

Playing with AWS Firecracker VMM

之大熱天捲起袖子動手玩

...

Ernest Chiang @ COSCUP 2020, Track: Cloud Native Hub

Give me a place to stand on, and I will move the Earth.

—Archimedes

sli.do

- #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm
- #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm
- #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm #awsvmm
- 議程中有任何問題、好奇、疑問，都可以隨時丟進 sli.do
- US\$25 AWS Credits 問券連結，也放在 sli.do 裡頭喔



Ernest Chiang

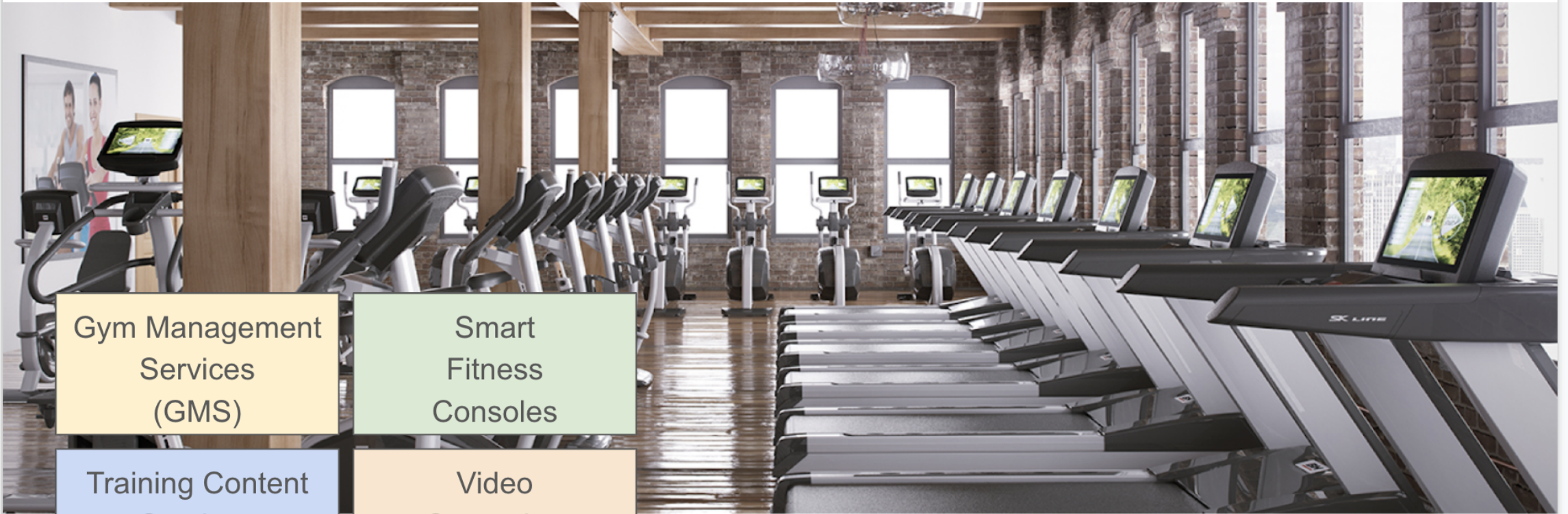
Worked on process integration engineering in semiconductor industry @tsmc.

Doing product and technology integration in fitness industry @pafers.

Off Work TGO Networks Taipei. AWS Community Hero. Mozillian. AIESECer.

PAFERS Fitness Services (PFS)

The integrated products and services for fitness.



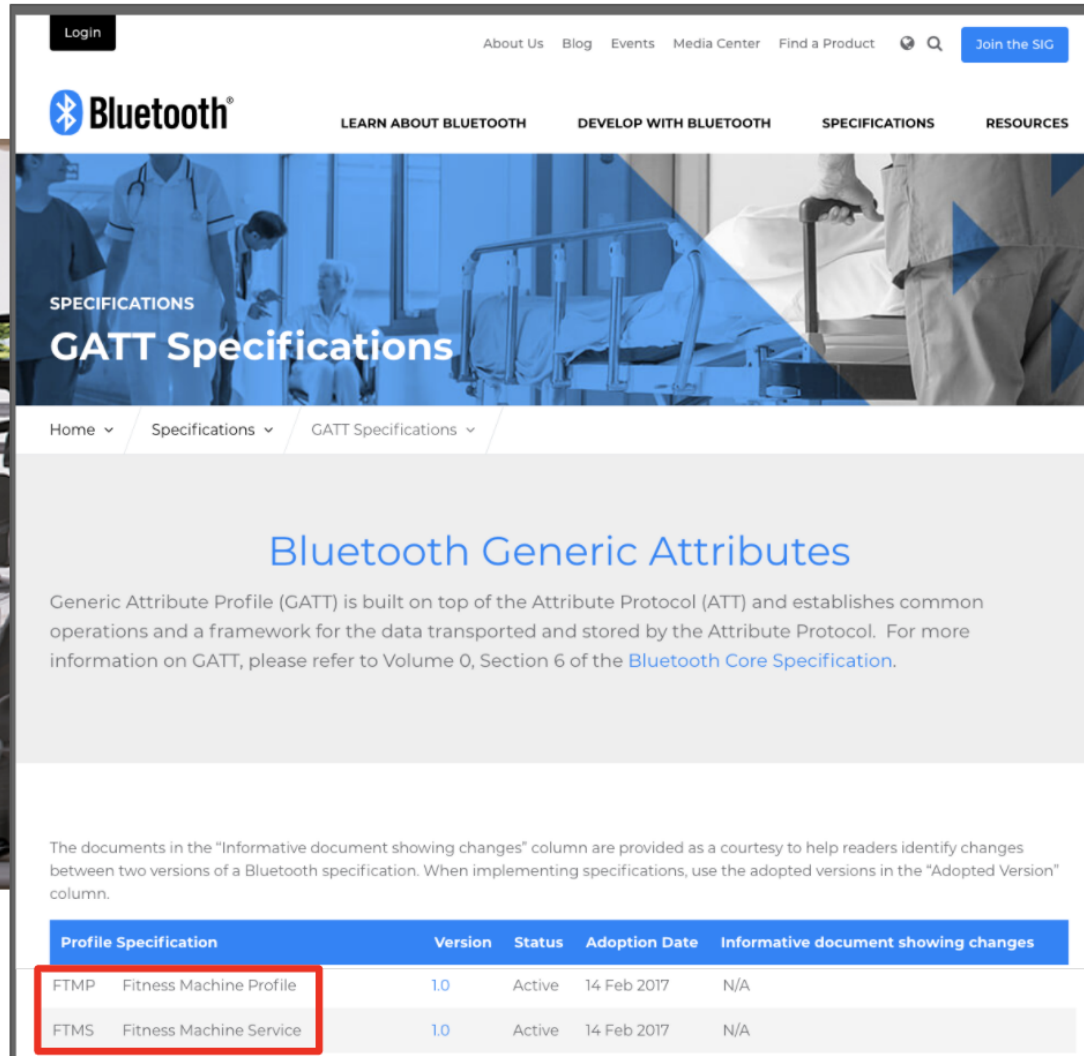
Gym Management
Services
(GMS)

Smart
Fitness
Consoles

Training Content
Services
(TCS)

Video
Streaming
Services

PAFERS Contributes Industry Knowledge to Bluetooth Spec



Bluetooth

LEARN ABOUT BLUETOOTH DEVELOP WITH BLUETOOTH SPECIFICATIONS RESOURCES

SPECIFICATIONS
GATT Specifications

Home Specifications GATT Specifications

Bluetooth Generic Attributes

Generic Attribute Profile (GATT) is built on top of the Attribute Protocol (ATT) and establishes common operations and a framework for the data transported and stored by the Attribute Protocol. For more information on GATT, please refer to Volume 0, Section 6 of the [Bluetooth Core Specification](#).

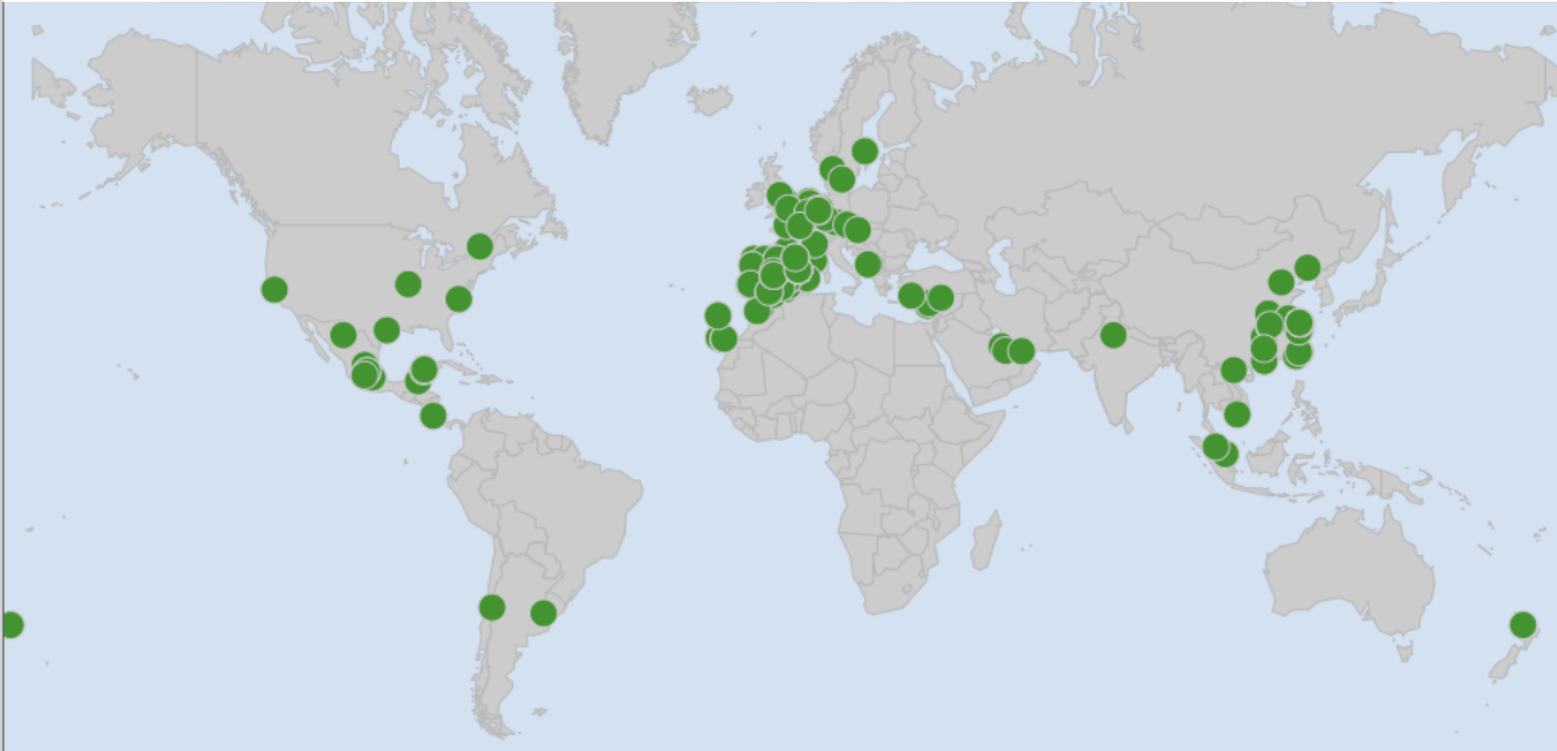
The documents in the "Informative document showing changes" column are provided as a courtesy to help readers identify changes between two versions of a Bluetooth specification. When implementing specifications, use the adopted versions in the "Adopted Version" column.

Profile Specification	Version	Status	Adoption Date	Informative document showing changes
FTMP Fitness Machine Profile	1.0	Active	14 Feb 2017	N/A
FTMS Fitness Machine Service	1.0	Active	14 Feb 2017	N/A



<https://www.bluetooth.com/specifications/gatt/>

Global PAFERS Edges Footprints



Outline

- Problems & Solutions
- Firecracker
- Virtualization & Containerization
- Lambda & Fargate
- Firecracker & container **d**
- Live Demo
 - Getting started with Firecracker in 2 Minutes
 - Creating 4,000 microVMs in 90 seconds
- Firecracker & Open Source Projects

Problems & Solutions

Firecracker, Part 1



What is Firecracker

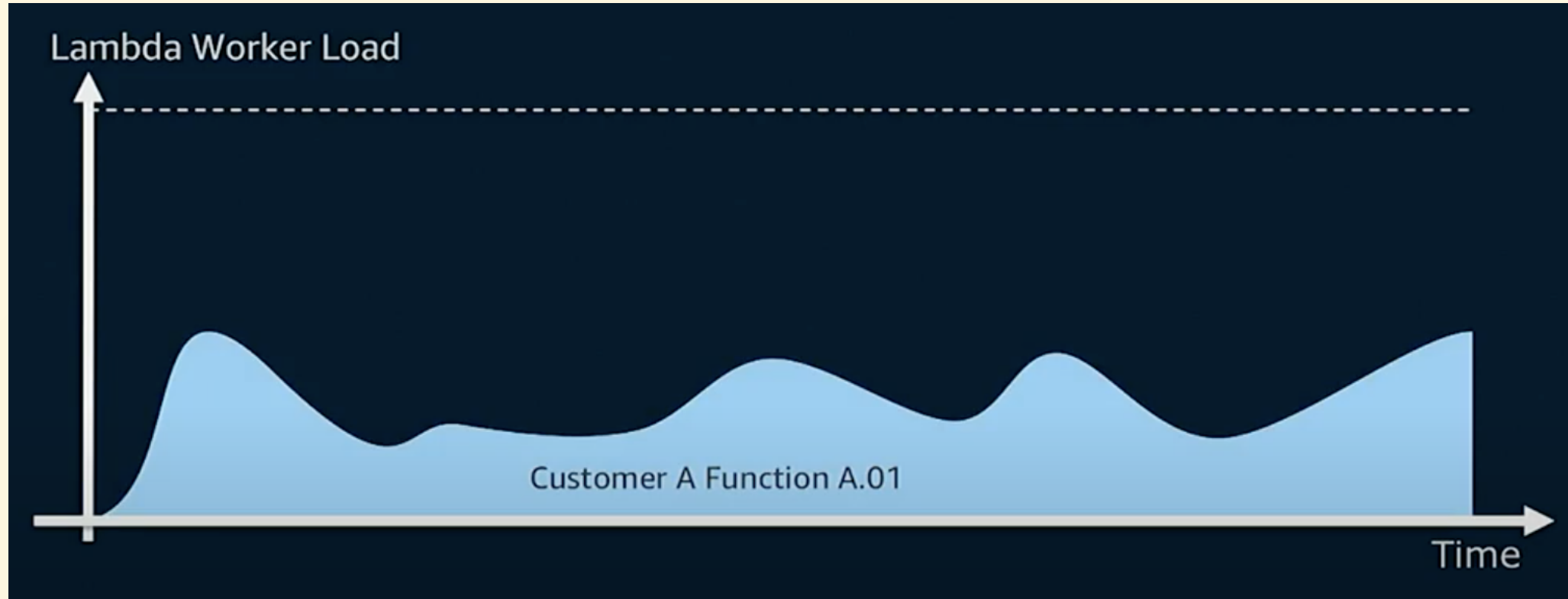
Firecracker is an open source VMM that is purpose-built for creating and managing secure, multi-tenant container and function-based services.



