

UCloud



容器与Serverless架构实践

叶理灯

UCloud创新产品线研发总监

QCon

全球软件开发大会

10月17-19日 上海·宝华万豪酒店



扫码锁定席位

九折即将结束

团购还享更多优惠，折扣有效期至9月17日

扫描右方二维码即可查看大会信息及购票



如果在使用过程中遇到任何问题，可联系大会主办方，欢迎咨询！

微信：qcon-0410

电话：010-84782011

ArchSummit

全球架构师峰会 2017



扫码锁定席位

12月8-9日 北京·国际会议中心

七折即将截止立省2040元

使用限时优惠码AS200，

以目前最优惠价格报名ArchSummit

仅限前20名用户，优惠码有效期至9月19日，

扫描右方二维码即可使用



如果在使用过程中遇到任何问题，可联系大会主办方，欢迎咨询！

微信：aschina666

电话：15201647919

极客搜索

全站干货，一键触达，只为技术

s.geekbang.org



扫描二维码立即体验

有没有一种搜索方式，能整合 InfoQ 中文站、极客邦科技旗下12大微信公众号矩阵的全部资源？

极客搜索，这款针对极客邦科技全站内容资源的轻量级搜索引擎，做到了！

扫描上方二维码，极客搜索！

这里只有 技术领导者

EGO会员第二季招募季正式开启



E小欧

报名时间：9月1日-9月15日

扫描添加E小欧，
邀您进入EGO会员预报名群

立即报名



UCLLOUD

TABLE OF CONTENTS

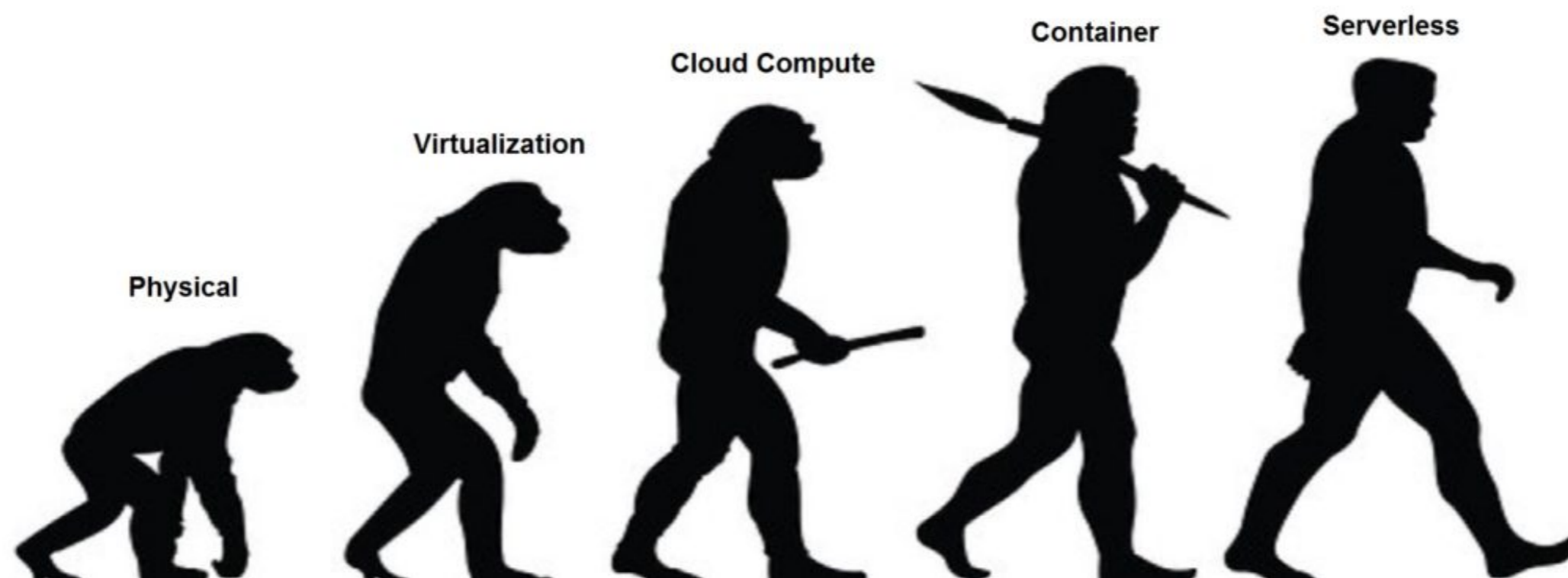
我们的问题

解决方案

Serverless

我们的实践

我们的问题 1/2



我们的问题 2/2

- 98%的物理机CPU峰值使用率不超过50%
- 据《福布斯》杂志统计，在商业和企业数据中心的典型服务器仅有5%-15%的平均最大利用率

UCLLOUD

TABLE OF CONTENTS

我们的问题

解决方案

Serverless

我们的实践

解决方案 1/3

- 基于闲置资源推出一个Pay as you go的计算类产品，计算服务化

解决方案 2/3

- IaaS
 - 装箱算法
 - NP完全问题
 - 非按需使用

解决方案 3/3

- PaaS
 - 供应商绑定
 - 不彻底的按需使用
- Serverless?

