阿里异地多活与同城双活的架构演进

谢吉宝（花名：唐三）
阿里巴巴高级技术专家

CNUTCon 2017
全球运维技术大会

上海·光大会展中心大酒店 | 2017.9.10–11

智能时代的新运维

大数据运维
DevOps
安全
SRE
Kubernetes
Serverless
游戏运维
AIOps
智能化运维
基础架构
监控
互联网金融

主办方 Geekbang. 极客邦科技 InfoQ

SPEAKER
INTRODUCE

唐三　阿里巴巴高级技术专家

- 谢吉宝，花名唐三
- 2010年加入阿里，10余年技术研发和系统架构经验
- 阿里期间主导设计了灰度发布系统、共享服务化平台、中间件运维平台、建站平台
- 目前负责阿里异地多活和同城双活的高可用体系建设和中间件的DevOps

TABLE OF
CONTENTS 大纲

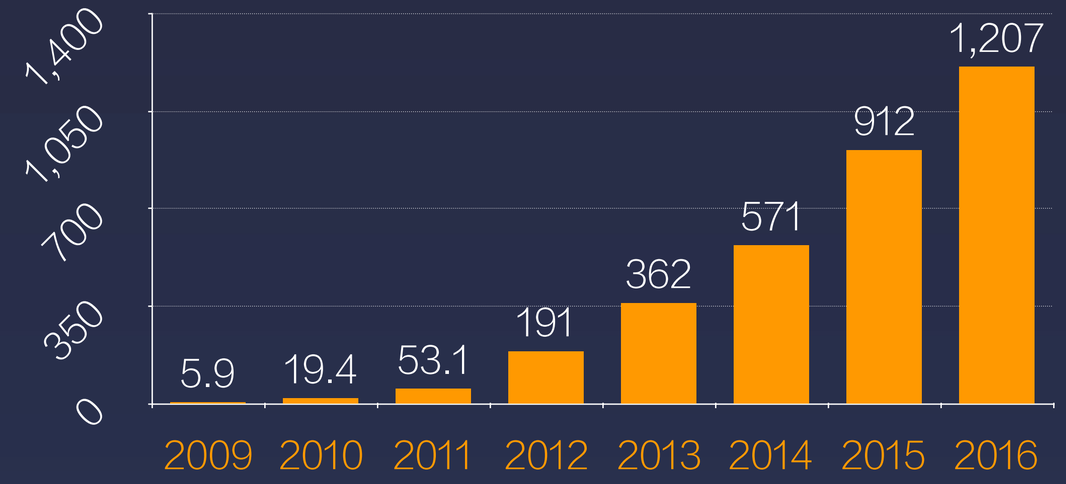
- 阿里异地多活的背景及解决方案
- 阿里异地多活的演进历程
- 阿里同城双活的新挑战
- 阿里基于多活架构的容灾及扩展能力
- 阿里异地多活技术未来展望

阿里技术架构的演进史

- 1.0 → 2.0时代
  - LAMP向单体Java应用演进（性能）
- 2.0 → 3.0 时代
  - 单体应用向大型分布式架构演进（效率）
- 3.0 → 4.0 时代
  - 单IDC架构向多IDC架构演进（容量、稳定）

