aws | 在线研讨会

AWS Firecracker 与 Amazon EKS on Fargate
基于 Kubernetes 的 Severless 容器计算平台

CNCF Webinar 2020

AWS 解决方案架构师
莫梓元

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

会议议程

- 虚拟化与容器运行时基础知识，KVM 及 Firecracker
- AWS Firecracker 设计与功能特性
- AWS Firecracker 如何实现多租户安全容器
- AWS Fargate Serverless 容器介绍
- Amazon EKS on Fargate 基于 Kubernetes 的无服务器工作节点介绍
- 安全容器运行时，数据平面实现及相关项目介绍
    - Firecracker-containerd, Kata-containers, Ignite, Rust-VMM 等

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Firecracker

什么是虚拟机

- 虚拟机是对物理计算机系统的仿真
- 虚拟机基于计算机体系结构，提供物理计算机的功能。
- 虚拟机监视器(VMM，也称为Hypervisor)是提供虚拟机抽象的应用软件。
- 历史悠久:
- 全虚拟化和半虚拟化: Mainframes, Unix, x86 环境 VMWare 和 Xen
- 硬件辅助虚拟化(Intel VT-x): KVM-QEMU, Hyper-V, Nitro, Firecracker

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

什么是 KVM

- Liunx 操作系统中内建的虚拟化基础组件
  - 构建 Hypervisor的工具集
- 利用处理器的硬件虚拟化特性
  - Intel VT-x 技术
  - 没有完整的 PV 模式(不同于 Xen)
- 将设备仿真交给进程实现
  - 大多数人使用 QEMU 进行仿真
  - AWS 构建了 Nitro 和 Firecracker

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

什么是 Firecracker VMM

- 设置 KVM，VirtIO 存储，网络，套接字等
- 基于 crosvm 项目
- 负责响应VM_EXIT
    - 有限设备仿真
    - 一些特权指令

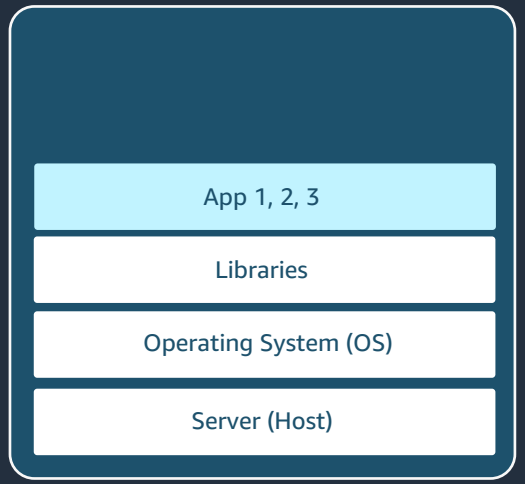
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

什么是容器

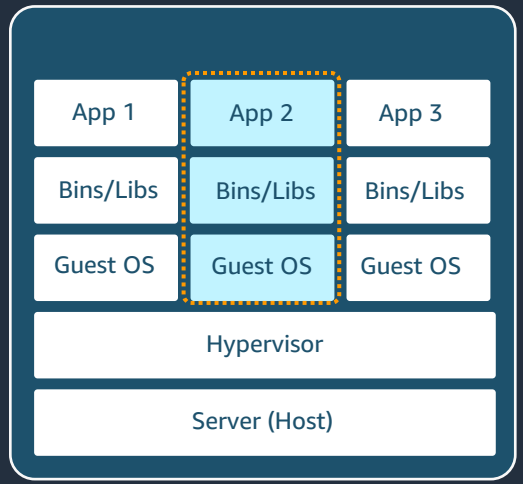
- 一种 OS 级别的虚拟化技术。
- 一个轻量级的、独立的、可执行的应用软件包，包括所有依赖项: 代码、运行时、系统工具、系统库、配置。
- 容器将应用与其环境隔离开来
    - 适合于开发和验证环境。
    - 帮助减少在相同基础环境上运行不同应用软件之间的冲突。
- 历史同样悠久：Chroot, FreeBSD Jails, OpenVZ, LXC, libcontainer 等
- Docker 及 ContainerD简化了容器的生命周期管理难度，也建立的 OCI 标准镜像规范

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

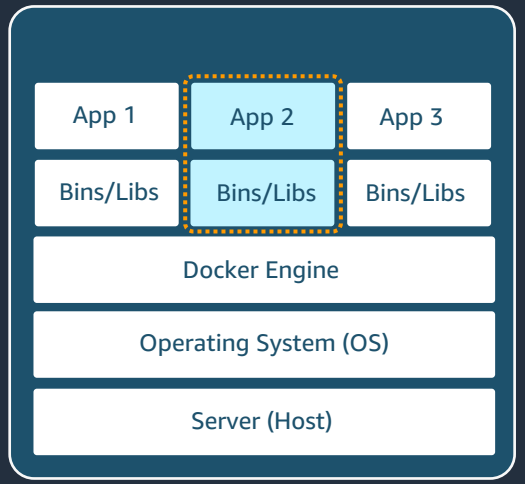
运行应用的不同方式



裸金属服务器



虚拟机



容器

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

容器 vs 虚拟机

- 资源隔离（命名空间）
  - 进程
  - 挂载点
  - 网络
  - 硬件 (cgroups)
- 减少攻击面
- 启动快 (敏捷)
- 更少的开销

- 资源隔离
  - 硬件
  - 网络
- 更好的资源可见性
- 不共享内核
- 隔离的故障域
- OS 不可见
  (Host OS vs. Hypervisor)