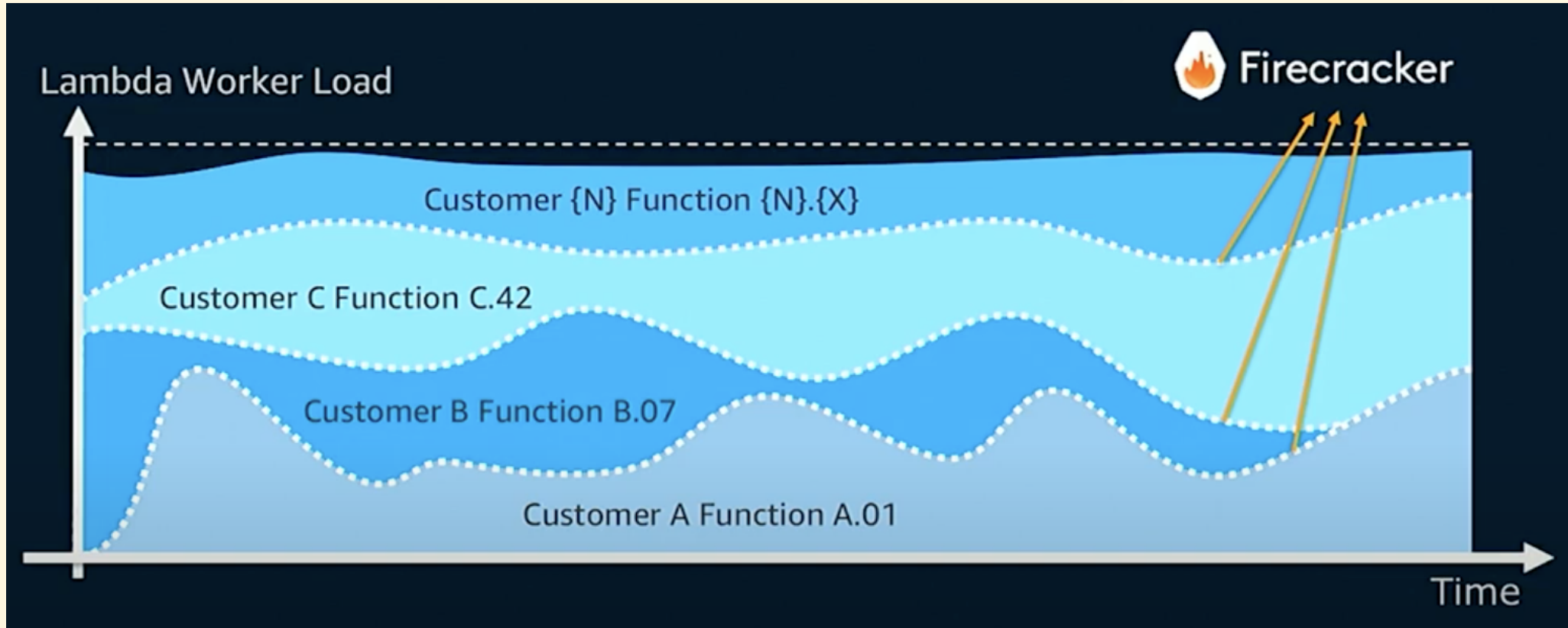
What is Firecracker

Firecracker is an open source **VMM** that is **purpose-built** for creating and managing secure, multi-tenant container and function-based services.

What problem is AWS helping to solve?



What problem is AWS helping to solve?



What problem is AWS helping to solve?

Multiple functions
on multiple environments
from multiple accounts.

What is Firecracker

- Open source virtualization technology (microVM)
- Security and isolation of traditional VMs
- Speed and density of containers
- Low resource overhead
- Developed at Amazon

Benefits of Firecracker



Security



Startup time



Utilization

Benefits of Firecracker



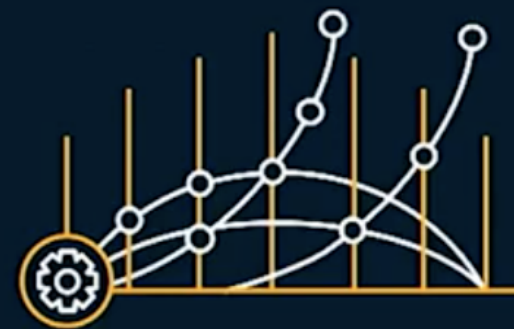
Security from
the ground up

KVM-based
virtualization



Speed
by design

<125ms to launch 150
microVMs per second/host



Scale
and efficiency

<5MB memory
footprint per microVM

安全隔離好

啟動時間短

產能效率高

#像極了愛情

-- AWS Firecracker VMM

Virtualization & Containerization

Virtualization (1/3)

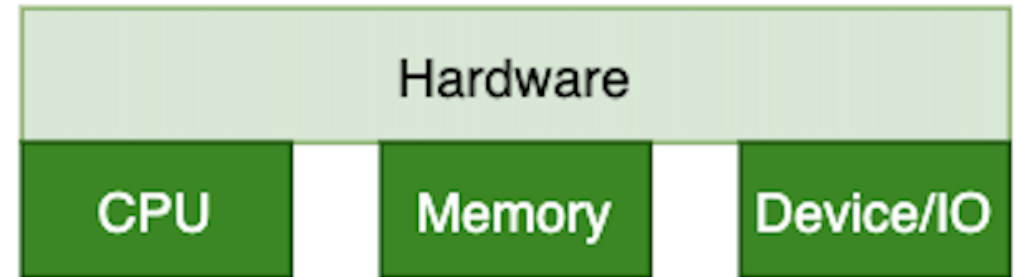
In computing, virtualization refers to the act of **creating a virtual (rather than actual) version of something**, including virtual computer hardware platforms, storage devices, and computer network resources.

Virtualization

(2/3)

Creating a **virtual** version of **something**:

- CPU
- Memory
- Device/IO (Storage, NIC)



Virtualization (3/3)

V · T · E		Virtualization software		[hide]
Comparison of platform virtualization software				
Hardware virtualization (hypervisors)	Native	Adeos · CP/CMS · Hyper-V · KVM (Red Hat Enterprise Virtualization) · LDoms / Oracle VM Server for SPARC · Logical Partition (LPAR) · LynxSecure · PikeOS · Proxmox VE · SIMMON · VMware ESXi (VMware vSphere · vCloud) · VMware Infrastructure · Xen (Oracle VM Server for x86 · XenServer) · XtratuM · z/VM		
		Hosted	Specialized	Basilisk II · bhyve · Bochs · Cooperative Linux · DOSBox · DOSEMU · PCem · PikeOS · SheepShaver · SIMH · Windows on Windows (Virtual DOS machine) · Win4Lin
	Independent		Microsoft Virtual Server · Parallels Workstation · Parallels Desktop for Mac · Parallels Server for Mac · PearPC · QEMU · VirtualBox · Virtual Iron · VMware Fusion · VMware Server · VMware Workstation (Player) · Windows Virtual PC	
	Tools	Ganeti · oVirt · System Center Virtual Machine Manager · Virtual Machine Manager		
OS-level virtualization	OS containers	FreeBSD jail · iCore Virtual Accounts · Linux-VServer · LXC · OpenVZ · Solaris Containers · Virtuozzo · Workload Partitions		
	Application containers	Docker · Imctfy · rkt		
	Virtual kernel architectures	Rump kernel · User-mode Linux · vkernel		
	Related kernel features	BrandZ · cgroups · chroot · namespaces · seccomp		
	Orchestration	Amazon ECS · Kubernetes · OpenShift		
Desktop virtualization	Citrix XenApp · Citrix XenDesktop · Remote Desktop Services · VMware Horizon View · Ulteo Open Virtual Desktop			
Application virtualization	Ceedo · Citrix XenApp · Dalvik · InstallFree · Microsoft App-V · Remote Desktop Services · Symantec Workspace Virtualization · Turbo · VMware ThinApp · ZeroVM			
Network virtualization	Distributed Overlay Virtual Ethernet (DOVE) · Ethernet VPN (EVPN) · NVGRE · Open vSwitch · Virtual security switch · Virtual Extensible LAN (VXLAN)			
See also: List of emulators				

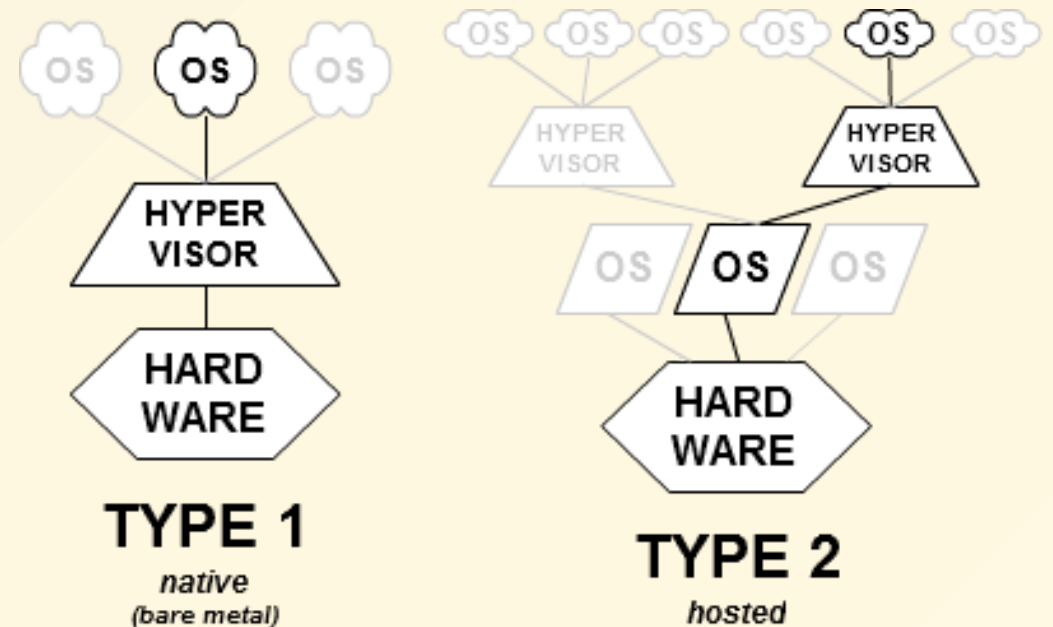
Hypervisor (1/6)

A `hypervisor` (or `virtual machine monitor`, `VMM`, virtualizer) is computer software, firmware or hardware that creates and runs virtual machines.

Hypervisor (2/6)

In 1974, Gerald J. Popek and Robert P. Goldberg classified two types of hypervisor:

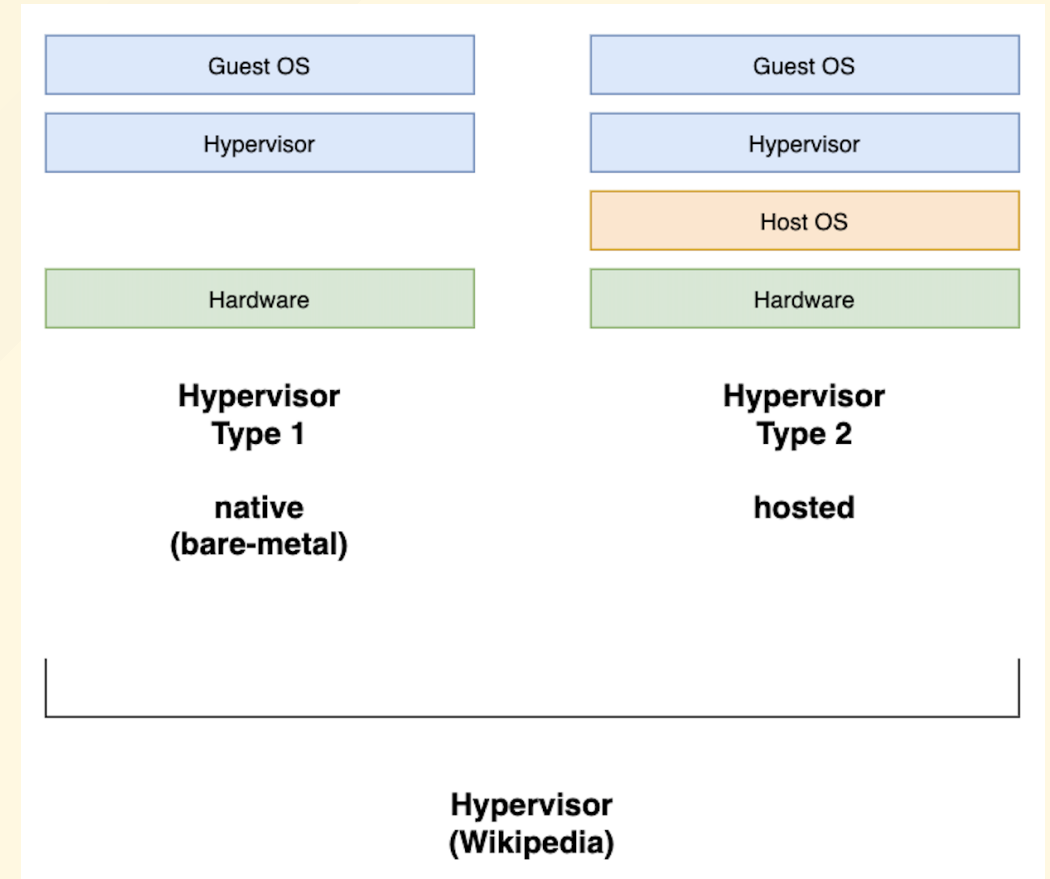
- Type-1, native or bare-metal hypervisors
- Type-2 or hosted hypervisors



