



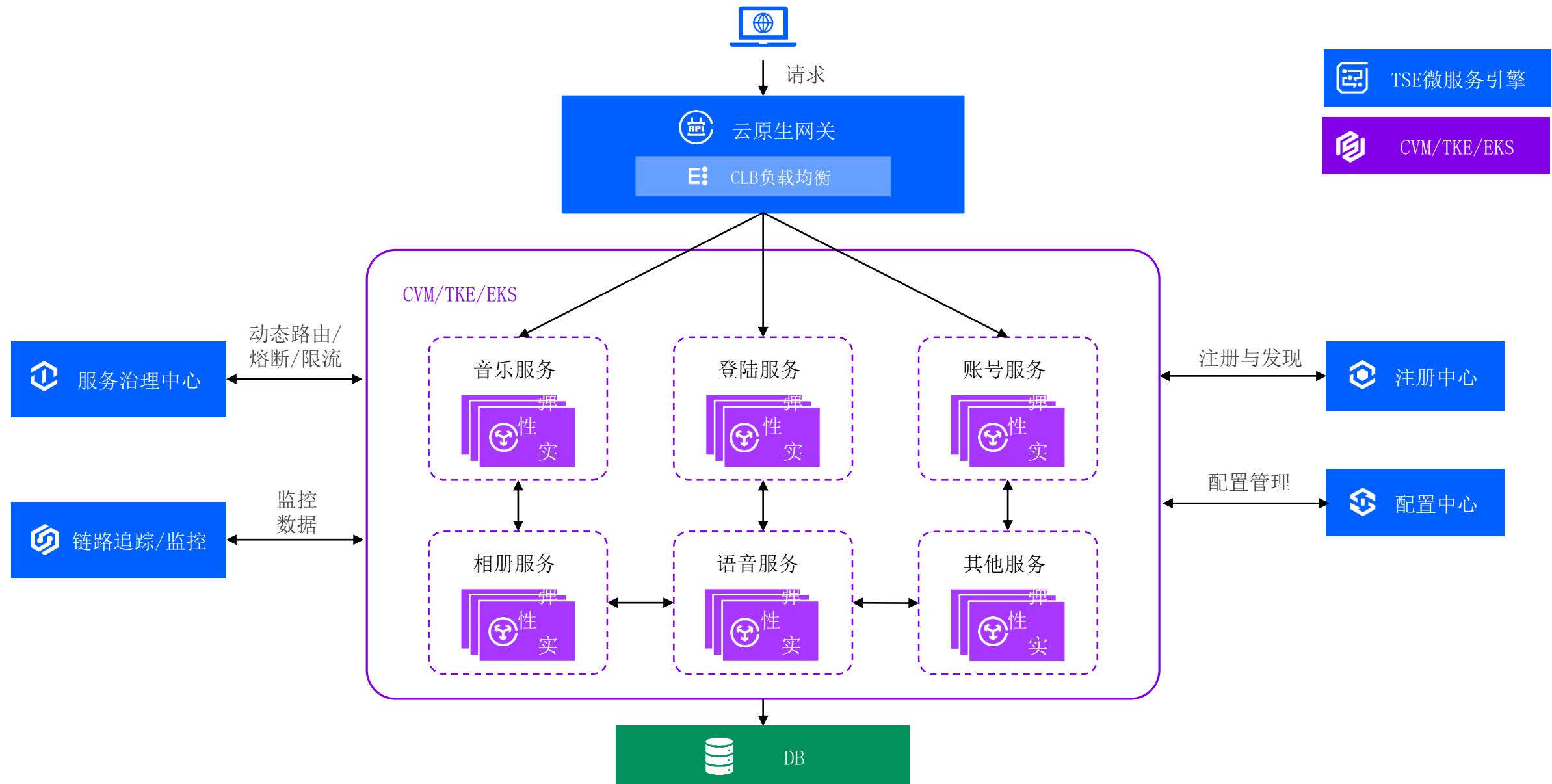
# 微服务引擎 TSE

— 微服务架构整体解决方案

颜松柏 / 微服务架构师



# 微服务架构



01

# TSE 微服务引擎

云原生网关/注册配置中心/服务治理中心

# 微服务引擎 TSE：简介

微服务引擎 TSE 是面向业界主流开源微服务项目的全家桶。提供：

- **注册配置中心**（原生支持Nacos/ZooKeeper/Consul/Eureka/Apollo）；
- **云原生网关**（原生支持Kong/Nginx/Nginx Ingress）；
- **微服务治理**（原生支持Spring Cloud/Dubbo/gRPC等框架，提供遵循Linux NextArch基金会微服务SIG服务治理规范的北极星PolarisMesh）能力。

## TSE的组成



兼容开源

易用

功能增强

运维简单

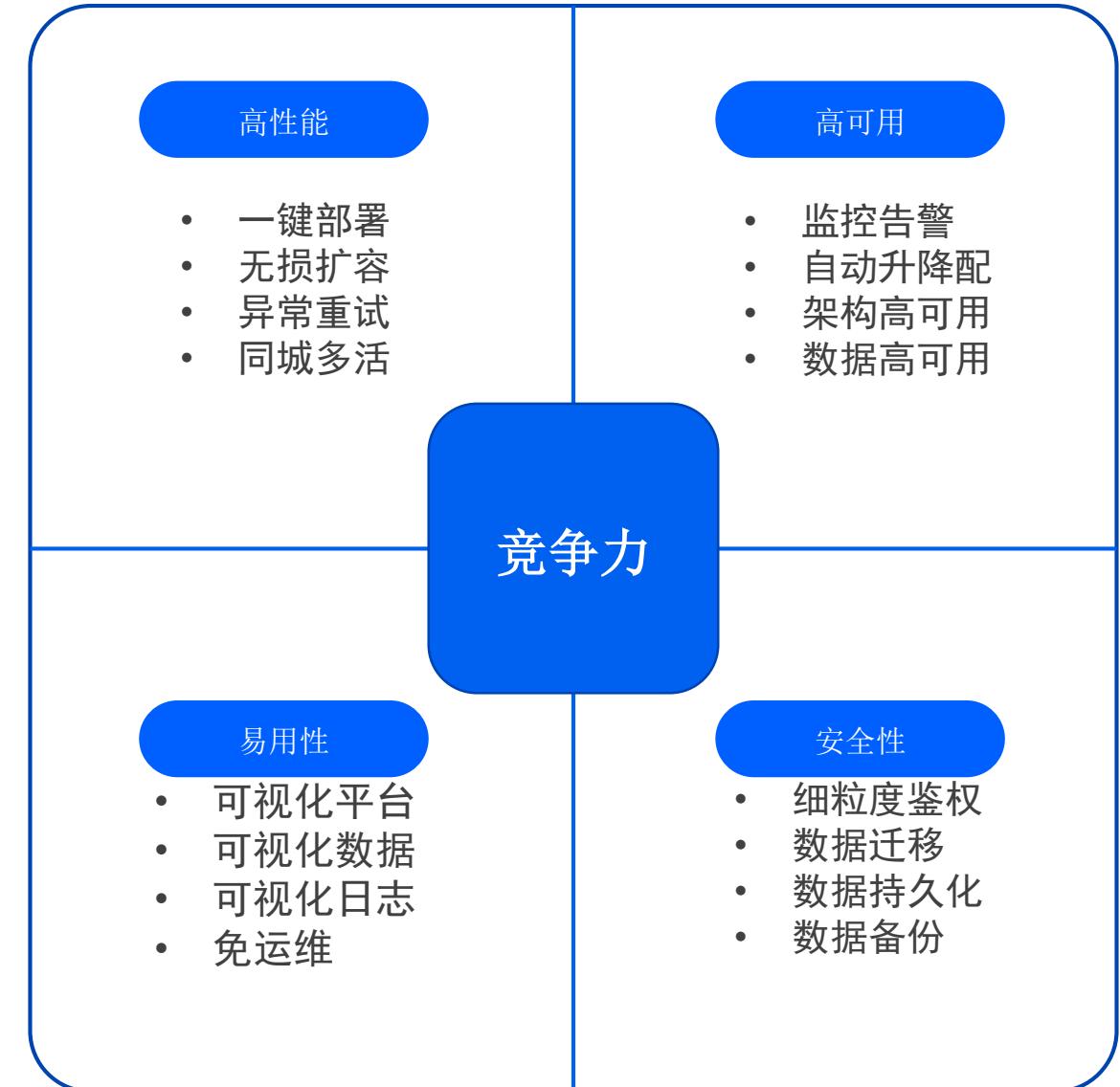
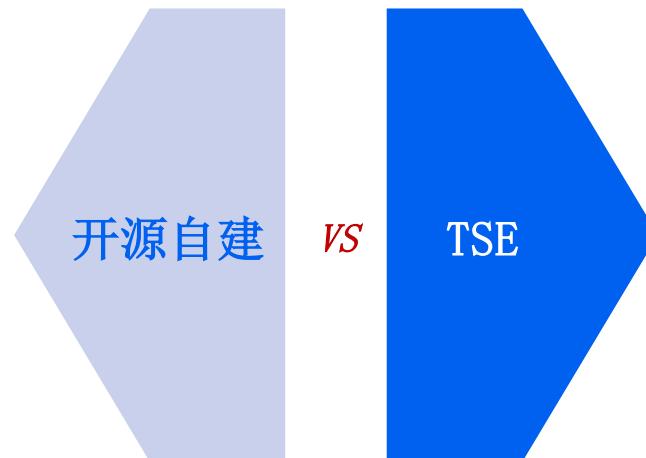
安全

高可用

# 微服务引擎 TSE：商业化进展

TSE	开源内核	开源内核维护组织	商业化进展
云原生网关	 NGINX	F5	即将商业化
	 Kong	Kong	即将商业化
注册配置中心	 Apache ZooKeeper™	Apache	商业化
	 NACOS.	阿里巴巴	商业化
	 Apollo	携程	商业化
	 NETFLIX EUREKA	Netflix	公测
	 Consul	HashiCorp	公测
	 etcd	CoreOS	规划中
服务治理中心	 POLARIS MESH	腾讯	商业化

# 微服务引擎 TSE：相比开源自建的价值



# TSE 相比开源自建的优势

## 对比项

搭建及运维

易用性

性能

安全

监控告警

云产品集成

## TSE

开箱即用/免运维

可视化操作与不停服重启

深度优化

CAM权限管理

已包含、无额外费用

与TEM等产品深度集成

## 开源自建

自行搭建及运维

手动修改配置文件并重启

需自行调优

单独建设

单独建设

不支持

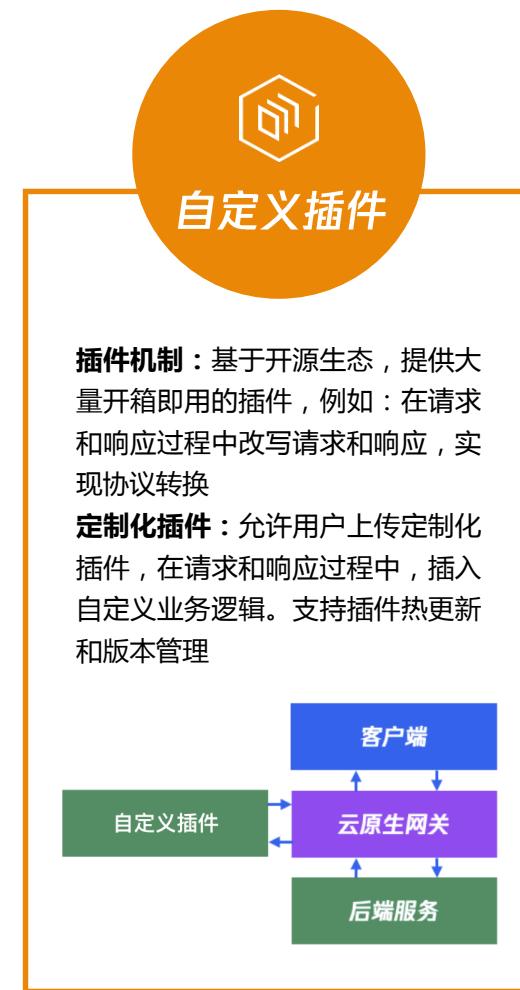
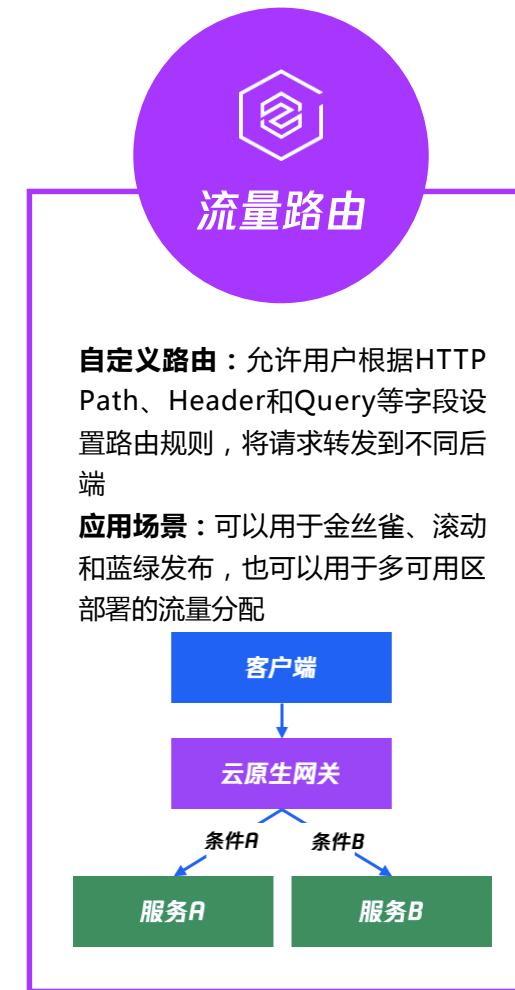


