
# 京东云跨集群大规模应用管理实践

JDCloud

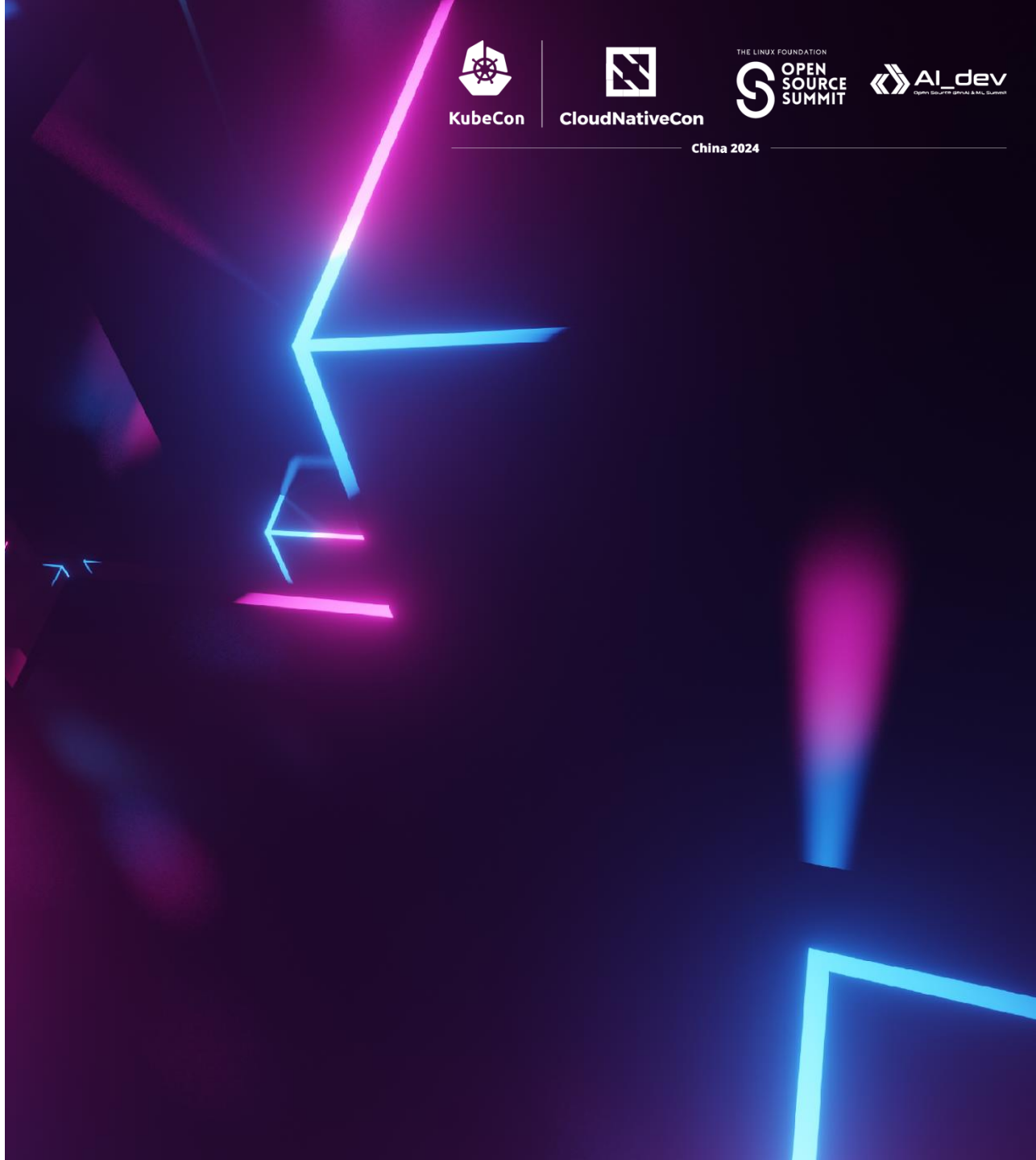
2024

# 优化人员简介

王晓飞
JDCloud
云原生开发工程师
wangxiaofei67@jd.com

# 目　录



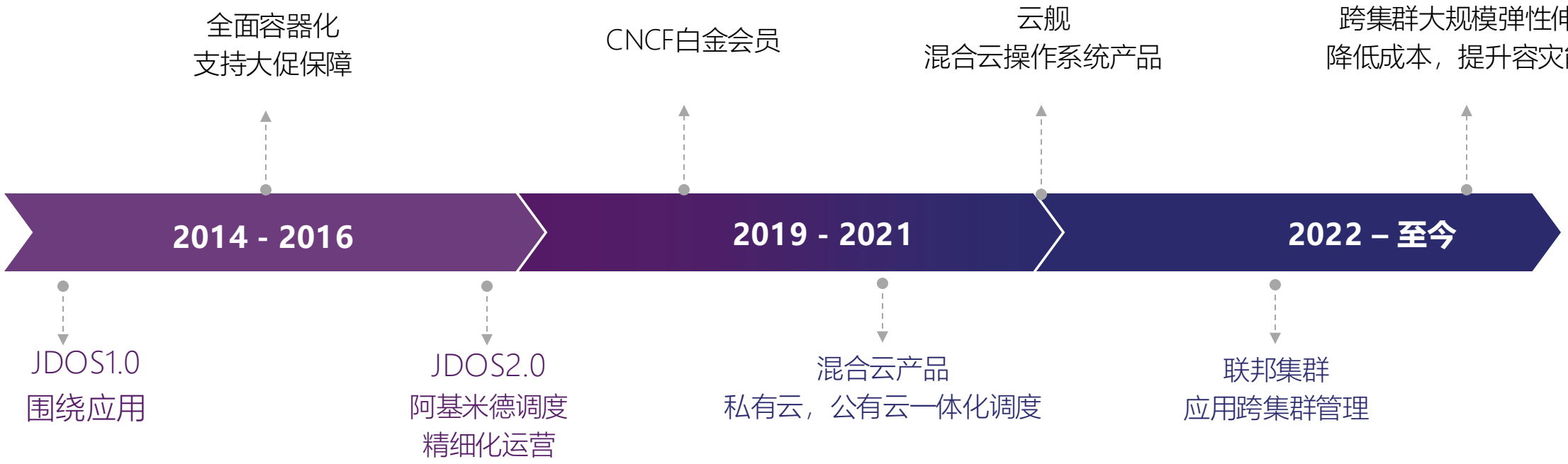


# 01

# 京东云容器化发展历程

# 京东容器化发展历程






**最早实施**
大规模混合云容器集群

**容器化最彻底**
业务应用，中间件，数据库，大数据，混部，AI

**100%覆盖**
承载京东零售，物流，金融，健康，工业，保险等业务

**效率提升**
实现多集群管理下的效率，成本，弹性的最优解

# 一切围绕应用

## 单集群困境

- 故障时爆炸半径过大，影响应用SLA。
- 应用跨集群弹性伸缩存在难度，人工运维。

## 跨集群调度

- 资源统一调度