UCLLOUD

TABLE OF CONTENTS

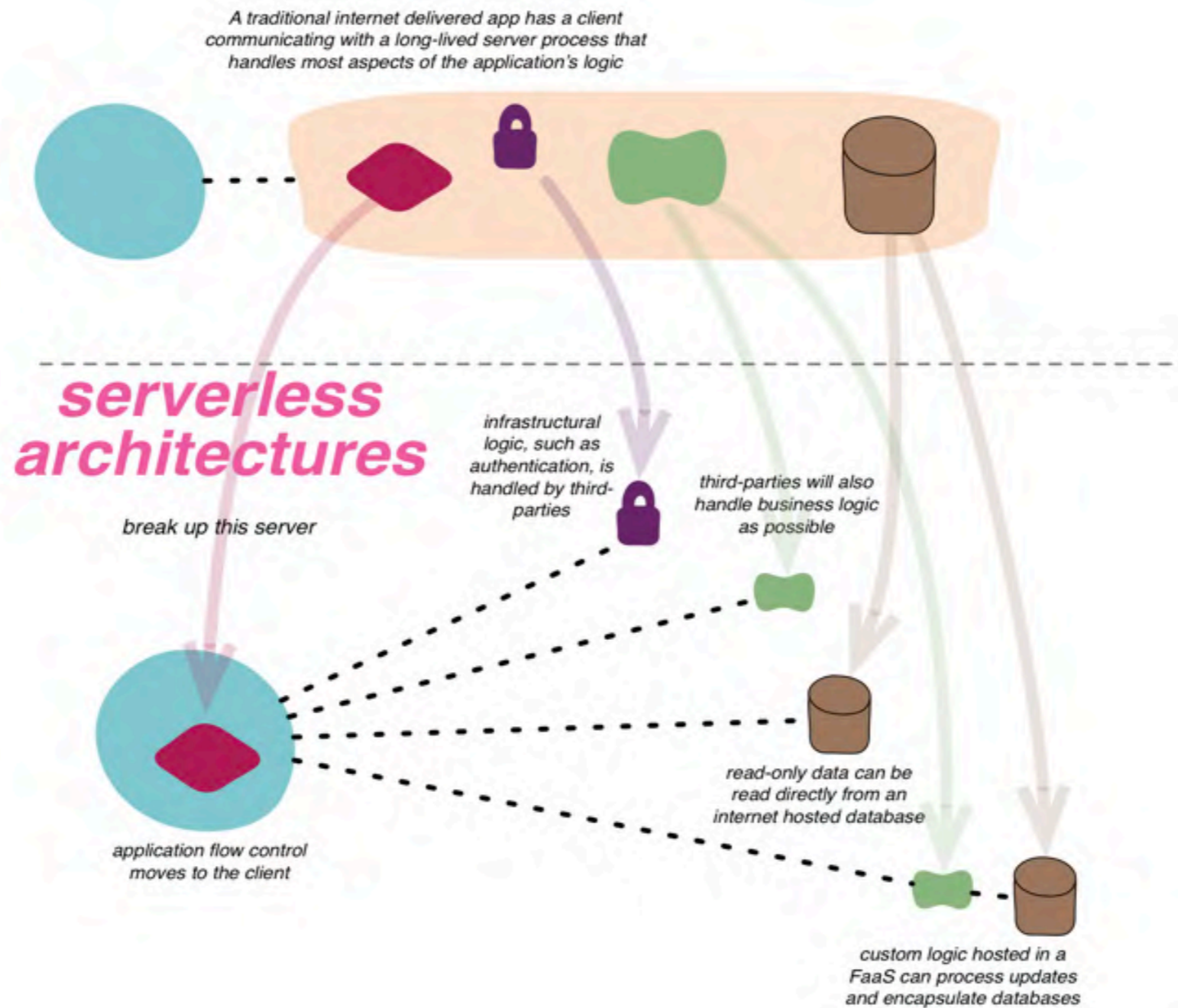
我们的问题

解决方案

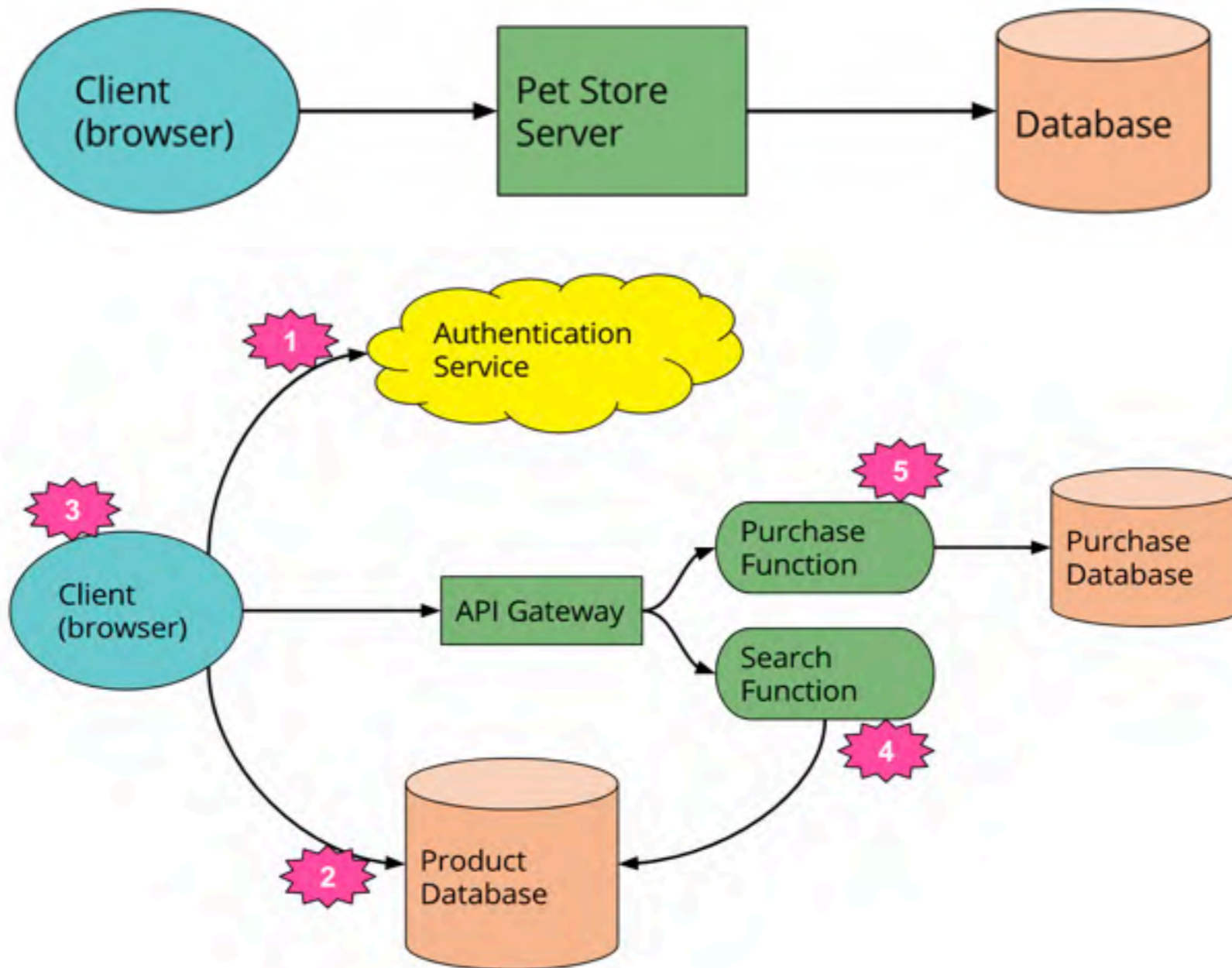
Serverless

我们的实践

Serverless 1/2