为什么做异地多活

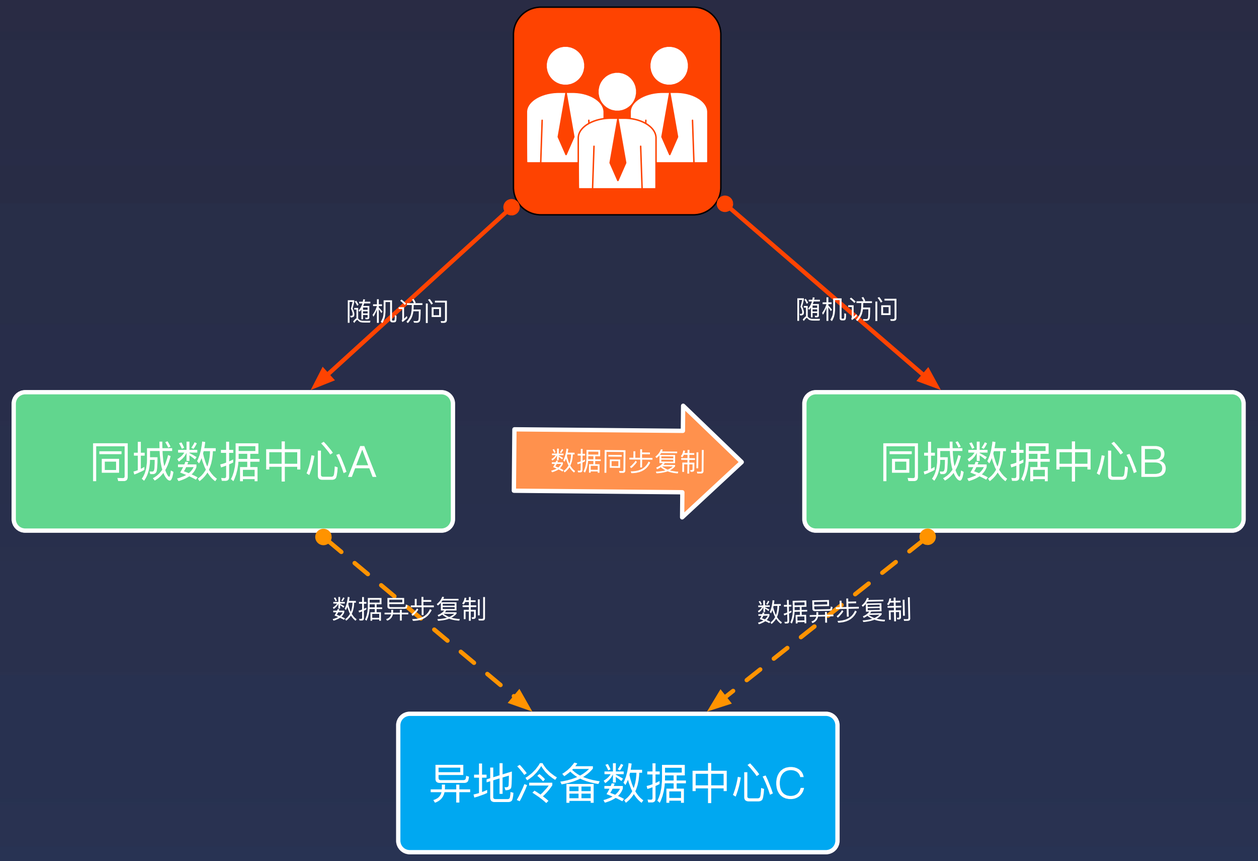
历年双十一销售额汇总



历年双十一秒级峰值汇总

- 资源：电商、阿里云、大数据等业务高速发展，单地资源容量受限
- 扩展：业务多元化对异地部署需求
- 容灾：天灾、人祸都会影响业务的可用性

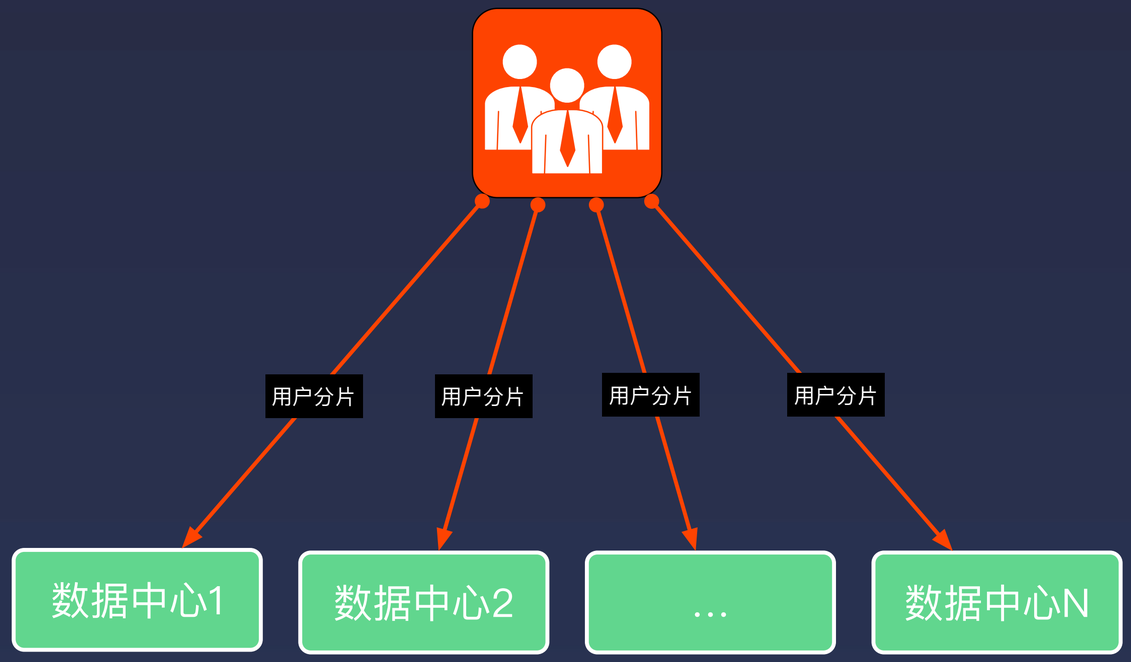
业界主流容灾方案



两地三中心方案

- 冷备中心不工作，关键时刻不敢切
- 冷备中心不工作，成本存在严重浪费
- 本质上数据仍然单点写，数据库瓶颈无法解
- 资源、容灾、扩展仍然未得到解决

理想化的解决方案



- 按用户分片，访问不同数据中心，随意切换
- 数据中心内的业务完成自调用闭环
- 数据中心无限水平复制

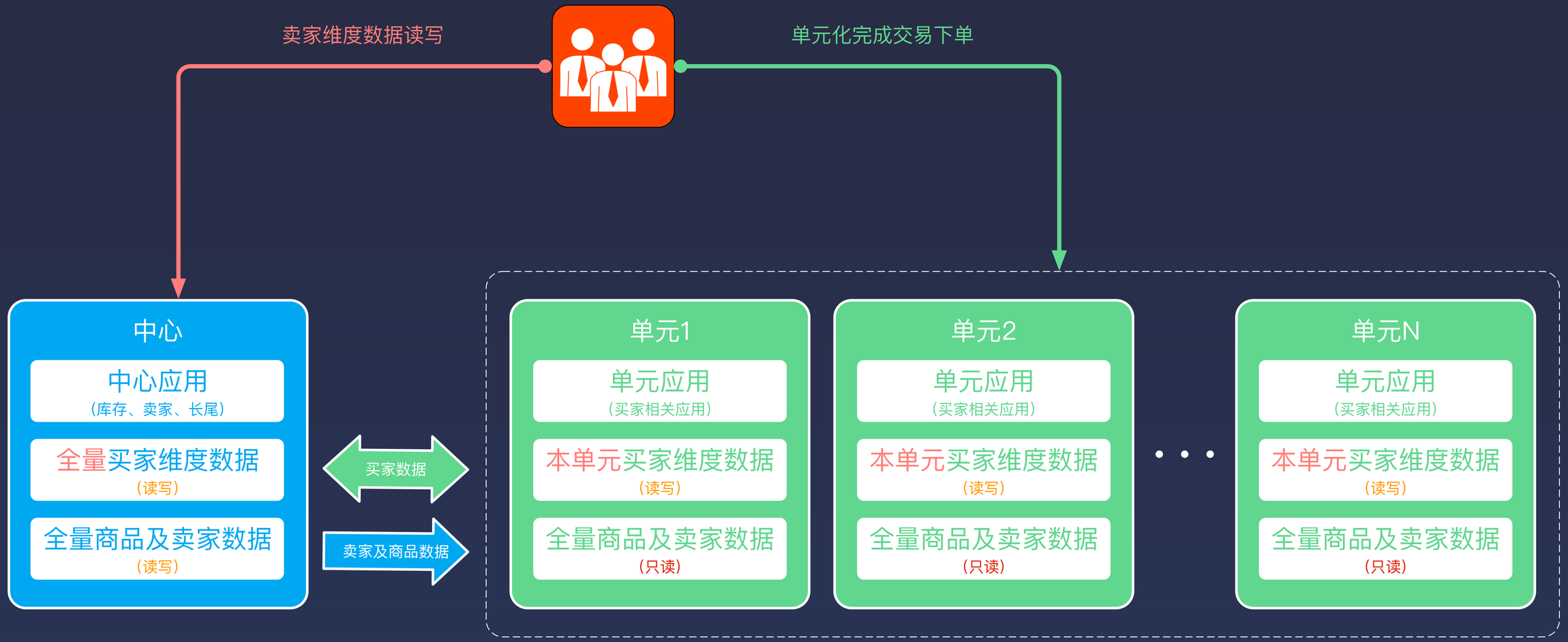
现实很骨感

- 数据维度多
  - 买家、卖家、商品三个维度
- 业务多且复杂