## TSE 应用场景一

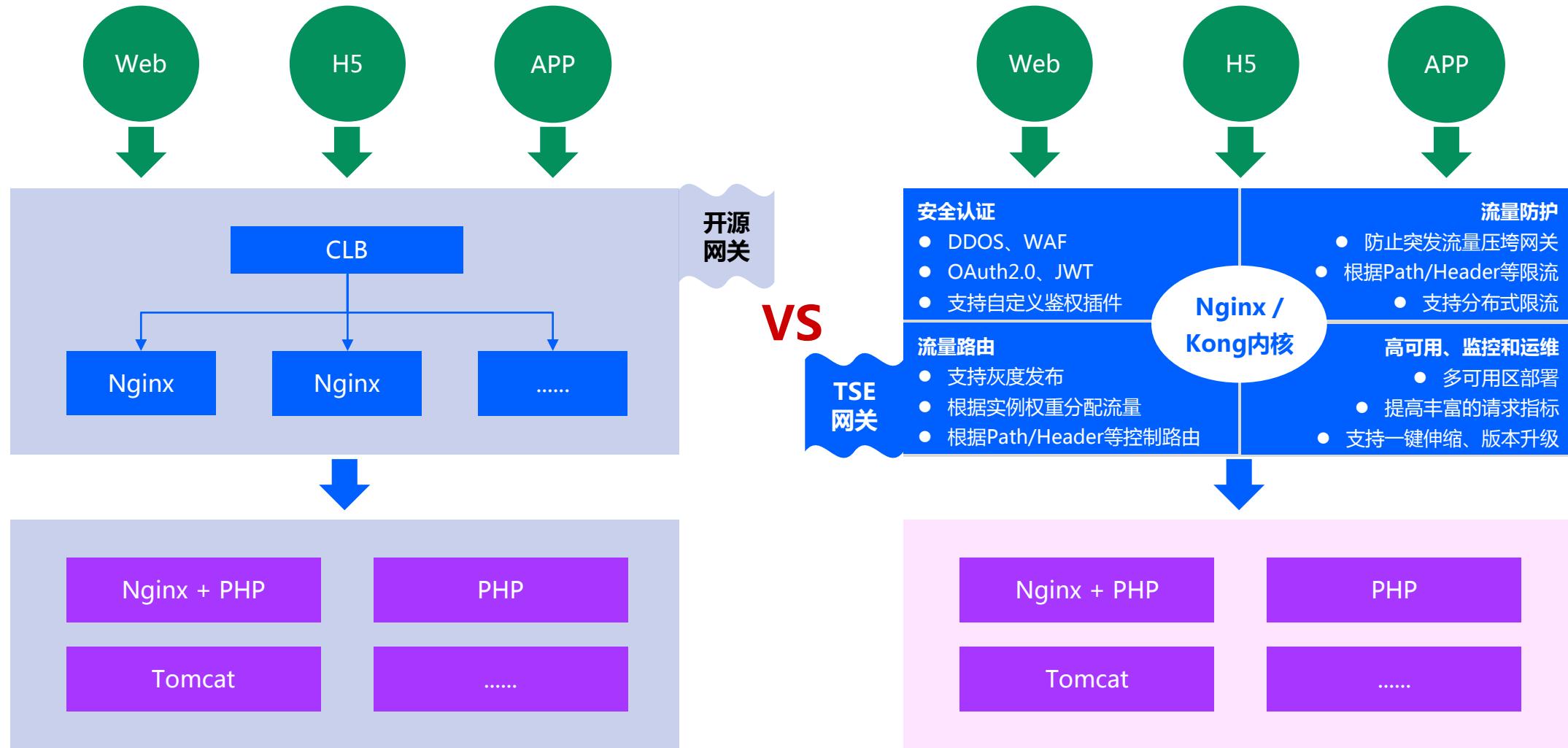
构建安全可靠、功能强劲的业务网关



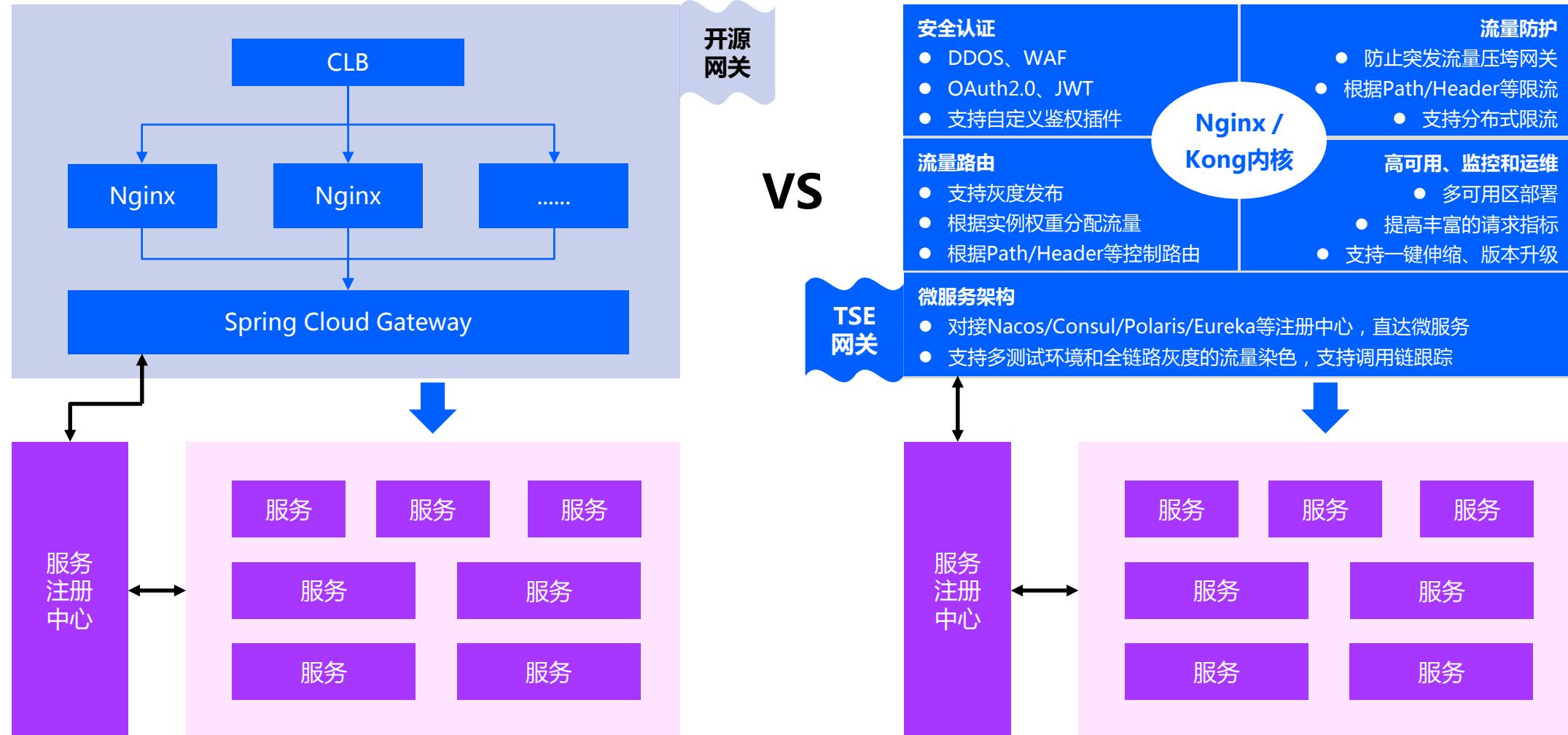
# TSE 云原生网关 = 流量网关 + 安全网关 + 服务网关



# 典型Web架构：开源网关 VS TSE 云原生网关



# 典型微服务架构：开源网关 VS TSE 云原生网关

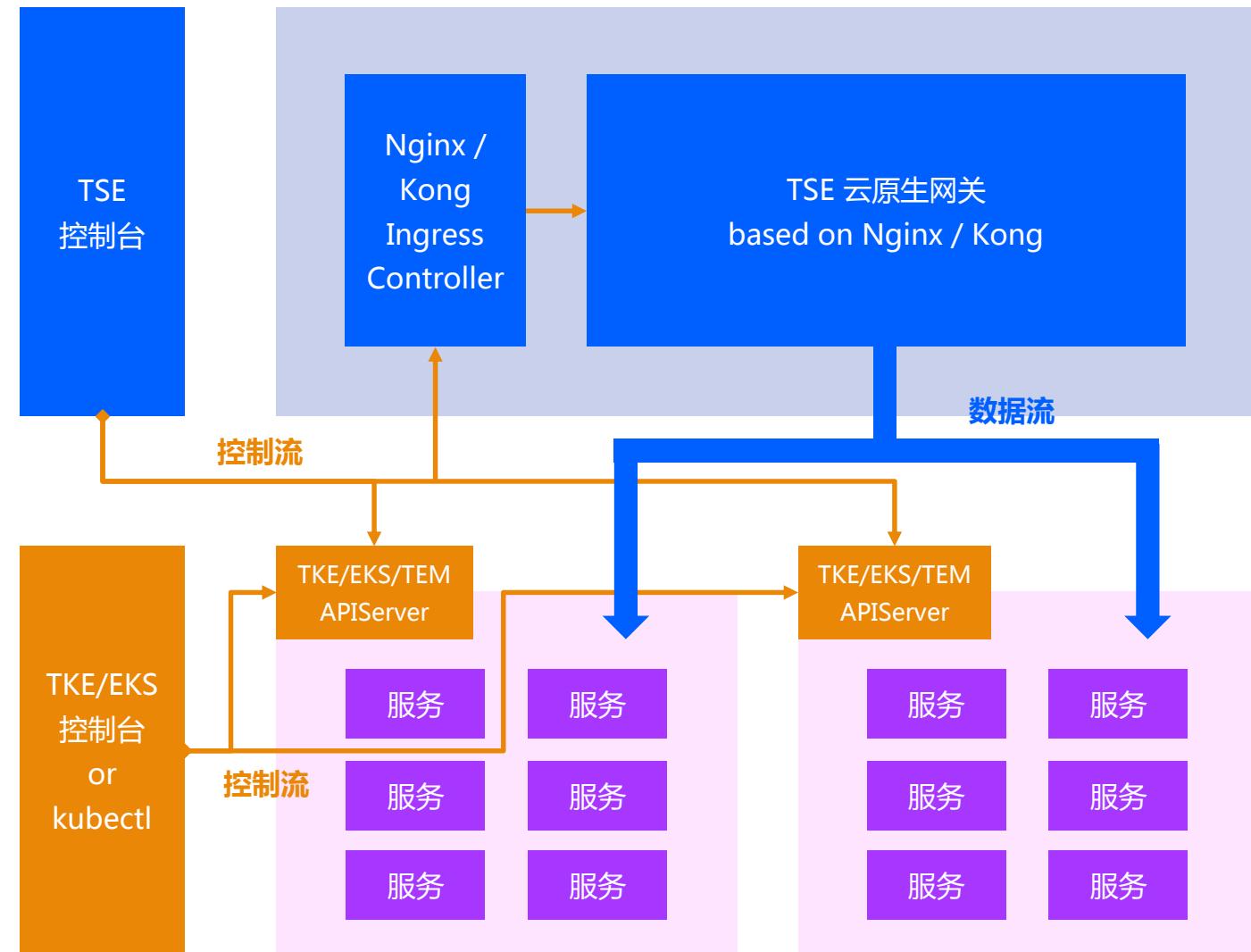


# TSE 云原生网关支持 Nginx / Kong Ingress

如果接入层服务在 **TKE/EKS/TEM** 等容器平台上 , TSE 云原生网关支持 **Ingress** 模式 :

