
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CMP328

Lower your IT carbon footprint with AWS compute services

Paul Moran (he/him)

Principal Technical Account Manager
AWS Enterprise Support

Zi Shen Lim (he/him)

Sustainability GTM, EC2 Graviton
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS sustainability journey

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon’s take on sustainability

Meeting the needs of the present without compromising the ability of future generations to meet their needs*

United Nations definition of sustainability

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

*https://www.un.org/en/academic-impact/sustainability

Sustainability is evolving

From ...	To/and ...
Bolt on	Built in
The right thing to do	The smart thing to do
Less harm	More good
Risk driven	Opportunity led
Managing past impacts	Forward-looking strategy
Sustainability of and in the cloud	Sustainability through the cloud

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Essential terminology

The GHG Protocol defines how to convert greenhouse gases to CO_2 equivalent (CO_2e)

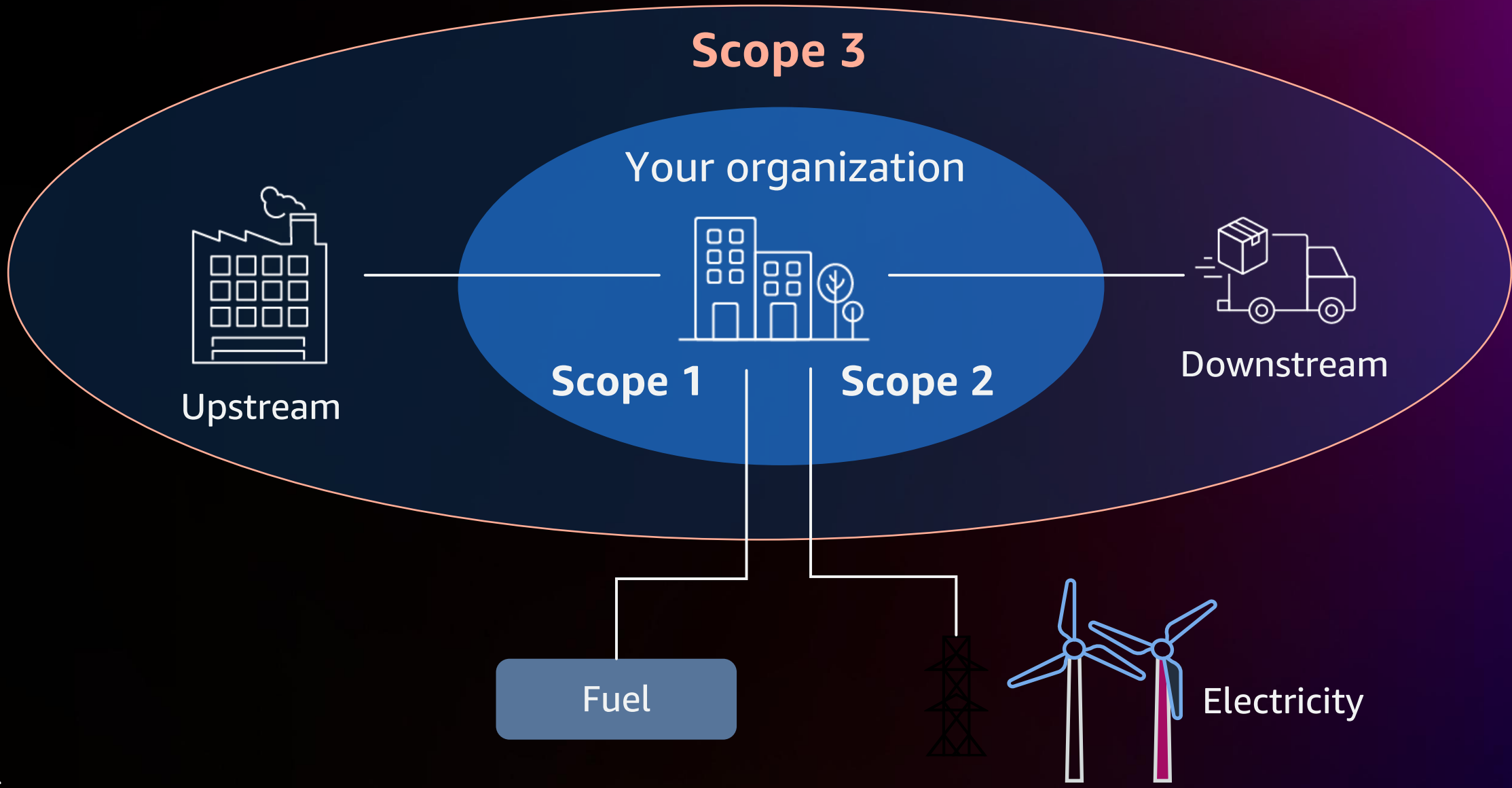
and the three scopes of emissions

Greenhouse Gas Protocol https://ghgprotocol.org

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GHG scopes



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS take on sustainability

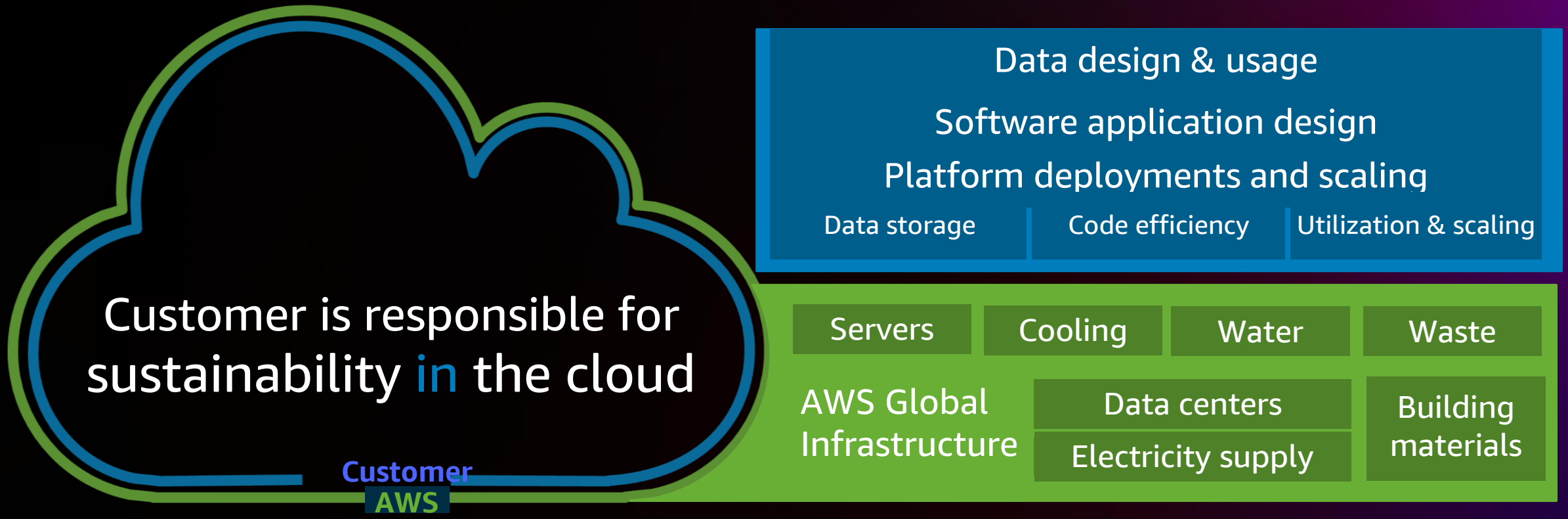
Sustainability in the cloud is a continuous, focused effort on

energy reduction and efficiency across all components of a workload

But there is also on the cloud and through the cloud

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Shared sustainability responsibility