  - 业务太多，应用之间依赖关系错综复杂
  - 一次业务调用对应上百次原子调用
- 业务实时性要求高
  - 同城 < 2ms
  - 异地6ms - 100ms之间；

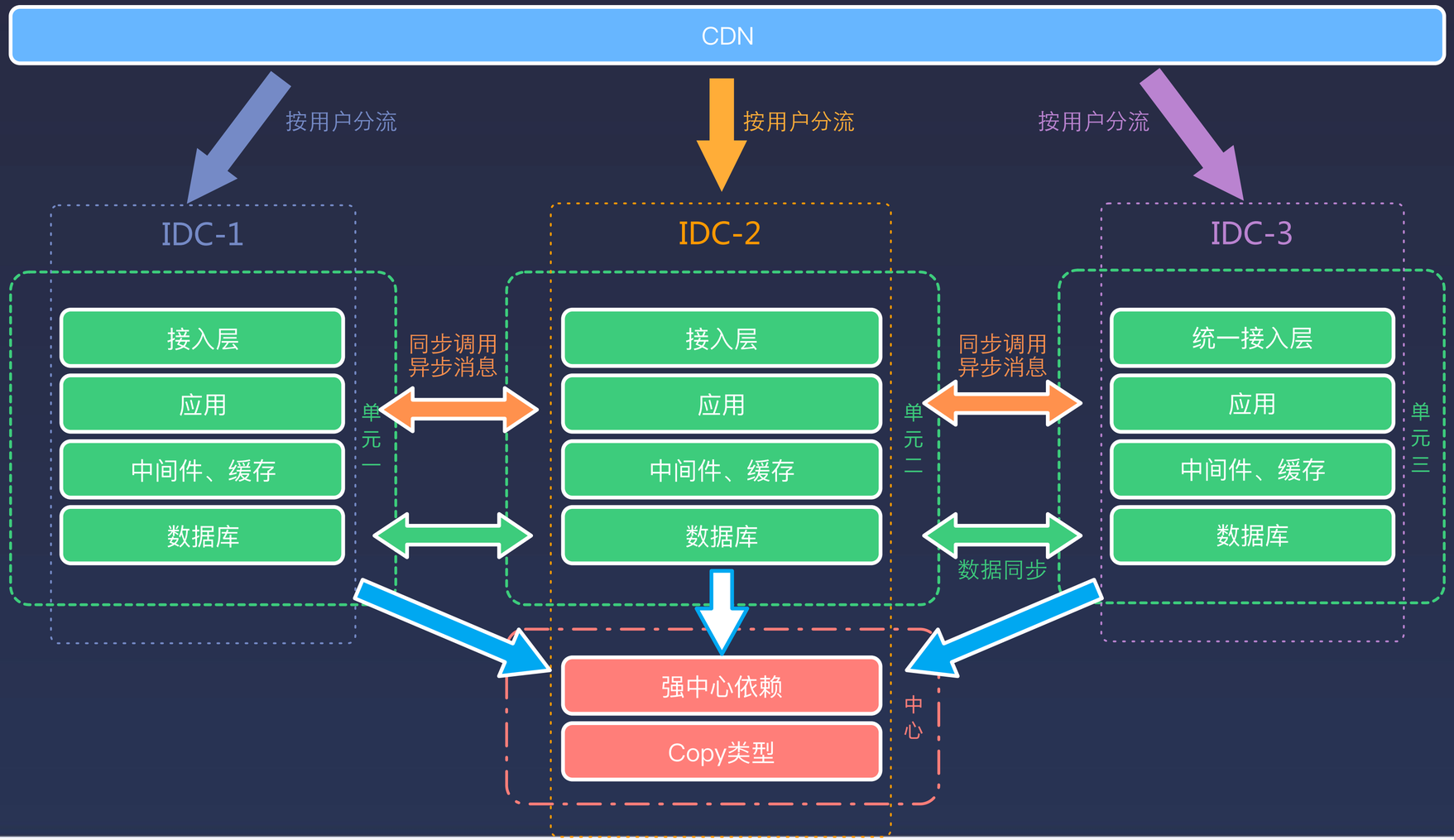
基本原则

- 按买家维度来进行数据切片
- 只取与买家链路相关的业务（单元）做“多活”
- 单元内最大限度的封闭
- 无法接受数据最终一致的跨单元单点写

业务架构


卖家维度数据读写
单元化完成交易下单
中心
中心应用
（库存、卖家、长尾）
全量买家维度数据
（读写）
全量商品及卖家数据
（读写）
买家数据
卖家及商品数据
单元1
单元应用
（买家相关应用）
本单元买家维度数据
（读写）
全量商品及卖家数据
（只读）
单元2
单元应用
（买家相关应用）
本单元买家维度数据
（读写）
全量商品及卖家数据
（只读）
单元N
单元应用
（买家相关应用）
本单元买家维度数据
（读写）
全量商品及卖家数据
（只读）

