
AWS
re:Invent

CON216

AWS Fargate: Are serverless containers right for you?

Saloni Sonpal
Senior Product Manager
Amazon Web Services

Roy Penn
VP of Engineering
Amenity Analytics

AWS re:Invent

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

Agenda

Introduction to containers on AWS and AWS Fargate

Benefits: serverless, secure, and savings

Capabilities: logging, monitoring, and more

Workloads that are a good fit

What’s new and what’s next?



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

AWS container services landscape

Management

Deployment, scheduling, scaling, and management of containerized applications



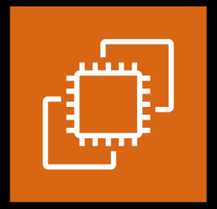
Amazon Elastic Container Service (Amazon ECS)



Amazon Elastic Kubernetes Service (Amazon EKS)

Hosting

Where the containers run

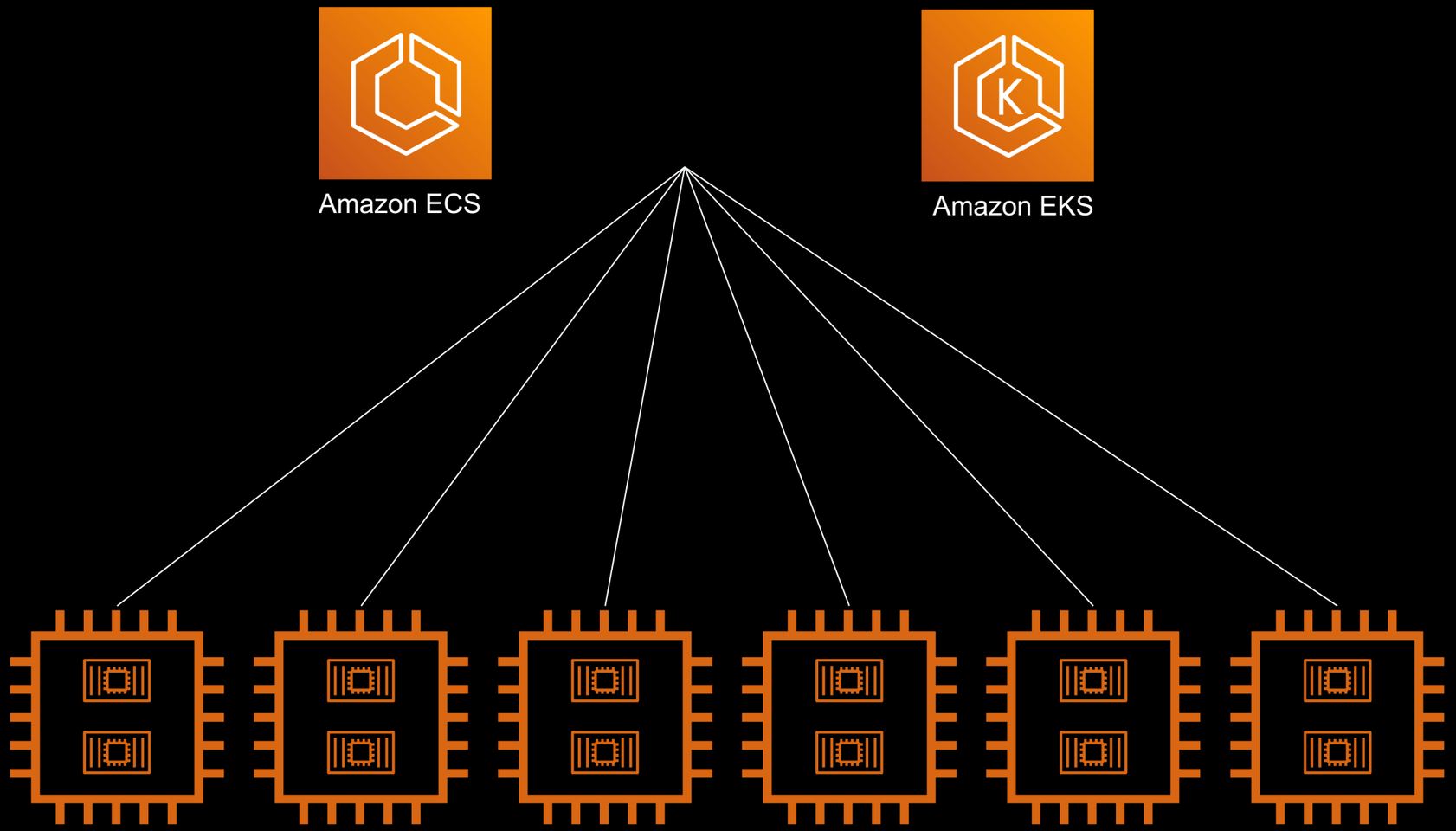


Amazon Elastic Compute Cloud (Amazon EC2)



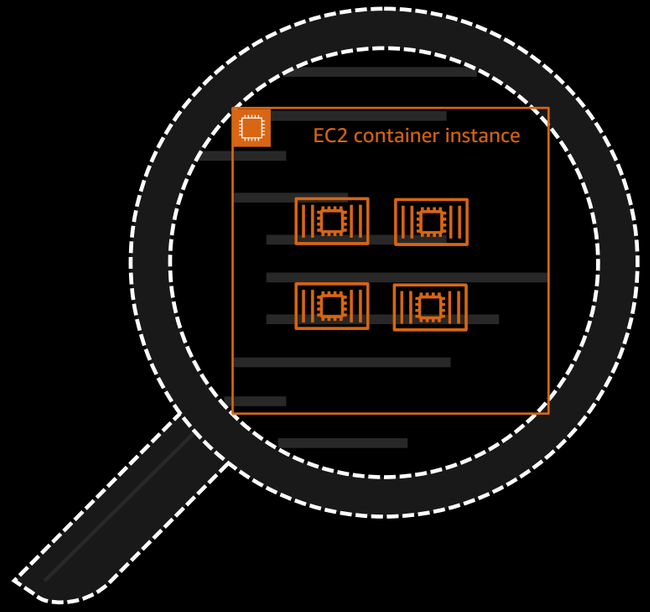
AWS Fargate

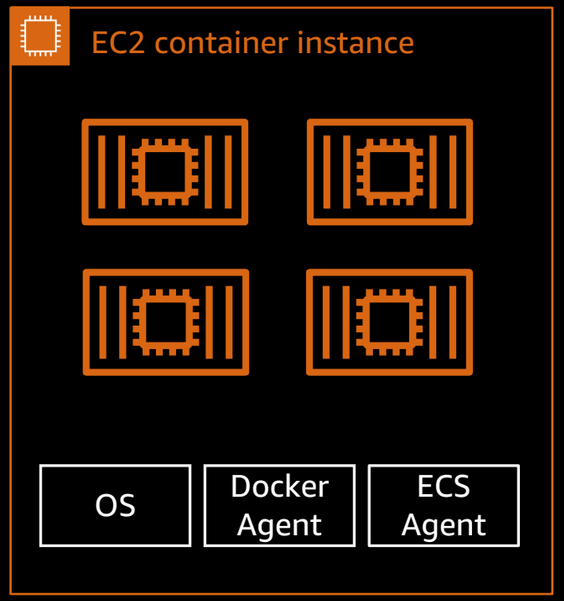
Fully managed container orchestration and scheduling



