



## 甘文浩

26岁 男

求职意向：大数据架构师 大数据工程师

☎ 13242851363

✉ qq349074225@live.com

🌐 github.com/archongum

### 基本信息

居住地：广东 广州

籍贯：广东 江门

身高：180cm

体重：65kg

### 求职意向

- 期望职位  
大数据架构师/大数据工程师
- 期望地点  
广州市

### 语言技能

英语六级

客家话/普通话/粤语

### 职业技能

Linux | 熟练

Java/Python | 熟练

Ambari生态(Hadoop/Hive等) | 熟练

Docker/Kubernetes | 掌握

Presto/MySQL/SQL | 熟练

### 兴趣爱好

排球 / 单机游戏 / 看新闻

## 自我介绍

热爱编程，热爱开源，热爱新技术。

个人成就：Presto社区成员/贡献者，贡献了PR一个；Apache NiFi社区成员/贡献者，贡献了PR三个；

学习方面：学习能力强，主动学习新晋技术。例如最近的kubernetes；

英语方面：自2018下半年以来，坚持使用Google（英语）作为日常生活、学习、工作的搜索引擎；学习工作上坚持通过官方英文文档、Stack Overflow等渠道提高技术水平；每天追踪Chrome首页推荐、Google News上的文章；

工作方面：有良好的沟通能力和逻辑思维能力，能够和团队成员进行有效的沟通，通过技术手段解决问题；对技术有热情，能够不断深入研究新技术。

## 教育背景

○ 华南农业大学 本科  
2012-09 至 2016-07



## 工作经历

○ 广州市小蓝车网络科技有限公司

中级大数据工程师 2018-04 至 2019-12

1. 负责整套大数据平台/数仓平台的搭建以及维护工作；
2. 负责Python机器学习的开发工作；
3. 探研大数据的发展方向，优化现有技术框架；
4. 参与Spring Boot数据对外接口开发；
5. 曾获得季度优秀奖。



○ 南京烽火星空通信发展有限公司

Hadoop大数据工程师 2016-07 至 2018-01

- 1, 根据客户的需求，做MR、HiveQL、ETL等开发；
- 2, 负责深入学习MR开发，优化问题；
- 3, 日常使用Linux命令行编写数据处理脚本、管理Hadoop组件等；



## 项目经历

○ 开源项目经历 2019-01 至 2019-12

开源项目：Presto，Apache NiFi

加入社区：Presto社区成员/贡献者，Apache NiFi社区成员/贡献者

主要贡献：

1. prestosql/presto的pull/1264：优化相关
2. prestosql/docker-images的pull/27：Bug相关
3. apache/nifi的NiFi-6305，NiFi-6442，NiFi-6480：优化，Bug相关

○ 大数据平台/数仓平台搭建 2018-04 至 2019-12

组件使用：Ambari(HDFS/Hive/Kafka)、Presto、NiFi、Spark、DataX、Azkaban、Prometheus、Grafana等。

平台构思：存储与计算分离。

平台介绍：

存储层使用HDFS；

计算引擎以Presto为主，DataX为辅助；

调度使用NiFi/Azkaban；

指标监控使用Prometheus和Grafana；

利用Presto SQL跨数据源以及高效计算等特点，实现整个平台OLAP、数仓的ETL、对外接口的

ad hoc即席查询开发工作；

两个测试环境使用Docker一键搭建，省去分别搭建的时间。

## 商品自动补货/商品推荐 2018-04 至 2019-12

工具算法：Python+SQL、sklearn的随机森林回归（商品自动补货）、jieba分词+相似度计算（商品推荐）；

实现步骤：

1. 结合数仓，使用Presto SQL取数；
2. Python读取SQL返回的结果为DataFrame并做适当的预处理（离群点检测、归一化、标签编码等）；
3. 调用对于接口sklearn/jieba，训练集进行训练，得到模型；
4. 根据实际需求，选择是否持久化模型到Redis存储层；
  - 4.1. 如果选择不持久化模型，则直接调用模型获得预测结果并保存到数仓；
  - 4.2. 如果选择持久化模型，则通过外部触发来调用模型接口（使用Spring Boot自行开发）做预测并返回。

## 在校经历

社团经历

♀ 校排球队 副攻 2012-09 至 2016-07

奖学金

三等奖学金 2013-12