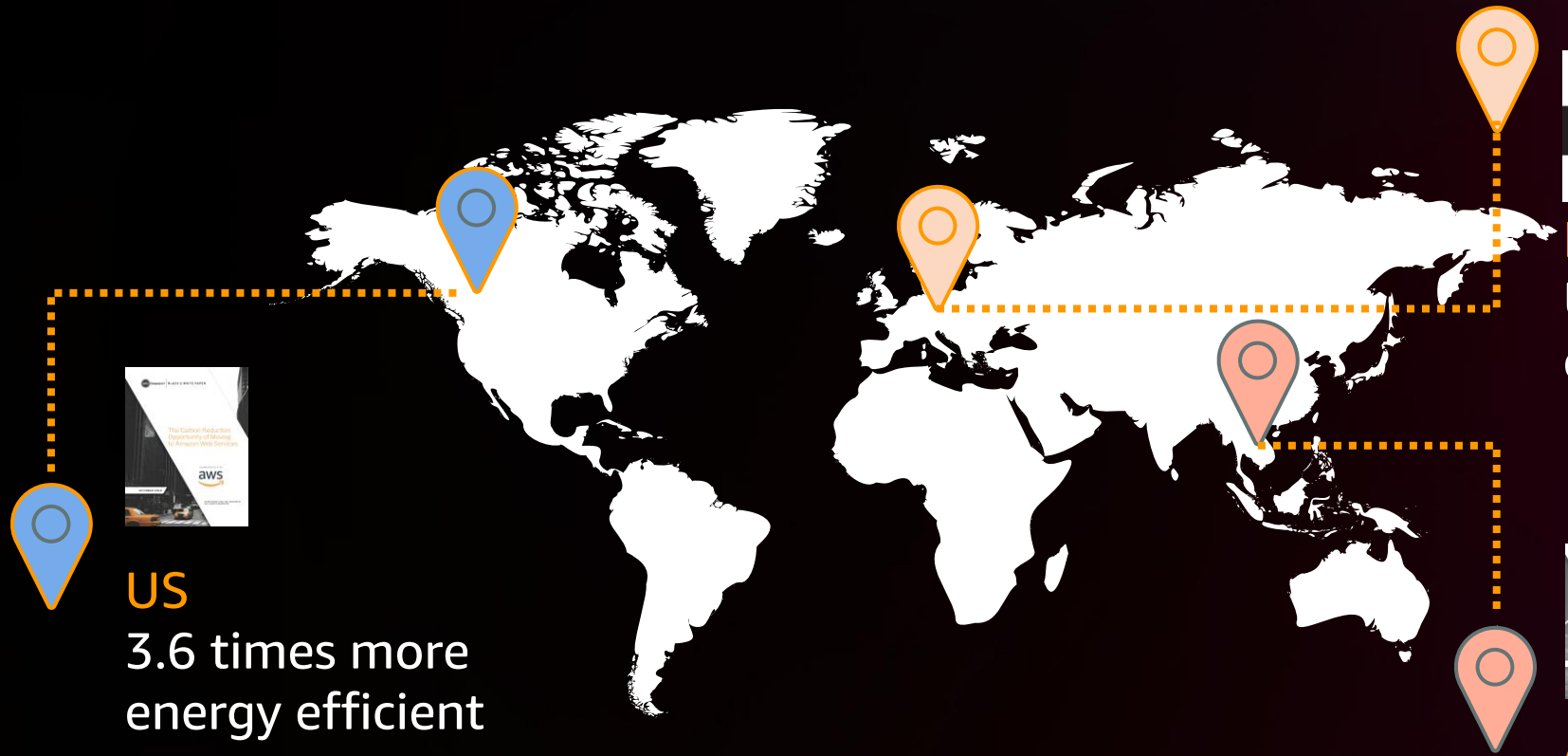


AWS is responsible for sustainability **of** the cloud

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Carbon reduction opportunity

AWS can lower the carbon footprint of average on-premises data center workloads by nearly 80% today and up to 96% once AWS is powered with 100% renewable energy



US
3.6 times more energy efficient



Europe
Up to 5 times more energy efficient

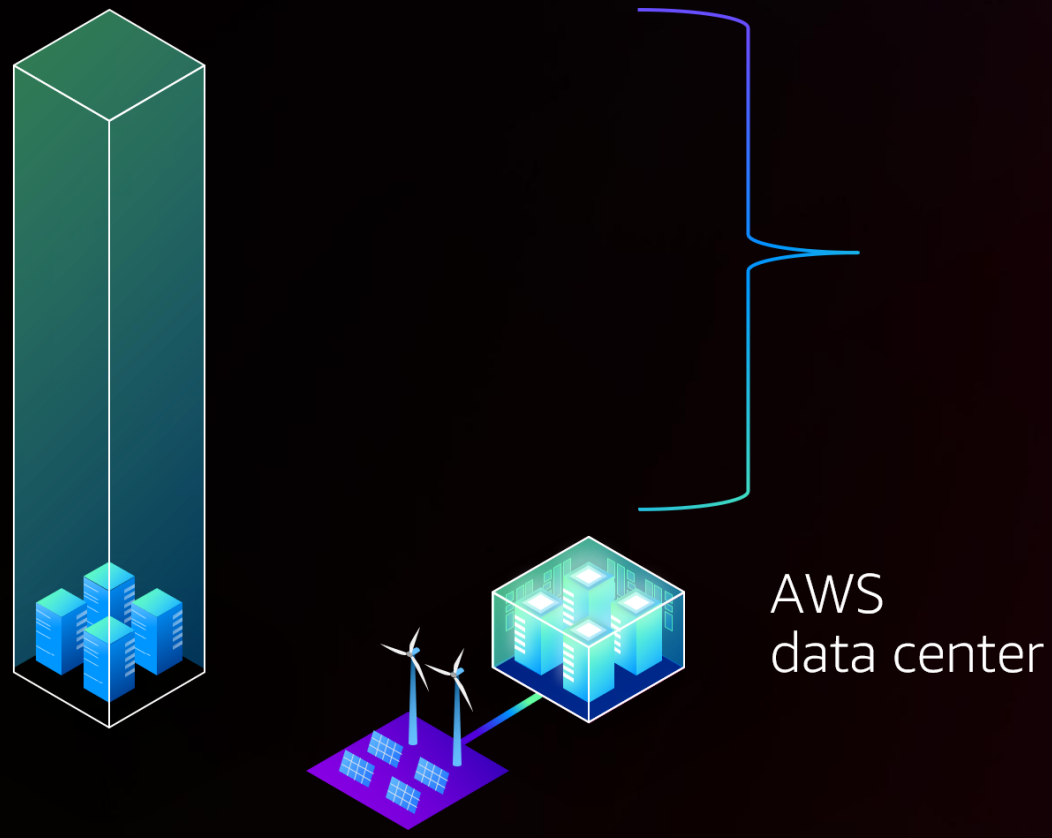


Asia Pacific
5 times more energy efficient

Find all the reports on aws.amazon.com/sustainability/resources/

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Report findings



Up to 80% reduction in energy usage by moving workloads to AWS

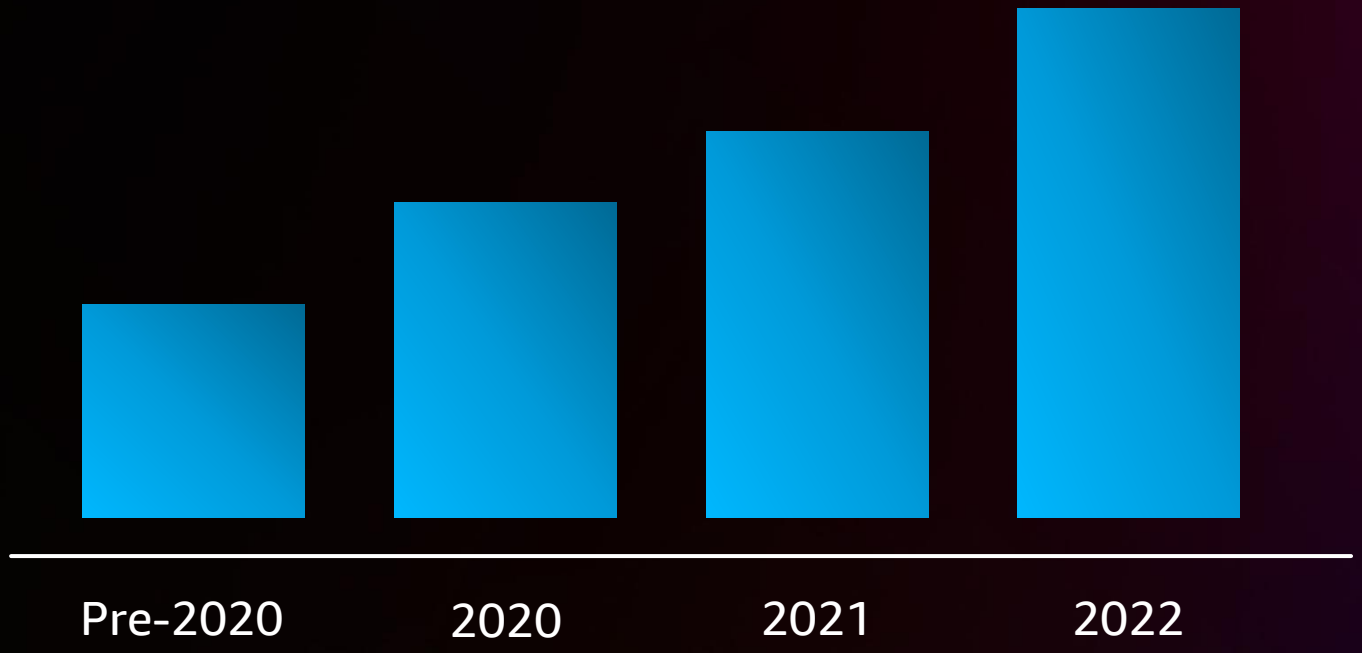
- Efficient facilities
- Efficient servers
- Higher utilization
- Renewable energy