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Serverless 需要新的计算平台

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Serverless 计算的需求

安全

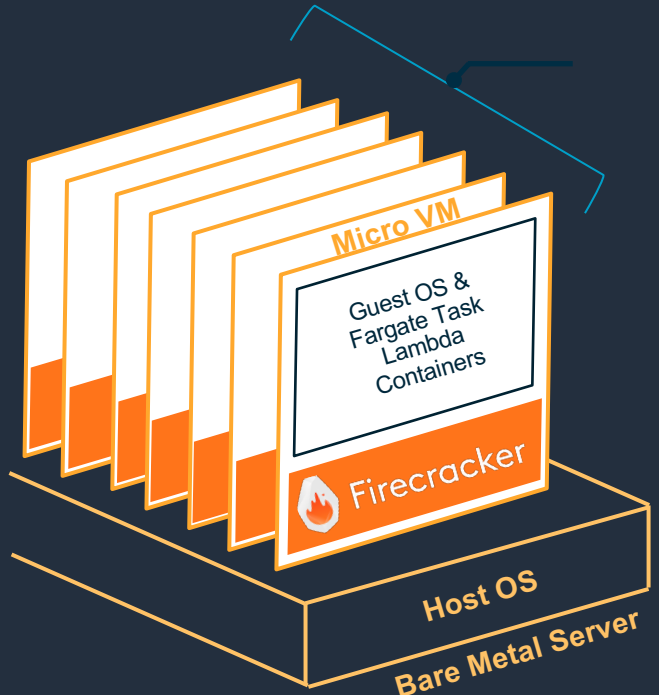
启动时间

资源利用率

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

什么是AWS Firecracker？

为容器工作负载专门构建
的开放源码虚拟化机

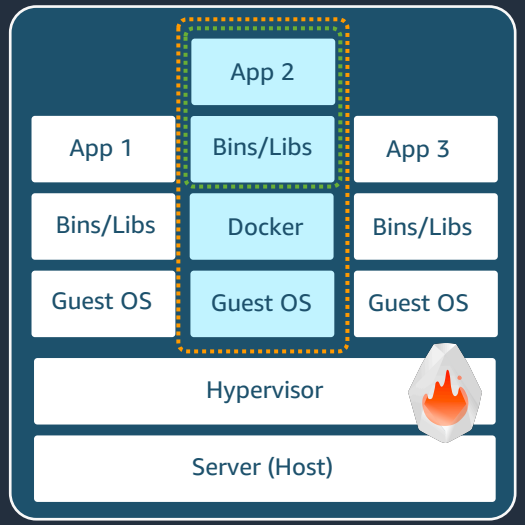


硬件虚拟化隔离保证多租户安全

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

为什么使用AWS Firecracker？

每个容器享有独立隔离内核，以确保多租户安全

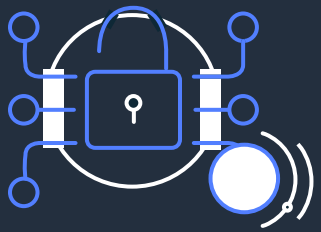


Container in Micro VM

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

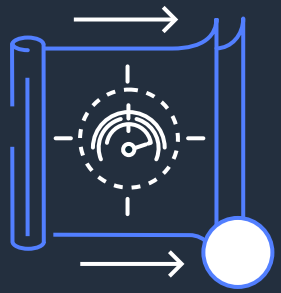
AWS Firecracker 在 Serverless 计算下的优势

用于构建无服务计算的开源虚拟化技术



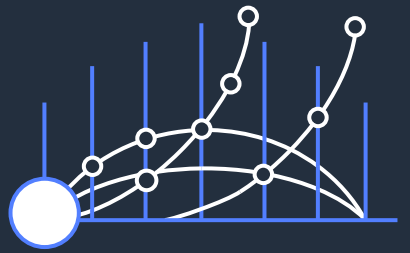
安全隔离

深度防御
最小的设备模型
速度限制
默认的安全



速度源于简单

Micro VM启动时间: **4ms**
Guest OS Linux用户空间引导时间:**125 ms**

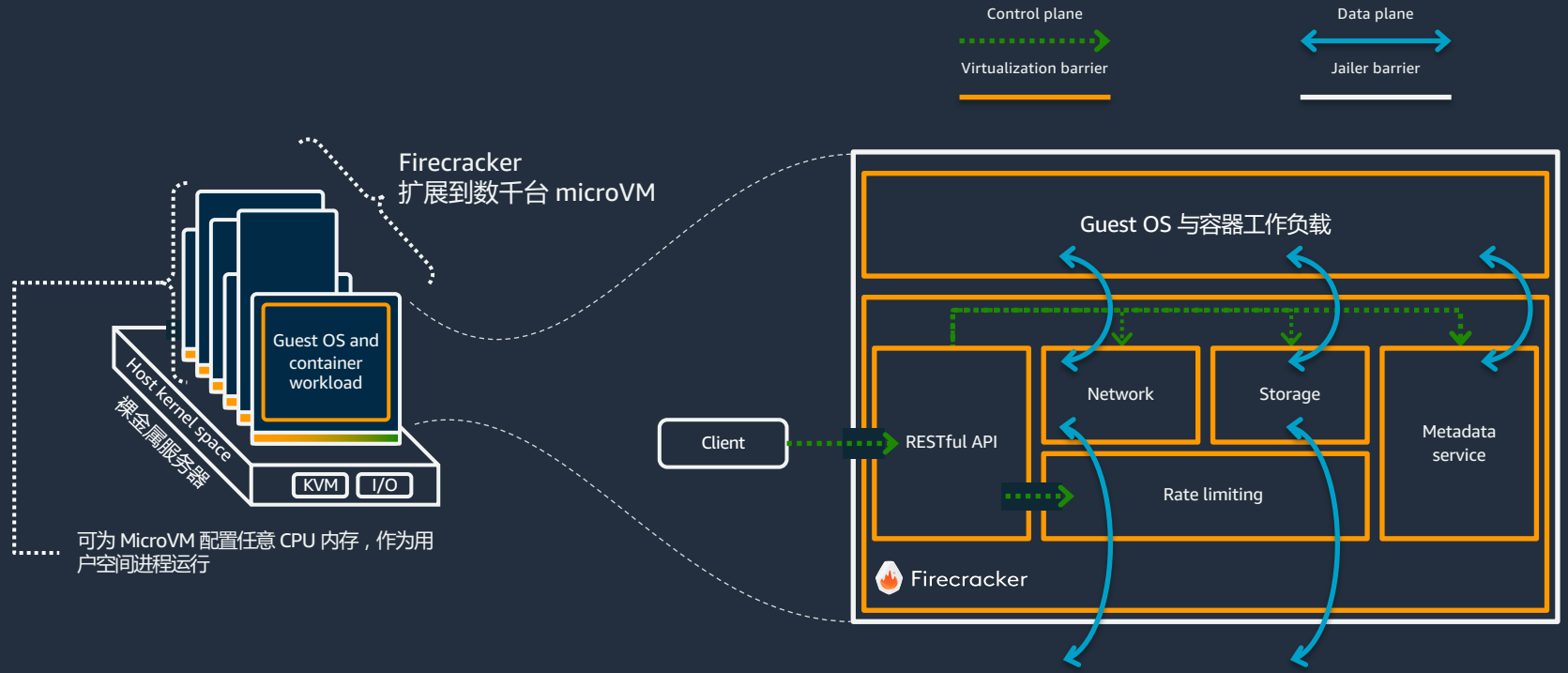


大规模下的效能

Firecracker 的内存开销: **< 5 MiB**
