AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

OPN312

Building a multi-architecture CI/CD pipeline for Kubernetes

Scott Malkie (he/him/his)
Sr. Solutions Architect
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

What are we doing here?

Why multi-architecture?

What is multi-architecture?

How to multi-architecture?

To the whiteboard!

Next steps and survey link

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Who are you and why are you here?



- Been with AWS for 6 years
- Previously an Amazon Elastic Kubernetes Service (Amazon EKS) Specialist Solutions Architect
- Currently working in Business Development for AWS Graviton
- Help customers solve business and technical challenges
- Based in Virginia (HQ2)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-architecture

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Broadest and deepest compute platform choice

CATEGORIES

General purpose

Burstable

Compute intensive

Memory intensive

Storage (High I/O)

Dense storage

GPU compute

Graphics intensive

+

CAPABILITIES

Choice of processor
(AWS, Intel, AMD)

Fast processors
(up to 4.0 GHz)

High memory footprint
(up to 12 TiB)

Instance storage
(HDD, SSD, NVMe)

Accelerated computing
(GPUs and FPGA)

Networking
(up to 100 Gbps)

Bare Metal

Size
(Nano to 32.xlarge)

+

OPTIONS

Amazon Elastic Block Store
(Amazon EBS)

Amazon Elastic Inference

=

MORE THAN

475

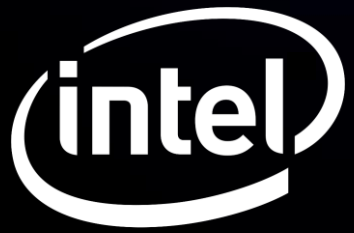
INSTANCE TYPES

for virtually every workload and business need

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Broadest choice of processors



Intel® Xeon Scalable processors



AMD EPYC processors



AWS Graviton processors

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton2-based Amazon EC2 instances

UP TO 40% BETTER PRICE-PERFORMANCE OVER COMPARABLE X86-BASED INSTANCES

M6g, M6gd

General purpose workloads

FREE TRIAL

T4g

Burstable general purpose workloads

C6g, C6gd, C6gn

Compute-intensive workloads

R6g, R6gd, X2gd

Memory-intensive workloads

NEW!

Im4gn, Is4gen

Storage-intensive workloads

NEW!

G5g

GPU-based graphics and machine learning workloads

AVAILABLE ACROSS 23 AWS REGIONS GLOBALLY*

* Not all instances available in all regions

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

New: Amazon EC2 G5g instances powered by AWS Graviton2

THE FIRST GRAVITON-BASED INSTANCES TO FEATURE GPU ACCELERATION



Most cost-effective instances in Amazon Elastic Compute Cloud (Amazon EC2) for Android game streaming with up to 30% lower cost per stream per hour

Get significantly lower cost-per-inference for machine learning inference over x86-based GPU instances

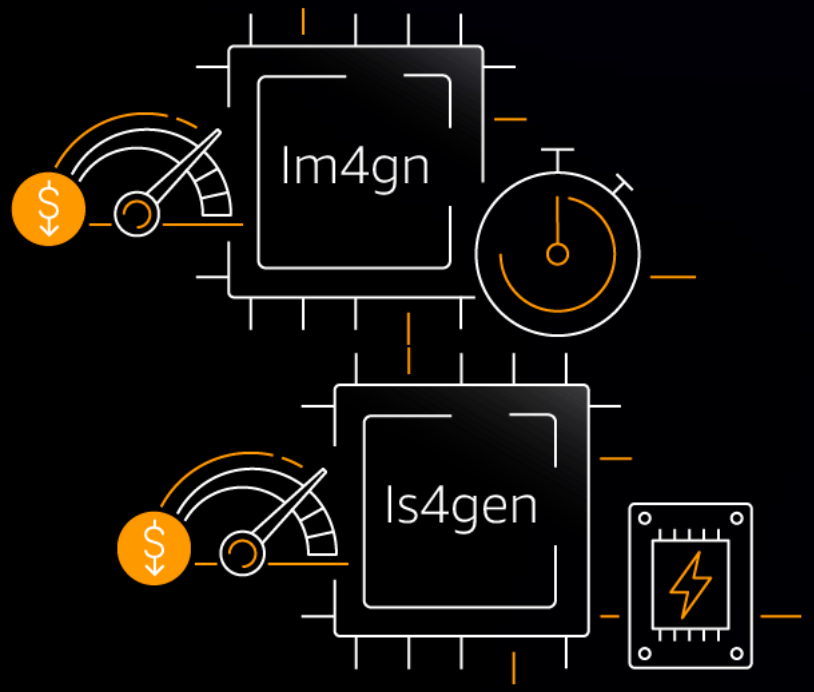
Powered by AWS Graviton2 and featuring NVIDIA T4G Tensor Core GPUs

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

New: Amazon EC2 Im4gn and Is4gen instances powered by AWS Graviton2

THE BEST PRICE-PERFORMANCE FOR STORAGE-INTESIVE WORKLOADS IN AMAZON EC2



Im4gn instances provide up to 40% better price performance compared to I3 instances

Is4gen instances offer the lowest cost per TB of SSD storage in Amazon EC2

Both instances feature the new AWS Nitro SSDs, custom-designed by AWS to provide high I/O performance, low latency, and low latency variability for fast, predictable performance

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lower TCO with Graviton2-powered instances



