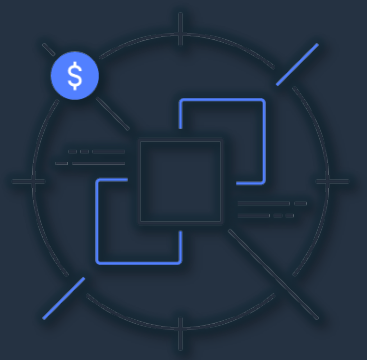


Amazon EC2 :: Optimize EKS cost with Spot & A1 Instances

Arthur Petitpierre – arthurpt@amazon.com

Specialist Solutions Architect – EC2 A1 Instances

November 2019

Who am I ?

Arthur Petitpierre – arthurpt@amazon.com / @ArthurPtP
Specialist Solutions Architect – A1 Instances @AWS

Previously:

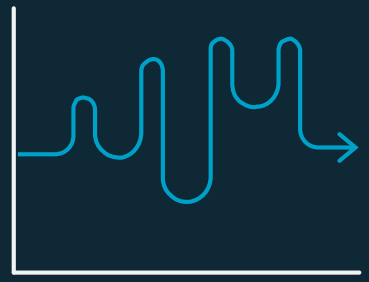
- HPC Specialist SA @AWS
- HPC Services CTO @ATOS
- And a few other stuffs…

Occasionally deliver Snowballs around ~~Paris~~ Seattle on a cargo-bike

Amazon EC2 purchase options

On-Demand

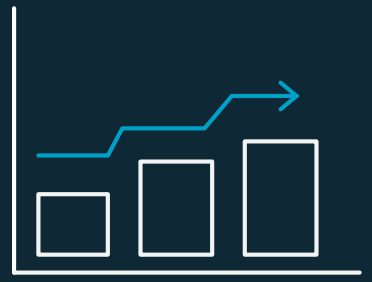
Pay for compute capacity **the second** with no long-term commitments



Spiky workloads, to define needs

Reserved Instances

Make a 1 or 3-year commitment and receive a **significant discount** off On-Demand prices



Committed & steady-state usage

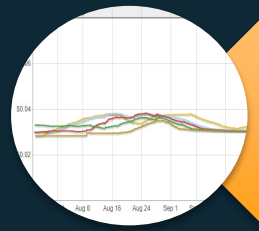
Spot Instances

Spare EC2 capacity at **savings of up to 90%** off On-Demand prices

Fault-tolerant, flexible, stateless workloads

aws

Spot is easy



Price changes infrequently based on *long term* supply and demand of spare capacity in each pool independently

Just request capacity and pay the current rate. No Bidding

Interruptions only happen when OD needs capacity. No outbidding

aws

Large customer base

Research

Caltech
GRAIL
CLEMSON UNIVERSITY
OpenAI
illumina
SevenBridges
NASA JPL
GENOME ONE
HUMAN LONGEVITY, INC.

Consumer apps

yelp
Quora
mercado libre
Aol.
Zillow
reddit
UBER
Pinterest
lyft
Expedia

AdTech & MarTech

ironSource
LKQD
codewise
AdRoll
TellApart
NetSeer
Quantcast
bloomreach
MirrorWeb
COMPLEX

Sports, media, & entertainment

COMCAST
FUSE FX
ESPN
Theory Studios
SONY
milkshake
NEXON
scrippsnetworks interactive