Hypervisor (3/6)

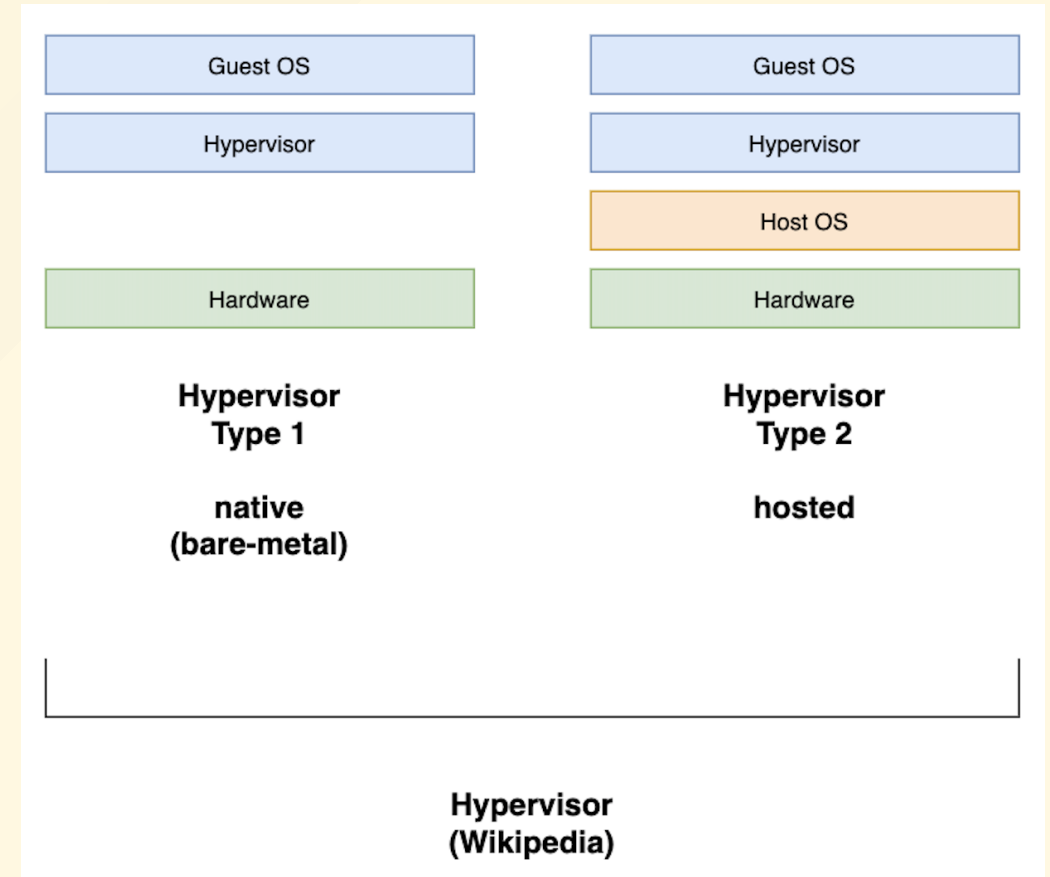
The distinction between these two types is not always clear.

For instance, Linux's Kernel-based Virtual Machine (KVM) and FreeBSD's bhyve are kernel modules that effectively convert the host operating system to a type-1 hypervisor.

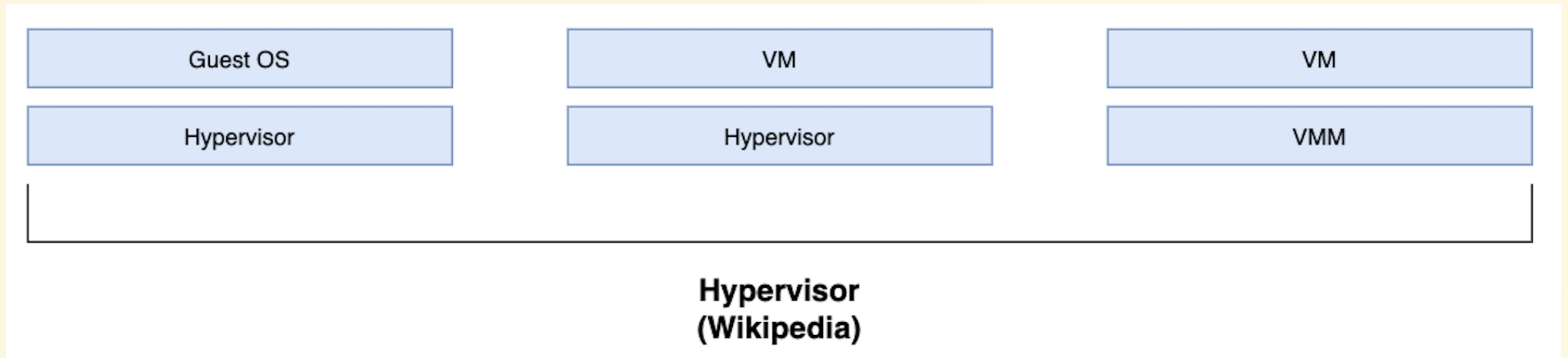


Hypervisor (4/6)

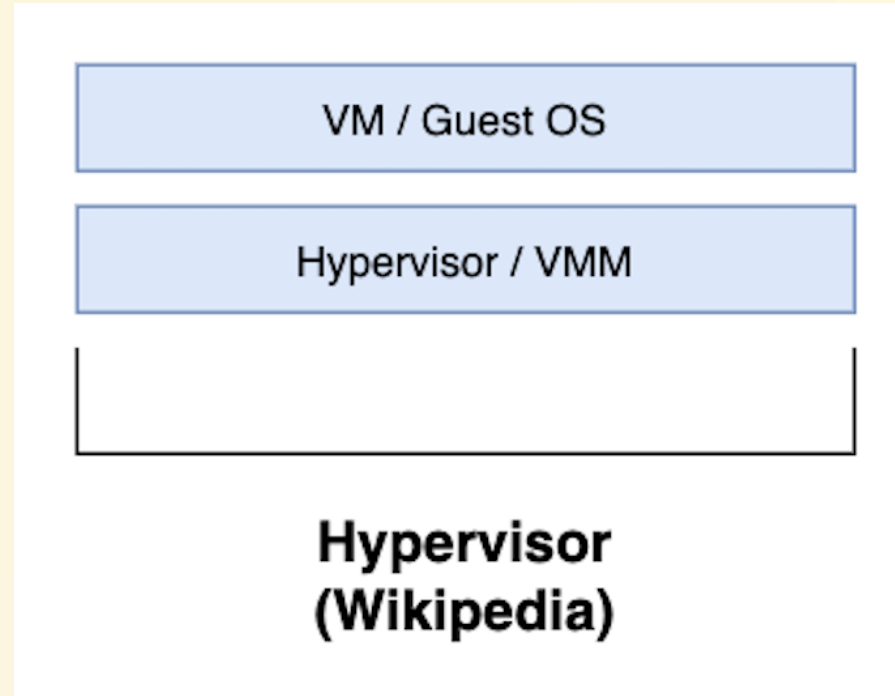
At the same time, since Linux distributions and FreeBSD are still general-purpose operating systems, with applications competing with each other for VM resources, **KVM** and **bhyve** can also be categorized as type-2 hypervisors.



Hypervisor (5/6)



Hypervisor (6/6)



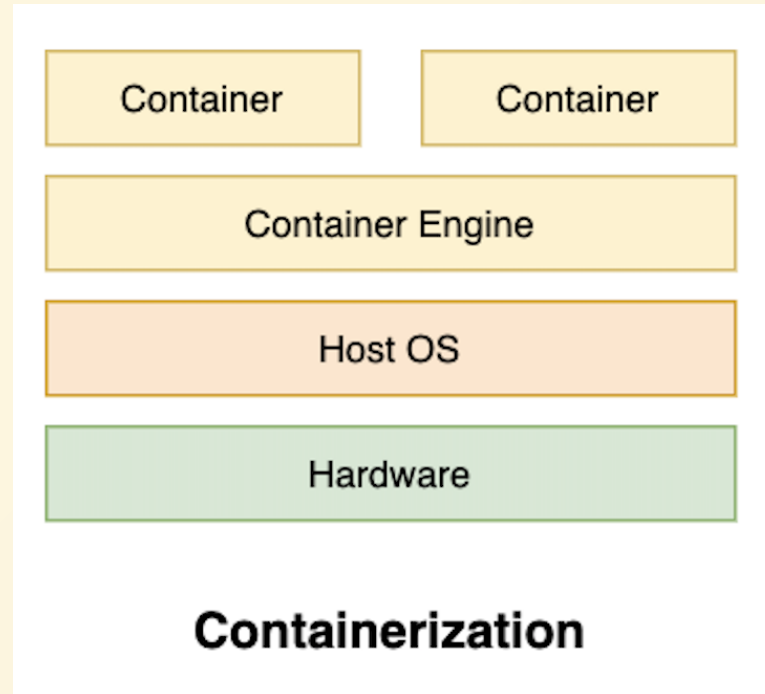
KVM

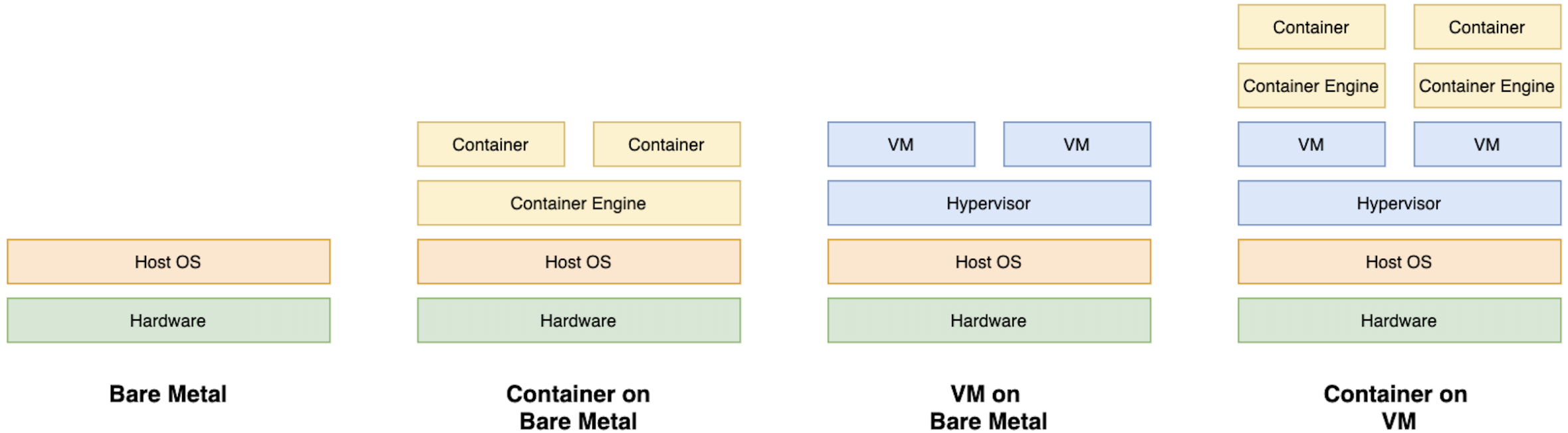
Kernel-based Virtual Machine (KVM) is a virtualization module in the Linux kernel that allows the kernel to function as a hypervisor.

Containerization

Operating-system-level virtualization, also known as containerization, refers to an operating system feature in which the kernel allows the existence of multiple isolated user-space instances. Such instances, called **containers**, partitions, virtual environments (VEs) or **jails** (FreeBSD jail or chroot jail), may look like real computers from the point of view of programs running in them.