- 兼容原生 Ingress 架构和使用方式
- 享有 TSE 云原生网关全部能力
- 支持多 K8s 集群的流量路由

```
apiVersion: networking.k8s.io/v1
kind: Ingress
metadata:
  name: ingress-myservicea
spec:
  rules:
    - host: myservicea.foo.org
      http:
        paths:
          - path: /
            pathType: Prefix
            backend:
              service:
                name: myservicea
                port:
                  number: 80
  ingressClassName: tse-nginx
```



# TSE 云原生网关相比开源网关的优势

功能类型	功能特性	TSE 云原生网关	开源 Nginx	开源 Kong
安全认证	<b>DDos</b>	支持4层和7层DDos高级防护	需要自研	需要自研
	<b>WAF</b>	防SQL注入等常见Web攻击、防爬虫、防刷	需要自研	需要自研
	<b>身份认证</b>	支持 OAuth2.0、JWT 等协议	需要自研	插件支持
		支持对接腾讯云的数字身份管控平台	需要自研	需要自研
	<b>自定义鉴权</b>	支持用户设置鉴权插件，对接自己的鉴权系统	需要自研	插件支持
流量防护	<b>单机QPS限流</b>	防止突发流量压垮网关服务器	需要自研	插件支持
	<b>根据Path/Header等限流</b>	支持针对Path/Header等字段设置限流策略	需要自研	插件支持
	<b>分布式限流</b>	内置分布式限流服务器，实现高精度的分布式限流	需要自研	需要自研
流量路由	<b>根据Path/Header等控制路由</b>	支持针对Path/Header等字段设置路由策略	需要自研	插件支持
	<b>根据实例权重分配流量</b>	根据权重将一条路由上的流量分到不同实例	需要自研	需要自研
	<b>灰度发布</b>	根据Path、Header和实例权重进行流量灰度	需要自研	需要自研

# TSE 云原生网关相比开源网关的优势

功能类型	功能特性	TSE 云原生网关	开源 Nginx	开源 Kong
高可用、监控和运维	开箱即用	全功能一键创建，简单易用	需要自研	需要自研
	运维简单	一键伸缩、升降配、升级版本，漏洞修复	需要自研	需要自研
		内置基础监控和日志，支持对接 <i>Prometheus</i> 和 <i>CLS</i>	需要自研	需要自研
		腾讯云提供 7X 24 小时紧急问题定位和恢复服务	需要自研	需要自研
	多活容灾	支持同城、跨城多可用区部署	需要自研	需要自研
微服务架构	对接注册中心	对接Nacos/Consul/Polaris/Eureka等注册中心	需要自研	需要自研
	支持流量染色	流量染色用于多测试环境和全链路灰度的路由	需要自研	需要自研
	支持调用链跟踪	支持开启调用链跟踪，对接 <i>SkyWalking</i> 和 <i>TAPM</i>	需要自研	需要自研
自定义插件	插件机制	支持	支持	支持
	插件热更新	支持	不支持	支持
	插件管理平台	支持	不支持	不支持

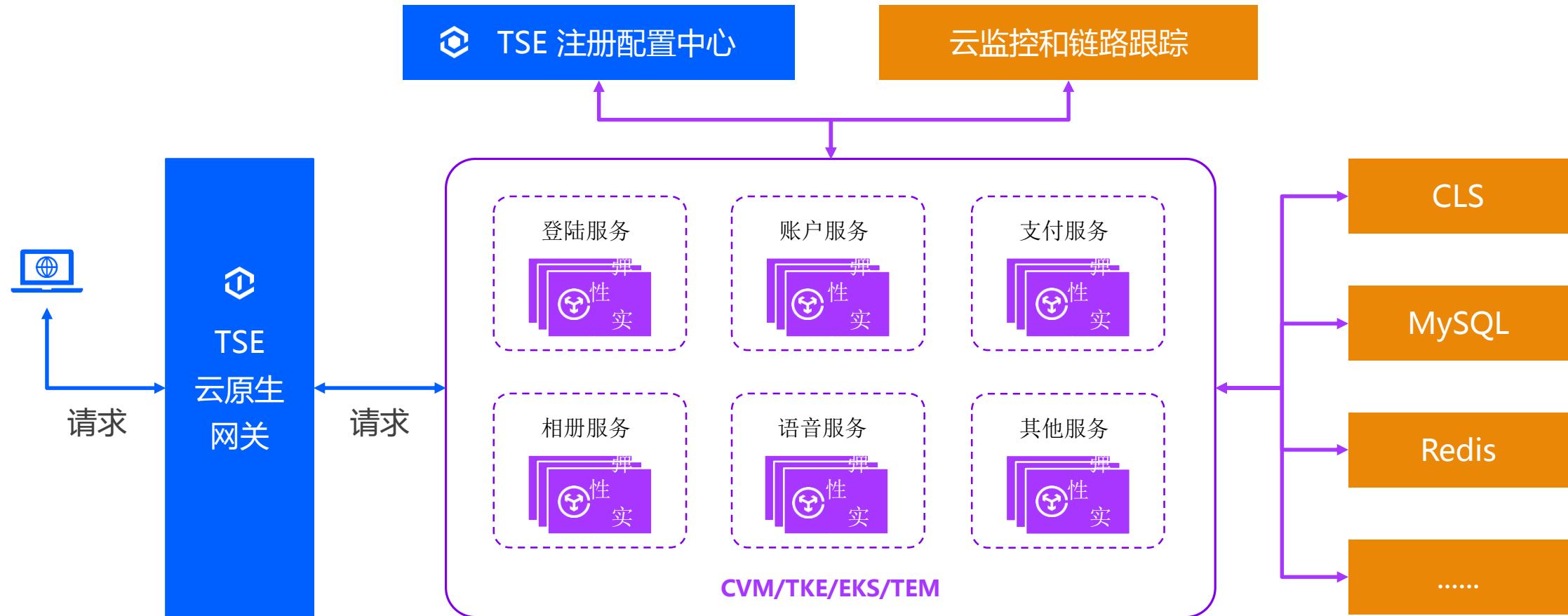


## TSE 应用场景二

构建轻量、高可用、易伸缩的微服务架构



# 微服务核心组件 = 网关 + 注册中心 + 配置中心



# TSE 注册中心：使用场景



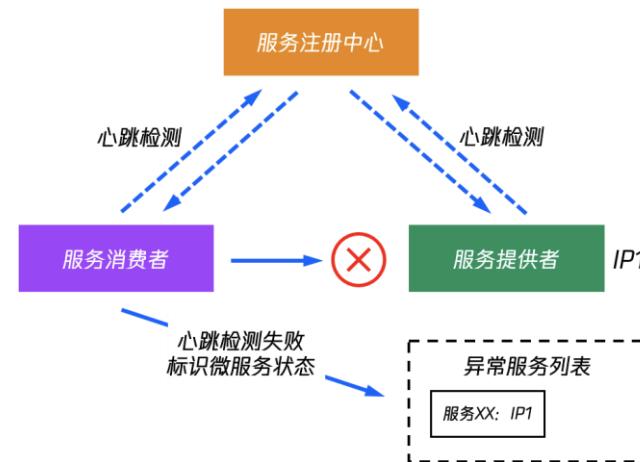
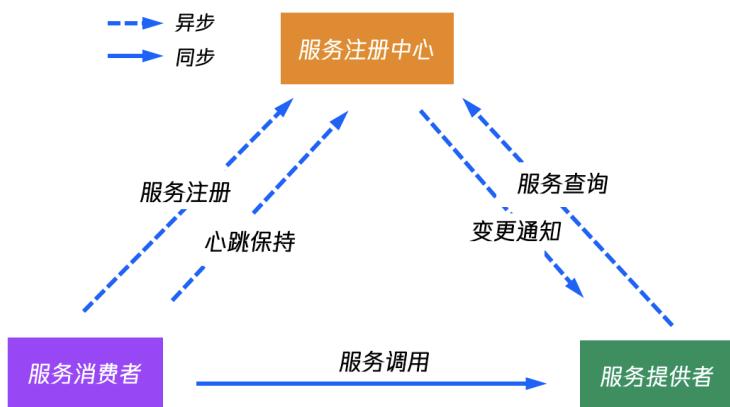
服务注册发现



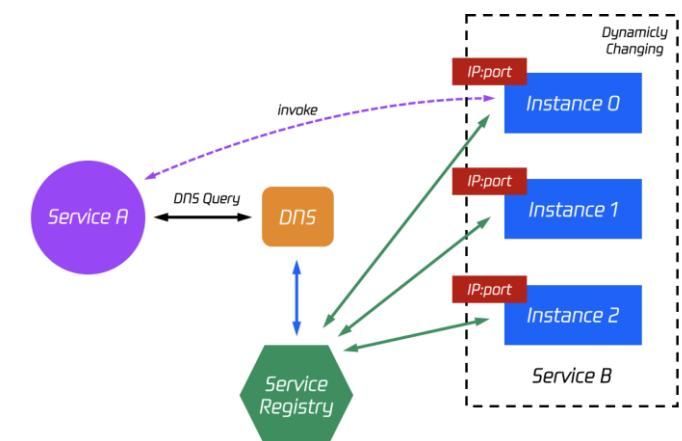
客户端健康检查



内网DNS访问



通常注册中心提供健康检查的能力，基于心跳上报方式维持活性，服务端在x秒内如果没收到客户端的心跳请求，会将该实例设置为不健康，在x秒内没收到心跳，会将这个临时实例摘除。以确保服务的可用性。



注册中心提供DNS接入能力支持将注册中心的数据应该以域名的格式暴露出来，客户端可以直接通过域名寻址后进行直连的服务调用，无需使用CLB转发。

# TSE 注册中心：功能对比

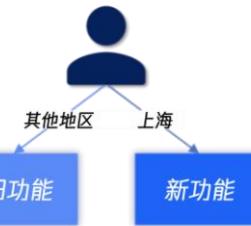
	Zookeeper	Nacos	Etcd	Consul	Eureka
服务注册	CP	CP+AP	AP	CP	CP
健康检查	Keep Alive	TCP/HTTP/MYSQL/Client Beat	Client Beat	TCP/HTTP/gRPC/Cmd	Keep Alive
负载均衡策略	RoundRobin	权重/metadata/Selector	Ribbon	Fabio	RoundRobin
雪崩保护	无	有	有	无	无
自动注销实例	支持	支持	支持	支持	支持
访问协议	TCP	HTTP/DNS	HTTP	HTTP/DNS	TCP
监听支持	支持	支持	支持	支持	支持
配置中心	支持	支持	不支持	支持	支持
跨注册中心同步	不支持	支持	不支持	支持	不支持
服务框架 原生集成	spring cloud	支持	支持	支持	支持
	dubbo	支持	支持	不支持	不支持
	go-zero	不支持	支持	支持	不支持

