

China · Beijing

京东核心中间件是如何支撑业务快速发展

何小锋

中间件负责人



[北京站]

主办方 **Geekbang** & **InfoQ**
极客邦科技



促进软件开发领域知识与创新的传播



关注InfoQ官方微信
及时获取ArchSummit
大会演讲视频信息



全球软件开发大会 [北京站]

2017年4月16-18日 北京·国家会议中心

咨询热线: 010-64738142



全球架构师峰会 2016 [深圳站]

2017年7月7-8日 深圳·华侨城洲际酒店

咨询热线: 010-89880682

大纲

邦之利器，今可示人

——京东核心中间件介绍

宝剑锋从磨砺出

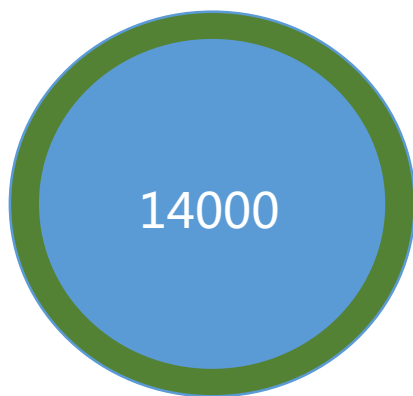
——京东中间件的架构演进

饮水思源

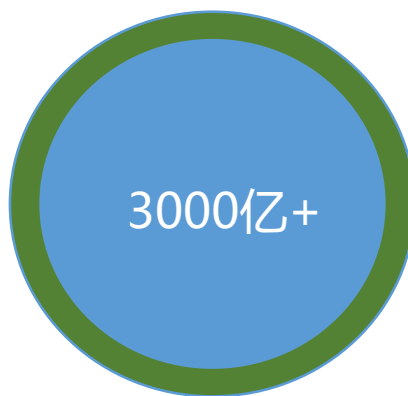
——京东中间件的开源计划

JSF

自主研发高性能分布式的RPC微服务框架，是京东服务化、开放化的技术标准



服务接口



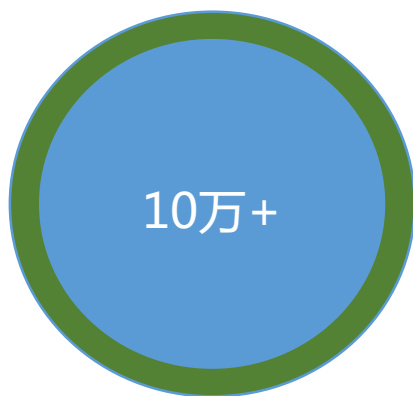
调用量
(11.11)



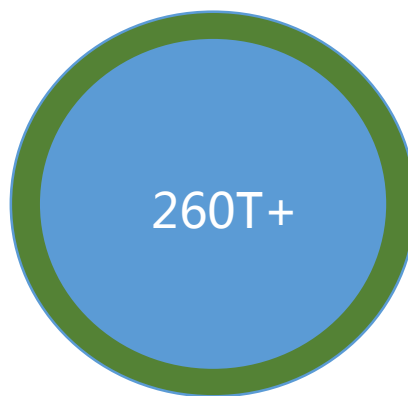
实例数

JIMDB

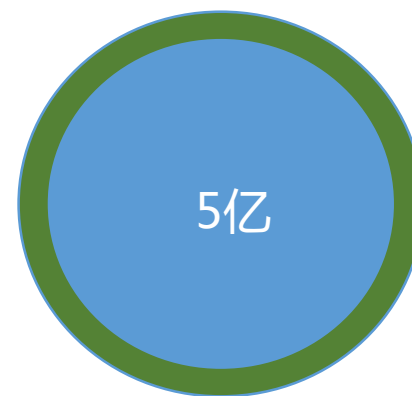
自主研发高性能分布式缓存平台



实例数量



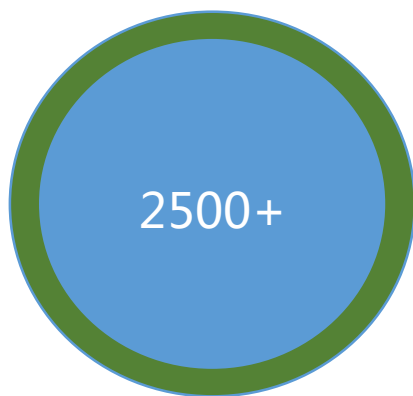
内存量



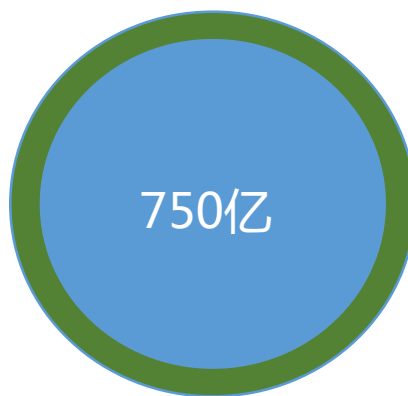
OPS峰值
(11.11)

