

# 薛鹏

## Python 开发工程师

### 教育经历

- 南京理工大学 · 本科 · 计算机科学与技术 · 2015年毕业

### 工作经历

#### 青云科技有限公司 | 大数据开发工程师

2021年12月 - 至今

- 负责中后台部门流批一体实时数仓技术方案选型与落地实践。
- 负责业务部门临时取数，数据业务报表开发工作。

#### 上海德拓信息技术有限公司 | Python 开发工程师

2018年11月 - 2021年11月

- 负责云计算产品产品重构以及分布式系统高可用架构设计。
- 负责基于 OpenStack 进行二次开发。

#### 宜高美信息服务成都有限公司 | C 开发工程师

2016年11月 - 2018年06月

- 负责富士通 5G 和 NB-IoT 新项目的开发
- 负责富士通 4G 项目的性能优化，提升了30%的运行性能。

#### 华为成都研究所 | C 开发工程师

2015年07月 - 2016年10月

- 负责 IAS 和 CAMC 项目功能的开发与验证工作。

### 项目经历

#### 流批一体实时数仓建设

2021年12月 - 至今

旧的离线数仓开发维护成本高，自助分析能力弱且数据实时性不够，负责实时数仓建设。通过积极与业务部门沟通，完成渠道转化趋势表，用户留存率等8个数据业务报表开发；通过调研行业最佳实践，完成实时数仓架构设计，包括计算引擎（Flink）、数据集成（Flink CDC）、数据存储（Hudi）、数据分析（Trino）、BI 工具（Metabase）技术选型，并使用 Flink SQL 在资源维表，用户维表进行落地。

#### 基于 OpenStack 进行二次开发

2020年02月 - 2021年11月

针对企业客户特殊需求，负责对 OpenStack 进行二次开发。积极与售后和客户沟通分析需求，完成了虚拟机强制关机、虚拟机可追溯性、系统盘查询、CDROM 管理功能等17个功能的开发。特别针对金融客户进行虚拟机性能调优，虚拟机高可用架构设计。通过 CPU 绑定、大页内存等技术将虚拟机性能提升30%。通过 Pacemaker 监测集群状态，在物理机宕机及脑裂时触发虚拟机迁移，将虚拟机可用性从90%提升至99.99%。

#### 云计算产品重构

2019年09月 - 2020年01月

线上问题频发，严重影响用户使用体验，负责解决产品已知问题、基础架构改造以及功能重构。梳理已知高优先级问题并制定项目计划协调研发进行修复，将 Bug 数从 363 降到 43；使用 Flask-SQLAlchemy、Flask-HTTPAuth、Flask-Caching、Flask-Mail、Celery 进行基础架构改造实现 RESTful API；梳理不合理功能 23 个，新增功能 36 个，完成集群管理、项目管理（Keystone）、网络管理（Neutron）功能重构，通过引入自动化测试、code review，将开发周期缩短了30%，新开发代码 bug 数从 80 多个降至个位数。

#### 构建高可用云计算产品系统

2019年06月 - 2019年08月

生产环境特定场景下不可用，负责系统高可用架构设计。通过 Pacemaker 实现服务的高可用，使用 MariaDB Galera 多主集群、RabbitMQ 镜像队列集群实现数据的高可用；定位并解决 MariaDB Galera 集群脑裂，Keepalived VIP 抢占和脑裂，RabbitMQ 主节点问题等核心问题，将服务的可用性从 90%提升至99.99%。