



QCon 全球软件开发大会
INTERNATIONAL SOFTWARE
DEVELOPMENT CONFERENCE

BEIJING 2017

百度外卖从IDC到云端服务迁移历程

SPEAKER / 赵晓燕

目录

1

※ 自我介绍

2

※ 百度外卖平滑迁移百度云的路程

3

※ 基于云的运维平台建设



自我介绍

- 大连理工大学，计算机专业研究生
- 入职百度运维部



2011.1月

2011.1月~2014.10月



- KS-OP技术负责人
- 负责百度知道、百度文库、百度旅游
- 负责社区基础架构部服务运维

- LBS-OP技术负责人
- 负责百度地图、百度外卖、百度旅游



2014.10月~2015.9月



2015.9月~至今



- 加入百度外卖
- 担任运维总监

目录

1

※ 自我介绍

2

※ 百度外卖平滑迁移百度云的历史

- 背景&挑战
- 迁移目标
- 解决方案
- 迁移效果

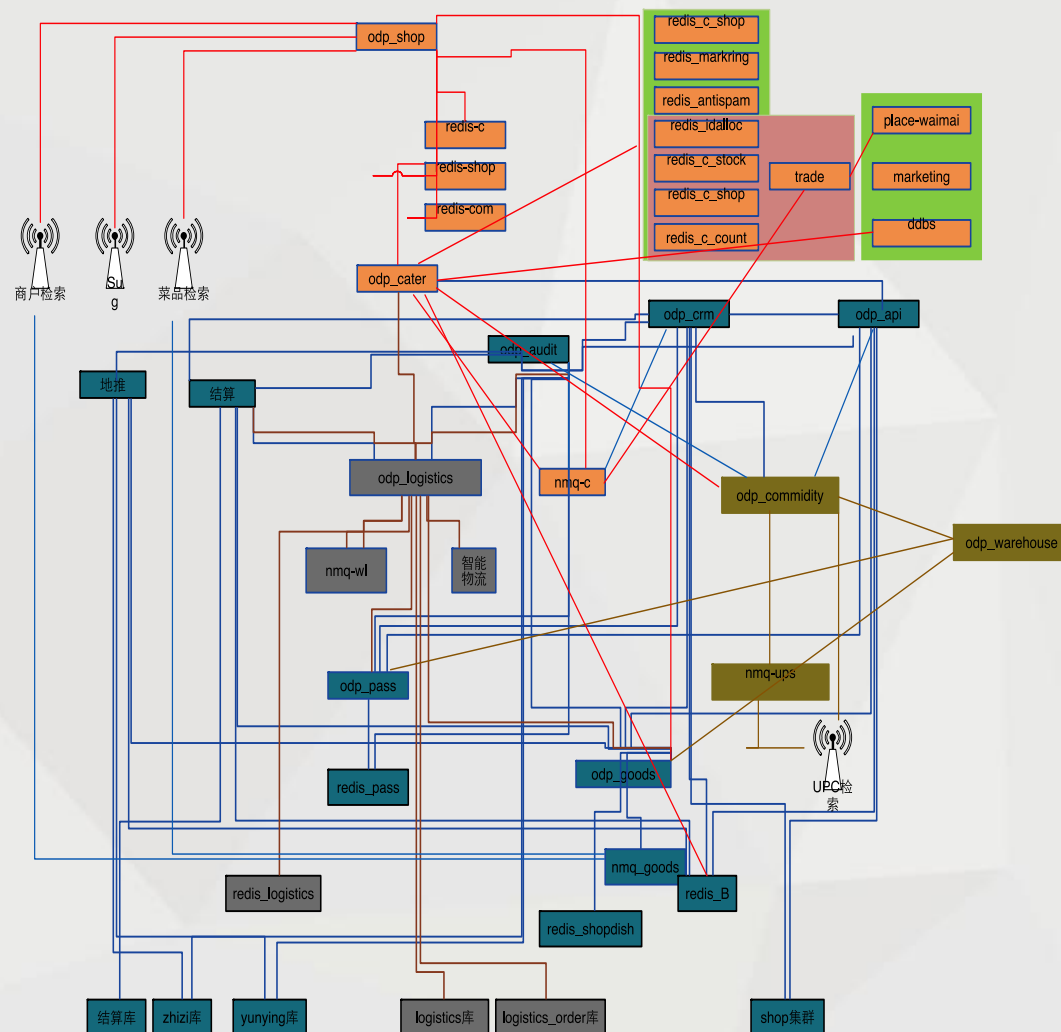
3

※ 基于云的运维平台建设

2015.9 百度外卖从百度spinoff

挑战&难点：

- 数千台服务器规模
- 数百 T mysql数据，数T Redis数据
- 业务架构复杂：用户端/商户端/物流端
- 业务抢占市场份额期，不允许服务中断

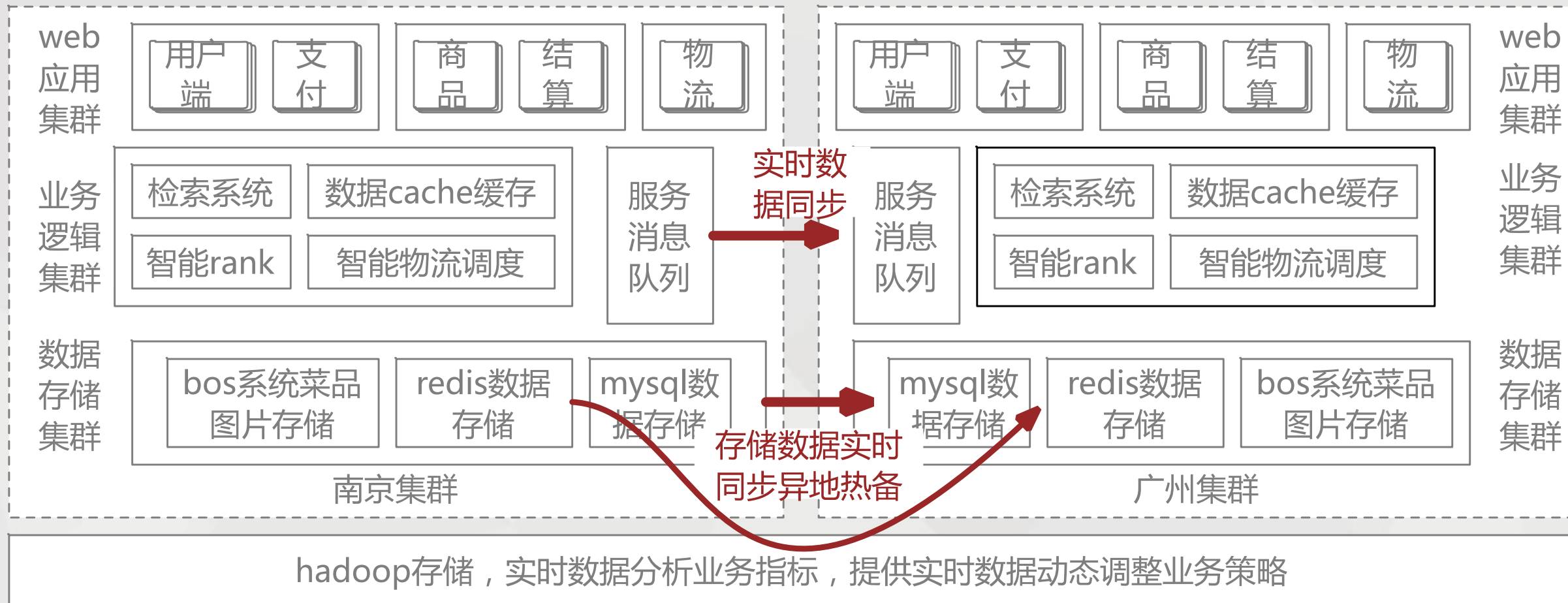




迁移百度云| 背景&挑战



百度外卖 技术团队
BAIDU WAI MAI TECHNICAL TEAM





平滑迁移、分钟级切换、业务零感知



云环境准备

- 机房选择
- 网络打通
- 云主机准备



业务架构迁移

- 架构解耦
- 业务拆迁移



数据同步迁移

- 数据级联
- 实时同步
- 降低延迟

■ 迁移百度云| 云环境准备

- **百度云机房选择**：多地域、低延迟
- **网络打通**：百度IDC与百度云打通专线，数据实时同步
- **云主机准备**：专属机，镜像安装(内核裁剪[镜像2.6.32])，puppet初始化系统参数，DNS服务



