

- Serverless 容器从入门到精通 -

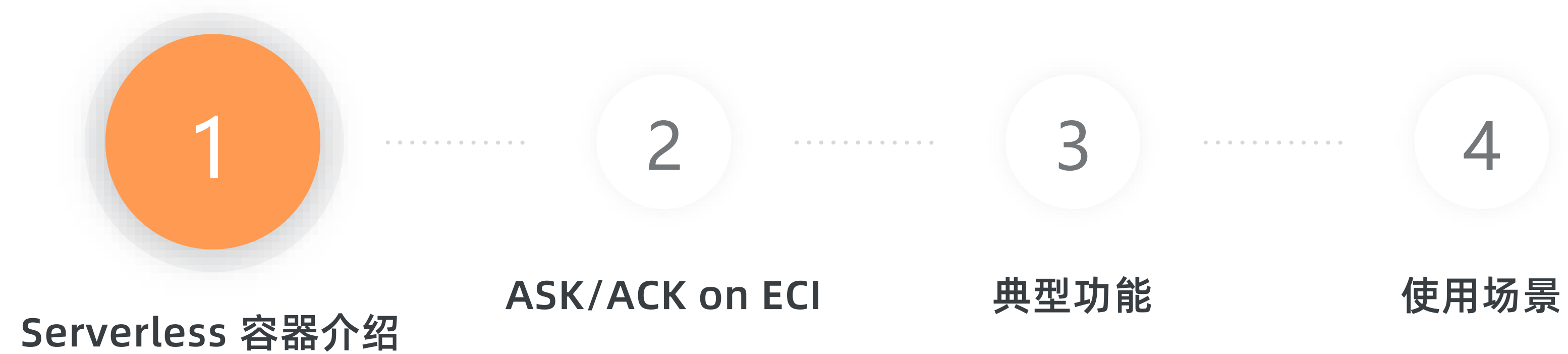
第 01 讲

Serverless Kubernetes 容器服务介绍

张维 阿里巴巴高级技术专家



关注“Serverless”公众号
获取第一手技术资料



Developers love Serverless



敏捷开发

- 无需购买和安装机器
- 无需管理服务器
- 无需升级和更新 OS
- 快速部署和更新应用
- 快速发布



极致弹性

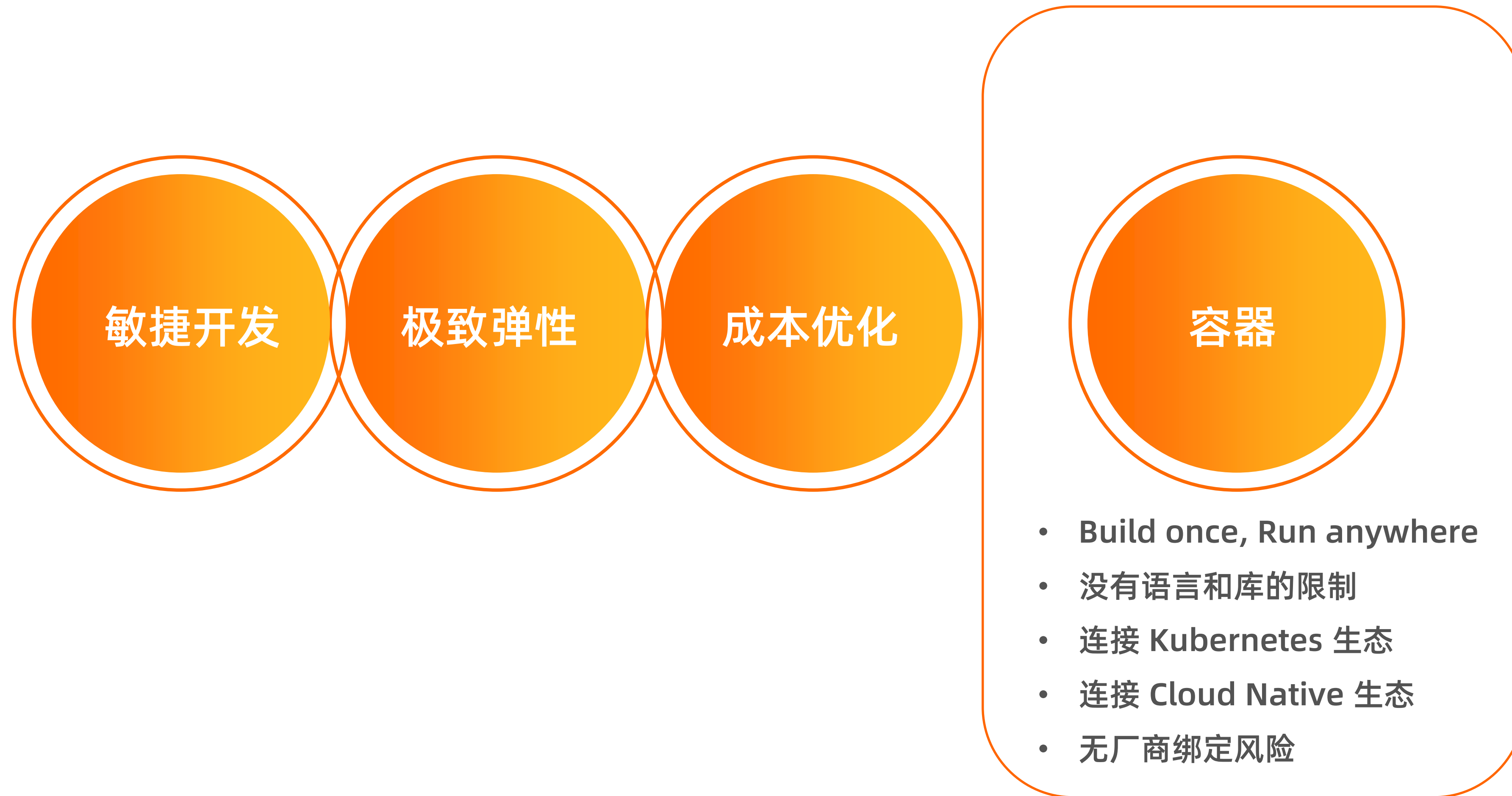
- “无限” 容量
- 秒级弹性
- 更好的扩展性
- 更好的灵活性



成本优化

- 按需创建
- 无资源闲置

Developers love Serverless Container



Serverless Container Landscape

Competitive Landscape of Public Cloud Container Service Market

Analysis Criteria		Alibaba	AWS	Google	IBM	Microsoft	Oracle
Container Orchestration Services	Kubernetes	Alibaba Cloud Container Service for Kubernetes (ACK)	Amazon Elastic Kubernetes Service (EKS)	Google Kubernetes Engine (GKE)	IBM Cloud Kubernetes Service (IKS)	Azure Kubernetes Service (AKS)	Oracle Container Engine for Kubernetes (OKE)
	Other Orchestration Technologies	Alibaba Cloud Container Services ^c	Amazon Elastic Container Service (ECS)				
Serverless Container Services	Serverless Kubernetes	Alibaba Cloud Serverless Kubernetes (ASK)	AWS EKS on Fargate			AKS Virtual Nodes	
	Serverless Container Instance	Alibaba Cloud Elastic Container	AWS ECS on Fargate	Google Cloud Run Google Cloud Run on GKE		Azure Container Instances (ACI)	
Kubernetes- Based Managed aPaaS	Red Hat OpenShift		Red Hat OpenShift Dedicated ^b	Red Hat OpenShift Dedicated ^b	Red Hat OpenShift on IBM Cloud OpenShift Service Mesh IKS Istio	Azure Red Hat OpenShift	
Container- Related Other Services	Service Mesh (Managed)	Alibaba Cloud Service Mesh	AWS App Mesh	Anthos Service Mesh			
	Registry Marketplace						
Hybrid and Multicloud Solutions	On-Premises Deployable Container Solutions	ACK on Apsara Stack	ECS/EKS on AWS Outposts	Anthos	IBM Cloud Paks	AKS Engine on Azure Stack ^d	Oracle Linux Cloud Native Framework
	Multicloud Container Management Solutions	ACK		Anthos (for the Other Clouds) ^e	IBM Cloud Pak Multicloud Management	Azure Arc	
General Criteria	First Production Support Date	May 2016	April 2015	August 2015	March 2017	March 2018	May 2018
	Pricing Model for Kubernetes Master Services ^a	Free	10 Cents per Hour per Cluster (EKS)/Free (ECS)	10 Cents per Hour per Cluster ^{f, g}	Free	Free	Free