每台实例同时运行数千台 microVM

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Firecracker 总体设计



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Firecracker 功能特性

- 适用 Rust 语言开发的基于KVM虚拟化技术的VMM
- 基于 Apache 2.0 许可证开源
- 针对 Serverless 工作负载
- 非通用的 VMM
- 良好的多租户支持
- 支持任意的 vCPU 和 内存 组合
- 支持资源超分

- 快速启动 (用户空间初始化 125 ms)
- 快速恢复 (快照恢复 5 ms)
- 稳定的启动速度: 150+ microVMs/host/sec
- 低内存开销 (<5 MiB)
- 低硬件资源开销 (4000+ VMs in i3.metal)
- 宿主机提供 REST API
- 最小化 Guest 设备模型

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Firecracker
如何提升安全性与资源利用率

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws | 在线研讨会

虚拟化同样面临的安全隐患

- 虚拟机并不意味着真正安全
  - VENOM: QEMU 漏洞 (CVE-2015-3456)
  - 一个影响软盘控制器(FDC)设备仿真的“缓冲区溢出”漏洞
- 使用 Rust 语言开发新 VMM
  - 内存安全 – VirtIO, 内存模型
  - 提供安全且简单的并发控制 – vCPU 线程, I/O 工作线程
  - 手动垃圾回收 – 减小虚拟化时延开销

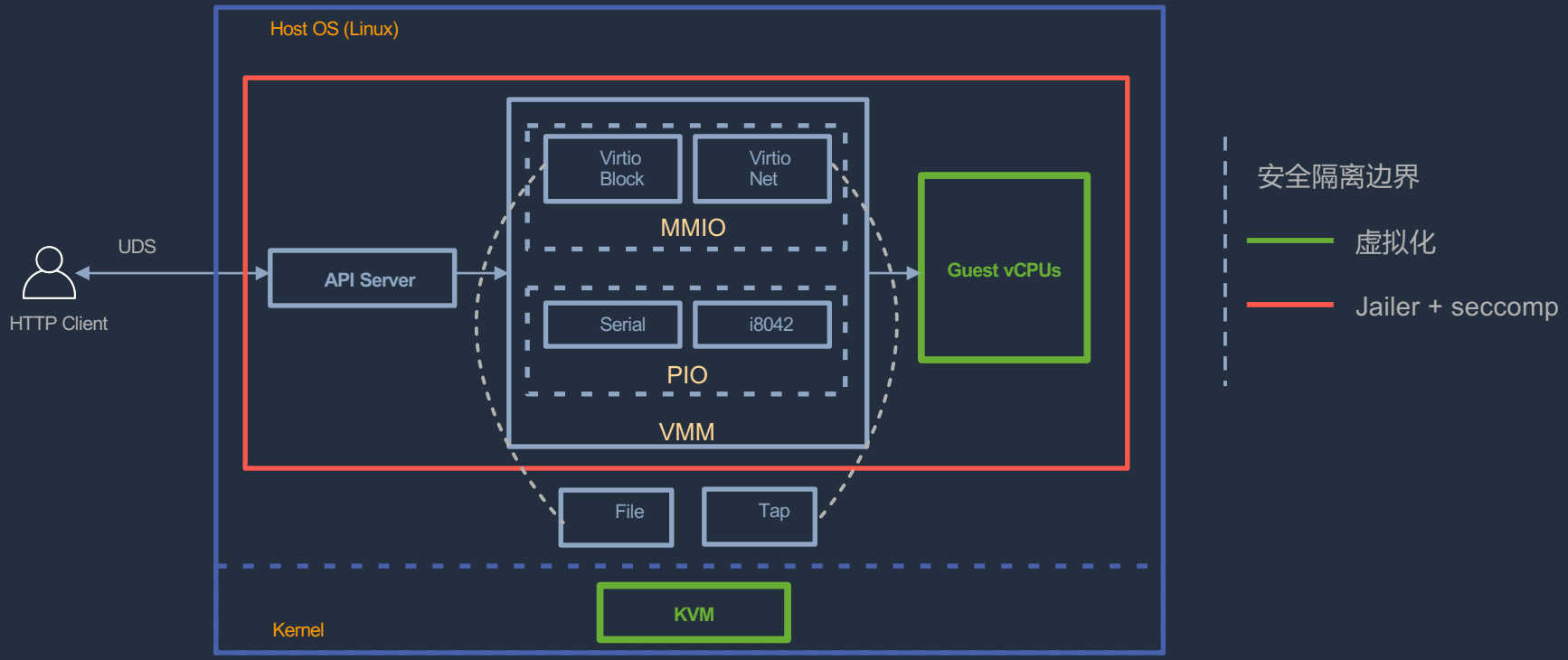
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