- 多分组灵活调度策略

## 跨集群弹性伸缩

- 跨集群弹性伸缩
- 一键迁移

## 多活和高可用

- 应用多活
- 跨集群网络

# 02

# 联邦集群

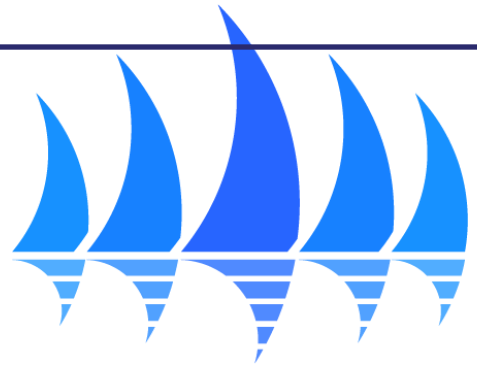
# 产品化



京东云

京东云·云舰

https://www.jdcloud.com/cn/products/yunjian

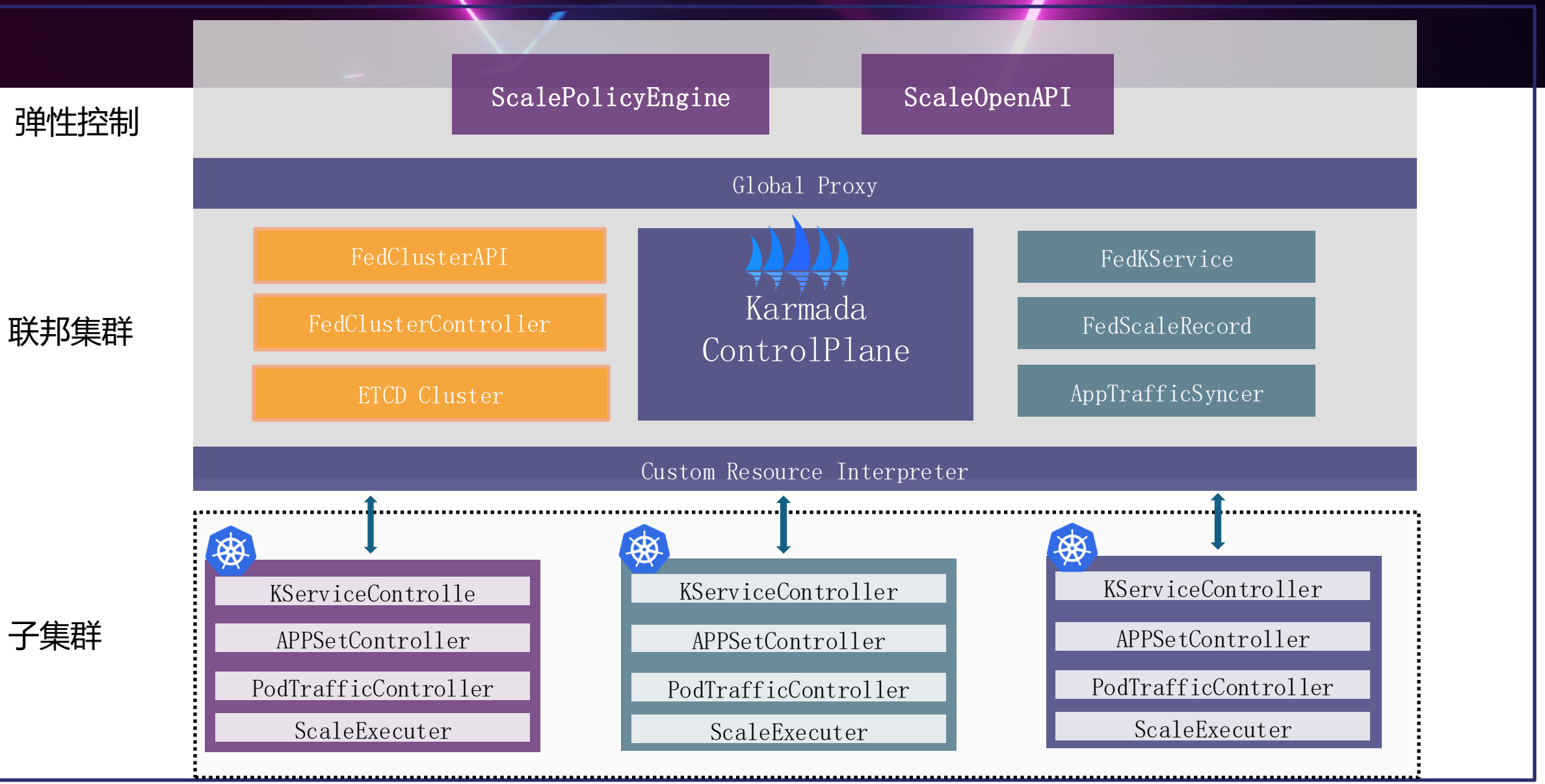


基于Karmada进行深度开发和增强

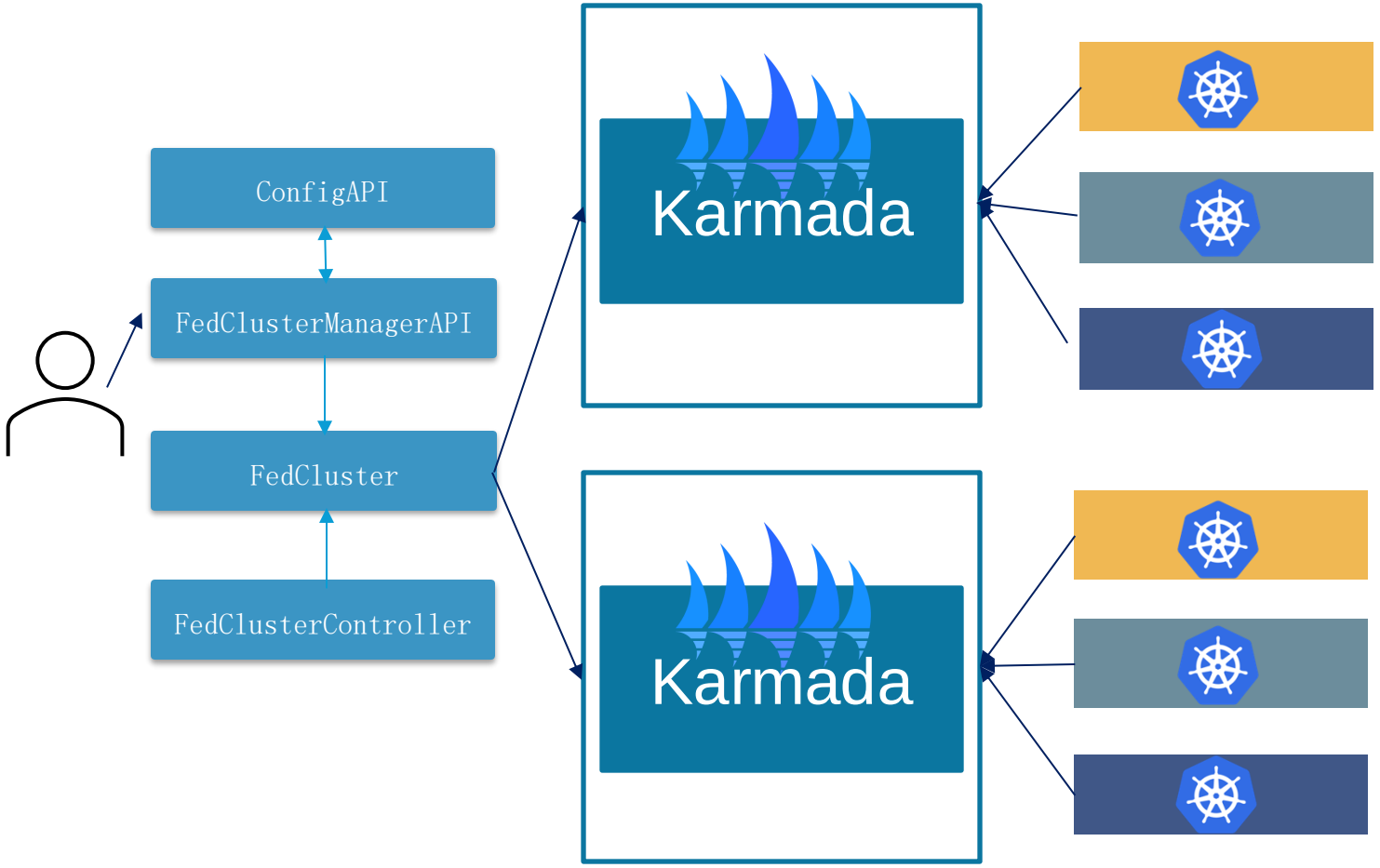
https://github.com/karmada-io/karmada

# 整体架构





# 联邦集群服务部署




**新建**

FedClusterManagerAPI处理用户创建联邦集群请求，和ConfigAPI 服务交互，获取创建上下文和参数。创建FedCluster CR。

