

李福春	求职意向: 明途知产 · AI应用架构师 (主管级) 15年一线研发 · 7年架构团队管理 · 腾讯云架构师认证架构师 / 名人堂专家			
	出生年月 1988年01月 性别 男 籍贯 湖南省常宁市 期望月薪 30-50K 毕业学校 北华大学 · 软件工程 · 统招本科 (2011.07) 英语 大学英语四级 515分 电话/微信 13682320515 邮箱 505847426@qq.com 地址 深圳市宝安区西乡街道安居鸿栖台			
岗位要求匹配优势	1. 方案一生产架构: 与产品/业务共创, 把 AI 落地方案转为可生产技术架构与实施路径; 覆盖选型 (GenAI/LLM vs 规则检索 vs 混合)、架构设计、PoC、落地、成本与风险控制, 并输出文档/架构图/trade-off。 2. AI 能力边界判断: 明确何时用 RAG / Agent, 何时优先规则引擎/检索; 方案兼顾准确率、可解释性、延迟、合规与成本——拒绝「凡事上大模型」。 3. 企业级 GenAI 落地: 主导 AgentX / DigitalHuman / HeroDash, 覆盖智能体、知识库、复杂工作流从概念到生产; 向量库 Milvus + 微服务 + 监控告警。 4. 生产难题与 Eval: 攻坚幻觉、响应速度、Token成本、长文本与 Agentic Workflow; 以效果评测/案例闭环把不确定模型输出变成可上线服务标准。 5. 带队与技术路径决策: 指导/带领工程师从原型到生产; 带过 15人交付团队; 金融 KA 私有化合规经验; 可独立拍板技术路径并跨部门协同。			
技能要求匹配优势	1. GenAI 技术栈: 精通 LLM 应用开发、RAG、Agent 框架与向量库 (Milvus, 对齐 Pinecone/Weaviate/Chroma 能力); 熟悉 GPT / Claude / DeepSeek / Llama 等主流模型优劣与生产排障。 2. 编排与工具调用: 自研 DAG Workflow 对齐 LangChain / LangGraph / LlamaIndex / Dify; 深度实践 MCP / Tool Calling; 支撑复杂任务编排与系统集成。 3. 工程底座与 MLOps 相邻: 精通 Java + Python; Docker/K8s/CI-CD 与后端系统集成; 熟悉微服务、网关、MQ 与可观测 (软对齐 MLOps 部署交付, 不编造训练平台专项)。 4. 效果 / 成本 / 合规: 持续做准确率-延迟-Token 成本权衡; 权限/审计/私有化隔离; 金融/保险合规交付; 对齐 Eval 与上线门禁思路。 5. 商业与决策: 从投入产出比做技术选型; 拥抱变化、能把前沿能力快速判断为可工程价值; 独立决策技术路径并推动落地。			
工作经历	2025.07 – 2026.07 微软小冰 (AI 数字人 / Agent) Java 高级 / KA 交付负责人			
	1. 以 AI应用架构角色把产品方案转为生产架构: 选型 LLM/RAG/Agent vs 规则检索混合路径; 从 0 到 1 交付 AgentX (Runtime + Agentic Workflow + MCP), 对齐 LangChain / LangGraph。 2. 主导 DigitalHuman ASR/TTS/LLM/RAG 全链路; 攻坚幻觉/延迟/Token成本, 端到端 2s→500ms, 可用性 99.9%; 金融 KA 私有化合规交付。 3. 落地 Docker/K8s + 双网关隔离 + 监控告警; 指导工程师闭环上线; 持续做效果评测与成本权衡 (对齐 Eval / ROI)。			
	2024.11 – 2025.06 纵腾集团 (跨境物流 / 末端派送) 后端 Leader			
	1. 负责技术方案评审、Code Review 与研发规范; 主导微服务拆分与跨模块协同交付。 2. 支撑美/法/荷/意/摩尔多瓦五国部署, 日运单 10万→100万、站点 100→500、快递员 1000→5000。 3. 落地 Flink + StarRocks 流批一体数仓与多通道消息中心, 日均千万级推送、触达率 99.99%。			
	2023.05 – 2024.11 深圳客诺信息技术有限公司 (全球客服 SaaS / CRM / AI) 架构师			
	1. 主导 HeroDash 企业级 AI 应用: RAG 知识库、智能客服/智能问答与复杂对话工作流; 向量检索 (Milvus) + 效果调优, 支撑 600+ 并发与 500万+ 用户。 2. 整合 Llama3 / Whisper / Azure / Google 等多模型; 沉淀 Prompt Engineering 与检索链路, 提升准确率/可控性, 控制调用成本。 3. 基于 SpringCloud + K8s + GitOps 完成可扩展高可用架构; AI 与业务系统 API 深度集成, 符合 GDPR/CCPA 合规基线。			
	2021.11 – 2023.04 深圳市禅游科技股份 (游戏发行运营) Java 架构师			
	1. 从 0 到 1 打造低代码「魔方」(表单/流程/BI), 培训 10+ 研发小组, 开发周期缩短 70%。 2. 重构发行平台高并发链路 (用户 14亿、峰值 3万 QPS), 落地 WebFlux + Dubbo3 与容器化 CI/CD。 3. 全链路监控告警 (ELK / Prometheus / Zabbix), 故障秒级发现; 推动资源成本下降约 30%。			
	2020.04 – 2021.11 深圳奥哲网络 (低代码 PaaS) 架构师			
	1. 主导《云枢》从 0 到 1 云原生架构 (Docker+K8s+Istio+SpringCloud), 支撑高可用与规模化扩展。 2. 从 0 到 1 组建建行北京 15人交付团队, 完成 3 次里程碑并推动分行规模化落地。 3. 研发周期缩短约 40%, 自动化部署比例提升至 90%, 客户满意度达 99%。			
	2018.09 – 2020.04 深圳乐心医疗 (可穿戴医疗物联网) 资深工程师			
	1. 主导乐心健康云 / 医生平台 / 保险健康模块架构, 覆盖穿戴设备数据接入与多租户云服务。 2. Dubbo + CAT + Zebra 升级后日请求破 亿级、可用性 99.99%, 医生日活突破 30万。 3. 牵头天安保险健康模块落地, 企业客户从 0 拓展至 300+。			
	2015.01 – 2018.09 深圳前海映像 (跨境电商供应链) Java 高级			
	1. 基于 Dubbo 从 0 到 1 搭建跨境供应链微服务 (用户/订单/仓储/物流/财务等), 支撑日均百万级订单处理能力建设。 2. 全链路性能调优, 订单查询从 2.3s 降至 0.7s, 综合吞吐与稳定性显著提升。 3. 推动后端 Code Review 与技术分享, 团队协作效率提升约 40%。			
	2011.07 – 2015.01 深圳第七大道 (网页游戏) Java 工程师			
	1. 负责页游平台用户/充值/客服/运营推广等核心模块开发与调优。 2. 支撑峰值并发约 2000、积累用户约 2亿, 服务年营收超 2 亿的运营体系。 3. 完成《神曲》《弹弹堂》官网与公司官网 CMS/SEO 建设, 保障活动投放与品牌转化。			
代表项目	AgentX 智能体引擎 AI应用架构 / Agentic Workflow 2025.07 – 2026.07 背景与功能 : 企业级 AI Agent 生产平台。以 Agent Runtime + Agentic Workflow + MCP Tool Calling + 多模型路由, 把产品侧 AI 方案工程化为可扩展、可观测、可控成本的服务; 支撑播报、交互、营销、 智能问答 、问数据等业务智能化接入。			