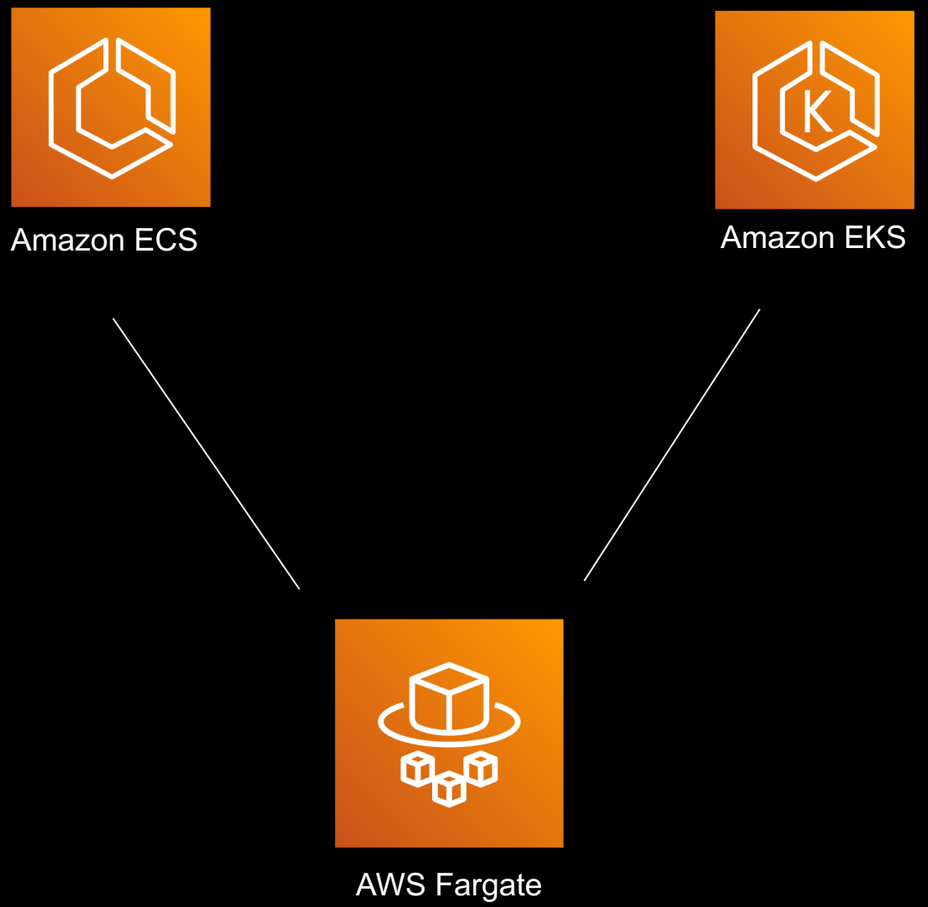
* For simplicity, we will use Amazon ECS terminology. Unless stated otherwise, the content applies to both EKS and ECS

But you still end up managing more than just containers





Amazon ECS and Amazon EKS on Fargate

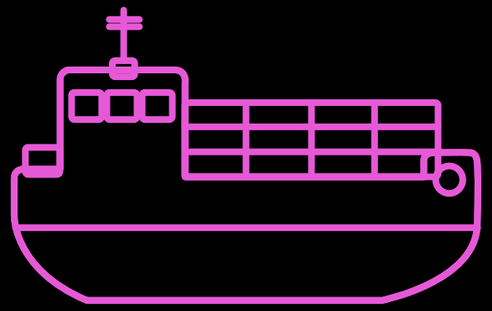


Fargate at a glance



3x

growth for Fargate in one year



Hundreds of millions

of tasks / pods started each week



First

serverless compute option that has a committed use discount plan

Benefits



Serverless

Managed by AWS; no AWS EC2 instances to provision, scale, or manage – simple operations model