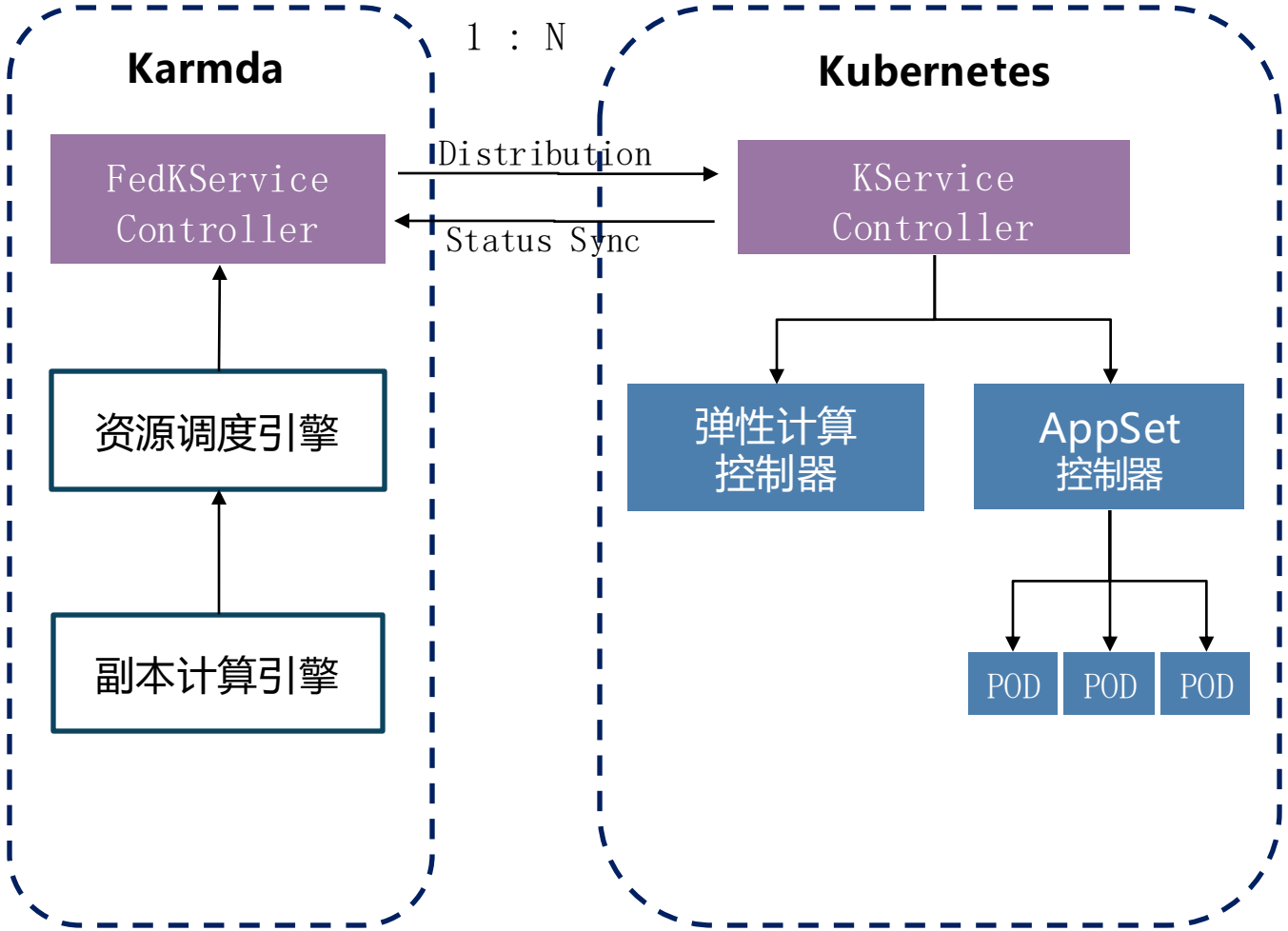
**运行**

FedClusterController监听到FedCluster，按照Spec声明，在指定Kubernetes集群创建联邦集群
ControlPlane。

**管理**

FedClusterController监听FedCluster，声明Spec加入指定联邦集群，
联邦集群开始管理来自多云的Kubernetes集群。

# FedKService 控制器




**资源调度引擎**

基于Hippo服务，进行多集群资源精准测算，得出资源画像。

**副本计算引擎**

按照资源画像以及调度策略，进行副本多集群拆分。

**弹性计算控制器**

执行弹性策略，执行扩缩容计划。

**AppSet控制器**

直接控制pod的工作负载

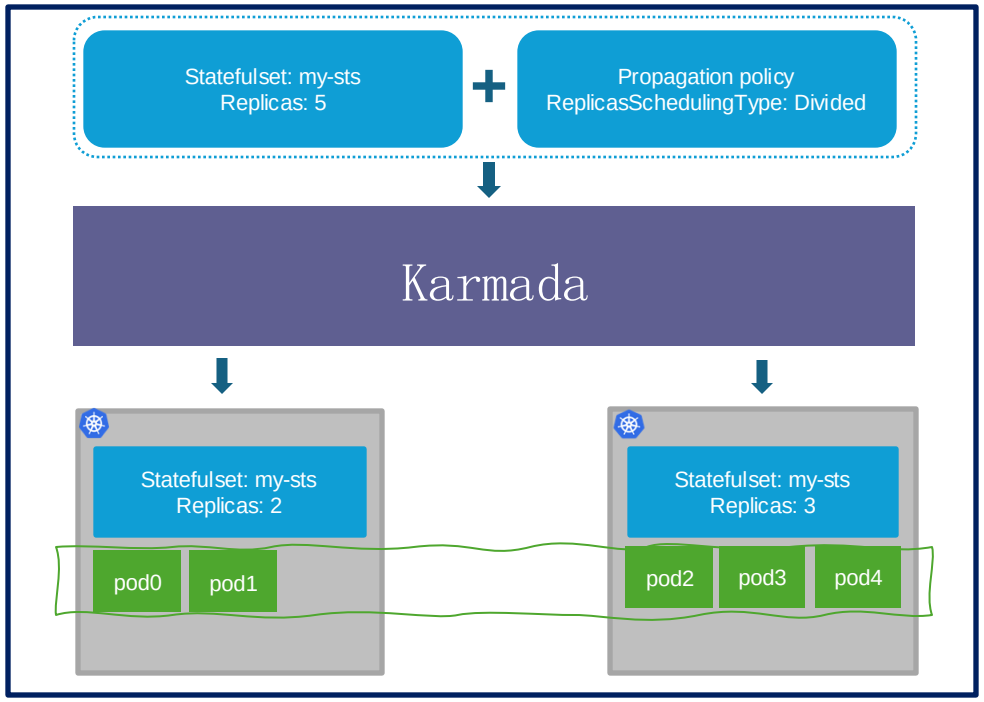