AgentX 智能体引擎架构

接入层

API Gateway 双路隔离

Sa-Token 鉴权

AgentX API

Workflow API

拖拽设计器

编排层

Agent推理引擎(Java)

DAG工作流(Python)

多轮会话/记忆

任务规划/工具选择

能力层

多模型LLM

MCP工具适配

RAG/ES检索

业务工具

中间件层

MySQL

Redis

RabbitMQ

Elasticsearch

OSS/MinIO

基础设施

Docker/K8s

CI/CD

Prometheus/Grafana

核心技术点:

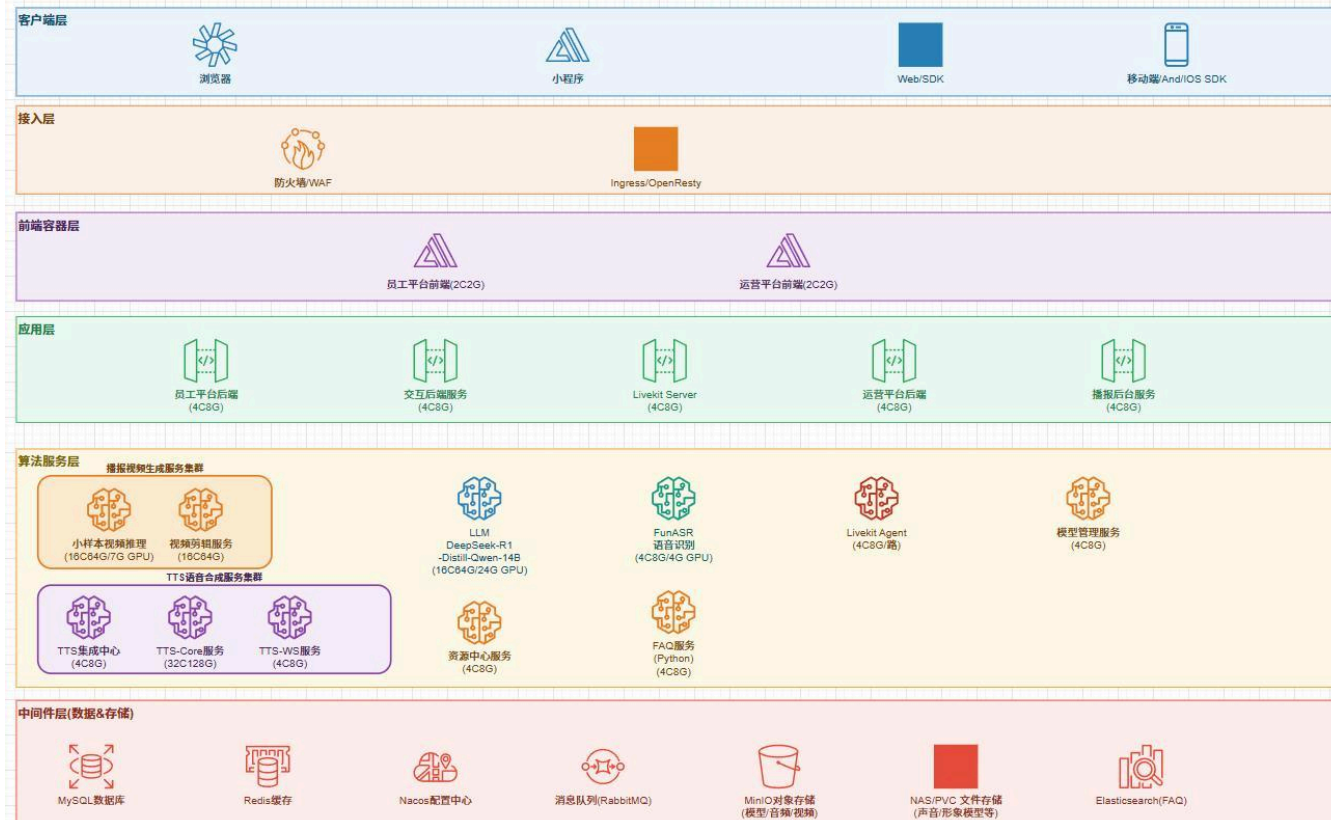
- 基于 Java 构建 **Agent Runtime** / 推理编排引擎, 集成多模型 **LLM**, 经 **MCP** 完成 **Tool Calling** 标准化接入 (对齐 **LangChain / LangGraph**) 。
- Python DAG Workflow** + 拖拽设计器, 支持多步任务、节点调度、状态管理与 RabbitMQ 异步执行。
- Docker/K8s** 工程化 + 双 API 网关隔离; 鉴权/重试/审计; 持续做延迟/成本/稳定性迭代, 保障生产可用。

技术栈: SpringBoot + FastAPI + MyBatisPlus + SaToken + Redis + RabbitMQ + ES + OSS + Celery + Docker + K8s + Prometheus + LLM/MCP

数字员工平台 DigitalHuman | LLM 应用 / RAG / 私有化 | 2025.07 – 2026.07

背景与功能: 面向企业的数字人生产与运营平台。覆盖声音/视频/交互数字人入驻、播报音视频创作、LiveKit 实时交互、**RAG** 知识库问答、TTS/ASR、**智能问答**与智能交互; 支撑银行/保险等 KA 私有化与信创交付 (权限隔离与审计) 。