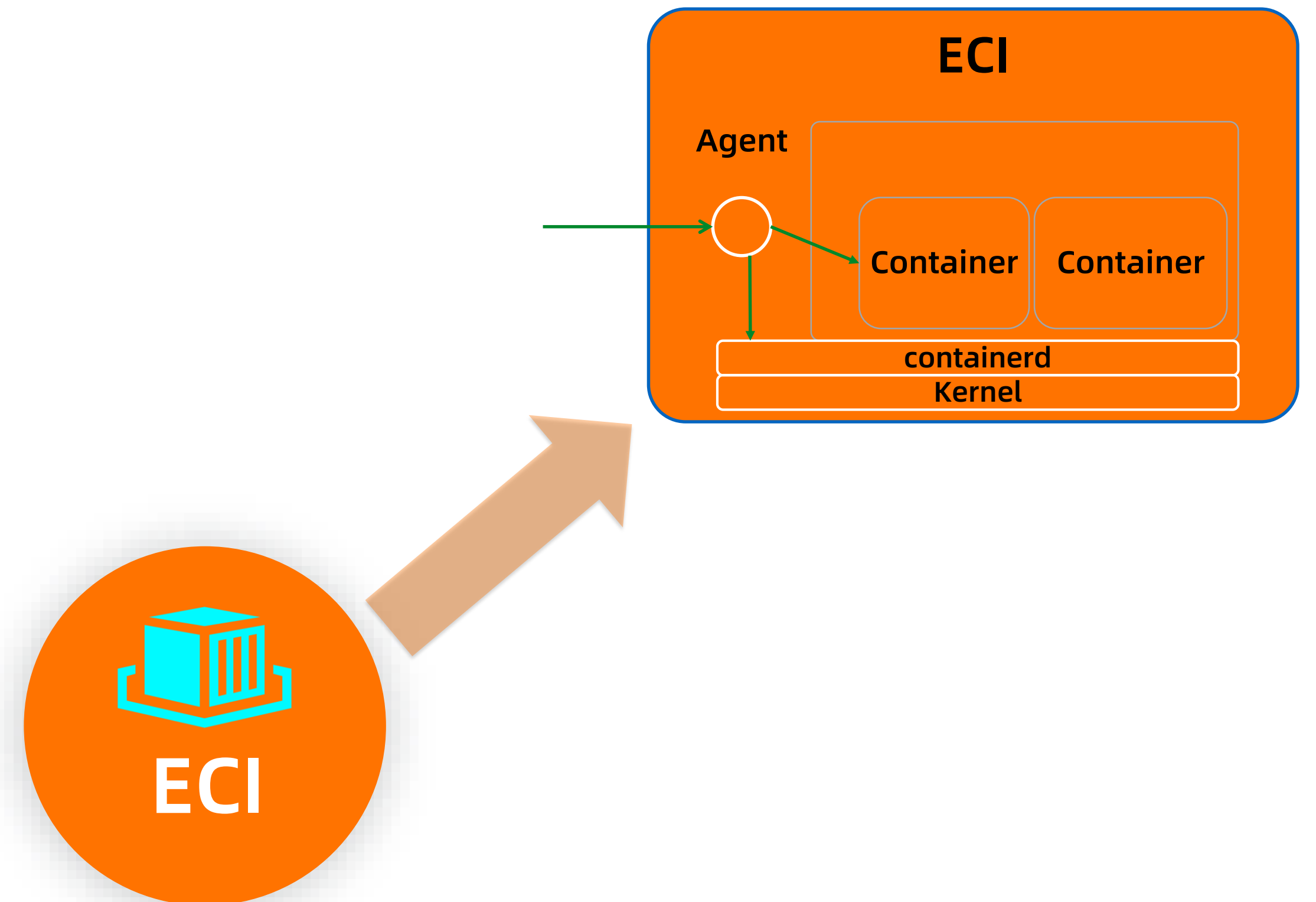
Source: Gartner

- Alibaba Cloud Serverless Kubernetes (ASK), ECI
- AWS Fargate
- Azure ACI

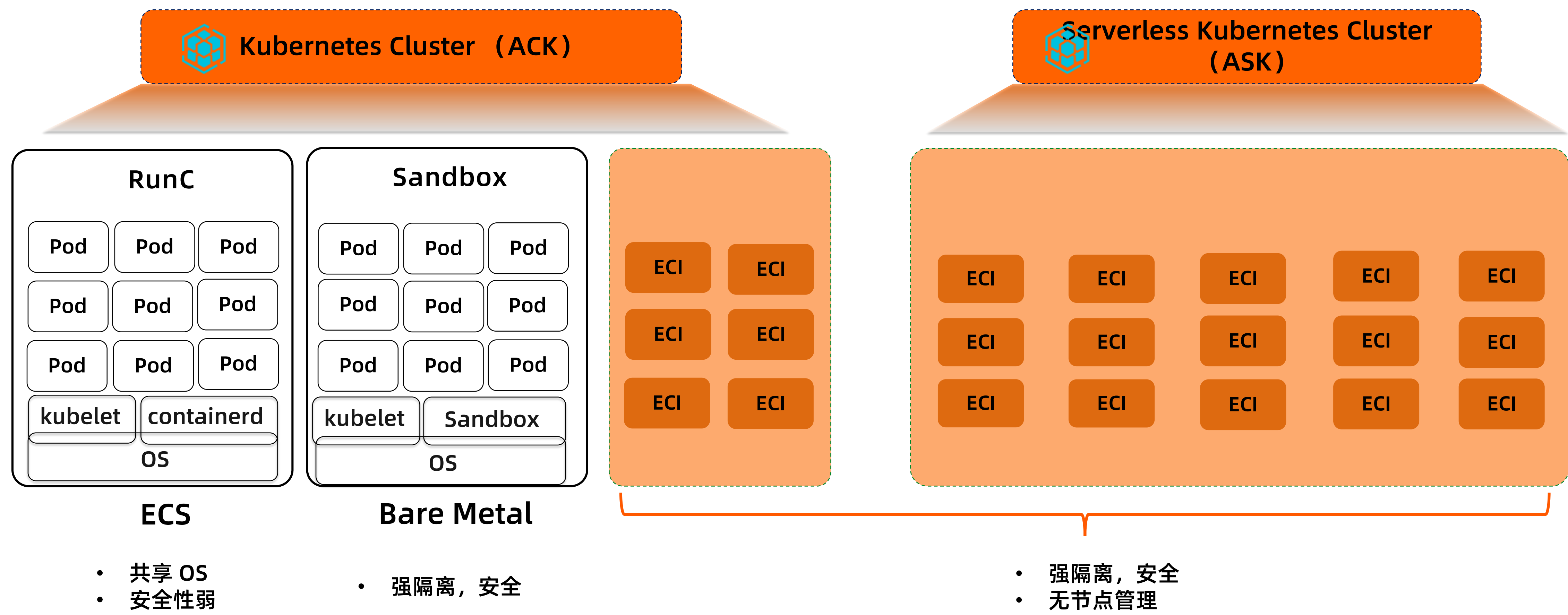


ECI: Elastic Container Instance

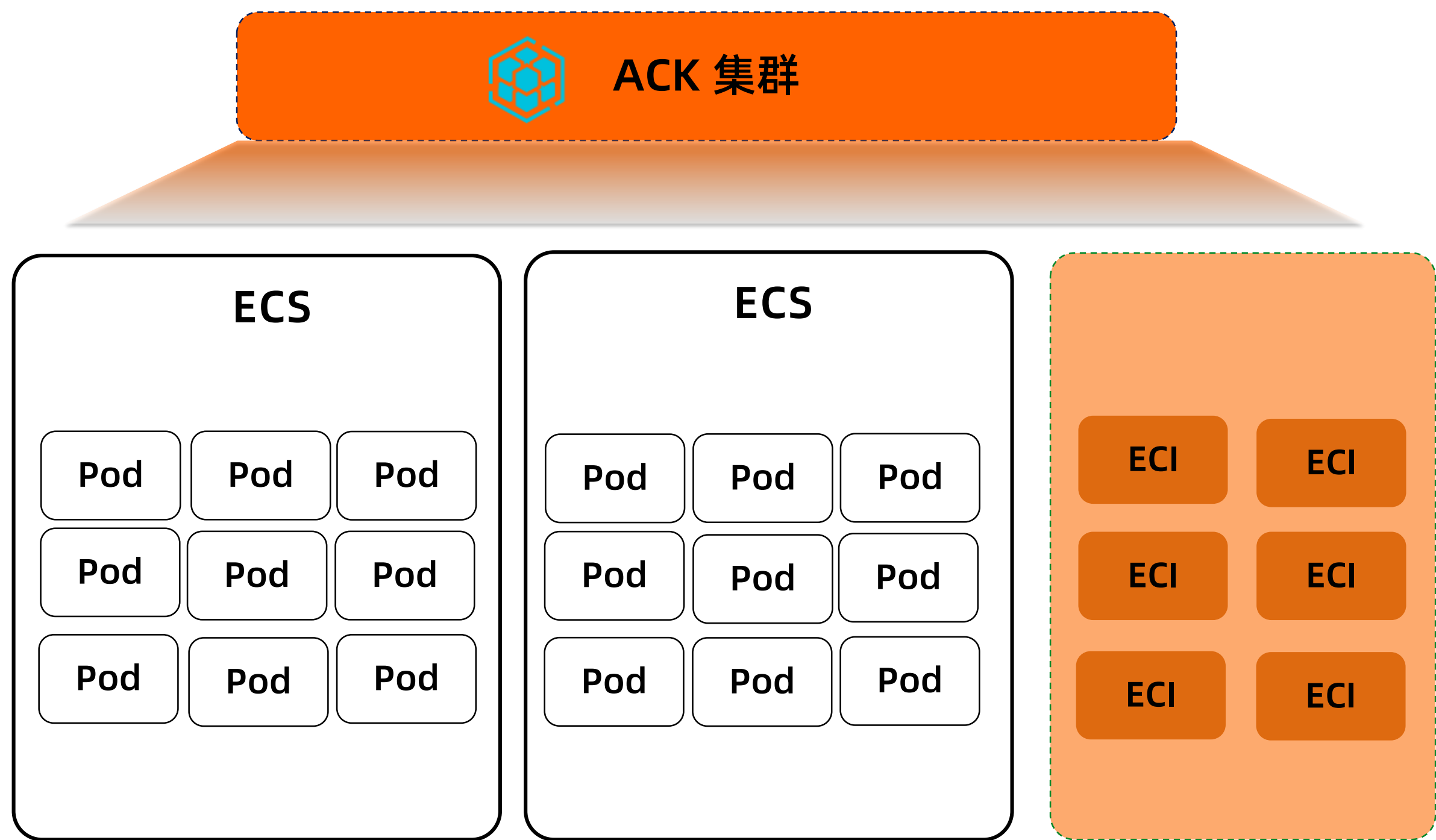
- **Run Containers without Managing Infrastructure**
- 容器成为云上的一等公民
- 安全隔离的容器运行环境
- 支持 CPU 0.25c - 64c, GPU, 按需创建按秒收费
- Spot Instance
- Startup time: ~10s
- 镜像缓存: 无需从远端拉取镜像
- 与 ECS 并池: 大规模资源池, 弹性能力保障
- **在 Kubernetes 中使用 ECI: ACK/ASK on ECI**