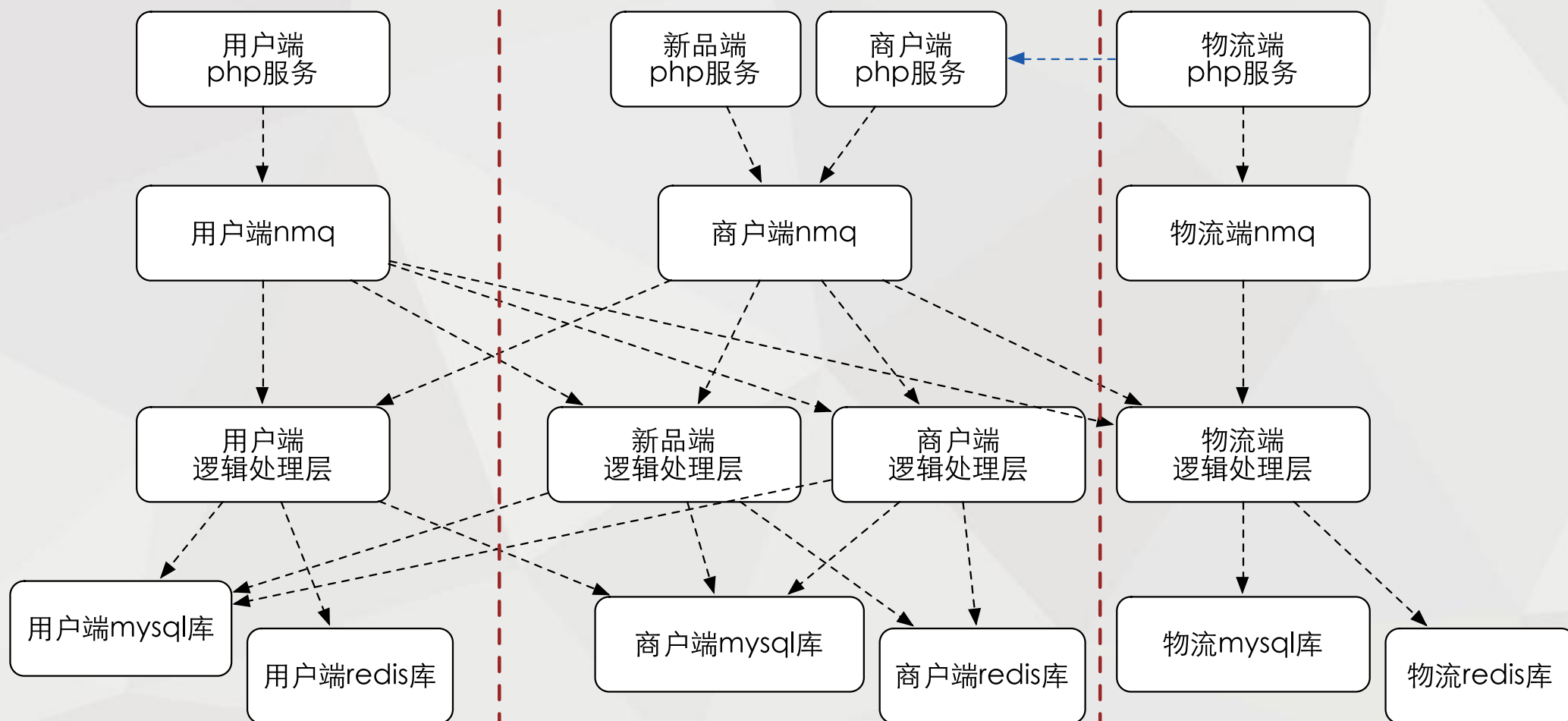


迁移百度云| 业务架构迁移



百度外卖 技术团队
BAIDU WAI MAI TECHNICAL TEAM

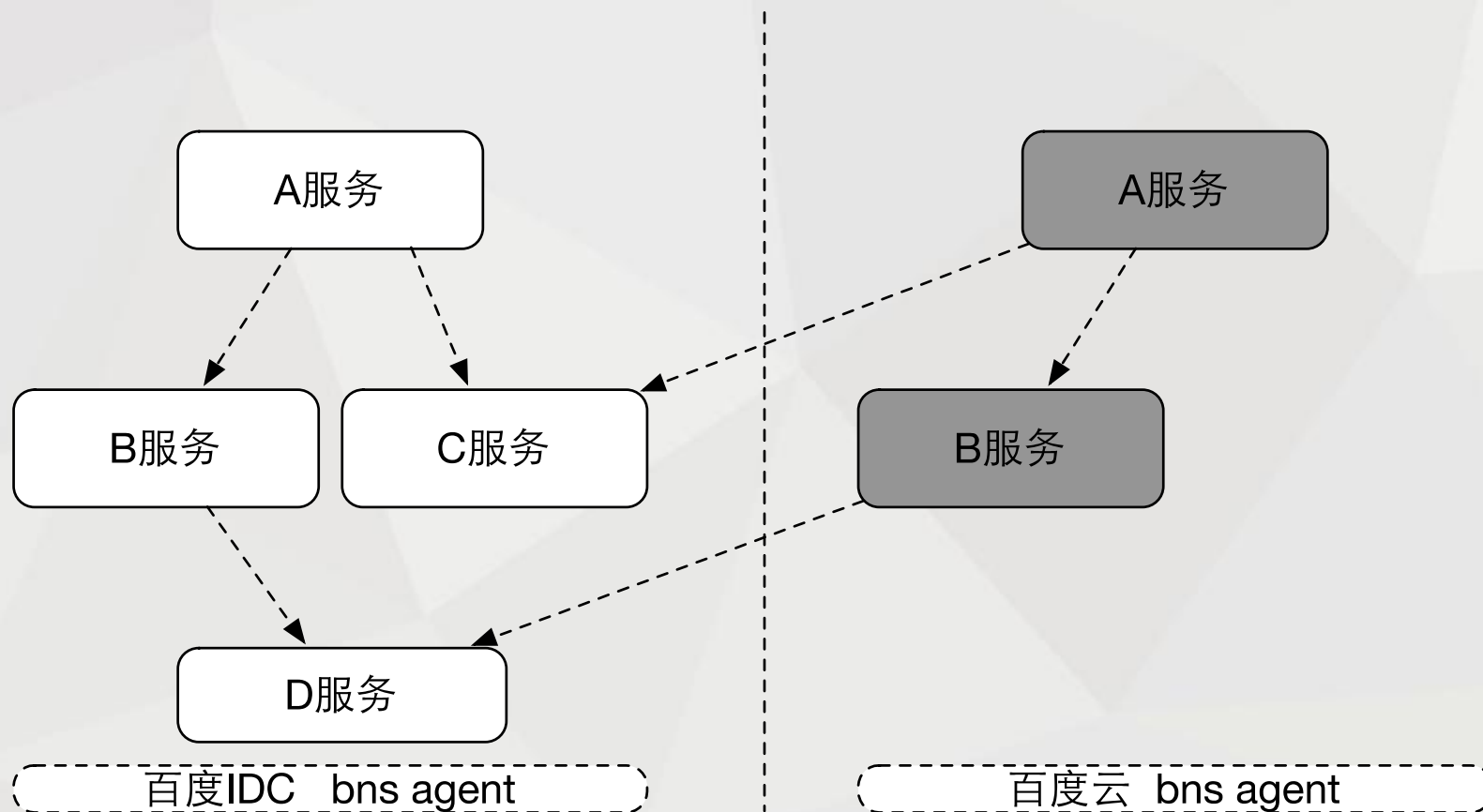
• 业务解耦：业务拆分独立迁移





迁移百度云| 业务架构迁移

- **业务解耦**：bns(Baidu Naming Service)打通，各业务拆分独立迁移
- **自顶向下**：业务迁移、数据迁移拆分，独立切换

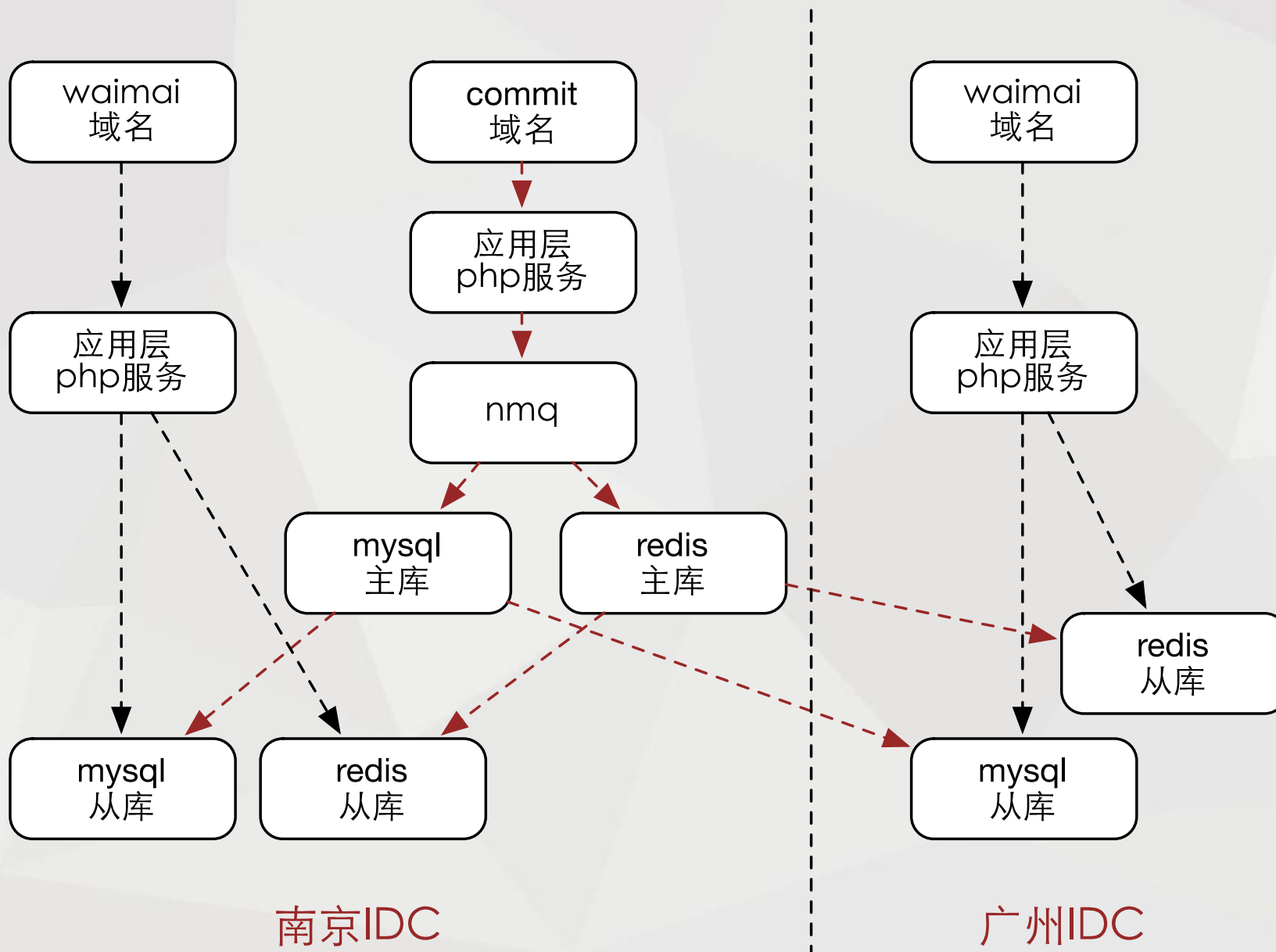




迁移百度云| 业务架构迁移



百度外卖 技术团队
BAIDU WAI MAI TECHNICAL TEAM



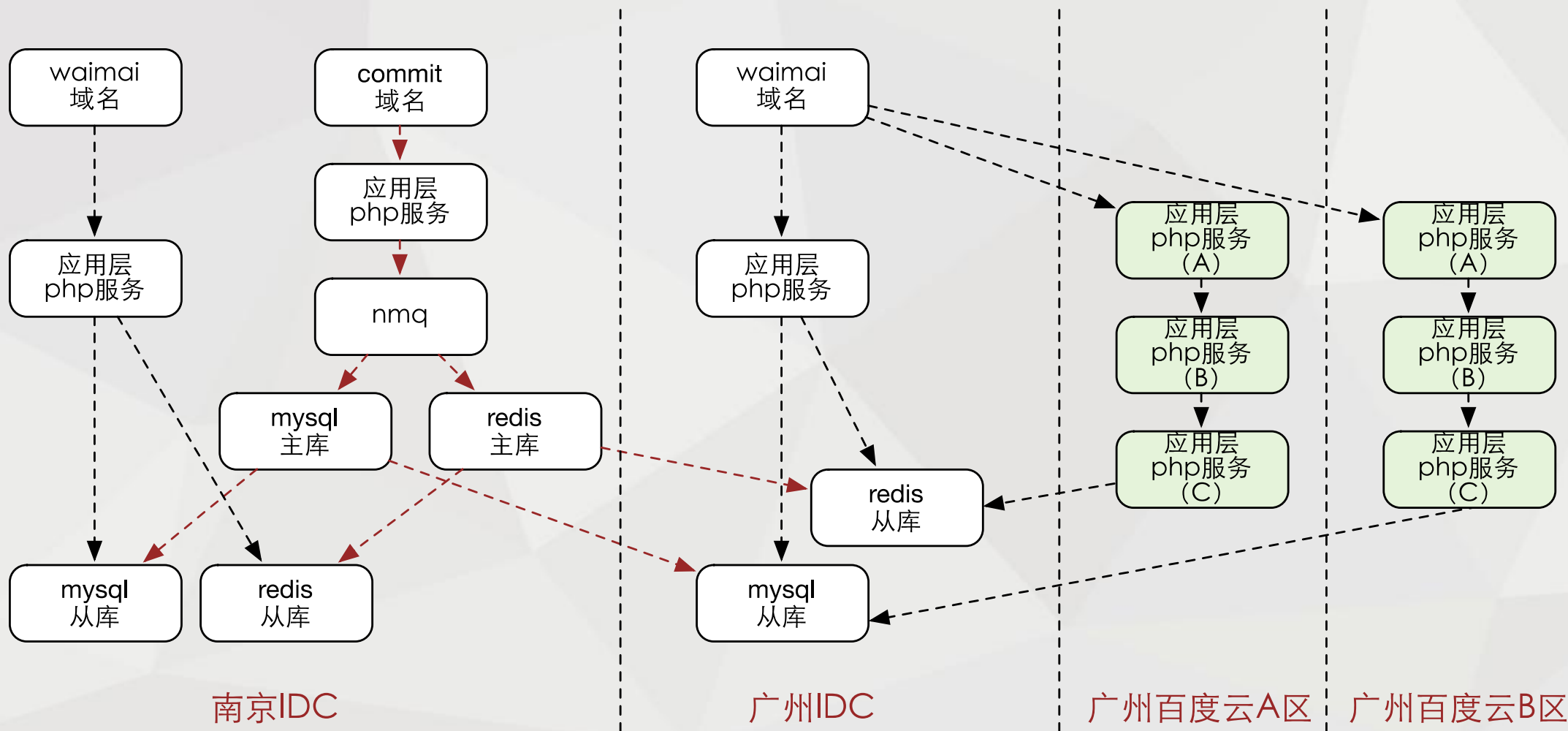


迁移百度云| 业务架构迁移