数字员工平台架构图



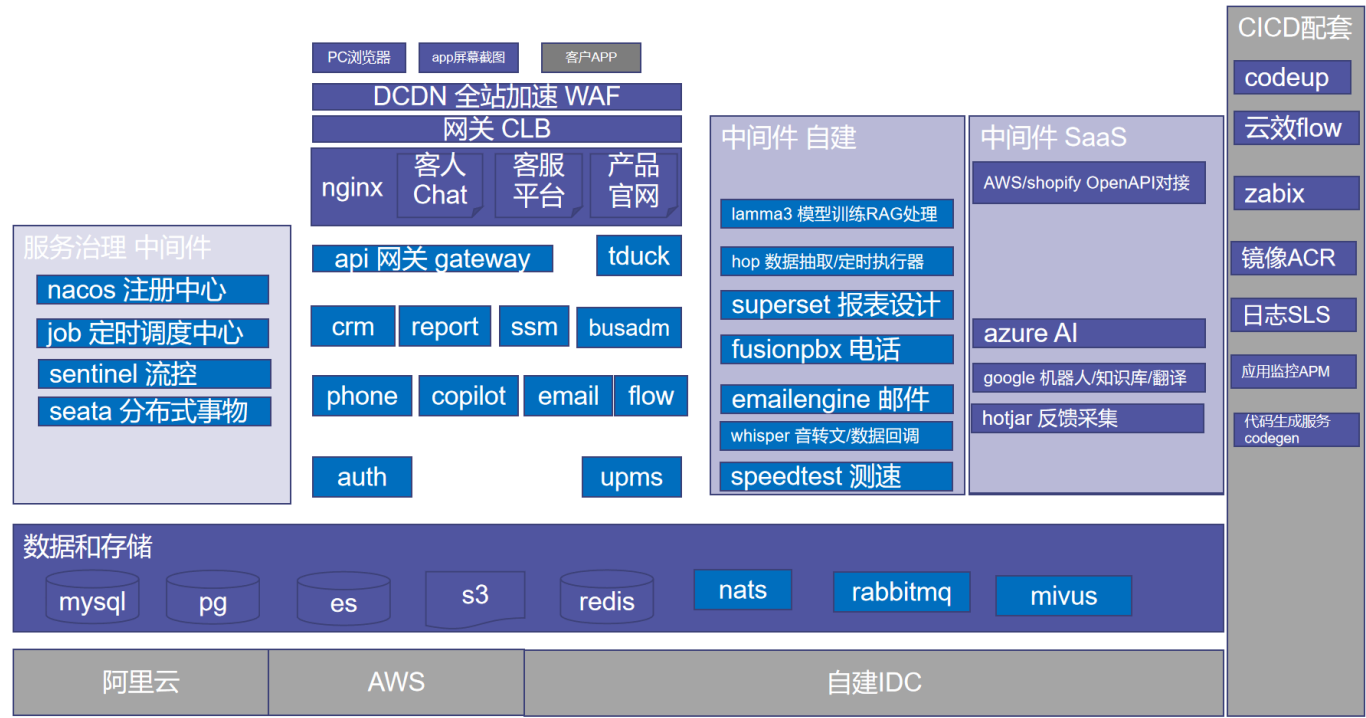
核心技术点:

- 打通 ASR/TTS/**LLM/RAG**/LiveKit 实时交互全链路, 支撑万级用户场景下可用性 **99.9%**。
- 多引擎 TTS 聚合与 SSE 流式调度, 端到端延迟从 2s 优化至 **500ms** (降 75%) ; 配合观测指标持续压延迟与成本。
- 主导私有化标准版与信创适配 (兴业/国泰/瑞众) ; 权限隔离 + 审计闭环, 贴近企业 AI 安全治理。

技术栈: SpringBoot + MySQL + Redis + RabbitMQ + ES + Nacos + MiniO + Docker + K8s + LLM/ASR/TTS/LiveKit

全球 AI 智能客服平台 HeroDash | RAG / 智能客服 | 2023.05 – 2024.11