JMQ

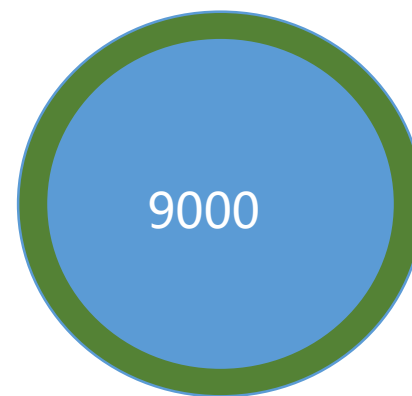
自主研发的高性能分布式消息队列



接入应用

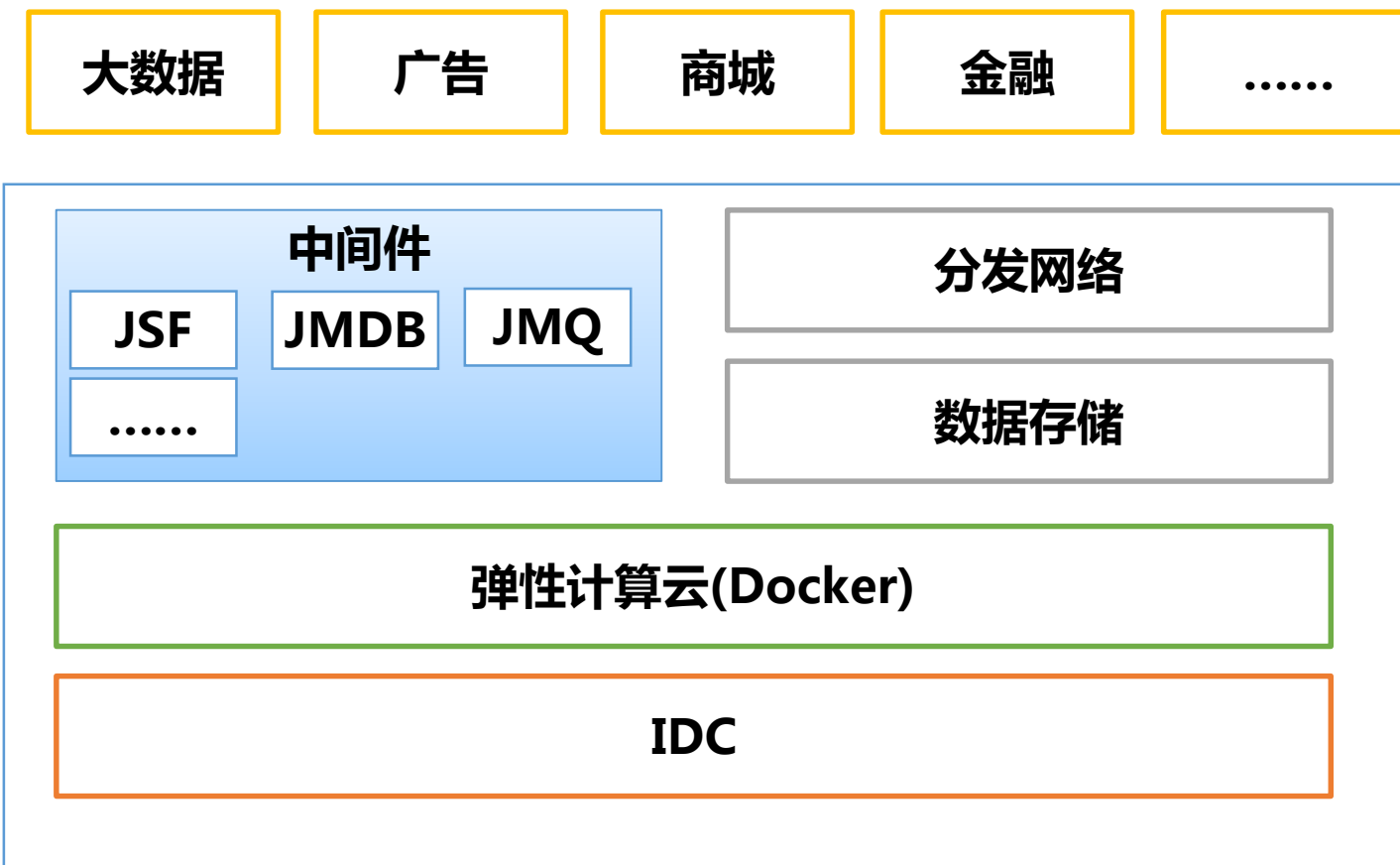


吞吐量
(11.11)

