



2025 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站





2025 AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

08/08-09 | 北京站

11/14-15 | 深圳站



查看会议详情

北京站论坛设置

大模型和 AI 应用评测

智能存储与检索技术

下一代知识工程

AI+ 金融业务创新

智能需求工程

智能体与研发效率工具

AI 产品运营与出海策略

大模型安全与对齐

大模型应用开发框架与实践

智能体经济 (Agentic Economy)

智能测试工具的开发与应用

具身智能与机器人

代码生成及其改进

AI+ 新能源汽车

AI 前沿技术探索与实践

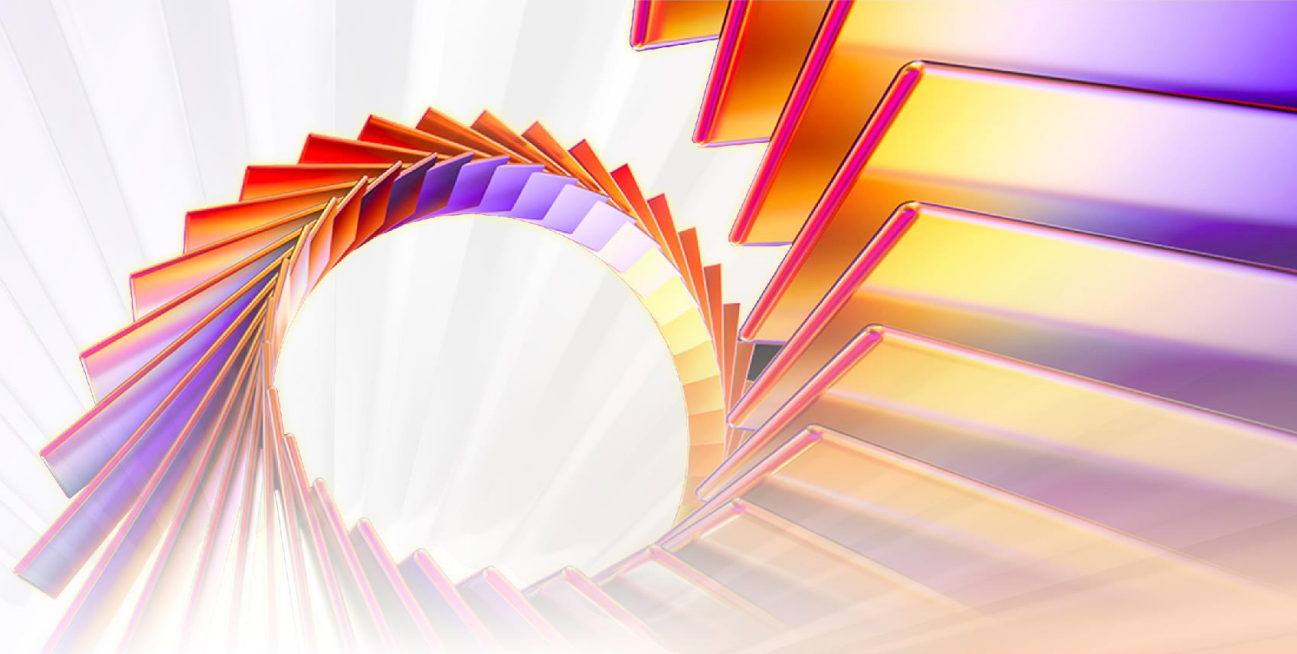


| 05/23-24 | 上海站

2025 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



云函数 SCF 解码车企智能出行 三大最优解

王瑞明 | 腾讯云



王瑞明

腾讯云专家工程师

现任腾讯云Serverless云函数专家工程师，主导函数计算核心引擎的架构演进与技术攻坚，完成从技术架构设计到商业化落地的全链路实施。深耕云计算领域十余年，专注于分布式系统与弹性算力模型创新，主导的冷启动优化方案推动行业延迟指标突破百毫秒量级，创新设计毫秒级弹性资源调度体系推动资源利用率突破70%水位线。当前技术聚焦于云函数在多级智能调度、轻量化、云原生结合、AIInfra的创新落地等方面。

目录 CONTENTS

I. 智能出行三重挑战

—出行行业高速发展，智能辅助驾驶、车联网、智舱场景的成本、可用性、效率三重挑战

II. 云函数SCF介绍

—新一代的云原生算力平台，无需运维，极致弹性，按量计费

III. 解决方案

—云函数SCF在智能辅助驾驶、车联网、智舱三大应用场景的解决方案

- a. 智能辅助驾驶：百Gb吞吐数据处理，算力错峰复用实现最优成本
- b. 联网：百万车辆并发连接，毫秒级数据实时处理
- c. 智舱：跨部门拉通协作效率，微服务开发模式轻松提高

IV. 总结与展望

—云函数SCF关键应用能力升级，为出行行业筑牢底座

PART 1

智能出行三重挑战

出行行业高速发展，智能辅助驾驶、车联网、智舱场景的成本、可用性、效率

▶ 出行行业高速发展带来智能辅助驾驶、车联网、智舱场景 成本、可用性、效率三重挑战

2024年国内新增新能源车1200万辆，同比增速40%。随着新能源汽车电气化升级联网，应政策要求需要接入新能源汽车国家监测体系，出行行业有着海量数据处理需求，业务量级巨大，核心在以下三大场景：