背景与功能: 混合云全球客服 SaaS (work.herodash.ai) 。整合邮件/电话/聊天/社媒多渠道, 工单智能调度、**知识库**自助、**智能客服**与 AI 质检; 打通信息孤岛, 符合 GDPR/CCPA。



核心技术点:

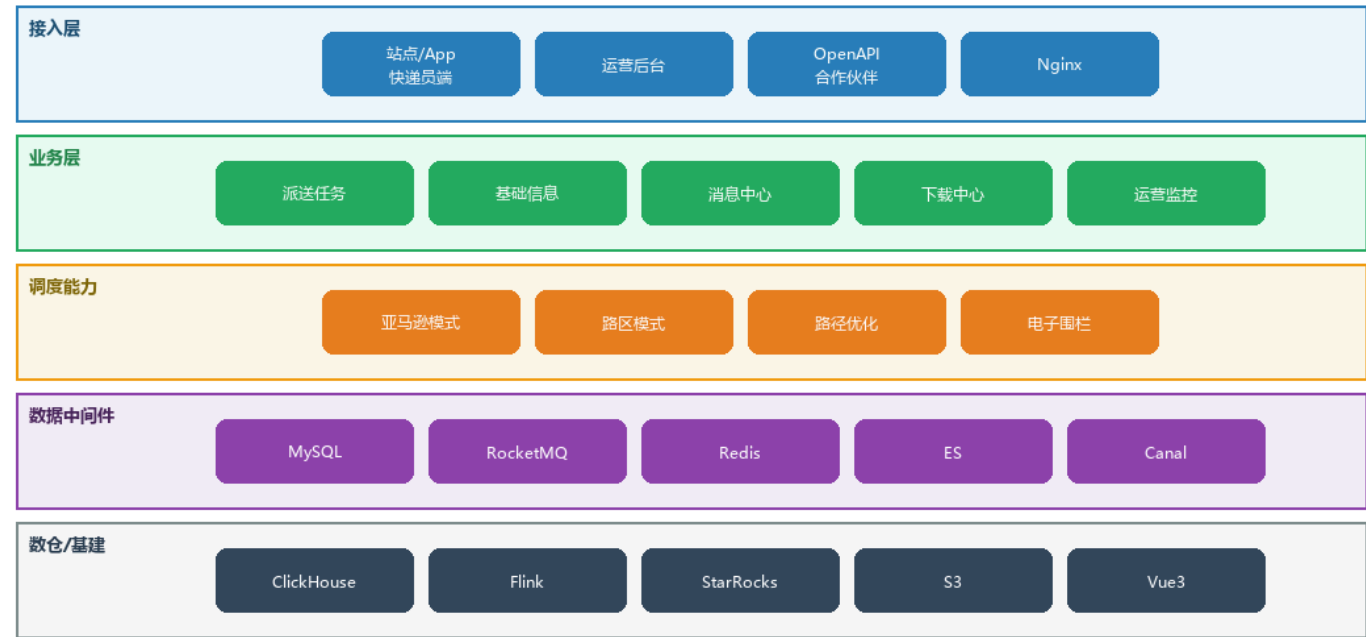
- 1. SpringCloud 微服务 + 脚手架提效, 迭代周期缩短约 **50%**; 支撑 **600+** 并发、**500万+** 用户。
- 2. **RAG 智能问答/智能客服**与语音质检落地; 向量检索 (**Milvus**) + **Prompt** 效果调优; FusionPBX 电话集群化, 稳定性 **99.99%**。
- 3. 整合 Llama3/Whisper/Azure/Google AI; 持续做检索与评测闭环 (准确率/成本/延迟), 沉淀数据/案例分析能力。