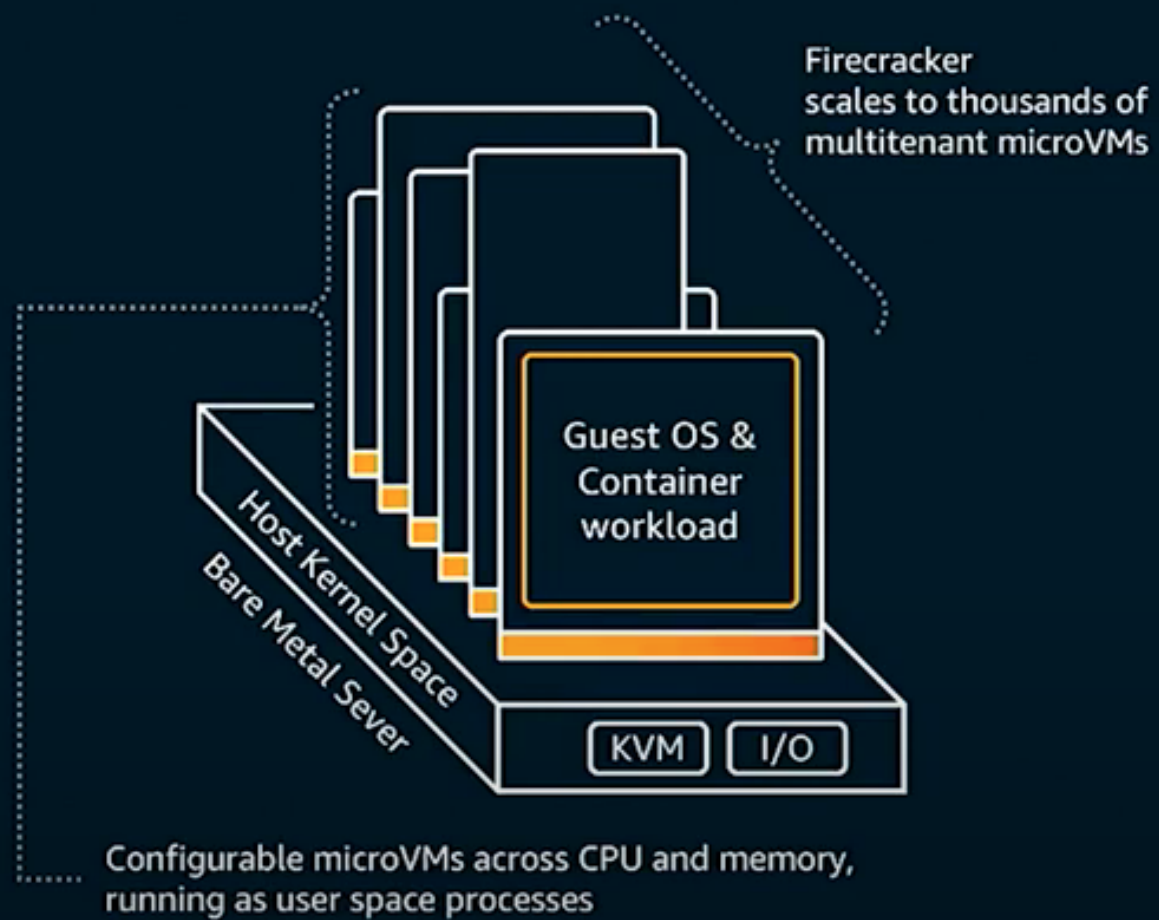
Containerization

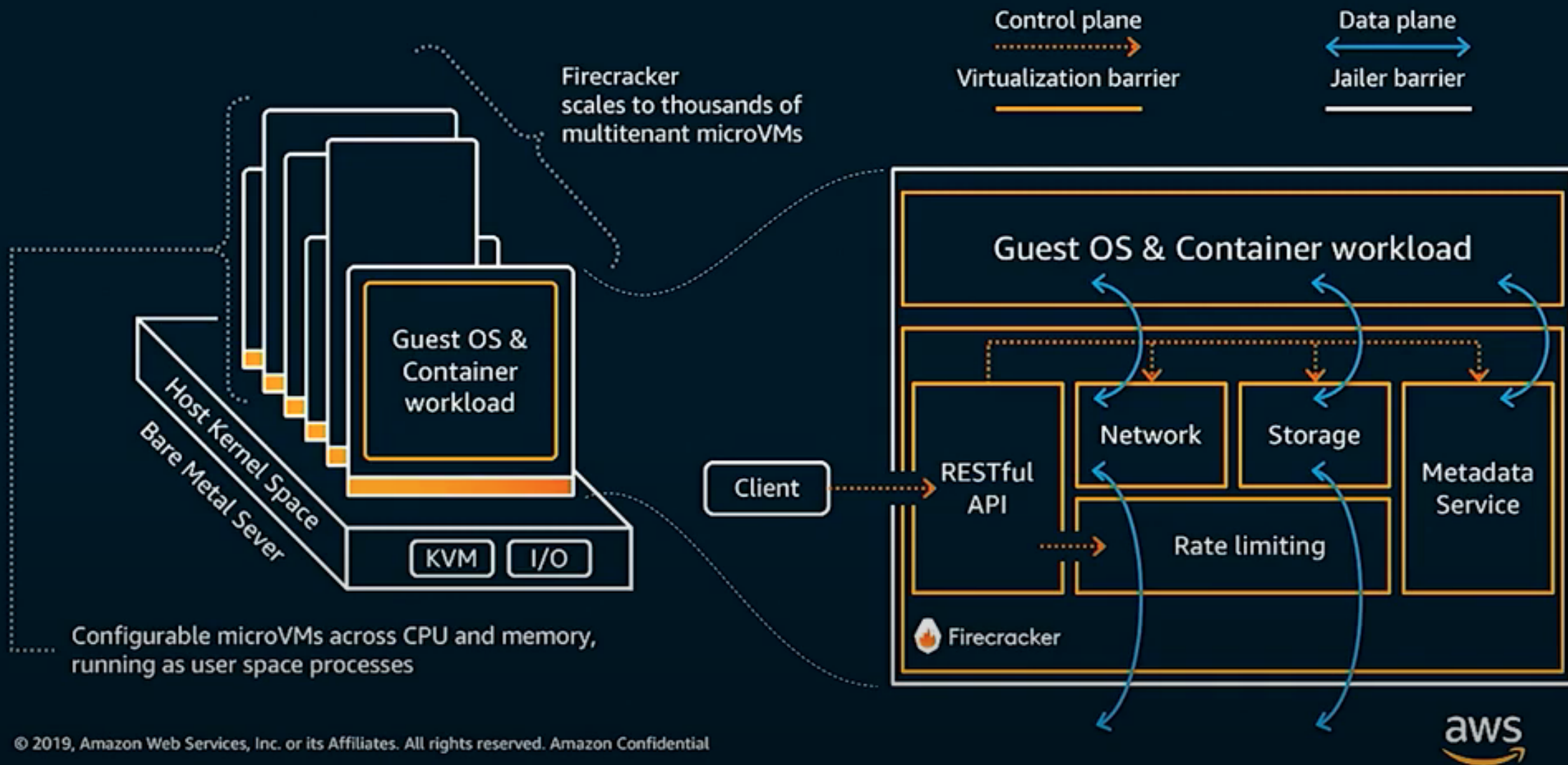


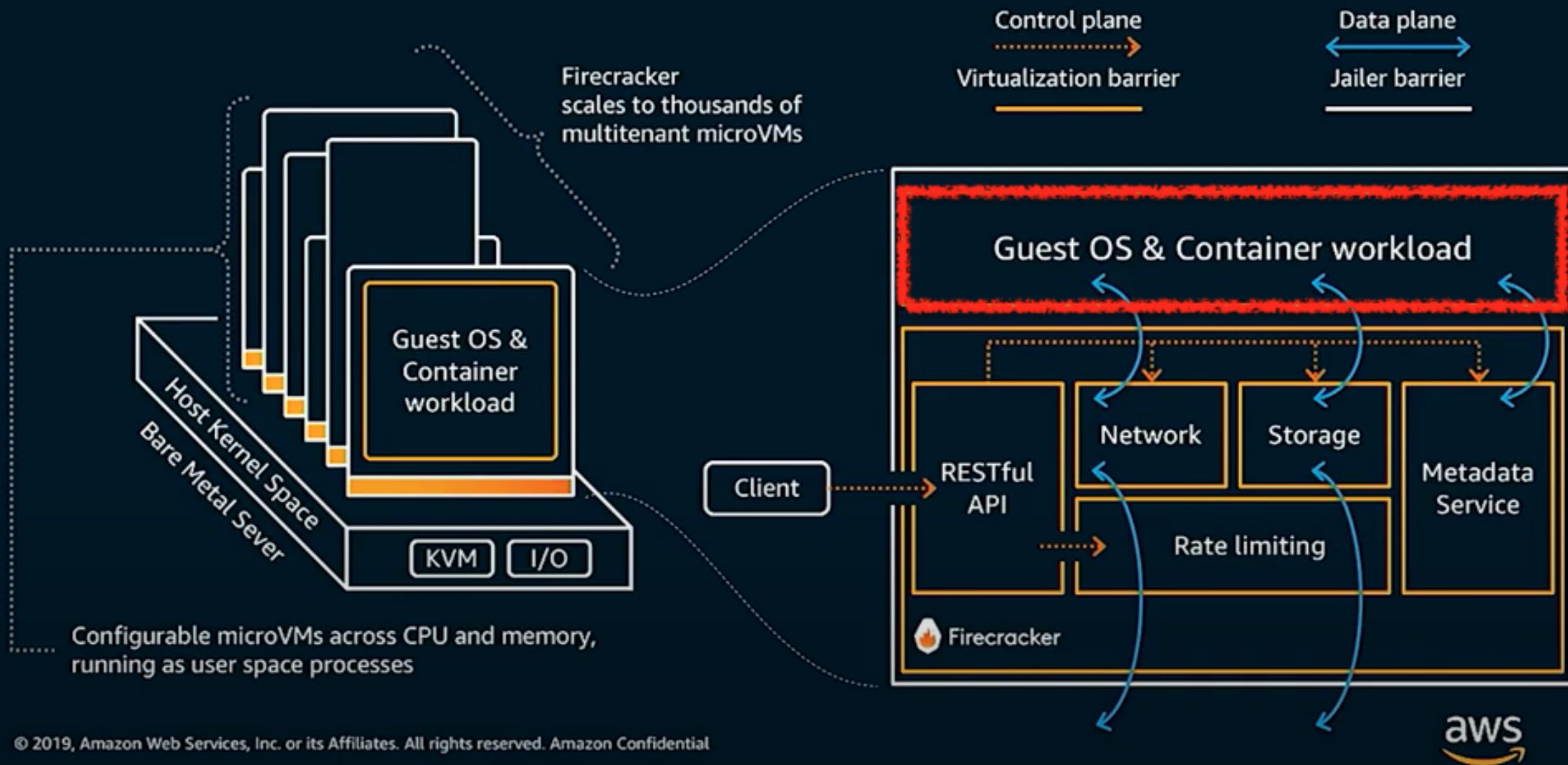


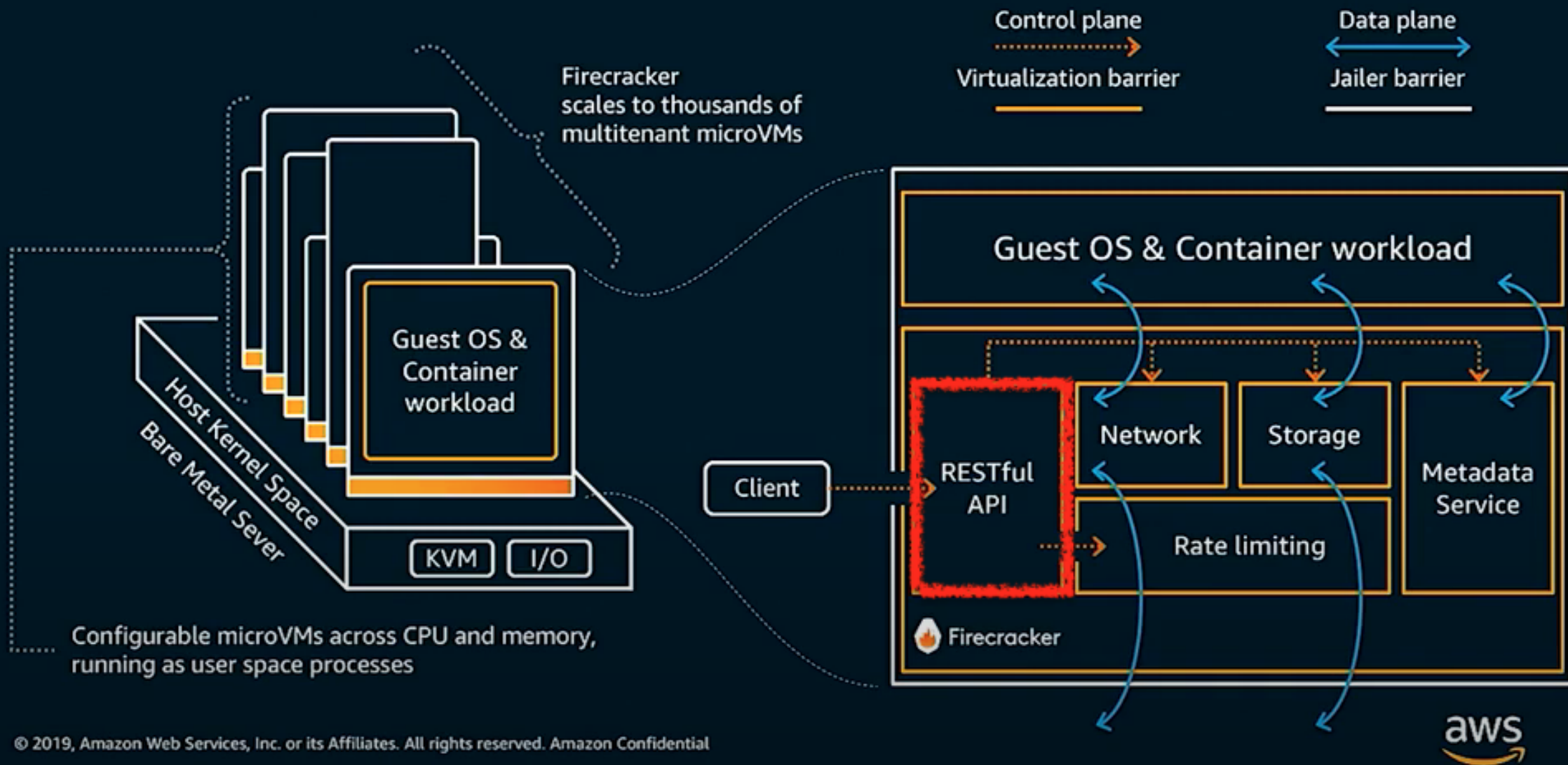
**Common
Combinations**

Firecracker, Part 2









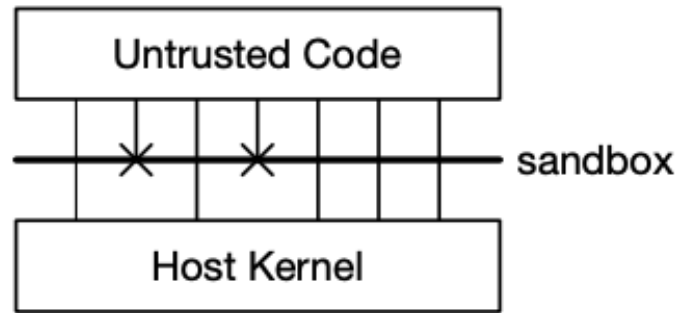
Host-facing REST API

GET / Returns general information about an instance.	GET / machine-config Gets the machine configuration of the VM.
PUT / actions Creates a synchronous action.	PUT / machine-config Updates the Machine Configuration of the VM.
PUT / boot-source Creates or updates the boot source.	PUT / mmds Creates a MMDS (Microvm Metadata Service) data store.
PUT / drives/{drive_id} Creates or updates a drive.	PATCH / mmds Updates the MMDS data store.
PATCH / drives/{drive_id} Updates the properties of a drive.	GET / mmds Get the MMDS data store.
PUT / logger Initializes the logger by specifying two named pipes (i.e. for the logs and metrics output).	PUT / network-interfaces/{iface_id} Creates a network interface.