Serverless 2/2



UCLLOUD

TABLE OF CONTENTS

我们的问题

解决方案

Serverless

我们的实践

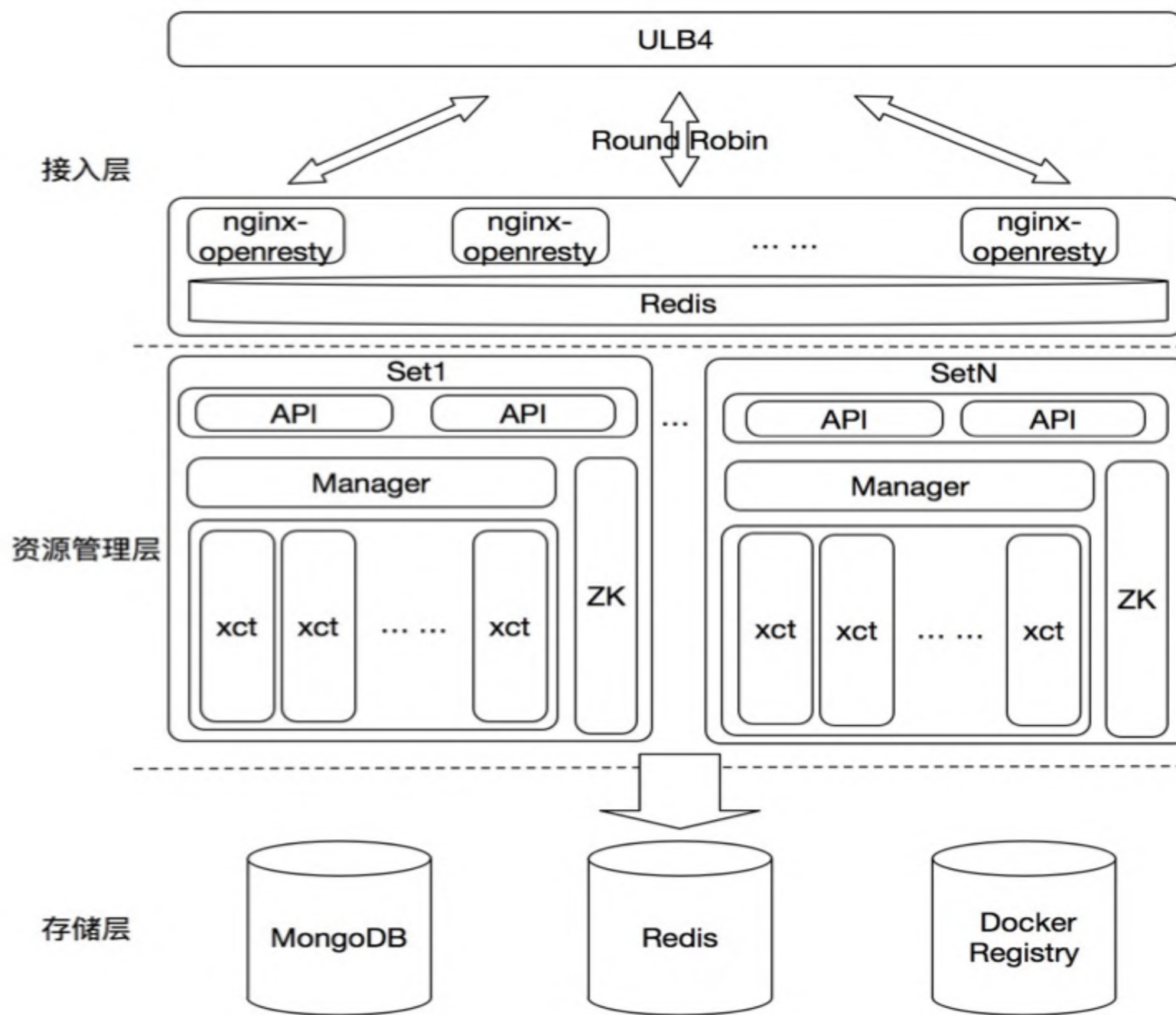
U CLOUD 我们的实践 1/7

- 闲置资源+计算服务化
 - 闲置资源通过受限虚拟机提供
 - 把计算抽象为:输入 + 算法 + 输出
 - 提供API

UCLLOUD 我们的实践 2/7

- 算法以Docker方式打包
 - 提供Docker 镜像仓库
 - 具备标准化+跨语言优点
- 以Restful API方式提供服务
 - 通过API传递输入+拉起算法
 - 输出通过API返回