在 Kubernetes 中管理 ECI Pod

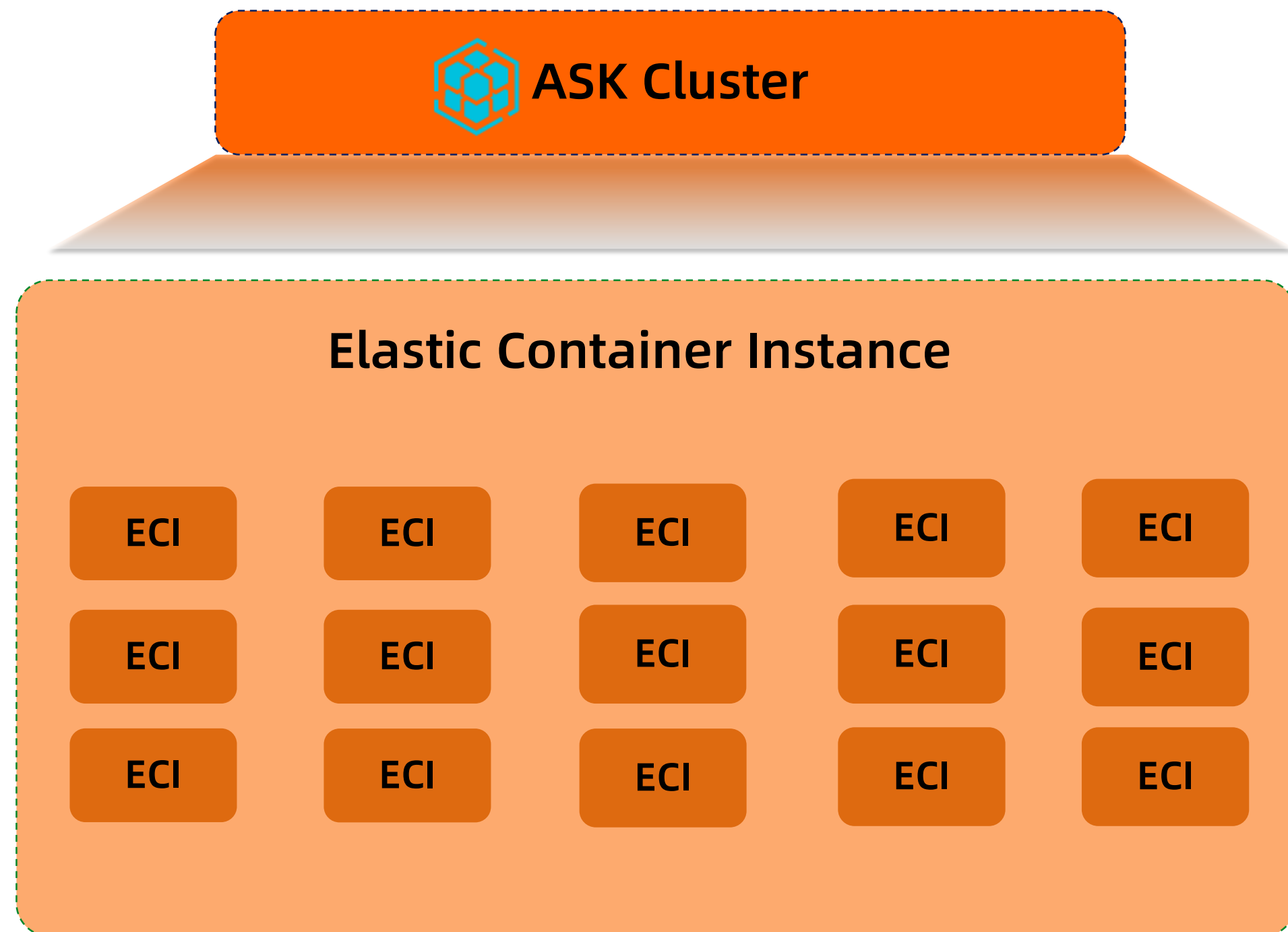


ACK on ECI



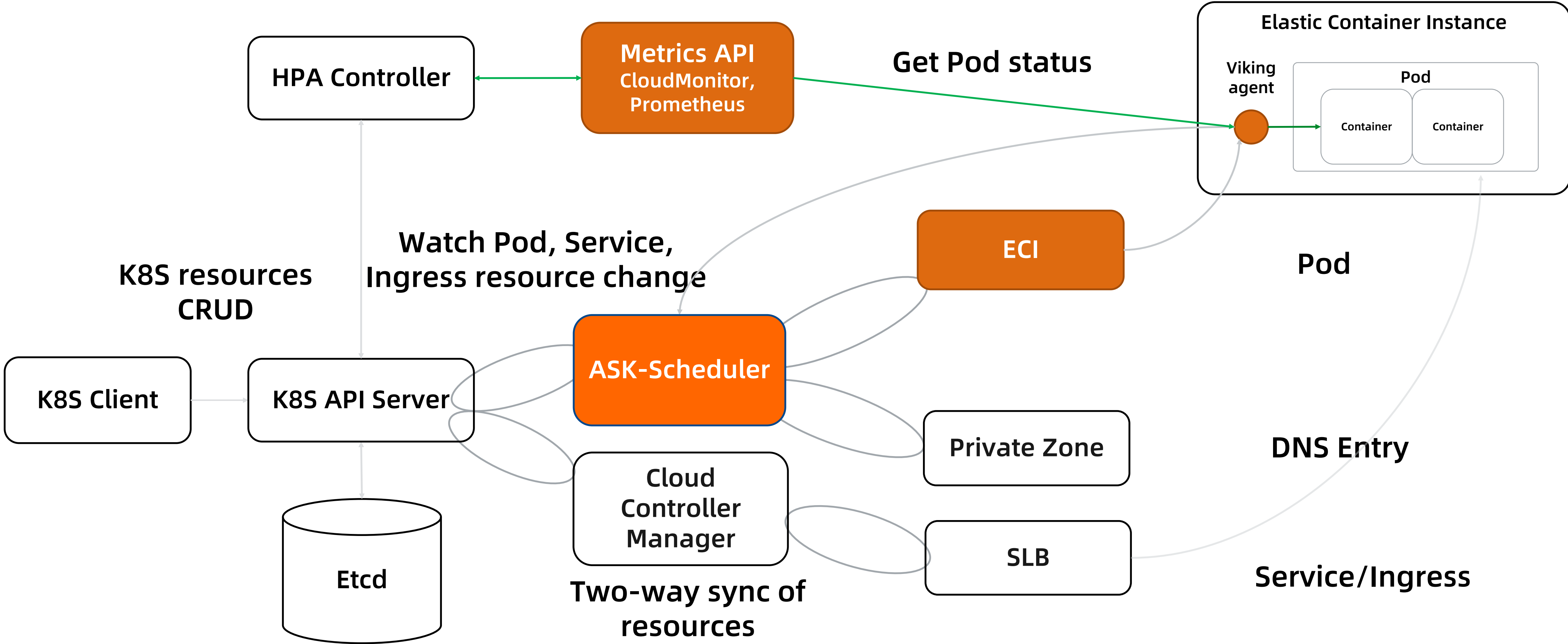
- **ECS + ECI 混合部署**：将 long run 应用运行在 ECS 上，弹性和任务应用运行在 ECI 上。
- “无限” 容量
- 极致弹性
- ECI 和 ECS Pod 之间 **互联互通**（支持 ClusterIP）