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Path to 2025

Total renewable energy capacity

September 2022 18.5 GW



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

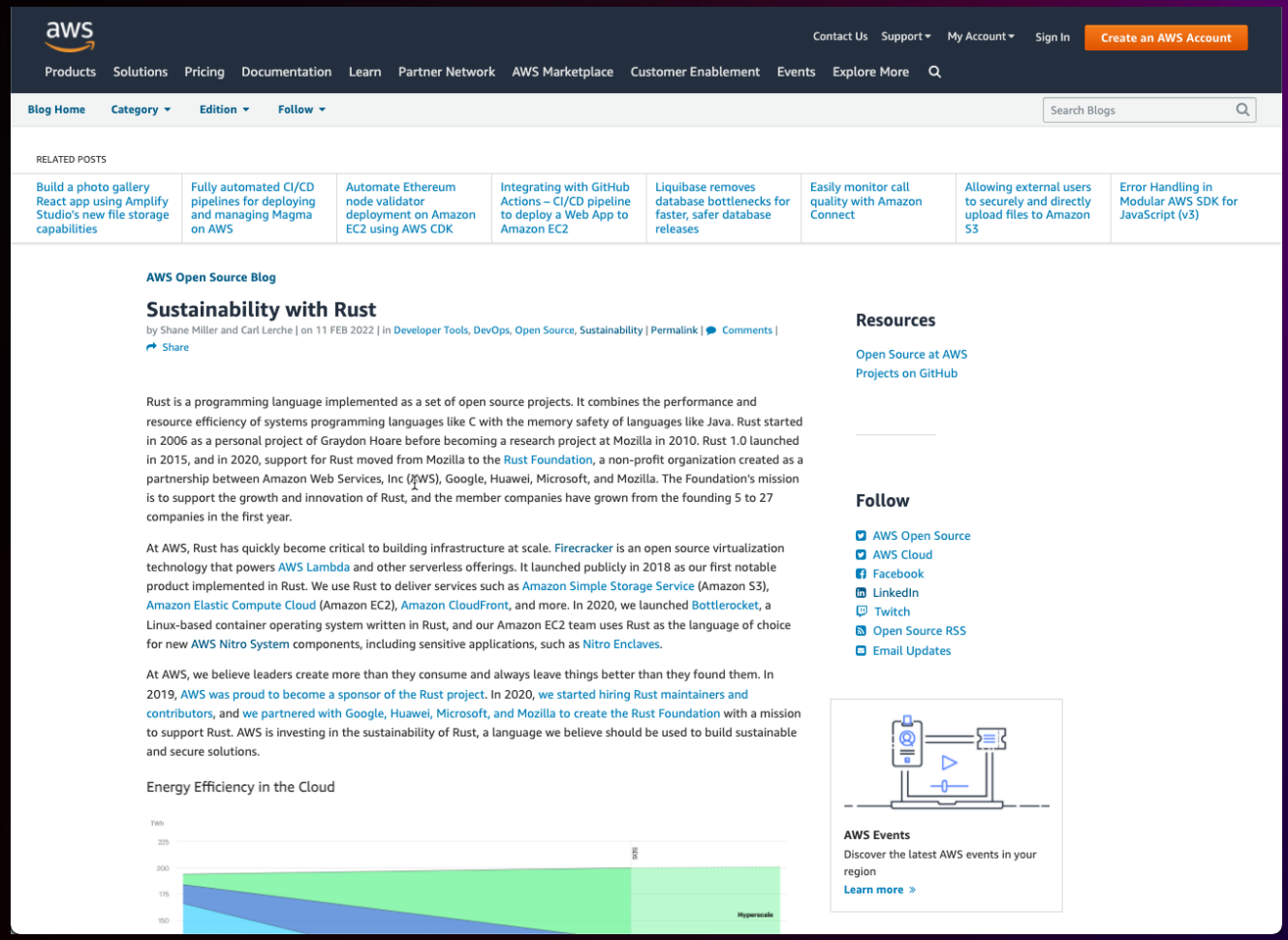
We need to go further

"Renewables should not replace energy efficiency as a design principle"

Shane Miller, Rust Foundation Chair, Sustainability with Rust blog

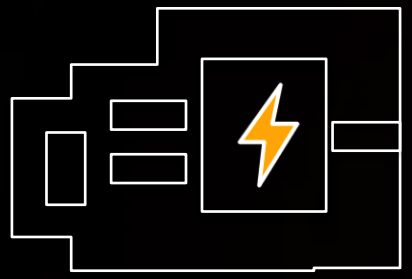
"The greenest energy is the energy we don't use"

Peter DeSantis, SVP, AWS, re:Invent 2020



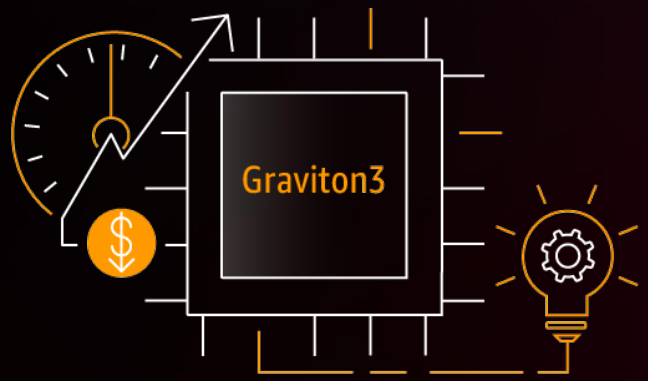
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Journey of silicon innovation at AWS



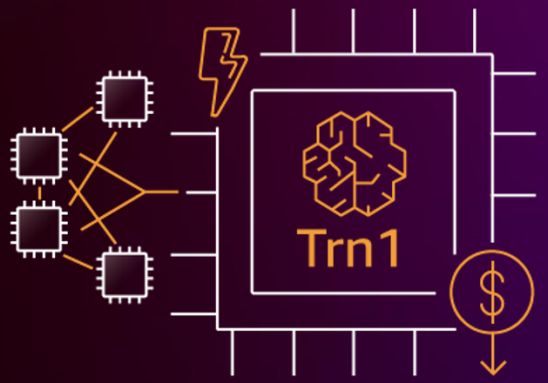
AWS Nitro System

Hypervisor, Nitro Cards, network, storage, SSD, and security



AWS Graviton

Powerful and efficient, modern applications



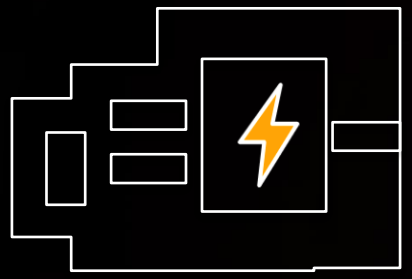
AWS Inferentia and AWS Trainium

Machine learning acceleration

aws

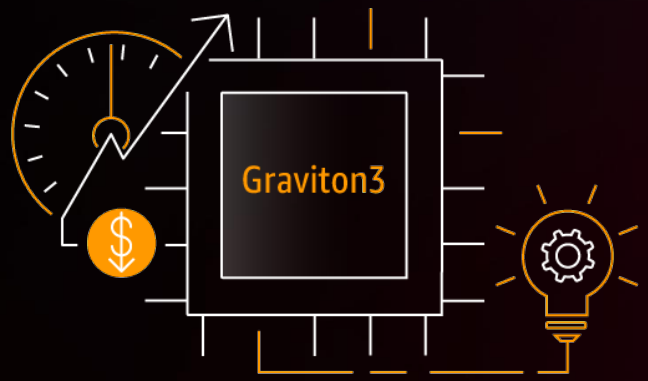
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Journey of silicon innovation at AWS



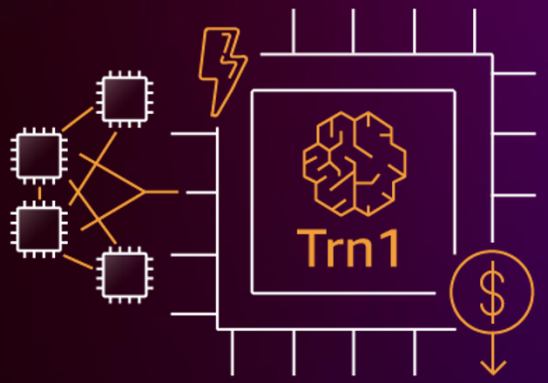
AWS Nitro System

Hypervisor, Nitro Cards, network, storage, SSD, and security



AWS Graviton

Powerful and efficient, modern applications



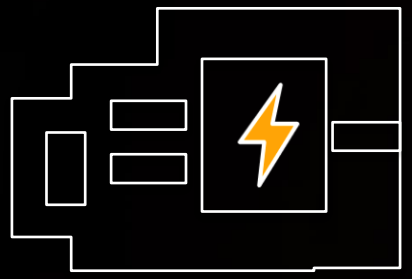
AWS Inferentia and AWS Trainium

Machine learning acceleration

aws

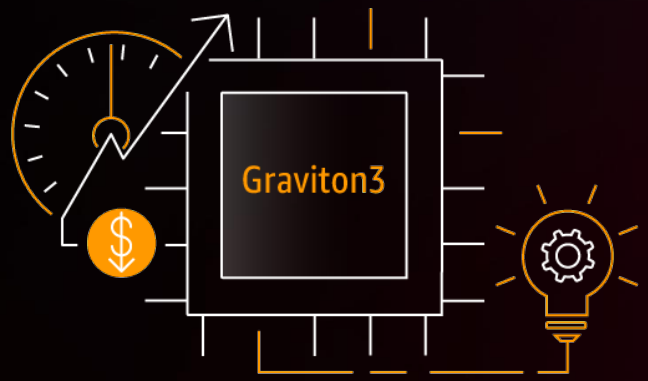
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Journey of silicon innovation at AWS