技术架构

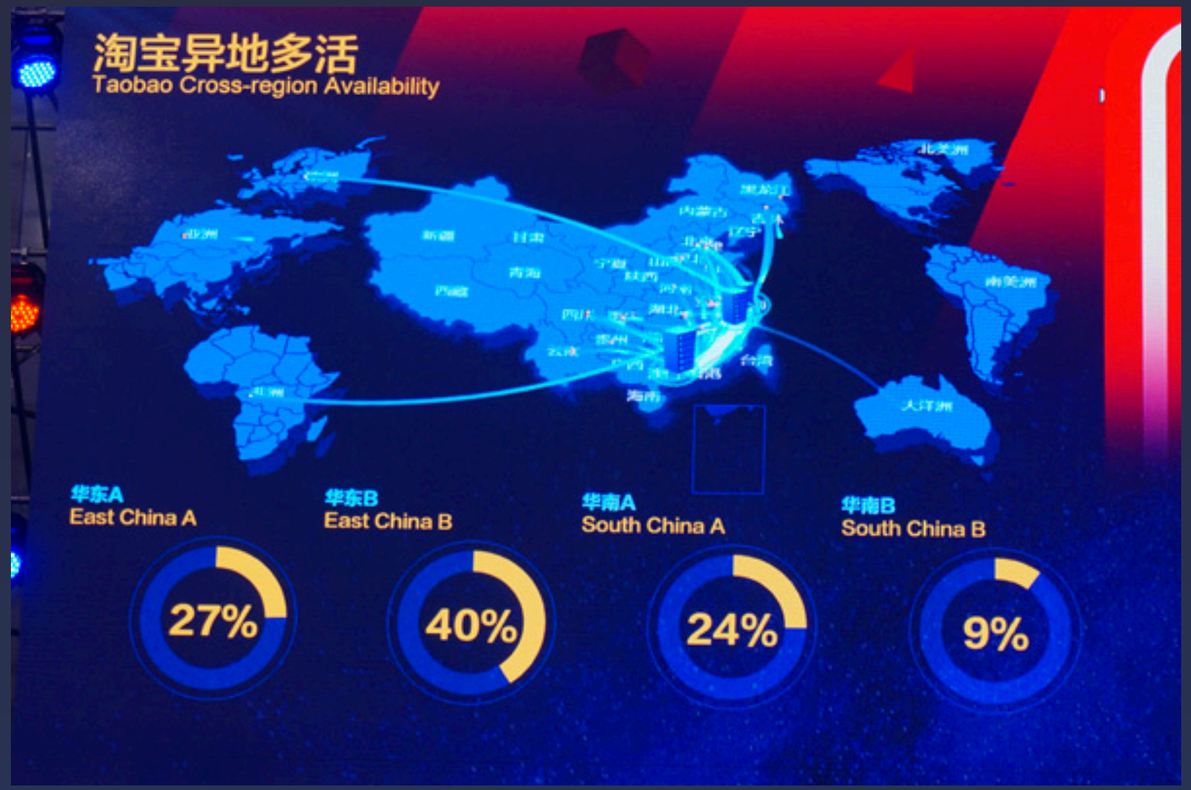

CDN
按用户分流
按用户分流
按用户分流
IDC-1
IDC-2
IDC-3
接入层
应用
中间件、缓存
数据库
同步调用
异步消息
同步调用
异步消息
统一接入层
单元一
单元二
单元三
数据同步
强中心依赖
Copy类型
中心

技术挑战

- 路由一致
- 数据延时
- 数据的正确性

为期三年的单元化项目

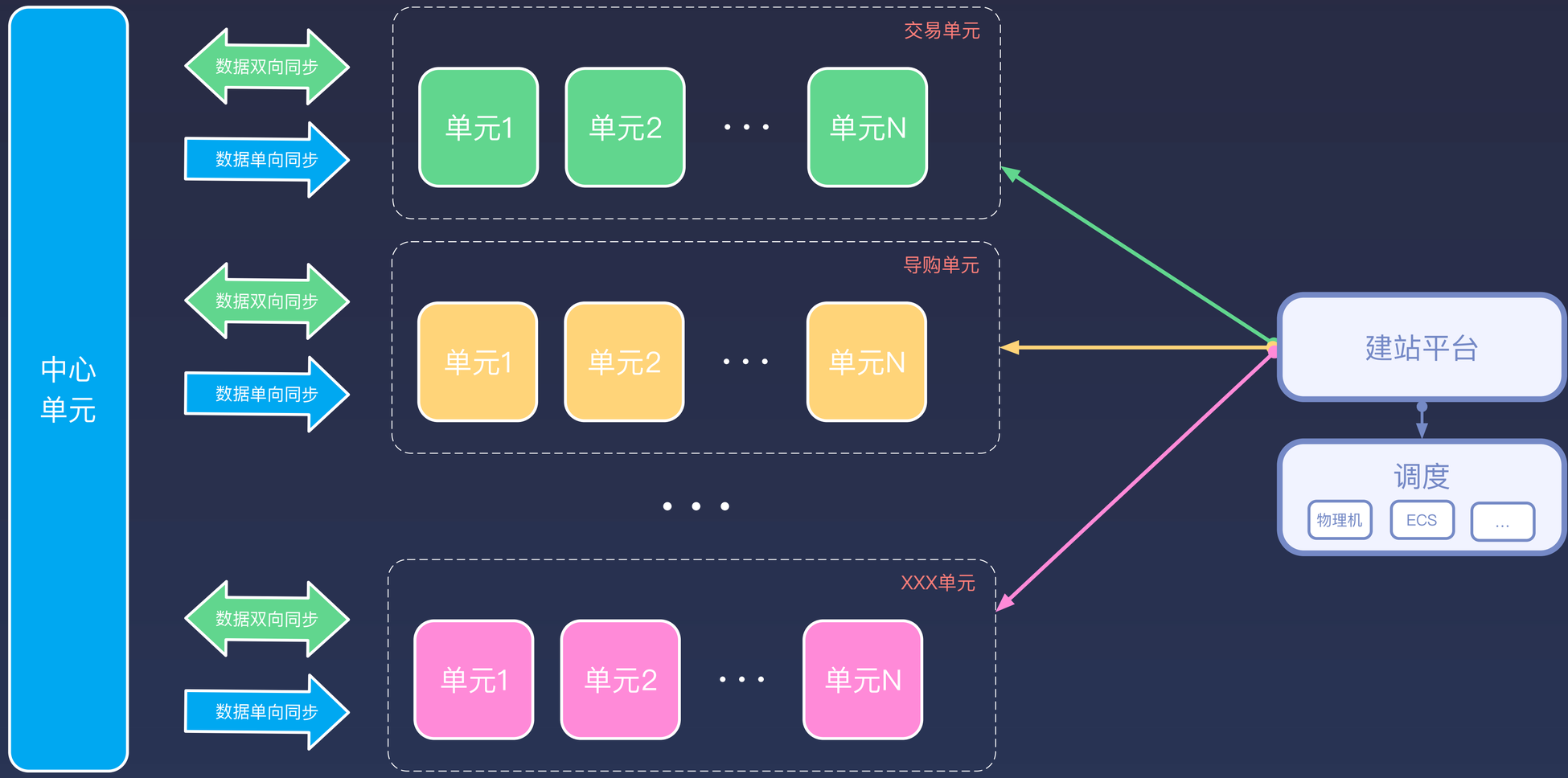
- 2013：杭州同城两个POC验证
- 2014：杭州、上海近距离两个单元
- 2015：千里之外的三地四单元架构



