

陈啸天

18015528893 xtchen.pitt@gmail.com

北京朝阳区

后台开发

在职

教育背景

2014.9-2018.6 大连理工大学 工商管理 本科

技能评价

- ◇ 熟悉 Python 和 Go 语言，编程基础扎实，能够有效拆解复杂业务模块，合理划分架构层次
- ◇ 具备微服务化，可观测体系建设，容灾体系建设和研发效能提升相关项目的落地经验
- ◇ 熟悉内容的接入、处理和分发相关链路，有落地大型爬虫系统的经历，了解内容管线、内容池和推荐系统
- ◇ 对内容和数据敏感，擅长内容、作者相关数据的分析挖掘与策略应用

工作经历

2021.5-至今 腾讯视频 体育平台研发中心 高级后台开发

- ◇ 持续迭代升级赛事和内容相关的核心服务，包括图文直播、评论/弹幕、内容池、推荐接入层等
- ◇ 负责体育后台的微服务化改造，由单体 PHP 服务拆分为 API 网关、接口适配层和微服务层，大幅提升架构性能、可观测性和可维护性
- ◇ 从 0 构建体育的降级容灾能力，包括限流、降级缓存和全链路过载保护，提高资源利用率的同时保障核心页面和互动行为的高稳定性
- ◇ 负责提升团队研发效能，搭建全链路灰度、接口录制回放等基础能力，主导并落地了监控告警治理，全埋点数据上报，测试环境治理等项目

2019.4-2021.4 一点资讯 内容智能部 负责人

- ◇ 负责外站自媒体库，全网内容池等项目，负责爬虫整体架构和基础组件的设计
- ◇ 基于全网内容池挖掘内容、作者的相关特征，例如高热、地域、时效性、稀缺性等，建立了一套内容、作者的分级体系，同时也提供了相似度比对、原创识别等通用能力
- ◇ 分析站内内容供给与用户消费数据，结合实验优化内容供给策略，提升内容利用率和人均时长

2018.7-2019.4 百度视频 技术平台 1 部 数据研发

- ◇ 担任长短视频、PGC 视频和机器剪辑视频业务线的的数据爬取和收录工作
- ◇ 完成视频处理流程调度的重构，提高视频处理的透明度和稳定性
- ◇ 负责机器剪辑视频业务表字段设计、视频处理调度设计和相关功能开发

项目经历

2018.9-2018.12 机器剪辑视频 核心成员

- ◇ 独立负责机器剪辑项目中的爬取和视频处理部分，支持机器自动生产视频和辅助人工制作视频
- ◇ 设计视频信息存储和视频的处理流程，包括长视频精彩看点、横屏短视频剪竖屏、短视频合成等类型，接入视频处理平台，日均生产 1000 余条视频
- ◇ 负责热点事件剪辑平台项目的表设计、爬虫及后端编写，实时/离线爬取热点话题及其文字、图片和视频素材，实现辅助人工制作视频、文本转视频、自动生产热点话题视频

2020.4-2020.10 全网内容池 负责人

- ◇ 覆盖大内容平台、垂类 TOP、传统新闻门户在内的 30 多个站点，提供海量内容，每日更新量 5000 万+，主流站点自媒体作者覆盖率达 95%
- ◇ 基于 Airflow 分布式调度框架，定义出细分任务和组合链路，细分任务是最小粒度的爬取目标，组合链路将细分任务灵活地拼接，满足具体的爬取需求，增加可复用性。大幅提高开发效率和可维护性
- ◇ 针对反爬和风控，设计一系列反反爬服务和策略，如代理 IP 模块、Cookie 模块、验证码识别、手机群控和浏览器集群等，突破公众号、小红书、抖音等主流平台
- ◇ 设计了一套海量内容存储方案，包括去重、冷热分离和动态指标拉链等
- ◇ 爬虫系统 PaaS 平台化，针对不同场景的内容获取需求，提供个性化爬取链路配置、任务调度和内容投放能力