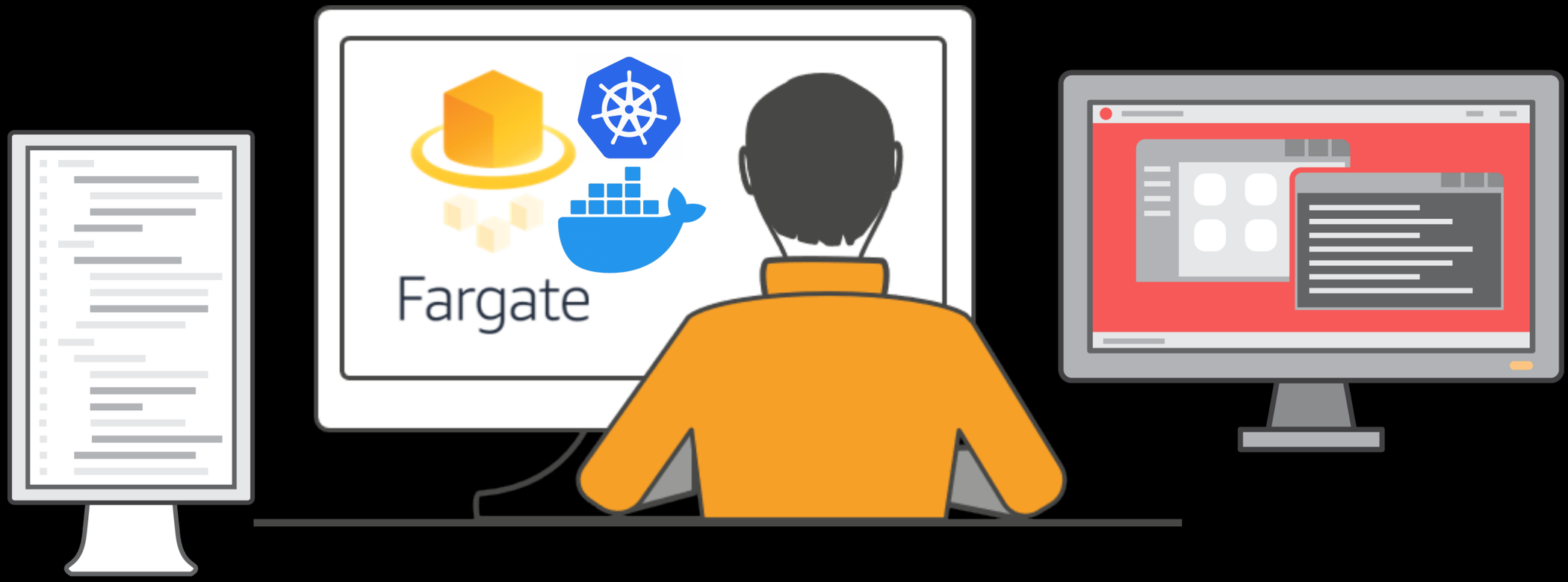
Secure by default

Isolated, patched, and compliant for running the most sensitive workloads

Savings

Elastic on-demand pricing model with granular resource allocations giving customers the ability to procure only the vCPU and memory needed when they need it without any excess

Focus on applications, not infrastructure



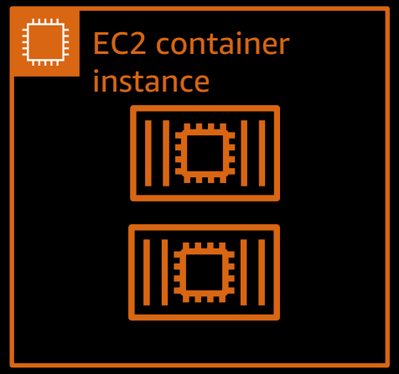
Serverless



AWS re:Invent

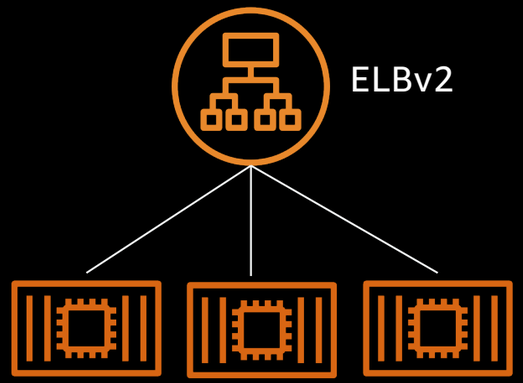
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Constructs



Task

- A running instantiation of a task definition
- Use FARGATE launch type
- Task contains one or more containers



Service

- Maintain and run copies
- Integrated with ALB or NLB
- Unhealthy tasks automatically replaced

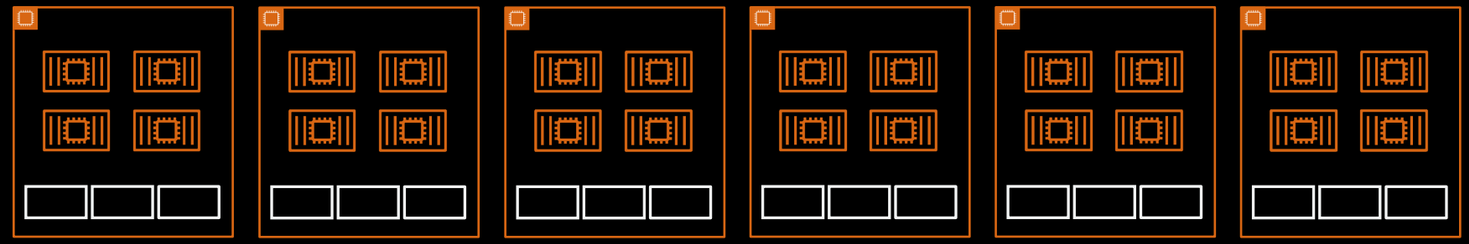


Cluster

- Infrastructure Isolation boundary
- IAM Permissions boundary

Fargate does operational heavy lifting for you

- Patching and upgrading OS, agents, etc.
- Scaling the instance fleet for optimal utilization





“By removing the complexity of infrastructure management and shifting some responsibilities, like OS hardening and patching, onto AWS [Fargate], we can reduce the amount of resources we spend on these tasks and instead focus on adding value to our Vodafone customers.”

Maria Farrugia

Head of Digital Architecture,
Vodafone UK

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

Secure by default



AWS re:Invent

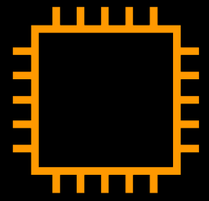
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Sharing is not caring

Each task or pod running on Fargate has its own isolation boundary and does not share the underlying kernel, CPU resources, memory resources, or elastic network interface with another task or pod.

https://docs.aws.amazon.com/AmazonECS/latest/developerguide/AWS_Fargate.html
https://docs.aws.amazon.com/eks/latest/userguide/fargate.html