B2B enterprise tech

AUTODESK
HGST
red violet
LOTAME
IFTTT
Mapbox
zuora

Financial services

FINRA
CoreLogic
AIG

aws

Amazon EC2 Spot integrations

Auto Scaling

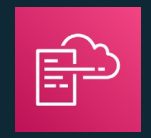
AWS Batch

Amazon EMR

AWS Data Pipeline

Amazon Elastic Container Service

Amazon Elastic Container Service for Kubernetes



AWS CloudFormation

---

cloudera

AWS Thinkbox Deadline

docker

kubernetes

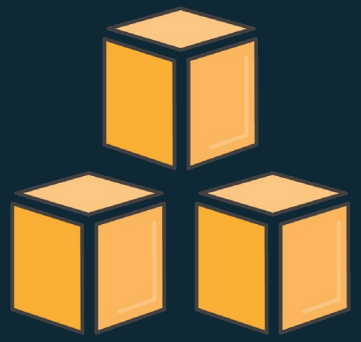
Bamboo

Jenkins

MESOS

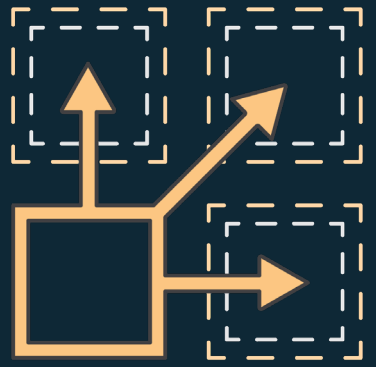
aws

Is my workload Spot Ready?