技术栈: SpringCloud + Nacos + Sentinel + Seata + MySQL + PostgreSQL + Redis + ES + RabbitMQ + NATS + FusionPBX + Milvus + K8s

末端配送物流系统 GFS | 后端 Leader | 2024.11 – 2025.06

背景与功能: 纵腾美国本土末端配送与揽收核心系统 (gofoexpress.com)。覆盖订单接收、运单生成、中心操作、派件调度、路径优化、履约签收、问题件处理与实时监控, 支持亚马逊模式与灵活路区模式。

末端配送物流系统 GFS 架构



核心技术点:

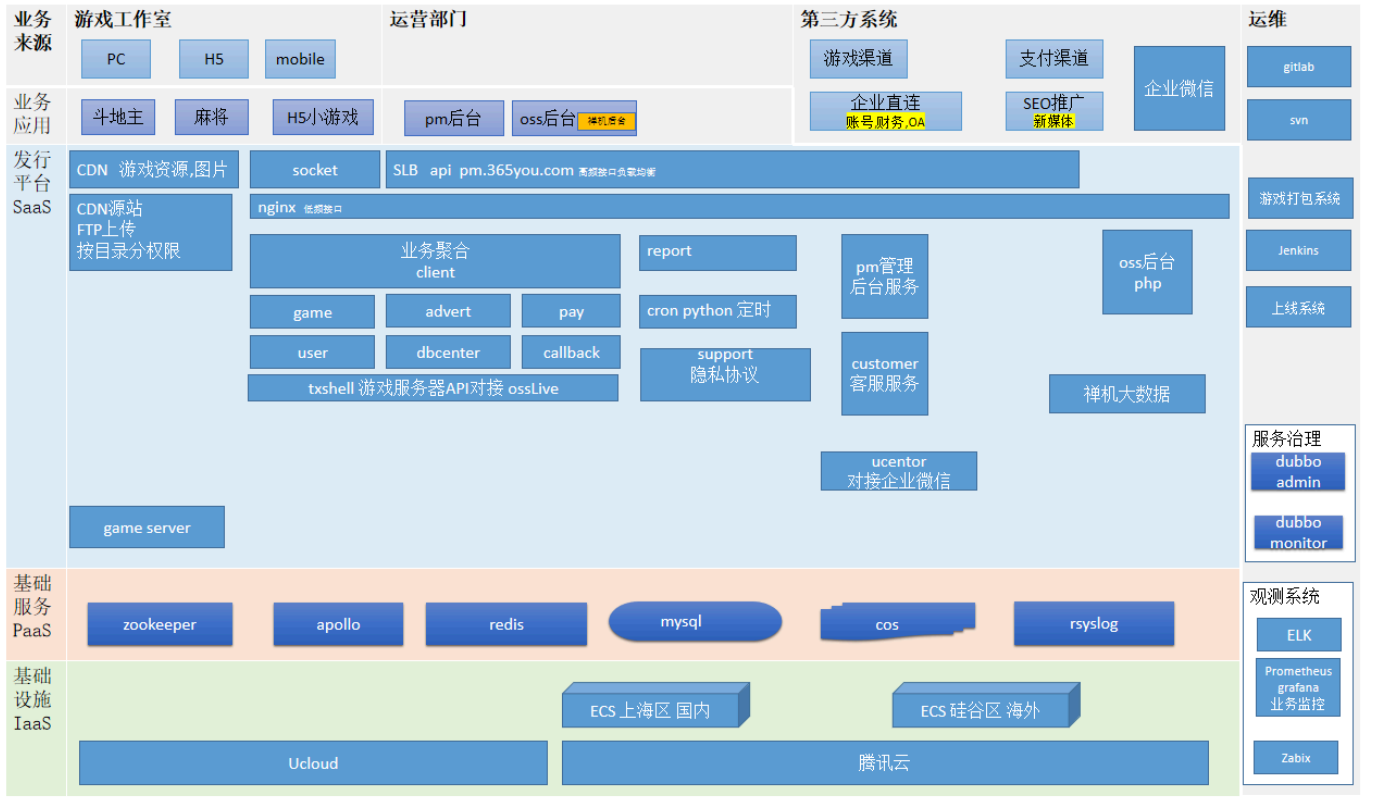
- 1. 派送任务多模式落地 + 微服务拆分, 支撑日运单 **10万→100万**、站点 100→500、快递员 1000→5000。
- 2. Canal 解耦 + ClickHouse 指标查询, 演进至 Flink+StarRocks 流批一体数仓, 实现分钟级运营监控。
- 3. 多通道消息中心 (短信/邮件/站内信/WhatsApp 等) 日均千万级推送, 触达率 **99.99%**; 协同解决积压/慢 SQL 等产线问题。