AWS Fargate security model



Hardware virtualization

Tasks get isolated compute



Network isolation

Tasks get dedicated ENIs



Storage isolation

Tasks get dedicated storage



Credentials isolation

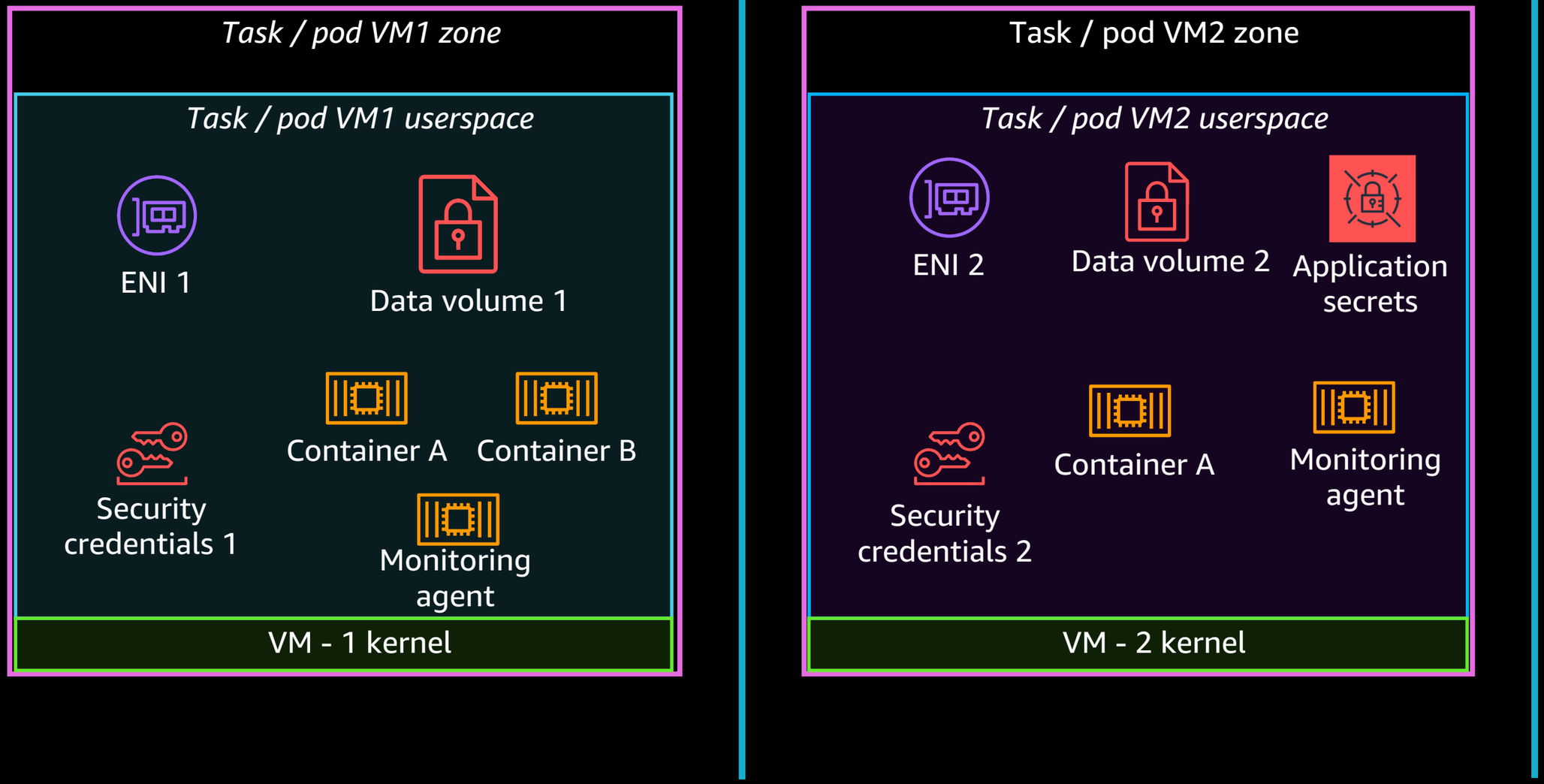
Tasks get credentials scoped to applications



Hands-off patching

Infrastructure is always patched

Strong trust boundaries by default



Savings and cost optimizations



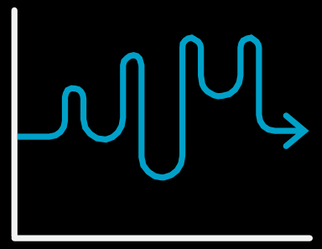
AWS re:Invent

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

Fargate purchase options

Fargate

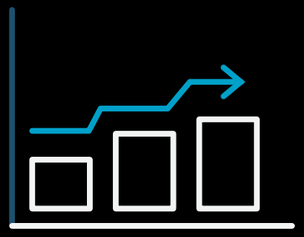
Pay for containers **per second** with no long-term commitment



Capacity needs can change rapidly

Compute Savings Plans

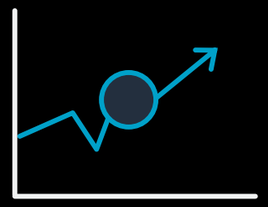
Make a 1- or 3-year commitment and receive a **significant discount**



Baseline compute needs known in advance

Fargate Spot

Spare capacity with **savings up to 70%** off Fargate standard pricing (Amazon ECS)



Fault-tolerant, flexible workloads

Pricing

Pay for what you provision

Billed for Task / Pod level CPU and Memory

Per second billing; 1 minute minimum

Task / Pod configurations

CPU	Memory
256 (.25 vCPU)	512MB, 1GB, 2GB
512 (.5 vCPU)	1GB, 2GB, 3GB, 4GB
1024 (1 vCPU)	2GB, 3GB, 4GB, 5GB, 6GB, 7GB, 8GB
2048 (2 vCPU)	Between 4GB and 16GB in 1GB increments
4096 (4 vCPU)	Between 8GB and 30GB in 1GB increments

50 different CPU / Memory configurations to choose from

What are Savings Plans?

- Commit to $ / hr usage level of compute power for 1 or 3 years
- All usage up to commitment level covered by Savings Plans
- Compute Savings Plans: apply to all Amazon EC2 instances and cover Amazon EMR, Amazon ECS, Amazon EKS, Fargate, and AWS Lambda
- EC2 Instance Savings Plans: apply to a specific instance family within a Region

Fargate Savings Plans

Term length	Payment option	CPU discount	Memory discount
1 Year	No upfront	15-20%	15-20%
1 Year	Partial upfront	20-25%	20-25%
1 Year	All upfront	22-27%	22-27%
3 Years	No upfront	40-45%	40-45%
3 Years	Partial upfront	45-50%	45-50%
3 Years	All upfront	47-52%	47-52%

https://aws.amazon.com/savingsplans/pricing/

Fargate Spot

WITH AMAZON ECS ON AWS FARGATE

- *Unused* Capacity
- Price up to 70% less than standard Fargate
- Interruptible and can be reclaimed
  - Two-minute warning
  - Default stopTimeout 30s (can be increased to 120s)
- Automatic diversification