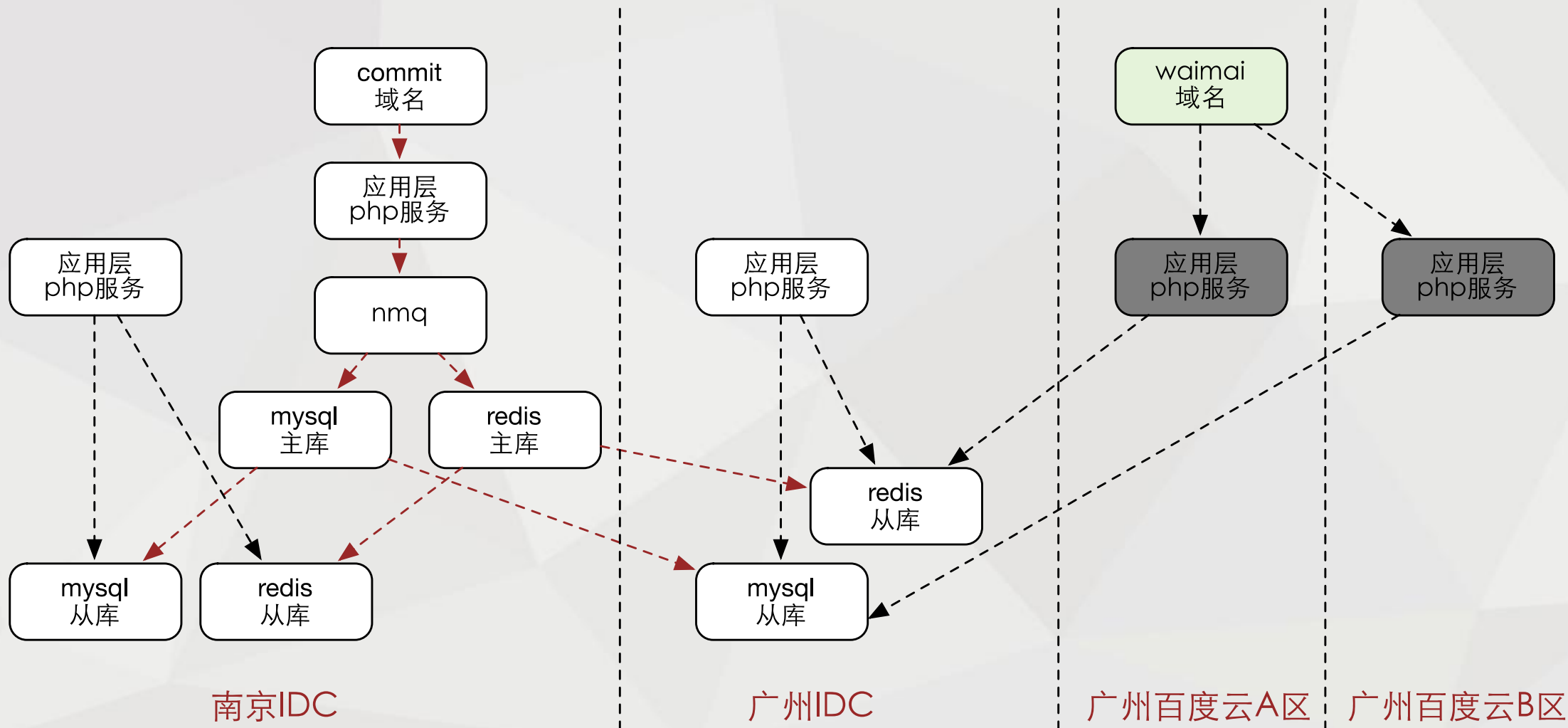
百度外卖 技术团队
BAIDU WAI MAI TECHNICAL TEAM

- **业务调用原则**：避免跨region访问，垂直关联
- **业务改造点**：WODP版本迭代，代理改造，bos基础数据迁移，cdn迁移



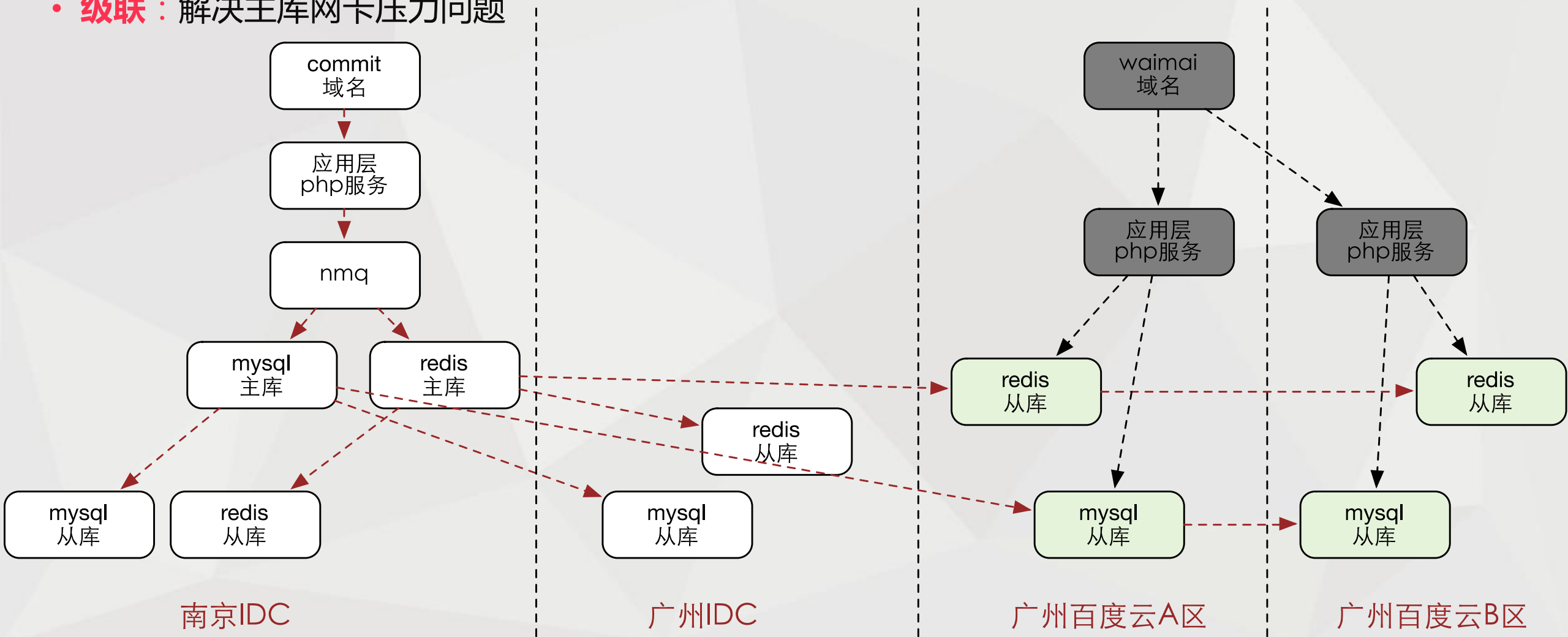
迁移百度云| 业务架构迁移

- 读全流量切至百度云，保留IDC应用层服务，做好回滚预案



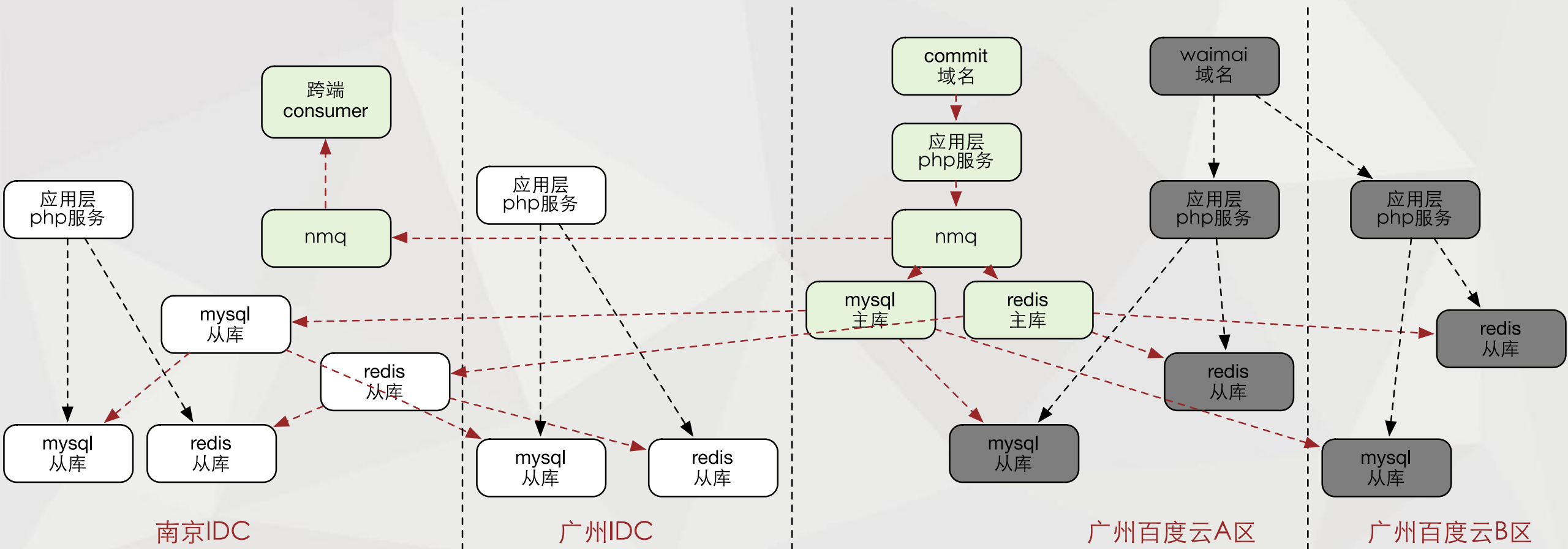
迁移百度云| 数据同步迁移

- **存储层**：读写分离，读写分步迁移
- **级联**：解决主库网卡压力问题



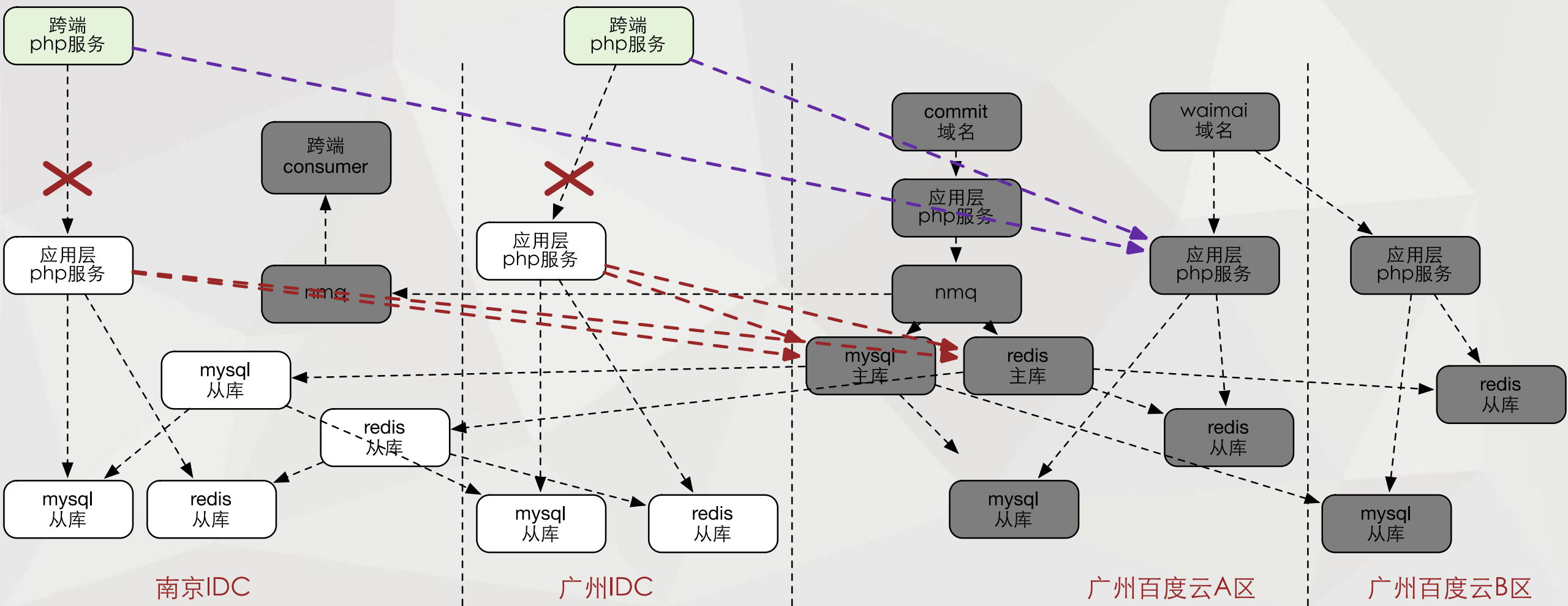
迁移百度云| 数据同步迁移

- **写流量迁移**：避免跨地域写；
- **操作顺序**：redis -> mysql -> nmq -> commit域名



■ 迁移百度云| 数据同步迁移

- **跨端依赖**：减少跨地域请求，降低服务性能损失
- 一次跨地域？多次跨地域？





迁移百度云| 迁移效果



百度外卖 技术团队
BAIDU WAI WAI TECHNICAL TEAM

- 历经7个月时间
- 外卖自研21个系统
- 方案打磨近10次
- 百度云支持解决了47个问题
- 平滑、无损、对业务无感知

【redis操作】

耗时：60s左右

- Waimai-logistic:
 - 第一个分片：持久化61秒，同步用时21分钟，业务零感知，无报错；
 - 第二个分片：持久化58秒，同步用时19分钟，业务零感知，无报错；
 - 第三个分片：持久化68秒，同步用时21分钟，业务零感知，无报错；
 - 第四个分片：持久化58秒，同步用时20分钟，业务零感知，无报错；
- Waimai-logistic-cex:
 - 一个分片：持久化2秒，同步用时2分钟；业务零感知，无报错；