# TSE 配置中心：使用场景



## 业务配置

- 新上线功能根据地域、用户等信息灰度
- 促销活动开关，中奖率参数动态控制
- 动态公告，控制公告展示以及公告文案



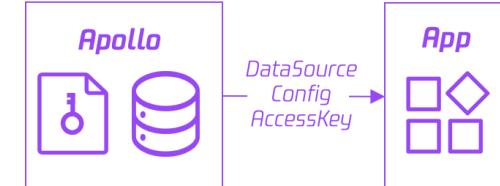
中奖率 40% -> 20%  
活动时间 13:00 ~ 15:00

❗ 公告：我是一条限时公告



## 基础组件配置

- 通过配置中心管理应用数据源，访问私钥等
- 动态调整日志级别
- 作为其它基础组件配置下发通道，例如服务治理规则、xds 协议



# TSE 配置中心：功能特性

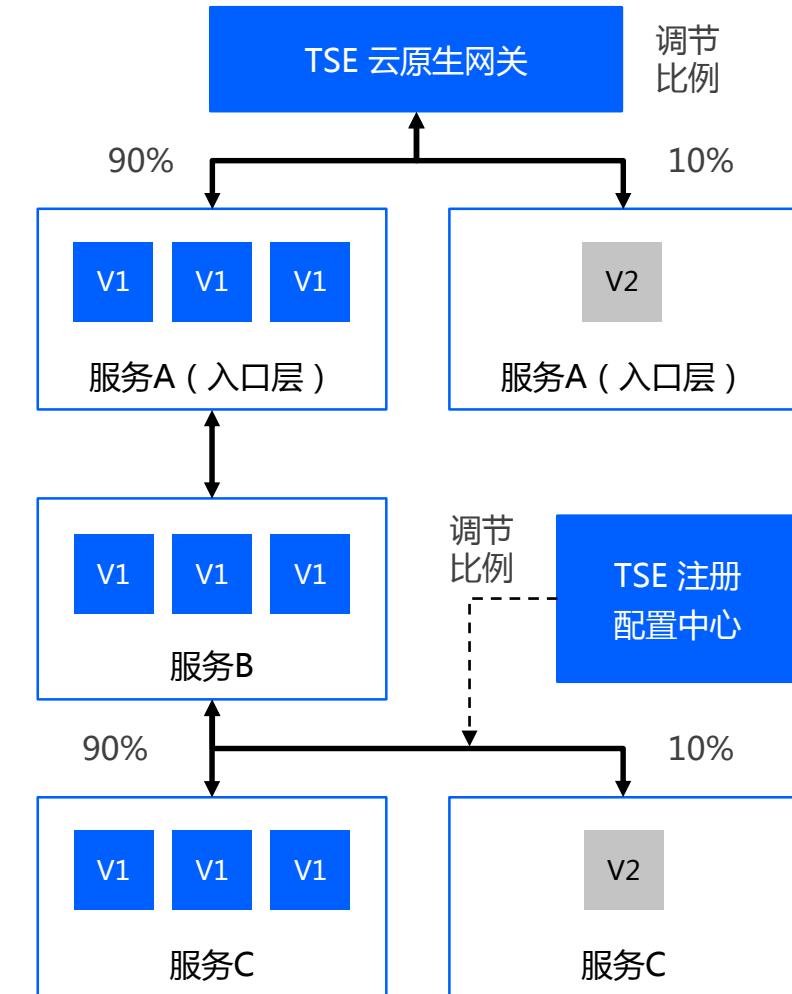
	功能点	优先级	Apollo	Nacos	Zookeeper
配置中心	完善的配置管理平台	高	支持	支持	不支持
	实时推送	高	支持	支持	支持
	版本管理、回滚	高	支持	支持	不支持
	灰度发布	高	支持	支持	不支持
	多环境管理	高	支持 [物理隔离多环境]	支持 [逻辑隔离多环境]	不支持
	可维护性	高	强 [完全无状态]	中 [有状态]	中 [有状态]
	权限管控	高	支持	支持	不支持
	多语言 <b>SDK</b> 支持	高	支持所有主流语言	支持所有主流语言	支持大部分语言
	服务部署复杂度	中	高 [组件多]	低 [单一组件]	低 [单一组件]
	配置继承	中	支持	不支持	不支持

# 发布阶段：实现金丝雀、滚动、蓝绿发布

TSE 云原生网关和注册配置中心可以帮助业务实现蓝绿、金丝雀或者滚动发布：

- 金丝雀发布：对于本次发布的服务，先升级一个实例，如果没有问题，再升级剩余实例
- 滚动发布：对于本次发布的服务，先升级一个/批实例，再分批升级剩余实例
- 蓝绿发布：旧版本实例保持不动，另外部署新版本实例，流量切到新版本实例

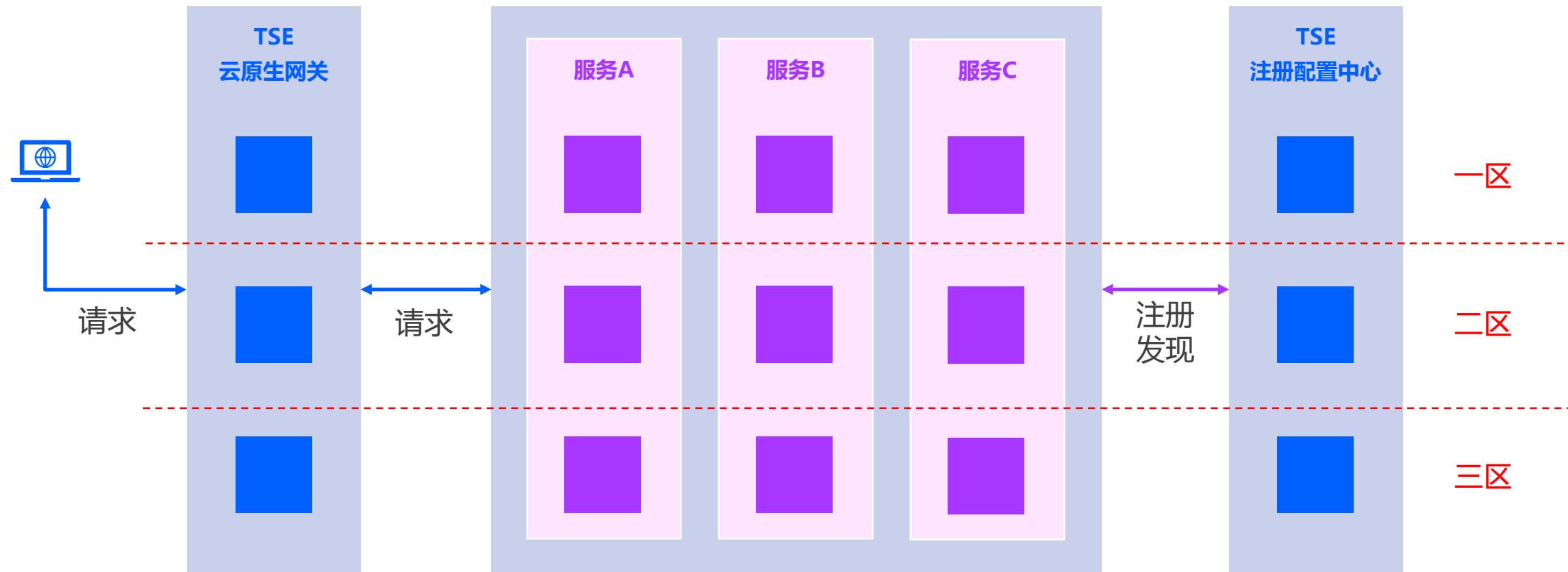
	入口层服务	非入口层服务
指定流量比例进行灰度发布	金丝雀发布 支持	支持
	滚动发布 支持	支持
	蓝绿发布 支持	支持
指定部分用户、地域或者其他条件进行灰度	金丝雀发布 支持	-
	滚动发布 支持	-
	蓝绿发布 支持	-



# 生产阶段：实现多活容灾

TSE 云原生网关和注册配置中心可以帮助业务架构实现多活容灾：

- TSE 云原生网关和注册配置中心的服务端采用同城三可用区部署
- 业务应用可以采用同城多可用区部署。一个应用的多个节点部署在不同的可用区，注册到同一个服务下





## TSE 应用场景三

实现全链路的流量和服务治理



# TSE 北极星 = 注册中心 + 服务网格 + 配置中心

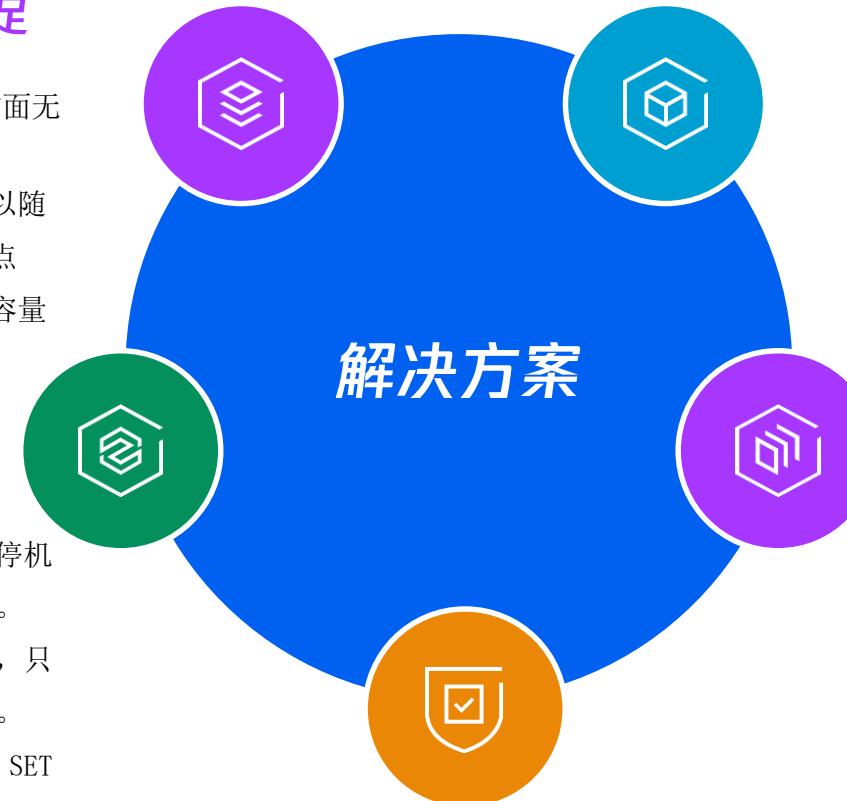
## 开源组件性能/功能不满足

随着业务规模的增加，开源组件在可用性和性能方面无法满足客户诉求。

- 北极星支持计算存储分离、控制面无状态，可以随着接入节点的增加平行扩展，轻松支持百万节点
- 北极星服务端支持数据分片，并发性能和存储容量也可平行扩展

## 东西流量治理

- **优雅上下线**: 服务启动到正式上线或者服务下线到停机这个过程中，切走访问流量，提升服务访问成功率。
- **灰度发布**: 新版本服务发布时，通过流量治理能力，只切换少部分流量，避免新功能不稳定导致全网故障。
- **单元化**: 服务按照不同的逻辑SET进行单元化部署，SET之间会进行逻辑隔离，容灾情况下允许跨SET调度。



## 服务间调用高可用场景

- **故障熔断**, 基于服务调用的失败率和错误数等信息对故障资源进行剔除。
- **访问限流**, 支持服务/接口/标签多级限流能力, 提前限制超过阈值的流量。

## 统一异构系统服务治理

- 打通K8s和框架（SpringCloud/Dubbo等框架）的服务发现和治理体系
- 打通K8s和虚拟机的服务发现和治理体系

## 容灾场景

- 支持同城或者跨城多中心部署，实现跨地域的全局服务发现和流量调度。
- 就近访问和容灾切换。

# TSE 北极星服务治理 – 功能特性

北极星以**服务注册中心、配置中心**为基础，扩展了**服务治理功能**以及相应的控制面

- **基础**：服务发现、服务注册、健康检查、DNS；配置管理、配置下发。
- **扩展**：流量调度（动态路由、负载均衡）、熔断降级（实例、接口、服务三级熔断）、访问控制（限流、鉴权）
- **生态**：多语言、多框架、异构系统；联动上下游生态组件，如：springcloud gateway、nginx等