Logging and monitoring



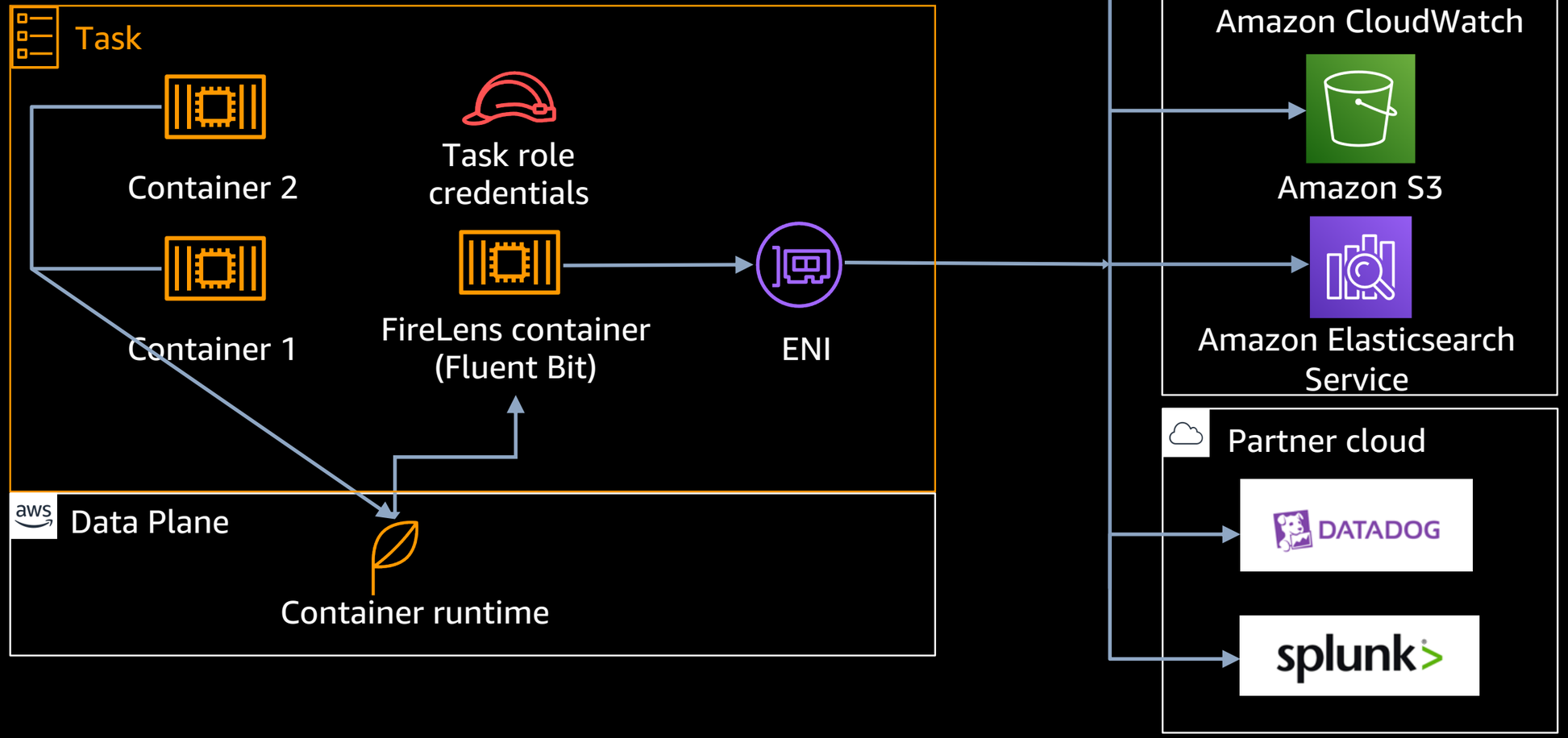
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Logging