ASK: Serverless Kubernetes



- **Nodeless**: 简单易用, 无节点管理, 运维成本低
- **极致弹性**: 30s 500 pod
- **成本优化**: 按需创建, 支持 spot 和预留实例券
- **Kubernetes兼容性**: deployment / statfulset / job / service / ingress / CRD
- **ALB Ingress**: 基于 SLB 7 layer
- **Knative serving on ASK**: automatic scaling in knative
- **集成 ARMS, SLS**

ASK Architecture: Cloud-scale Nodeless Kubernetes





GPU Instance

- 预装 GPU 驱动，比如 TensorFlow 和 CUDA Toolkit

vCPU	Memory (GiB)	GPU type	GPU count	InstanceType
2	8.0	P4	1	ecs.gn5i-c2g1.large
4	16.0	P4	1	ecs.gn5i-c4g1.xlarge
8	32.0	P4	1	ecs.gn5i-c8g1.2xlarge
16	64.0	P4	1	ecs.gn5i-c16g1.4xlarge
32	128.0	P4	2	ecs.gn5i-c16g1.8xlarge
56	224.0	P4	2	ecs.gn5i-c28g1.14xlarge
8	32.0	V100	1	ecs.gn6v-c8g1.2xlarge
32	128.0	V100	4	ecs.gn6v-c8g1.8xlarge
64	256.0	V100	8	ecs.gn6v-c8g1.16xlarge

```
apiVersion: apps/v1beta2
kind: Deployment
metadata:
  name: nginx-gpu-demo
  labels:
    app: nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
      annotations:
        k8s.aliyun.com/eci-use-specs: ecs.gn5i-c4g1.xlarge
    spec:
      containers:
        - name: nginx
          image: tensorflow/tensorflow:1.13.1-gpu-py3
          resources:
            limits:
              nvidia.com/gpu: '1'
          ports:
            - containerPort: 80
```

Spot Instance

- 极大降低按量计算成本
- 一个小时的运行时常保障
- 适用场景: 大数据、媒体处理、科学计算、弹性web站点、测试

SpotAsPriceGo

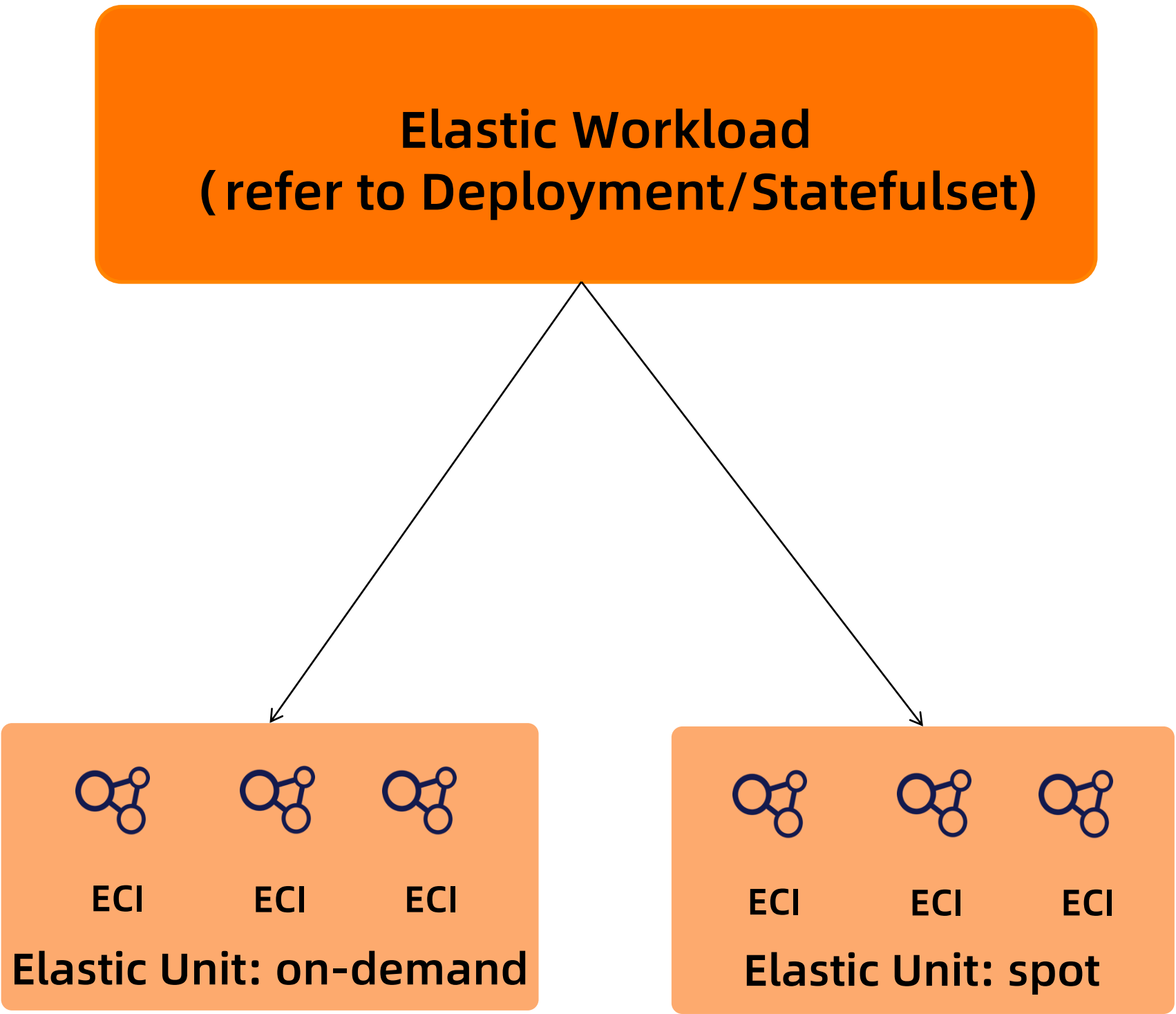
```
apiVersion: apps/v1beta2
kind: Deployment
metadata:
  name: nginx
  labels:
    app: nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
      annotations:
        k8s.aliyun.com/eci-use-specs : "2-4Gi"
        k8s.aliyun.com/eci-spot-strategy: "SpotAsPriceGo"
    spec:
      containers:
        - name: nginx
          image: nginx:1.7.9
          ports:
            - containerPort: 80
```

