

个人信息

- 甘家城 / 男 / 1996
 - 手机/微信: 15968182251
 - 邮箱: 935669873@qq.com
 - 本科 / 杭州电子科技大学 / 电子科学与技术专业 / 19届毕业生
 - 技术博客: <https://ganjiacheng.cn>
 - Github: <https://github.com/webgjc>
 - 期望职位: 大数据平台开发 / 后端开发
 - 期望薪资: 13-16k
 - 期望城市: 杭州
-

工作经历

• 大搜车软件技术有限公司 (数据研发工程师) (2019-04 ~ 至今)

◦ 元数据采集与数据地图

该项目主要是对hive元数据进行采集、管理,同时也支持mysql类型元数据。采集的元数据主要包括表字段信息、分区信息、表字段血缘信息、关联信息、变更记录、热度信息、表负责人、表业务划分、权限信息等,从metastore、hivehook、离线调度平台、ranger采集到整合到后端存储,进行了统计与分类,并支持在页面上的展示搜索。项目使用python/django/celery实现,存储用的mysql和redis,本人在项目中为后端开发,项目基本为从零搭建。项目中的元数据采集使用celery的定时任务,采用新表和全量两种同步策略,也提供手动同步的方式。具体可以参考我的原创博客[讲讲元数据与元数据采集](#)

◦ 数据标准与数据建模

该项目一方面是对数据标准、数据码值进行定义、录入、管理和对外提供。另一块是数仓建模工具,通过表单,hql,excel等方式导入建表信息,提供ER图画图模块对导入表进行建模作图。建模的时候可以直接应用到数据标准和码值信息。项目使用java/springboot(optimus)/mybatis搭建,存储使用mysql,本人为项目的后端开发,项目为从零搭建。

- 项目中在码表中因为码表是自建的,数值字段是自定义的,因此使用了json存储,且用到分表存储;
- 在hql导入中用到了hql的create语句的解析,使用的是antlr4和hive的hplsql.g4然后用java实现visit模式的create_table;
- 渲染导出hql使用musache模板引擎;
- excel导入使用easyexcel实现;

◦ 自助查询与敏捷分析

该项目是主要是hive的查询工具,类似hue,同时也支持连接其他各类数据库。功能有筛选库表,hql查询,自配参数查询,执行日志/进度获取,hql格式化,hql血缘,hive资源队列信息,

结果集的(存储, 下载, 分析计算做图), sql模板(管理, 分享等)。本人为项目的后端开发之一, 从中间开始接手并开发新功能。项目后端使用python/django/celery实现, 与数据地图为一个大后端。

- 查询使用ranger控制查询权限, 到表;
- 筛选库表使用数据地图采集数据;
- hql查询使用pyhive库连接hiveserver2(celery/redis异步队列);
- 查询日志/进度使用pyhive异步模式下poll获取到的progress信息中的mapreduce信息和进度, containerLog参考yarn client内部实现的获取日志;
- hql格式化和hql血缘使用antlr4解析hive的hqlsql.g4, 并使用java手动实现格式化和血缘;
- 结果集使用json格式存储至oss, 使用pandas对数据处理计算作图可视化;

• 默安科技有限公司(实习) (运维开发) (2018-07 ~ 2019-03)

◦ 产品打包部署更新等自动化脚本和平台

为各个产品编写打包的脚本, 部署脚本, 更新脚本等。通过搭建的平台使用可以直接将产品打包部署到机器上, 可以更新该环境的代码, 和查看环境中产品运行的状态。平台项目使用python/django构建, 日常的各个脚本使用shell编写完成, 同时在版本迭代有变更或依赖包有变动时实时更新。

◦ 其他一些服务如vpn, 文档服务器, 官网等的维护

搭建维护vpn平台使用openvpn实现, 对接钉钉群完成在群内获取验证码登录, 并做登录记录和统计。

文档服务器和官网做日常维护和更新。

• 艾锡信息技术有限公司(实习) (python开发) (2017-12 ~ 2018-04)

◦ 量化策略与相关平台工具

主要实现一些金融研报的策略, 在开放平台完成回测。

搭建金融相关使用工具的平台, 主要使用python/flask。

技术之谈

- 关于编程语言, 原来python用的比较多, 后来做数据产品的时候java也学了很多, 框架的话分别为django和springboot;
- 对前端有所涉猎, 在校期间的大学生创业项目自学了html css和js, 后来也在大搜车学了react和内部用的前端框架, 并在实际项目中动手写了一部分;
- 之前做过段时间运维, 日常的运维也能自己完成, 打包部署装机, shell脚本等;
- 版本管理git用的比较多, svn很早前也用了一段时间;
- 对于算法一直在跟进, 偶尔也算在leetcode有做题吧;
- 日常会有很多奇思妙想实现一些demo, 在github和博客里有体现;
- 没有直接很深入的研究框架和底层, 但很多都是用到了会去深挖一下;
- 之前在校一直有写博客, 中间工作断了一段时间, 后来心血来潮继续写;

最后一席话

- 毕业刚一年，很多东西也是现学现用，希望能有很多的机会在技术更多的深入，在行业上更多的发展，让自己尽快成为某个领域的专家。之前一年做了大数据相关的平台开发，虽然不是接触最难的任务调度的部分，但也感受到这块的潜力，希望能继续在相关的领域贡献力量。最后感谢阅读，欢迎交流！