FIRELENS



Fluent bit

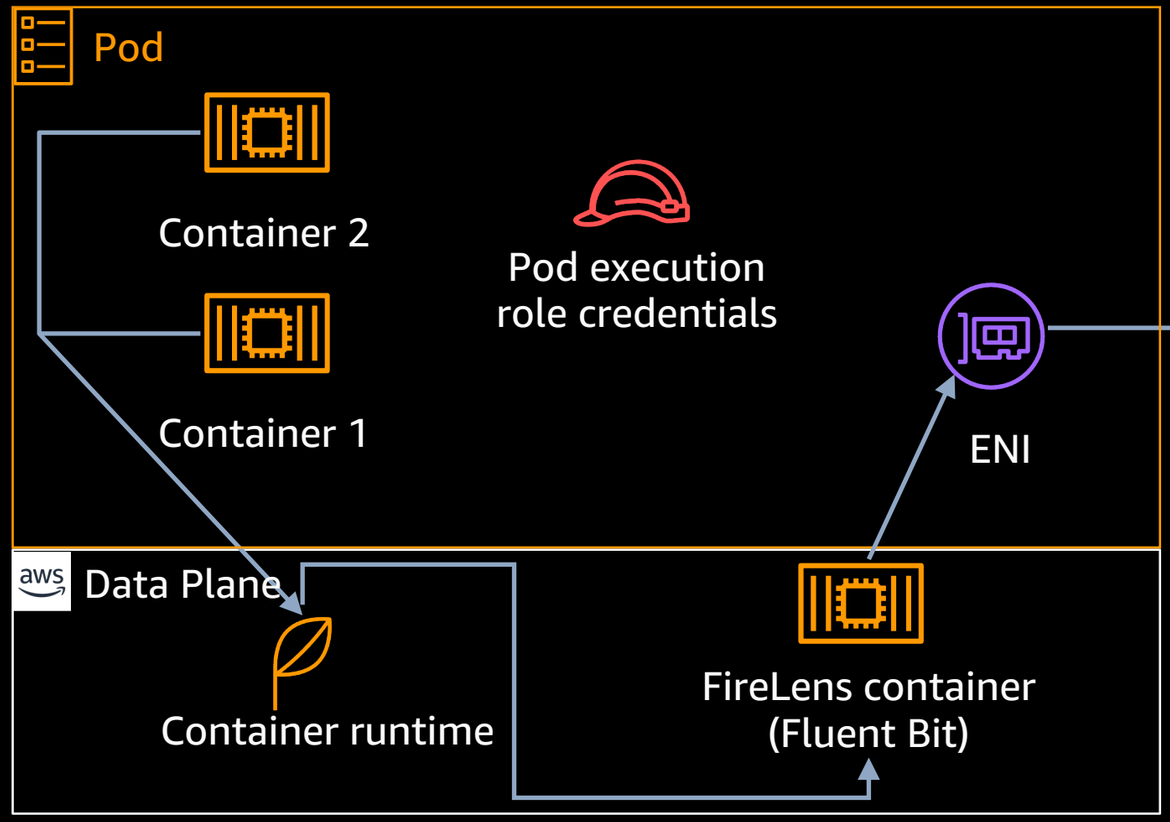


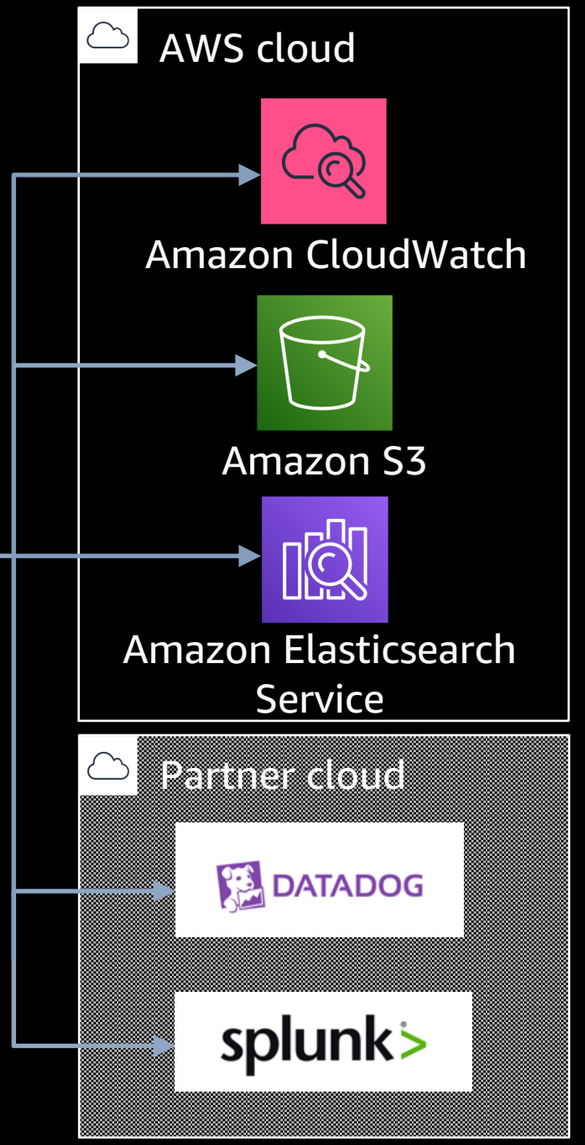
Logging for Amazon EKS on Fargate (New!)

FIRELENS



Fluent bit





Monitoring

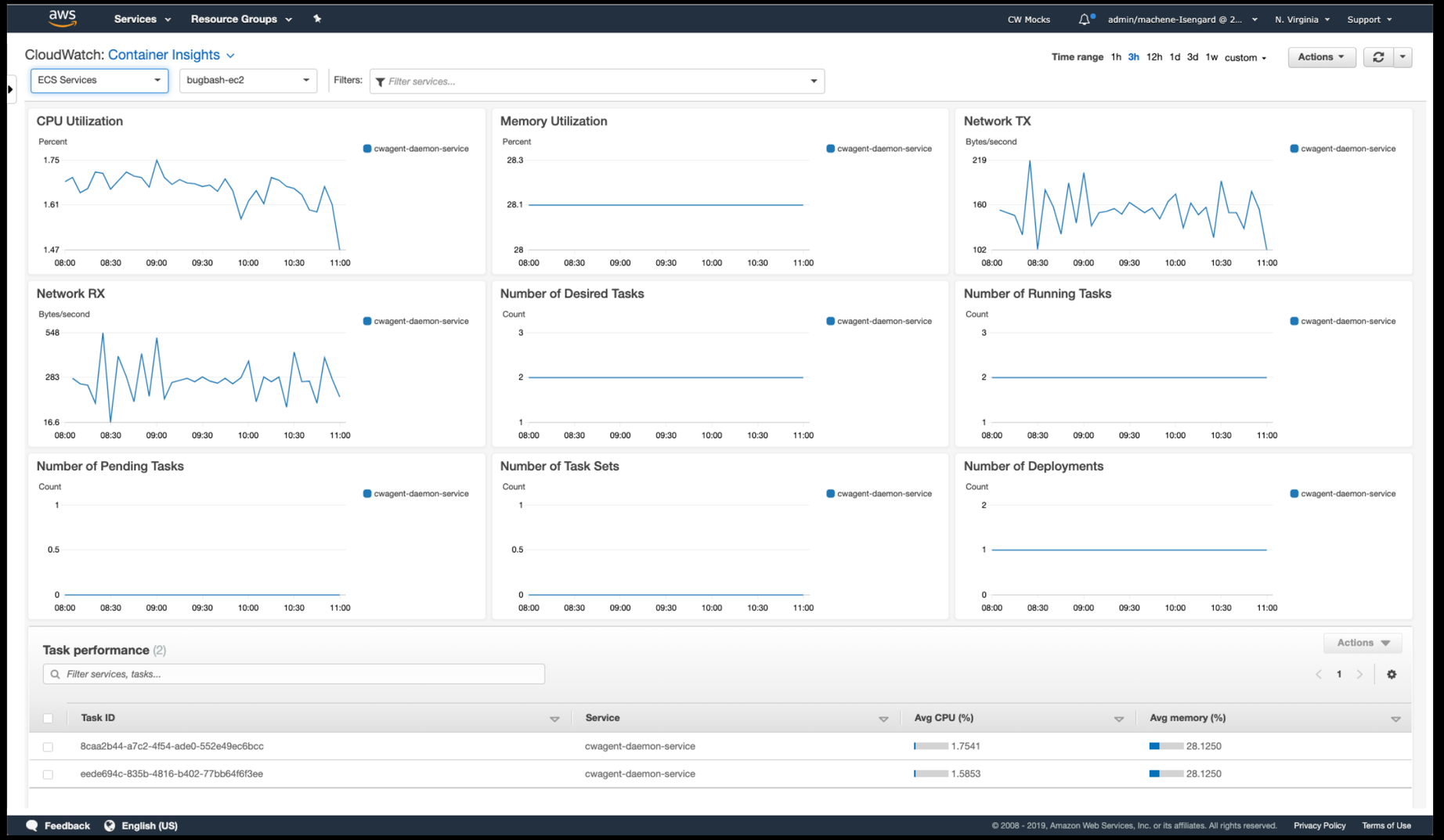
CLOUDWATCH CONTAINER INSIGHTS

A fully managed observability service for monitoring, troubleshooting, and alarming on your containerized applications and microservices

- ✓ Collects, aggregates, and summarizes
- ✓ Reliable, secure metrics and logs collection
- ✓ Automated summaries and analysis
- ✓ Observability experience across metrics, logs, traces
- ✓ Ad hoc analytics

Monitoring

WITH AMAZON ECS ON AWS FARGATE



Partners – extending Fargate's capabilities

Foundation	Monitoring & Logging	DevOps & CI/CD	Security & Networking
docker	DATADOG	GitLab	ALERT LOGIC
HashiCorp	dynatrace	GitHub	aqua
spot	moogsoft	pulumi	paloalto NETWORKS
	New Relic	XebiaLabs Enterprise DevOps	portshift
	splunk>	weaveworks	sysdig
	solarwinds		TREND MICRO
	sumo logic		

Workloads that are a good fit



AWS re:Invent

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

Web services



Vanguard®

Vanguard runs their Web services and APIs on Fargate and are going all in with Fargate for security sensitive workloads.

“We chose ECS Fargate for a variety of reasons: It’s a fully distributed architecture, it enable DevSecOps patterns, was backwards compatible with our legacy container orchestration platform, is fully automated and automatable, secure by default, is cost effective and enable NoOps patterns.”