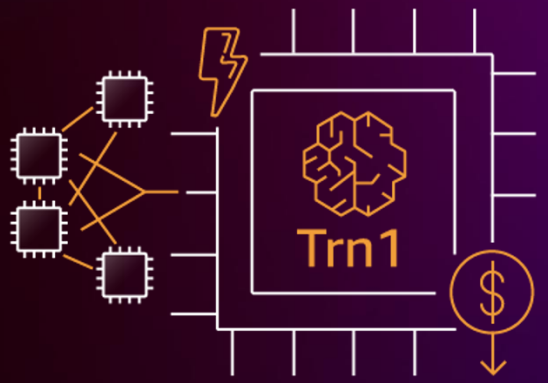
AWS Nitro System

Hypervisor, Nitro Cards, network, storage, SSD, and security



AWS Graviton

Powerful and efficient, modern applications



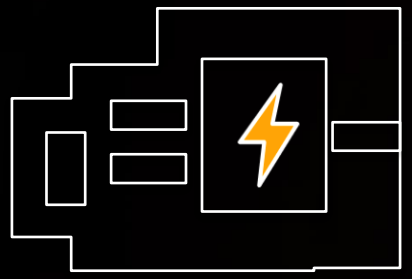
AWS Inferentia and AWS Trainium

Machine learning acceleration

aws

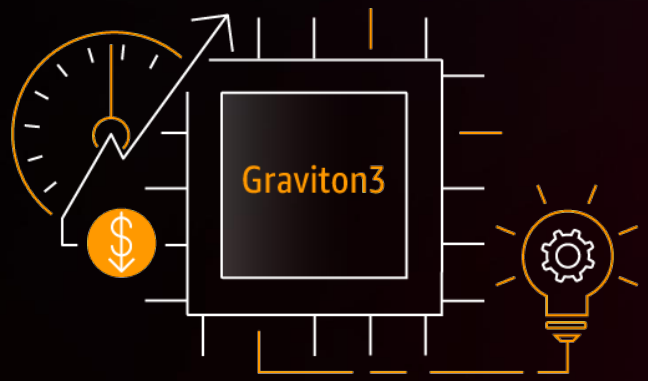
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Journey of silicon innovation at AWS



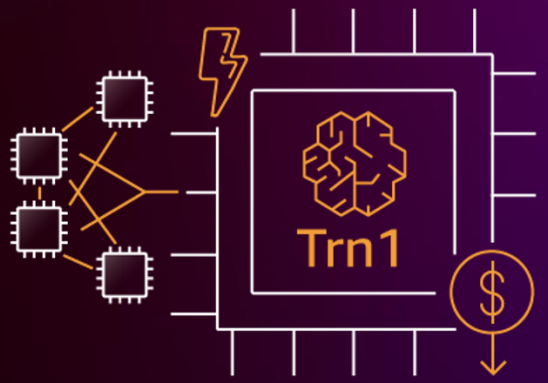
AWS Nitro System

Hypervisor, Nitro Cards, network, storage, SSD, and security



AWS Graviton

Powerful and efficient, modern applications



AWS Inferentia and AWS Trainium

Machine learning acceleration

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why build our own chips?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why build our own chips?



Specialization

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why build our own chips?



Specialization



Speed

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why build our own chips?



Specialization



Speed



Innovation

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why build our own chips?



Specialization



Speed



Innovation



Security

aws

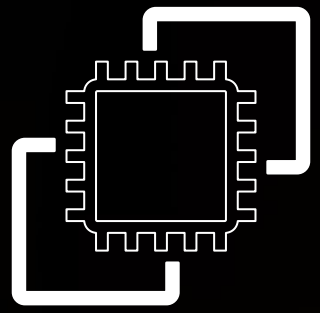
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Nitro System

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Nitro System

If we applied all of our learnings, how would we change our server platforms?



AWS Nitro System



Improve throughput

Simplify hypervisor

Reduce latency & jitter

Bare-metal instances



Transparent encryption

Hardware root of trust

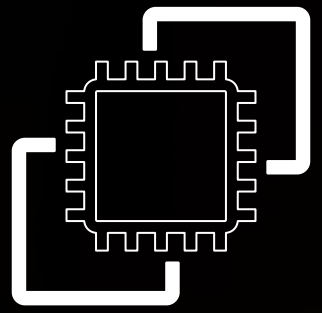
No operator access

Narrow auditable APIs

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Nitro System



AWS Nitro System

Purpose-built hardware/software

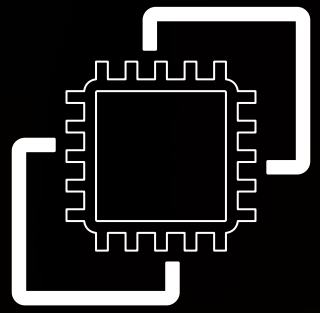
5 generations of custom chips

Hypervisor built for AWS

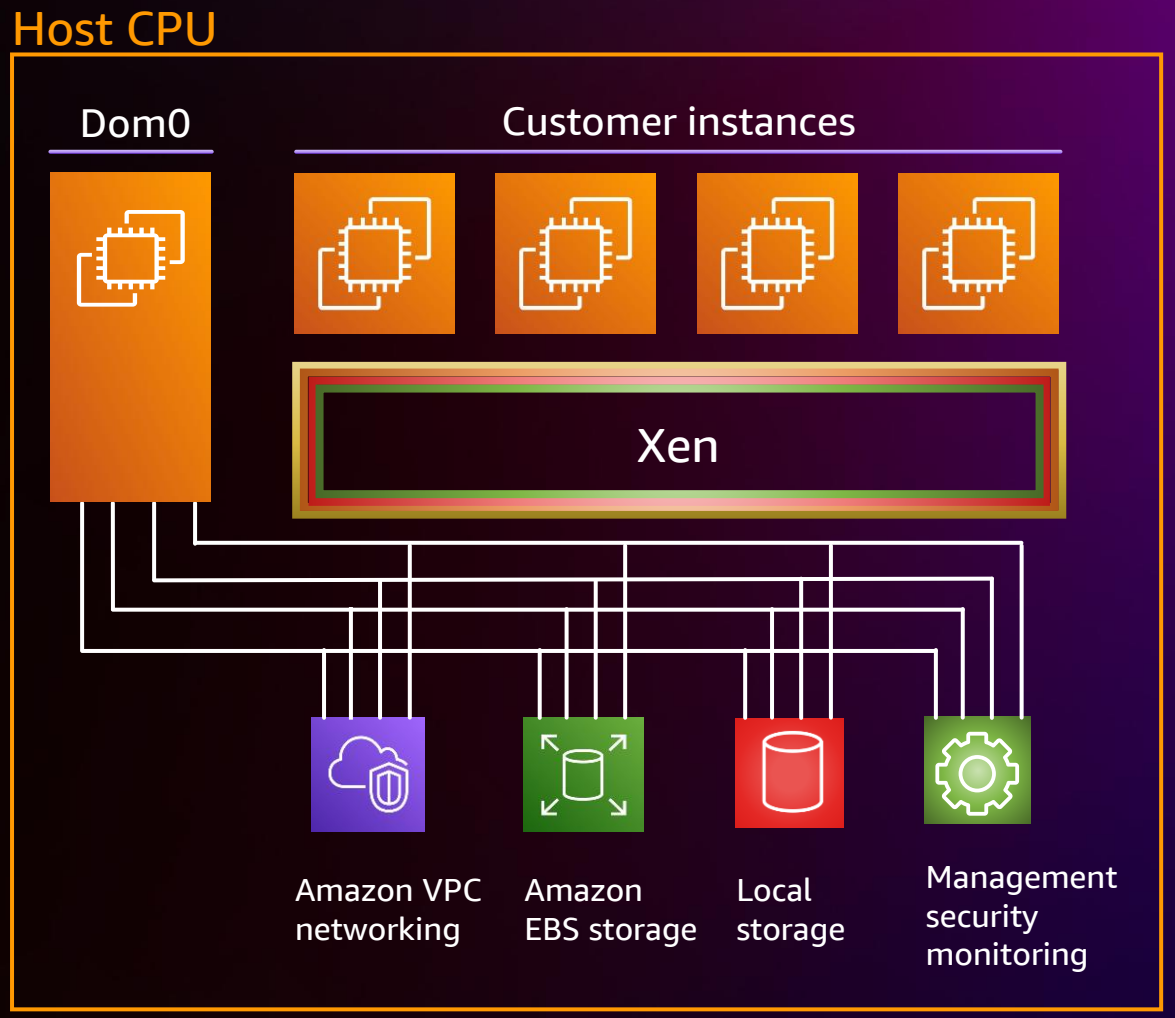
Powering over 500 instance types

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Before Nitro ...