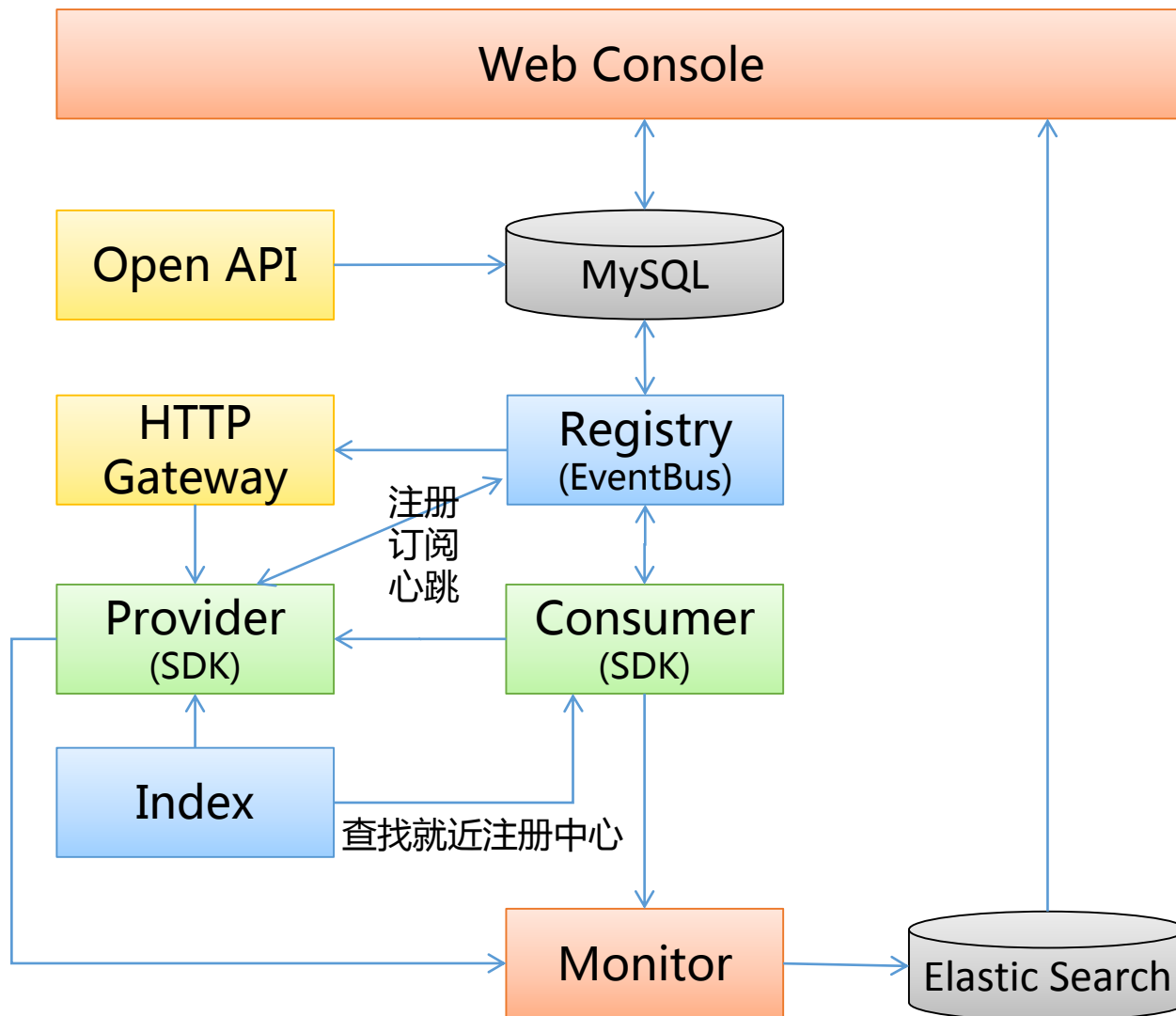


队列数量

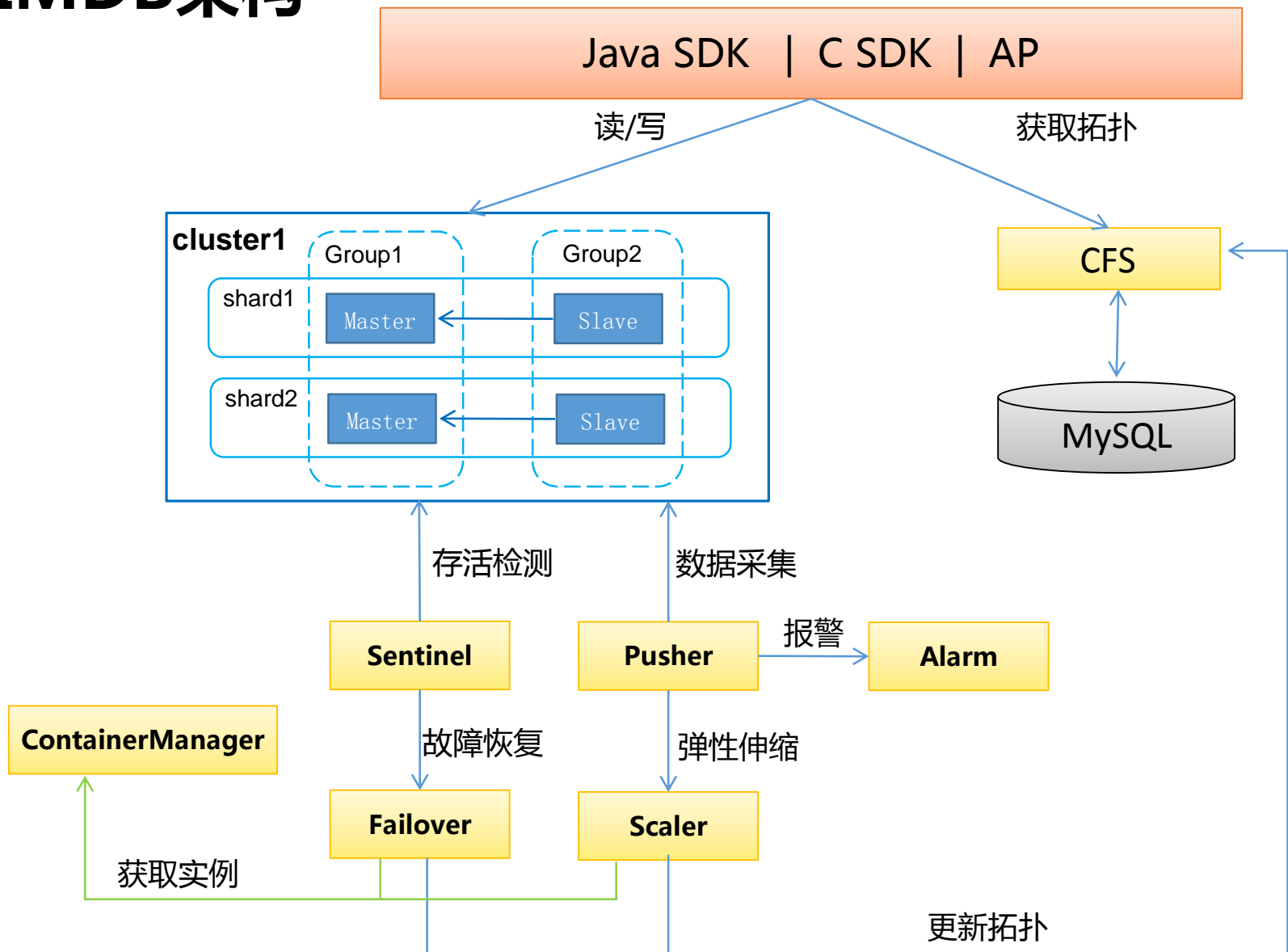
中间件核心支撑



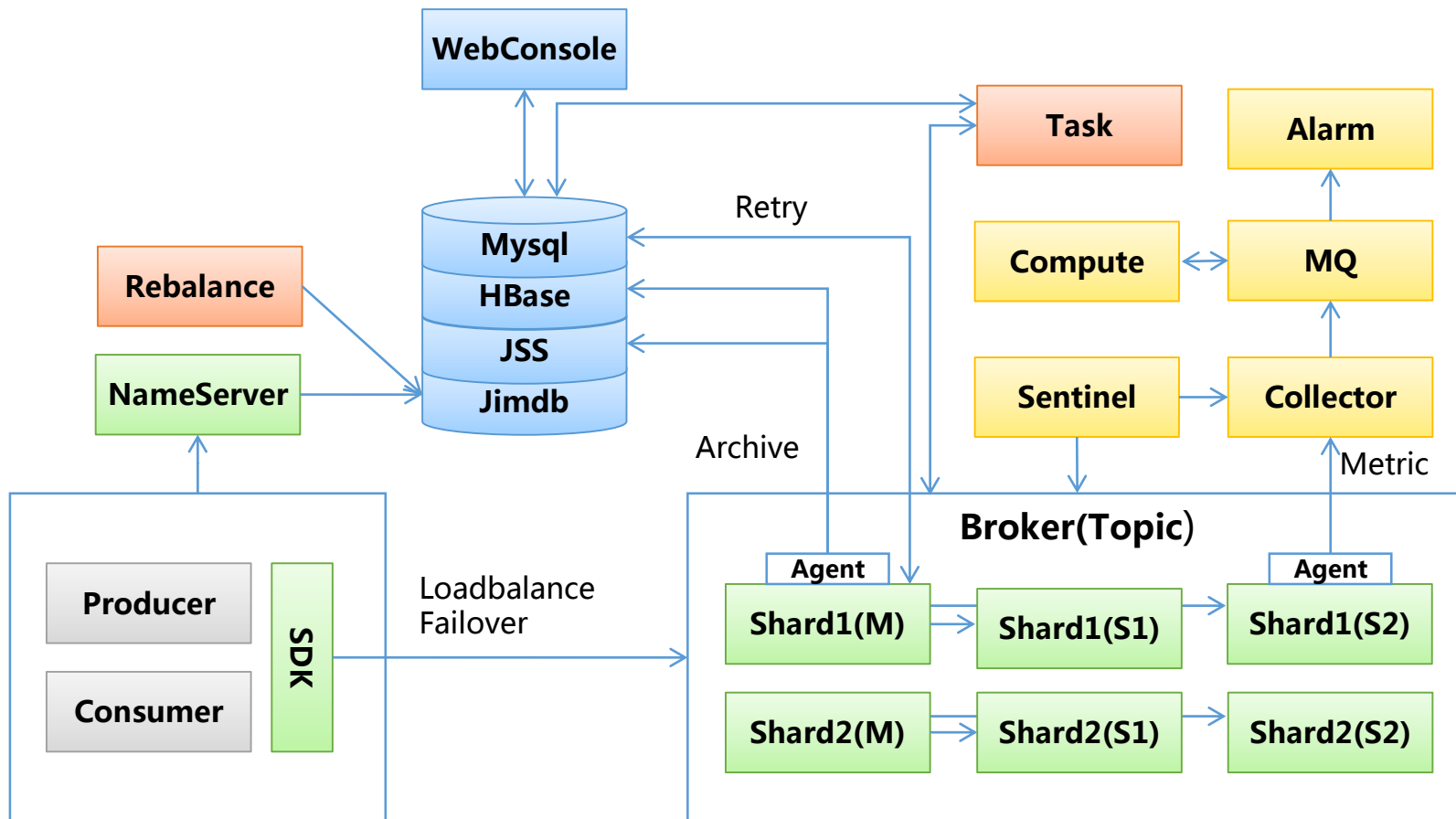
JSF架构



JIMDB架构



JMQ架构



大纲

邦之利器，今可示人

——京东核心中间件介绍

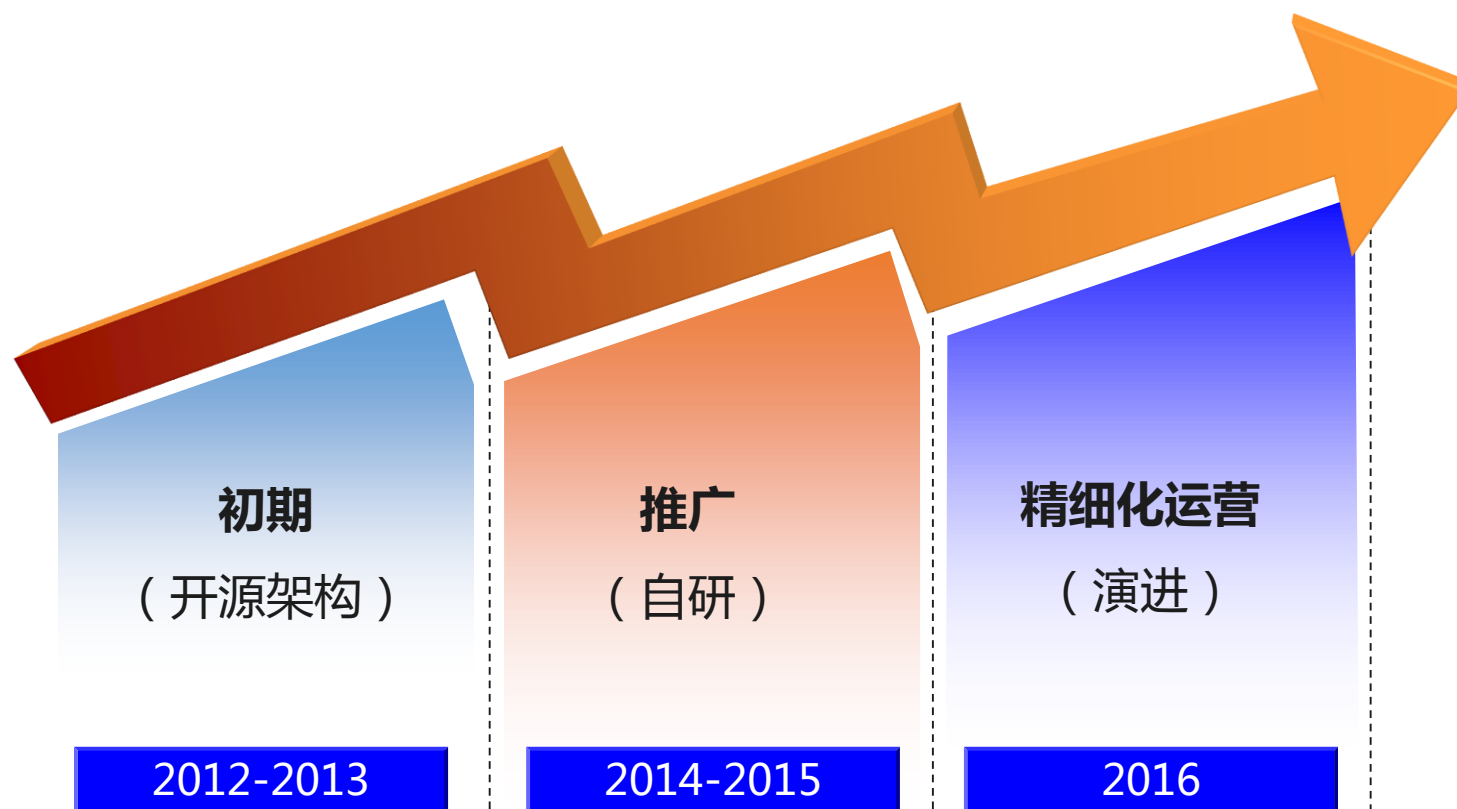
宝剑锋从磨砺出

——京东中间件的架构演进

饮水思源

——京东中间件的开源计划

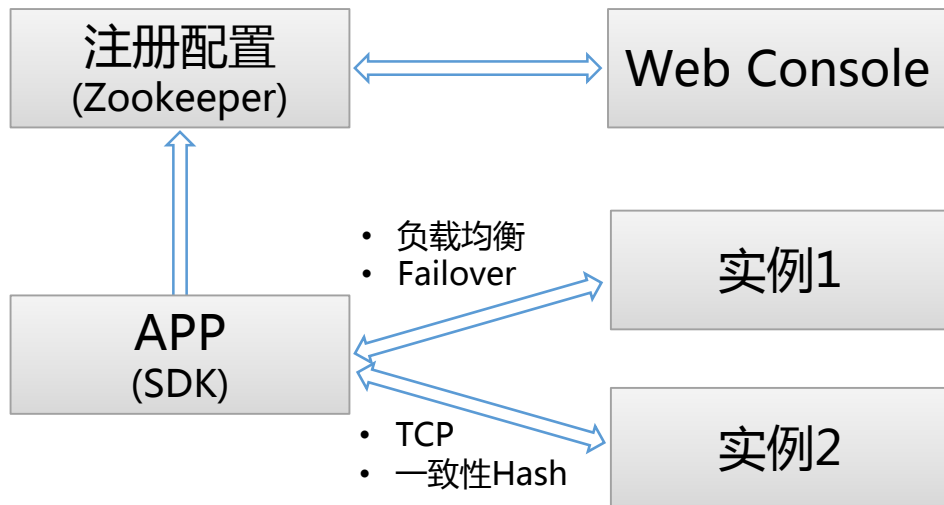
架构演进



襁褓初期，蹒跚起步

从.NET转向JAVA，SOA

痛点		解决方案
性能慢	直接访问数据库，DB压力山大	分布式Redis方案，保护数据库
	串行同步调用多个接口，性能低	分布式ActiveMQ方案，异步调用
