

▶ Halo Pan

北京市

电话: 15510004144

电子邮件: halo9pan@hotmail.com

在线简历: <http://cv.halo9pan.cn>

目标职位

架构师, 技术专家

学历

硕士 (2004/9 - 2007/7) 北京大学

软件与微电子学院 IT 项目管理

学士 (1999/9 - 2003/7) 华中科技大学

电气与电子工程学院 电气工程及其自动化

工作经历

ADVISORY SOFTWARE ENGINEER (2005/10 - 2013/10)

IBM 中国开发中心(CDL)

IBM Software Group, Application and Integration Middleware Software,
Information Management

系统架构师、研究员、技术委员会委员 (2013/10 - 2016/9)

爱奇艺

技术产品中心 - 后台开发部

数据系统架构师、机器学习技术专家 (2016/9 -)

网易

电子商务部

个人网站

<http://halo9pan.cn>

个人博客

<http://blog.halo9pan.cn>

GitHub

<https://github.com/Halo9Pan>

Kaggle

<https://www.kaggle.com/halo9pan>

开发技能

▶ 编程技能 80%

熟悉 JVM 内存模型、GC 策略等，阅读过 Java 虚拟机规范，能非常熟练的使用并了解运行机制。对多线程和 I/O 部分做过深入研究。Java EE 和 Spring、Hibernate 等框架有丰富的使用经验并阅读过部分源代码。能熟练使用常用的原生 JavaScript API，了解过浏览器运行机制。对流行前端框架停留在使用层面，少部分框架或者库阅读过代码。也有过后端 Node.JS 的使用经验。配合前端数据可视化，了解常见 HTML 标签语义，CSS 选择器、定位、盒模型、动画、位移等。熟悉 Linux 下 gcc 和 clang 的使用和编译过程，能写简单的 makefile、CMakeFile。也有过使用 OpenCV 和 ffmpeg 的经验。主要 Python 用来写系统配置脚本和作为数据分析工具，常见 API 能熟练使用。目前机器学习主力语言，熟悉相关三方库如 numpy、scikit-learn、pandas 等的 API，对性能调优有一定了解。

▶ 高性能计算 85%

熟悉 GPU 编程模型，常用接口以及优化策略。OpenCL 比 CUDA 略熟悉。发表过四篇相关专利。了解锁机制、函数式编程、Actor、GPU SIMD 等并发、并行编程模型。发表过一篇相关专利。

▶ 机器学习 75%

熟悉常见机器学习算法，实现过朴素贝叶斯、线性回归、随机森林、kNN，能使用协同过滤、关联规则、SVM 等算法。小规模使用过 Caffe，熟悉 TensorFlow 并阅读过部分代

码, 熟悉 TensorFlow 生态系统。使用过 Theano、PyTorch、Keras、Mahout、Spark ML。发表过两篇相关专利。

▶ **大数据 70%**

理解 HDFS、编写 MapReduce。配置 Hadoop 环境并调优参数, 阅读过部分代码。实现过 TeraSort。理解 RDD, 配置过 Spark 环境, 阅读过部分代码。实现过 TeraSort。常用 MLlib 模块, 使用过 Streaming 模块, 对 SQL 模块也有过接触。也有使用 HBase、Hive、Flume、Sqoop、Kylin、Druid 等的经验。

▶ **数据可视化 90%**

长期从事相关工作。拥有管理学背景, 能根据商业模型分析数据模型并决定展示重点。能以需求为基本出发点, 依靠自己的技术栈, 从头至尾实现数据可视化。发表过两篇相关专利。

▶ **Docker 80%**

有部署 Kubernetes 及 Swarm 的经验, 目前开发环境全部基于 Docker 或 Vagrant。熟练编写 Dockerfile。设计并实现了基于 Docker 的机器学习系统。

▶ **数据库 70%**

在 IBM 长期从事数据库相关工具的开发。MySQL、PostgreSQL、MongoDB 等配置及简单调优。熟悉数据库建模。

项目

▶ **网易: 大数据平台 (2016/9-2017/6)**

设计大数据平台架构、按照数据使用的层次, 划分为统计、分析、挖掘三个部分、根据每个层次使用的场景和特点针对性的提供不同粒度及时效性的数据服务。组织团队按照计划完成相关工作。