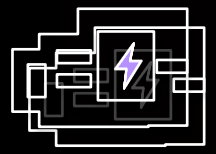
AWS Nitro System



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro Card for VPC networking

Nitro Cards



VPC networking

Amazon EBS

Nitro SSDs

System controller

VPC data plane offload

Encapsulation, security groups, flow logs, routing, port mirroring, DHCP, DNS

VPC encryption

Transparent end-to-end 256-bit encryption

Elastic Network Adapter (ENA)

Extendible host interface

Elastic Fabric Adapter (EFA)

Kernel bypass

Low latency at scale

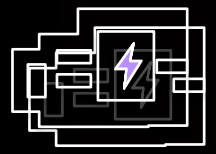
Multi-pathing (SRD)

aws

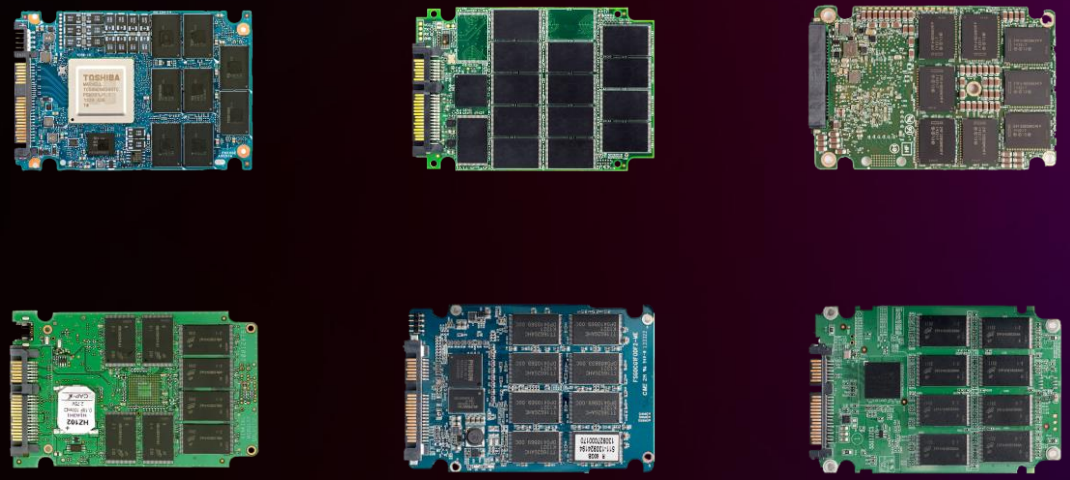
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro SSDs

Nitro Cards



VPC networking
Amazon EBS
Nitro SSDs
System controller

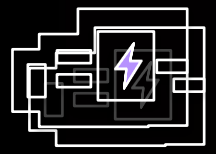


Each vendor has their own FTL and differing performance

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro SSDs

Nitro Cards

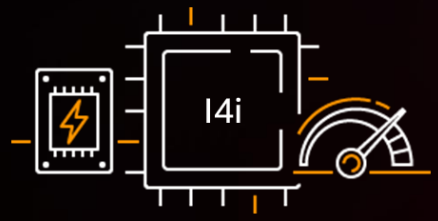


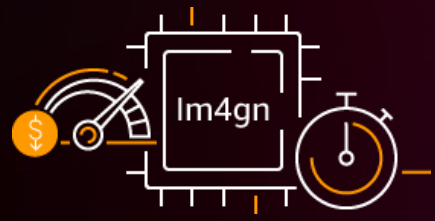
VPC networking

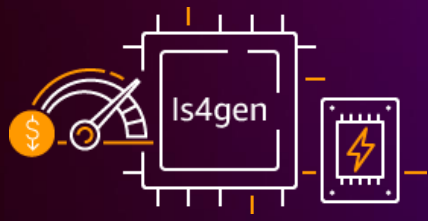
Amazon EBS

Nitro SSDs

System controller







60%

LOWER AVERAGE I/O LATENCY

75%

LOWER TAIL LATENCY

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Reducing storage costs

UP TO 50% DECREASE IN TCO



Splunk is a leading data platform provider and is designed to investigate, monitor, analyze, and act on data at any scale

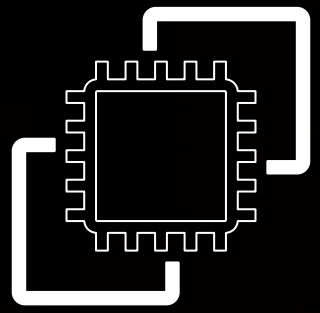
"When evaluating the new Im4gn/Is4gen instances powered by AWS Graviton2, we observed an up to 50% decrease in search runtime compared to I3/I3en instances, which we currently use"

Brad Murphy, VP of Engineering, Splunk

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Nitro System



AWS Nitro System

Nitro Card

Amazon VPC networking

Amazon EBS storage

Local storage

Host CPU

Dom0

Customer instances

Xen

Amazon VPC networking

Amazon EBS storage

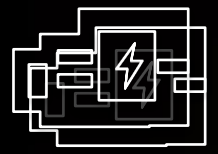
Local storage

Management security monitoring

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Nitro System

Nitro Cards



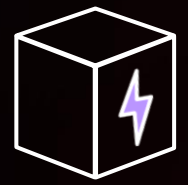
VPC networking

Amazon EBS

Nitro SSDs

System controller

Nitro Hypervisor



Lightweight hypervisor

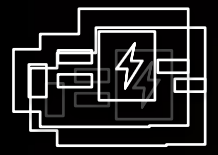
Memory and CPU allocation

Bare metal–like performance

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

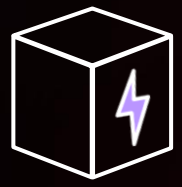
AWS Nitro System

Nitro Cards



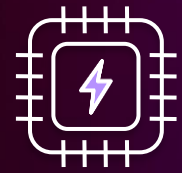
VPC networking
Amazon EBS
Nitro SSDs
System controller

Nitro Hypervisor



Lightweight hypervisor
Memory and CPU allocation
Bare metal–like performance

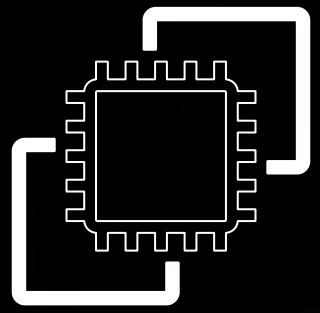
Nitro Security Chip



Integrated into motherboard
Traps I/O to nonvolatile storage
Hardware root of trust

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

After Nitro …



AWS Nitro System

Nitro Card



Amazon VPC networking



Amazon EBS storage



Local storage



Management security monitoring

Host CPU

Customer instances

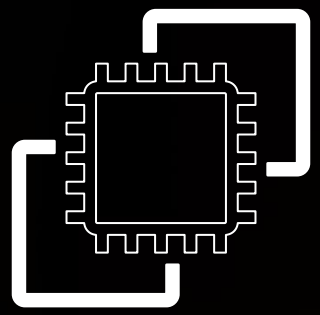


Nitro Hypervisor

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro security – Confidential compute