Ucloud 我们的实践 3/7



UCLLOUD 我们的实践 4/7

- 架构要求
 - 高可用
 - 可扩展性强

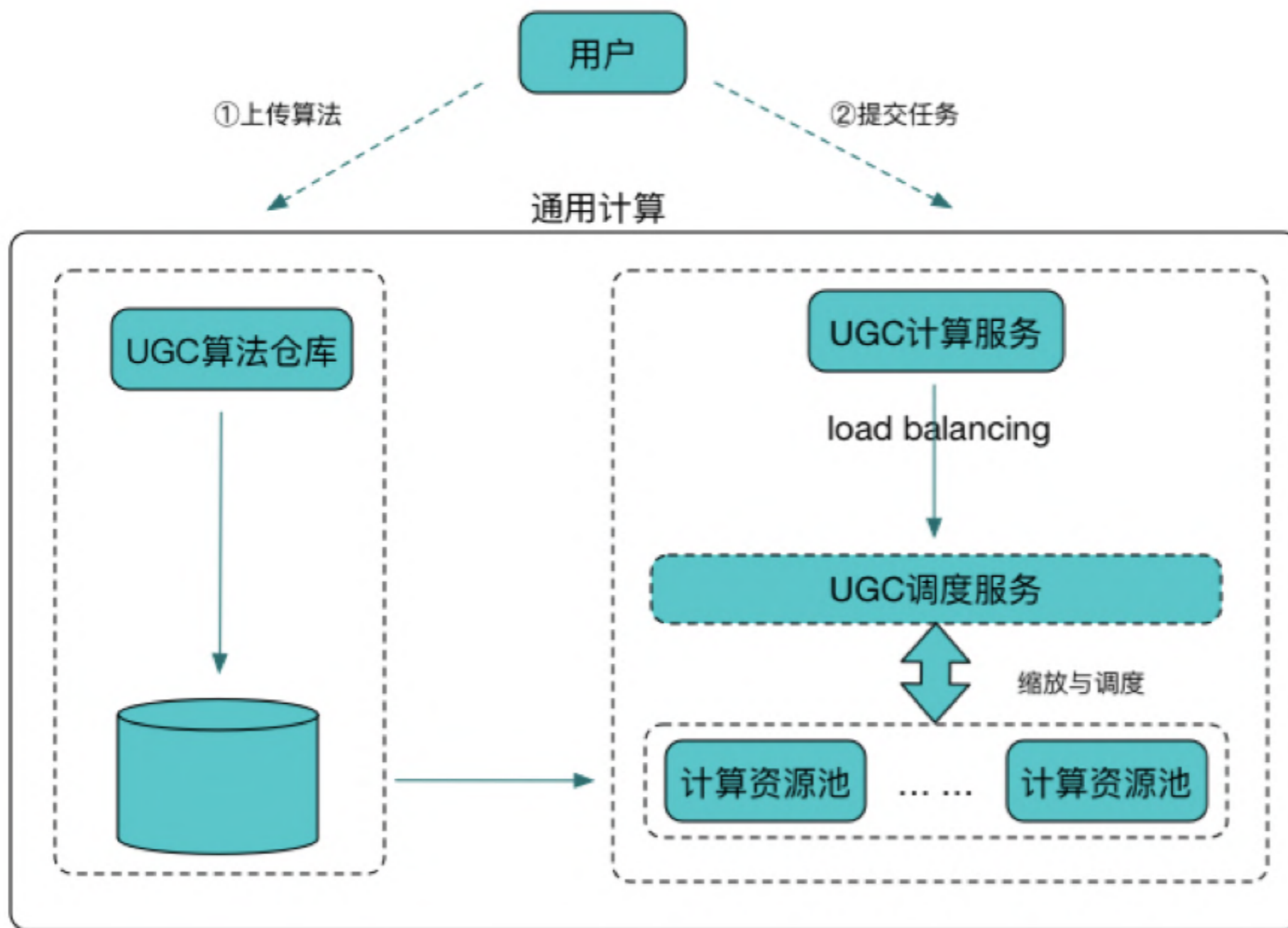
UCloud 我们的实践 5/7

- 支持跨可用区容灾
- 按Set部署
- 灰度发布
- 所有模块可以平行扩展

UCLLOUD 我们的实践 6/7

- 好架构的标准
 - 控制软件复杂度
 - 避免机器瓶颈
 - 控制人的错误的影响的范围

Ucloud 我们的实践 7/7





案例-图片处理 1/3

