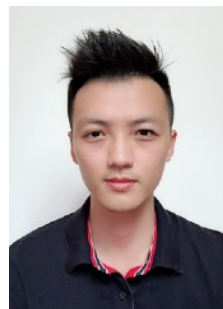


个人资料

姓名: 杨昌俊 性别: 男
年龄: 30 英语: CET-6
学历: 本科 工作年限: 5年
邮箱: yangchangjun010@hotmail.com
电话: 18682426031
个人主页: <https://aprilyoungs.github.io>



求职意向

目标岗位: 大数据开发工程师
期待薪资: 面议
期望地点: 深圳
目前状态: 离职, 一周内可上岗

专业技能

- 熟练使用Java开发语言, 面向对象编程, 接口式编程, 常用核心类库
- 熟练使用Scala编程语言, 包括面向对象编程, 函数式编程, 模式匹配, 隐式机制
- 熟悉Spark的原理和流程, 并能用Spark API开发应用程序, 熟悉Spark Streaming组件, 并能用其开发实时计算系统, 了解Spark源码
- 熟练使用Linux / Unix系统, 编写shell脚本
- 熟悉Hadoop的系统架构, 掌握Hadoop集群的搭建, 并能用CDH快速部署Hadoop集群
- 熟悉MapReduce的原理和流程, 并能用其API开发应用程序
- 熟悉Zookeeper的架构、运行的原理, 并能应用其为集群提供协同服务, 如HA
- 熟悉Hive的工作原理, 数据仓库理论, 完成对数据主题抽取和多维分析
- 熟悉使用MySQL, 掌握其存储原理及查询速度优化
- 熟练使用Sqoop, DataX工具, 实现非关系型数据库与关系型数据库表数据互导
- 理解Hbase的存储原理, Hbase存储架构, 实现数据的毫秒检索
- 掌握Redis内存数据库的基本原理, 实现数据的毫秒查询
- 掌握Kafka、Flume数据采集工具的原理, 实现流式数据的过滤和分析
- 理解Impala的工作原理, 对海量数据即席查询
- 掌握Airflow, Azkaban等任务调度系统
- 熟练使用Git、SVN等版本管理工具
- 无障碍阅读英文技术文档, 听说良好, 有外企工作经验

工作经历

2020.03 ~ 2021.04 Medicus AI GmbH

开发工程师

部门: 产品研发部

工作描述: 负责新产品研发及项目的实施; 定期汇报项目进展, 提出实验方案改进建议; 与其他部门配合保证产品落地质量; 编写工作相关的技术文档, 单元测试, 保证产品质量; 配

合运营等其他部门，提供产品相关技术支持。

2019.03 ~ 2020.03 深圳乐智教育科技有限公司 开发工程师

部门: 产品研发部

工作描述: 负责新产品研发及项目的实施; 定期汇报项目进展, 提出实验方案改进建议; 与其他部门配合保证产品落地质量; 编写工作相关的技术文档, 单元测试, 保证产品质量; 配合运营等其他部门, 提供产品相关技术支持。

2016.10 ~ 2019.02 康美健康云服务有限公司 iOS开发工程师

部门: 软件研发部

工作描述: 基于iOS端技术的开发, 负责iPhone和iPad应用程序的编码工作; 协助产品经理进行应用及人机界面软件模块的设计和实现; 完成与工作相关的技术文档编写; 配合运营等其他部门, 提供产品相关技术支持。

2015.04 ~ 2016.05 富士康科技集团 采购员

部门: 富金机全球采购部

工作描述: 基于负责塑料加工件的采购工作, 包括询价, 比价、议价, 评估, 索样, 决定, 请购, 订购, 协调与沟通, 催交。

项目经验

➤ 项目1: 智慧物流大数据分析调度平台项目

开发环境: IDEA+maven+JDK

系统架构: Sqoop+Hive(Hive on spark) +Airflow+Hadoop+Spark Mlib+Kafka+Redis+HBase+Spark Structured Streaming+SpringBoot