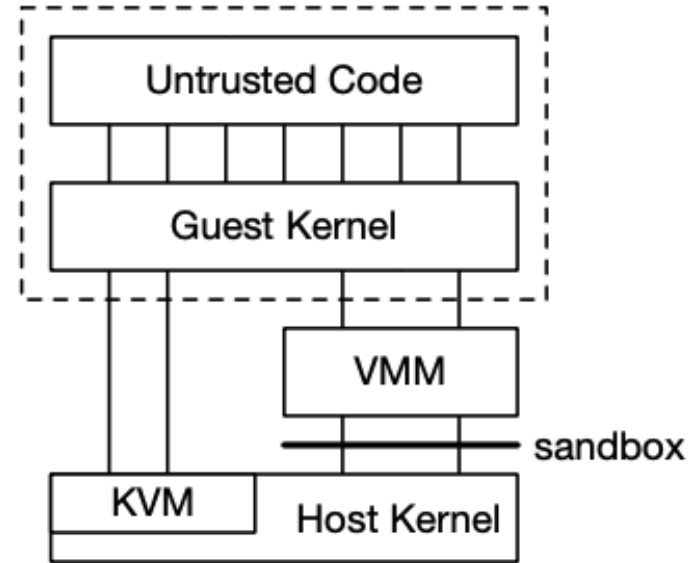
Firecracker

- Started with a branch of [crosvm](#)
 - Removed >50% of the code
- 96% fewer lines of code than QEMU
- Simplified device model
 - no BIOS, no PCI, etc
- Apache 2.0 license

Security Models (1/2)

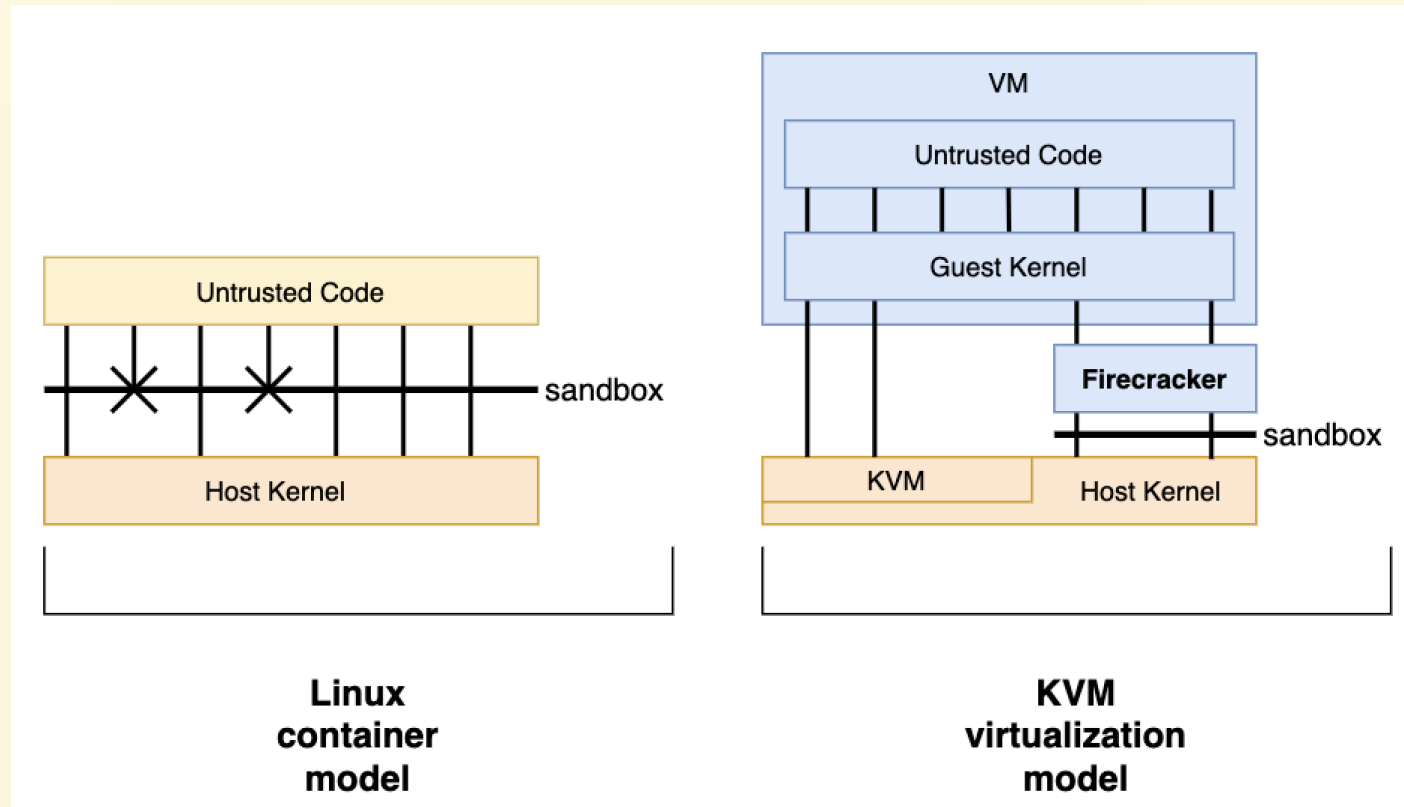


(a) Linux container model



(b) KVM virtualization model

Security Models (2/2)

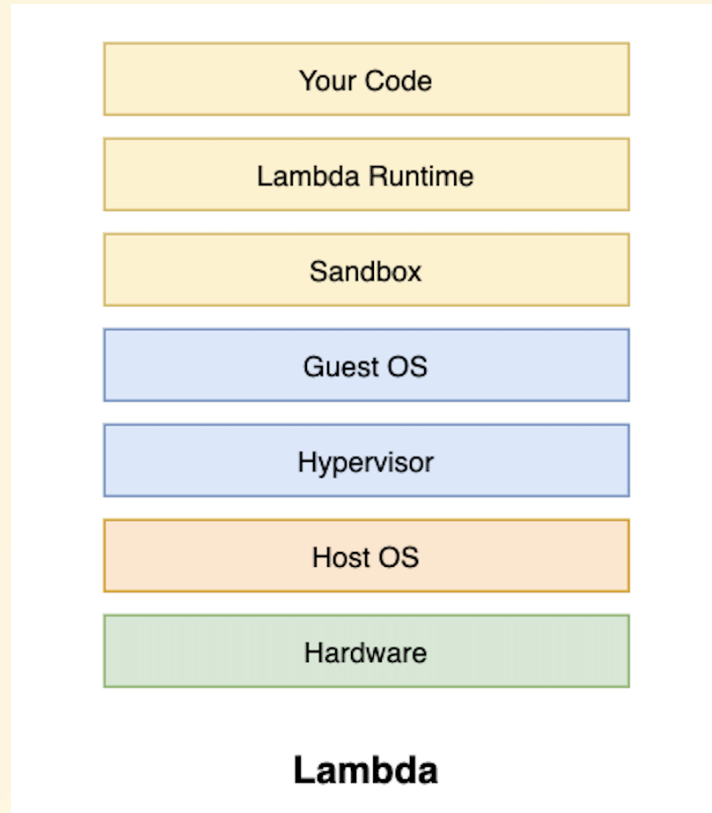


Firecracker

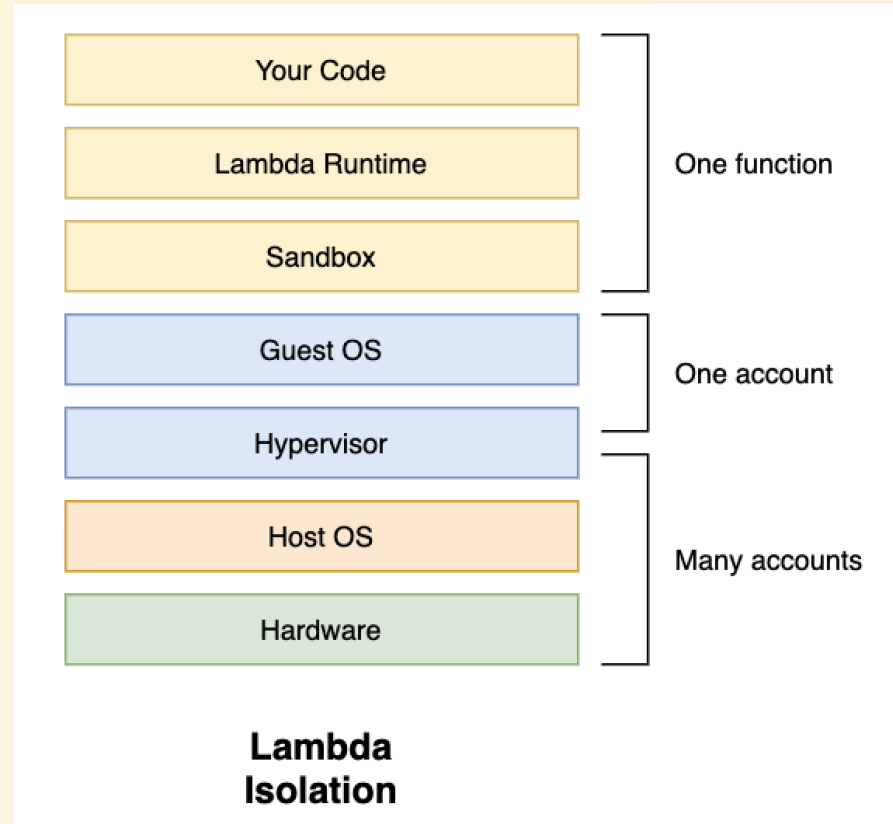
- In production in AWS Lambda
 - Millions of workloads
 - Trillions of requests/month