# 联邦增强

## KService status sync

解决KService status资源从子集群同步合并到联邦集群。

## StatefulSet起始序号在多集群控制

在联邦集群使用StatefulSet和单Kubernetes集群序号控制行为相同。

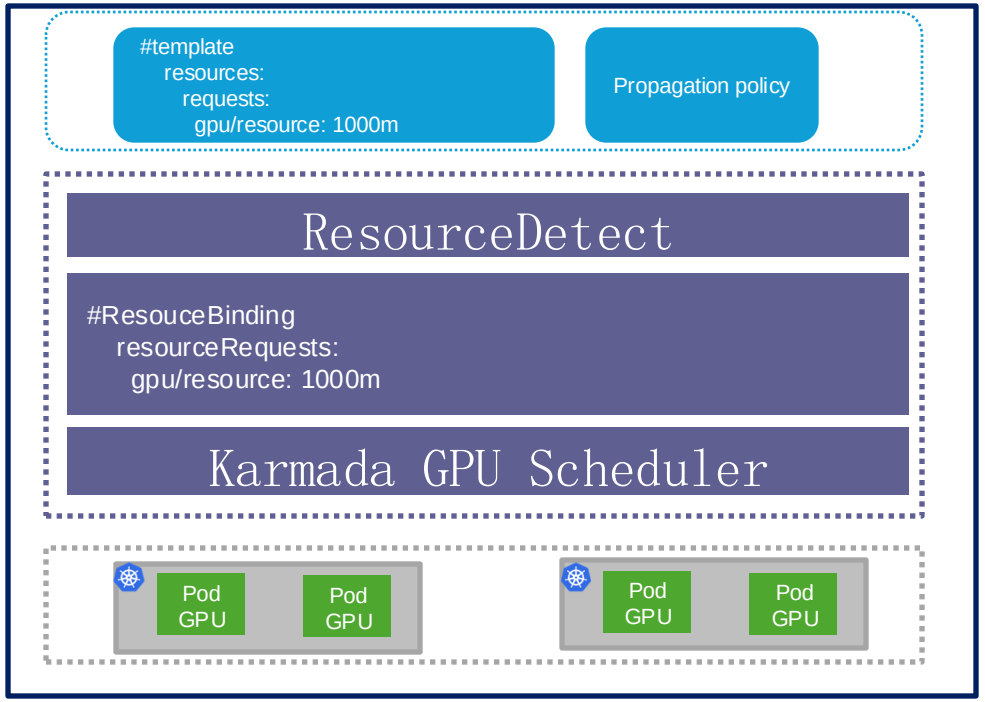



## 全局代理性能提升

1.子集群查询时使用pb协议；　　2.从单独查询，修改为并发查询；
3.存储过程优化，使得存取数据性能更高。　　4.性能提升30%

## Karmada scheduler增强调度

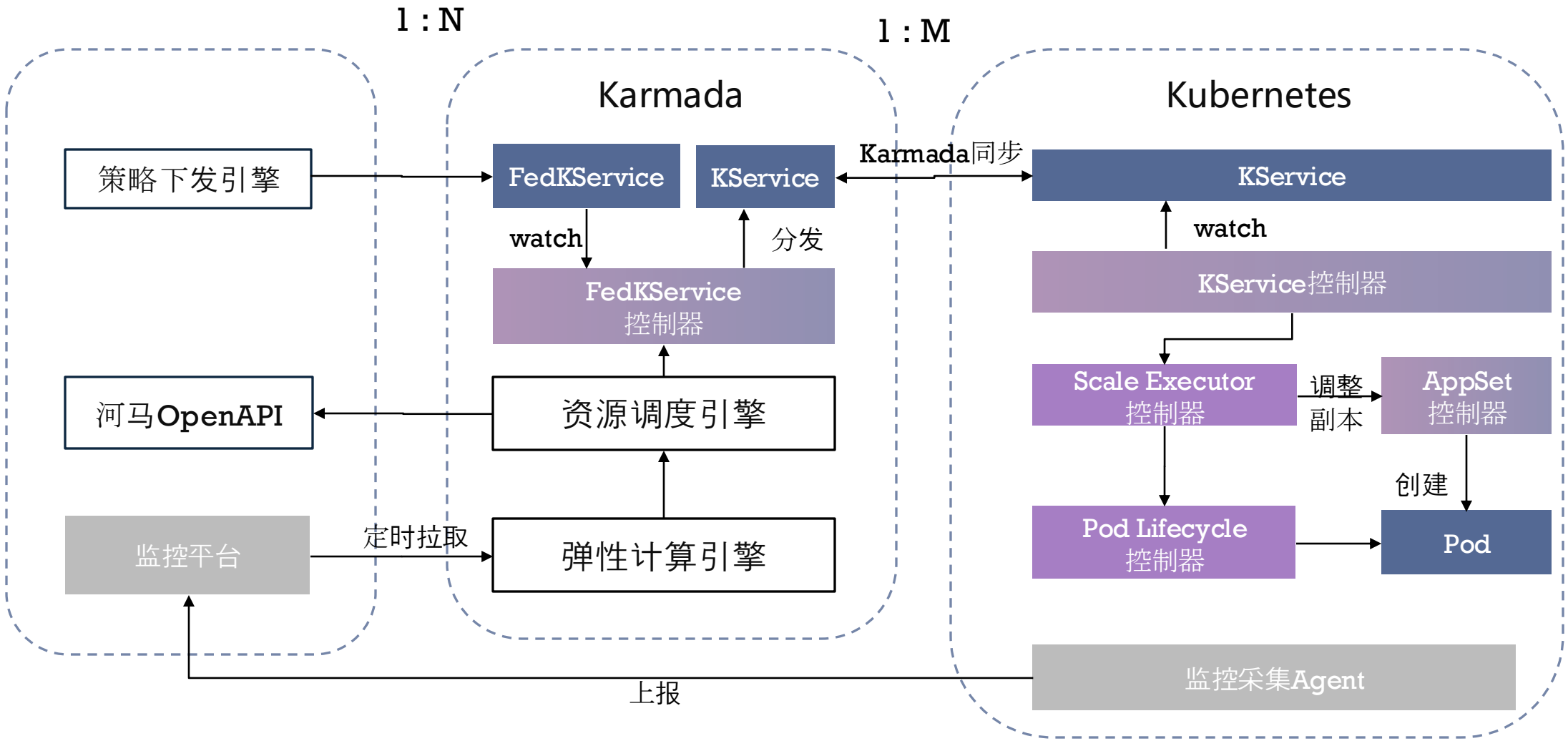
支持GPU集群调度，支持异构集群调度。



# 03

# 跨集群弹性伸缩

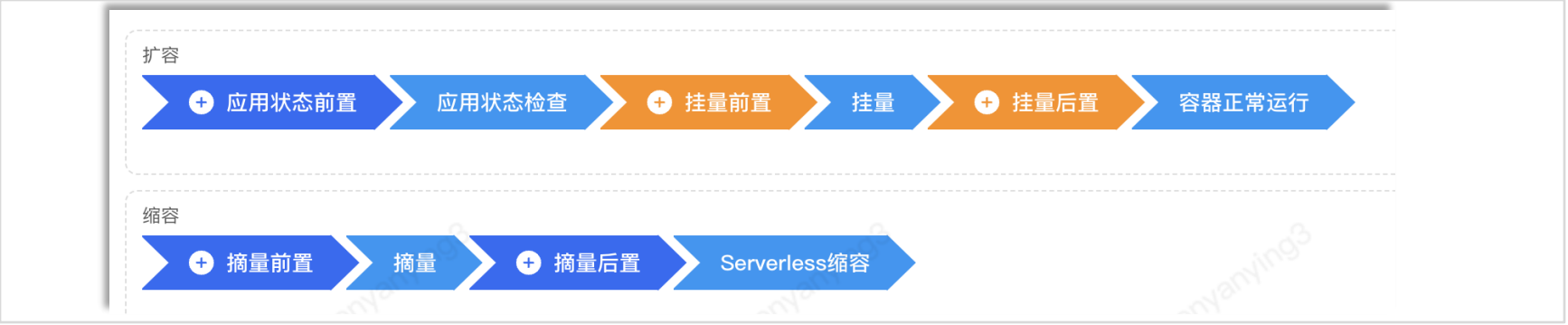
# 弹性伸缩架构


1 : N
1 : M
Karmada
Kubernetes
策略下发引擎
FedKService
KService
Karmada同步
KService
watch
分发
watch
FedKService
控制器
KService控制器
河马OpenAPI
资源调度引擎
Scale Executor
控制器
调整
副本
AppSet
控制器
创建
监控平台
定时拉取
弹性计算引擎
Pod Lifecycle
控制器
Pod
监控采集Agent
上报

# 联邦弹性伸缩过程



| 弹性策略计算 | 资源分发调度 | 集群下发 | 扩缩容执行 | Pod管理 |
|---|---|---|---|---|
| 联邦 | 联邦 | 联邦 | 集群 | 集群 |
| 实时匹配弹性策略，计算应用需要副本数 | 资源测算，分发调度副本 | Karmada下发调度结果 | 根据弹性策略反复执行扩缩容，直至挂量Pod数达到期望值 | 配置化生命周期管理异常Pod GC机制 |