	HTTP短链接调用	RPC框架，实现服务治理，TCP长连接，通过Zookeeper实现注册中心
缺乏服务标准和治理，找不到服务提供方和消费方		



试水取经，带痛成长

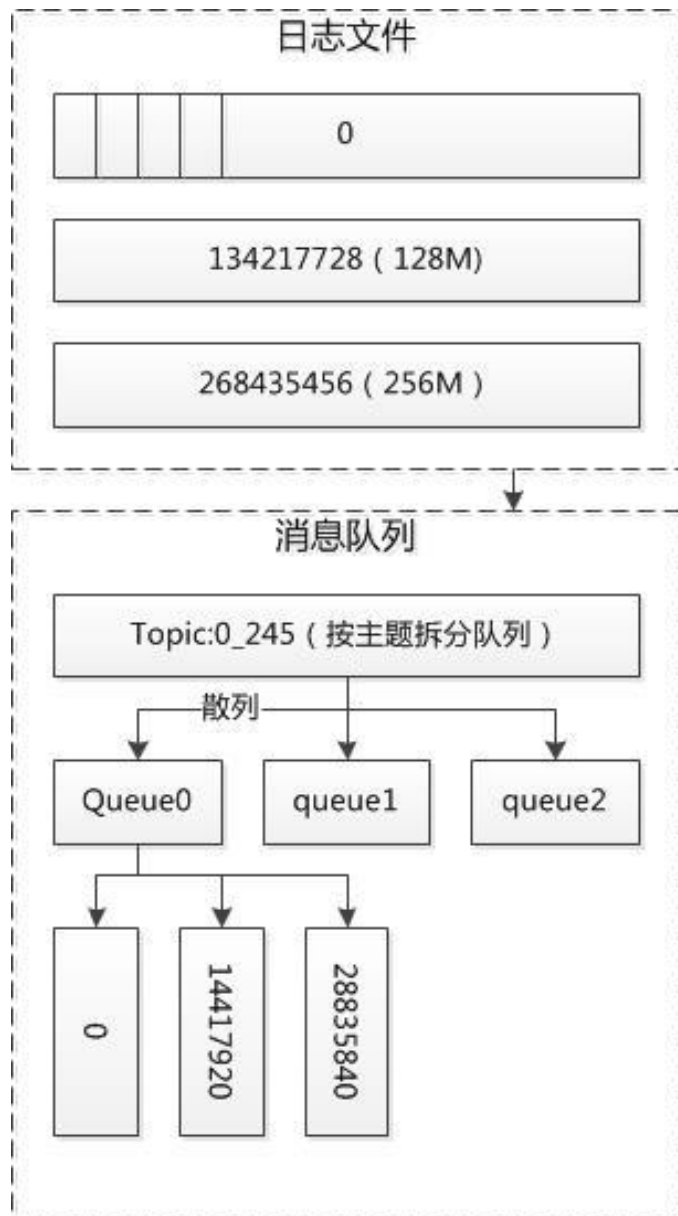
大促无法满足，代码掌握不深，响应不能及时，开始自研中间件

痛点	解决方案
性能满足不了大促的需求	<ol style="list-style-type: none"> 1. 减少网络传输：优化通信协议，减小数据包大小，数据压缩，批量传输 2. 异步处理(NIO，异步事件) 3. 减少内存数据拷贝 4. 优化文件存取：顺序追加，组提交，减少拷贝，内存映射文件 5. 优化复制协议：并行复制，增量复制

案例 AMQ压测1K消息刷盘，写入TPS是4000；生产环境，订单管道消息，几十个消费者，写入TPS只能达到几百。

分析 1) 存储模型有缺陷，采用B-Tree索引，性能随消息积压量的上升急剧下降；多个消费者订阅，消费者多个实例部署，只能把Topic转换成Queue模型，一份数据多次拷贝。

