

秦续业

个人简历

杭州市文一西路 969 号西溪园区 1 号楼小邮局

+86 186 161 09836

✉ qin@qinxuye.me

🌐 www.qinxuye.me

📍 chineking



教育背景

- 2011-2014 计算机科学与技术硕士, 自适应分布式实验室, 上海交通大学.
- 2005-2009 软件工程学士, 东南大学.

工作经历

- 2014.3-2015.9 阿里妈妈, 研发工程师
- 2015.9-今 阿里云, 资深研发工程师

专业能力

- 编程语言 Python, Java, C/C++, JavaScript, Go, C#
- 开发环境 Linux, Emacs, Eclipse
- 开源软件 Hadoop, Spark, Lucene, MongoDB, Django, Tornado, OpenStack, Neo4j

社区

- 博客 <http://qinxuye.me>
- GitHub <https://github.com/chineking>
- BitBucket <https://bitbucket.org/chineking>

项目经历

- 2015-至今 **PyODPS DataFrame** 框架, 公司项目 (Python, ODPS), 现有的 pandas 库提供丰富的 API 尤其是 DataFrame API 来操作结构化数据; 同时阿里云 ODPS 本身作为大数据处理平台, 提供了海量数据的能力, 其中 ODPS SQL 是 ODPS 上主要的结构化数据处理语言。然而, Pandas 作为单机的库, 计算能力有限; ODPS SQL 能处理大量数据, 但受限于 SQL 的表达能力。因此, PyODPS DataFrame 框架提供了一种类似于 Pandas DataFrame 的 API, 但是能运用 ODPS 的海量数据计算能力, 对结构化数据来执行查询。DataFrame 框架目前将所有操作编译成 ODPS SQL 来执行, 也能在本地使用 pandas 计算。
项目地址: <https://github.com/aliyun/aliyun-odps-python-sdk>
- 2015 营销效果平台归因 API, 公司项目 (Java, ODPS), 在广告的营销效果分析中, 常常需要将一个效果比如成交, 归因到一个或者多个用户行为比如点击上, 而归因的来源可能多种多样, 归因的优先级以及维度也不尽相同。归因 API 提供 XML 方式的配置, 支持视图、维度和优先级的配置, 并提供函数的支持, 使得不需要任何单独的代码编写, 就可以在 ODPS 上完成各种不同场景的归因计算。其中函数的支持使用 JavaCC 来进行词法和语法分析。
- 2014 **ODPS Python SDK**, 个人项目 (Python), ODPS SDK 的 Python 实现, 支持对表、资源、函数等的操作, 也支持 Tunnel API 来进行表数据的上传下载。
- 2013 **Cola** 分布式爬虫框架, 研究项目 (Python), Cola 是一个分布式爬虫框架, cola 集群由一个 master 和多个 worker 组成。用户只需编写 url 的匹配规则, 以及为每个规则编写解析函数。任务被提交后, 就会被 master 分配到各个 worker 上来运行。每个 job 运行时, 各个 worker 间会共享一个分布式消息队列, 待抓取的 url 会被放置在队列中。队列还担任去重的任务。我独立完成了整个架构设计和全部代码的编写。
项目地址: <https://github.com/chineking/cola>

- 2012 新浪微博用户推荐, 研究项目 (Java), TF-IDF 常常在网页搜索中用来衡量一个词在网页中的重要程度, 我们将 TF-IDF 应用到新浪微博用户推荐, 首先我们扫描用户的所有微博, 然后算出 TF-IDF 值最高的几个词, 最后通过这些词来计算出两个微博用户间的相似度, 并根据相似度进行推荐。我在实验中负责了数据的获取以及 Hadoop 程序的编写。
- 2012 云备份工具, 课程项目 (Python), 在用户设定了备份文件夹后, 云备份工具会自动将文件备份到多个云存储中。包括 Amazon S3、微盘、以及 Google Cloud Storage。在项目中我负责了整个架构的设计, 并完成了各个云存储的客户端 API 的实现。
项目地址: <https://github.com/chineking/cloudbackup>
- 2010 个人博客, 个人项目 (Python, Django), 博客支持包括文章的管理、评论等基本功能, 且支持多种第三方登录。所有的前端和后端都由我独立完成。
项目地址: <https://bitbucket.org/chineking/chineblog>

实习经历

- 2012.3–2012.6 英特尔亚太研发中心, 开源技术中心 (OTC), Intel Atom 处理器上的虚拟化, 包括 KVM 的修改和移植。
- 2013.7–2013.9 阿里巴巴中国有限公司, 无线搜索事业部离线系统, 参与了增量更新索引的数据库的设计和实现。