1. 智能辅助驾驶场景

采集地图信息、摄像头信息上传到 cos，进行 T+1 处理（解压、解析、脱敏、打包，投递相关信息到数据库、大数据等下游）



海量数据，超大量的GPU和CPU 算力需求，如何降低成本？

 低成本

2. 车联网场景

车的行驶数据、状态数据实时上报到 MQTT，一部分进行准实时处理（故障分析等），一部分进行T+1处理（轨迹查询等）



百万车机设备接入，超高并发连接，数据处理实时性要求高，当前架构能否支撑？

 高可用

3. 智能座舱场景

采集用户基础信息上传到推荐平台，用以生成用户画像，进行准实时数据处理，最终生成与用户高度相关的个性化推荐信息



一个功能跨车机、用户主数据、大数据、算法多个部门团队，拉通协作效率低，OTA 按年计？

 高效率



PART 2

云函数SCF简介

新一代的云原生算力平台，无需运维，极致弹性，按量计费

Serverless是什么？新一代云原生算力平台

按实际用量收费，最大限度节约机器资源和运维成本

类似出行方案

付费方式

特点

服务器

运维成本

网约车

按量付费

不用不花钱

Serverless

免运维

租车

租约期付费

不用也要收钱

虚拟机

运维软件

私家车

一次性付费

资源独占

物理机

运维硬件

云函数SCF是什么？新一代云原生算力平台

按实际用量收费，最大限度节约机器资源和运维成本

 - 上传业务代码/镜像，声明业务运行的触发规则

开发者

- ★ **无改造**：支持上传容器镜像，兼容传统Web 框架
- ★ **更易用**：声明业务执行条件即可自动运行

- 实时感知事件源，自动触发函数运行

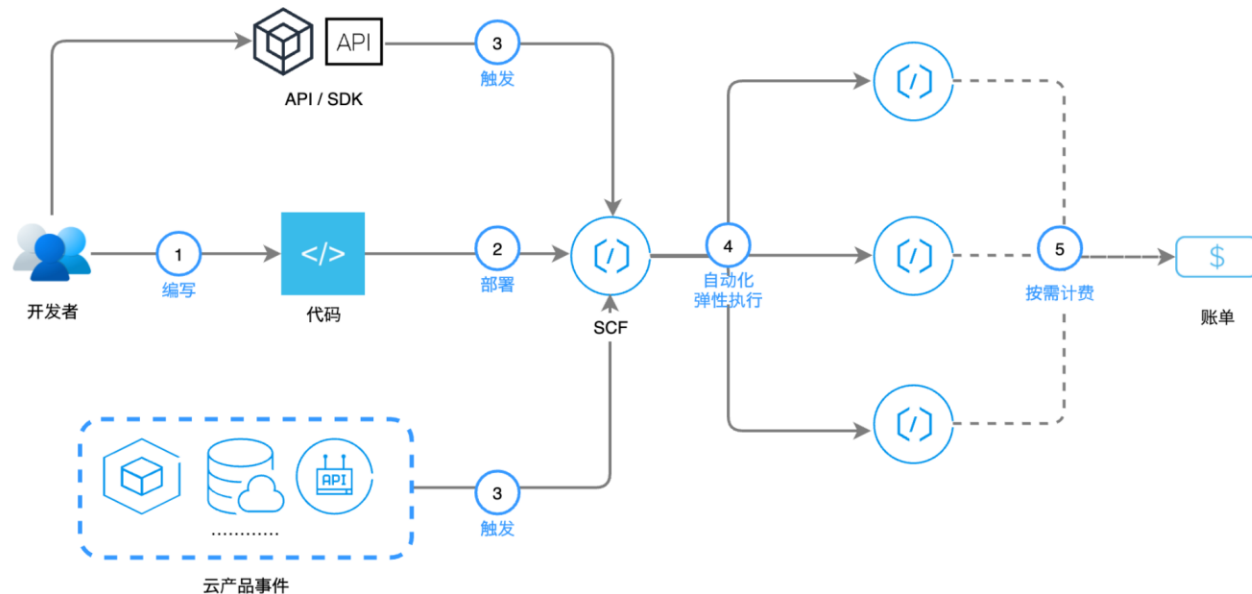
SCF

- ★ **连接万物**：丰富的触发器类型，各种场景、各类事件源、自定义触发规则（定时/Kafka/数据库/HTTP URL/Cos/Cls/CLB/ ...）
- ★ **实时消费**：平台组件常驻实时感知事件源行为，满足触发规则 函数自动执行

 - 基于实际请求动态拉起实例，运行业务逻辑

SCF

- ★ **极致弹性**：基于实际请求数量自动拉起对应实例执行，毫秒级极速动态扩缩容，拉起实例速度可达10000个/分钟
- ★ **超高可用**：自动跨Zone容灾，多集群部署容灾机制
- ★ **安全隔离**：自带租户级强隔离，保障业务安全运行
- ★ **实例复用**：连续请求时底层复用实例，无需冷启动
- ★ **按量付费**：按量毫秒级精确计费，只为代码实际运行时长付费，无资源常驻成本



聚焦核心业务逻辑

无需运维底层资源

不为闲置资源付费

极致弹性扩缩容

自带容灾机制

存量业务 无需改造

- ✧ 兼容传统Web框架，无缝适配原生Web Servers
- ✧ 支持Job镜像，无需对已有镜像进行任何改造
- ✧ 内置Runtime，开发代码只需聚焦核心业务逻辑
- ✧ 函数插件模式，兼容原Sidecar容器使用模式

灰度发布 管理体系

- ✧ 灰度发布，完备的别名+函数版本管理体系
- ✧ 流量控制，支持按照百分比或者请求参数路由

CI/CD 全流程工具

- ✧ 标准CLI命令行
- ✧ 声明式YAML蓝本
- ✧ IDE/代码仓库/Jenkins 等集成支持
- ✧ COS 对象存储/TCR 镜像仓库 集成

可观测 白盒排障

- ✧ 函数日志，请求粒度日志记录，集成 CLS
- ✧ 指标监控，函数维度监控+实例维度监控，集成云监控
- ✧ 自定义告警，对接云监报告警平台，自定义规则，多渠道触达
- ✧ 实例登录，支持登录生产实例排障调试



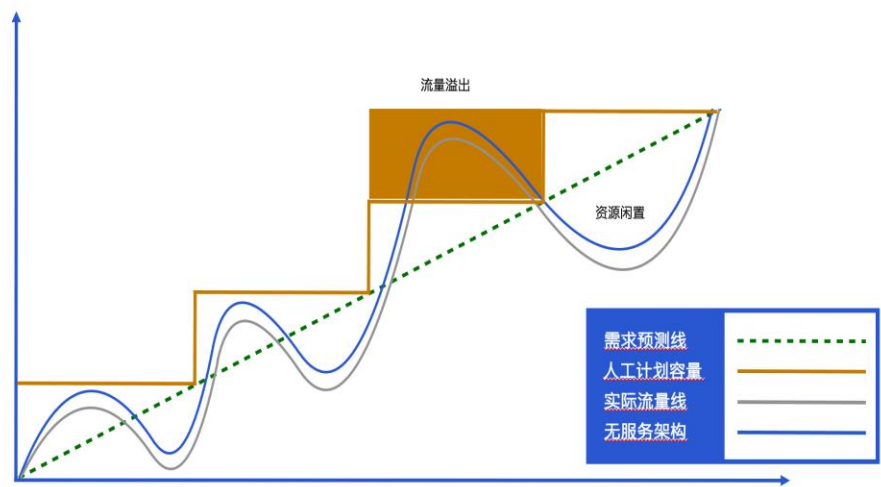
云函数事件触发机制



云函数运行机制

跨AZ动态扩缩容

云函数使用**基于实际请求在多可用区主动扩容**的策略，相较于容器基于CPU使用率/内存使用率的HPA被动扩容策略，可保证资源极致弹性，能根据调用量自动扩缩容，完美贴合调用曲线，最大程度节省资源成本。