需求

- 实时
- 高并发
- 跨语言的算法

痛点

- 物理机机群维护
- 算法更新频繁
- 容量管理困难



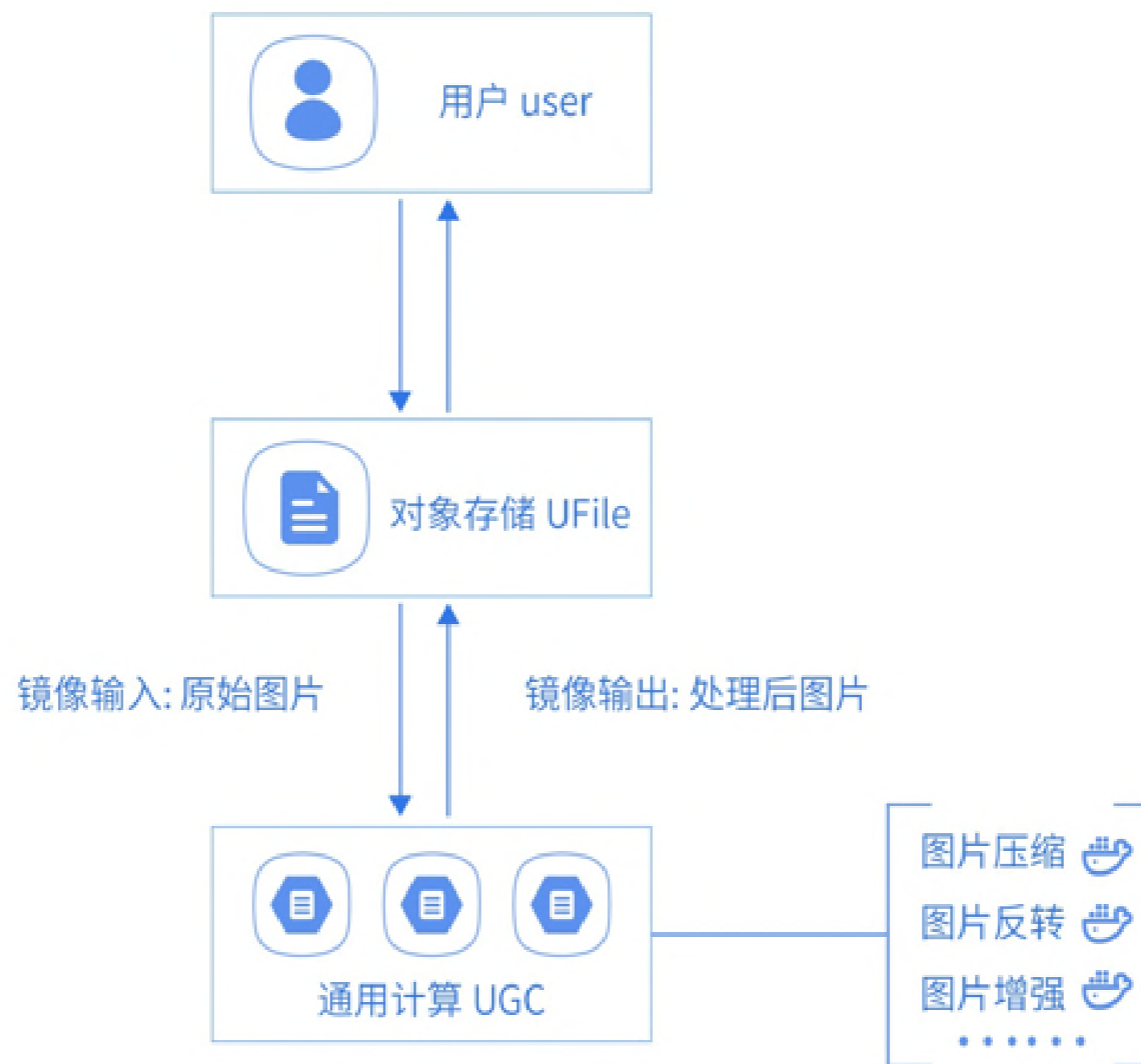
案例-图片处理 2/3

解决方案

- 将图片算法独立出来
- 利用UGC做Serverless架构改造
- NoOps



案例-图片处理 3/3





案例-基因数据分析 1/3

需求

- 小文件处理
- 自定义算法
- 高并发
- 开放平台

痛点

- 自己搭建架构过于复杂
- 资源维护及算法维护工作繁重
