AWS re:Invent

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

CMP201

What's New in Amazon EC2

Adam Boeglin
Principal Specialist
Amazon Web Services

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

Everything New in EC2

- New instance types & capabilities
- New Intel, AMD, and AWS processors
- Hybrid & Edge computing updates
- EC2 Hypervisor Updates
- AWS-Designed Custom Silicon

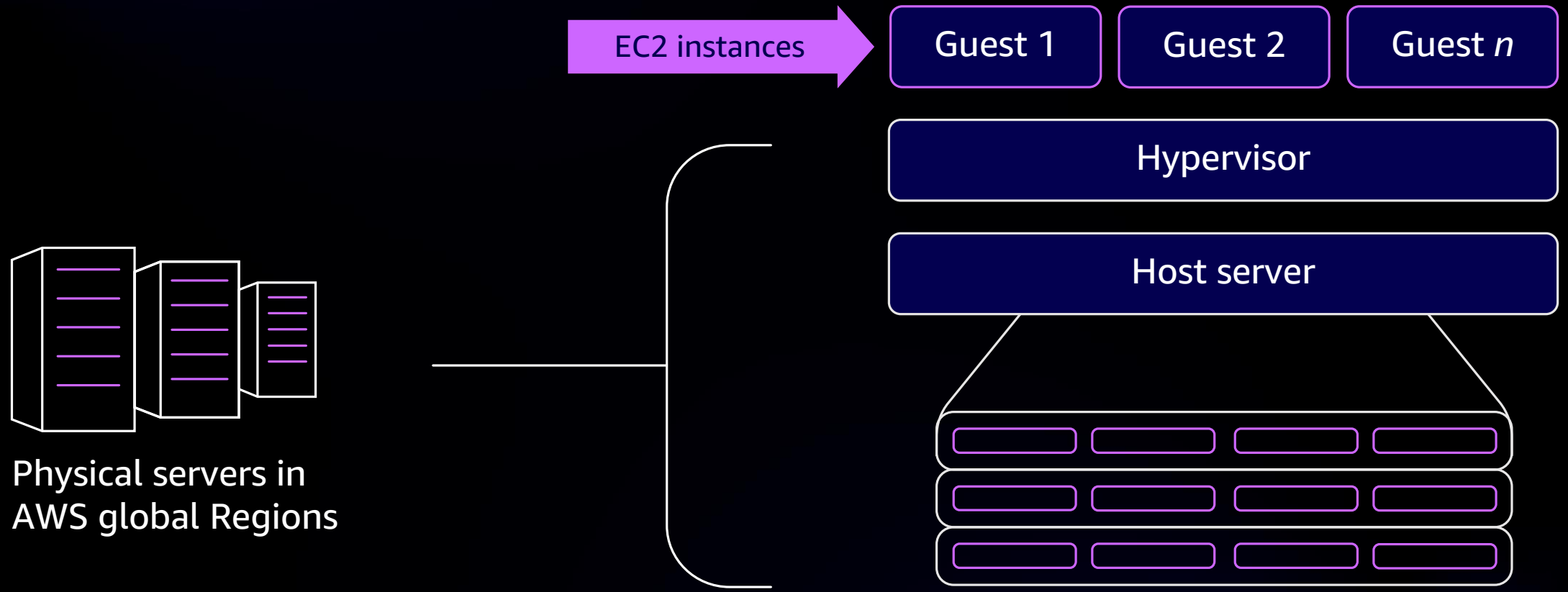
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Adam's Top 8 EC2 Announcements of 2021

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

Amazon Elastic Compute Cloud (Amazon EC2)

VIRTUAL SERVERS IN THE CLOUD



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

Instance naming & configuration

family

c6gn.8xlarge

generation | capabilities | size

Categories

- Burstable (t3, t4)
- General purpose (m5, m6)
- Compute Optimized (c5, c6)
- Memory Optimized (r5, r6, x1, x2)
- Storage Optimized (d3, i3, i4, im4, is4)
- Accelerated (p4, g4, f1, inf1)
- Mac (mac1)

+

Capabilities

- Choice of processor: Intel (i), AMD (a), AWS Graviton (g)
- High Frequency: up to 4.5+ GHz (z)
- Extra Memory: up to 16 TiB (e)
- Instance storage: Local NVMe disks (d)
- High Networking: up to 400 Gbps (n)
- Enhanced EBS Performance: up to 60Gbps & 260,000 IOPS (b)

+

Instance Size

- Size (.nano to .32xlarge)
- Bare Metal (.metal)

+

Options

- Linux, Windows, macOS
- Amazon EBS
- Amazon Elastic Inference
- Amazon Elastic GPU
- Elastic Fabric Adapter