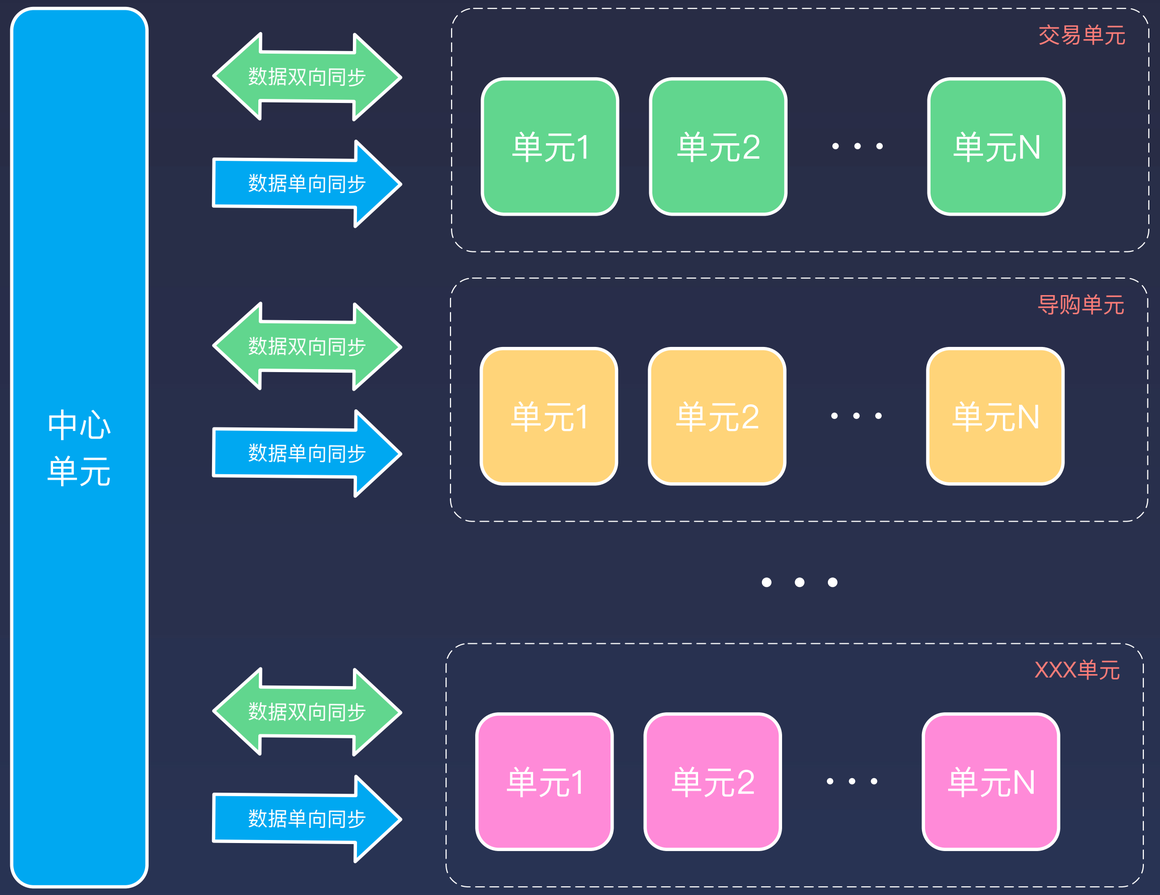
听起来好像很牛逼的样子，

这样架构升级后，容量、扩展和容灾的问题解决了么？

资源和扩展能力


中心
单元
数据双向同步
数据单向同步
交易单元
单元1
单元2
单元N
数据双向同步
数据单向同步
导购单元
单元1
单元2
单元N
数据双向同步
数据单向同步
XXX单元
单元1
单元2
单元N
建站平台
调度
物理机
ECS
...