最大程度节省资源闲置成本

资源无需常驻，仅在产生有事件触发时调度实例运行

真正的按量计费

毫秒级精确计费，只为代码**实际运行时长**付费

	云函数	容器产品
扩容机制	单实例单并发, 基于请求主动扩缩容	基于CPU/内存使用率的HPA扩容机制
业务侧表现	根据实际请求实时扩缩容, 贴合业务曲线	按照波峰流量预备资源, 造成大量资源闲置, 成本提升; 按照波谷流量预备资源, 用户请求无法及时响应, 影响体验
高可用机制	平台自带容灾机制, 支持跨可用区主动扩容, 故障时流量自动无感切换	业务方自行实现跨可用区容灾, 需要部署双倍资源应对机房单点故障等情况
业务侧表现	无需容灾投入, 单点故障时自动切换可用区, 用户无感知	额外进行容灾规划, 资源部署成本增高, 运维难度高; 单点故障时紧急切流, 业务侧感知明显





云函数SCF可以给企业带来巨大的降本增效价值

极致降本：70%以上

- 单价更低：函数算力单价对标cvm实例；
- 按量计费：只为代码**实际运行时长**付费，无资源常驻成本；
- 计费粒度更精细：**毫秒级**精确计费；

对于正在使用容器或CVM承载的存量业务，迁移至函数后，资源成本降幅可达70%以上

高可用质量保障

- 超高并发：支持横向扩展，单集群 **100 万 QPS** 稳定运行
- 弹性伸缩：动态扩缩容，**速度可达 10000 个/分钟**；
- 跨Zone容灾：**多集群部署**容灾机制，实现业务的不间断运行。

显著成效：

Top的车企客户

腾讯地图、腾讯文档、腾讯会议、腾讯游戏、腾讯数据中台、腾讯AI智能平台、腾讯视频等等



根据业务对响应时间的要求、单次处理的数据量级分为四种场景

近线场景

一般基于消息队列进行数据的**准实时**处理。可秒级或分钟级响应，不要求立即返回结果。如媒体转码、数据处理、模型推理、消息通知

在线场景

毫秒级响应，与前端实时交互处理用户请求。运行时间长，服务流量及资源利用率有潮汐特征，对服务 SLA 要求极高。



离线场景

算力**吞吐量**大，批量发起，不要求立即进行处理，对处理时间有一定的宽泛预期空间。运行期间资源利用率较高。如压测/拨测、数据归档、自动化(RPA)、其它定时任务

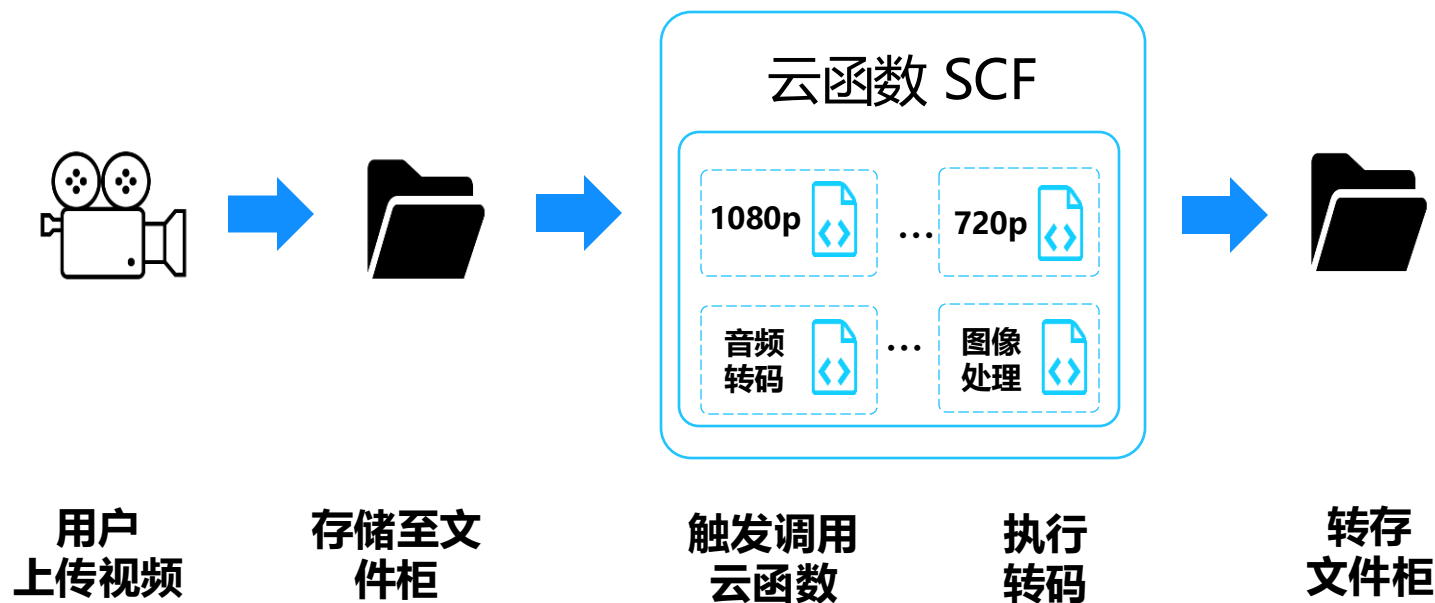
定时场景

固定执行周期或单次执行



场景一:GPU智能转码

智能转码主要是指采取AI技术能力，对视频进行智能化护理，类似于视频极速高清、视频超分、HDR优化、细节增强等，底层采取GPU提供对应的算力支撑。



使用 Serverless gpu音视频转码方案:

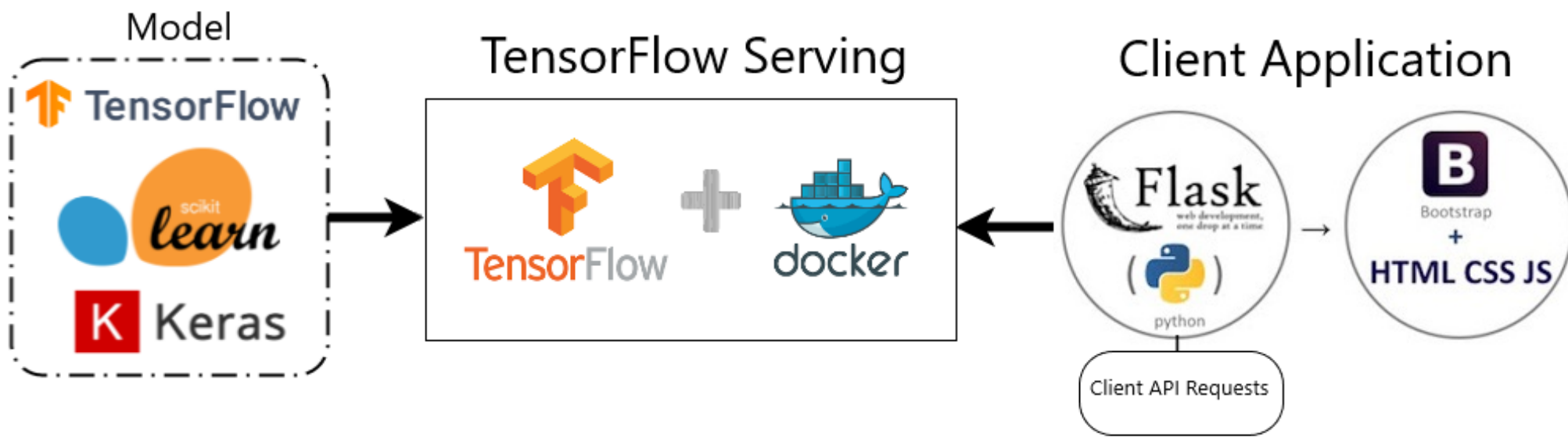
- **转码算法先进**: 采用腾讯云多媒体实验室的转码算法，提供算法模版，应用封装，快速部署适;
- **灵活可定制化**: 基于现有的功能进行扩展，实现定制化的转码或者业务需求;
- **成本更低**: 只有在转码期间才收费、免运维;
- **效率更高**: 可以通过切分视频并发转码，且不产生额外费用;



GPU场景和案例

场景二：GPU 推理

GPU普遍利用率在50%以下，甚至只能用到20%左右，而GPU的价格比较昂贵，客户的成本比较高。通过云函数运行封装好训练的模型，对外提供推理服务，**利用函数的弹性伸缩、按量计费提供性价比更高的AI推理方案**；此外云函数可以直接支持容器化的应用，可以把模型打包到容器镜像，创建函数的时候可以直接从TCR中加载镜像，降低客户的改造成本。

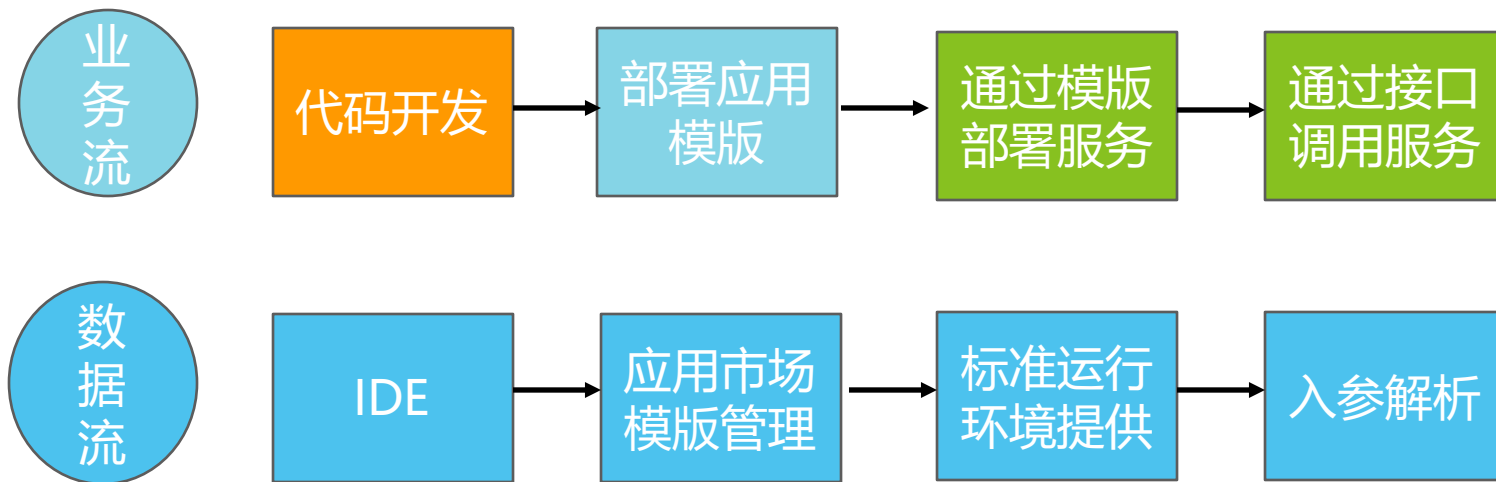


场景三:智能模型定制化交付

一些SAAS平台或者智能模型的提供会有一些定制化的交付，采取serverless应用可以形成标准的交付流程，极大程度上降低客户的交付成本。底层提供gpu算力的灵活调度，满足不同模型的算力要求（Serverless MCP）

方案优势：

- 1、资源隔离，完全归属客户小账号
- 2、标准化交付，降低交付成本
- 3、客户灵活调整算力，最大程度优化使用效果



Serverless与MCP融合创新，构建AI应用全新智能中枢

mcp server这类事件驱动型业务尤其适合部署在serverless函数（functiongraph）上，通过python/go等语言快速构建mcp server，实现毫秒级弹性扩容。4. 现存业务接入：
华为云 26天前

把MCP和AI代理部署在无服务器架构上，大幅提升业务性能

2. mcp server on serverless2.1. 效果展示先看看效果，模仿mcp 官方server例子开发一个天气查询的mcp server，同时部署到腾讯云函数。2.2. 天气查询MCP Server代码
腾讯技术工程 1个月前

Serverless MCP 运行时业界首发，函数计算让 AI 应用最后一公里提速

第二步：测试 MCP Server 提供的工具能力部署成功之后，通过触发器页面，拿到测试URL可对当前 MCP Server 进行测试。如果希望将部署的 MCP Server 用于生产，建议使...
阿里云原生 1个月前

MCP Server On FC之旅2: 从0到1-MCP Server市场构建与存量OpenAPI转MCP Server

STDIO MCP Server一键转为 SSE MCP ServerCloud Native效果1. 通过 FunctionAI 平台的模版 <https://cap.console.aliyun.com/create-project?template=start-mcp-fast-de...>
阿里云原生 4天前