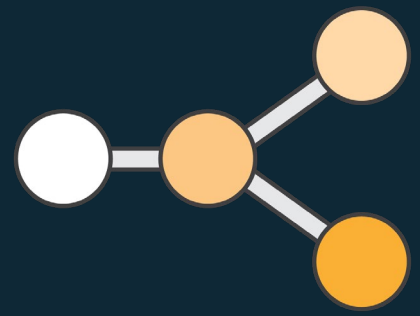


Stateless

Fault-Tolerant



Flexible: Multi-AZ and Instance Flexibility

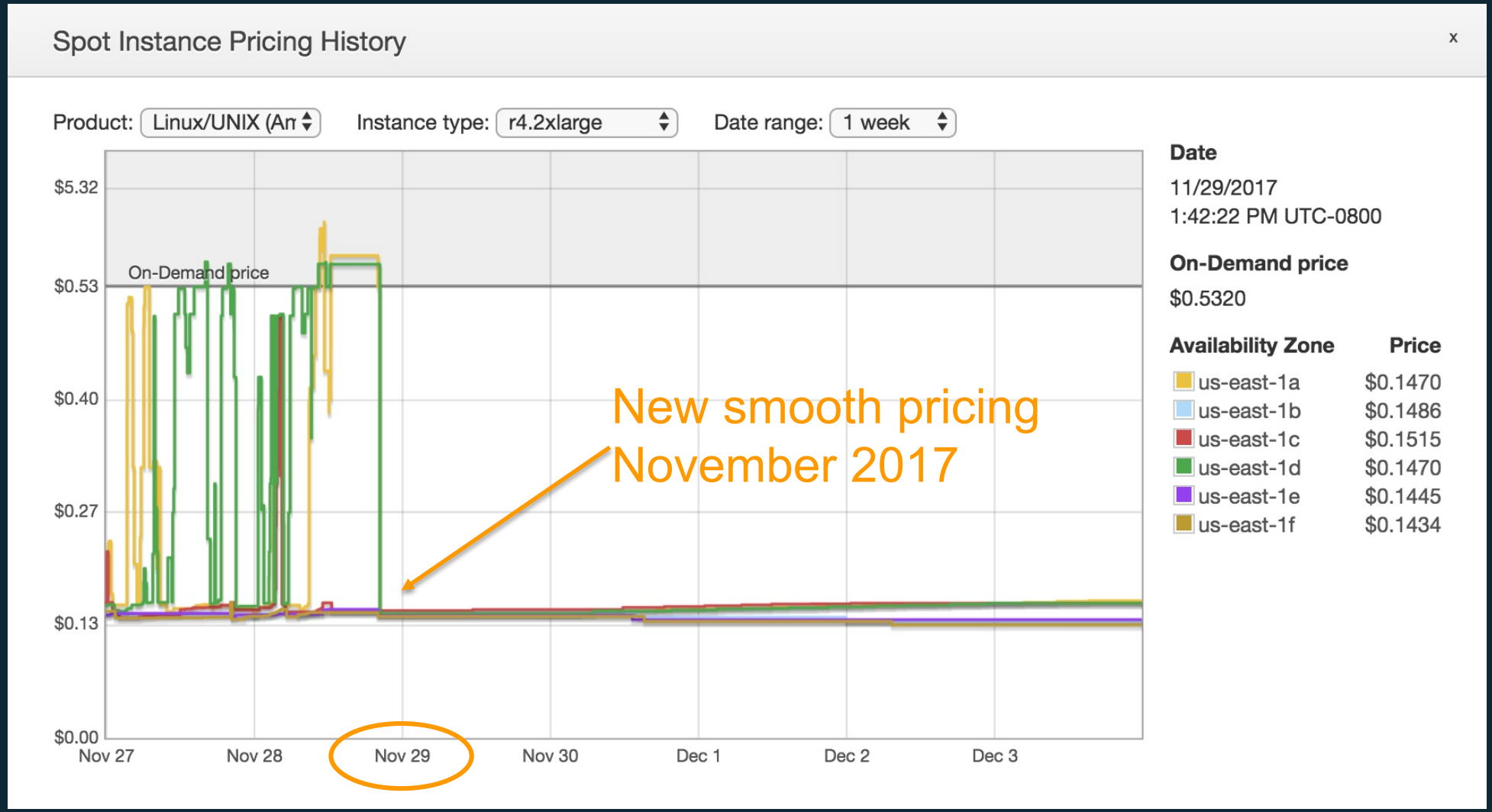


Loosely Coupled

Looks familiar?

aws

Spot pricing model change – no more bidding



EC2 Spot pools – instance flexibility

C4	1a	1b	1c	On Demand
8XL	$0.50	$0.27	$0.29	$1.76
4XL	$0.21	$0.30	$0.16	$0.88
2XL	$0.08	$0.07	$0.08	$0.44
XL	$0.04	$0.05	$0.04	$0.22
L	$0.01	$0.01	$0.04	$0.11

Each instance family

Each instance size

Each Availability Zone (69)

In every region (22)

Is a separate ***Spot pool***

Monitoring Spot usage – Savings Summary

Savings Summary

A high-level summary of your savings across all of your **running** Spot Instances.
For detailed reporting on your account-level Spot usage, visit Cost Explorer.

Spot usage and savings

47	178	487.5	~~$9.54~~	$2.66	72%
Spot Instances	vCPU-hours	Mem(GiB)-hours	On-Demand total	Spot total	Savings

Average cost per vCPU-hour: **$0.0150**
Average cost per mem(GiB)-hour: **$0.0055**

Details

c3.large (2)	88 vCPU-hours	165 mem(GiB)-hours	$1.37 total	74% savings
c4.large (6)	12 vCPU-hours	22.5 mem(GiB)-hours	$0.20 total	71% savings
c5.large (3)	6 vCPU-hours	12 mem(GiB)-hours	$0.11 total	63% savings
m5.large (6)	12 vCPU-hours	48 mem(GiB)-hours	$0.22 total	66% savings
r5.large (6)	12 vCPU-hours	96 mem(GiB)-hours	$0.23 total	73% savings
t2.medium (12)	24 vCPU-hours	48 mem(GiB)-hours	$0.18 total	70% savings
t2.large (12)	24 vCPU-hours	96 mem(GiB)-hours	$0.36 total	70% savings

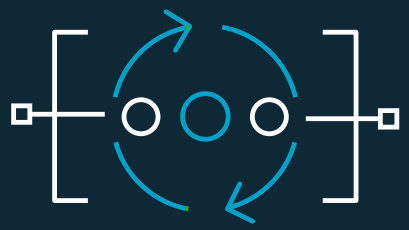
* Spot savings are estimated savings and may differ from actual savings. This is because the savings shown on this page do not include the billing adjustments for your usage.

What about interruptions?

Minimal interruptions

Over **95%** of the instances were not interrupted in the last 3 months

The work you are doing to make your applications fault-tolerant also benefits Spot



Spot is optimized for stateless, fault-tolerant, or flexible workloads.

Any application that can can have part or all of the work, paused and resumed or restarted, can use Spot.