如何消除虚拟化安全隐患

- 非常有限的设备模型
- 非常有限的功能集合
- 消除 Guest 与 Host 内核的直接交互
- 不同虚拟机进程间沙箱隔离
- Rust 开发语言
- 每个虚拟机一个 Firecracker 进程

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Firecracker 采用有限的设备模型



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Fargate

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

什么是 AWS Fargate

Fargate 是一个 AWS 上为容器设计的无服务器计算平台

使用 EKS 和 ECS 的 Fargate 之间的差异，是由业务编排系统所驱动的

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

You should be able to write your code and have it run, without having to worry about configuring complex management tools.

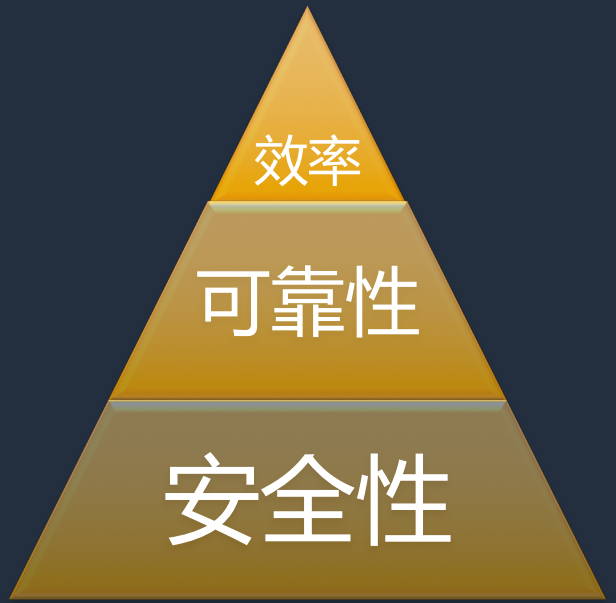
This is the vision behind AWS Fargate.

Dr. Werner Vogels

CTO, Amazon.com

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Fargate 设计原则



- 保障客户工作负载的安全始终是头等大事。
- 接下来，我们将竭尽全力确保服务的可靠性和可伸缩性。
- 一旦达到上述目标，我们将努力提高集群资源利用率和运行效率。