当 MCP 遇上 Serverless，AI 时代的最佳搭档

第二步：测试 MCP Server 提供的工具能力部署成功之后，通过触发器页面，拿到测试URL可对当前 MCP Server 进行测试。如果希望将部署的 MCP Server 用于生产，建议...
阿里云原生 1个月前

PART 3

解决方案

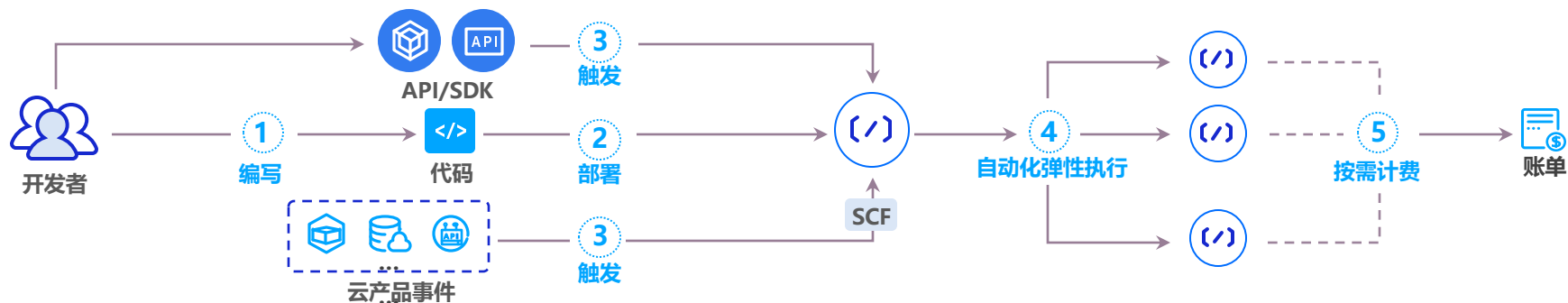
云函数SCF在智能辅助驾驶、车联网、智舱三大应用场景的解决方案

► 如何降低成本？Serverless，全新云模式，按实际用量收费，最大限度节约机器资源和运维成本

类似出行方案	网约车	租车	私家车
付费方式	按量付费	租约期付费	一次性付费
特点	不用不花钱	不用也要收钱	资源独占
服务器	Serverless	虚拟机	物理机
运维成本	免运维	运维软件	运维硬件



► 如何提高效率和可用性？云函数SCF新一代的云原生算力平台，无需运维+极致弹性



开发

- 上传业务代码/镜像，声明业务运行的规则

- 无改造** 支持上传容器镜像，兼容传统Web 框架
- 更易用** 一键部署，全自动运行

聚焦核心业务逻辑
无需运维底层资源



最高效率



SCF

- 基于实际请求动态拉起机器，运行业务逻辑

- 极致弹性** 支持上传容器镜像，兼容传统Web 框架
- 超高可用** 自动跨Zone容灾，多集群部署容灾机制
- 安全隔离** 自带租户级强隔离，保障业务安全运行

极致弹性扩缩容
跨区容灾机制



最高可用

- 按量付费** 按量毫秒级精确计费，只为代码实际运行时长付费，无资源常驻成本

自来水式计费模式
不为闲置资源付费



最优成本

▶ 智能辅助驾驶：百Gb吞吐数据处理，算力错峰复用实现最优成本

智能辅助驾驶-合规脱敏业务背景：

车辆在行驶过程中，会通过传感器采集各类环境数据，涉及各类敏感数据。根据政府合规要求，车企需对车牌/地理/人脸/轨迹等敏感信息进行脱敏处理，业务上可分为量产车在线脱敏和研采车离线脱敏，脱敏后的数据再继续应用于下游平台。



智能辅助驾驶：百Gb吞吐数据处理，算力错峰复用实现最优成本

智能辅助驾驶-量产车合规脱敏在线处理业务流程：



业务特征：

数据吞吐规模大，波峰波谷明显：节假日高峰时百万车辆规模，峰值带宽高达 339 Gbps，谷值 3.5 Gbps

处理时效要求高： T+1小时

脱敏成功率要求高： 95%以上



智能辅助驾驶：百Gb吞吐数据处理，算力错峰复用实现最优成本

智能辅助驾驶-量产车合规脱敏在线处理解决方案：

事件触发：根据文件类型将脱敏任务分发到不同函数上

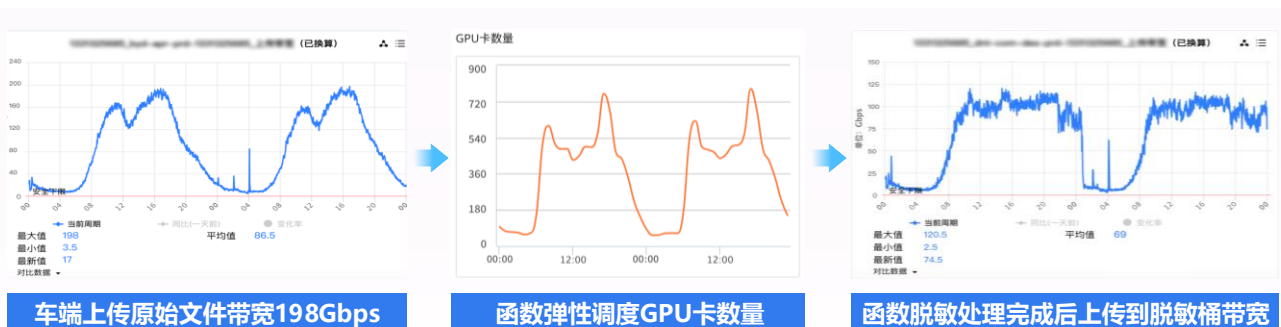
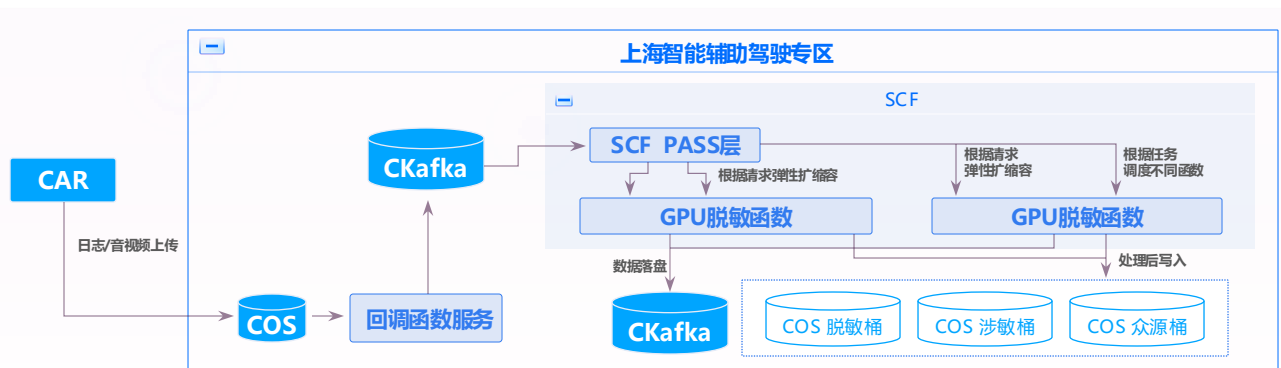
弹性扩缩容：根据实际数据量弹性调度资源