# 测试阶段：解决多测试环境的流量路由问题

**适用场景：**某个微服务应用，在开发测试时，不需要部署所有的服务，仅部署本次有变更的服务，其他服务通过流量动态路由的方式复用基线环境服务资源。

**优势：**

- 节约资源成本**，开发/测试按需申请，用完即弃
- 提升研发效率**，摆脱大量域名本地绑定 hosts 等配置化工作

## 实现方案

### 1、实例打标

K8s注册场景：在workload上通过添加annotation打上环境标签。

微服务框架注册场景：TSE支持对服务下所有实例进行分组，通过标签能够区分部署的环境。

### 2、流量染色

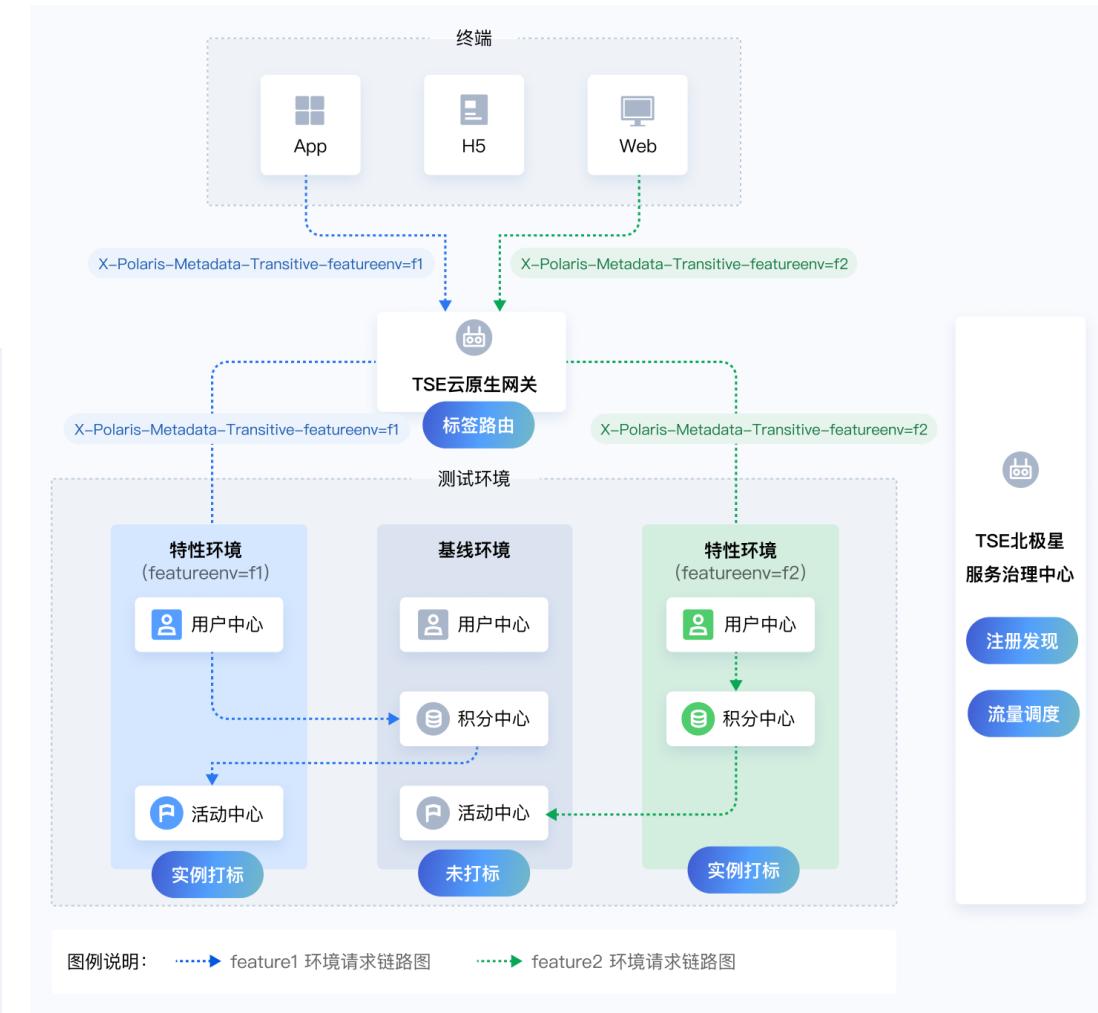
TSE 云原生网关可以对流量特性进行染色。例如：给特定uin的请求进行染色。

### 3、网关到后端服务的流量路由

TSE云原生网关支持标签路由，按照请求中的测试环境信息进行动态路由。

### 4、后端服务与服务间的路由

北极星治理中心支持链路上各服务能够根据请求流量特征对不同测试环境中的服务进行动态路由。

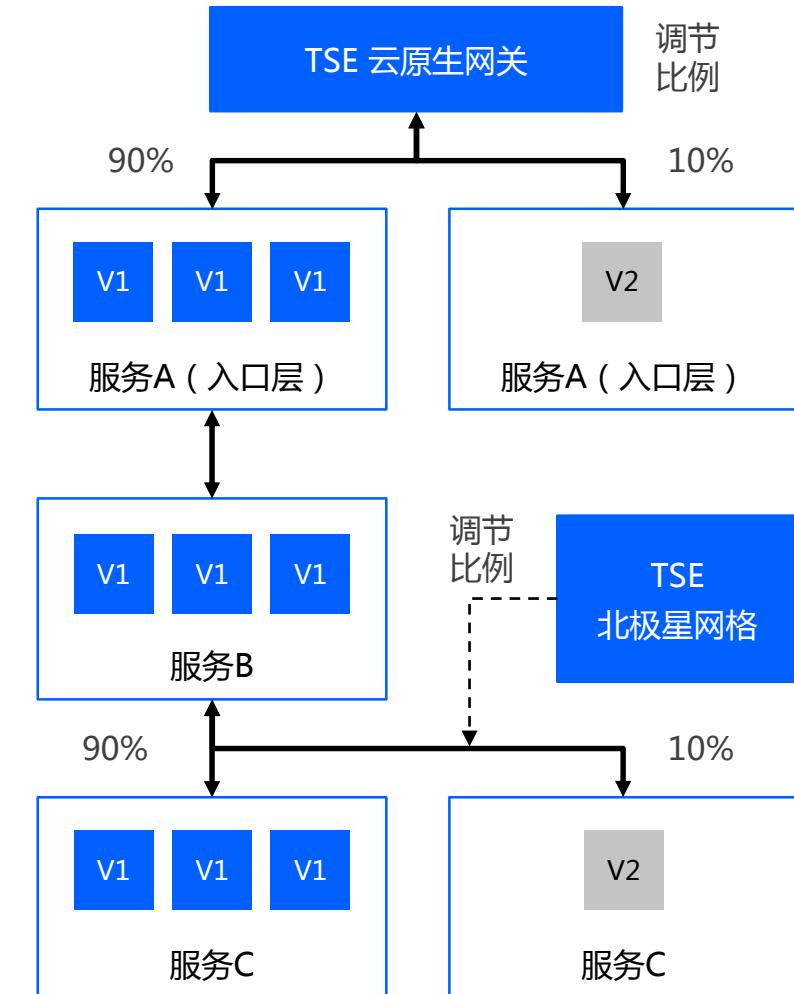


# 发布阶段：实现金丝雀、滚动或者蓝绿发布

基于 TSE 云原生网关和注册配置中心可以实现蓝绿、金丝雀或者滚动发布：

- 金丝雀发布：对于本次发布的服务，先升级一个实例，如果没有问题，再升级剩余实例
- 滚动发布：对于本次发布的服务，先升级一个/批实例，再分批升级剩余实例
- 蓝绿发布：旧版本实例保持不动，另外部署新版本实例，流量切到新版本实例

	入口层服务	非入口层服务
指定流量比例进行灰度发布	金丝雀发布 支持	支持
	滚动发布 支持	支持
	蓝绿发布 支持	支持
指定部分用户、地域或者其他条件进行灰度	金丝雀发布 支持	支持
	滚动发布 支持	支持
	蓝绿发布 支持	支持



# 发布阶段：实现全链路灰度

- 全链路隔离流量泳道
- 端到端的稳定环境
- 流量一键切流
- 可观测能力

## 实现方案：

### 1、实例打标

- K8s注册场景：在workload上通过添加annotation打上版本标签。
- 微服务框架注册场景：TSE支持对服务下的所有实例进行分组，通过标签能够区分版本。

### 2、流量染色

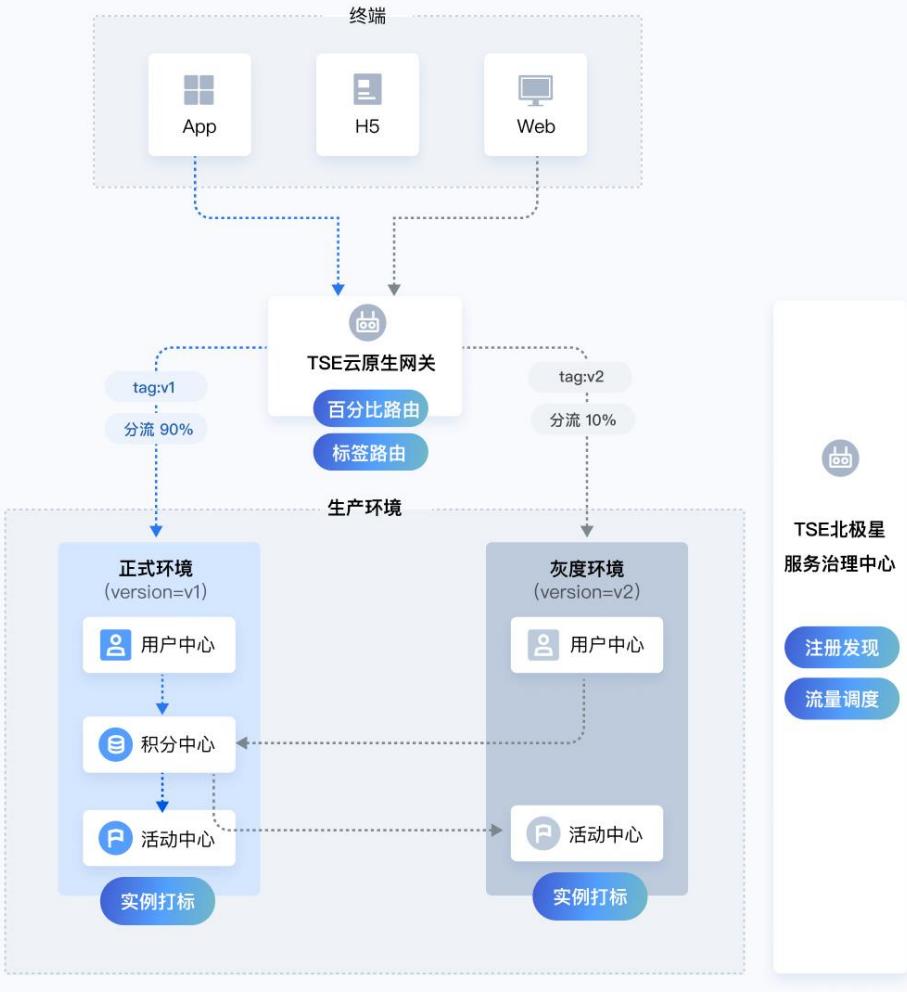
TSE 云原生网关可以对流量特性进行灰度染色。支持**动态染色与静态染色**两种方式。