AWS Lambda

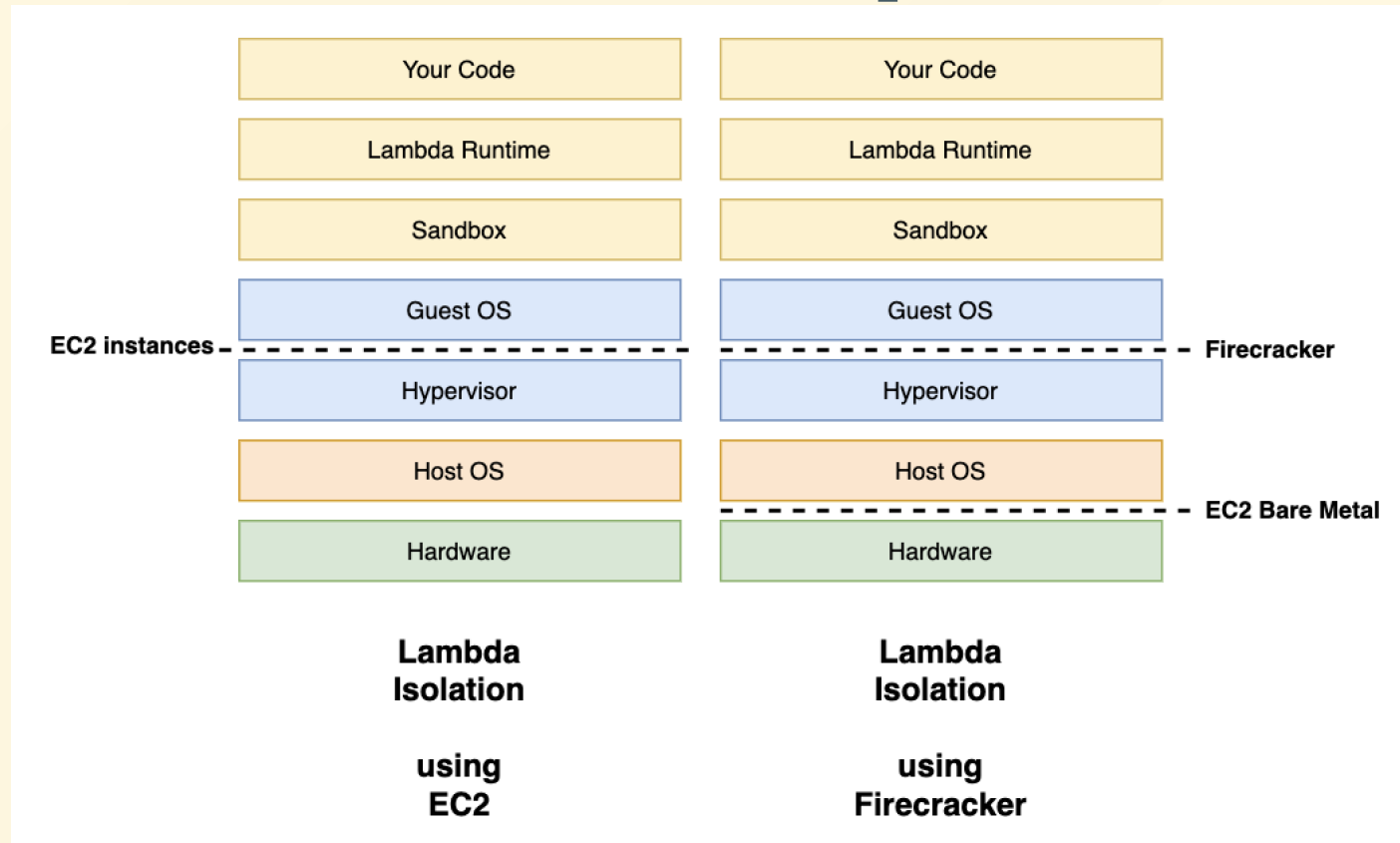
Lambda worker architecture



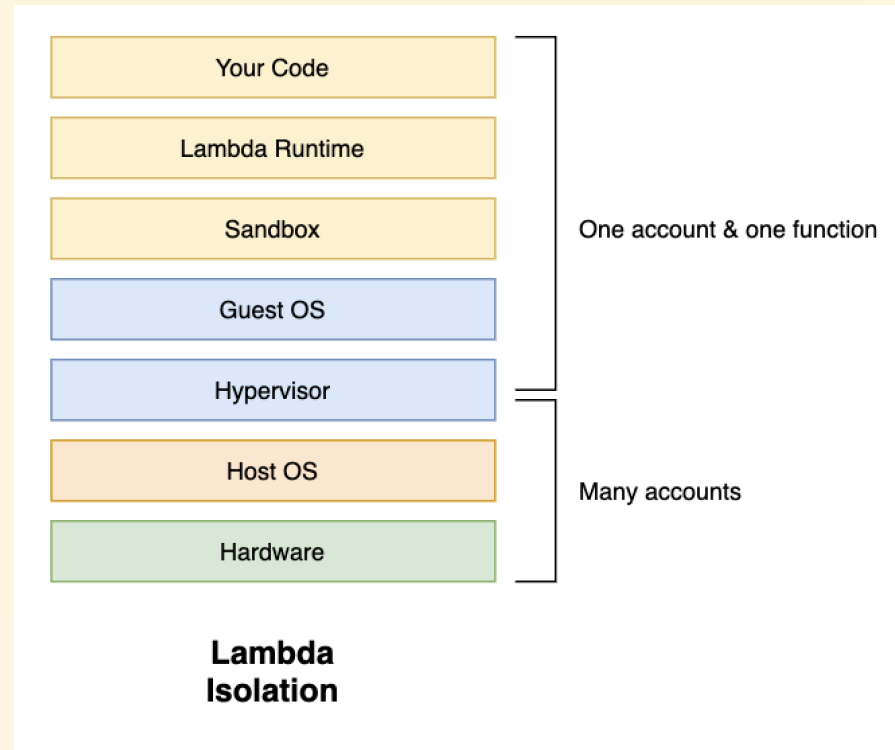
Lambda worker isolation



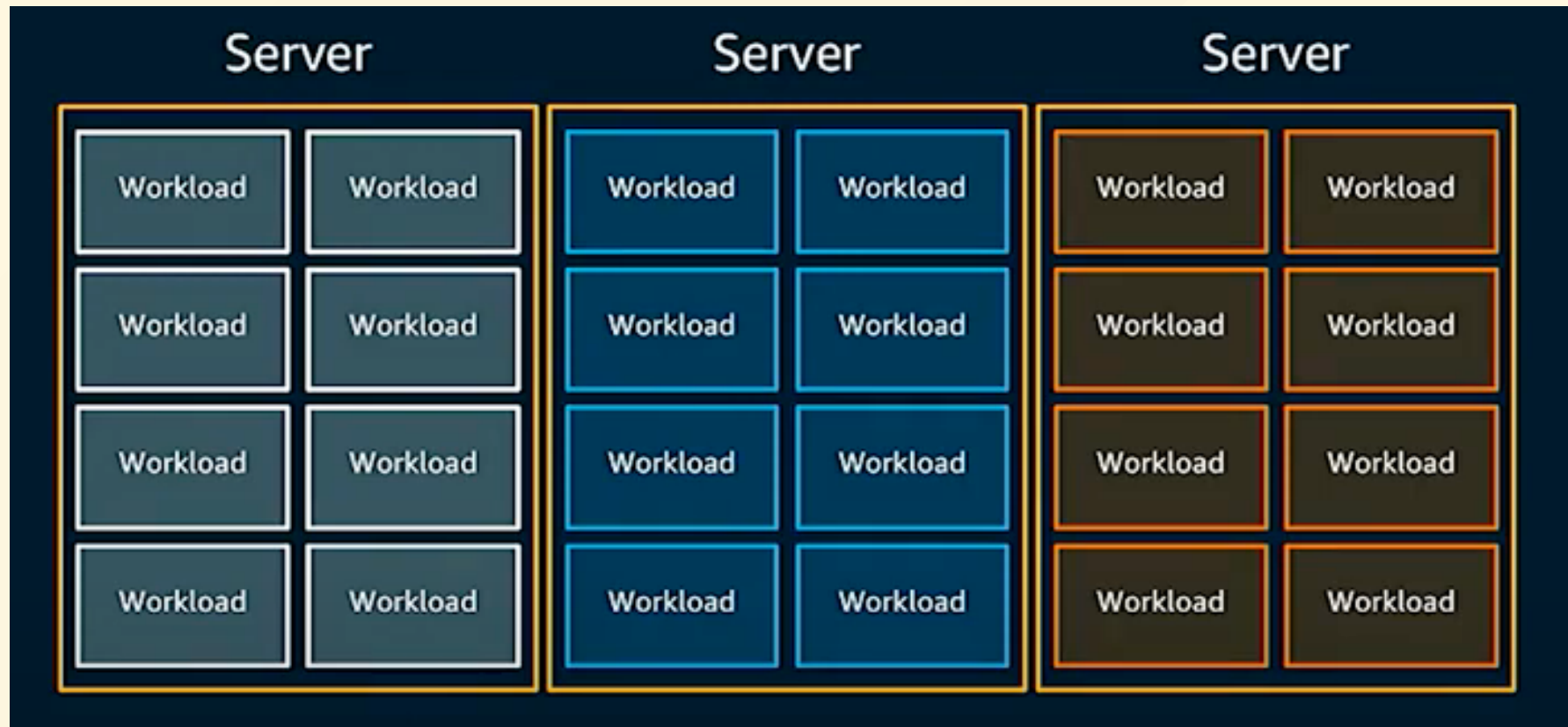
Lambda isolation comparison



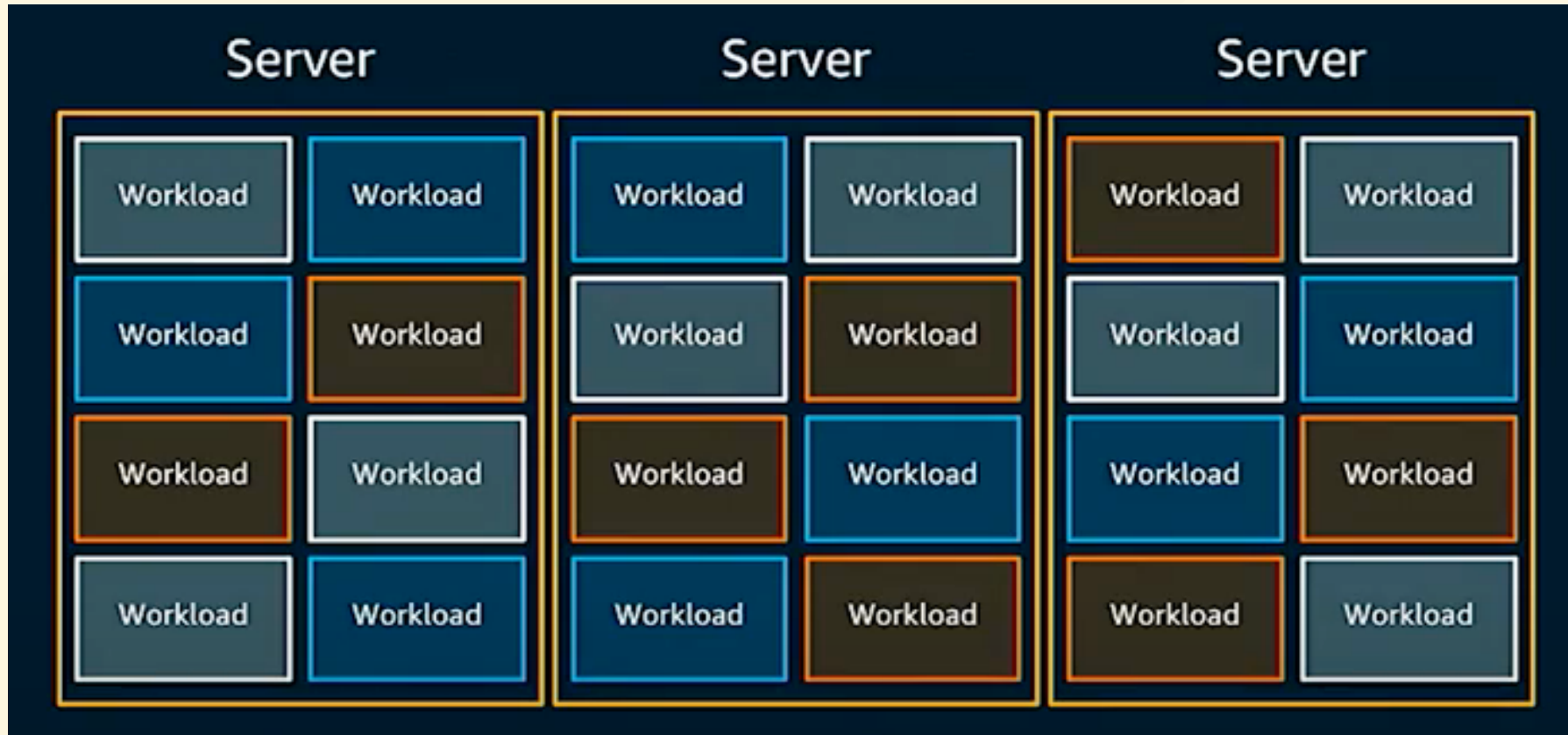
Lambda isolation using Firecracker



Allocate Workloads:



More efficient:



AWS Container Services landscape

Management

Deployment, scheduling, scaling & management of containerized applications



Amazon Elastic Container Service



Amazon Elastic Container Service for Kubernetes

Hosting

Where the containers run



Amazon EC2



AWS Fargate

Image Registry

Container image repository

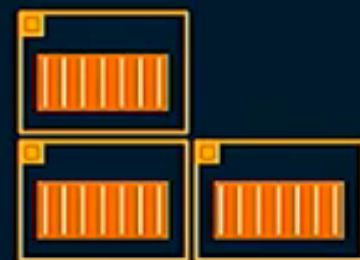





Amazon Elastic Container Registry

AWS Fargate

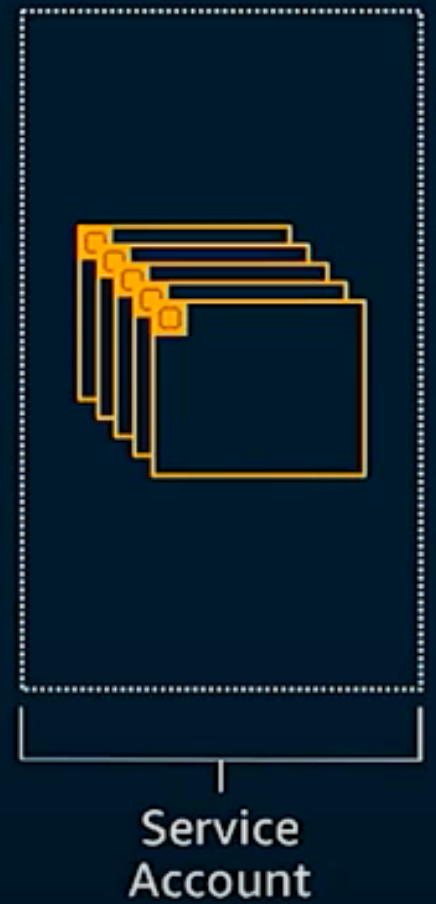
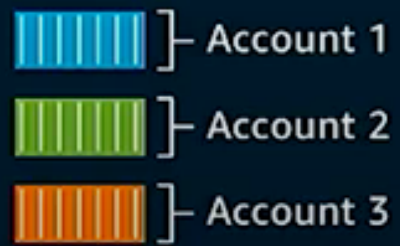
Fargate configurations

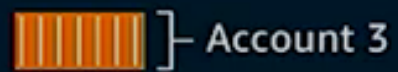
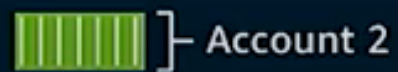
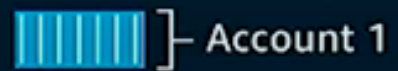
CPU (vCPU)	Memory Values (GB)
0.25	0.5, 1, 2
0.5	Min 1GB, max 4GB, in 1GB increments
1	Min 2GB, max 8GB, in 1GB increments
2	Min 4GB, max 16GB, in 1GB increments
4	Min 8GB, max 30GB, in 1GB increments