落地效果：

GPU单卡性能极致压榨：多个脱敏任务调度到同张卡运行，单 L20卡视频文件处理速度达到30MB/s

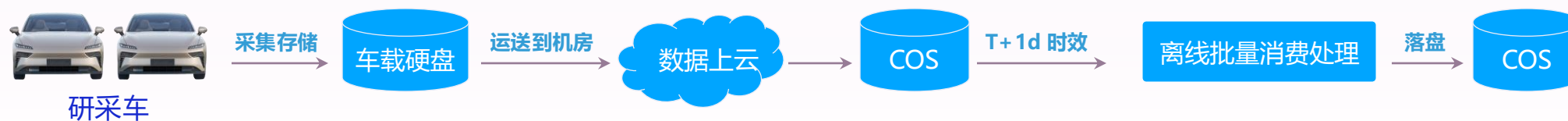
GPU资源弹性供给：晚高峰18点，数据传输平均带宽 180Gbps，供给750张卡；凌晨四点，数据传输平均带宽24Gbps，供给100张卡

包月+按量灵活计费，降低成本：相较于传统包月750张卡，综合资源成本降低30%+



智能辅助驾驶：百Gb吞吐数据处理，算力错峰复用实现最优成本

智能辅助驾驶-研采车合规脱敏离线处理场景业务流程：

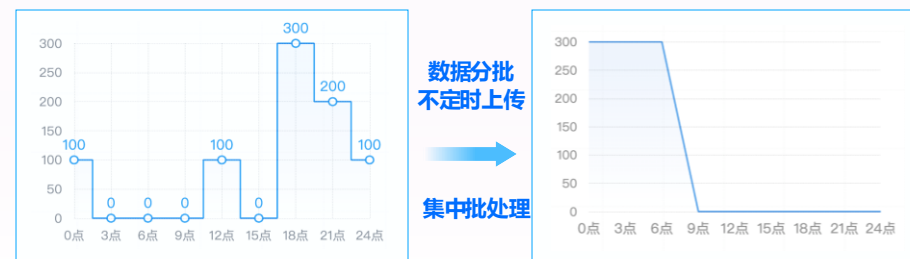


业务特征：

数据分批不定时上传：研采车数据通过物理硬盘运送到机房，在0点、12点、18点、21点不定时分批上传云端

数据量大：单硬盘容量4TB，峰值数据量可达800TB/天

离线批处理：每天0点~6点、16点、22点进行离线批处理，时效要求T+1天



在线脱敏和离线脱敏，资源诉求相同、使用时段不重叠 —— 资源调度优化的终极武器：错峰复用机制

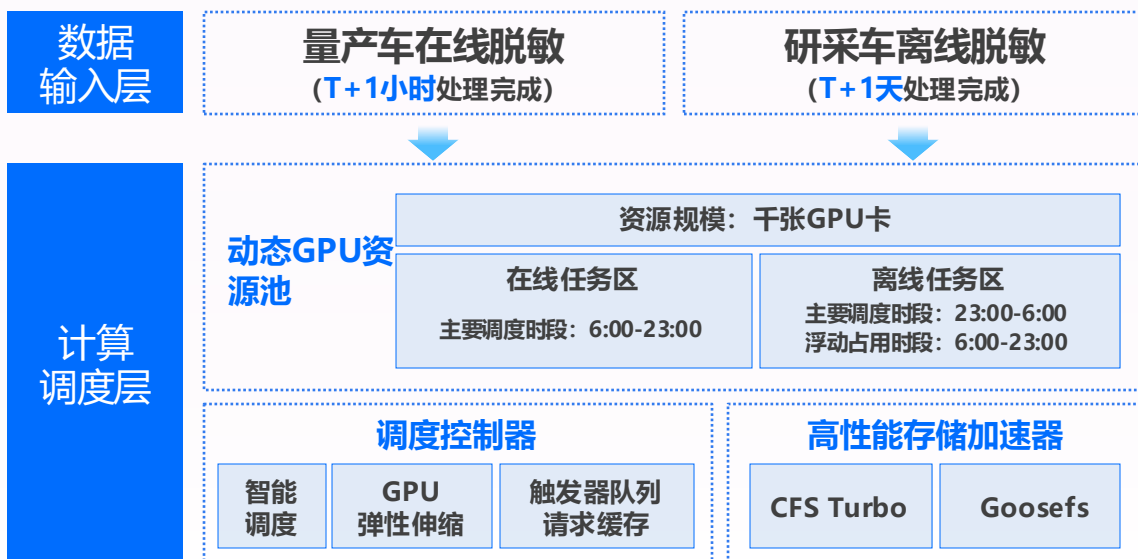


智能辅助驾驶：百Gb吞吐数据处理，算力错峰复用实现最优成本

智能辅助驾驶-研采车合规脱敏离线处理场景解决方案：

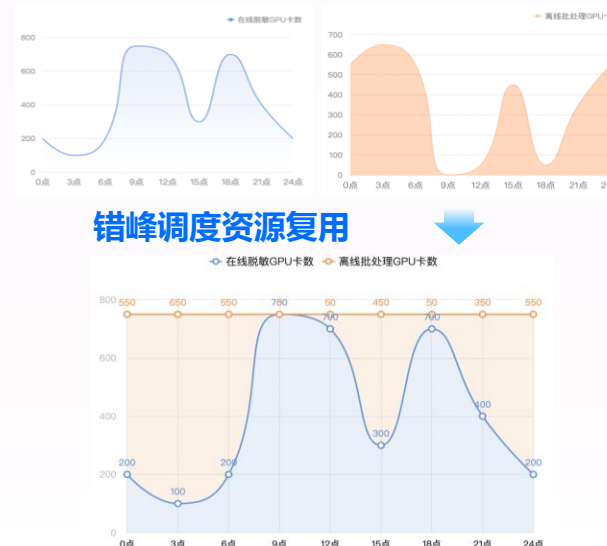
任务实时接收进入待执行队列：离线数据上云后，先进入队列，等待在线业务释放GPU资源后进行批处理