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EKS on Fargate

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws | 在线研讨会

选择 EKS 运行 Kubernetes 的优势

适合生产负载

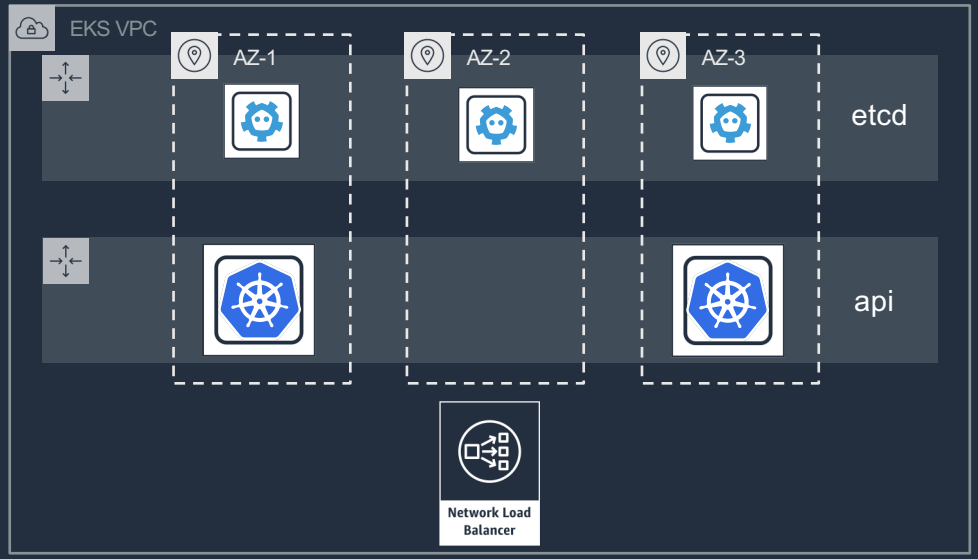
原生与上游一致

OSS 项目贡献

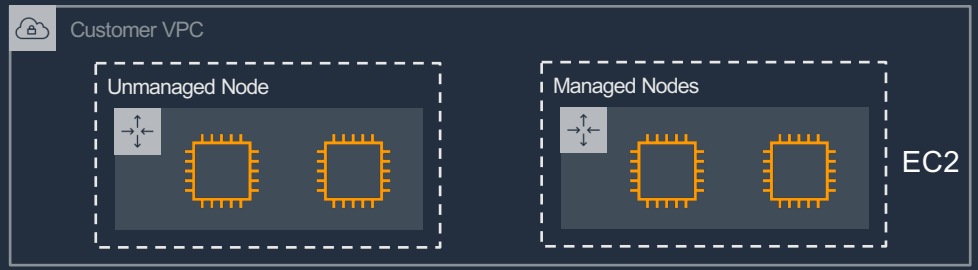
无缝集成

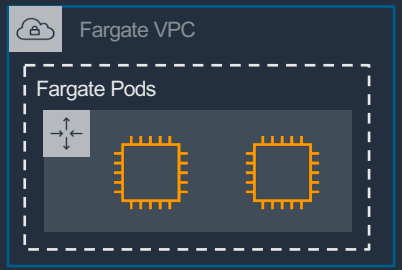
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EKS 整体架构



EKS 托管控制平面





EKS 数据平面

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws 在线研讨会

Amazon EKS on Fargate

Bring existing pods

- 你不需要改变现有的 Pod
- Fargate能与Kubernetes上运行的现有工作流和服务协同工作

Production ready

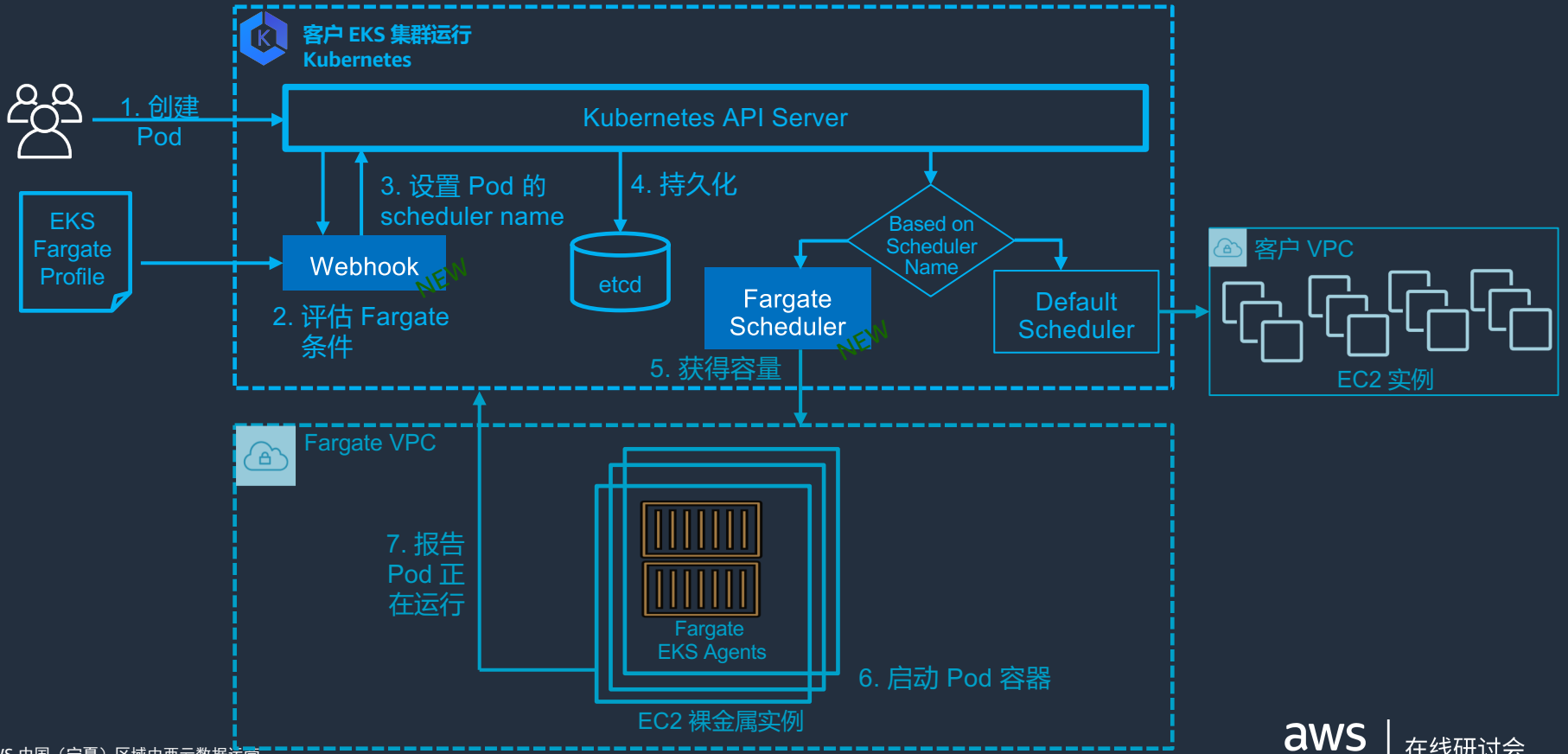
- 快速启动 Pod,并可以轻松的跨多个可用区运行 Pod实现高可用
- 每个 Pod 都运行在独立的隔离运行环境