2) 网络模型不够鲁棒，NIO实现不好



1. 存储消息体，同步刷盘
2. 文件名为起始全局偏移量
3. 顺序追加
4. 消息ID组成：*IP:端口:日志偏移量*

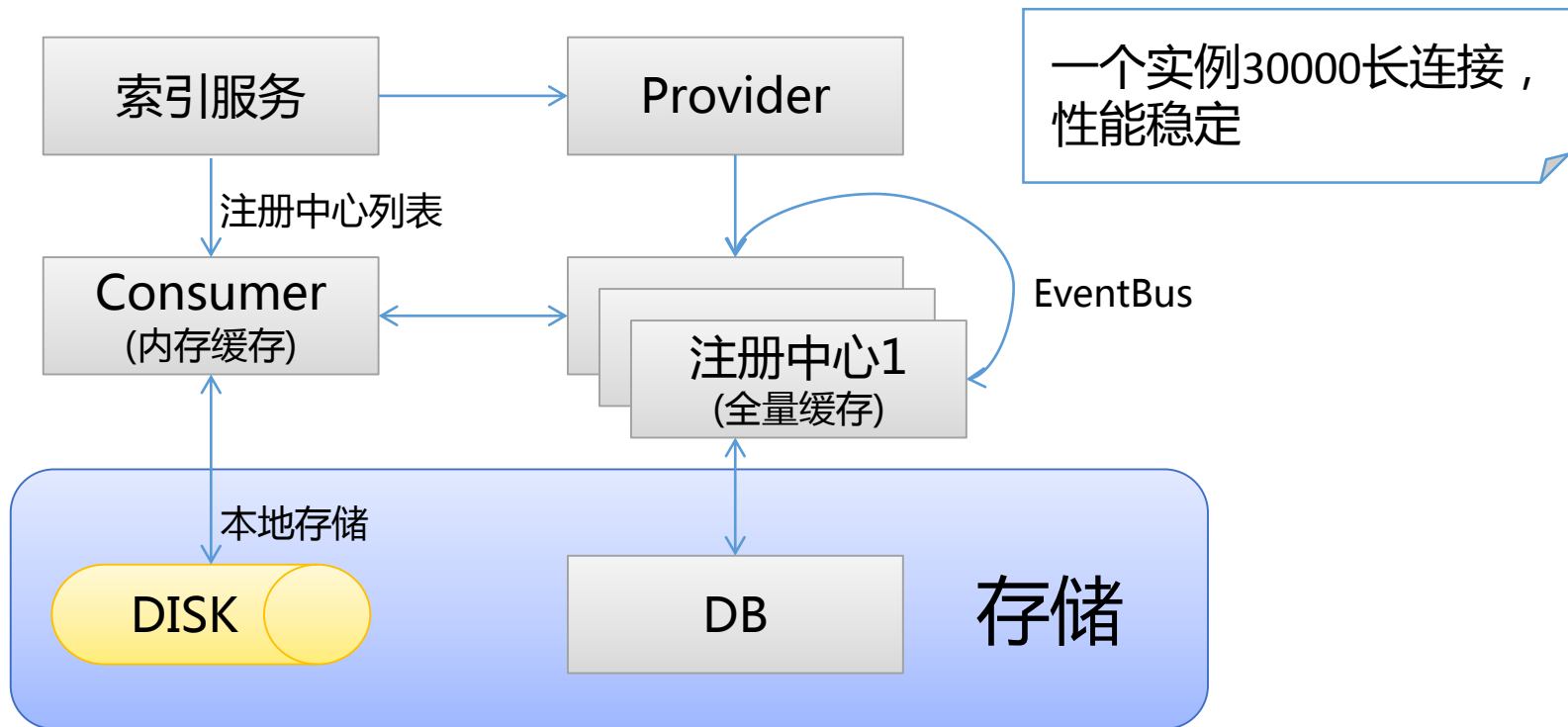
1. 存储消息位置，快速索引，异步刷盘
2. 文件名为起始全局偏移量
3. 顺序追加
4. 支持按业务ID散列，顺序消费
5. 一个队列只对应一个消费者

经过各种措施优化后，1K消息刷盘，写入
TPS达到**40000**

渐进推广，循序优化

成倍增长的规模

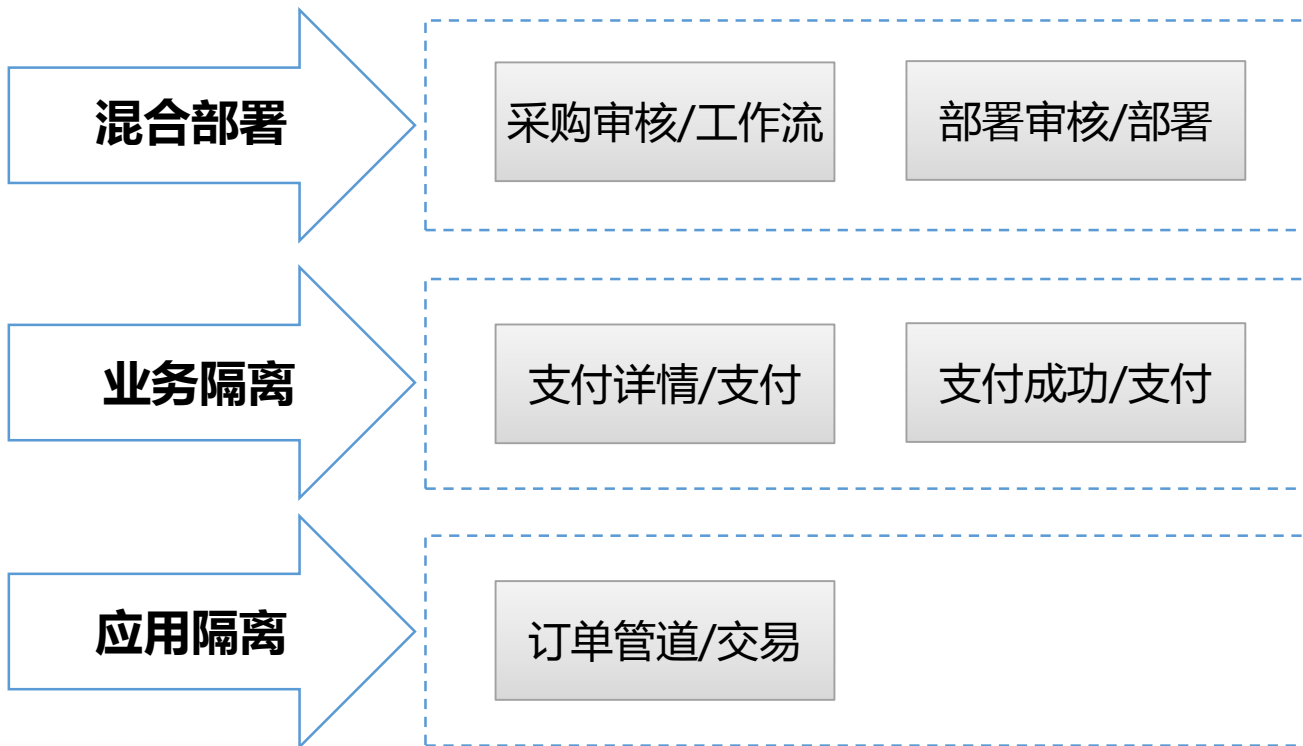
痛点	解决方案
应用实例暴增，Zookeeper连接数过大(16000)，雪崩效应	自研注册中心，采用DB+本地缓存模式，分布式部署横向扩展



渐进推广，循序优化

应用接入增多，资源有限，探索共享机制

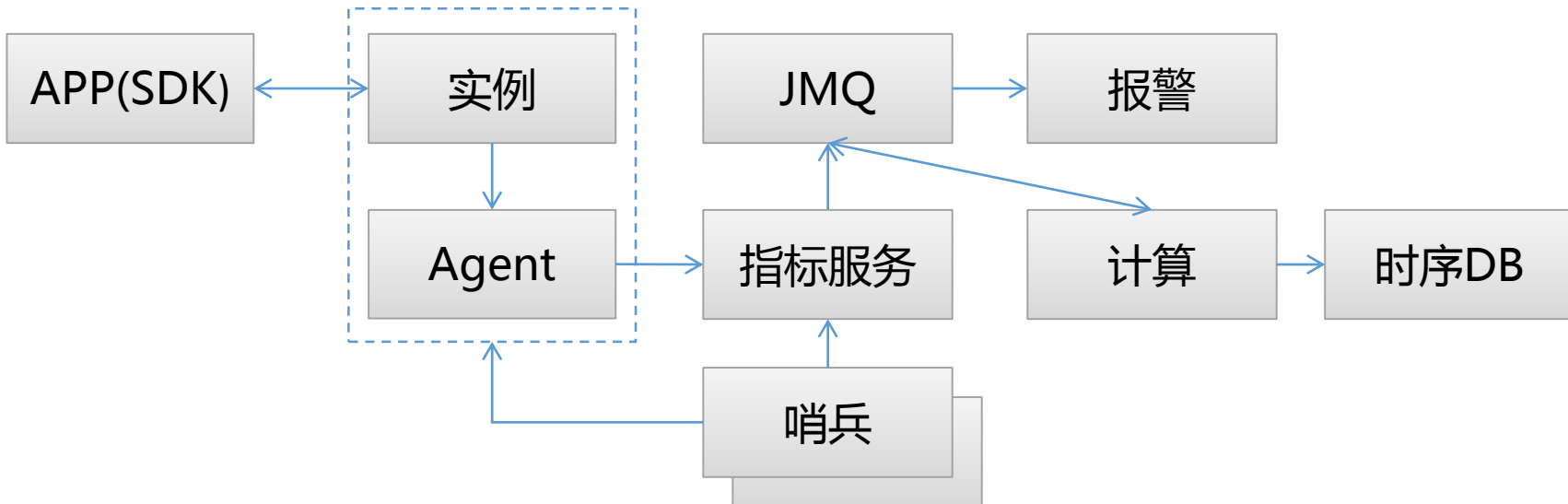
痛点	解决方案
核心应用和非核心应用混合部署，互相影响	<ol style="list-style-type: none"> 1. 注册的时候提供应用级别和业务信息 2. 可以按照业务进行物理隔离 3. 可以按照核心应用进行物理隔离