### 3、网关到后端服务的流量路由

TSE云原生网关支持标签路由，按照请求中的服务版本信息进行流量转发。

### 4、后端服务与服务间的路由

北极星治理中心支持链路上各服务能够根据请求流量特征进行动态路由。



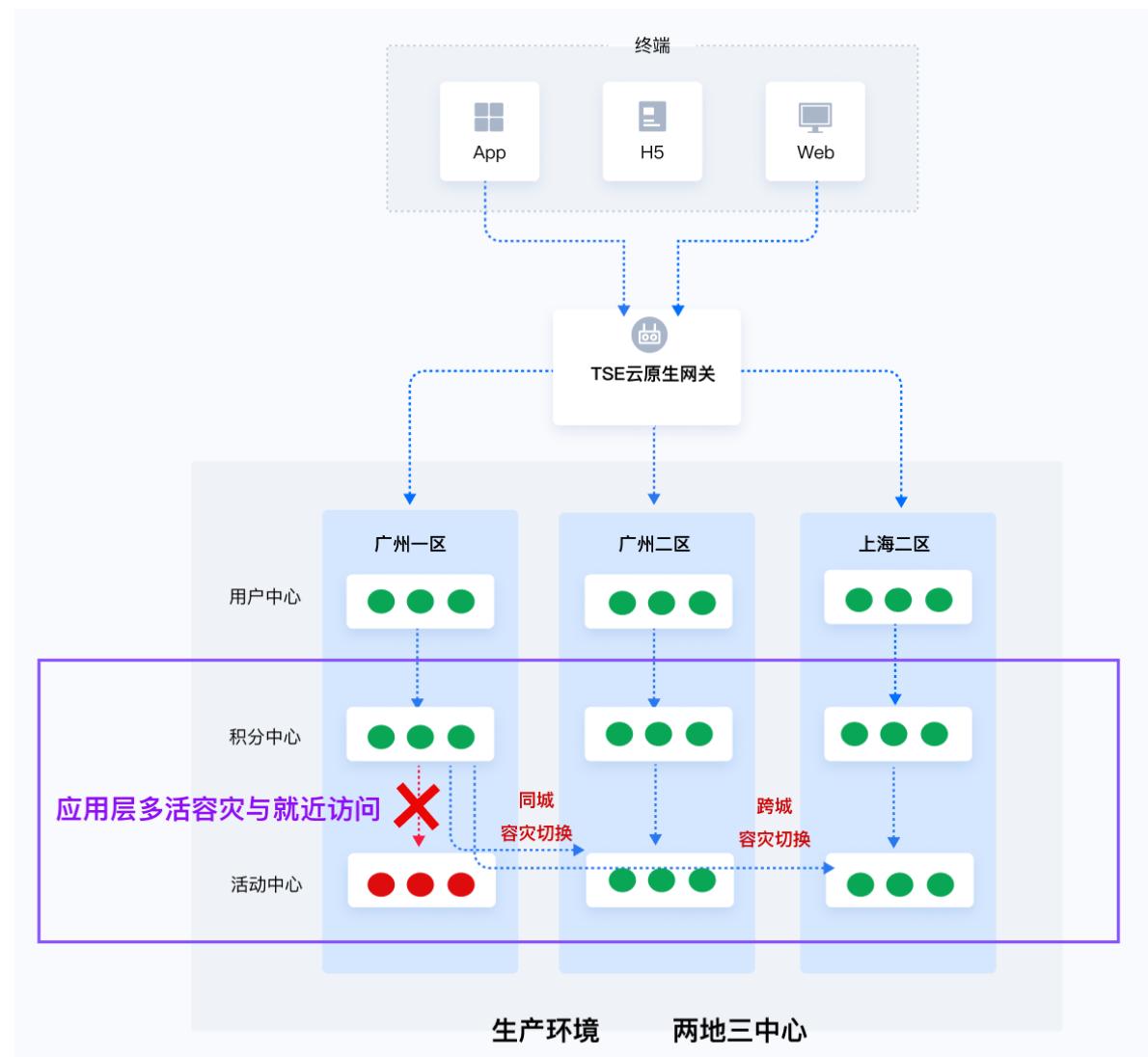
# 生产阶段：实现多活容灾和就近访问

TSE通过云原生网关和北极星服务治理中心提供接入层与应用层的多活容灾与就近访问。实现故障快速恢复、容量快速扩容。

- TSE支持自动获取服务实例的地域信息
- 支持自动根据地域信息进行就近路由
- 支持跨可用区、跨地域容灾切换



支持黎明觉醒全球开服，实现跨zone、跨地域容灾，就近访问。



生产环境    两地三中心

# 生产阶段：支持单元化架构的路由

## 适用场景：

- 资源无法满足横向扩展需求
- 业务可支持SET代码改造
- 业务的访问用户在地理位置上分布较广，希望解决地理位置问题

## 单元化方案：

- 业务系统部署在多地域的不同数据中心中。
- 通过SET化实现不同区域的流量在本区域内闭环操作。
- 某SET出现故障后业务流量切换至其他SET。

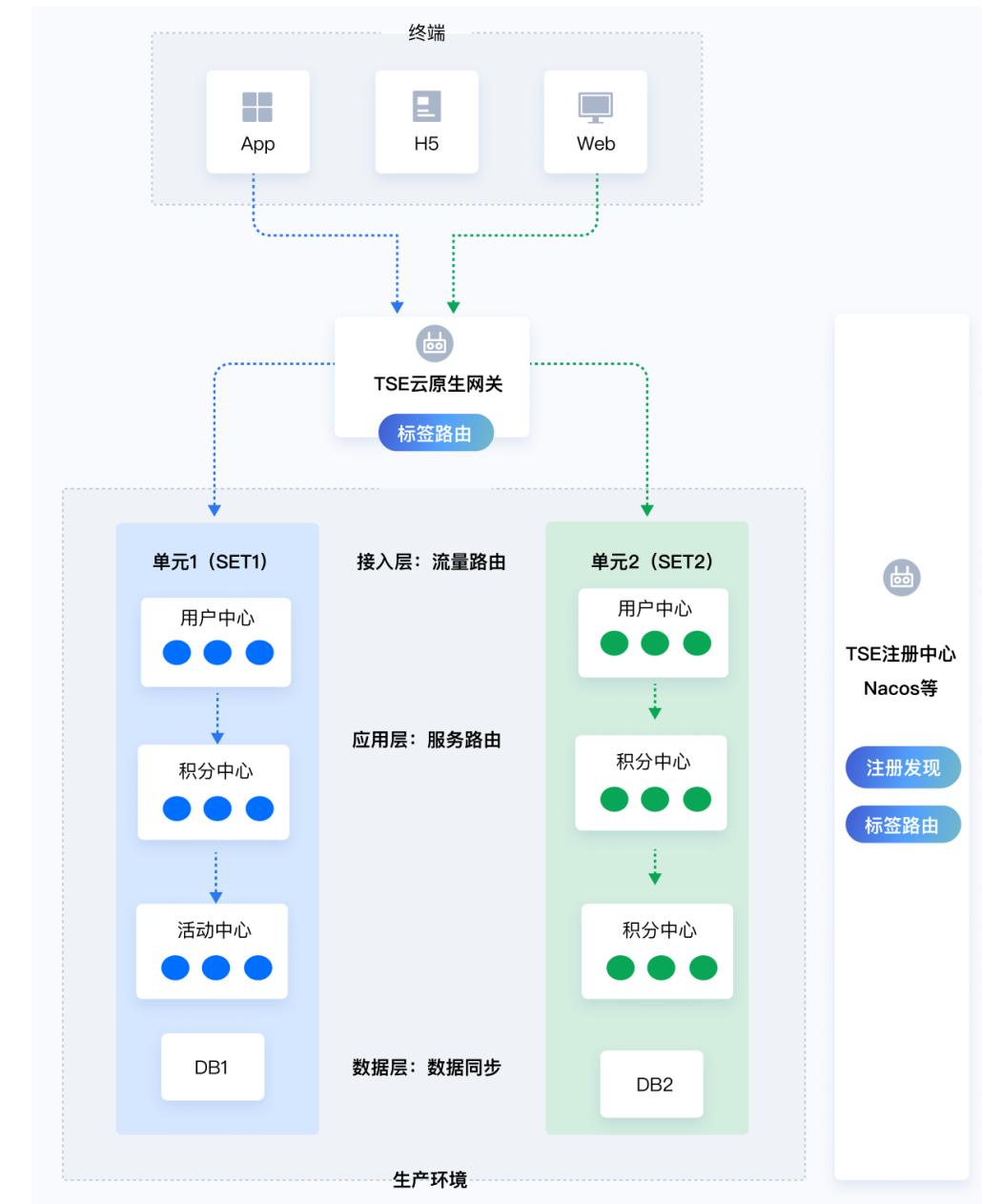
## 实现方案：

- 实例打标**：TSE支持对服务下的所有实例按照单元（SET）进行分组，通过标签能够区分单元（SET）。
- 动态路由**：TSE网关和北极星支持根据请求流量特征对不同单元模块（SET）中的服务进行动态路由。
- 隔离**：支持SET间服务调用的强隔离。



微信支付

微信支付借助北极星成为具备跨机房多活秒级容灾能力和快速伸缩扩容能力的金融级安全可靠的交易系统。



# 生产阶段：限流场景

支持接入层服务流量限流和服务间调用限流场景

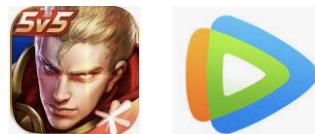
## 多维度精细化限流能力

- 支持服务/接口/标签的限流能力
- 支持秒、分钟、小时、天等时间微服的限流能力。

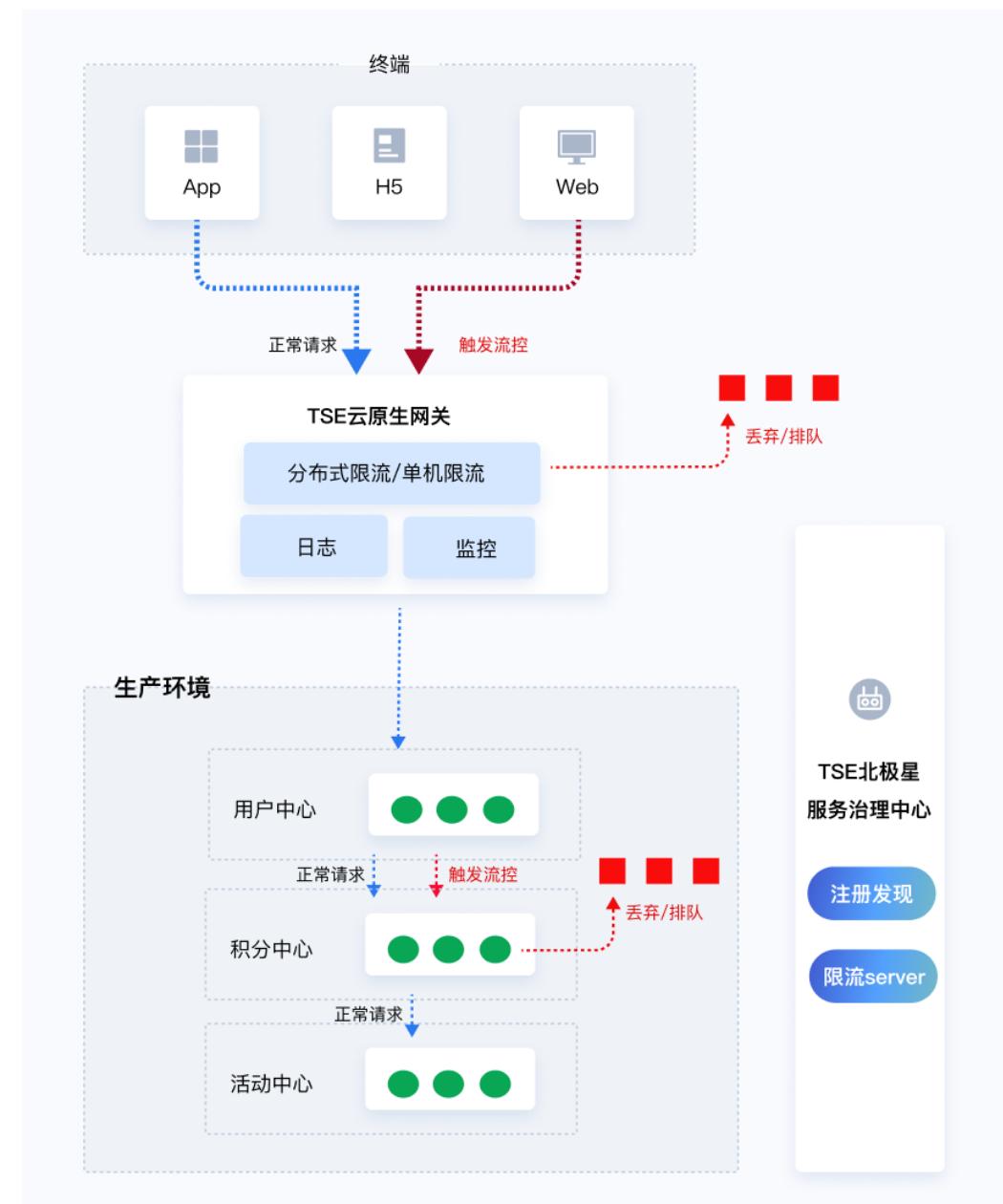
## 应用场景

- 单机限流：针对单个被调实例级别的限流，流量限额只针对当前被调实例生效，不共享。
- 分布式限流：针对服务下所有实例级别的限流，多个服务实例共享同一个全局流量限额。

## 清晰的流量监控、日志能力



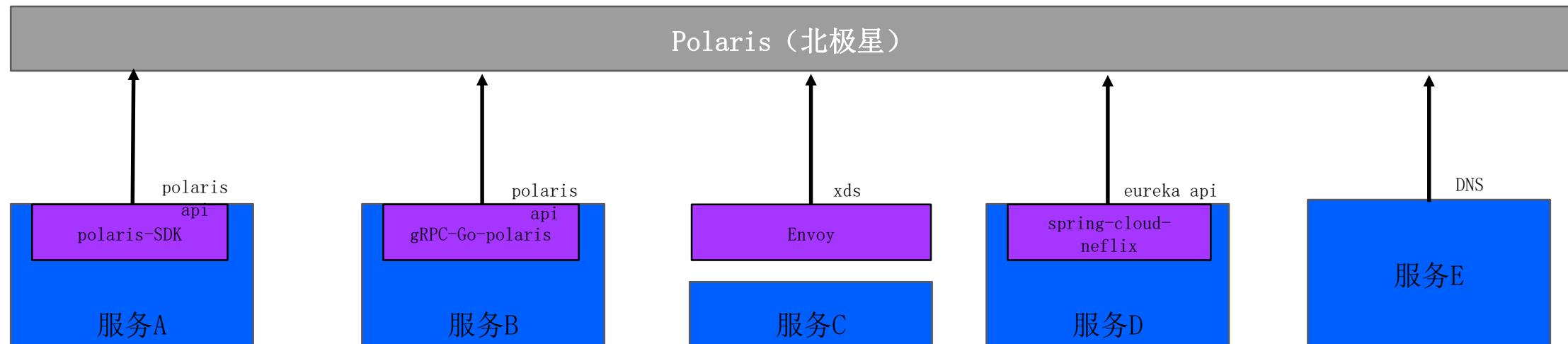
王者荣耀/腾讯视频每年都会举行运营活动，比如点券领取等，为了防止资源被突发流量冲垮，依赖北极星/云原生网关设置了多级流量管控策略，保障运行活动可以顺利进行。



# 支持的接入方式

北极星注册配置中心支持以下5种接入方式：

- 通过多语言SDK接入，应用代码需要引入北极星SDK。支持语言：Go , C++, Java , PHP , LUA
- 通过框架组件进行接入，应用只需要更新框架依赖。支持框架：Spring Cloud, gRPC, dubbo-go, go-zero, CloudWeGo/kitex等
- 通过sidecar进行接入，应用只需要更换sidecar镜像。支持：Envoy
- 通过注册中心API兼容方式进行接入，应用只需要修改服务端地址。支持注册中心：Eureka
- 通过DNS的方式进行接入，应用只需进行DNS接入相关配置。