项目描述:

依托大数据技术的智慧化物流平台, 保证物流系统各个环节高效的运转。实现功能: 仓储预测, 智能运单调度(车货匹配), 指标统计, 实时车辆监控管理, 车辆轨迹回放

责任描述:

1. 根据项目需求选择合适的技术实现方案
2. 根据业务需求设计数据仓库每一层的表格和字段
3. 使用数据迁移工具Sqoop把MySQL业务层的数据导入到Hive中
4. Hive数据ETL, 处理成机器学习算法需要的数据, 最终把数据导出为csv格式供负责机器学习的同事使用
5. Hive数据ETL, 统计各种指标并导出到MySQL, 供前端展示
6. 编写shell脚本自动化处理Sqoop数据导入Hive, 并处理Hive数据
7. 使用Airflow任务调度系统每天定时调用shell脚本
8. 使用仓储预测的数据生成车辆调度方案, 并输出到MySQL, 供Web端使用
9. 使用Nginx服务器接收车载客户端的监控数据, 并转发到Kafka
10. 使用Spark Structured Streaming实时拉取Kafka数据, 分别把处理后的数据存到Redis和HBase, 供Web端展示; 把检测到的数据异常发送到Kafka供Web端拉取
11. 使用Impala查询临时提出的数据统计指标

➤ 项目2: 企业电商离线数仓项目

开发环境: IDEA+maven+JDK

系统架构: MySql+Flume+Hive(Hive on spark)+DataX+Hadoop+Azkaban+Impala+Atlas+Griffin

项目描述:

离线数仓项目, 数据来源业务数据、日志数据, 定期清洗并分析数据, 包括日志数据分析和业务数据指标分析

责任描述:

1. 根据项目需求选择合适的技术实现方案
2. 根据业务需求设计数据仓库每一层的表格和字段
3. 使用数据采集工具Flume把日志文件聚合并存到HDFS, 再使用Hive外部分区表加载数据到Hive表
4. 使用数据迁移工具DataX把MySql业务层的数据导入到Hive中
5. Hive数据ETL, 统计各种指标并导出到MySql, 供前端展示
6. 编写shell脚本自动化处理DataX数据导入Hive, 并处理Hive数据
7. 使用Azkaban任务调度系统每天定时调用shell脚本
8. 使用Atlas管理数据仓库的元数据, 方便溯源, 知道最终结果是怎么计算出来的
9. 使用Griffin监控数据质量, 保证输出结果的准确性
10. 使用Impala查询临时提出的数据统计指标

➤ 项目3: 应用日志分析系统

开发环境: IDEA+maven+JDK

系统架构: MySql+Flume+Hive+DataX+Hadoop+Azkaban

项目描述:

离线数仓项目, 数据来源App埋点日志数据, 定期清洗并分析数据, 包括行为分析和业务数据指标分析

责任描述:

1. 根据项目需求选择合适的技术实现方案
2. 根据业务需求设计数据仓库每一层的表格和字段
3. 使用数据采集工具Flume把日志文件聚合并存到HDFS, 再使用Hive外部分区表加载数据到Hive表
5. Hive数据ETL, 统计各种指标并导出到MySql, 供前端展示
6. 编写shell脚本自动化处理DataX数据导入Hive, 并处理Hive数据
7. 使用Azkaban任务调度系统每天定时调用shell脚本

教育经历

2010/08 -- 2014/07 大连工业大学| 材料科学与工程|本科

自我评价

参与过多个项目的开发和迭代, 熟悉常见业务处理, 能快速融入团队; 逻辑思维能力强, 思路清晰, 热爱学习, 有整理学习心得的习惯, 对新技术有着强烈的好奇心; 乐于与同事和领导沟通, 以便快速解决项目遇到的问题; 英文良好, 能无障碍阅读英文技术文档。