# 弹性伸缩场景



## 故障自动迁移

集群节点故障迁移
集群下线
跨机房容灾迁移

## 大促压测容器热备

流量低时，热备容器压缩为低规格
流量高时，低规格热备容器恢复

## 指标监控弹性伸缩

支持机器性能、方法调用性能指标
个性化定制扩缩容规则

## 定时扩缩容

支持Cron表达式


# 生产实践分享：弹性伸缩和人工部署的切换



## 弹性伸缩场景

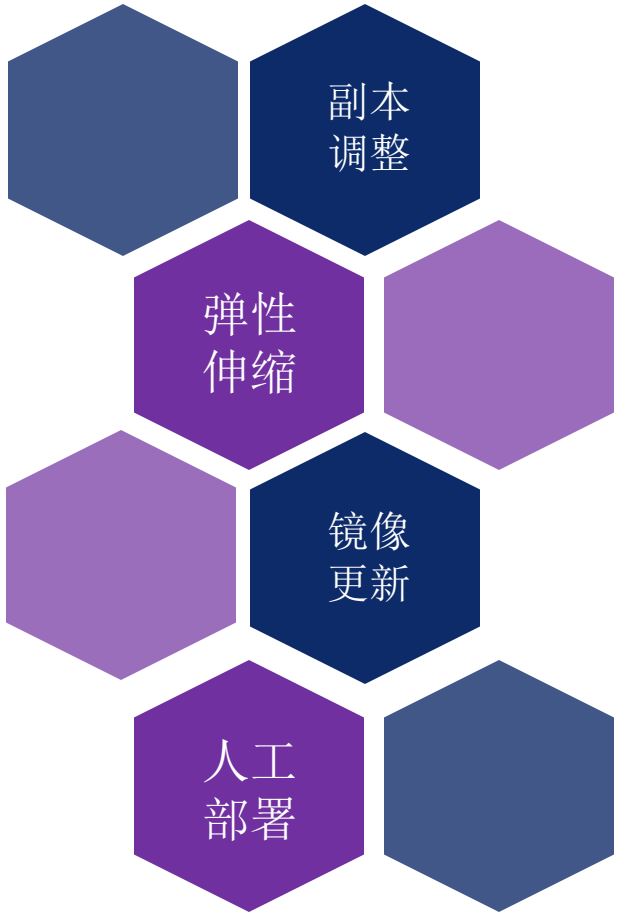
- 默认开启，自动伸缩
- 只调整副本数
- 不负责镜像版本、环境变量、配置文件等更新