技术栈: Java8 + SpringBoot + MyBatisPlus + MySQL + RocketMQ + Redis + ES + Canal + ClickHouse + Flink + StarRocks + Vue3

禅游发行平台 | 架构师 | 2021.12 – 2023.04

背景与功能: 禅游游戏体系核心现金流系统。对接 OVH/小程序/苹果等联运渠道的登录、支付、进游戏与广告投放, 并提供运营/商务/工作室统一管理门户。

02 金箍棒部署架构 V2



核心技术点:

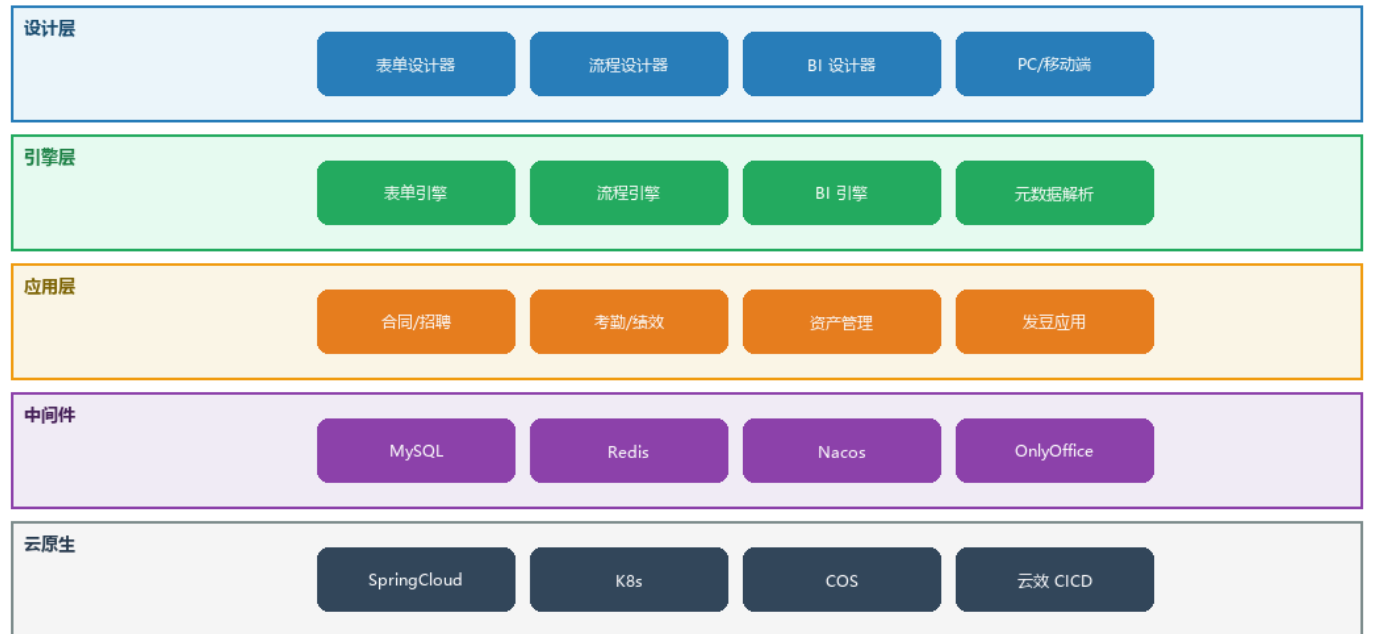
- 支撑用户约 **14亿**、日活约 **900万**、峰值秒并发约 **3万**; 腾讯云容器化 + GitLab CI/CD, 弹性伸缩与分钟级发布。
- WebFlux 网关异步化与熔断集群化, 服务调用隔离容灾, 稳定性显著提升。
- 全链路监控告警 (ELK / Prometheus / Zabbix), 故障秒级发现、分钟级处置。