Check for 2-minute instance termination notice via instance metadata or CloudWatch Events and ***automate*** *by*:

- ☑ Checkpointing
- ☑ Draining from ELB
- ☑ Using stop-start and hibernate to restart faster

aws

EC2 Spot with  Amazon EKS

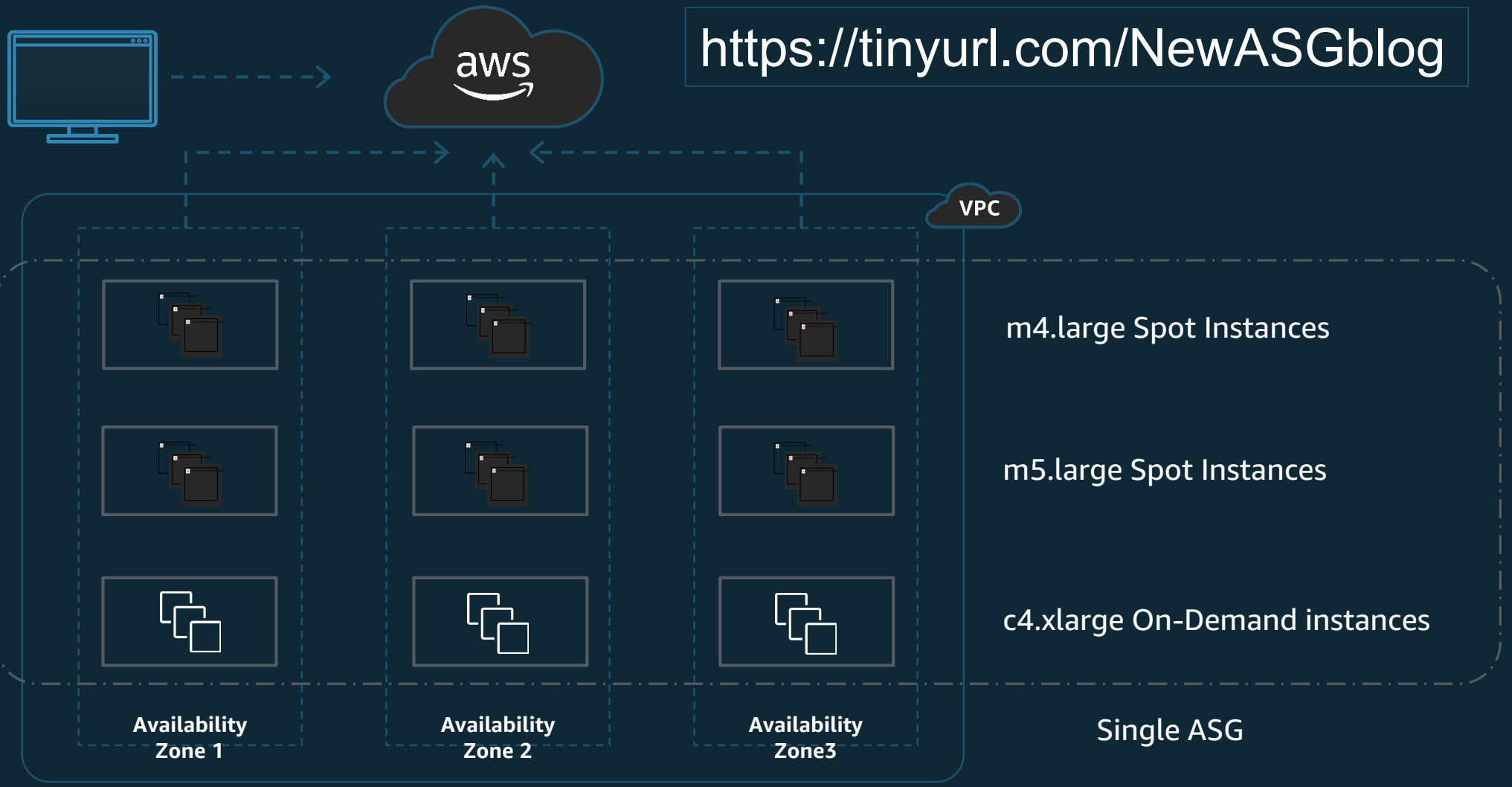
kubernetes

- Run a DaemonSet on every worker to catch the Spot interruption and coordon & drain the node
- Use labels to identify Spot nodes (for the DaemonSet, and other purposes – affinity & tolerations?)

