
3. 主量业务发展情况、精准营销转化情况、重点业务发展情况、精准营销转化商机类型
4. 咨询投诉万用户工单率、投诉热点、百万用户越级申诉率
5. 业务经理签退补偿与工作有效性判断

技术应用：Hadoop + Hive + MySQL + HBase + Oracle + OGG + Kafka

主要工作职责

1. 通过MR实现业务经理的签退功能
2. 通过Hive对业务人员进行工作范围确认
3. FTP同步源端数据, 导入HIVE中
4. Hive分析数据之后导入MySQL中, 用于展示
5. OGG同步Oracle数据至HBase中, 通过hive进行分析处理, 分析结果存入新的表中
6. 监控HDFS文件变化, 使数据同步到Kafka中
7. 通过Canal同步源端Mysql数据到Kafka中
8. 通过SparkStraming进行Kafka的数据消费。