【C端修改连物流bns，由IDC修改为公有云】

23:20开始操作，无异常；

【mysql操作】

耗时：30s左右

- logistic-order：主从切换时间30秒；业务零感知，符合预期
- logistic：主从切换时间26秒；业务零感知，符合预期
- DBA修改IDC上dbproxy的配置，把原有IDC的mysql 挂为公有云mysql的从库：00:07操作完毕；业务零感知，符合预期
- DBA切换mysql 的rdview库到公有云：00:07操作完毕，业务零感知，符合预期

【智能物流切换访问基础物流的请求bns；DS->WL】

23:55开始操作，00:05 操作完毕；服务回归正常，符合预期

【智能物流停nj的服务，启动公有云gzns的服务】

- 自动调度：00:19开始操作；00:21第一台智能物流服务启动，开始智能调度；00:23全部操作完毕
- 众包：00:25 开始操作；00:28 操作完毕；

【物流切换dstask、dsproxy、cs bns到公有云】

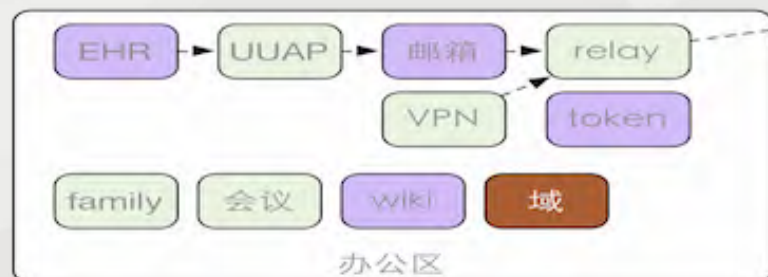
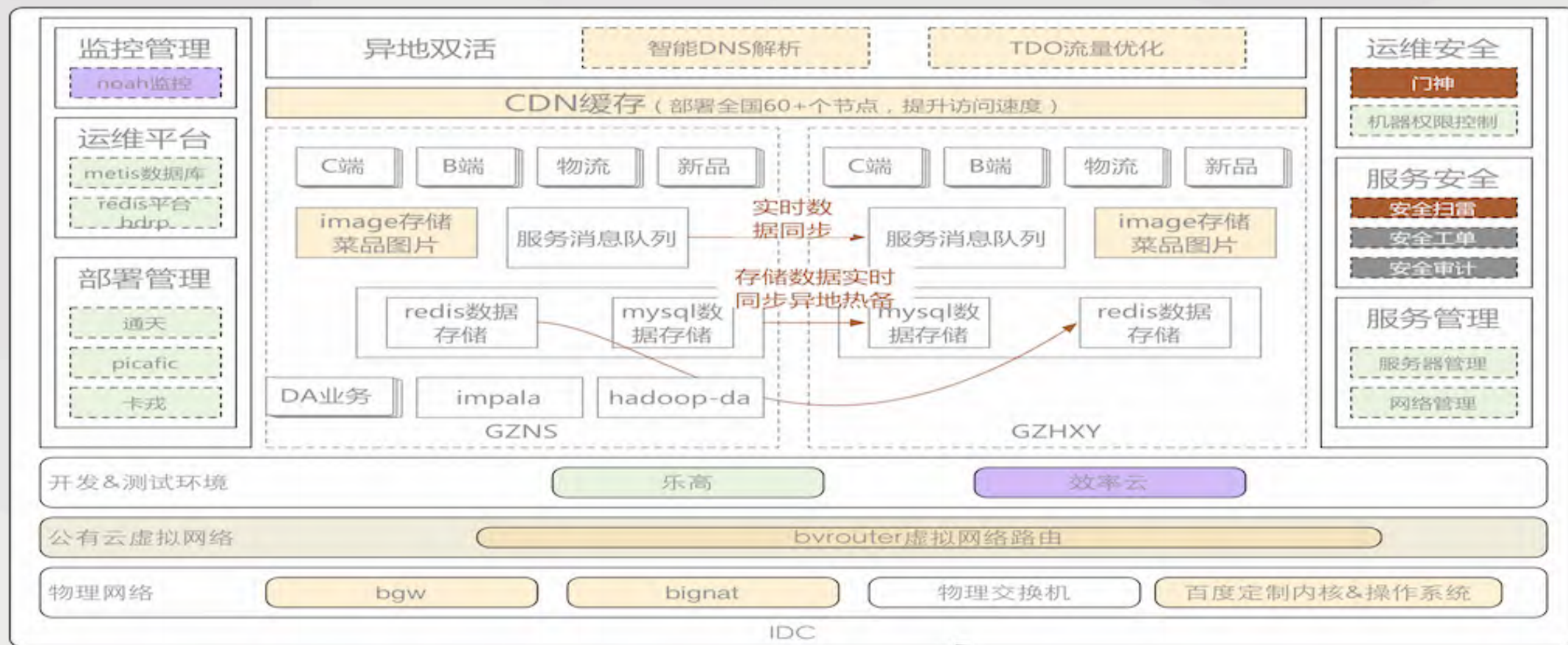
00:31开始操作，00:33操作完毕；



迁移百度云| 迁移效果



百度外卖 技术团队
BAIDU WAI WAI TECHNICAL TEAM



- 外卖自建系统 (Waimai self-built system)
- 公有云服务 (Public cloud service)
- 外卖缺失功能 (Waimai missing functions)
- 百度继续提供支持 (Baidu continues to provide support)
- 购买第三方服务 (Purchase third-party services)

目录

1

※ 自我介绍

2

※ 百度外卖平滑迁移百度云的过程

3

※ 基于云的运维平台建设

- 网络层
- 操作系统层
- 应用层：围绕业务生命周期
- 正在做...

Spinoff

背靠大山，白手起家

从法拉利到拖拉机？

UUAP	uu.baidu.com	编译发布	sc.baidu.com
VPN	v.baidu.com	代码提交	svn.baidu.com
邮箱	mail.baidu.com	服务器登陆	rel.baidu.com
域		开发机申请	qian.baidu.com
ERP	e.baidu.com	WIKI	wiki.baidu.com
入职		开发立项	cafe.baidu.com

PAAS平台			oxxp.baidu.com
监控			noah.baidu.com
域名			rmss.baidu.com
BNS			noah.baidu.com
VIP			rmss.baidu.com
代码上线			speed.baidu.com
服务安装			各种平台服务,bdrp等
系统初始化			forman.baidu.com
操作系统安装			
服务器硬件			rmss.baidu.com
GZ	NJ	SZWG	
IDC环境			