aws

Multiple instance types & purchase options in ASG

https://tinyurl.com/NewASGblog



The new ASG combines purchase options, instance types and AZs in a single ASG

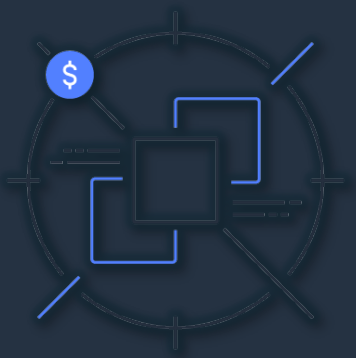
aws

Main takeaways for success with Spot

- Be instance type agnostic and let ASG/Fleet provide the required capacity at the lowest price
- Adopt Launch Templates to benefit from new ASG and Fleet features
- New instance families generally have higher interruption rates
- Architect for fault-tolerance to be Spot compatible and increase your availability

aws

+



EC2 Spot instance Workshops:
https://ec2spotworkshops.com

Choice of processors and architectures

Intel Xeon Scalable (Skylake) processor

AMD EPYC processor

AWS Graviton processor 64-bit Arm

\+ Choice of GPUs and FPGAs for compute acceleration

Right compute for each application and workload

aws

First instance powered by AWS Graviton Processor

Amazon EC2 A1

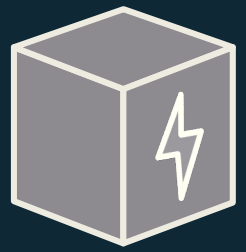
Run scale-out and Arm-based applications in the cloud

Up to 45% cost savings

AWS Graviton Processor
64-bit Arm Neoverse cores and
custom AWS silicon

Flexibility and choice for your workloads

Lower cost



Maximize resource efficiency with AWS Nitro System

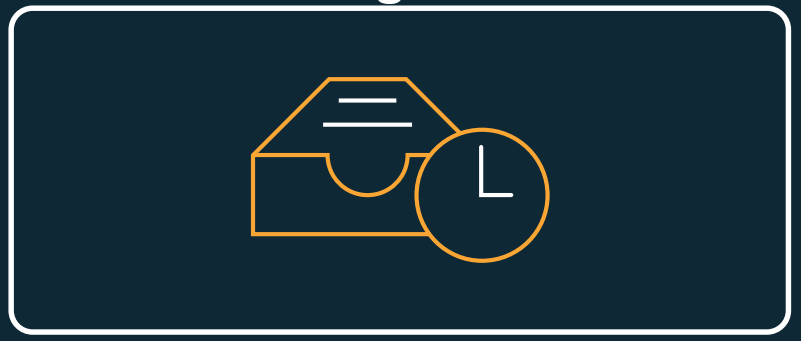
aws

Target applications for Amazon EC2 A1

Web tier

Containerized microservices

Caching fleets



IoT, Gaming, Arm workloads

aws

Arm software ecosystem for A1 instances

OSVs and ISVs

Amazon Linux 2

Ubuntu 16.04, 18.04 and newer

Red Hat Enterprise Linux 7.6 and 8.0

SUSE Linux Enterprise Server for Arm 15

Docker Desktop Community and Docker Enterprise Engine (Beta)

Added since launch:
Fedora Rawhide, Fedora Atomic, Debian 10, and Ubuntu 18.10 (Bionic)
More coming soon.

Containers

Most Docker official images support arm64

Amazon ECS

Available today!