渐进推广，循序优化

使用量持续增长，集群规模不断扩大，故障数增加，系统稳定性需要保障

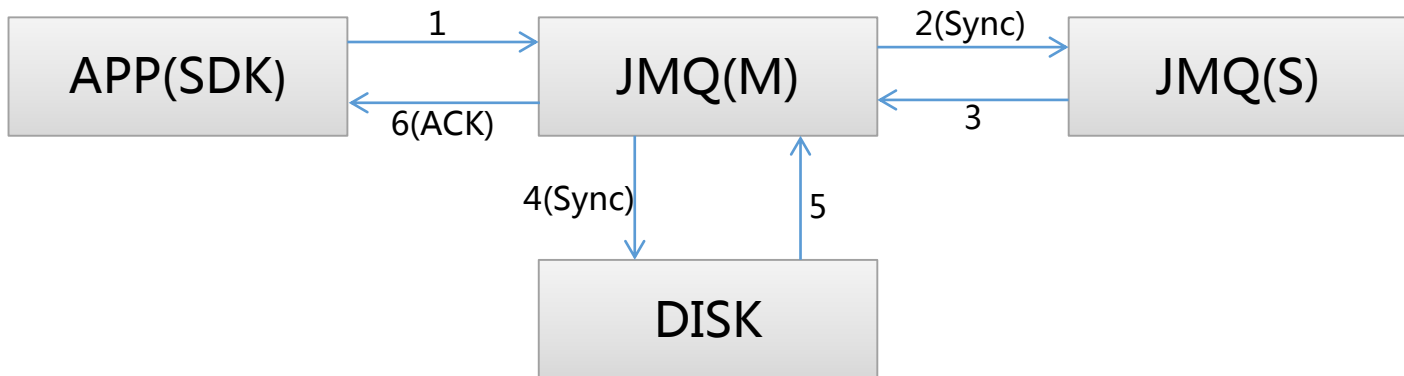
痛点	解决方案
监控指标不足，客户端性能较难采集	1. 尽可能采集系统指标，包括硬件和网络性能指标 2. 客户端生成TPS性能指标和成功指标，定时上传
报警策略僵硬	报警策略可自定义：根据指标、应用和集群灵活配置策略(阈值、邮件、短信、用户)
历史数据海量，存储困难	采用时序数据库来存储（OpenTSDB，Elasticsearch）
存活误报	通过本地机房多个机架上的探针进行宕机检测和投票



渐进推广，循序优化

诸如订单、收货地址等关键数据，要确保不能丢失

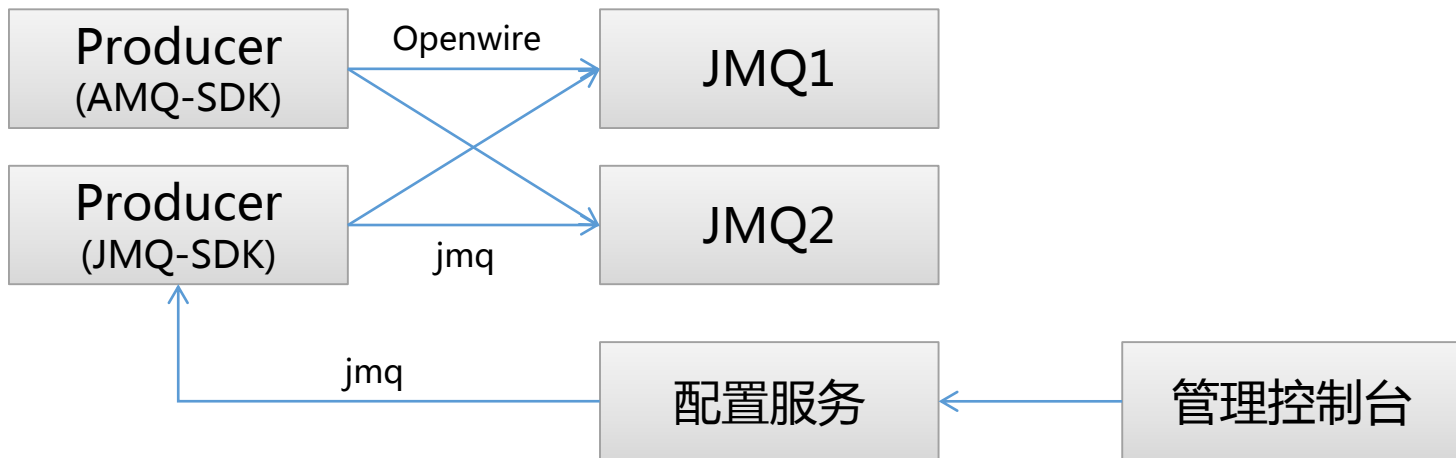
痛点	解决方案
MQ消息不能丢失	<ol style="list-style-type: none"> 1. 消息应答模式 2. 服务端收到消息，强制刷盘，防止掉电 3. 服务端收到消息，主从同步复制 4. 至少投递成功一次，消费者需要防重