- 请求峰值很难预测



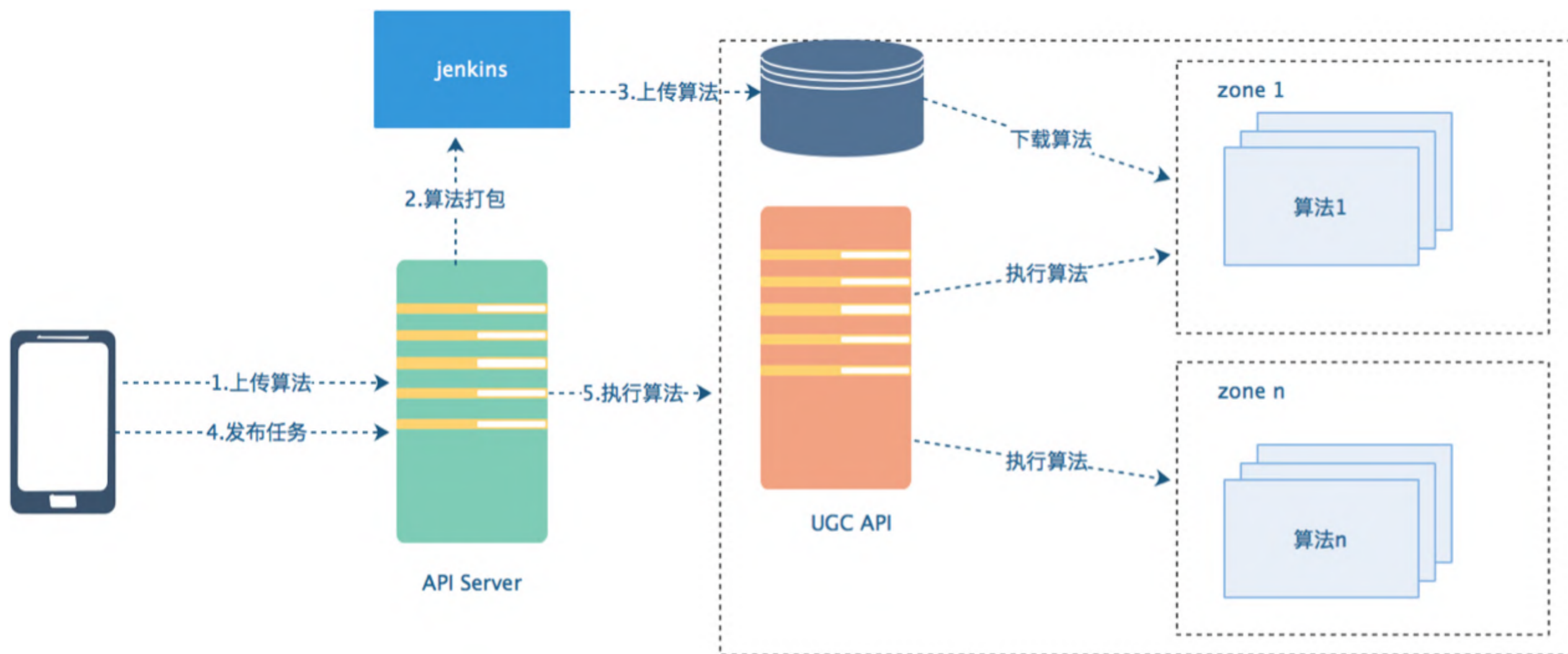
案例-基因数据分析 2/3

解决方案

- 算法打包上传至通用计算
- 终端调用转发到通用计算
- 简化架构免去维护资源成本



案例-基因数据分析 3/3





案例-OCR识别 1/3

需求

- 实时
- 机器学习模型打包
- 高并发

痛点

- 为了节省成本，需要优化Server性能
- 算法工程师不擅长Server端优化
- 为了应对峰值需准备一大批物理机器，浪费严重



