

联系方式

- 手机: 17315031651
- Email: q726575153@gmail.com
- 微信号: 17315031651

个人信息

- 代乐/男/1996
- 专科/无锡科技职业学院
- 工作年限: 3年
- 技术博客: mrdlontheway.github.io
- Github: <https://github.com/MrDLontheway>
- 期望职位: 大数据开发,实时计算,离线计算
- 期望薪资: 面议
- 期望城市: 上海

个人简介

- 1.能独立部署安装 cdh ambari 大数据集成环境 集成kerberos 以及其他相关权限认证
- 2.熟练使用 spark 有spark源码 修改以及阅读经验(version1.6 and mster 3.0.0), 包括 spark streaming、spark core、spark SQL 开发, 了解 DAG 划分, 熟悉其 shuffle 过程, 常用 rdd, df 及常见的性能调优方式, 能用 spark 对接各种数据源, 进行业务处理。能够用 scala 开发 spark 相应程序, 对接常见的各种数据源。
- 3.熟练使用 kafka, 熟悉其架构、实现原理 阅读过Producer 以及 Consumer 流程源码(version 1.0.1) 等。
- 4.熟练使用elk 全家桶, 了解es 实现原理,查询 以及 数据录入优化 以及 hbase + es 离线数仓+索引 解决方案
- 5.熟练使用flink blink 处理分析流式数据 熟练使用flink cep 实现复杂事件处理
- 6.熟悉使用 hive hbase 进行数仓建设 熟悉基本的数仓分层策略
- 7.熟悉使用spring ,spring mvc , springboot 开发java ee 后台相关项目
- 8.熟悉使用aliyun 相关产品 emr maxcompute odps rocketmq
- 9.有 kubernetes,docker 容器相关使用经验

Apache Accumulo Contributor

自我学习能力较强,可以快速定位问题,对大数据生态圈组件理解深入,阅读过 spark kafka netty 相关源码 经常在开源社区贡献

正在学习OLAP方面

工作经历

Paypal (大数据开发)[2020年7月-至今]

Data-Foundation [2020年7月-至今]

Quality Monitoring , onboard all pipeline jobs , distribute collect metric task daily according to plan , use kafka Mirror Maker Synchronize to the unreachable cluster , collecting service accept task and execute then report the result to monitor, the monitored data source types are Oracle Teradata Hdfs Hive Druid...

北京赛斯信安技术股份有限公司 (大数据开发)[2018年10月-2020年7月]

大数据计算分析平台 [2019年10月-2020年7月]

开发数据处理平台, 能让用户通过页面简单的拖拽完成复杂的数据处理过程 分布式并行处理简单化 兼容大部分离线 在线数据源 数据处理视图化 支持 数据源数据自定义解析 java scala 动态调用 udf 函数动态注册管理 python shell 脚本处理 支持计算引擎 spark flink

规则引擎 spaas(streaming Processing as a service) [2019年4月-2020年7月]

项目主要为用户以及其他系统提供可靠易用的流式及离线 分布式处理service

项目主要功能

- 支持可视化, 配置化, SQL 化开发实施流式项目
- 支持基于业务策略的事件时间戳流式处理
- 支持UDF的注册管理和动态加载
- 支持多目标数据系统的并发幂等入库
- 支持多层sql 业务流处理
- 支持自定义处理流程
- 任务资源动态管控
- 安全权限管控和多租户资源管理 适配Kerberos 权限认证

- flink-cep 模式匹配 复杂事件处理
- 支持流式任务处理过程热更新
- ...

目前支持数据源如下:

- DB库(mysql sqlserver oracle gbase 支持所有JDBC 驱动)
- 数据仓库(hive hbase hdfs odps(aliyun)))
- 消息队列 kafka (AVRO CSV JSON) aliyunMQ
- 索引库 elasticsearch

我在此项目中 负责 spark 及 flink 任务主体执行部分 项目过程中最大的问题是之前没使用flink在真实开发中应用 在短时间内快速掌握flink 计算相关技术 快速完成项目初期阶段flink 计算从0 到1 的过程

智能分析平台 (graph database and compute) [2018年10月-2020年7月]

项目主要从关联数据中建立关系 以Graph 图数据形式持久化存储 从大量图数据中帮助用户找到关心的数据联系

我在此项目中负责图库相关存储以及计算工作 为应用提供易用的图库查询操作API

Graph 技术实现

- accumulo元数据存储
- elasticsearch索引
- zk 存储 filed 属性 meta data
- vertex edge 数据visibility 与 用户user Auth 各用户对数据的可见性
- spark graphx 对接 graph databse 做并行计算 (进行中)

上海雁展科技有限公司[2018.06-2018.10] (大数据开发)

c2b 数字化供应链平台 [2018.06-2018.10]

scala spark hive elasticsearch

搭建一个跨企业间的全业务链数据平台, 通过贯通客户、整车与零部件供应商之间供应链业务数据, 实现零部件的即时制造与供货。

主要负责 数据处理 接入es部分开发 通过hive on spark 定时批量从hive拉取数据,分析 拆单 打散 导入es 后台从es获取数据

无锡实创科技有限公司 (java 开发)

葡萄塘电商平台 (直播+电商) [2017.06-2017.08]

负责商城团购 预售 满赠满减活动 以及用户积分商城的模块开发

主要实现 平台 团购 预售 自己商品满赠满减 积分商城 直播间分发红包功能.

个人作品

技能清单

- 大数据集成平台: CDH,Ambari,HUAWEI FusionInsight,aliyun 飞天
- 大数据计算: Spark,Flink
- 大数据存储: HDFS,Hive,Hbase,Accumulo
- ELK Kafka
- 容器相关: kubernetes,docker
- OLAP: Druid,Clickhouse (了解使用过)
- 服务器相关: Linux,常用Shell编写
- CI/CD: GitLab,Jenkins,Nexus
- 后台开发: ssh,ssm,mysql,tomcat,springboot,dubbo,spring cloud,oracle、mysql、redis等, 能够实现简单的sql优化
- 前台WEB: bootstrap,jquery,jsp,freemaker