Amazon EKS

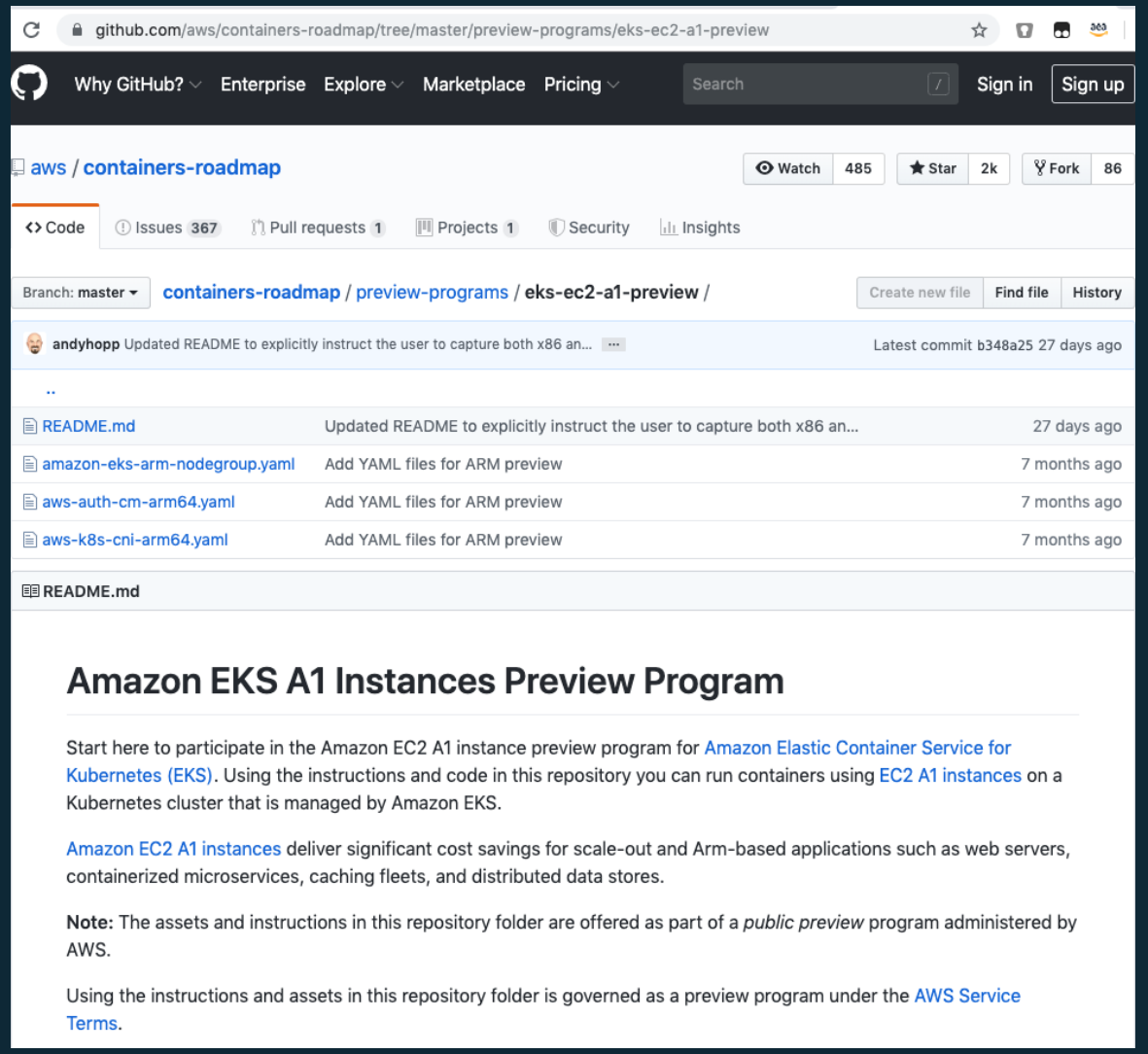
Public Preview!

Tools

- AWS Marketplace
- AWS Systems Manager
- Amazon CloudWatch
- AWS CodeCommit
- AWS Cloud9
- AWS CodePipeline
- Amazon Inspector
- AWS Batch