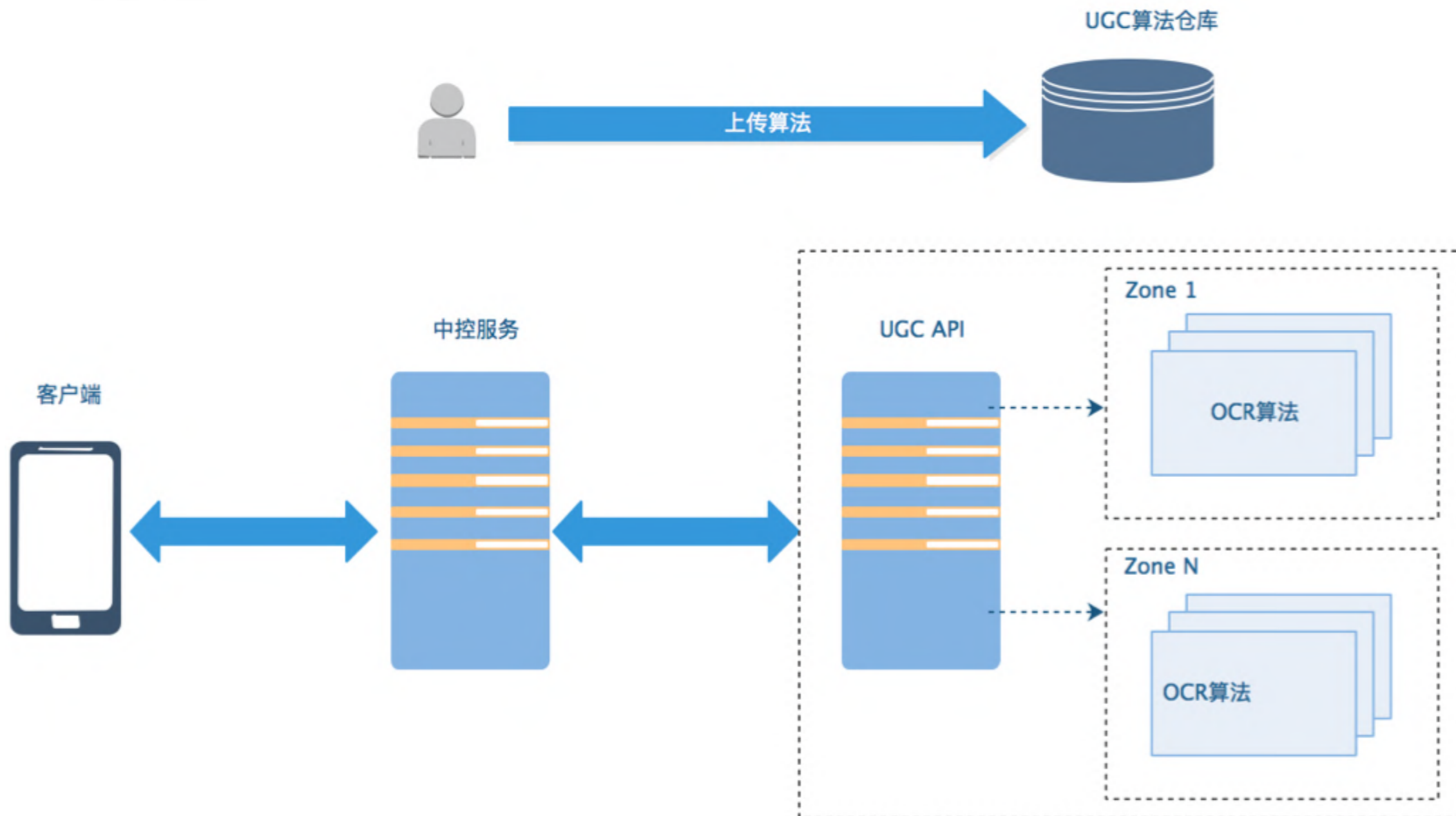
案例-OCR识别 2/3

解决方案

- 模型和代码打包上传到通用计算
- 通过通用计算API将OCR识别嵌入到业务逻辑
- 废除物理机机群



案例-OCR识别 3/3



没有银弹

- 无状态
- 相对高延时
- 粘合剂

关注 UCloud , 分享与交流更多

UCloud云计算

(微信号 : ucloud2012)



UCloud技术公告牌

(微信号 : ucloud_tech)



THANKS!

智能时代的新运维

CNUTCon 2017