容灾能力



- 单实例、集群层面的故障，有其它高可用手段保障
- 机房级的故障，可对同城进行秒级切换
- 异地机房或者地域的故障，可对单元秒级切换

等一会儿，你说得这么牛逼，

那你给我说说“5·27支付宝大规模宕机事故”是咋回事？

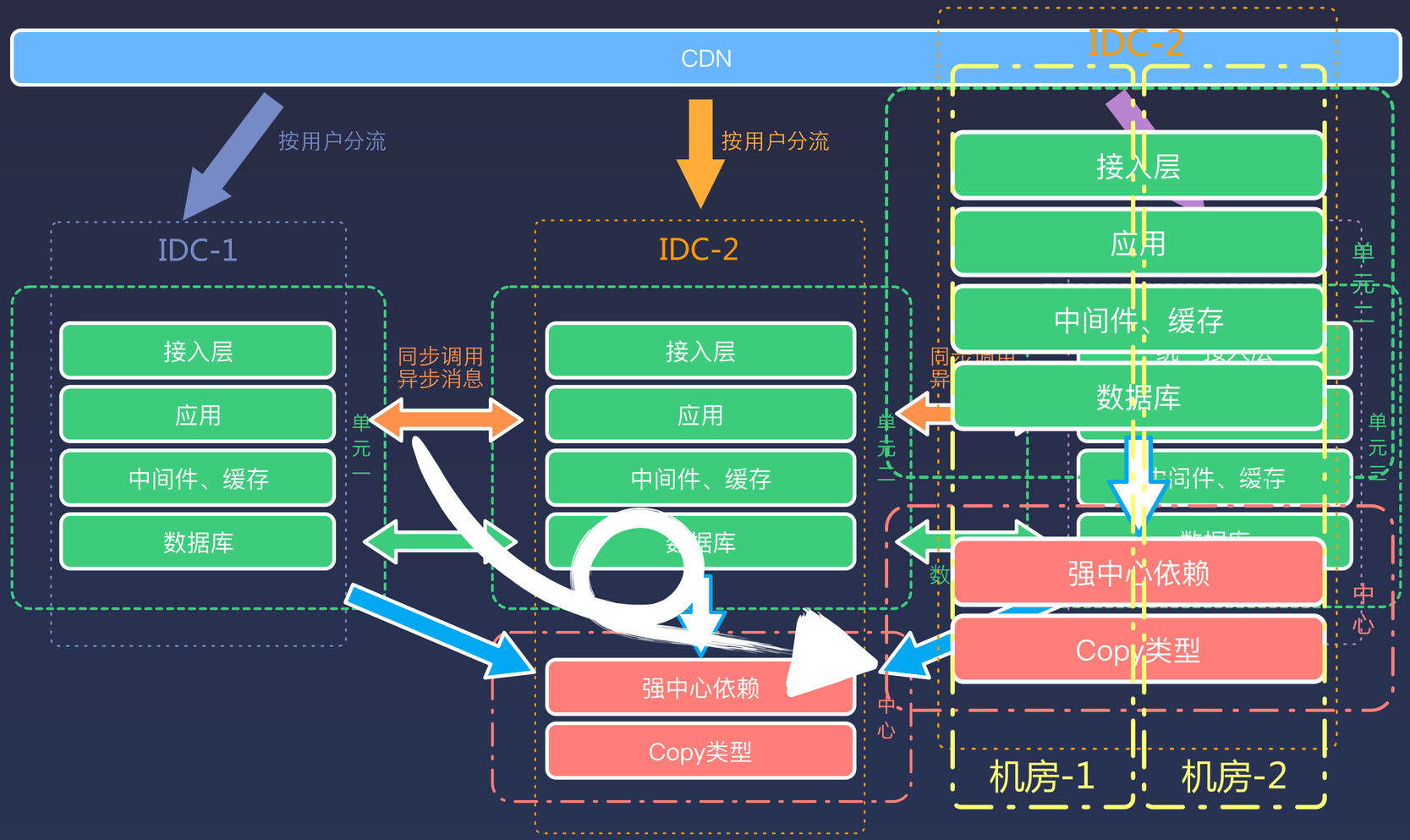
有种天鹅叫“黑天鹅”

黑天鹅事件（英文：Black swan event）指非常难以预测，且不寻常的事件，通常会引起市场连锁负面反应甚至颠覆。从次贷危机到东南亚海啸，从“泰坦尼克号”的沉没到9.11事件，瑞士央行放弃欧元兑瑞郎汇价下限后瑞郎的暴涨。黑天鹅存在于各个领域，无论金融市场、商业、经济还是个人生活，都逃不过它的控制。

在IT领域，黑天鹅主要指网络/电力问题，导致的机房大面积、长时间不可恢复，甚至极端场景，数据可靠性出现问题：

- 机房内大量或者全部服务器脱网或者掉电
- 机房间断网
- 人为失误或者程序Bug

“中心”是什么情况


CDN
按用户分流
按用户分流
IDC-1
IDC-2
IDC-2
接入层
应用
中间件、缓存
数据库
同步调用
异步消息
单元一
单元二
单元三
强中心依赖
Copy类型
中心
机房-1
机房-2

“中心”的问题

- 业务类型繁多，关系复杂
  - 实时性要求极高的业务
  - 长尾业务
- 很多业务不支持“双活”
  - 长尾业务
  - 离线任务
- 演练成本高，成功率低

轻量级故障演练

- 通过MonkeyKing平台，可以实现单实例的故障模拟；
- 故障模拟恢复平台通过DSCP（差分服务协议）打标来实现对IP五元组粒度的断网模拟；
- 通过通过故障模拟恢复平台，业务方可以自动生成自己的业务链路，可设定容灾等级并对其进行静态巡检；
- 通过故障模拟恢复平台，业务方实现自己对自己的业务进行断网模拟演练；
- 通过故障模拟恢复平台，业务方可以实现自己对自己业务的恢复操作；