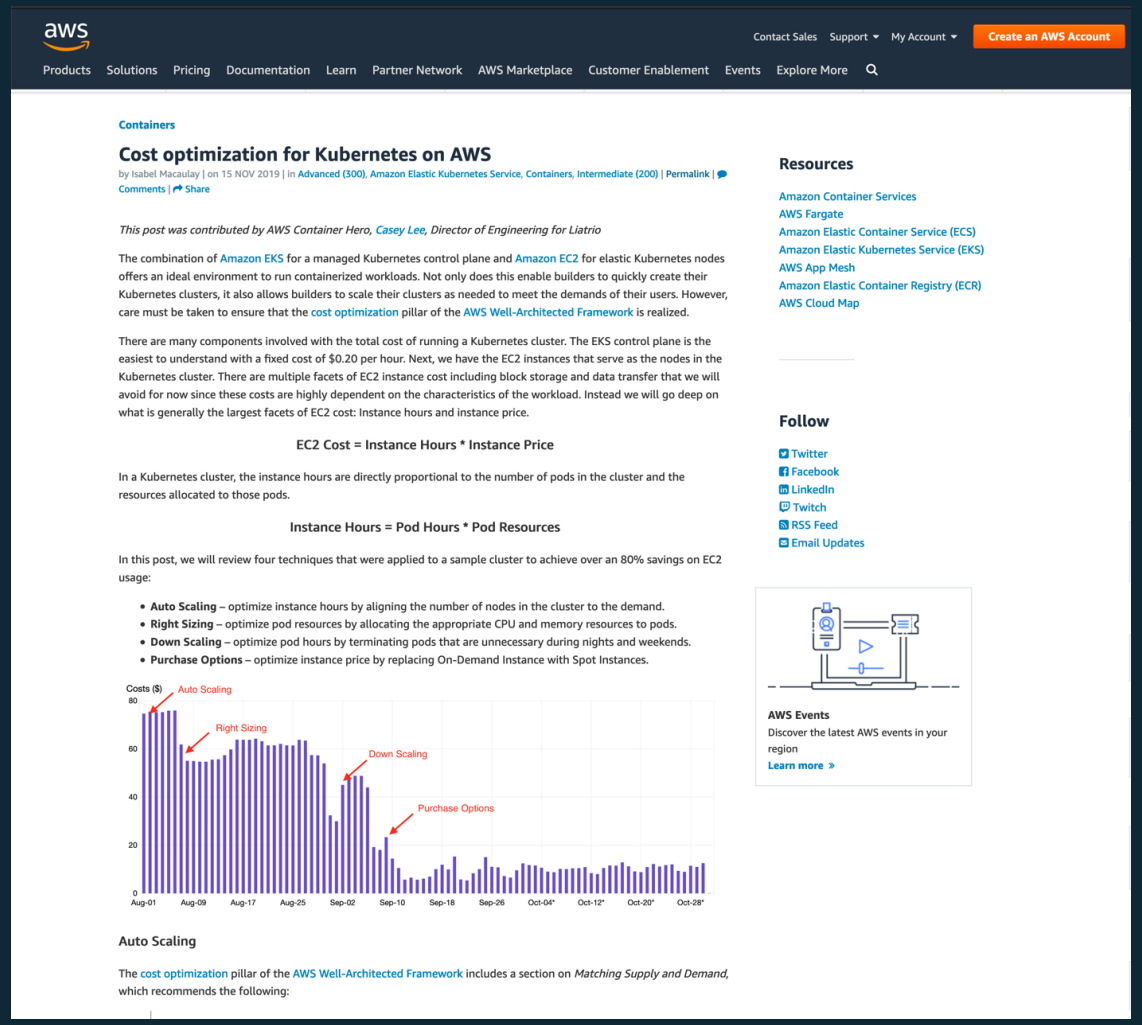
+ Amazon Corretto (OpenJDK)

Amazon EKS A1 Preview



Blog shout out - Cost optimization for Kubernetes on AWS

This post was contributed by AWS Container Hero, Casey Lee, Director of Engineering for Liatrio