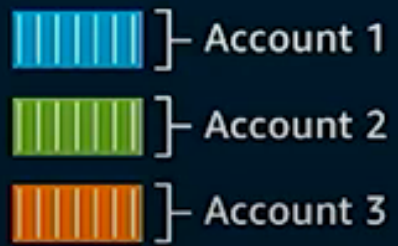
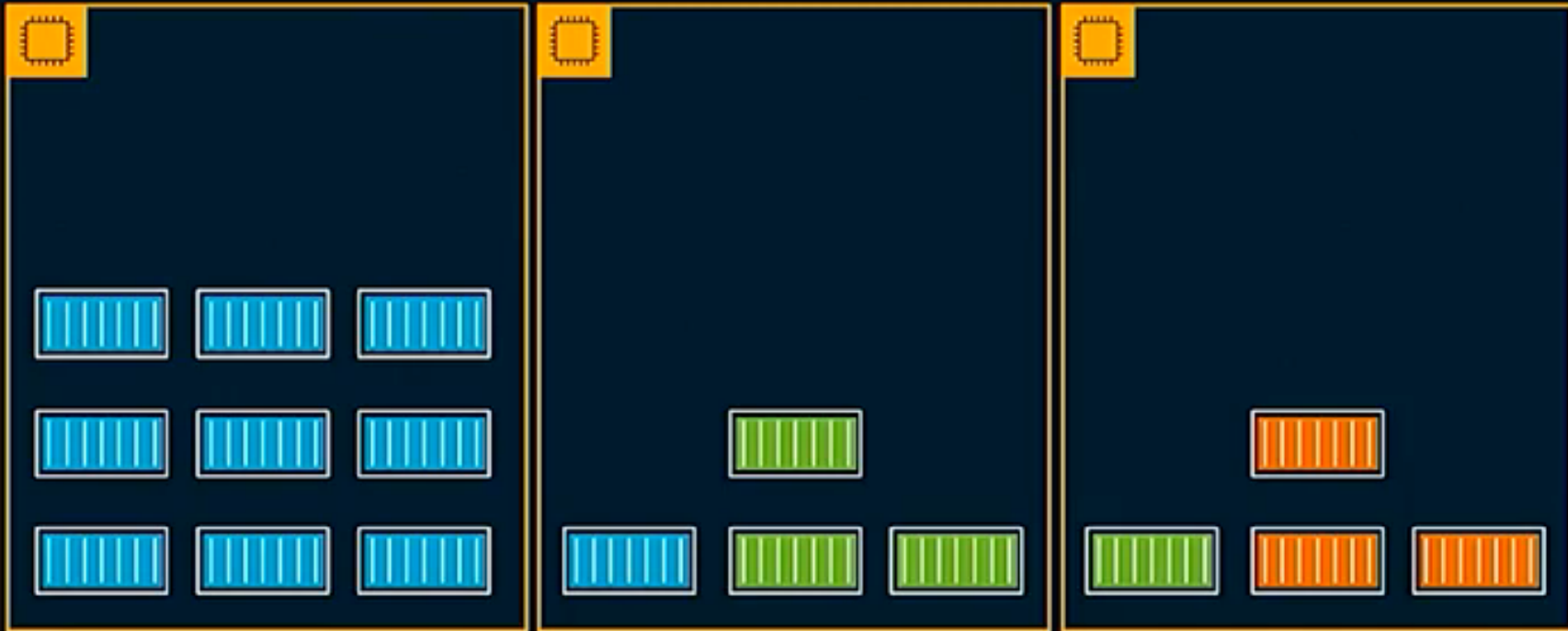


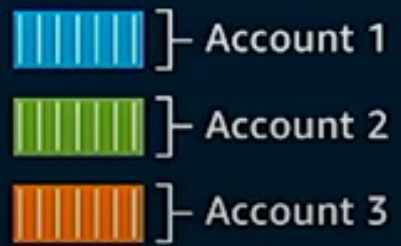
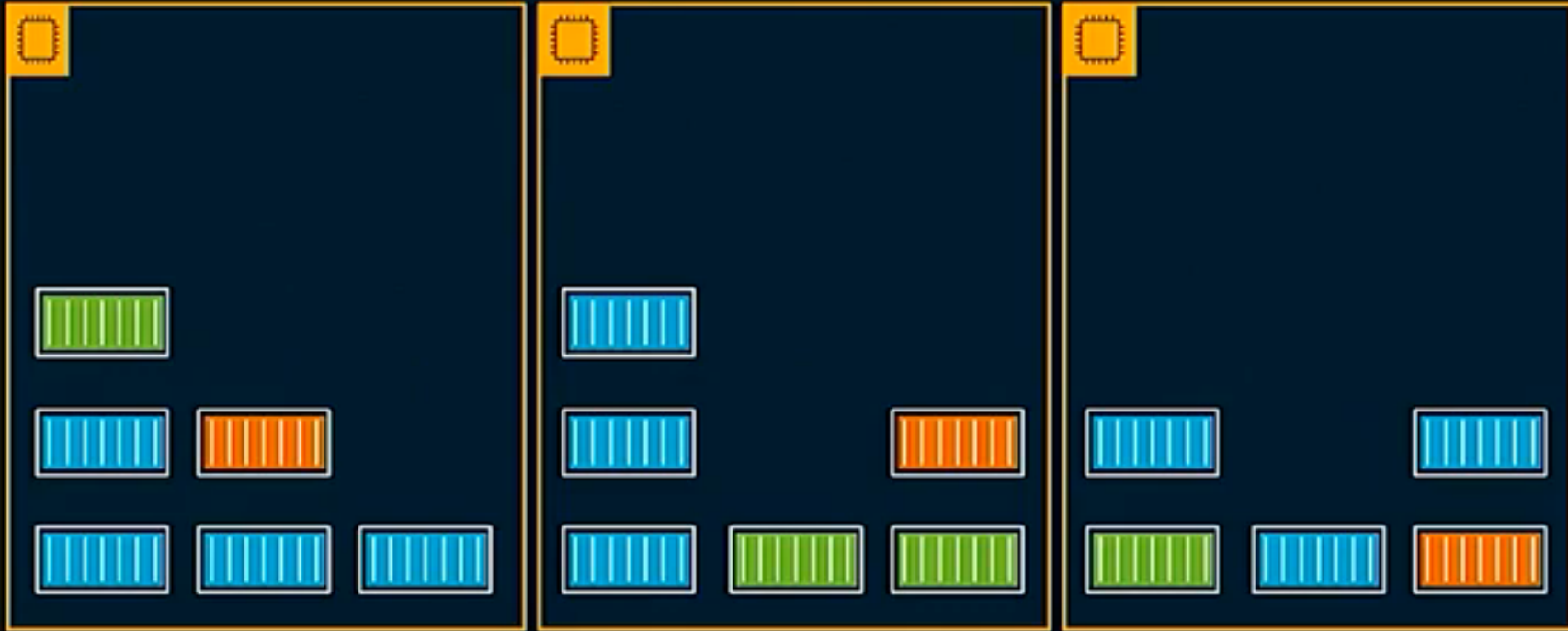
-  } Account 1
-  } Account 2
-  } Account 3









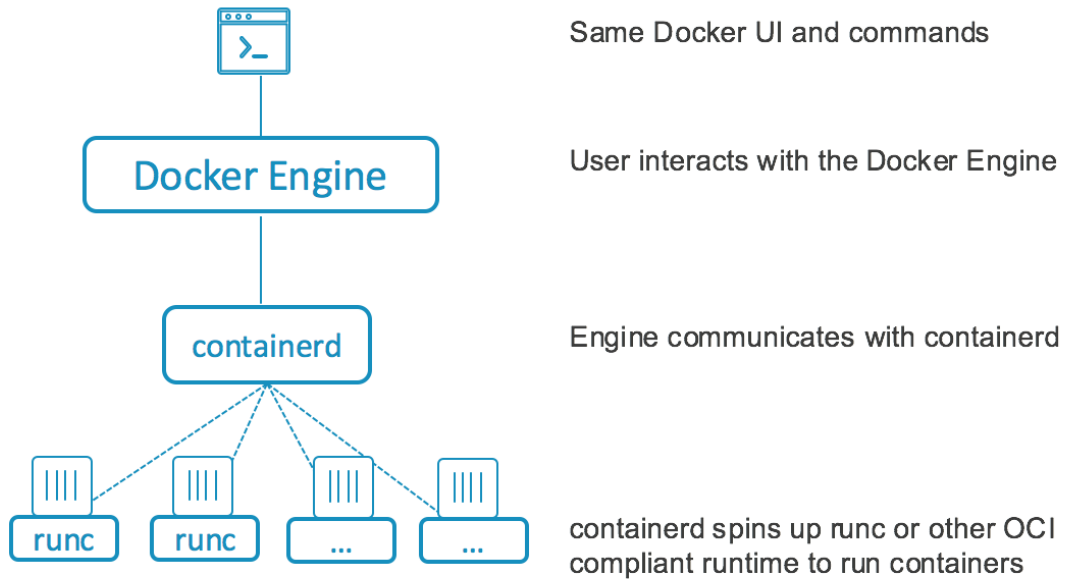


Firecracker & containerd

Firecracker & containerd

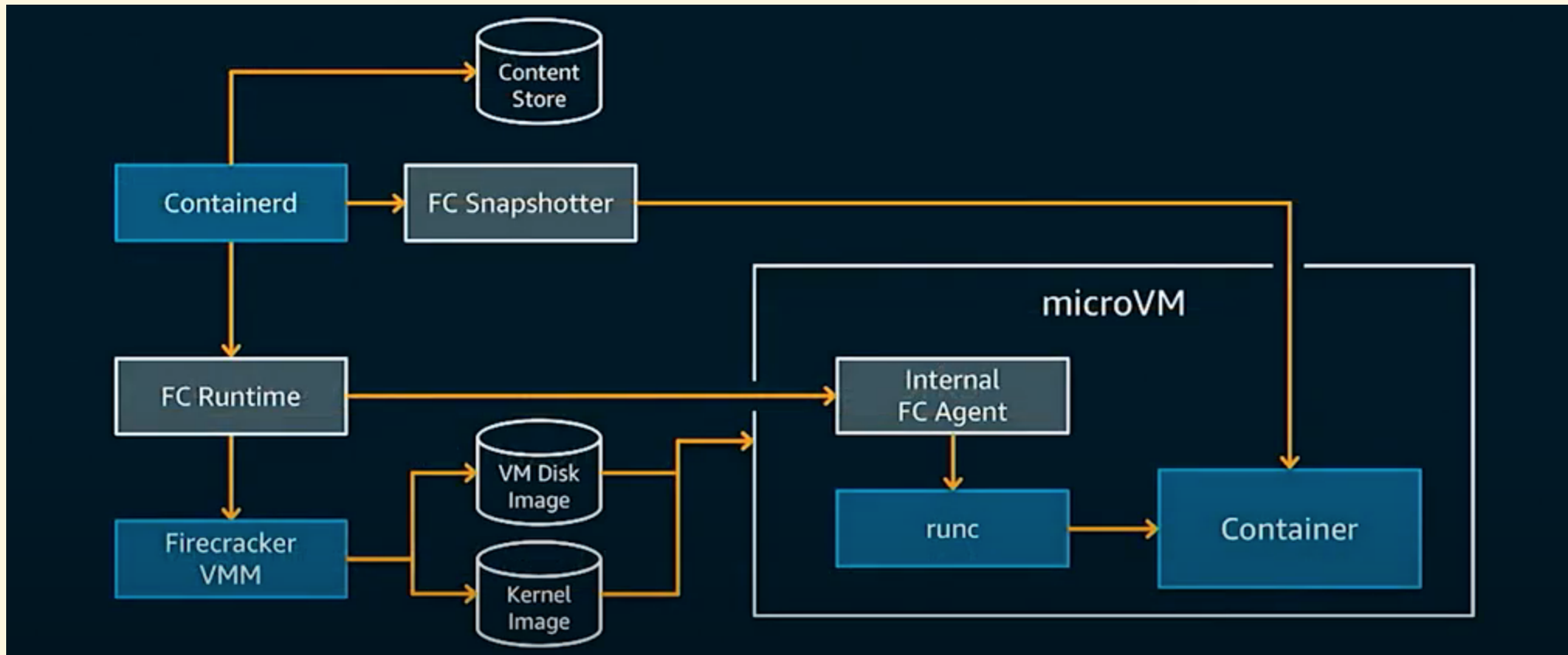
- containerd to manage containers as Firecracker microVMs.
- Multi-tenant hosts
- OCI image format
- Work with popular orchestration frameworks
 - Kubernetes and Amazon ECS
- Define a future: light as container, secure as VM

OCI Image & OCI Runtime



- container **d**
- runc
 - is a CLI tool for spawning and running containers according to the OCI specification.

Firecracker & containerd Architecture



Live Demo

Live Demo #1

Getting Started with Firecracker in 2 Minutes

Getting started with Firecracker

- Firecracker on AWS bare metal
- Firecracker on other clouds with bare metal (e.g., Packet)
- Firecracker on GCP nested-virt
- Firecracker on Azure nested-virt
- Firecracker on your dev machine (physical/nested-virt)

Getting started with Firecracker

- **Firecracker on AWS bare metal**
- Firecracker on other clouds with bare metal (e.g., Packet)
- Firecracker on GCP nested-virt
- Firecracker on Azure nested-virt
- **Firecracker on your dev machine (physical/nested-virt)**

VM

Firecracker

VM: Ubuntu

Hypervisor: VirtualBox

Host OS: macOS

Hardware: Macbook Pro

Demo #1

Live Demo #1

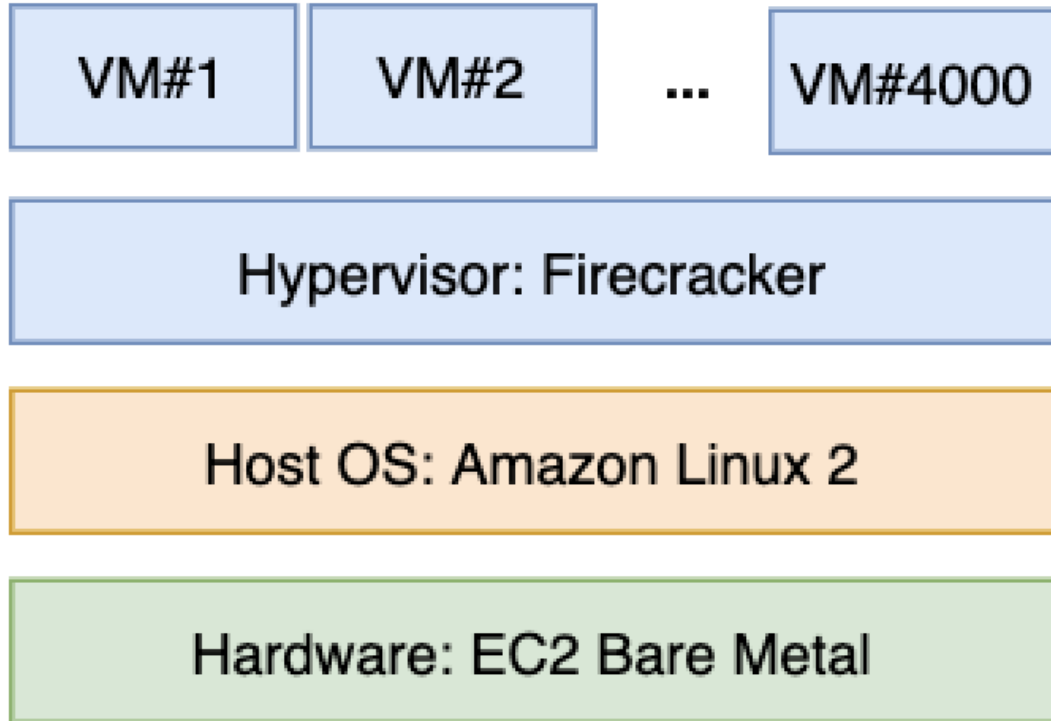
Getting Started with
Firecracker in 2 Minutes:

Firecracker on VirtualBox on
macOS on Macbook Pro

<https://github.com/dwchiang/firecracker-workshops/tree/master/01-getting-started>

Live Demo #2

Creating 4,000 microVMs in 90 Seconds



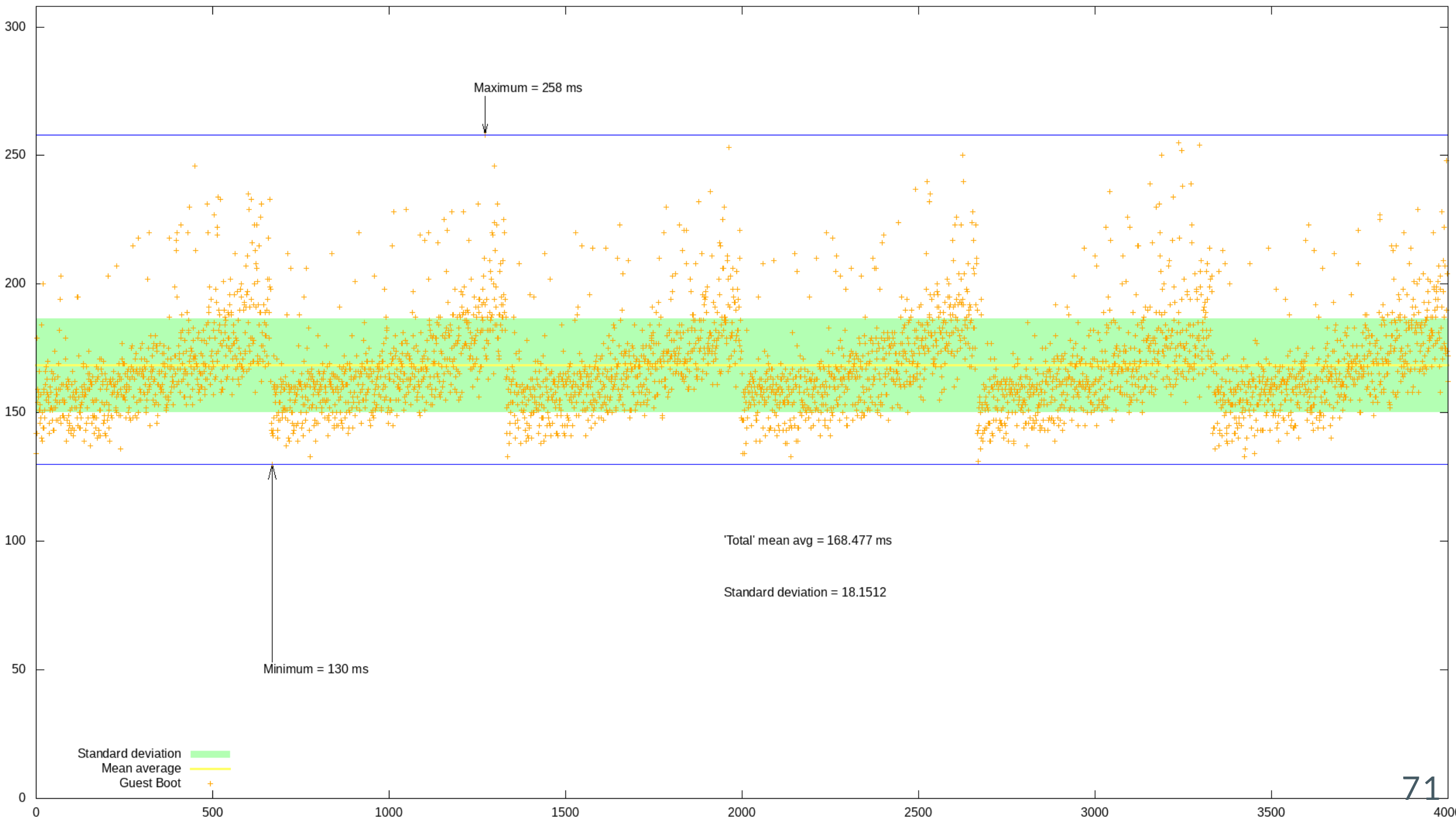
Demo #2

Live Demo #2

Creating 4,000 microVMs in
90 Seconds:

Firecracker on EC2 Bare
Metal instance

<https://github.com/dwchiang/firecracker-workshops/tree/master/02-4000-microVMs>



Type Name	vCPU	ECU	Memory	Instance Storage	Cost per hour
i3.metal	64	208	512 GiB	8 x 1900 NVMe SSD	\$4.992
m5.metal	96	345	384 GiB	EBS Only	\$4.608
m5d.metal	96	345	384 GiB	4 x 900 NVMe SSD	\$5.424
c5.metal	96	375	192 GiB	EBS Only	\$4.08
c5d.metal	96	375	192 GiB	4 x 900 NVMe SSD	\$4.608

Savings on Spot Instance

Savings

Last hour ▼

A high-level summary of your savings across all of your running and recently terminated Spot Instances. For detailed reporting on your account-level Spot usage, visit [Cost Explorer](#)

Spot usage and savings

1	72	512	\$4.99	\$1.50	70%
Spot Instances	vCPU-hours	Mem(GiB)-hours	On-Demand total	Spot total	Savings
				\$0.0208	\$0.0029
				Average cost per VCPU-hour	Average cost per mem(GiB)-hour

Details

i3.metal (1)	72 vCPU hours	512 mem(GiB)-hours	\$1.50 total	70% savings
--------------	---------------	--------------------	--------------	-------------

* Spot savings are estimated savings and may differ from actual savings. This is because the savings shown on this page do not include the billing adjustments for your usage.

73

Firecracker & Open Source Projects

Firecracker Integration with Open Source Projects

- Kata Containers
- UniK
- OSv
- Weave Ignite

Weave Ignite

- Open source VMM with a container UX
- Combines Firecracker microVMs with OCI images
- Works using **GitOps**
 - `ignite gitops <repo>`

Who would use Firecracker?

- Teams building compute services
- Teams integrating Firecracker with container stacks
- Developers & security engineers who want to contribute

Takeaways

安全隔離好

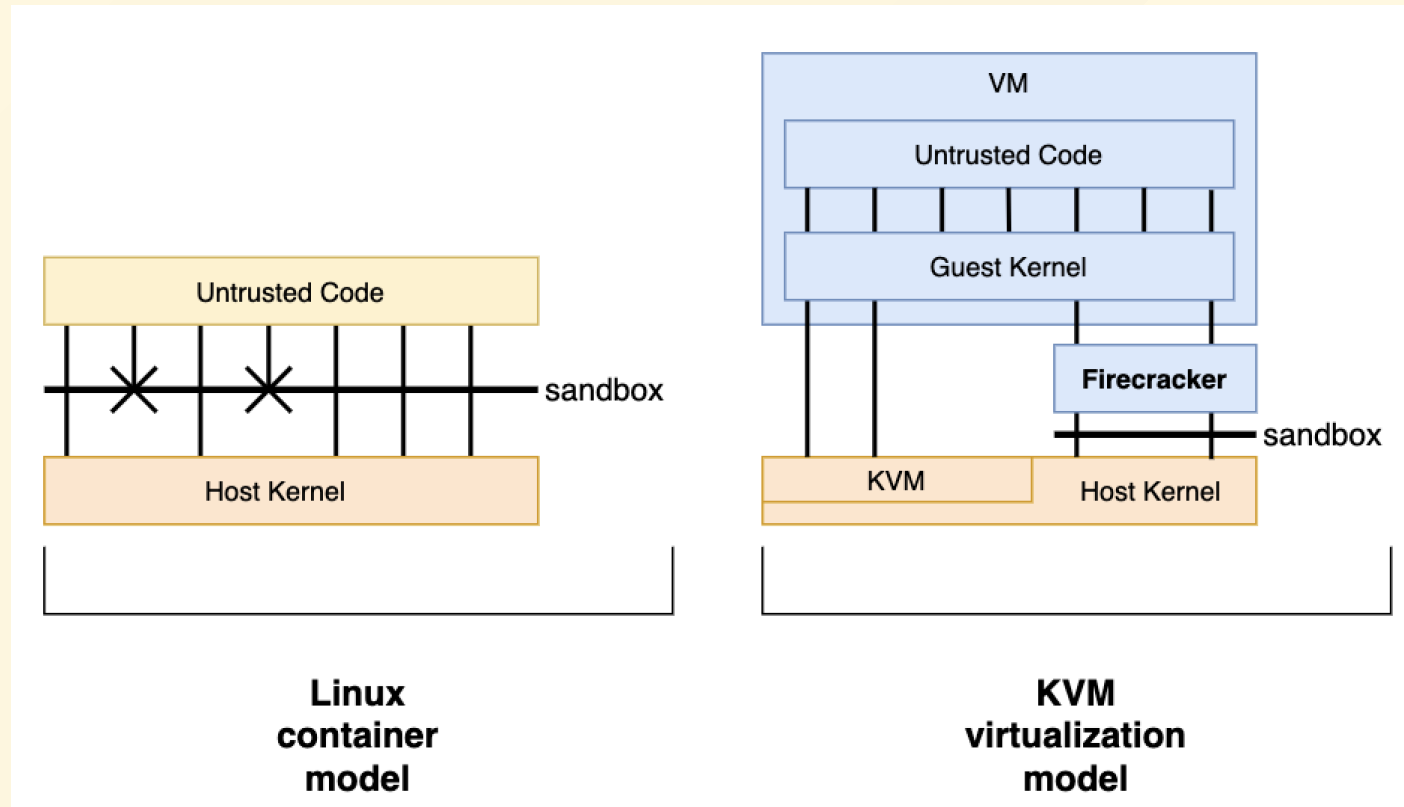
啟動時間短

產能效率高

#像極了愛情

-- AWS Firecracker VMM

Firecracker Security Model



Q&A & Thank you

Blog <https://www.ernestchiang.com>

Twitter [@dwchiang](https://twitter.com/dwchiang)

#CrossFieldIntegration

#TechnicalManagement

#Bluetooth #AWS

<https://bit.ly/awsvmm2020>

抽獎活動

&

\$25 AWS Credits

Community

Community

- Cloud Native Taiwan User Group
 - Facebook : <https://www.facebook.com/groups/cloudnative.tw>
- AWS User Group Taiwan
 - Facebook : <https://www.facebook.com/groups/awsugtw>
- Taiwan CDK Meetup
 - Facebook : <https://www.facebook.com/groups/cdkmeetuptw>

Reference

Reference: Firecracker

- Project Homepage : <https://firecracker-microvm.github.io/>
- Project GitHub : <https://github.com/firecracker-microvm/firecracker>
- Project Roadmap : <https://github.com/firecracker-microvm/firecracker/projects/13>

Reference: Firecracker

- Youtube : [Firecracker: A Secure and Fast microVM for Serverless Computing](#), 2019-0717, by Meena Gowdar (@meejamb) & Arun Gupta (@arungupta)
- Youtube : [NSDI '20 - Firecracker: Lightweight Virtualization for Serverless Applications](#), 2020-02, by Marc Brooker at [NSDI 20](#)
 - Paper (PDF) : [Firecracker: Lightweight Virtualization for Serverless Applications](#)

Reference: Firecracker

- **Blog**: [深度解析 AWS Firecracker 原理篇 - 虚拟化与容器运行时技术](#) by 莫梓元.
- **Blog**: [深度解析 AWS Firecracker 实战篇 - 一起动手点炮竹](#) by 莫梓元.
- **Workshop**: [IGNITE YOUR FIRECRACKER WORKSHOP - AWS TKO 2020](#)
- **Workshop**: [Firecracker Workshop Collections](#)
- **Slide**: [Deep Dive into Firecracker Using Lightweight Virtual Machines to Enhance the Container Security Boundary - AWS Summit Sydney, 2019](#)

Reference: Firecracker

- **Demo** : [A demo running 4000 Firecracker microVMs](#)
- **Docs** : [Firecracker Design](#) (firecracker-microvm/firecracker)
- **Docs** : [Getting started](#) (firecracker-microvm/firecracker)
- **Youtube** : [Running AWS Firecracker in your local machine](#), by Abhijith PK, 2018.

Reference: ecosystems

- **Weave Ignite** is an open source Virtual Machine (VM) manager with a container UX and built-in GitOps management.
 - <https://github.com/weaveworks/ignite>
- **OSv** is an open-source versatile modular unikernel designed to run single unmodified Linux application securely as microVM on top of a hypervisor, when compared to traditional operating systems which were designed for a vast range of physical machines.
 - <https://github.com/cloudius-systems/osv>

Reference: ecosystems

- **Kata Containers** is an open source project and community working to build a standard implementation of lightweight Virtual Machines (VMs) that feel and perform like containers, but provide the workload isolation and security advantages of VMs.
 - <https://github.com/kata-containers/kata-containers>

Reference: ecosystems

- [crosvm](#)
- [rust-vmm](#)
- ...
- [Cloud Hypervisor](#)

Reference: Virtualization

- Youtube : [Linux 核心設計 發展動態回顧 \(2020-05-23\)](#) by jserv
- Slide : [Embedded Virtualization applied in Mobile Devices](#) by jserv, 2012.

Open Source at AWS

- <https://aws.amazon.com/opensource/>

Firecracker design principles

- Multitenant
- Any vCPU and memory combination
- Oversubscription permissible
- Steady mutation rate: 100+ microVMs/host/sec
- Limited only by hardware resources
- Host-facing REST API
- Minimalist guest device model

Slido Poll Results

2020-0801

How long have you been in the industry?

051

1 to 2 years



3 to 5 years



Less than 1 year



5 to 10 years



More than 10 years



Have you heard of...

061

AWS (Amazon Web Services)



Amazon EC2



Container



VM (Virtual Machine)



Virtualization



Linux Kernel



Have you heard of...

061

AWS Lambda



Hypervisor



KVM



AWS Fargate



VMM



Firecracker