▶ **网易: 机器学习服务平台 (2017/4-)**

整理机器学习训练及预测流程, 设计并开发框架以便快速将机器学习模型部署上线并给业务方提供实时服务。

设计基于容器技术的机器学习平台, 快捷接入 Tensorflow 深度学习框架, 提供在线学习服务。

▶ **网易: 机器学习建模 (2017/6-)**

基于用户属性预测用户价值；基于用户行为数据创建机器学习模型，通过自编码的方法学习用户行为模型。用户流失预测，在线点击率预测。

▶ **爱奇艺：统一发布平台 (2014/10-2016/9)**

统一各个终端的发布流程以及模板制作的接口，统一数据接口服务发现。

▶ **爱奇艺：会员系统架构升级 (2015/05-2016/9)**

在会员业务出现瓶颈后，优化会员系统架构。数据库分表分库、降低事物粒度优化性能。通过分拆服务接口，通过微服务的方式提高业务灵活性。增强系统及业务层面的监控及报警。

▶ **爱奇艺：智能运营系统 (2014/07-2016/9)**

通过机器学习对资讯类视频进行自动化分类、生成推荐标签。简化编辑工作流程，提高生成效率。

▶ **爱奇艺：页面编译系统 (2014/03-2014/06)**

爱奇艺网站页面发布前需要编译成静态页面，原有系统性能低，资源利用不均衡。系统架构升级，采用多队列，提高性能和资源利用率。

▶ **爱奇艺：页面发布系统 (2013/10-2014/02)**

后台动态动态生成，页面内容需要调用多个内部接口，通过多线程并发调用提高性能。

▶ **OpenStack Horizon and Zabbix Integration (2014/02-2014/05)**

在 OpenStack Horizon 的 dashboard 里集成 Zabbix 物理机监控信息。

▶ **IBM：IBM Data Studio (2012/05-2013/10)**

[IBM Data Studio](#) 是一个基于 eclipse 的集成化数据库管理解决方案。让客户可以很方便的管理数据库，定制任务以及迁移数据。

▶ **IBM：IBM Smart Analytics System (2011/08-2012/04)**

[IBM Smart Analytics System](#) 是 IBM 的数据挖掘的整体解决方案，涉及 IBM 全球的多个软硬件部门。

▶ **IBM：Optim Query Tuner (2010/09 2011/08)**

一款数据库查询优化工具。通过对 DB2 查询语句的分析以及比较，对查询语句，索引以及统计提出相应的优化建议，使得查询语句有更高的性能，分别支持 [DB2 for z/OS](#) 和 [DB2 for Linux, Unix and Windows](#)

▶ **IBM：Optim Performance Manager (2008/03 2010/09)**

通过 Web 界面, 可以用来隔离并分析典型的数据库性能问题。还可以查看数据库运行状况的摘要, 并向下分析以获取更多详细信息。

▶ **IBM: Data Studio Administration Console (2006/04-2008/03)**

IBM Data Studio Administration Console 通过 Web 界面监控 DB2 数据库, 它提供了数据库健康和可用性监控模块, 可以获得系统的健康状况统计, 分析报警, 数据库健康信息仪表盘并且可以简单的分析问题所在。

专利

CN 201410415527 数据处理方法及装置

CN 201510541196 一种聚合页面中数据模块的方法及装置

CN 201510146258 一种基于图像的用户动作标识区域确定方法及装置

CN 201510149300 一种终端数据展现方法和系统

CN 201510629933 一种浮点数处理方法及装置

CN 201510629934 一种浮点数处理方法及装置

CN 201510727225 多视角视频采集系统

CN 201511020991 交互方法、装置和系统

CN 201511024848 并行运行 GPU 运算程序的实现方法及装置

CN 201511021303 视频标注方法和装置

CN 201511020414 一种基于字符编码转换的 GPU 内文本处理的方法及装置

CN 201511021485 一种基于图形处理器 GPU 的分词方法和装置