Rightsized and integrated

- 只需按照你实际运行 Pod 所消耗的资源付费
- 包含原生 AWS 网络和安全的集成

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EKS on Fargate 控制平面架构



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Firecracker-Containerd 开源数据平面

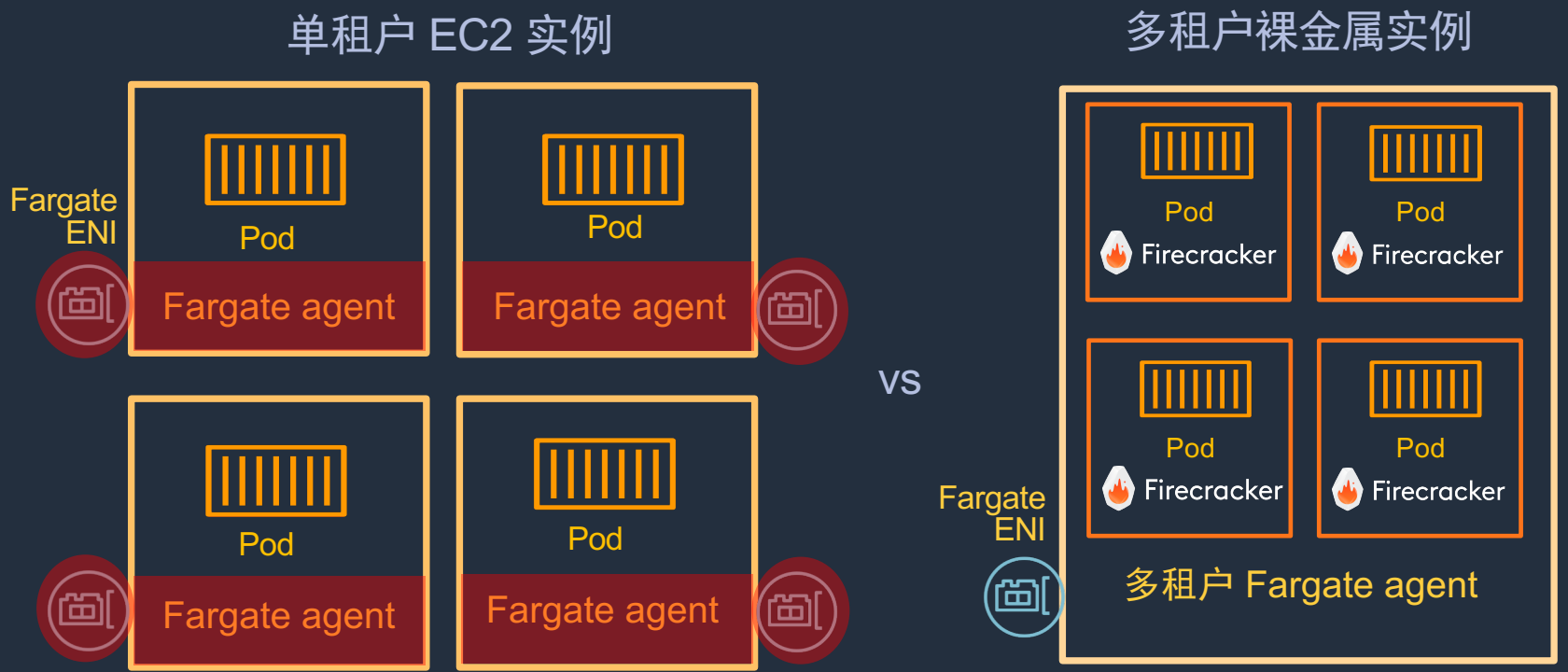
通过将 Docker 与 Conatinerd 等相关组件移到 Firecracker MicroVM外部从而进一步减少资源开销