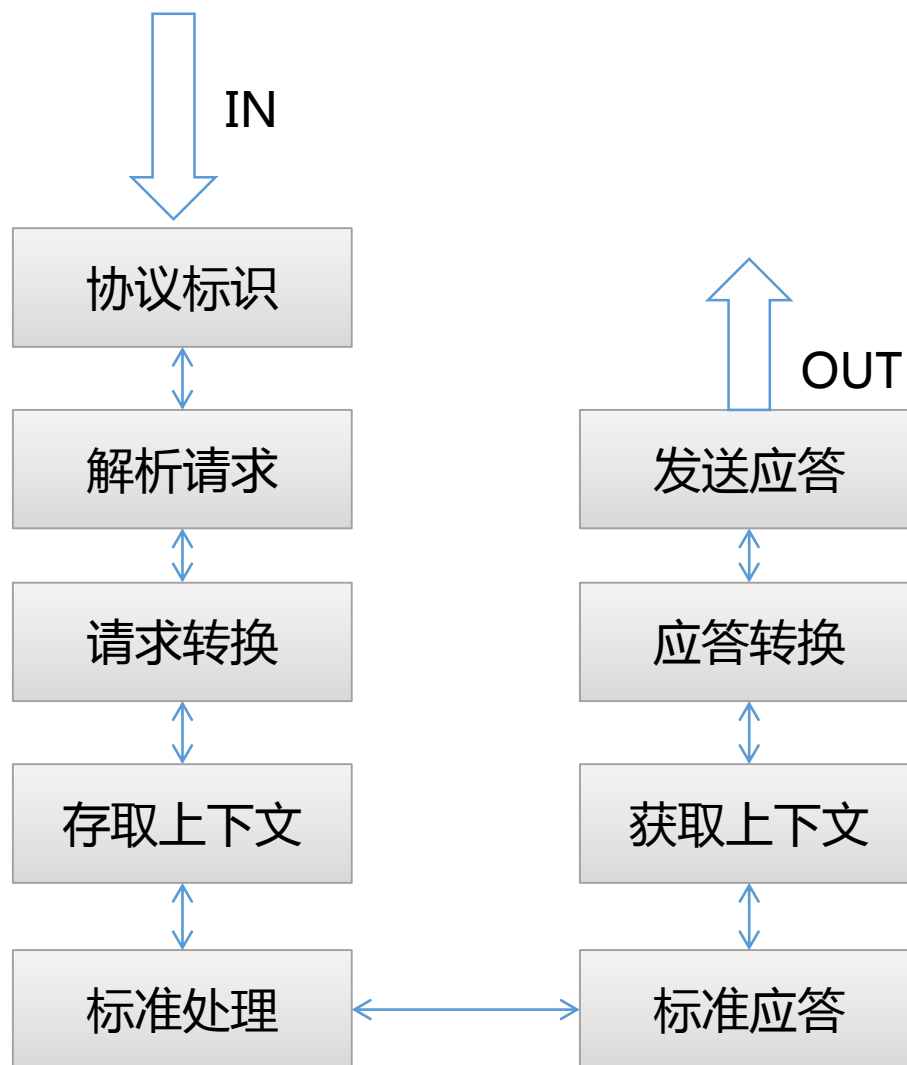


渐进推广，循序优化

升级步伐跟不上业务发展的速度，导致新老版本共存

痛点	解决方案
新老版本共存，升级困难	<ol style="list-style-type: none"> 1. 客户端协议增加版本标识 2. 服务端向后兼容 3. 轻客户端，重服务端； <ul style="list-style-type: none"> ✓ 减少客户端依赖，消费位置由服务端保存 ✓ 客户端路由策略，权重由服务端动态控制 ✓ 动态下发客户端参数，减少不必要的修改重启

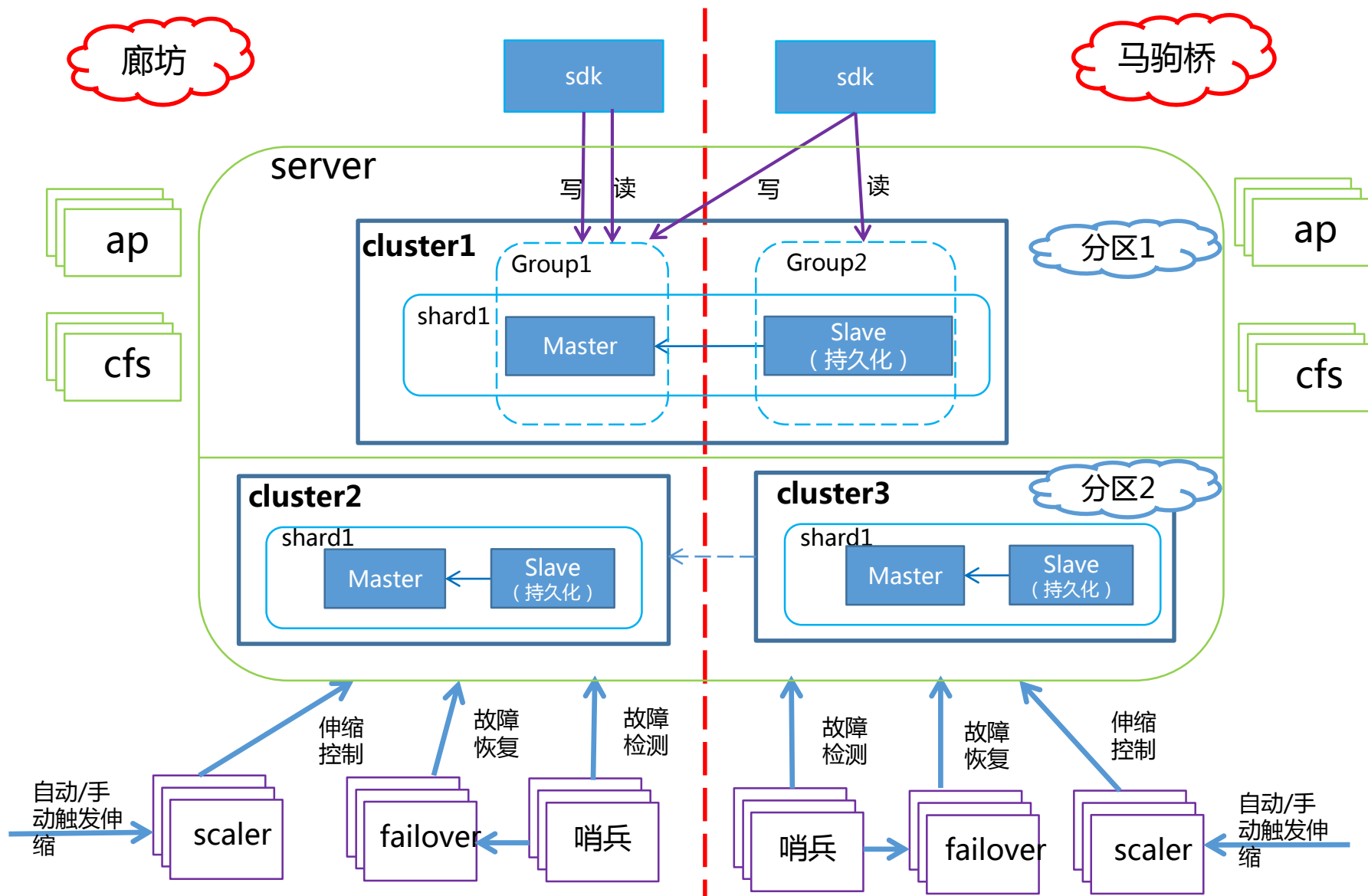




深耕精作，精细化运营

系统稳定性要求更高，实施容灾和应急预案

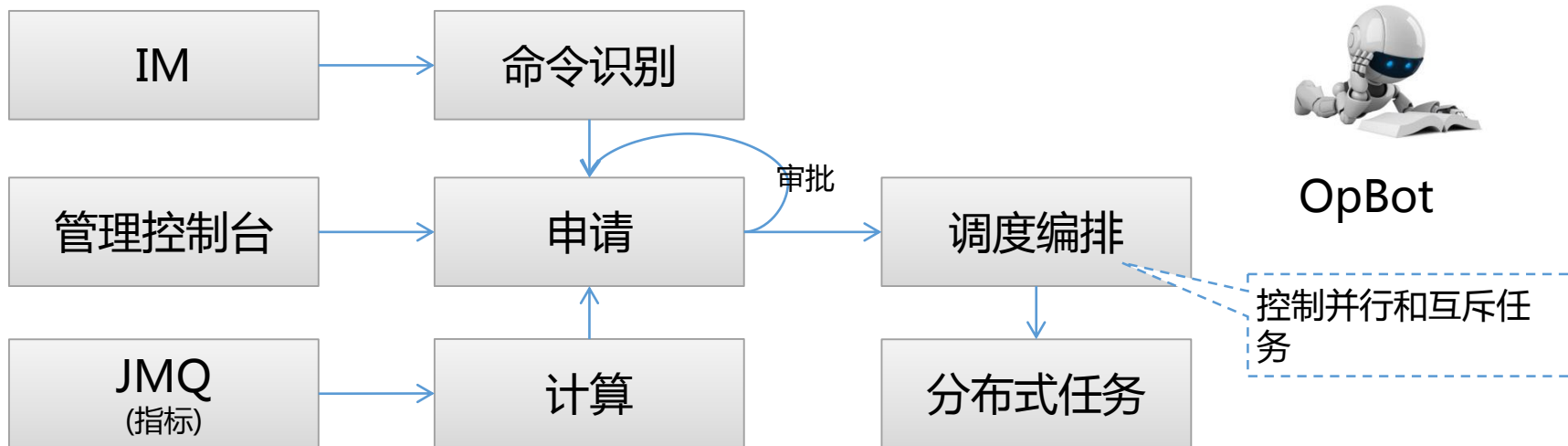
痛点	解决方案
容灾	1. 集群或分片跨物理机部署容灾 2. 集群或分片跨机架部署容灾 3. 集群跨机房部署容灾、双活、主备和复制 4. 定期巡检
跨机房网络延时	优先本地机房调用