PS:域名进行了更换

运维平台| 背景

近1年时间 6个人

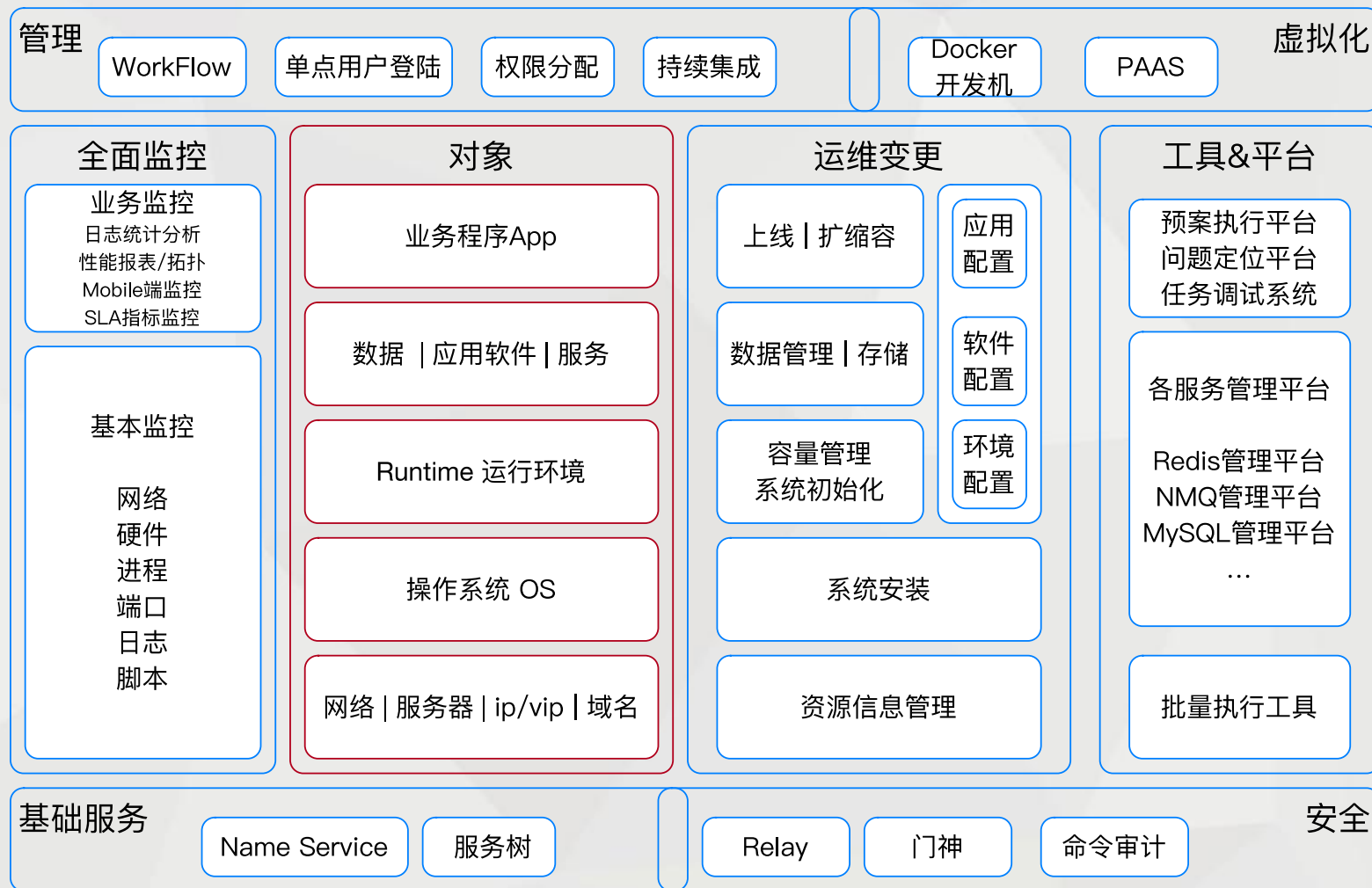
两大主体平台: **Pacific Atlantic**

业务平台 **11**个

通用服务 **6**个

Go lib库 **11+**

对外接服务 **30+**



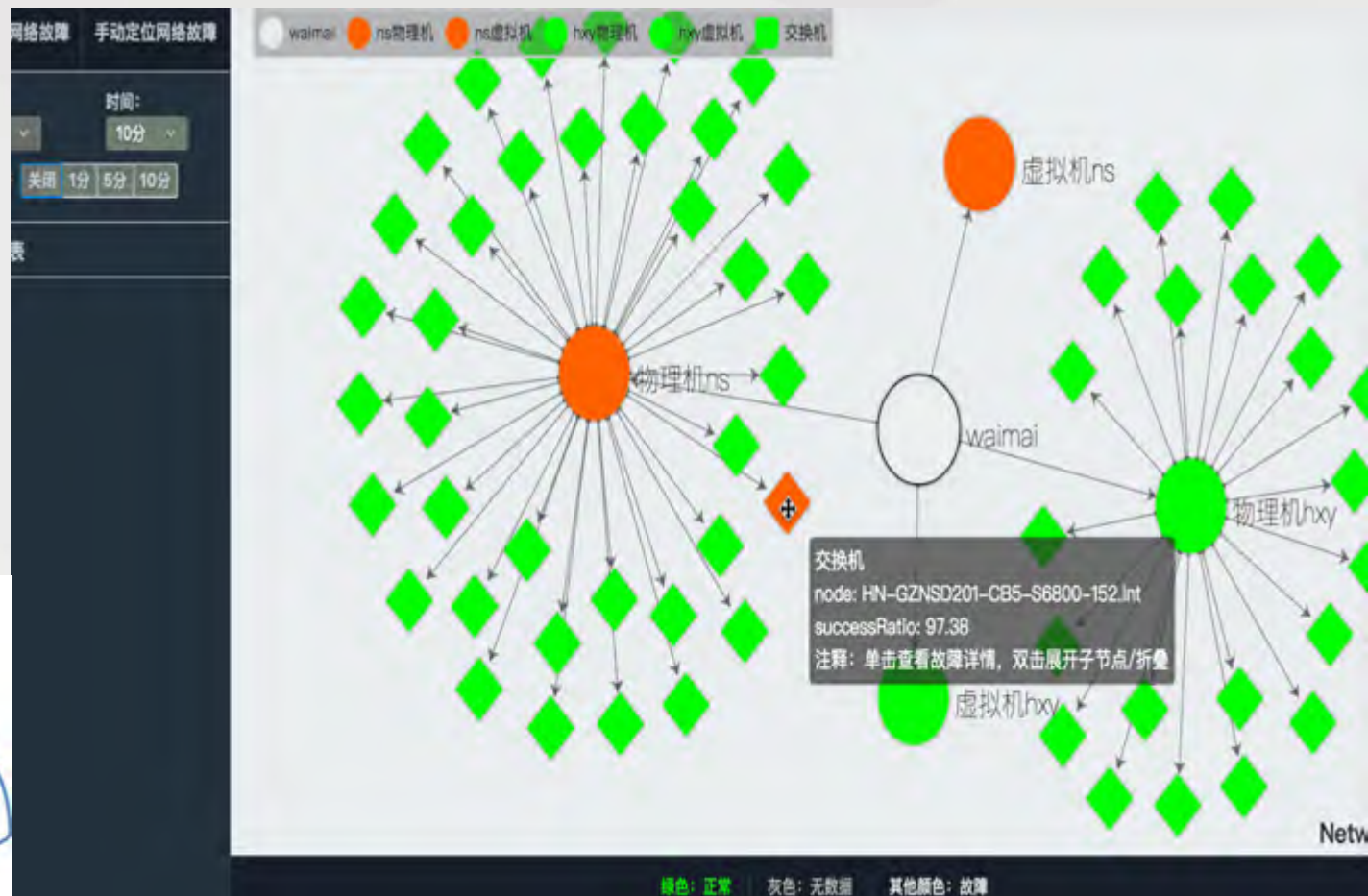
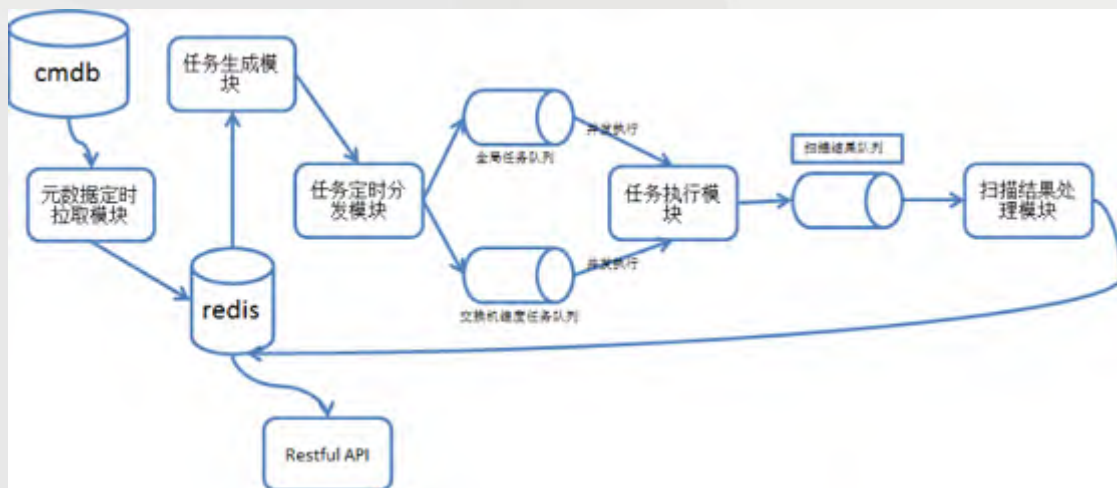


运维平台| 网络层



百度外卖 技术团队
BAIDU WAI MAI TECHNICAL TEAM

- **内网网络监控：**
- cmdb的agent里集成nmap的sdk
- 探测千台服务器耗时为3s
- Redis做存储，过期存储
- 屏蔽网络故障对检测服务的影响



运维平台| 网络层

- **外网网络监控**：公有云环境到各个城市ISP节点的链路质量监测
- 与apm监控互补，apm的数据在网络出现问题时不能上报





操作系统优化：

- cpu优化：关闭cpu降频，打开超线程
- 内存优化：刷脏页的频率优化、**transparent_hugepage**
- TCP/IP优化：SYN 连接请求、TIME_WAIT 套接字的最大数、最大的监听队列的长度、关闭ipv6等等
- SSD盘参数优化：磁盘调度策略、随机熵池

内核裁剪[镜像2.6.32]：

- CONFIG_NO_HZ 关闭
- CONFIG_SPARSE_IRQ 开启
- CONFIG_TRANSPARENT_HUGEPAGE 关闭
- CONFIG_KEXEC CONFIG_KEXEC_AUTO_RESERVE CONFIG_CRASH_DUMP 开启
- CONFIG_CPU_FREQ_DEFAULT_GOV_PERFORMANCE 开启
- CONFIG_KSM 关闭
- CONFIG_COMPACTION 关闭
- CONFIG_MIGRATION 关闭