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws | 在线研讨会

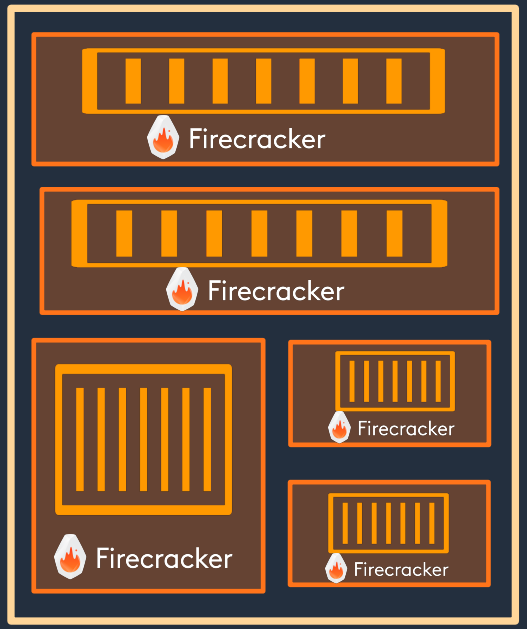
利用 Firecracker 云上数据平面开销

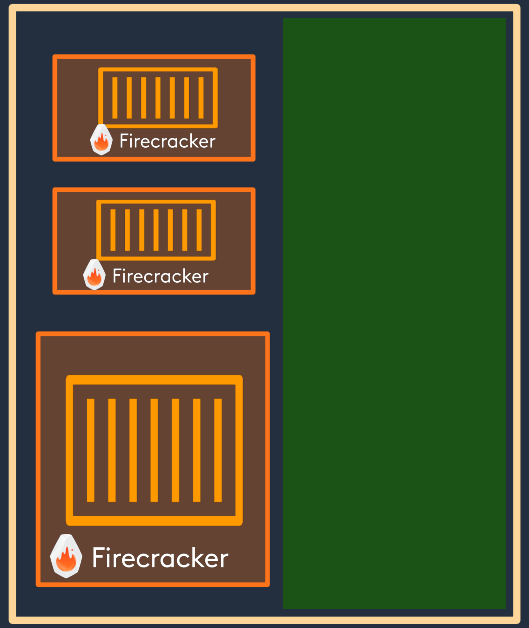


© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

利用 AWS Firecracker 提高资源利用率

多租户支持和各种规格的 MicroVM 混部获得更高的资源利用率

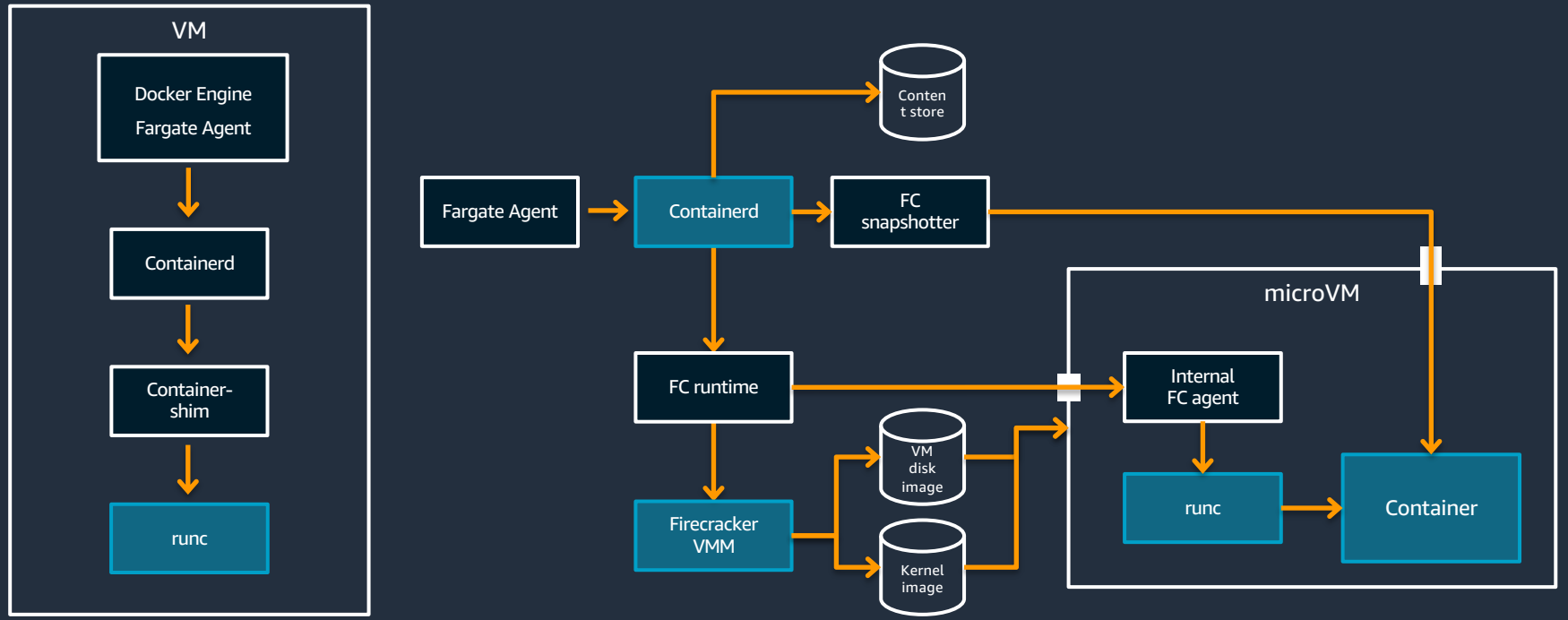




同构的 EC2裸金属实例集群

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Firecracker Containerd 实现原理



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

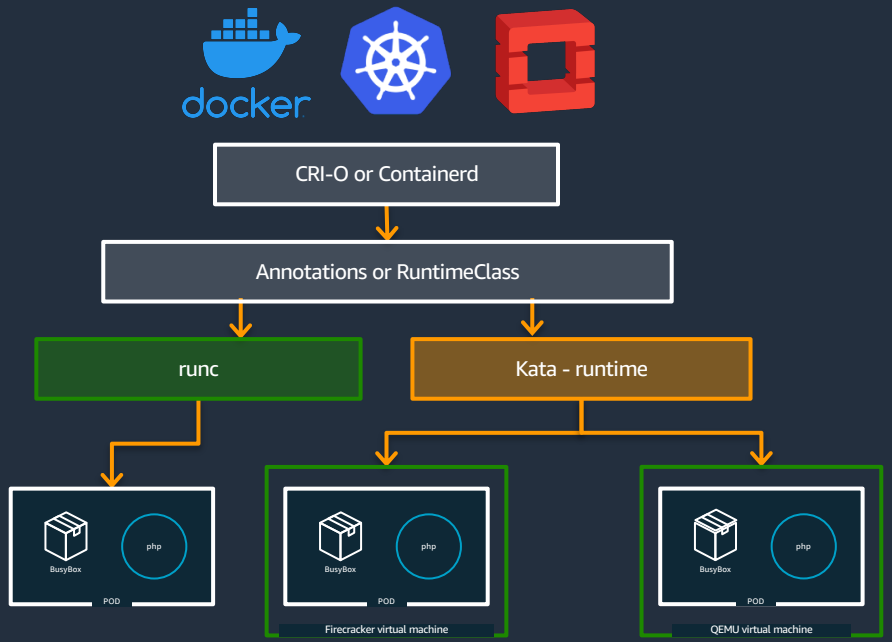
Kata Containers

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws | 在线研讨会

Firecracker 与 Kata Containers



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Weave Ignite

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

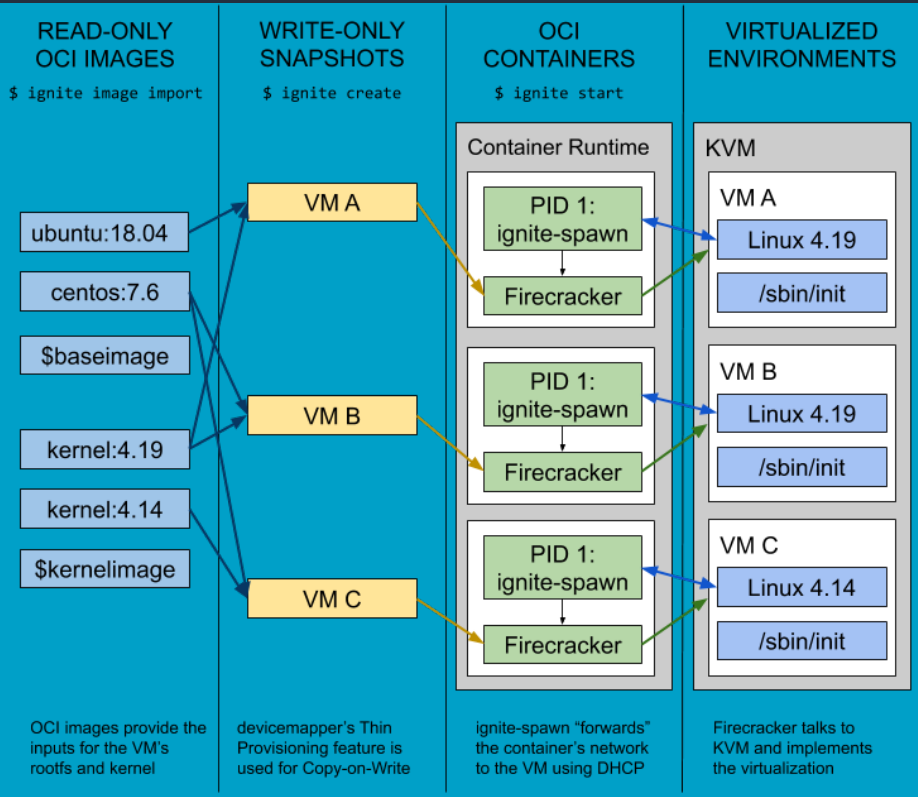
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

WeaveWorks 公司开源项目

- Ignite
  - 虚拟机管理工具提供类似容器的UX体验已经内建 GitOps 管理功能
  - Vagrant 和 Docker 的替代品
  - 提供 high-level API
- Footloose
  - 类似 Docker-compose 的方式创建 MicroVM 集群
- Firekube
  - 快速, 精简, 安全地创建 Kubernetes 集群
  - Firekube 使用 Weave Ignite 在 MicroVM 上运行 Kubernetes 就像运行容器一样简单

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Weave Ignite 设计



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

多元化的现代化容器运行时方案