AWS Nitro System

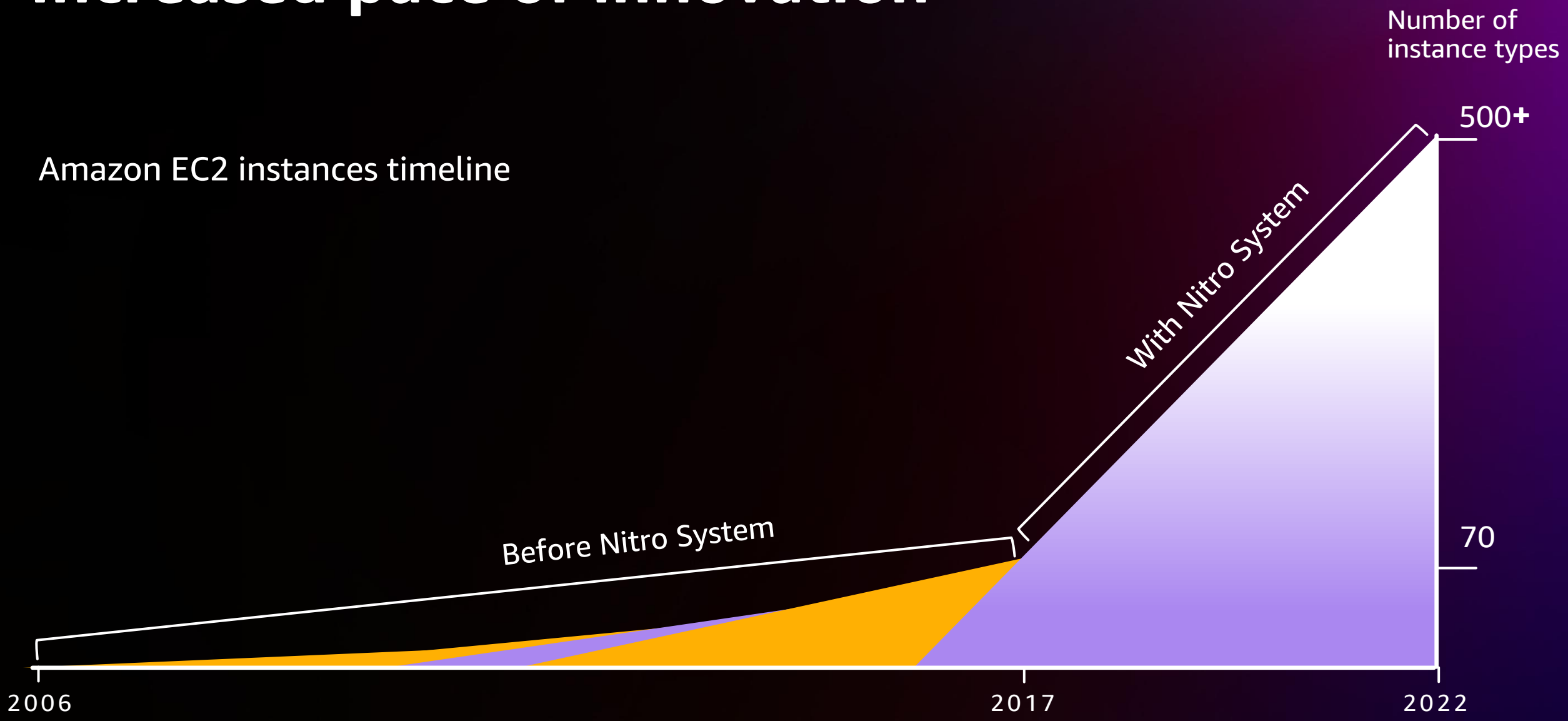
All interactions with the AWS Nitro System are through narrow, authorized, and authenticated APIs

There is no mechanism for any system or person to log in to the underlying EC2 host

There is no interactive access

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Increased pace of innovation



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Snap improved workload efficiency using Graviton



"Using Graviton2-based instances, we were able to reduce the Snapchat messaging fleet size and significantly lower cost compared to C5 instances. The move to C6g instances reduced CPU utilization by roughly 10% because of better performance on Graviton2. We are planning to move additional workloads to Graviton2 in the future and have found similar performance gains on other large workloads."

Aaron Sheldon, Software Engineer at Snap

Source: https://aws.amazon.com/ec2/graviton/customers/

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Graviton is helping Amazon meet sustainability goals

The same has been true with Graviton2. Because of this there has been a massive movement inside Amazon for general workloads to move to Graviton2, mainly to save on power, but also on costs. For the same workloads, Graviton2 will on average consume 60% less power than same-generation competitive offerings, and we're passing on those cost-savings to customers. Outside Amazon, at least 48 of AWS's top 50 customers have not just tested, but have production workloads running on Graviton2.

Amazon Science: **How silicon innovation became the "secret sauce" behind AWS's success** (July 27, 2022): https://www.amazon.science/how-silicon-innovation-became-the-secret-sauce-behind-awss-success

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton2: Prime Day 2021



- Increased compute by 12.5%, but due to increased efficiency, actually used about 6,000 fewer physical servers than Cyber Monday 2020
- Graviton2-powered EC2 instances supported 12 core retail services
- Datapath service is a key part of the Amazon site that supports lookups, queries, and joins across structured blobs of data
- Ran on a three-Region cluster composed of over 53,200 C6g instances
- The 40% price performance advantage of Graviton2, along with 20% lower cost, turns into a big win for us and for our customers

Source: https://aws.amazon.com/blogs/aws/prime-day-2021-two-chart-topping-days/

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton

BEST PRICE PERFORMANCE IN AMAZON EC2

AWS Graviton
2018



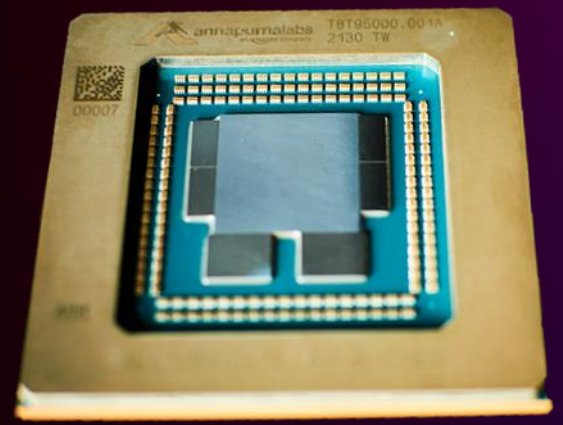
Significant cost savings for scale-out workloads

AWS Graviton2
2019



Up to 40% better price performance over comparable x86-based instances for a wide range of workloads

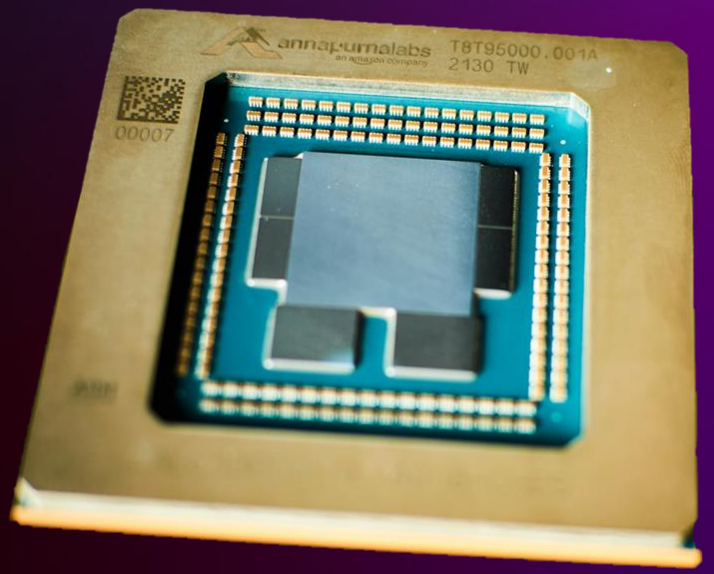
AWS Graviton3
2021



Up to 25% better performance over Graviton2 for compute-intensive workloads

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton3



Better compute performance vs Graviton2

- Up to 25% better compute performance
- Up to 2x higher floating-point performance
- Up to 2x faster cryptographic workload performance
- Up to 3x better performance for ML workloads, including support for bfloat16

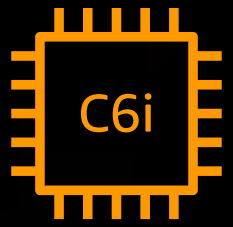
First DDR5 system in our data centers – 50% more DDR bandwidth

Up to 60% less energy for the same performance than comparable EC2 instances

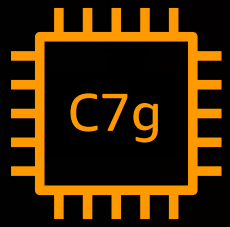
https://aws.amazon.com/ec2/graviton/

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Graviton3 – vCPU



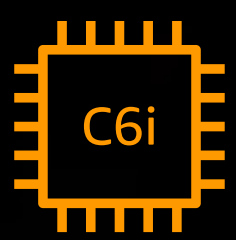
C6i instance



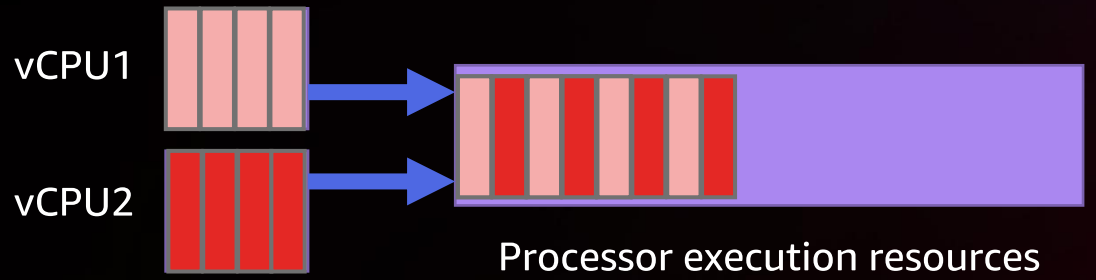
C7g instance

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

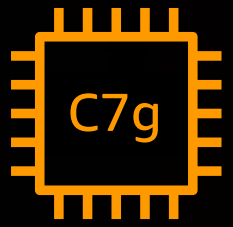
Graviton3 – vCPU



C6i instance



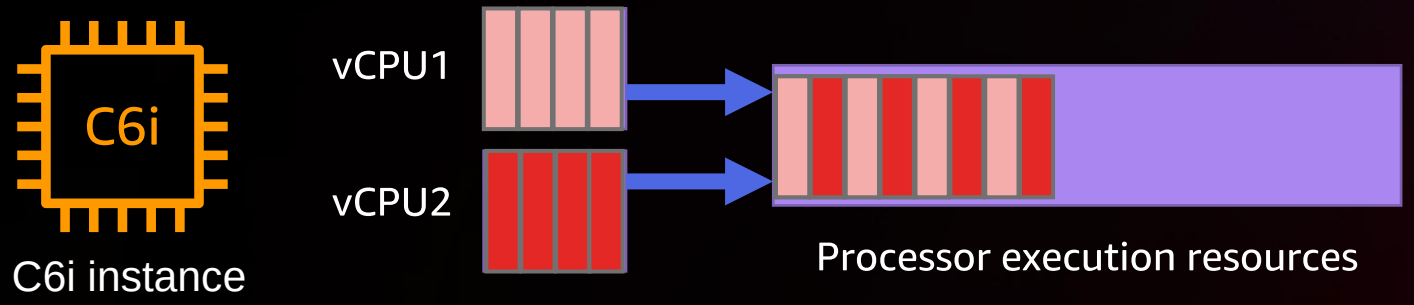
Simultaneous multithreading (SMT)