SpotWithPriceLimit

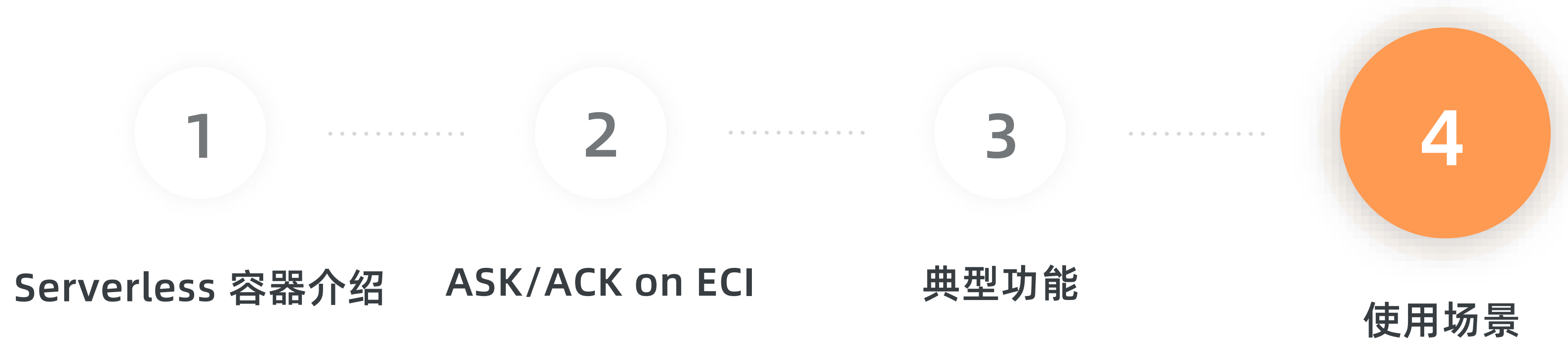
```
apiVersion: apps/v1beta2
kind: Deployment
metadata:
  name: nginx
  labels:
    app: nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
      annotations:
        k8s.aliyun.com/eci-use-specs : "ecs.c5.large"
        k8s.aliyun.com/eci-spot-strategy: "SpotWithPriceLimit"
        k8s.aliyun.com/eci-spot-price-limit: "0.250"
    spec:
      containers:
        - name: nginx
          image: nginx:1.7.9
          ports:
            - containerPort: 80
```

弹性负载Elastic workload

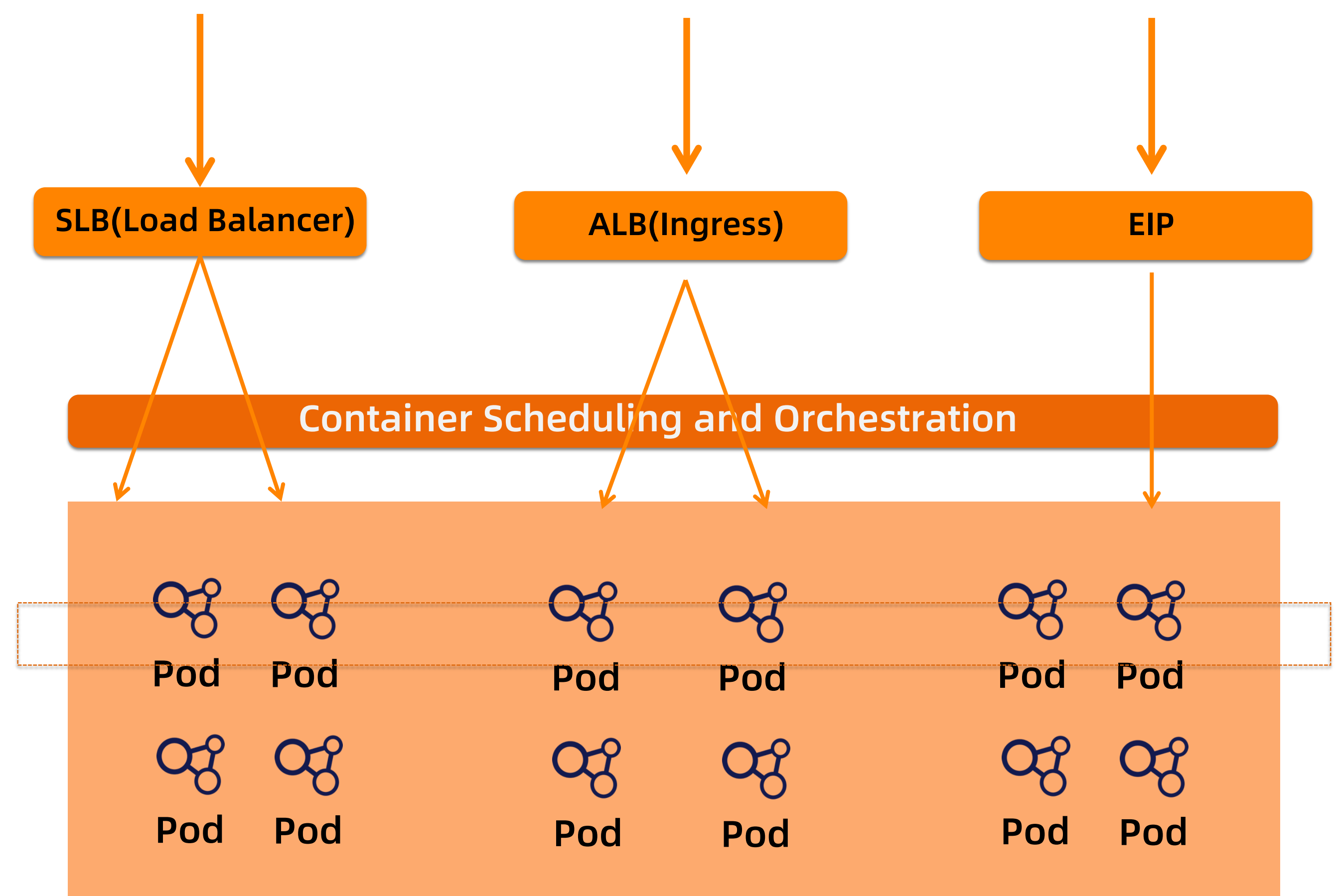
- 精细化配置应用负载使用eci和eci-spot实例的副本数量
- 降低deployment/statefulset的计算成本



```
apiVersion: autoscaling.alibabacloud.com/v1beta1
kind: ElasticWorkload
metadata:
  name: elasticworkload-nginx
spec:
  sourceTarget:
    name: nginx-deploy
    kind: Deployment
    apiVersion: extensions/v1beta1
    min: 0
    max: 2
  replicas: 6
  elasticUnit:
    - name: eci-spot
      annotations:
        k8s.aliyun.com/eci-spot-strategy: "SpotAsPriceGo"
```