- Micro VM:
  - AWS Firecracker-containerd
  - Microsoft runhcs
  - Alibaba Pouch
  - VMware VIC / Project Pacific
  - OpenStack Kata Containers
  - Hyper runV
- Sandbox
  - Google gVisor
- UniKernel
  - Nabla Containers

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

其它相关的开源项目

- Rust-VMM
  - 像玩乐高一样使用虚拟化模块
- UniK
  - Unikernel 和 microVM 的编译和编排平台
- OSv
  - 通用的模块化 unikernel 设计用于在云中的 microVM 上安全地运行未经修改的 Linux 应用程序

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Firecracker 实验

- https://ignite-your-firecracker.s3.amazonaws.com/quickstart.html

Amazon EKS on Fargate

- 安装 eksctl 工具
- > eksctl create cluster --fargate

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserve

Thank you!

莫梓元
moziyuan@amazon.com

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

感谢参加 AWS 在线研讨会

我们希望您喜欢今天的内容！
也请帮助我们完成**反馈问卷**。
欲获取关于 AWS 的更多信息和技术内容，可以通过以下方式找到我们：

- 微信订阅号：AWS 云计算（awschina）
- 微信服务号：AWS Builder 俱乐部（amazonaws）
- 新浪微博：https://www.weibo.com/amazonaws/
- 抖音：Amazon云计算（抖音号：266052872）
- 视频中心：http://aws.amazon.bokecc.com/
- 博客：https://amazonaws-china.com/cn/blogs/china
- 更多线上活动：https://aws.amazon.com/cn/about-aws/events/webinar/

AWS 中国账户注册

AWS 全球账户注册

AWS 中国（宁夏）区域由西云数据运营
AWS 中国（北京）区域由光环新网运营

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws | 在线研讨会