- **核心原则：初创团队**
- 围绕业务发展
- 支撑业务核心需求
- 深入各个环节细化



- **理念**：运维标准化、运维效率
- **历程**：百度云做镜像->puppet初始化(系统参数优化/常用工具)->blb/eip工具集->硬件故障检测->CMDB

服务树

pacific

▼

▼ wamr-out

▼ wamr-B

▼ wamr-C

▼ online

▶ cater

▶ nodeui

▶ shopui

▶ platform

▶ trevi

▶ index

▶ yunying

▶ cart

▶ c-static

▼ wamr-D

▼ wamr-E

▼ wamr-F

配置中心

上线系统

预案平台

工单系统

CMDB

Workflow

UDC

报警分析

帮助中心

zhangjian02

报表

服务器资源

EIP

BLB

套餐信息

输入 主机名/IP 后，点击快速查询按钮

快速查询

设置字段

批量查询

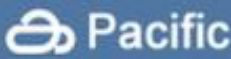
基本信息：☒ 主机名 ☒ 内网IP ☐ 外网IP ☐ Imageid ☒ IDC ☐ UUID ☒ 状态 ☐ 类型 ☐ 机架

配置信息：☐ 机器型号 ☒ OS ☐ Platform ☐ Gccversion ☐ CPU型号 ☐ CPU逻辑数量 ☒ CPU物理数量 ☒ OS内核 ☐ 内存厂商
☐ 内存大小 ☐ Disk厂商 ☐ Disk大小 ☐ 网卡一 ☐ RAID

其他信息：☒ 入库时间

序号	主机名	内网IP	IDC	状态	OS	CPU物理数量	OS内核	入库时间
1	genc-vmr-wamr-pay-ec2-genc-vmr-wamr	10.10.10.10						2016-12-27T07:29:08Z
2								2016-12-27T07:40:25Z
3								2016-12-27T07:40:25Z
4								2016-12-27T07:40:26Z

运维平台| 上线部署

 Pacific

配置中心

上线系统

预案平台

工单系统

CMDB

WorkFlow

UDC

报警分析

帮助中心

© zhangjian02

上线步骤	状态	操作
 打包	成功	详情 重做
 任务提交	成功	上线 重做
 mirror	成功	详情 重做
 暂停点	成功	详情 继续 重做
 全量	成功	详情 继续 重做

Pacific上线系统

“ 外卖自己的上线系统
安全 稳定 贴心交互 ”

部署系统：满足需求->精雕细琢

★ 抓住当前痛点,优先解决1-2个问题.