2020.11-2021.4	内容和作者特征挖掘	负责人
<ul style="list-style-type: none"> 支持站内外内容的点赞数、评论数等动态指标的分钟级监控。捕获竞品的 Push、热榜等信号 理解内容主体和分类，挖掘内容的时效、原创、地域、稀缺度等属性，分析评论情感，多维度打分，建立全面的内容分级体系 聚合作者近期的内容特征和表现数据，分析作者成长趋势，建设作者画像。挖掘出各领域的潜力作者，为自媒体平台提供价值线索。并提供创作者覆盖度、站内外表现对比等基础能力 基于内容的站内外表现和基础特征，沉淀冷启池、高热池。通过长期实验，优化平台的内容供给策略，并在分发环节提供了热度信号，大幅提升内容分发效率。rctr 提升 18.4%，人均时长增加 148s 		
2021.8-2022.6	体育接入层升级	负责人
<p>背景：体育后台最初是 PHP 单体应用，集中在一个高度复杂的大项目中。随着业务发展和技术架构的演进，单体应用逐渐演变为接入层的角色，存在框架老，代码乱，性能差，运营难等问题</p> <ul style="list-style-type: none"> 项目一期，主导体育 PHP 接入层改造方案的设计和评审，将架构分层为 API 网关，接口适配层和领域层。以体育核心接口比赛/内容底层页、推荐信息流的重构为标杆案例，沉淀出通用的代码框架和组件，并提供了从代码设计，到正确性保证，再到灰度上线的全流程指引 项目二期，以网关为起点建立了全链路可观测体系。设计并落地了一整套限流，降级和全链路过载保护的服务容灾方案 <p>收益：在该方案的指引下，体育在过去的一年中重构了 106 个接口，覆盖体育 98% 的流量，核心接口 QPS 提升 1 倍+，响应时间降低 57%，实现接口告警 3min 内触发，问题定位 10min 内完成，可用性提升至 99.99%</p>		
2021.8-至今	TAPISIX 网关	PMC 成员
<p>背景：TAPISIX 是司内基于 APISIX 网关定制开发的开源协同项目。担任 PMC，负责功能开发，CR 和发展方向规划。在司内多个业务落地，多次获得公司级别奖项</p> <ul style="list-style-type: none"> 适配公司特殊协议，支持多种协议转换 适配公司内部服务发现平台，支持动态路由能力，包括规则路由、就近路由和金丝雀路由等 完备网关的降级容灾能力，实现了访问限流、故障熔断、兜底缓存、请求优先级控制等特性 在体育业务落地 TAPISIX 网关，支撑日均 10 亿+流量，峰值 QPS 30w+，并以网关为起点建立了全链路可观测体系和容灾体系 		
2022.6-2022.12	测试环境治理	负责人
<p>背景：体育后台测试环境一直以来存在诸多痛点，严重影响了需求联调，走查和测试的效率</p> <ul style="list-style-type: none"> 结合网关、微服务框架、服务发现和容器平台，完成多环境泳道的方案设计与实现。支持随需求生命周期自动化管理测试环境，各特性环境可稳定运行并且流量隔离，在客户端侧可选择需求快速切换特性环境 抽象体育数据资产为比赛、内容、用户互动。根据不同类型数据产生的特点，采用不同的自动化生成手段，如通过比赛数据回放模拟比赛的全流程，配置内容接入白名单同步内容至测试环境，回放正式环境互动行为接口至测试环境等，保证了测试环境数据的完备性 <p>收益：需求已全部接入，新迭代中因环境阻塞测试，出现 bug 的情况降低至 0；测试数据的构造成本从 1-2 天降低至分钟级别</p>		
个人评价		
<ul style="list-style-type: none"> 善于通过沟通进行需求拆解，结果导向解决工作中各类难题 拥有良好的自我驱动力，乐于钻研前沿技术，不断强化自身软硬技能 		