Yoni Ryabinski

Chief Enterprise Architect

Vanguard

API servers



Ancestry runs their TREES workload with 100 million family trees on Fargate. They have standardized on GitLab for their CI / CD platform.

“We have a real time and on-demand system. We have thousands of EC2 instances running, we have Fargate instances running, sending logs into Cloudwatch, we pull those logs and we push them into our Kinesis stream for processing.”

Jim Bradshaw

Sr. Manager Platform Productivity & Tooling

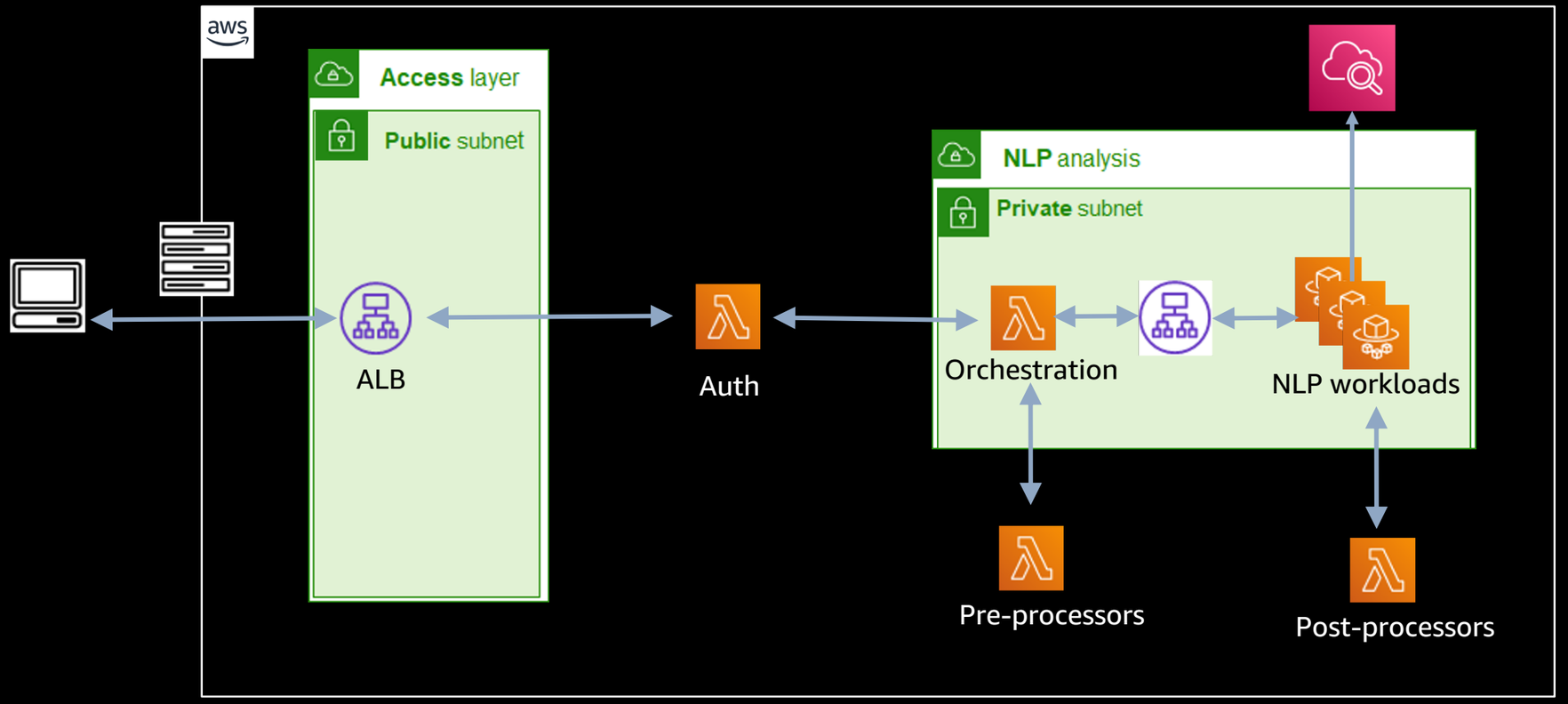
Ancestry

Data processing and inference

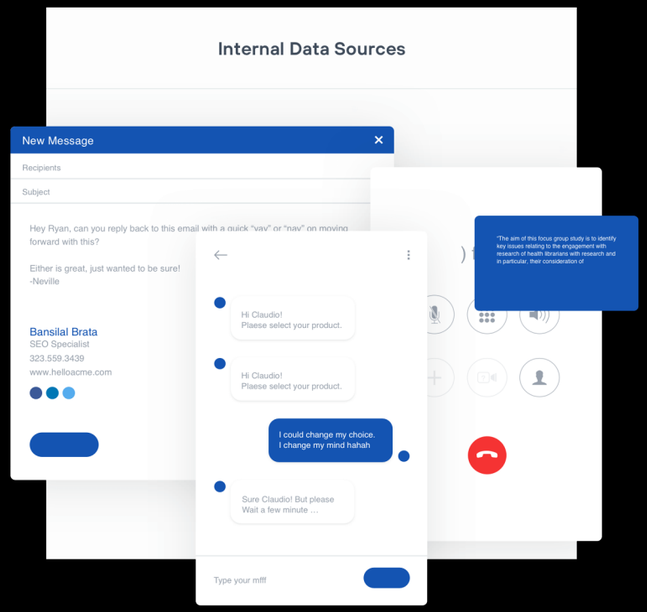


Amenity runs a significant portion of its NLP (Natural Language Processing) heavy lifting on Fargate Spot clusters. With multiple clusters running different NLP models variants, Amenity can process hundreds of millions of documents within a matter of hours with almost no ops attention required. They spin up thousands of ECS tasks in an almost unpredictable manner and manage to save money at the same time.

Architecture workflow



How Fargate benefited us?