Netflix的猴子家族

故障防范体系

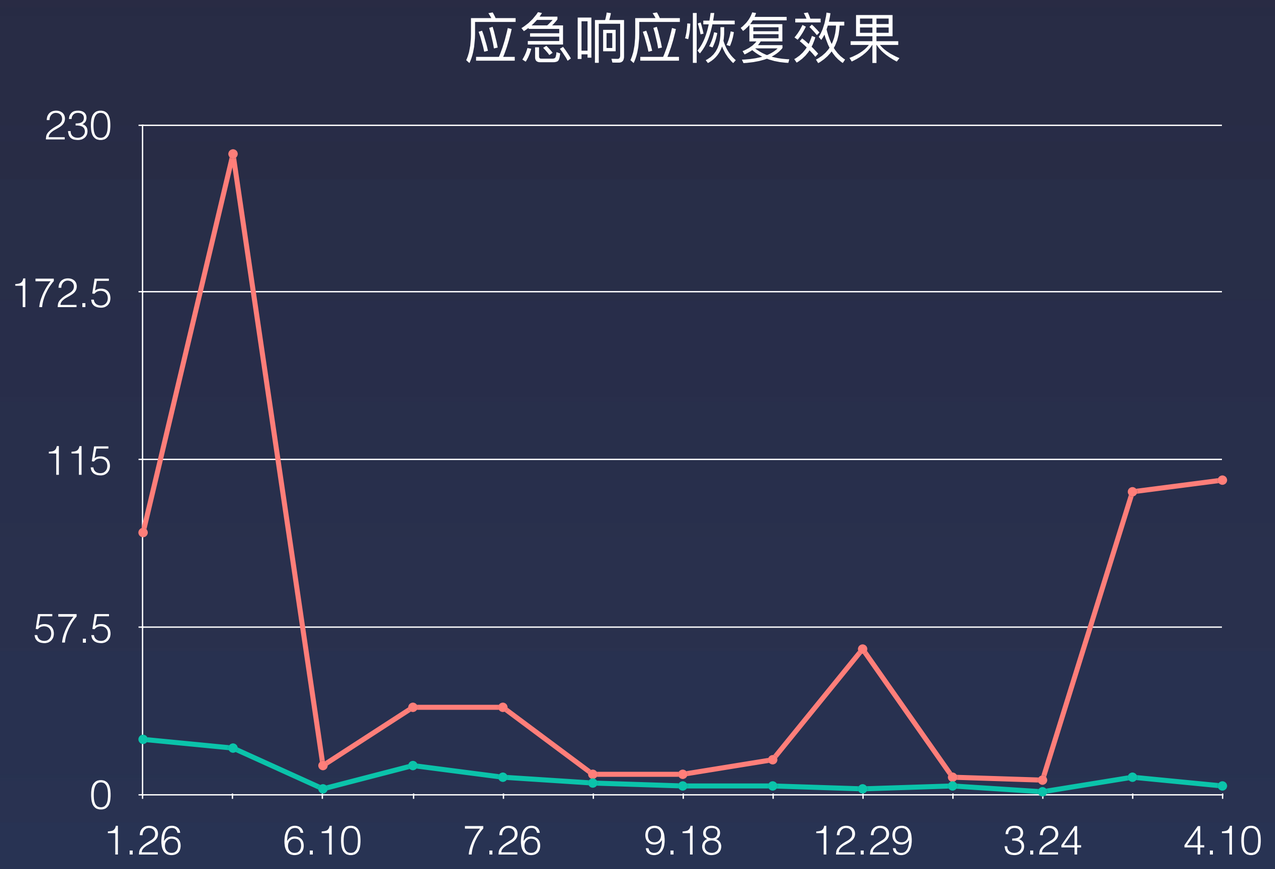
- 轻标准

1	不支持容灾
2	异地冷备
3	同城双活
4	异地多活

A	RTO <= 10min & RPO <= 10min
B	10min < RTO <= 30min & 10min < RPO <=30min
C	RTO > 30min & RPO > 30min
X	完全不具备恢复能力

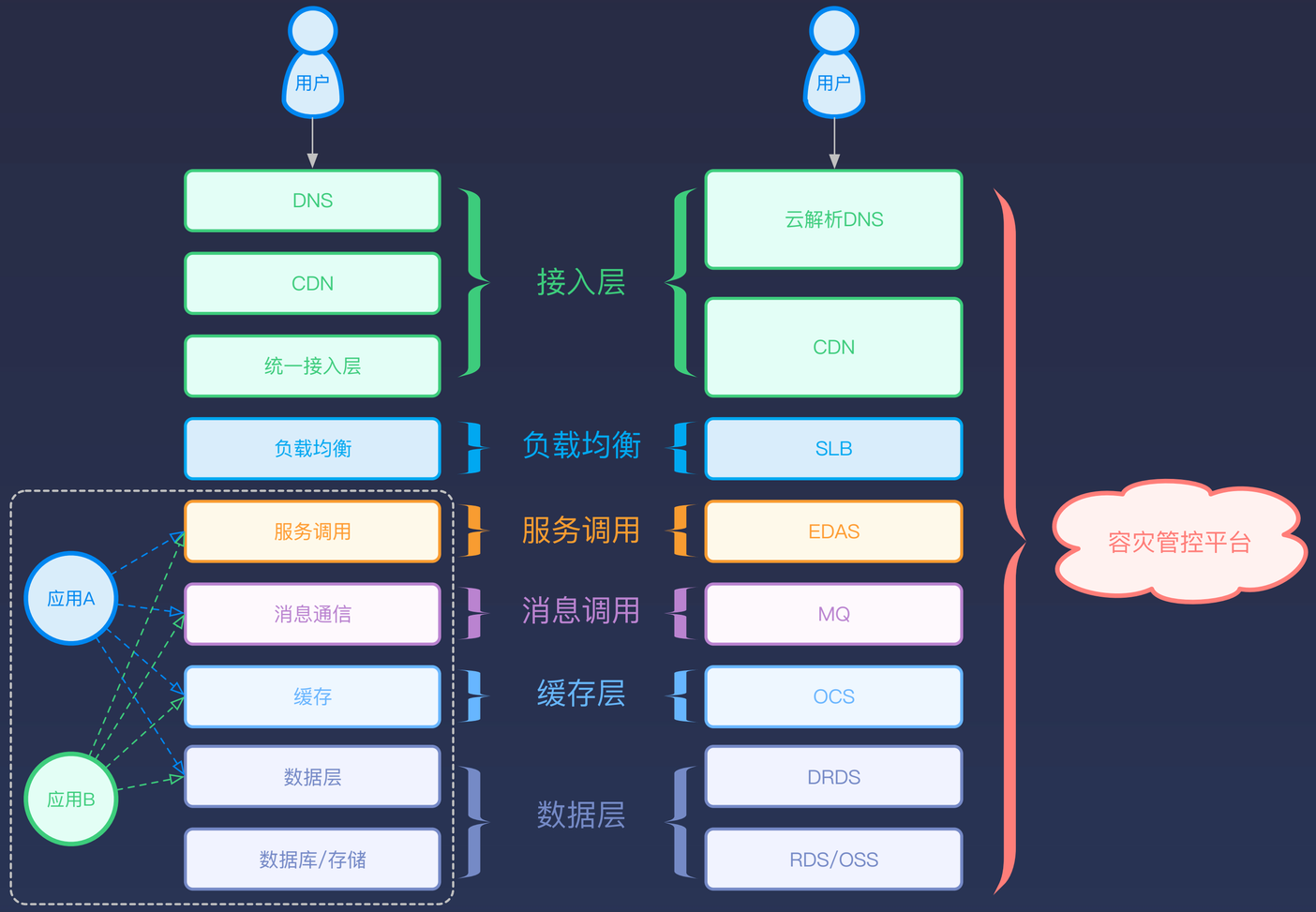
- 重管控
  - 故障快速发现、定位
  - 故障恢复平台
  - 链路自动生成、容灾等级巡检
- 常演练
  - 定期进行故障模拟、断网/断电演练、生产突袭

