

# 国庆假期任务安排

Author: zhouhan

Date: 2025-09-25 19:23:23

Desc : 2302A 专高 6

## 1. Minio 安装

可安装单节点或分布式的 minio , 本地部署或者 docker 都可

minio 上传一个 csv 文件, 使用 hive 映射该文件为表, hive 可查询, 能实现 ds 分区为佳, 使用时间分区, ds=20250925

## 2. CDH 安装 hbase

创建 hbase 表, 数据由 Flink 写入 FlinkAPI 或者 FlinkSQL 都可以

hbase rowkey 的设计, 符合企业原则

hbase 数据进入后, 使用 hive 和 hbase 的映射表进行映射, 数据需要可查询, 实现可分区为佳

## 3. python spider 爬虫

3.1 爬取中国气象数据, 数据要求 10 s 更新一次, 增量更新

3.2 爬取中国外汇当日市场数据数据 10 s 更新一次, 增量更新

ps : 环境可移植, 代码需要在 linux 中可运行, 使用 conda 环境进行配置, 配置反爬手段进行预防

## 4. 使用 VLLM 框架部署 Qwen3-0.6B 模型, 模型需要支持 RestFul 接口的调用

4.1 qwen 模型使用 2 种方式进行部署分别是基于 CPU 和 GPU 运行

4.2 qwen 模型使用云端进行部署, 开放远程调用的 callback

4.3 qwen 模型需要微调, 使用 LoRA 框架预先训练数据集, 针对地址信息进行调整, 数据处理使用 PySpark

## 5. 安装 PostgreSQL & SQLServer & Doris-2.1.6V 数据库 (linux)

分别创建 2 张表, 使用 FlinkCDC 读取这两张表的数据, 并写入下游的 Kafka 主题中, 使用 doris 的 Roadtime Load 方式进行数据加载, 使用动态分区, 使用分桶

## 6. 安装 FineReport 11 Version

所有代码, 所有实现逻辑, 需要写文档, 说明是怎么做的, 你做了什么工作, 怎么进行的实现, 有没有更好的方法去做  
代码需要上传 Github 中, 最后一天上传的 & 和别人代码类似的, 请说明实现逻辑和实现思路