任务智能调度：在线任务释放出可用资源后，智能调度离线任务



落地效果：

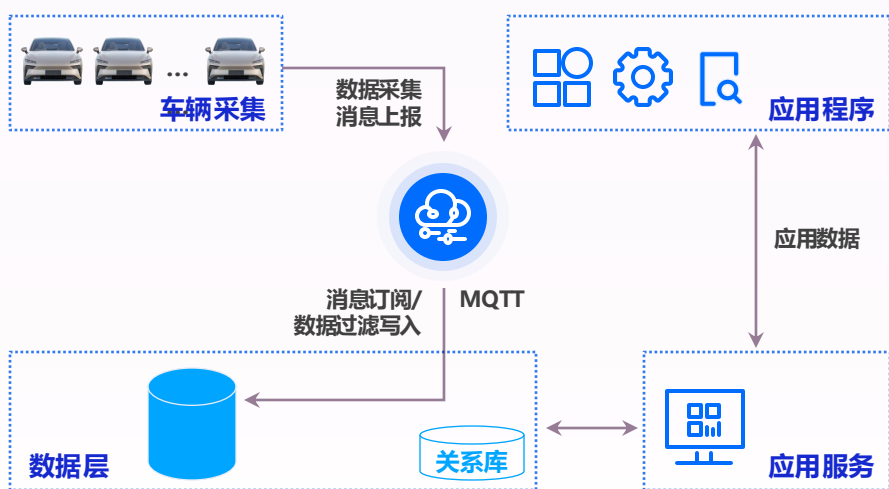
使用700+卡即可同时完成两个业务1600+TB/天的数据处理，无需为离线任务新增300+GPU卡，总成本降低30%



车联：百万车辆并发连接，毫秒级数据实时处理

车联网-EDR 数据实时处理场景业务背景：

行驶过程中，车辆通过传感器和控制器收集车辆位置、速度、油量、发动机状态等信息，通过MQTT 协议实时或定期传输，以进行数据分析、故障预警、远程控制等操作。



业务特征：

超大并发：高峰期10万+终端车辆同时在线

峰谷连接数差异大：早8点10万+连接数，凌晨4点3千+连接数，超十倍差异

数据处理实时性要求高：消息生产后要求秒级处理返回



车联网络：百万车辆并发连接，毫秒级数据实时处理

车联网-EDR 数据实时处理场景解决方案：

车辆EDR数据通过MQTT触发器毫秒级实时消费，瞬间拉起业务函数进行清洗处理，提高数据流转效率，满足百万辆终端的快速吞吐。

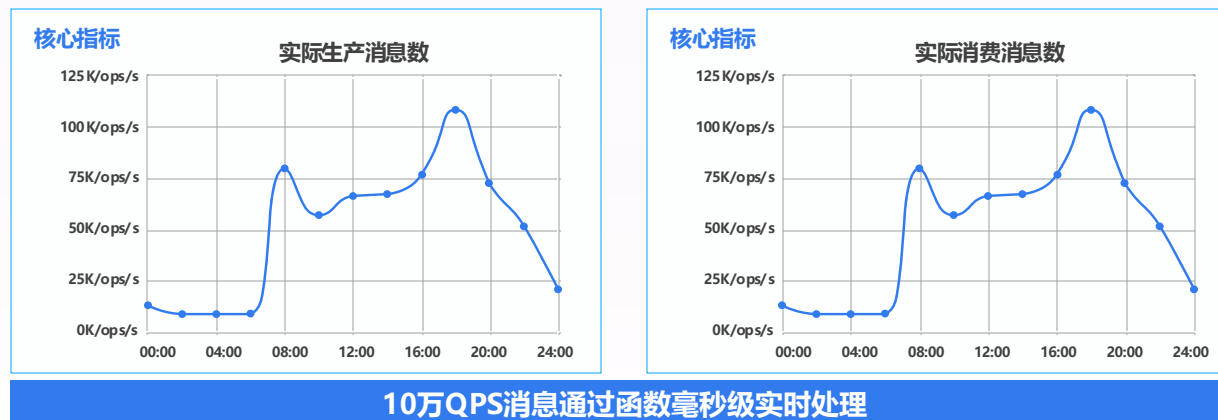


落地效果：

超高并发处理：支撑车端10万QPS数据上报

实时消费&超低时延：160ms内消费完成，投递
→消费→调用→执行全链路总耗时198ms

降低资源成本：资源高弹性，无常驻成本，相较
传统方案，节省70%+



▶ 智舱：跨部门拉通协作效率，微服务开发模式轻松提高

智舱-个性化推荐场景业务背景：

在车机埋点采集用户行为数据，分析用户画像并上报至推荐系统，产生与用户高度匹配的推荐信息（目的地周边推荐、精品推荐），实现“千人千车千面”的个性化服务。

落地效果：

跨部门多模块协作复杂：个性化推荐链路涉及车机系统、埋点采集、数据模型训练等7+模块，跨部门协调对齐工作量大

迭代周期长：单体架构模块耦合性强，迭代需同步评估上下游影响，需求响应周期长

系统稳定性依赖全局资源设计：业务高峰期流量波动时，全链路的资源分配与容错机制需统一规划，局部优化空间有限

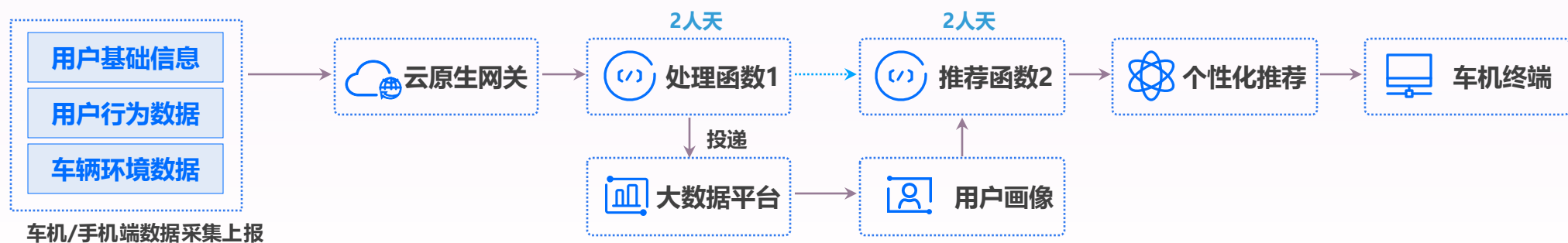
某车企，服务上线需拉通10+人，打通7+模块，花费20+人天



▶ 智舱：跨部门拉通协作效率，微服务开发模式轻松提高

智舱-个性化推荐场景解决方案：

函数作为微服务开发模式的最佳体现，通过**丰富的触发器组件**和**内置网关URL**，快速串联多个业务模块，打通应用端埋点采集→大数据实时处理→推荐服务响应全链路。