现阶段结果

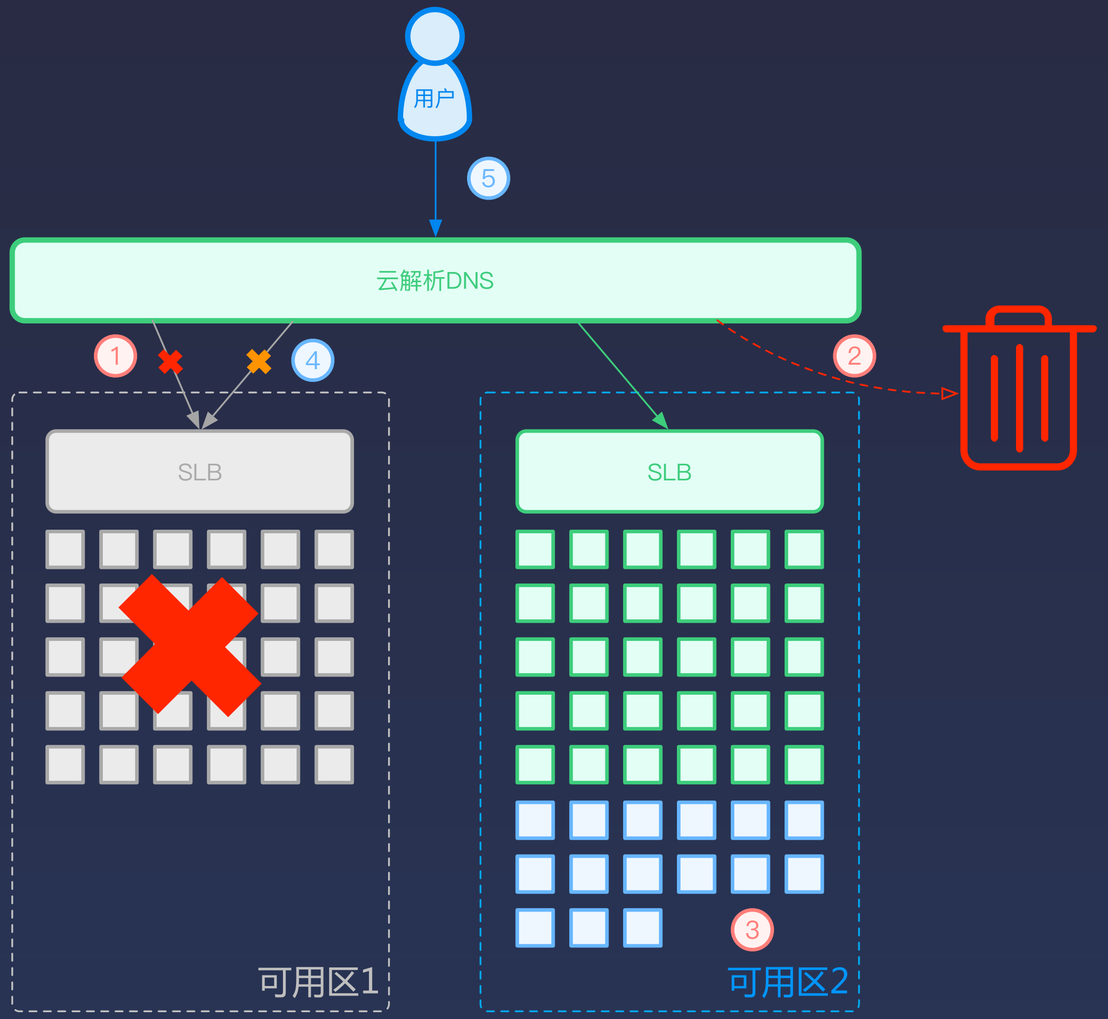


- 全年演练36次，白天10次，夜间26次
- 发现问题104个
- 业务恢复与故障恢复解耦
- 全年节省系统不可用时间591分钟

双活和多活的云端解决方案



高可用体系的生态闭环



- 容灾管控
  - 决策大盘、容灾切换、组织协助；
- 限流降级
  - 降级业务、丢弃过载流量，保证业务稳定；
- 弹性伸缩
  - 根据业务流量和机器水位动态扩容
- 故障模拟
  - 对故障改进完进行故障场景模拟，验证故障是否再现；
- 压测
  - 通过各个形式制造压测流量，来模拟真实大流量情况

联系方式 — 你懂的

欢迎技术交流

- 大规模、高并发、高可用
- 高可用、稳定性的方法策略
- 同城双活、异地多活
- 故障处理、限流降级、预案
- 监控、容量规划、调度

- 如何写简历
- 如何投简历

交流

“阿里技术”
官方微信公众号

扫一扫上面的二维码图案，关注公众号

公众号

钉钉

微信

Alibaba Group 阿里巴巴集团

THANKS

让创新技术推动社会进步

HELP TO BUILD A BETTER SOCIETY WITH
INNOVATIVE TECHNOLOGIES

Geekbang.

极客邦科技

InfoQueue

专注中高端技术人员的技术媒体

EGO EXTRA GEEKS' ORGANIZATION NETWORKS

高端技术人员学习型社交平台

StuQueue 斯达克学院

实践驱动的 IT 教育平台

地址：北京市朝阳区洛娃大厦C座8层1801室　　网址：www.geekbang.org