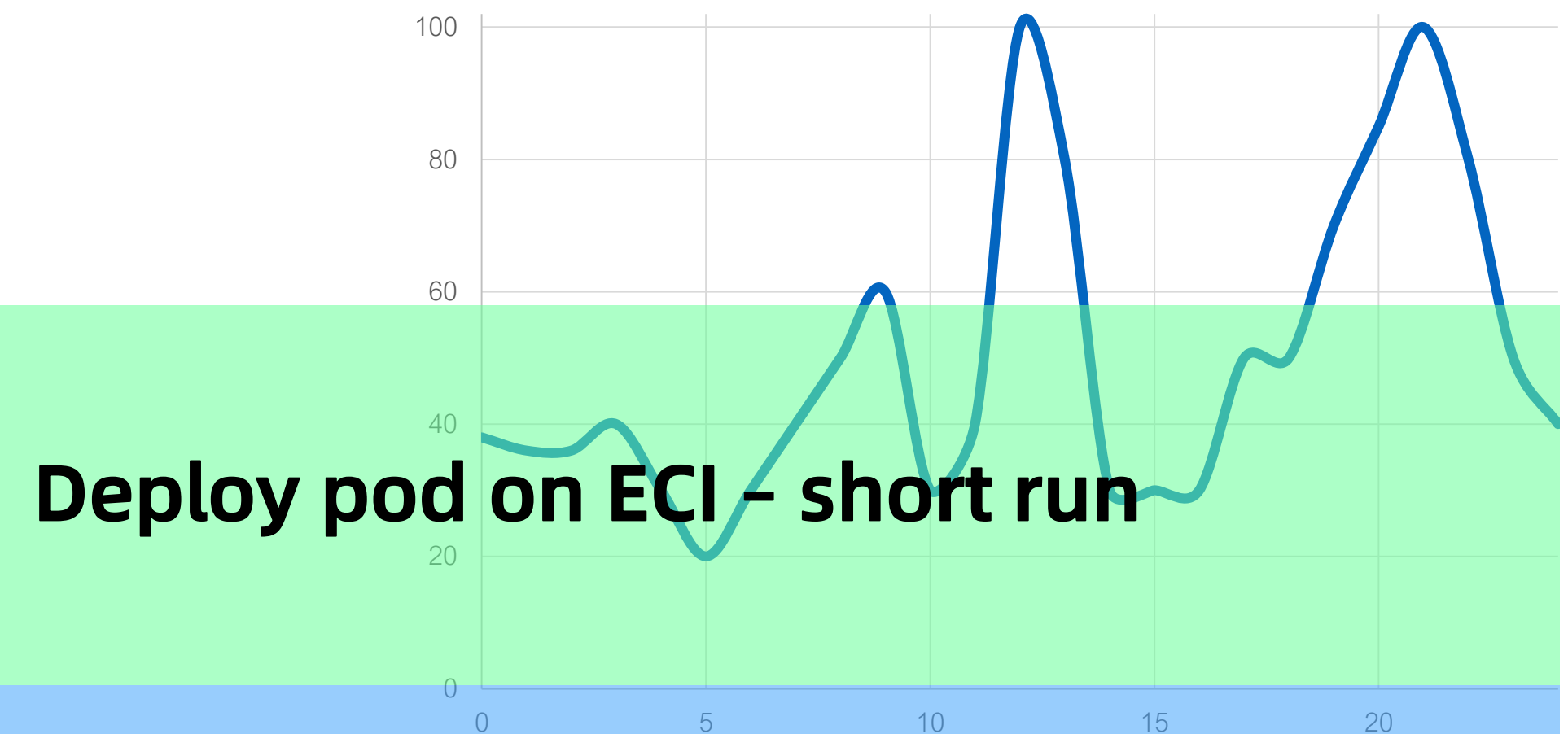
免运维应用托管



- 无需管理基础设施，专注应用开发
- 灵活弹性、按需使用
- 结合预留实例券大幅降低计算成本
- 支持原生 kubernetes 生态，包括 Helm Chart 部署等