https://aws.amazon.com/blogs/containers/cost-optimization-for-kubernetes-on-aws/

Demo: Docker Desktop x86_64/arm64 build

Demo goals

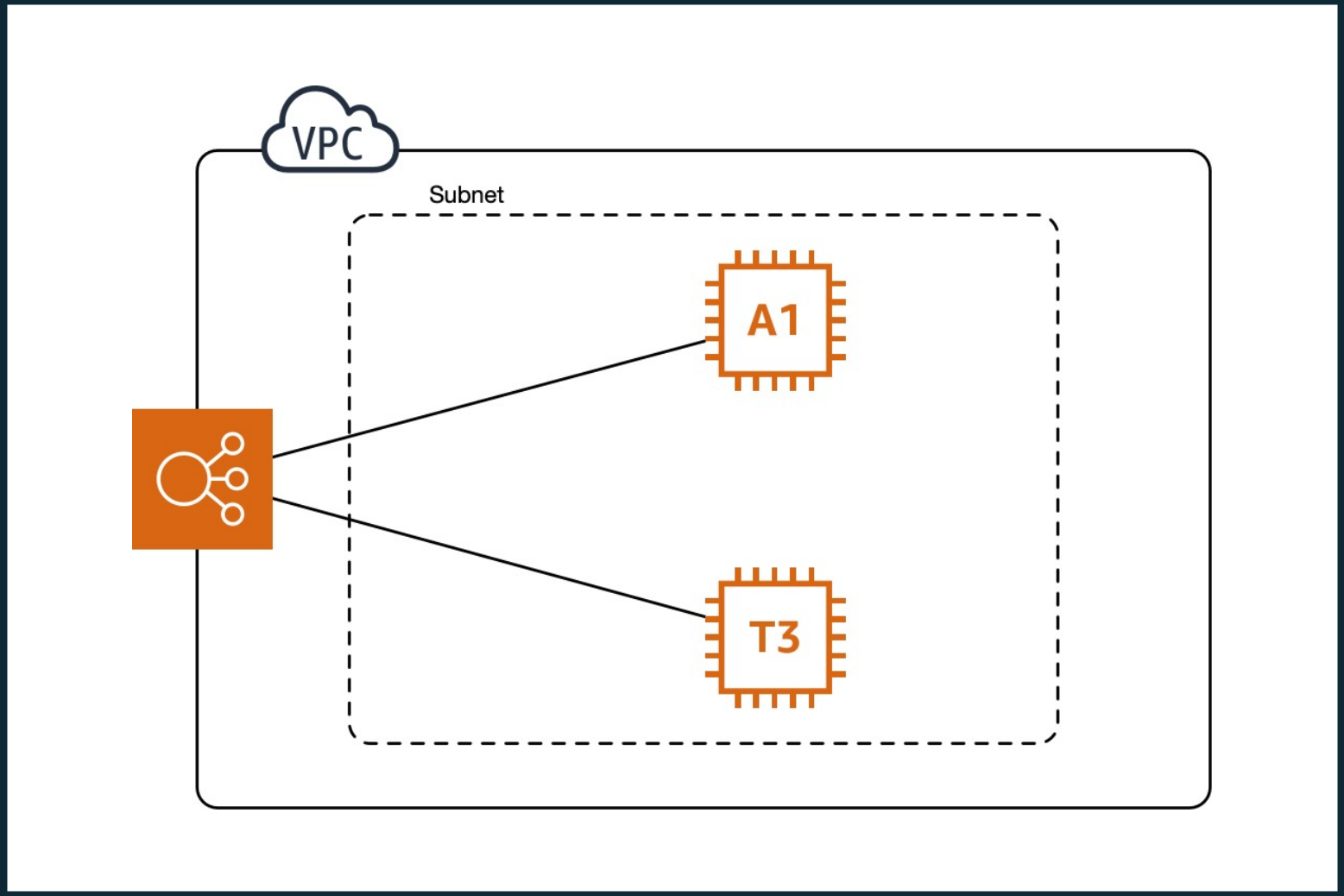
Show how seamless it is to build bi-arch containers and use them in a simple web application

What will we leverage ?

- Docker buildx
- AWS CloudFormation
- AWS Elastic Load Balancer
- Amazon EC2 A1 and T3 Instances

aws

Demo Architecture



Show me the code !

Wrap-up

What have we learnt ?

- There’s an easy transition path from x86_64 to arm64
- Bi-arch containers are easy to build with buildx
- AWS EC2 A1 instances are yet another cost reduction lever

aws

Thank you

Arthur Petitpierre – arthurpt@amazon.com
@ArthurPtP