技术栈: Dubbox + WebFlux + MySQL + Redis + Zookeeper + Apollo + Nginx + 腾讯云

低代码平台魔方 | 架构师 | 2021.12 – 2022.10

背景与功能: 自研面向业务与研发的专业低代码平台。提供可视化表单、流程、BI 设计 (PC/移动), 服务内部应用敏捷交付与研发降本增效。

低代码平台魔方架构



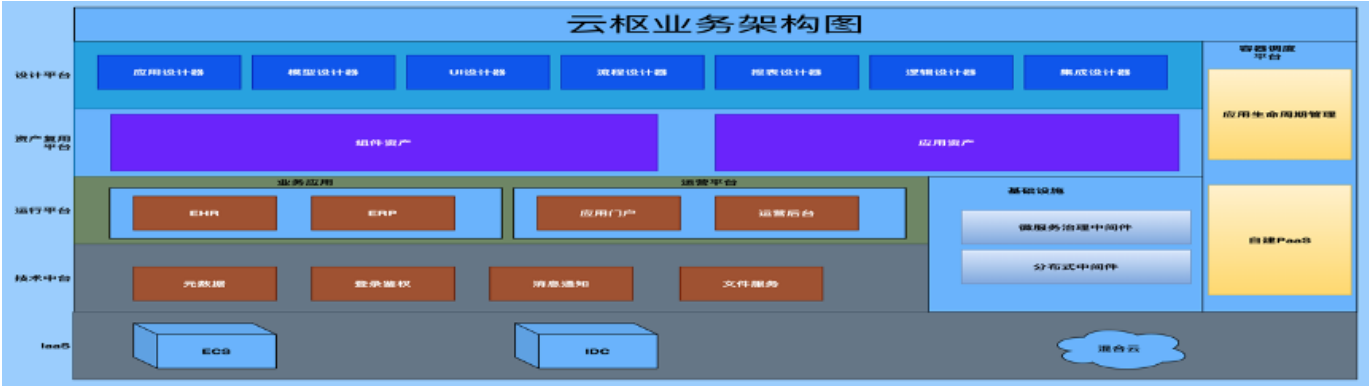
核心技术点:

- 从 0 到 1 完成蓝图、表单/流程/BI 引擎设计与核心编码, 交付合同/招聘/考勤等应用。
- 容器化 CICD 与自动化测试落地, 单周可交付关键应用, 开发周期缩短约 **70%**。
- 建立跨团队低代码标准, 培训 10+ 研发小组, 年均节省开发工时显著。

技术栈: DDD + SpringBoot + SpringCloud + K8s + MySQL + Redis + Nacos + OnlyOffice

奥哲云枢 | 架构师 | 2020.04 – 2021.12

背景与功能: 领域模型驱动的云原生低代码产品。含设计/运行/运营/运维四平台 + 资产/集成两大市场; 核心为模型/UI/流程/集成/逻辑/报表六大设计器, 元数据驱动快速搭应用。



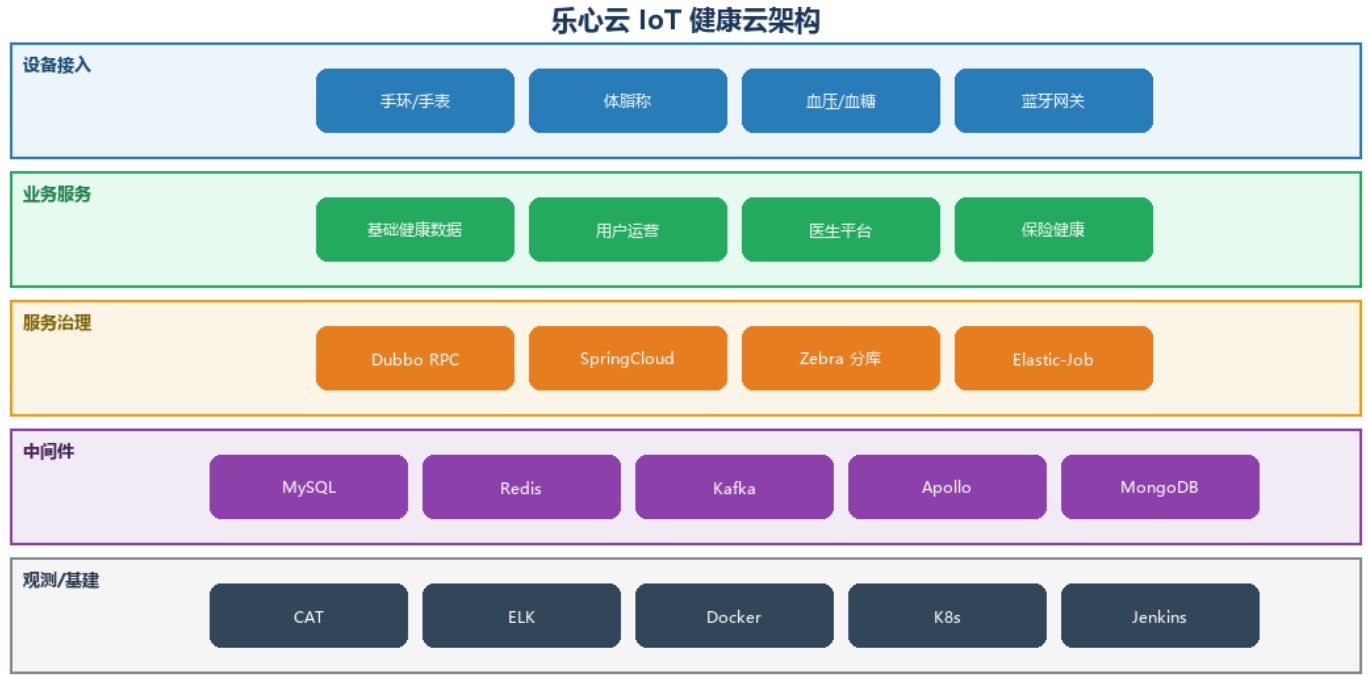
核心技术点:

- 1. 从 0 到 1 架构与云原生全链路落地 (Docker+K8s+Istio+SpringCloud), 支撑高可用与规模化扩展。
- 2. 封装核心 SDK, 输出规范与售前托底方案; 元数据扩展点与非功能 (吞吐/可用/扩展) 建设。
- 3. 从 0 到 1 组建建行北京 15 人交付团队, 完成 3 次里程碑并推动分行规模化落地。

技术栈: DDD + COLA + SpringBoot + SpringCloud + K8s + Docker + Istio + Vue

乐心云 | 资深工程师 | 2018.09 – 2020.12 | IoT

背景与功能: 乐心健康可穿戴设备数据接入网关与多租户微服务云 (手环/心电图表/体脂称/血压计/血糖仪)。覆盖基础健康数据、用户运营与医疗/保险行业方案, 日访问量过亿 (服务入口)。



核心技术点:

- 1. Dubbo/SpringCloud 服务治理 + Zebra 分库分表, 日请求破 **亿级**、可用性 **99.99%**。
- 2. 医生平台与天安保险健康模块落地, 医生日活 **30万+**, 企业客户拓展至 300+。
- 3. Docker/K8s + Jenkins 持续交付, CAT/ELK 全链路观测。

技术栈: Spring + Dubbo + Jersey + Kafka + Redis + MyBatis + Zebra + Apollo + Elastic-Job + Docker + K8s + CAT + ELK

跨境电商供应链系统 | Java高级 | 2015.01 – 2018.10

背景与功能: 从 0 到 1 建设跨境供应链: Facebook 单品站获单后经 MQ 推送, 完成订单审核、采购、仓储、物流与财务全链路; 前后端分离, 后端 Dubbo 微服务解耦八大业务域。

目标公司	跨境电商供应链架构 渠道接入 Facebook 单品站 MQ 推送 LayUI 前端 业务域 用户/产品 订单审核 采购 仓储 WMS 履约财务 物流 财务 客服 库存同步 中间件 Dubbo MySQL Redis RabbitMQ 运维保障 Nginx Tomcat 监控告警 CI/CD 核心技术点： 1. 八大业务域解耦与标准化框架，支撑百万级订单处理能力。 2. JVM/缓存/慢 SQL 调优，核心查询 2.3s→0.7s。 3. 引入 CI/CD、监控告警与 Code Review，发布周期缩短、故障率下降。 技术栈：Dubbo + MySQL + Redis + RabbitMQ + Nginx + Tomcat + LayUI
	第七大道网页游戏平台 Java工程师 2011.07 – 2014.12 背景与功能：页游平台高可用后台（7road.com）。提供登录注册、充值、客服、运营投放接口，并内联微信/微博/短信/邮件等第三方模块。 第七大道页游平台架构 接入层 官网/专题 游戏客户端 运营投放 Nginx LB 业务接口 登录注册 充值 客服 进游戏 运营能力 活动/广告 CMS/SEO 排行榜 第三方账号 数据层 MySQL Redis dbutils 部署 Tomcat 集群 GuiceServlet 可演进 Dubbo 核心技术点： 1. 支撑峰值并发约 2000、积累用户约 2亿 的运营体系。 2. 保障《神曲》《弹弹堂》等精品运营活动稳定投放。 3. 完成公司官网与 CMS/SEO，支撑招聘与品牌转化。 技术栈：GuiceServlet + MySQL + Redis + Nginx + Tomcat（可演进 Dubbo）
	择业标准 1. 具备持续的盈利能力 2. 技术团队的氛围好 3. 与时俱进的更新技术栈 技术布道 / 自媒体  腾讯云主页  今日头条  公众号  CSDN  Github