Reduce operational load improve ops excellence

Easier to spin up huge clusters and maintain them while also maintaining ultra high availability uptime

Optimize cost

Easier to realize spot cost savings and to pay for exactly what we need and use

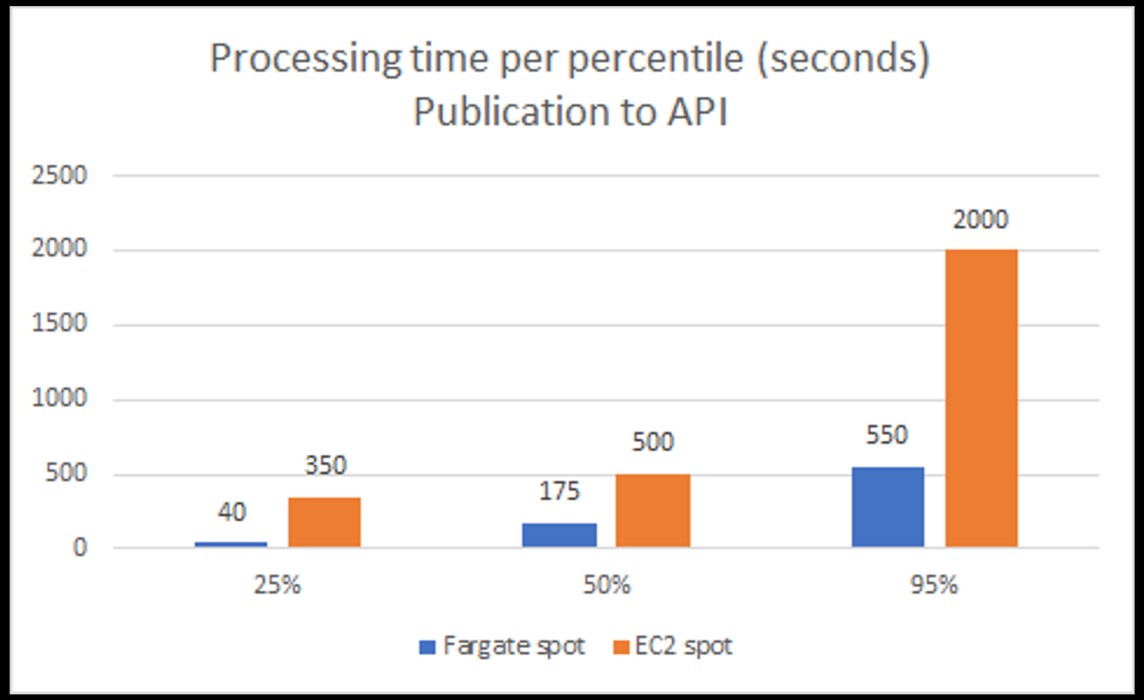
Time to innovation

We can now focus more of our efforts on generating insights and creating value for our customers

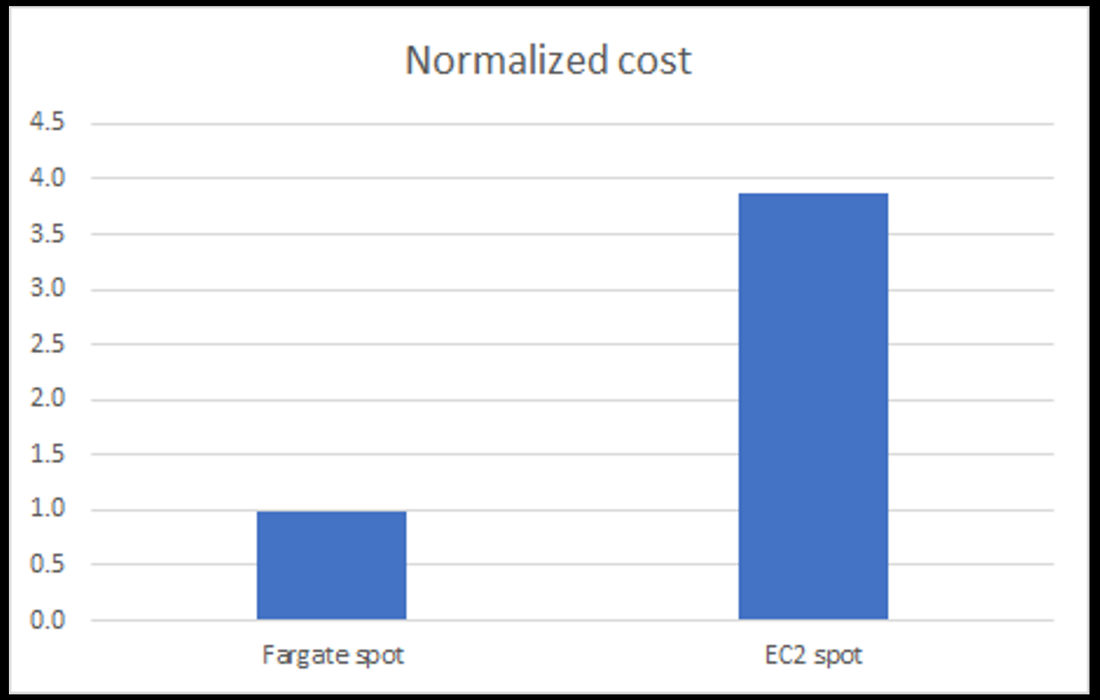
Business impact from using Fargate

We process hundreds of millions of documents on Fargate and have moved > 90% of our workloads to Fargate in the past year

Real time NLP workloads finish faster on Fargate Spot
Historic workloads finish in hours and not days



Workloads that are built around serverless Fargate Spot clusters are ~75% cheaper overall for Amenity



What’s new and what’s next?



AWS re:Invent

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

Launches this year

- Support for persistent task storage with Amazon EFS
- Ephemeral storage increased to 20GB
- Ephemeral storage encrypted by default
- Network performance metrics surfaced by Container Insights
  - Also available from task metadata
- Support for CAP_SYS_PTRACE
- Containerd replaced docker as the container runtime
- Usage metrics against service quotas (a.k.a. limits)
- Increased default concurrent task/pod service quotas
- AWS Batch Jobs for Fargate (ECS) **(New!)**

Amazon EKS specifically

- Included in Compute Savings Plans
- Available in additional commercial regions
- NLB support with IP targets
- Logging support via FireLens **(New!)**

Not supported today

- Privileged containers
- Add / drop Linux capabilities other than CAP_SYS_PTRACE
- Amazon EBS support for persistent storage
- Windows OS
- ARM architecture
- GPUs
- For EKS, DaemonSets (no worker nodes after all)

Looking ahead

- Expand Fargate capabilities and enable new workloads
- Improve performance and scaling
- Open-source engagements (Firecracker, containerd, fluentbit)
- Partner integrations (security and observability)
- Delight customers across multiple orchestrators (ECS, EKS, etc.)

Related sessions

CON218 – AWS Fargate: Roadmap and Vision

CON203 – Choosing your container data plane on AWS

CON304 – Amazon ECS on AWS Fargate deep dive with Affirm

CON303 – Amazon EKS on AWS Fargate deep dive

CON210 – Developing CI/CD pipelines with Amazon ECS & AWS Fargate

CON223 – AWS X-Ray and Amazon CloudWatch: Increasing visibility for ECS and Fargate

Thank you!

Saloni Sonpal
salonis@amazon.com

Roy Penn
roy@amenityanalytics.com

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



Please complete the session survey

AWS re:Invent

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