深耕精作，精细化运营

业务全部接入中间件平台，规模增长跃升，运维压力增大

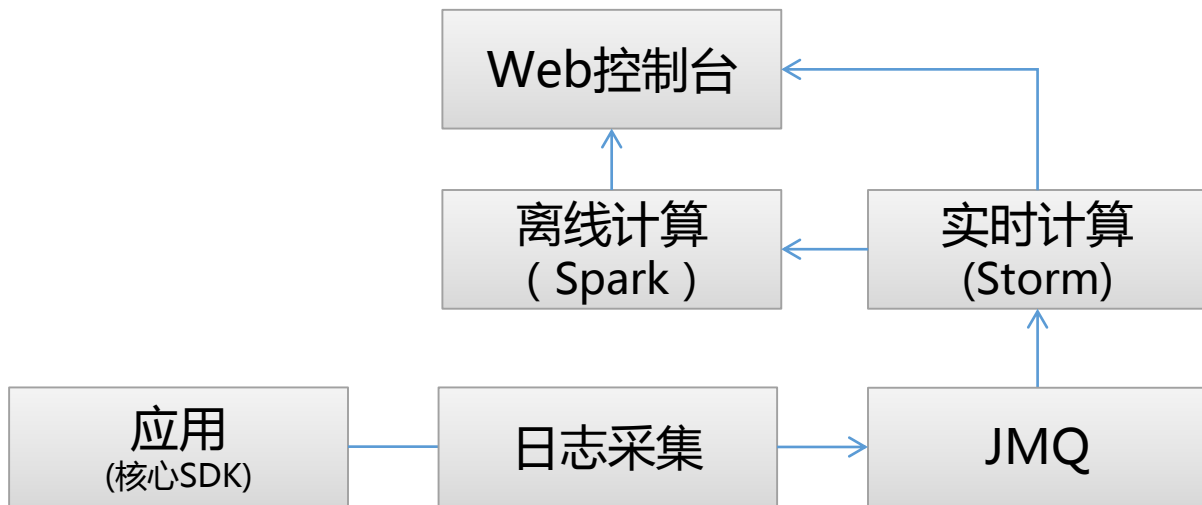
痛点	解决方案
接入工作量大；	自助申请，自动创建
机器数量增加，故障数增加；	故障检测，自动Failover
扩容工作量很大	自动弹性扩容，基于Docker弹性，丰富的弹性策略：容灾，应用级别，硬件需求，隔离需求，主机负载
历史原因造成机器负载不均匀	根据负载情况，自动迁移



深耕精作，精细化运营

微服务和异步架构，应用千种万样

痛点	解决方案
问题查询定位不方便	分布式跟踪



WHO: 调用链唯一标示
 WHEN: 调用链开始时间，结束时间
 WHAT: 环节序号，调用结果
 WHERE: 服务地址，服务名称，方法名称
 HOW: 类型，请求大小，应答大小

调用链详情

TraceId : ac1378aa14764397883893553d4f5c

开始时间 : 2016-10-14 18:09:48

总耗时 : 119ms

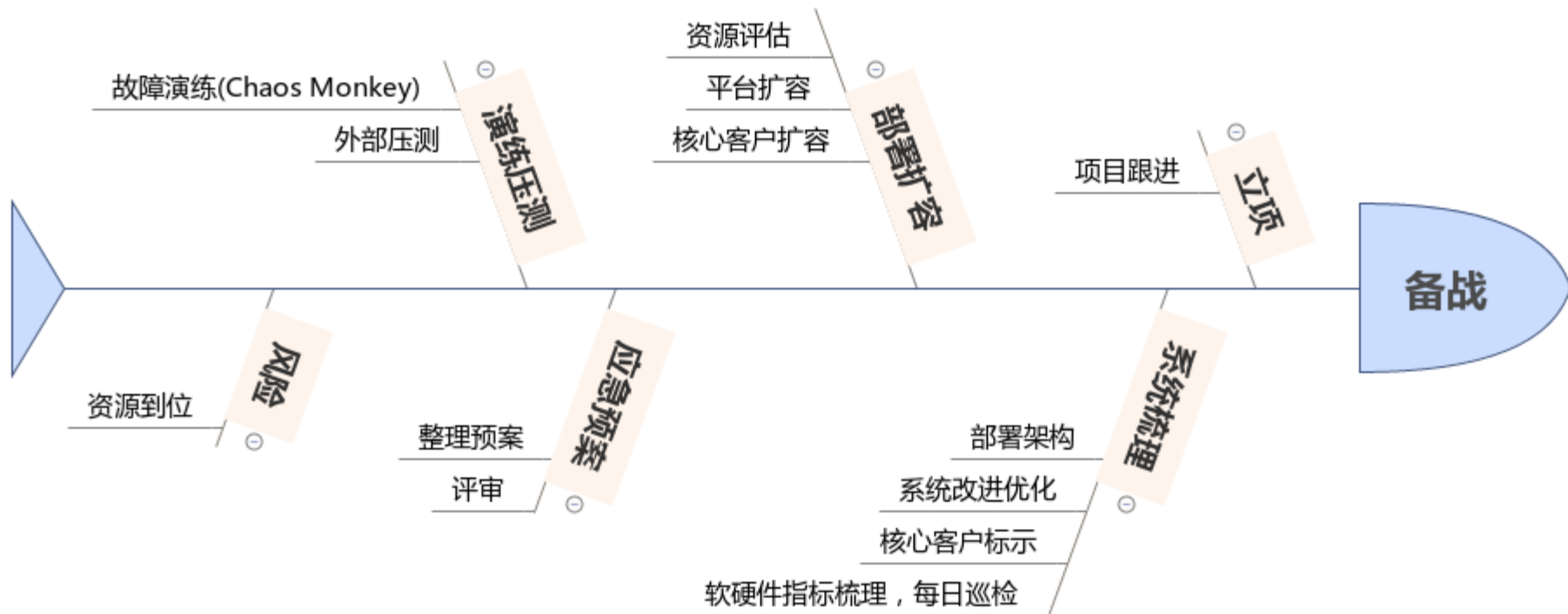
[日志原文](#)

TraceId ac1378aa14764397883893553d4f5c

查询

应用名	IP	类型	状态	数据量	服务/方法	耗时
▼ gamegw	172.19.120.1...	TRACE	成功	-	com.jd.card.soa.api.service.CardOrderA...	119ms
card	172.19.120.169	JSF	成功	0.3KB	com.jd.pop.ware.ic.api.service.read.Sku...	4ms
card	172.19.120.169	JSF	成功	0.4KB	com.jd.pop.ware.ic.api.service.read.War...	3ms
card	172.19.120.169	JSF	成功	2.6KB	com.jd.cart.client.service.CartService@g...	4ms
card	172.19.120.169	JSF	成功	0.1KB	com.jd.pop.vender.center.service.catego...	2ms
card	172.19.120.169	JSF	成功	0.3KB	com.jd.user.sdk.export.UserInfoExportSe...	2ms
card	172.19.120.169	JSF	成功	0.1KB	com.jd.gaj.sdk.service.VirtualBusinessR...	5ms
card	172.19.120.169	JSF	成功	0.1KB	com.jd.jmi.escort.api.service.SecurityAd...	8ms
card	172.19.120.169	JSF	成功	0.2KB	com.jd.vertical.trade.sdk.export.SubmitO...	17ms

火炼真金，大促备战



大纲

邦之利器，今可示人

——京东核心中间件介绍

宝剑锋从磨砺出

——京东中间件的架构演进

饮水思源

——京东中间件的开源计划

开源与自研，相辅相成



Q&A



THANKS



[北京站]

主办方 **Geekbang** > **InfoQ**
极客邦科技