# TSE 容灾与可用性



# TSE 高可用：同城多可用区容灾，SLA > 99.95%

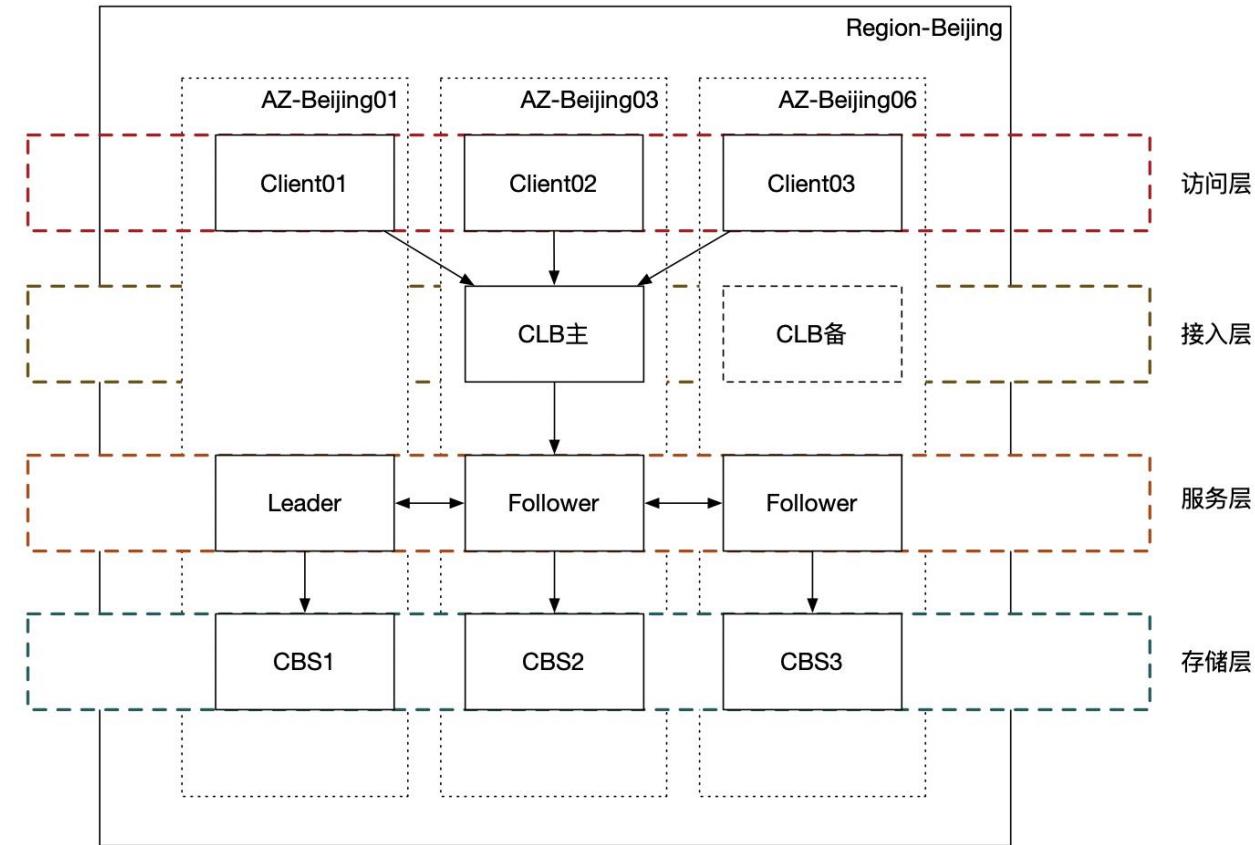
通过多可用区多实例部署，提供了至少N-1实例的容灾能力，接入层、服务层、存储都有高可用部署，满足客户生产需求。

基于CBS

数据持久化

三副本存储

数据高可用

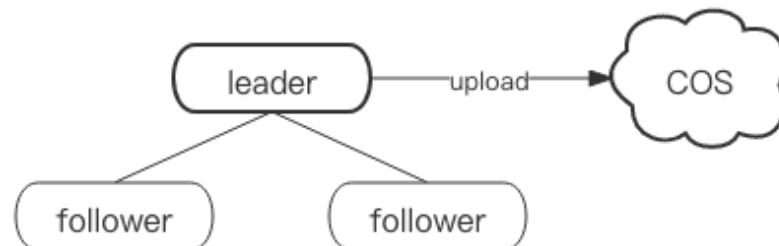


高可用架构图

# 可用性：数据备份和恢复

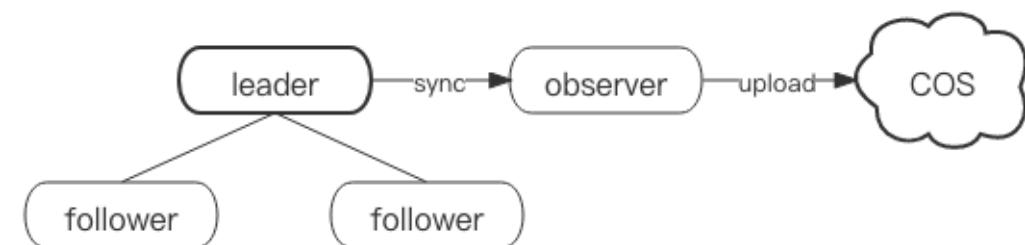
## 冷备

- 上传leader快照日志文件到COS中。
- 故障后可回溯至该快照文件的状态。



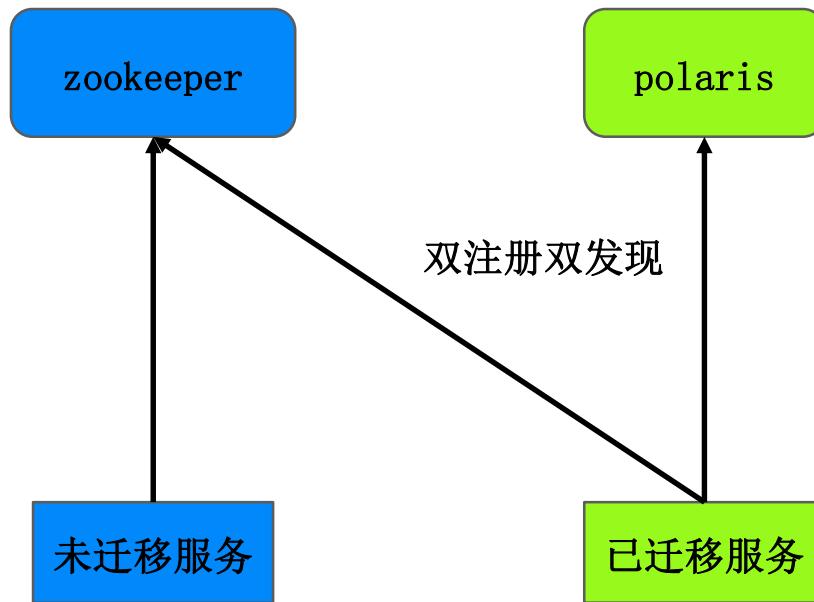
## 热备

- 变更observer快照输出逻辑，定时生成快照日志文件，并持久化到COS中。
- 通过observer输出快照文件，避免通过leader输出快照对性能产生影响；
- 自定义快照生成策略，故障恢复更灵活。



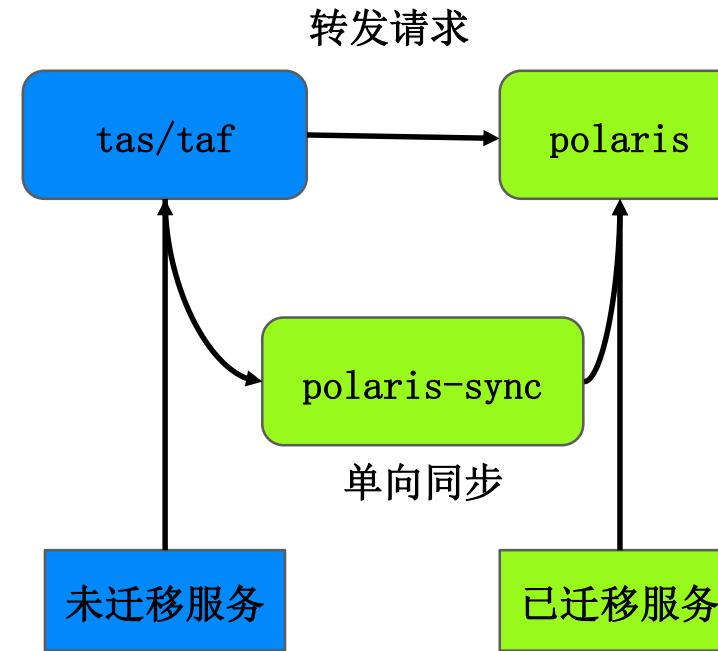
# 平滑迁移实践

## 方案1：客户端双注册双发现



适用场景：dubbo/SpringCloud等Java应用，  
北极星提供无侵入java-agent机制，用户无需  
修改代码即可使得已迁移服务的进行双注  
册双发现

## 方案2：单向同步+转发/协议兼容



适用场景：公司内部tas/taf/cl5等存量迁移使  
用单向同步+转发方式，公司外部eureka等  
存量迁移使用协议兼容的方案，适用于非  
Java应用迁移

# 注册中心故障案例

## 背景▶▶▶

存量的有些pulsar集群使用的是单机版本的zookeeper。该zk上有200W的znode，和10W的watch数，量还是比较大的。

存在两个问题：

1. 内存不够，触发fullgc频率较高，且时间较长，超过session时间，触发broker重启。
2. 单机版本的zookeeper，读写集中到一台server上，造成pulsar元数据写入的时延太高，影响了pulsar的整体吞吐。

所以，运维需要把单机转为使用TSE微服务引擎。

## TSE的优势▶▶▶

- CMQ的问题，本质上还是pulsar量太大，当前的配置都无法满足。升级规格和调整配置都是人工在参与，整体时间较长。
- 对于调整规格和配置方面，TSE只需要在控制台上修改下，就可以马上完成。整体上是不需要人工参与太多，调整较快。

## 迁移方案▶▶▶

### 迁移方案1

1. 运维第一次是直接启动了两个zk B和C，然后修改了原有单机zk A的配置。
2. zookeeper集群启动的时候，会有一次数据同步，保证server之间数据的一致性。在迁移方案一中，由于没有手动提前同步数据，所以zk B和C的数据依赖zk A的同步。这个同步过程是有最大同步时间initLimit控制的，默认是5秒。
3. 这次迁移失败的原因是，没有手动提前同步数据，200w+的量依靠server之间的同步是会超过默认最大同步时间5秒。导致的问题是，zookeeper无法启动起来，从而也引发了broker的重启。

### 迁移方案2

1. 调整brocker和zookeeper的sessionTimeout的配置，以延长session超时时间，防止自重启。
2. 同步快照日志和事务日志到zk B和C。
3. 修改initLimit和syncLimit的配置。
4. 停止A，修改A、B、C的配置，重启A、B，不启动C。等待集群稳定后，在启动C。