## 人工部署场景

- 第一次上线部署
- 镜像更新、配置更新、环境变量等更新
- 指定单个Pod更新，线上回归验证

## 场景矛盾

- 人工部署时，希望副本数稳定，停止弹性伸缩
- 弹性伸缩时，人工部署会干扰扩缩容的执行效果



# 弹性伸缩和人工部署切换

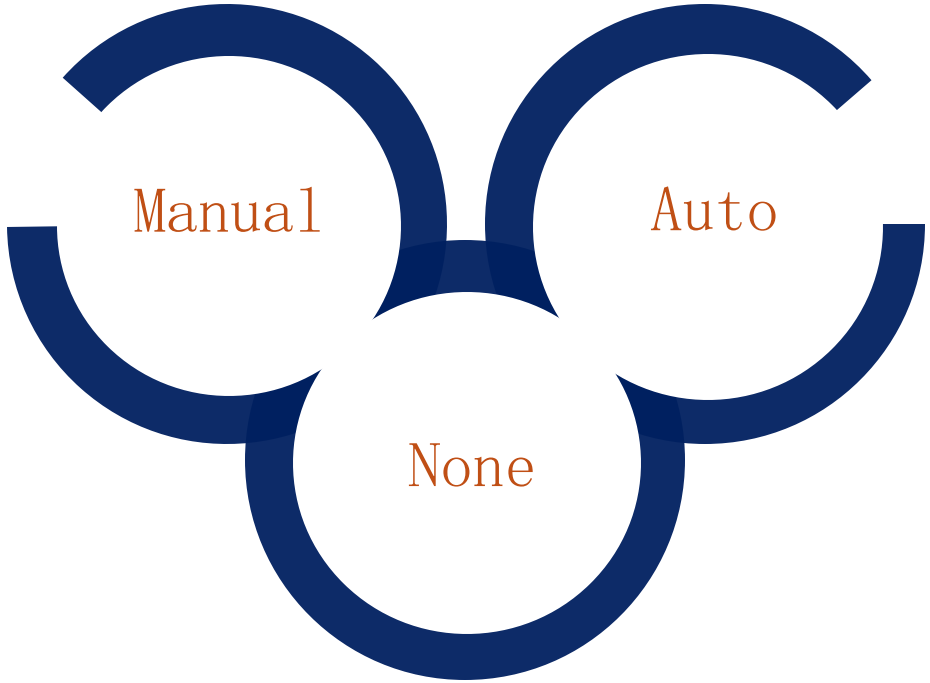


## Manual模式

人工部署场景

允许人工部署下发

屏蔽弹性伸缩控制




## Auto模式

弹性伸缩场景

允许弹性伸缩下发

屏蔽人工部署控制

## None模式

无模式状态，隔离场景

屏蔽弹性伸缩控制

屏蔽人工部署控制

# 全流程操作



Manual → None → Auto → Manual

## 01 第一次部署

第一次人工部署是Manual模式

## 02 确认人工部署完成

- 支持手动或自动确认上线
- 完成部署下发后才能切换

## 03 模式切换

切换条件：

- 配置并开启弹性策略
- 人工部署下发完成
- Pod版本一致

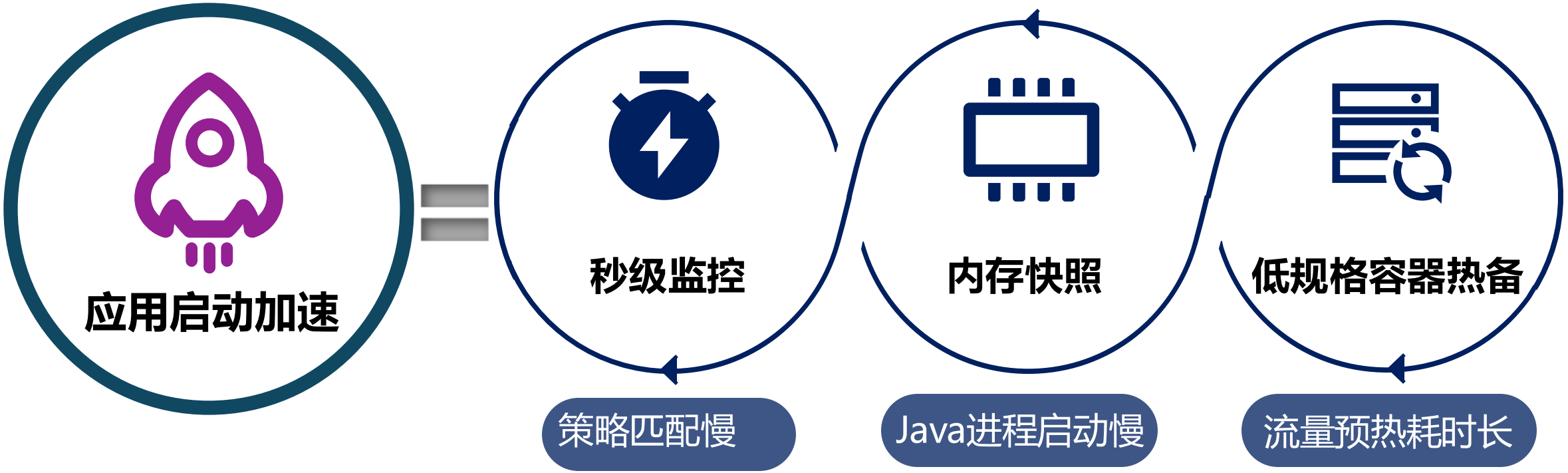
## 04 人工抢占

抢占后处理：

- 停止弹性控制器扩缩容工作
- 副本数修正

# 技术挑战：应用启动加速

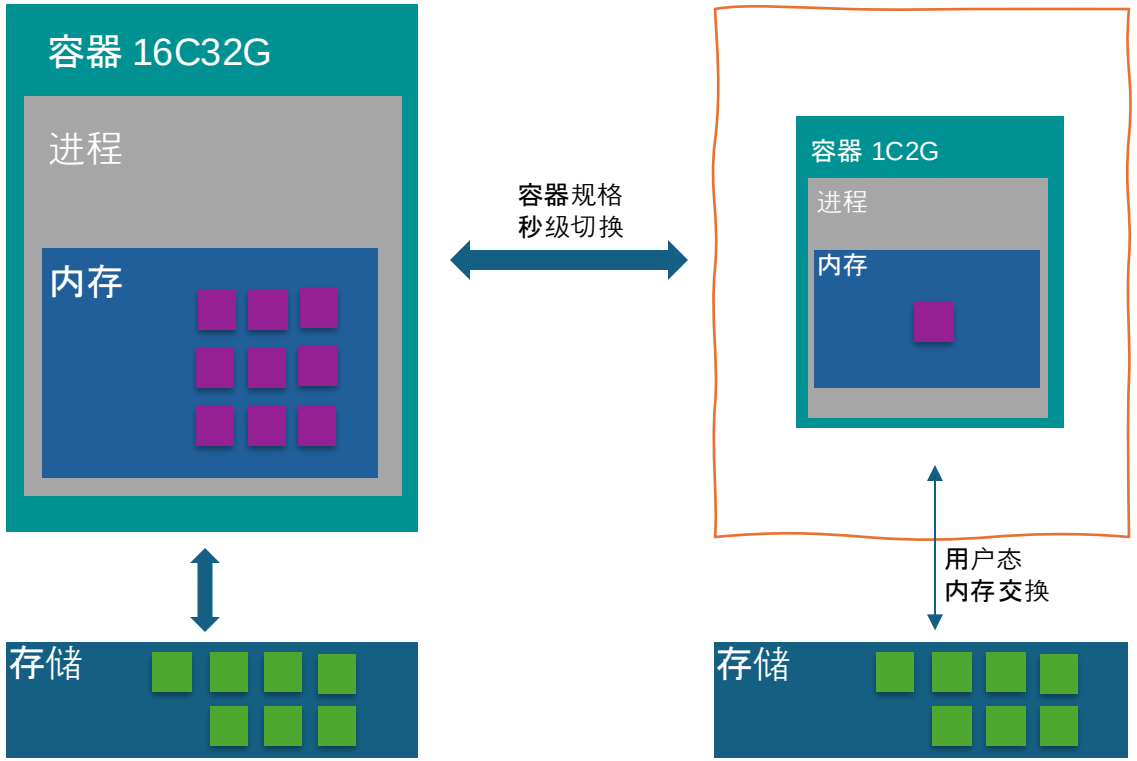




# 低规格容器热备

## 解决场景问题

大促压测前需要提前准备热备容器，占用大量资源




## 优点

- 无损，应用无需改造
- Pod启动快，启动时间减少80%以上