落地效果：

协作效率高：模块开发投入周期从20+人天缩短至4人天，节省80%人力投入

运维自动化水平提升：函数自带高可用自动化运维机制，业务聚焦单模块核心逻辑编写，运维投入减少50%

系统可靠性增强：弹性扩容机制应对10倍流量波动，并发管控模块隔离异常流量，整体服务可用性达99.95%

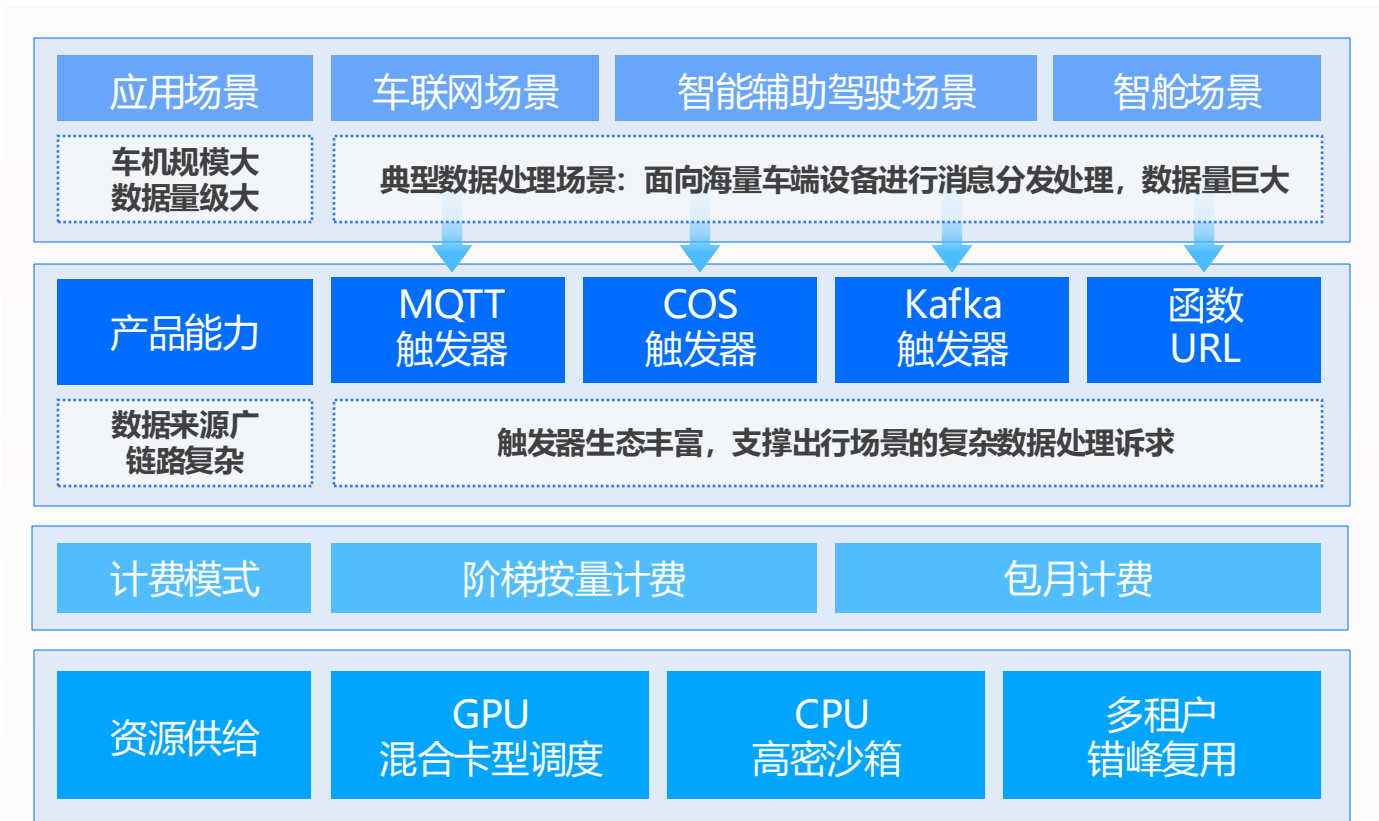
PART 4

总结与展望

云函数SCF关键应用能力升级，为出行行业筑牢底座



云函数SCF关键应用能力升级，为出行行业筑牢底座



-  兼容企业财务预算流程，
综合成本最优
-  弹性 GPU 资源供给，
保障业务高可用
-  免运维的应用托管平台，
提升开发效率

国内首个支持
Serverless
GPU混合卡型算力调度
GPU资源包月
GPU跨地域调度



参与调研您将优先获得



AiDD定制版
《AI+软件研发精选案例》



专属学习顾问
1对1需求对接

AiDD会后小调研

AiDD峰会致力于协助企业利用AI技术深化计算机对现实世界的理解,推动研发进入智能化和数字化的新时代。作为峰会的重要共建者,您的真知灼见对我们至关重要。衷心感谢您的参与支持!



扫码参与调研

2025 AI+研发数字峰会

拥抱 AI 重塑研发

科技生态圈峰会 + 深度研习

——1000+ 技术团队的选择



敦煌站

K+ 思考周®研习社

时间: 2025.08.29-30



上海站

K+ 金融专场

时间: 2025.09.26-27



香港站

K+ 思考周®研习社

时间: 2025.11.17-18



K+峰会详情



上海站

AI+研发数字峰会

时间: 2025.05.23-24



北京站

AI+研发数字峰会

时间: 2025.08.08-09



深圳站

AI+研发数字峰会

时间: 2025.11.14-15



AiDD峰会详情

AiDD 峰会



2025 AI+研发数字峰会
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