Highest performance
in their instance families



20% lower cost
vs same-sized, comparable fifth generation instances



Up to 40% better price-performance
vs comparable x86-based instances

Best price-performance within their instance families

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Graviton2 ease of adoption

AS A GENERAL RULE, THE MORE CURRENT YOUR SOFTWARE STACK THE BETTER

Difficulty	Workload	Actions
Virtually no effort	Amazon RDS, Amazon Aurora, Amazon ElastiCache, Amazon OpenSearch Service	Upgrade to latest and enjoy
Super easy	Amazon EMR	Typically, just works
Pretty easy	AWS Lambda	Typically, just works with Lambda managed runtimes or base images. Watch: JNI or Python-native modules
Quite easy	Linux – Interpreted and JIT'd languages (e.g., Java, PHP, Node.js)	Select Arm64 AMI and Install Bonus if using containerized Watch: JNI or Python-native modules
More involved	Linux – Compiled languages (e.g., C/C++, Python, Go)	Select Arm64 AMI and compile Watch: port any intrinsics or assembly
Some work, high reward	Microsoft Windows – .NET	Migrate to Linux + .NET core on Arm64
Sorry, not yet	Microsoft Windows	Microsoft Windows Server not yet available for arm64

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Arm64: Languages, toolkits, and runtimes

- Interpreted and compiled-bytecode languages can run without modification
  - Python, Java, Ruby, PHP, Node.js, many others
  - .NET Core supports Linux and arm64
  - Typically “just works”
- Compiled applications will need to be recompiled for arm64
  - Every major compiler supports arm64
  - C, C++, Go , Rust all support arm64
- AMIs: Most major Linux distributions including NetBSD/FreeBSD have arm64 versions

- Most AWS Tools and SDKs support Graviton2 transparently
  - AWS CLI v1, AWS CLI v2, CloudWatch agent, SSM agent
  - SDKs for C/C++, Node.js, Python, Go, Java, .NET

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Arm64 and container services

- **Amazon Elastic Container Service (Amazon ECS)**, **Amazon Elastic Kubernetes Service (Amazon EKS)**, and **AWS Fargate (ECS)** support AWS Graviton, including mixed x86 and Arm clusters
- **Amazon Elastic Container Registry (Amazon ECR)**, including **Amazon ECR Public** supports multi-architecture manifests
- **AWS CodeBuild** supports Graviton-powered instances and targets
- **Docker** itself supports a multi-arch toolchain through **buildx**

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step 1 – Identify dependencies and status

- **Identify** all libraries, dependencies, and agents in your production environment
  - Language-standard and open-source libraries
  - Paid-for proprietary or commercial libraries
  - Monitoring, logging, or security daemons or agents

- **Check** arm64 support status
  - Look for existing arm64 binaries or architecture-agnostic source code (OSS)
  - Ask vendor or provider for roadmap towards arm64 support
  - **Reach out to your AWS Account Team** who can help investigate or escalate

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step 2 - Test and deploy

- **Test!** All tests should support multiple architectures, and that may mean writing new tests; this is especially pertinent if you have to recompile your application
    - Perform the usual range of unit testing, acceptance testing, and pre-production testing
- **Update** your infrastructure as code (IaC) resources, leveraging tools like AWS CloudFormation or the AWS Cloud Development Kit (AWS CDK), to provision your application to arm64 instances
    - This will likely be a simple change - modifying the instance type, AMI, and user data to reflect the Graviton2 instance
- **Deploy** via a Green/Blue method; either manually (or via existing CI/CD tooling) create your new arm64-based stack alongside your existing stack; then, leverage weighted routing to send a small percentage of requests to the new stack
    - Monitor error rates, user behavior, load, and other critical factors in order to determine the health of the ported application in production

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Examples

…to the whiteboard!

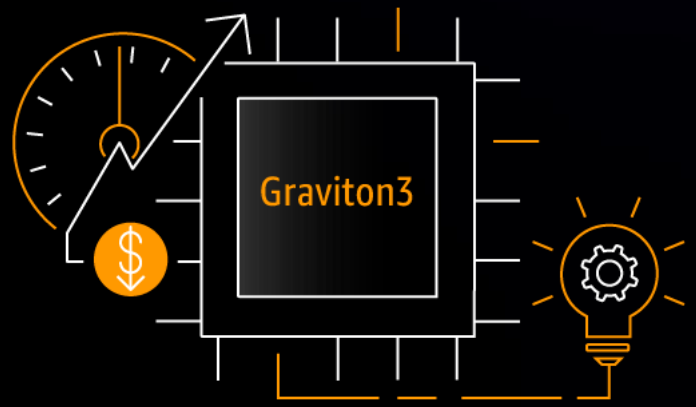
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Next steps

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Announcing AWS Graviton3 and Amazon EC2 C7g instances

ENABLING THE BEST PRICE PERFORMANCE FOR WORKLOADS IN AMAZON EC2



Up to 25% better performance compared to Graviton2

Up to 2x higher floating-point performance, up to 2x faster cryptographic workload performance, and up to 3x better machine learning performance compared to Graviton2

First in the cloud to feature DDR5 memory

60% more energy efficient over comparable EC2 instances

C7g instances, available in preview, will provide the best price performance for compute-intensive workloads in Amazon EC2

AWS Graviton3-based Amazon EC2 C7g instances are currently in preview

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton getting started guide

https://github.com/aws/aws-graviton-getting-started

This guide has been assembled by our Graviton team and is designed to help customers transition and optimize their applications

It covers various languages and libraries, including tips and tricks for each

In general, using latest versions of operating systems, compilers, and language runtimes will provide access to latest arm64 improvements and optimizations

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Scott Malkie

malkie@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.