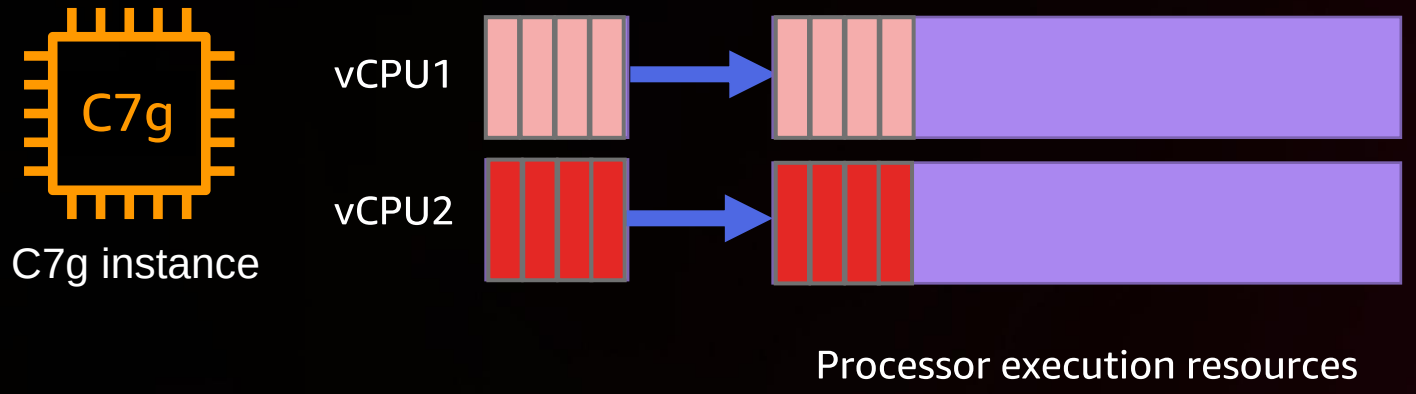
C7g instance

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Graviton3 – vCPU



Simultaneous multithreading (SMT)



Every vCPU is a physical core

No simultaneous multithreading (SMT)

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Innovation at the server level

Typical servers

- 2 sockets/server
- 42U rack

Traditionally run out of rack power before space

Graviton runs out of space before power

How can we put more sockets in a rack?

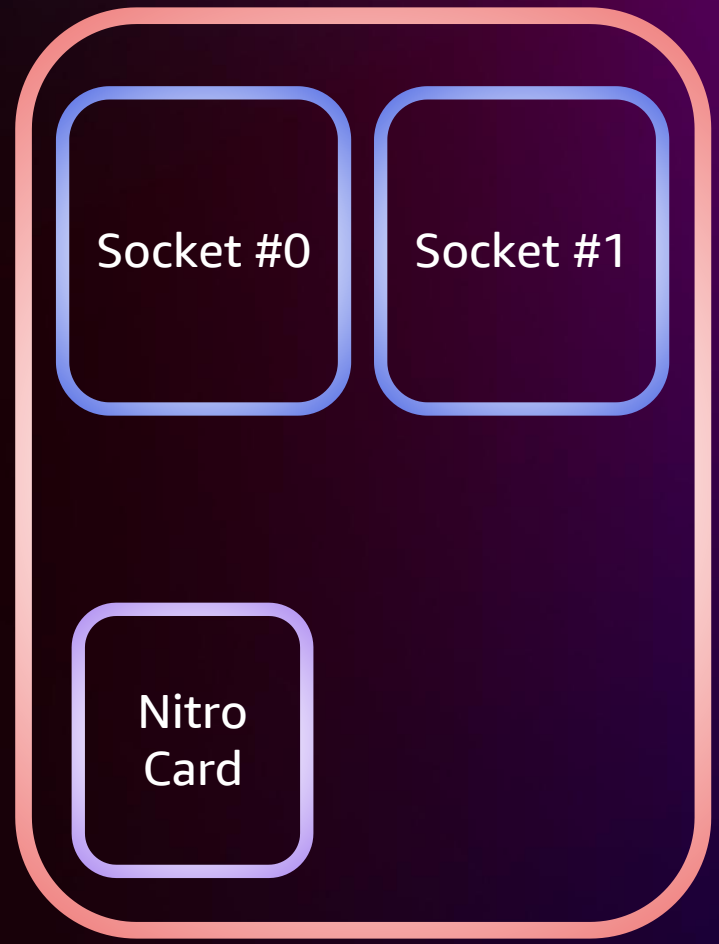


aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Graviton servers

Design the chip, package, and motherboard to fit …



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

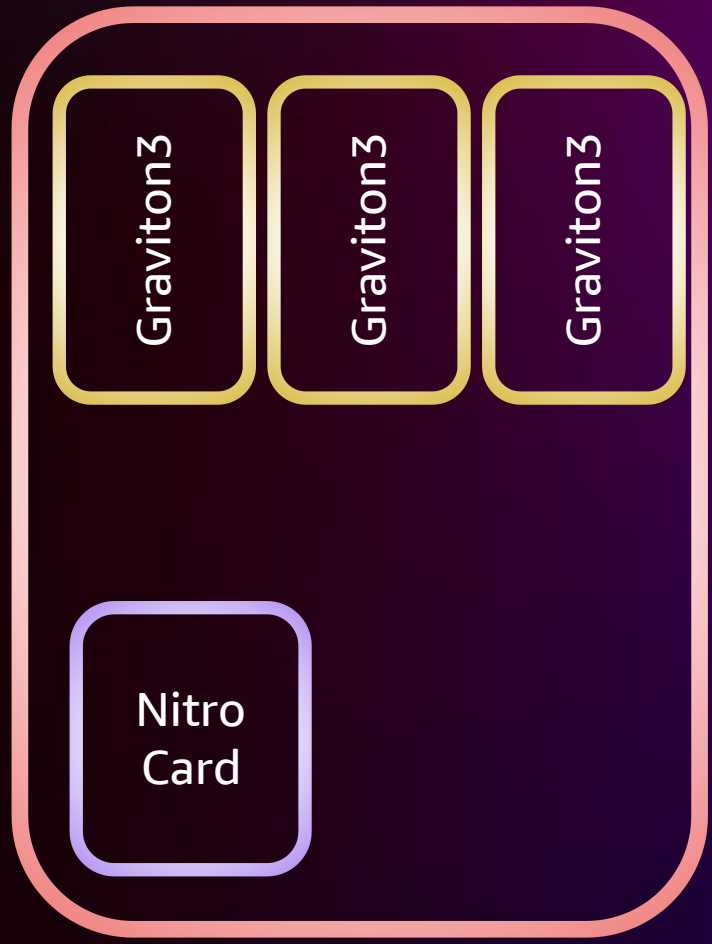
Graviton servers

Design the chip, package, and motherboard to fit three sockets per server

Increase the sockets/rack 50%

Come closer to the power requirements of other vendor solutions

Nitro Card is capable of managing three sockets simultaneously



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Graviton servers

Design the chip, package, and motherboard to fit three sockets per server

Incr

Com
vend

Nitr
simu

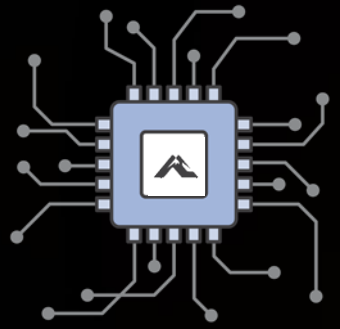


https://www.aboutamazon.com/news/aws/take-a-look-inside-the-lab-where-aws-makes-custom-chips

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

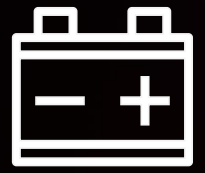
Graviton3 based seventh-generation EC2 instances

Graviton3



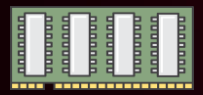
Industry-leading performance

Efficiency



Use up to 60% less energy for the same performance than comparable EC2 instances

DDR5



First DDR5 system in Amazon EC2

Best price performance



Best price performance in Amazon EC2

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton based Amazon EC2 instances

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

M6g, M6gd

General purpose workloads

T4g

Burstable general purpose workloads

C6g, C6gd, C6gn, C7g

Compute and network-intensive workloads

R6g, R6gd, X2gd

Memory-intensive workloads

Im4gn, Is4gen

Storage-intensive workloads

G5g

GPU-based graphics and machine learning workloads

AVAILABLE ACROSS 25+ AWS REGIONS GLOBALLY

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS managed services supporting Graviton

EASY TO GET STARTED WITH GRAVITON PRICE PERFORMANCE BENEFITS

Databases



Amazon DocumentDB



Amazon Aurora



Amazon RDS



Amazon ElastiCache



Amazon MemoryDB



Amazon Neptune

Machine Learning

NEW!