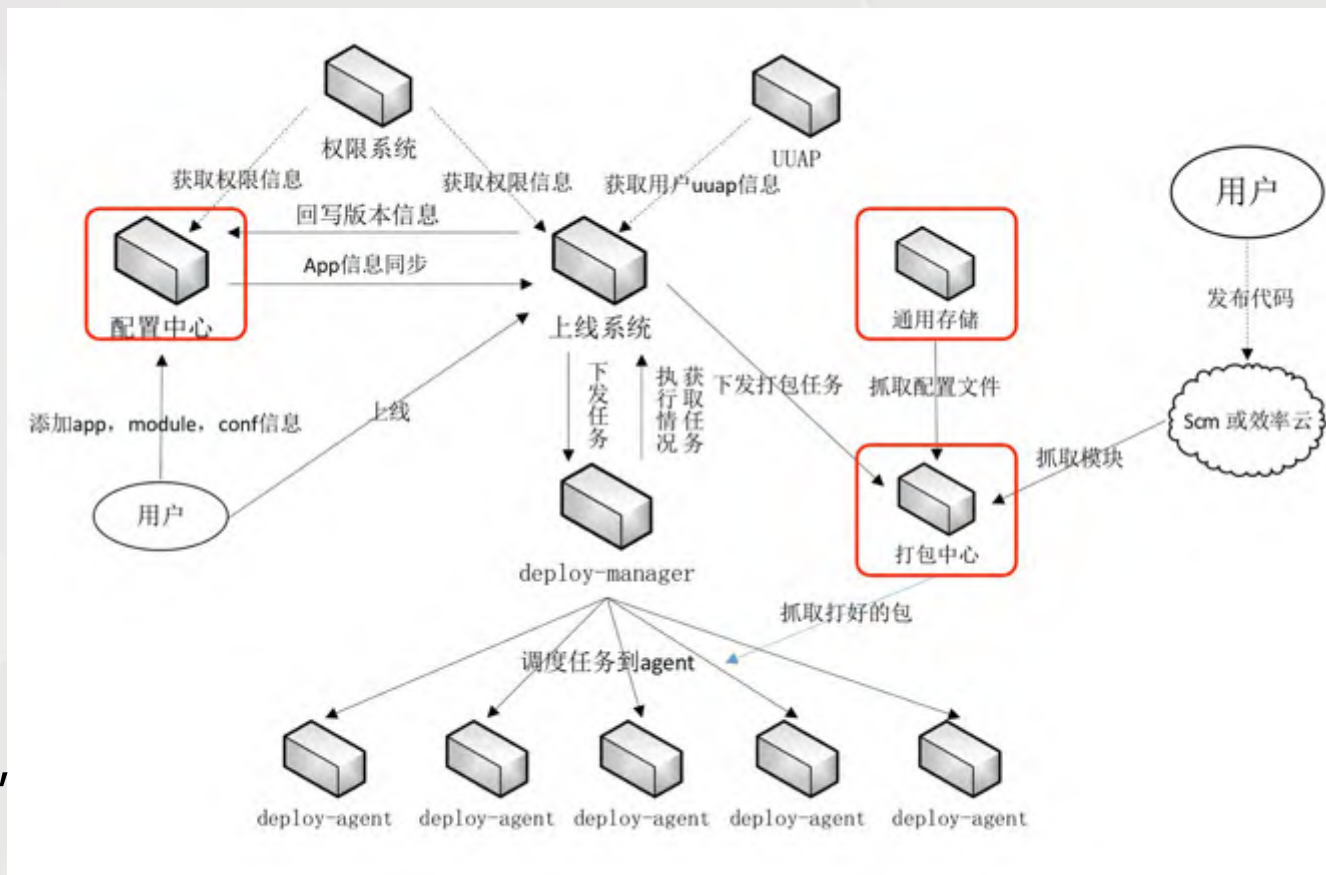
- 上线系统1.0 完成最简单的单模块上线

★ 关键时机,大胆重构

- 上线系统2.0 — 架构转折性变化.
- 执行效率从半小时以上降低到分钟级别.
- 成功率从89%提升到99%

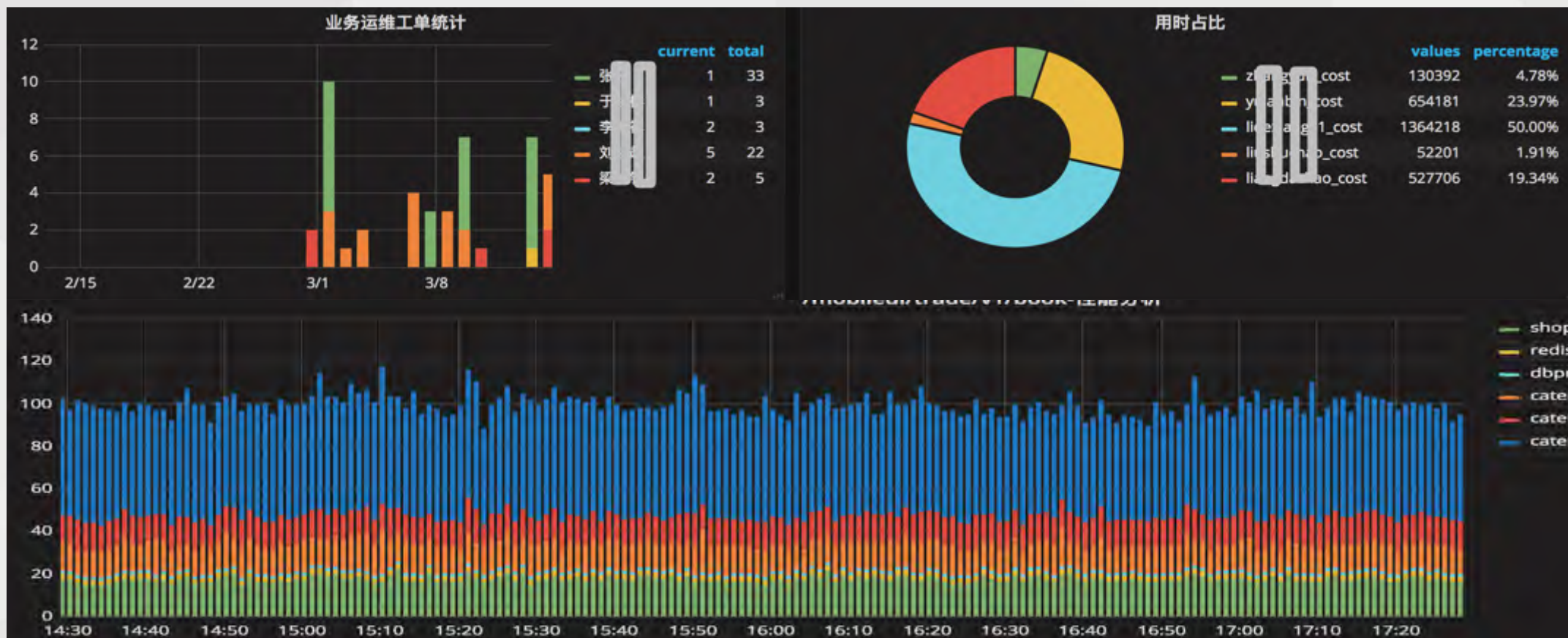
★ 精雕细琢,提升速度和功能完善

- 速度: 上线系统3.0 — 秒级打包,秒级上线,回滚优化,对接效率云等.
- 功能: 上线系统4.0 — diff 代码,配置派生等



运维平台| 正常运行

- **Atlantic平台**：系统性思维、挖掘需求的通用性、开源与创新
- **内容**：容量资源、工单统计、网络耗时、业务SLI指标等



Atlantic平台： open-falcon 及 grafana 的应用

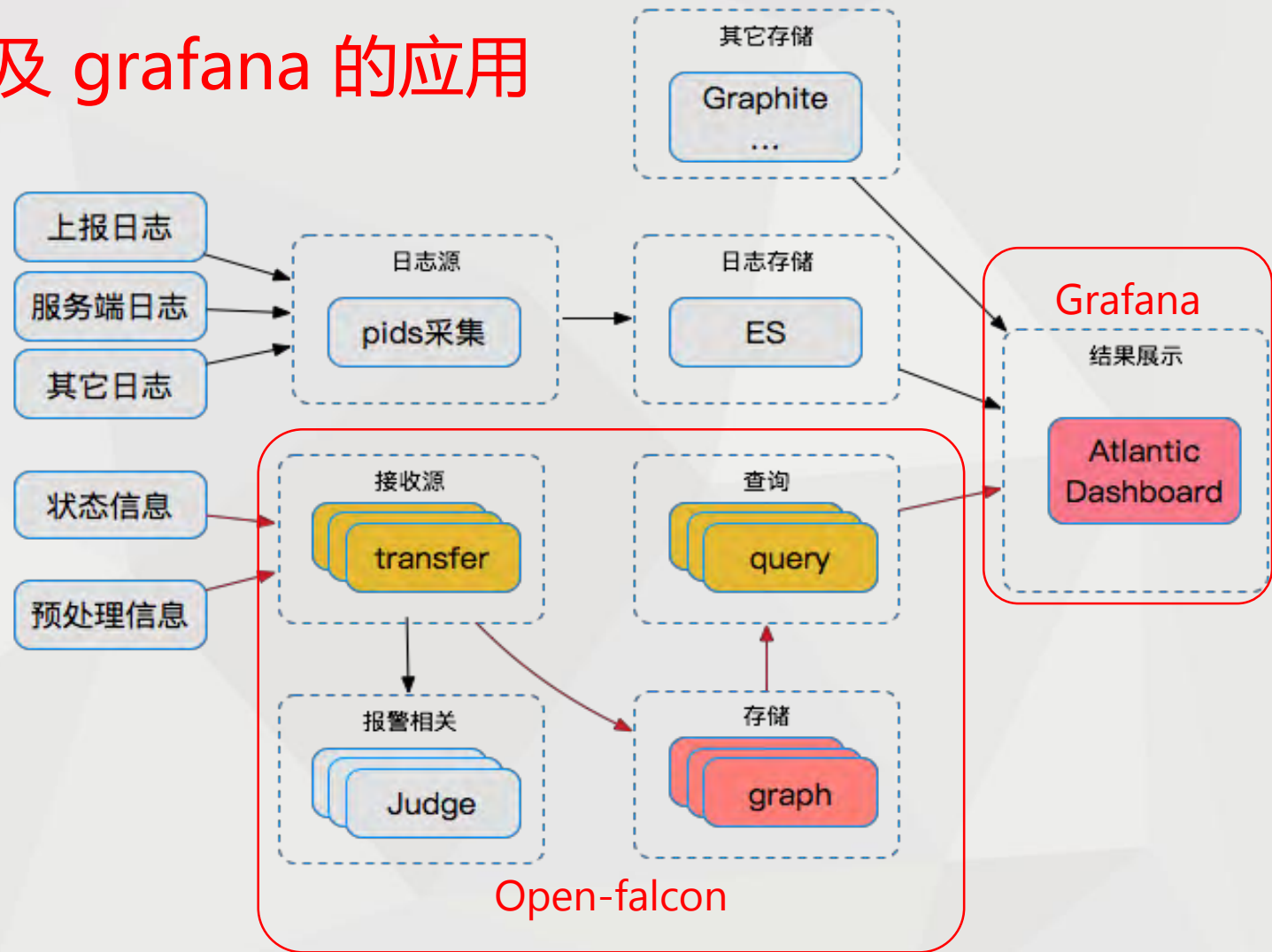
数据收集 — 业务方推送给open-falcon

数据存储 — graph 和 opentsdb等

数据展示 — grafana定制dashboard需求

数据报警 — open-falcon

- 修改graph query模块，增加多条件查询
- 增加同比、环比等函数功能.
- 增加求和计算、百分比计算等函数功能
- 增加和修复图表功能，如趋势图、中国地图等.
- 对接wm-uuap 权限等.



- **预案执行系统**：实现配置降级，已录入36个业务降级；可以实现分钟级别降级恢复
- 底层调用上线部署系统，进行降级配置的替换。

Pacific

配置中心 上线系统 预案平台 工单系统 CMDB WorkFlow UDC 报警分析 网络定位

帮助中心 zhaoxiaoyan

服务树

pacific

waimai-test

waimai-B

waimai-C

waimai-BI

waimai-DA

waimai-WL

waimai-INF

waimai-platform

waimai-nmq

未选app 全部

配置降级 服务降级 物理降级 自动预案

ID	名称	二级分类	APP名称	业务空间	前置命令	后置命令	并		
36	[预案]C端运营人文关怀入口降级	接口配置	cater	waimai-C			1	配置详情	操作 删除
35	[预案]C端运营专题入口降级	接口配置	cater	waimai-C			1	配置详情	操作 删除
34	[预案]C端运营开机屏广告入口降级	接口配置	cater	waimai-C			1	配置详情	操作 删除

刷新

执行 执行

历史上线单号:13818

恢复 恢复

历史回滚单号:13836



工单系统：支持可定制工单模板，支持移动审批



数据库操作：乐高+metis平台

- sql提交、sql审核、sql操作等

选择要修改的数据库及表

数据库: 表名:

集群: 业务方向:

子方向:

变更详情

原因说明:

项目高峰时间:

期望上线时间:

是否需要数据备份: ☐

是否需要RD回归测试: ☐

请输入需要评审的SQL:

需要评审的内容

TaskID	状态	请求者	审批者	集群	DB	Table	SQL	限时	操作
815	Done	m	jia	sa	waimu	p_order_feed_20170322	ALTER TABLE `p_order_feed_20170322` ADD INDEX idx_order_id (`order_id`);	0	已通过
814	Done	m	jia	sa	waimu	p_order_feed_20170315	ALTER TABLE `p_order_feed_20170315` ADD INDEX idx_order_id (`order_id`);	0	已通过
813	Done	m	jia	sa	waimu	openapi_order_monitor	CREATE TABLE `openapi_order_monitor` (`id` bigint(20) unsigned NOT NULL AUTO_INCREMENT COMMENT 'ID', `source_name` varchar(32) NOT NULL DEFAULT '' COMMENT '账号', `log_time` bigint(20) unsigned NOT NULL DEFAULT '0' COMMENT '记录时间', `create_time` int(10) unsigned NOT NULL DEFAULT '0' COMMENT '创建时间', `amount` int(4) unsigned NOT NULL DEFAULT '0' COMMENT '订单量', `ext` text NOT NULL COMMENT '扩展', PRIMARY KEY (`id`), KEY `index_source_name` (`source_name`), KEY `index_log_time` (`log_time`)) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='api订单流量监控';	0	已通过
812	Done	w	jia	sa	waimu	p_statement_order_20170315	CREATE TABLE `p_statement_order_20170315` (`id` bigint(20) NOT NULL AUTO_INCREMENT COMMENT '主键', `trade_code` varchar(64) NOT NULL COMMENT '交易流水号', `order_id` bigint(20) unsigned NOT NULL COMMENT '订单号', `account_id` bigint(20) unsigned NOT NULL COMMENT '结算账户id', `wid` bigint(20) unsigned NOT NULL DEFAULT '0' COMMENT '商户id', `statement_id` varchar(32) NOT NULL DEFAULT '' COMMENT '对账单唯一id', `state_time` int(10) unsigned NOT NULL DEFAULT '0' COMMENT '记录属于的账期',	0	已通过

2017年3月27日 14:52



Pacific_Robot

任务[855]--开始时间[2017-03-27 14:52:12]--结束时间[2017-03-27 14:52:28]--用时[16秒]--成功



Pacific_Robot

任务[853]--开始时间[2017-03-27 14:53:07]--结束时间[2017-03-27 14:53:23]--用时[16秒]--成功

运维平台| 正在做...



运维平台| 正在做...





关注QCon微信公众号，
获得更多干货！

Thanks!



INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

主办方 **Geekbang** **InfoQ**
极客邦科技