ECI 弹性资源池

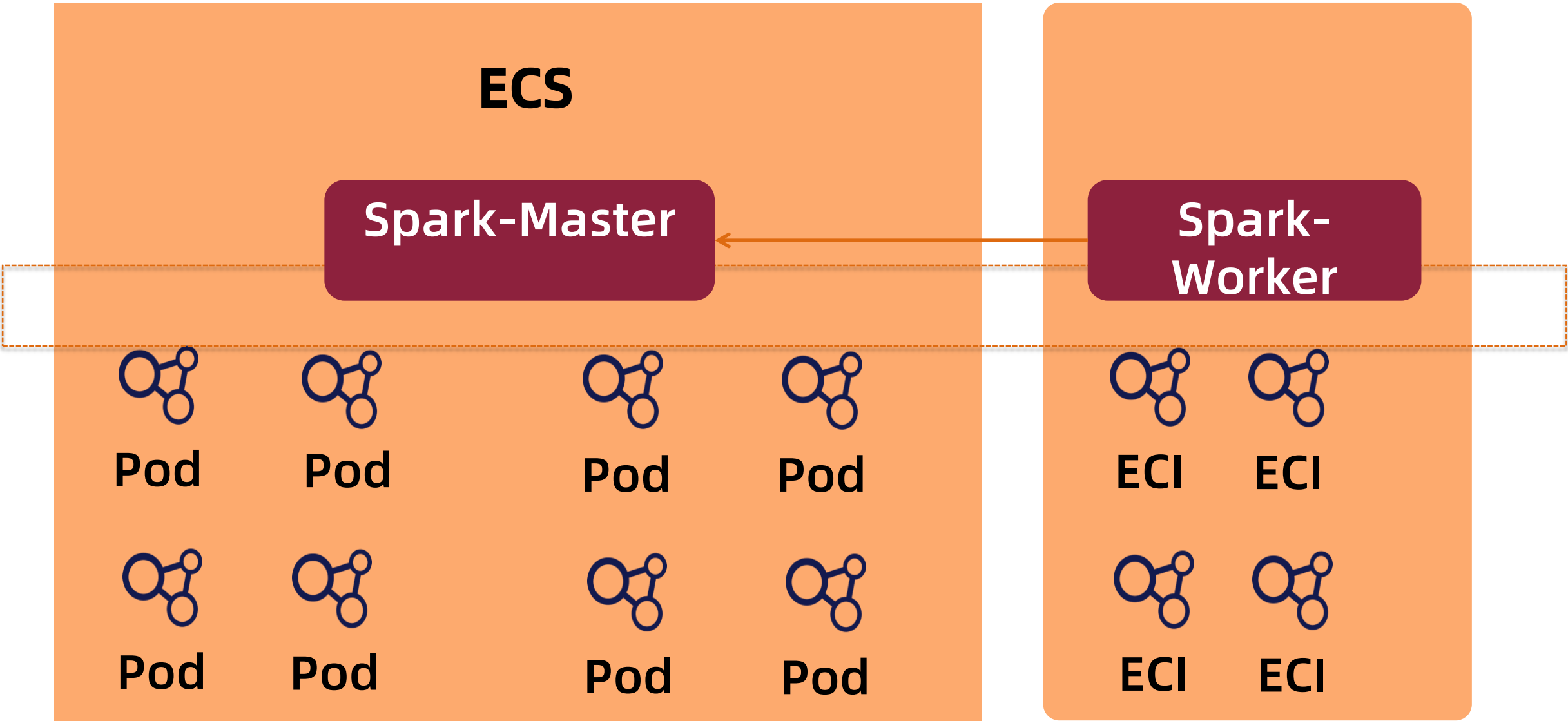
- 适用于有明显波峰波谷弹性特征的场景，比如电商、在线教育等。
- 更好的弹性效率。(30s 500 pod)
- 无需维护大量节点的固定资源池，节省成本。



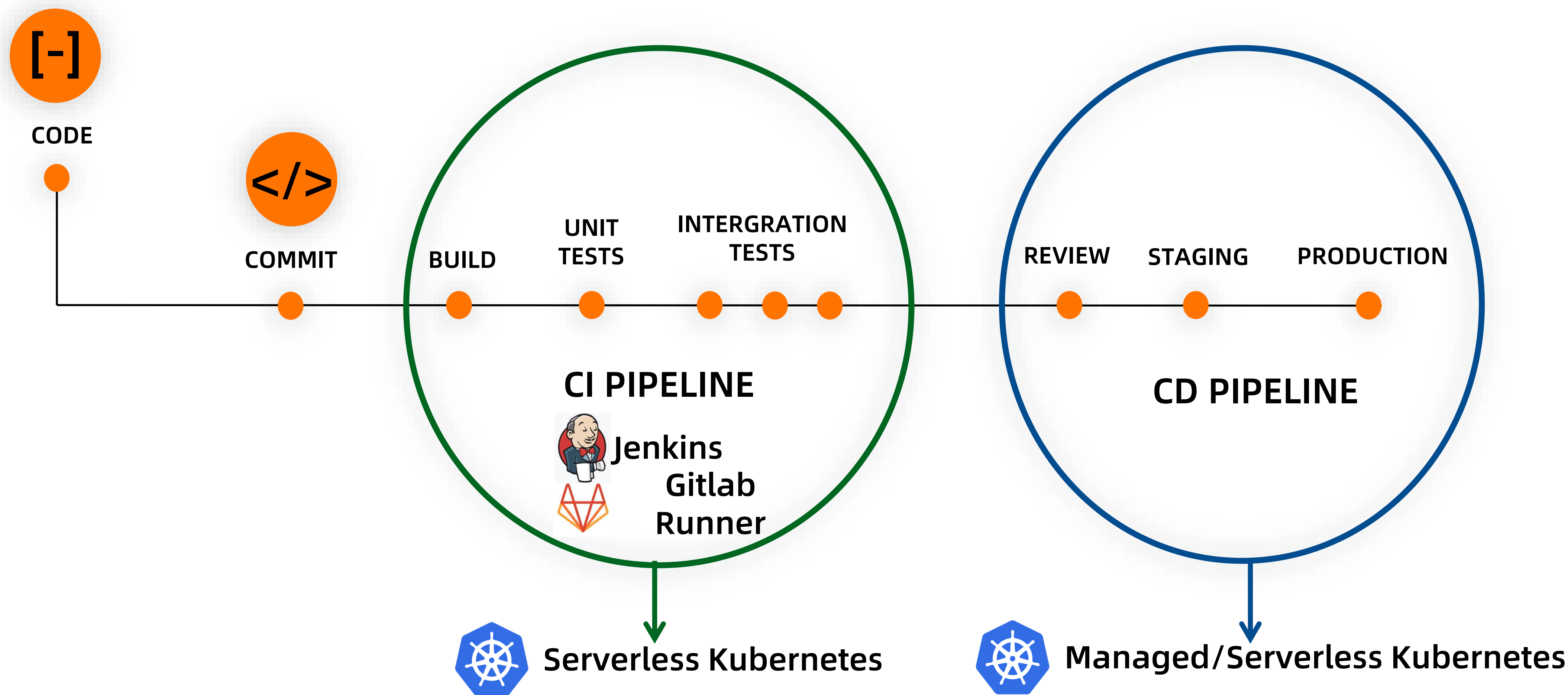
Kubernetes cluster with ECS node – long run

Kubernetes API Server

- 使用 ECI 运行 Spark / Presto / AI 计算任务
- “无限” 容量
- 节省成本



CI/CD 持续集成



开始使用

- 控制台: <https://cs.console.aliyun.com/ask>
- 产品文档: <https://www.alibabacloud.com/help/doc-detail/86366.htm>
- 示例: <https://github.com/AlibabaContainerService/serverless-k8s-examples>

谢谢观看

THANK YOU



关注“Serverless”公众号
获取第一手技术资料