Amazon SageMaker

Analytics



Amazon OpenSearch Service



Amazon EMR

Compute and Containers



AWS Lambda



AWS Fargate



AWS Elastic Beanstalk



Amazon Elastic Kubernetes Service (EKS)



Amazon Elastic Container Service (ECS)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

https://github.com/aws/aws-graviton-getting-started/blob/main/managed_services.md



Amazon is the largest corporate buyer of renewable energy globally

---

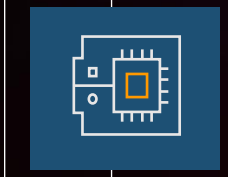
More than 379 projects around the world, on the path to power operations with 100% renewable energy by 2025

Once online, enough power for more than 4.6 million U.S. homes/year or 13.4 million European homes/year

AWS journey highlights



Water stewardship: evaporative cooling, recycled water, and onsite water treatment



Graviton processors provide better performance per watt than any other AWS processor



New data center construction incorporates use of low-carbon concrete



Renewable energy to power data centers



Launch Well-Architected Sustainability Pillar
Launch AWS customer carbon footprint tool



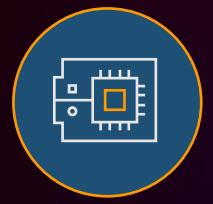
Amazon Sustainability Data Initiative (ASDI) provides free access to satellite data and climate models













aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Data center water stewardship

AWS utilizes real-time sensor data to adapt to changing weather conditions

---

AWS was the first data center operator in northern Virginia approved to use reclaimed water with direct evaporative cooling technology

---

In one US region, AWS partnered with local utilities to build infrastructure enabling reuse of 96% wastewater we discharge from our data centers

Further and faster, together



Commitment to a sustainable future

On September 19, 2019, Amazon and Global Optimism announced The Climate Pledge, a commitment to meet the Paris Agreement 10 years early

Net zero carbon by 2040

Path to 100% renewable energy by 2025

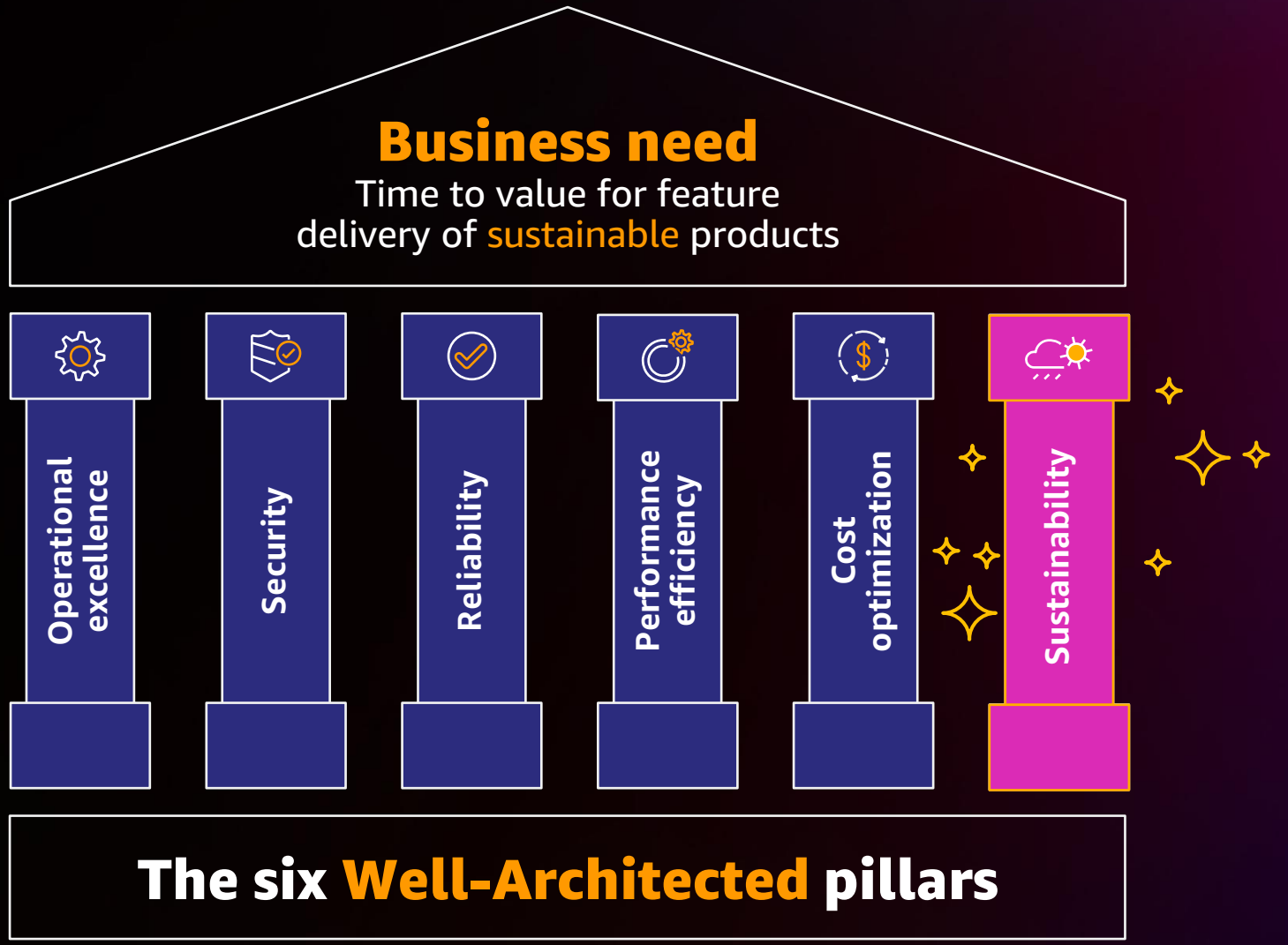
$2 billion Climate Pledge Fund

Over 375 Signatories

aws

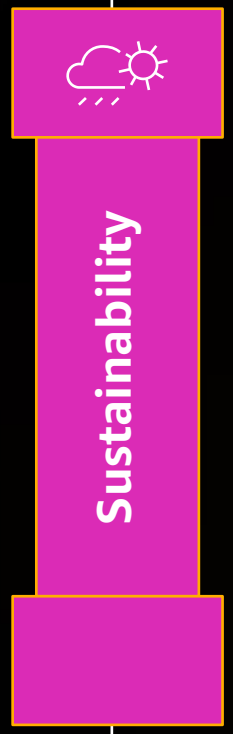
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Well-Architected



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Well-Architected for sustainability



Understand environmental impacts of the services used

Quantify impacts through the entire workload lifecycle

Apply best practices to reduce these impacts

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

re:Set



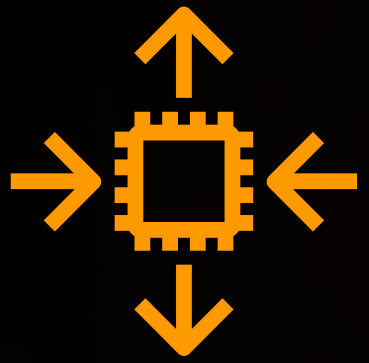
Reduce



Reuse



Recycle







© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Additional Resources



Sustainability Pillar - AWS Well-Architected Framework



re:Invent 2022 - Attendee Guide: Sustainability



AWS Graviton Technical Guide



re:Invent 2022 - Sustainability



AWS Graviton webpage

re:Invent 2022 - Sessions: search “Graviton”

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Graviton ease of adoption

As a rule, the more current your software stack the better

Difficulty	Workload	Actions
Virtually no effort	Amazon RDS, Aurora, ElastiCache, 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, Shared Objects 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 containerized Watch: JNI, Shared Objects or Python-native modules
More involved	Linux – Compiled languages (e.g., C/C++, Python, Go)	Select Arm64 AMI and compile Watch: port any intrinsics, assembly, or Python-native modules
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

aws © 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Paul Moran (he/him)
@pdmoran_aws

Zi Shen Lim (he/him)
@zishenlim



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.