- Java应用无需预热即可承接流量

## 缺点

需要提前部署低规格实例，占用少量资源

## 适用场景

工作负载主要由流量触发，摘流后CPU利用率低

# JoyLive

## 单元化多活

- 服务治理框架

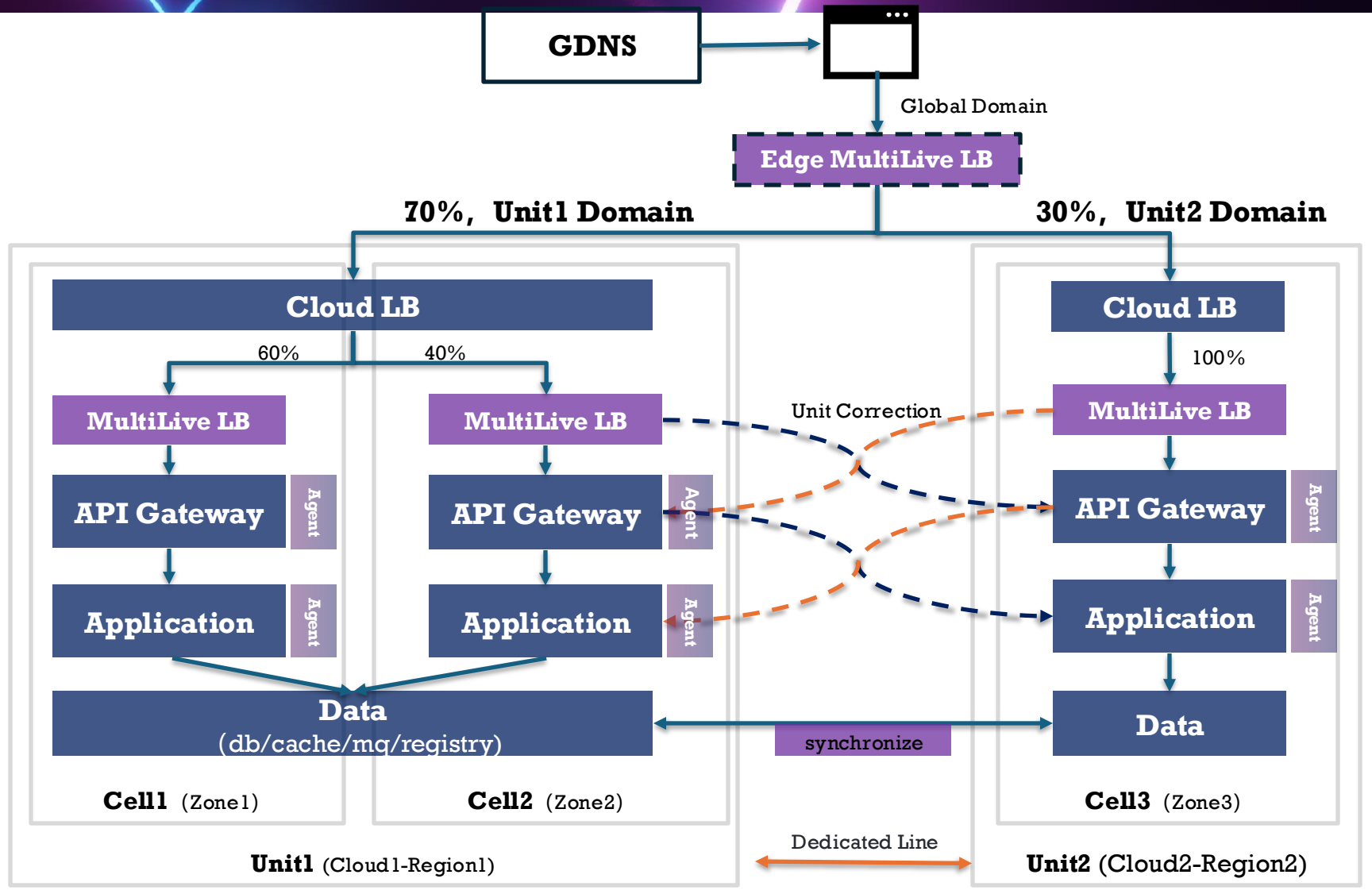
  基于字节码增强的面向应用多活和单元化的微服务流量治理框架

- 高可用

  提升联邦集群集成多活Unit和Cell等属性

- 开源

  https://github.com/jd-opensource/joylive-agent



# 跨集群服务发现和跨集群网络



## 跨集群通信

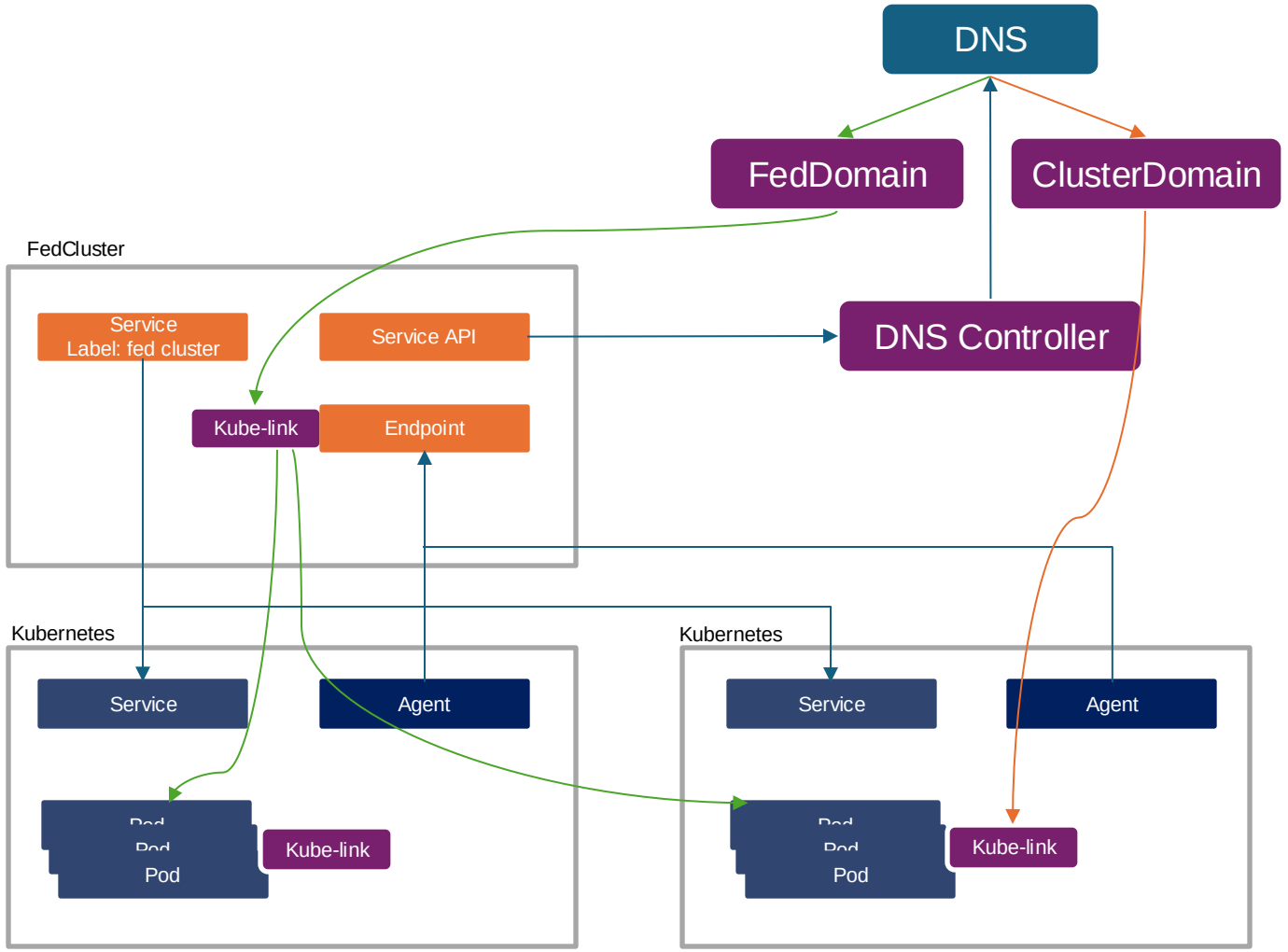
- 跨集群服务发现

面向多集群的四层LB和联邦域名注册

- 跨集群网络通信

基于eBPF的高性能负载均衡

- 开源准备中



# 04

# 总结

# 落地效果



春晚红包项目

京东春晚红包
支撑亿次数量抢红包

电商大促

支持多次京东618、京东11.11,电商大促，支撑**万亿级**交易额业务扩缩容

未来主要部署方式

未来是业务上线部署、自动迁移的主要方式之一
应用规模20万核+

KubeCon CloudNativeCon China 2024

THE LINUX FOUNDATION OPEN SOURCE SUMMIT

AI_dev Open Source GenAI & ML Summit

JDCloud 2024

# 谢谢

Contact us

JDCloud Dev

Tech Group