# TSE 客户案例

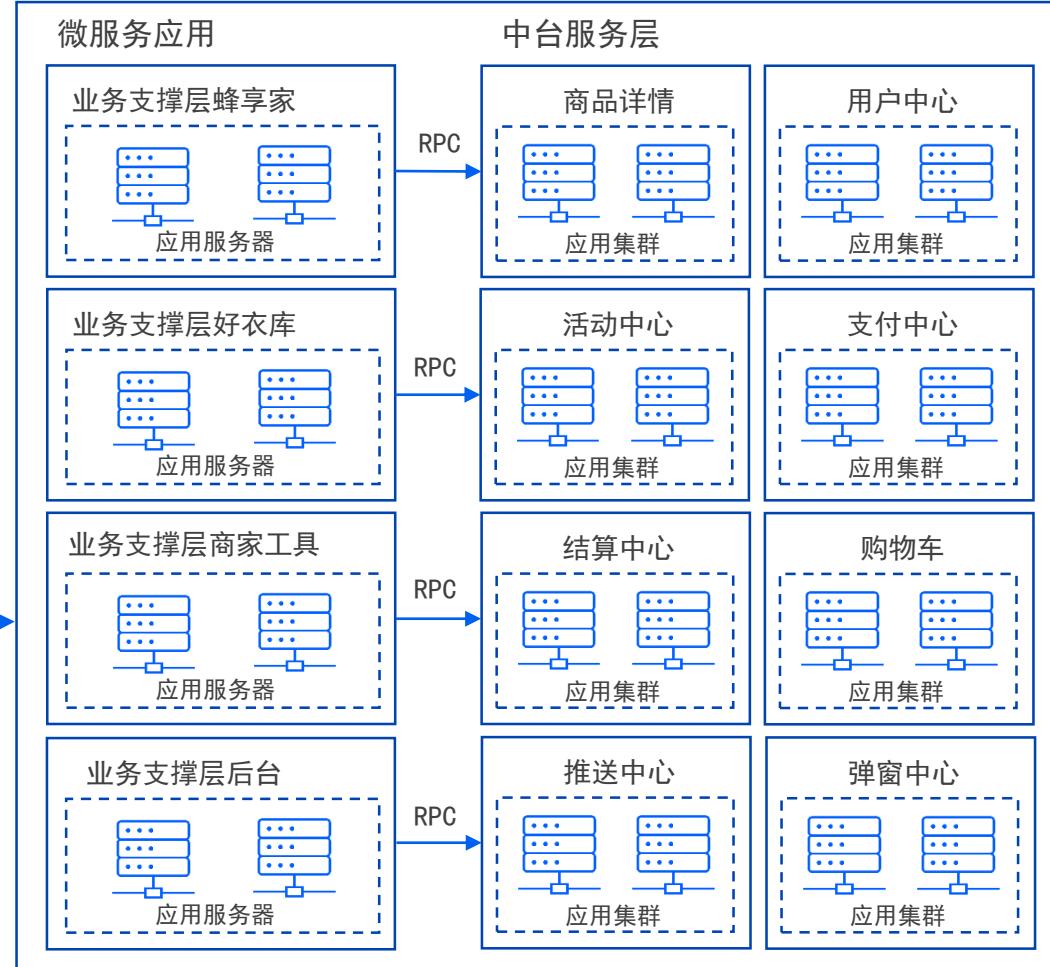
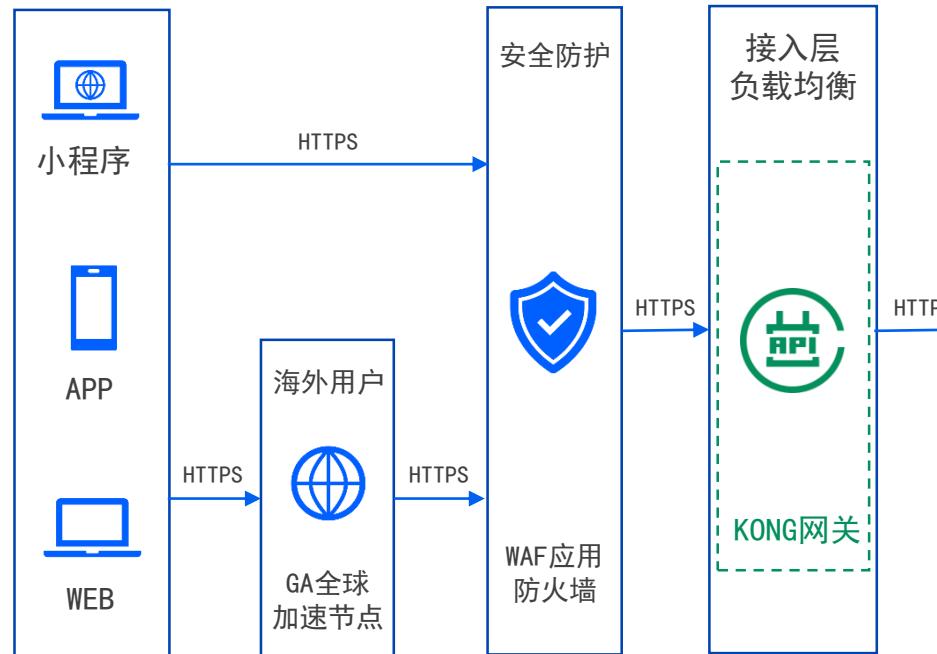


# 微服务架构典型案例：鲸灵科技

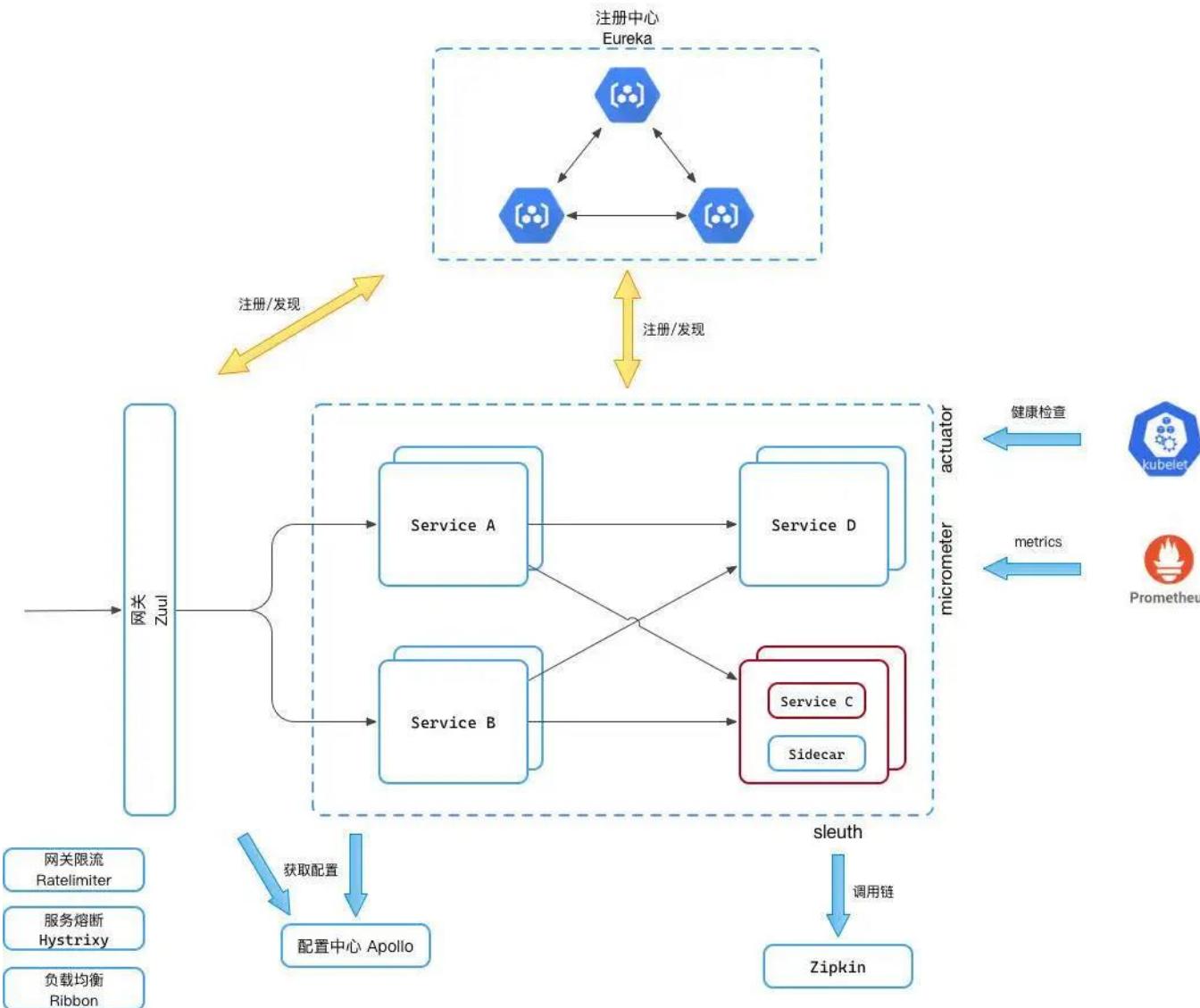
- 鲸灵科技通过使用TSE解决了他们原来管理微服务组件复杂、运维成本高、高可用建设难等问题。
- TSE给他们带来了微服务组件免运维、高可用、一键升降配等价值。

## 鲸灵微服务技术选型

- ✓ 微服务框架: Dubbo
- ✓ TSE: 云原生网关KONG
- ✓ TSE: 注册中心Zookeeper
- ✓ TSE: 配置中心Nacos
- ✓ TSE: 服务治理北极星



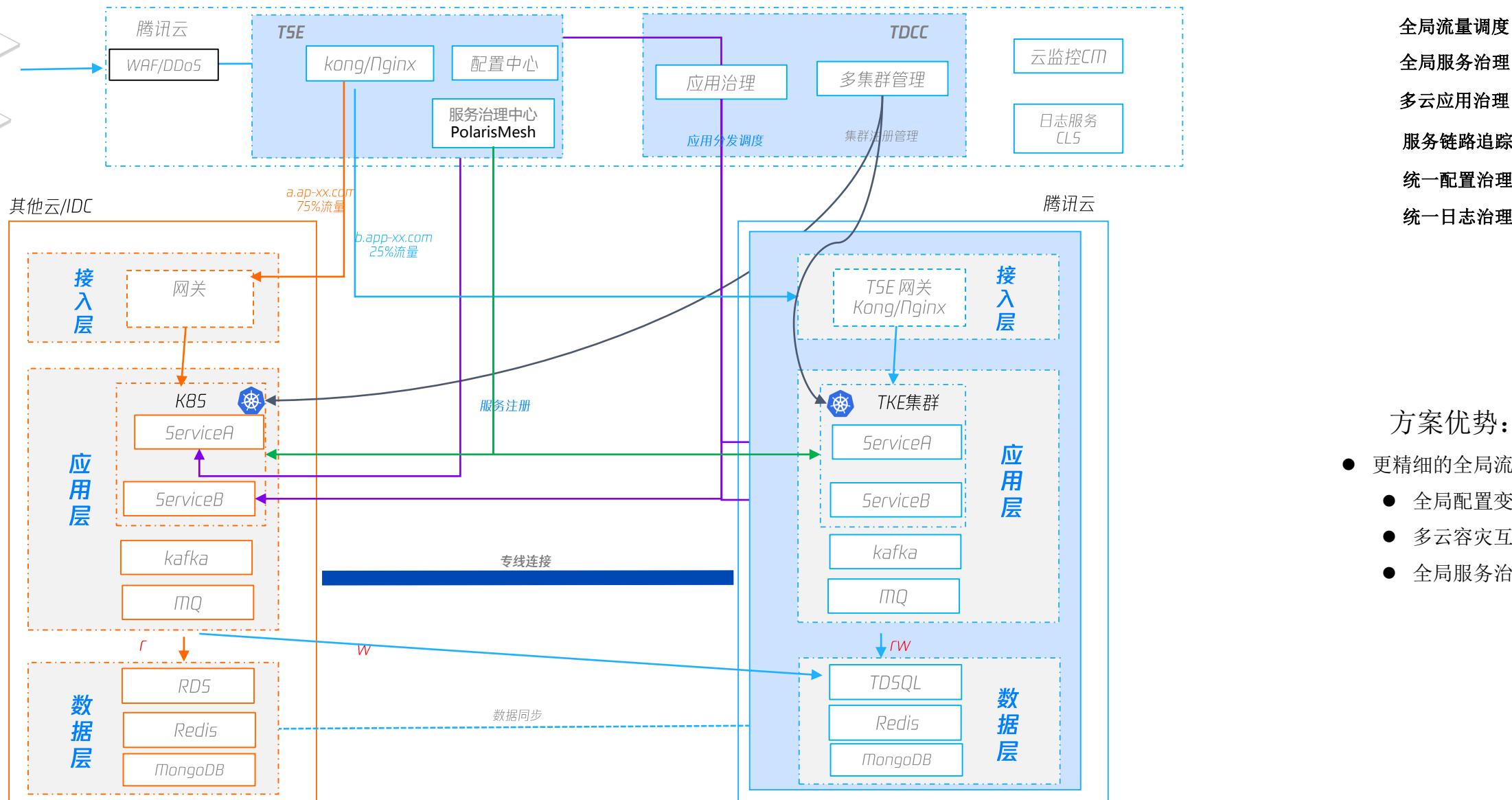
# 微服务架构典型案例：小鹏汽车



## 小鹏汽车微服务技术选型

- ✓ 微服务框架: SpringCloud
- ✓ 网关: Zuul
- ✓ TSE: 注册中心Eureka
- ✓ TSE: 配置中心Apollo
- ✓ 服务治理: Hystrix (熔断)、Ribbon (负载均衡)、RateLimiter (限流)
- ✓ 可观测性: SpringCloud Sleuth(客户端)、Zipkin (服务端)

# 基于云原生构建业务高可用架构，实现多云战略



200+活跃用户



腾讯音乐娱乐集团  
TENCENT MUSIC ENTERTAINMENT



极客学院  
jikexueyuan.com



POP MART



# 北极星客户案例

截止2021年9月，在线节点超过  
**5,000,000个**

- 每日服务调用量超过30万亿
- 接口的调用成功率超过99.999%
- 覆盖腾讯内部90%以上的业务部门



# 感谢倾听