=

Nearly 500

INSTANCE CONFIGURATIONS

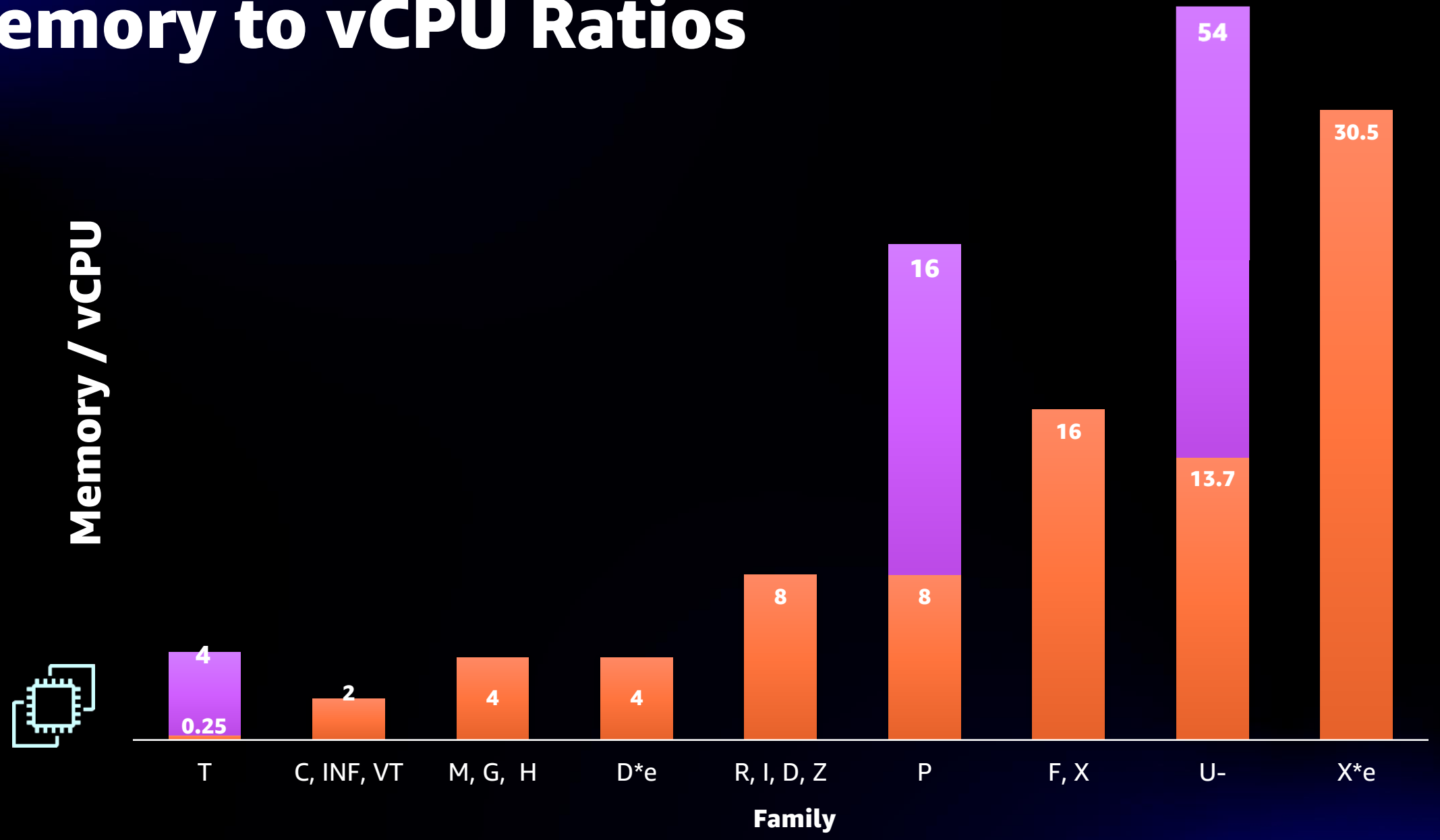
for virtually every business need

examples: c6gn.8xlarge, m5zn.metal, g4ad.16xlarge, r5b.large

aws

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

Memory to vCPU Ratios



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

#8 – New AWS Outposts, AWS Local Zones, & AWS Wavelength capabilities

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

Compute where you need it

	AWS Outposts	AWS Local Zones	AWS Wavelength
			
Overview	AWS infrastructure and services on premises	AWS infrastructure and services in large metro centers	AWS infrastructure and services in CSP 5G networks
Use cases	Migration, local critical apps, data residency	Migration, low latency, local data processing	Ultra-low latency, local data processing
Service model	Expandable capacity in customer's DC, collocation, on-premises location	Scalable capacity in AWS managed and operated facility	Scalable capacity in CSP DC managed and supported by AWS

aws

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

Supported in 23 Regions

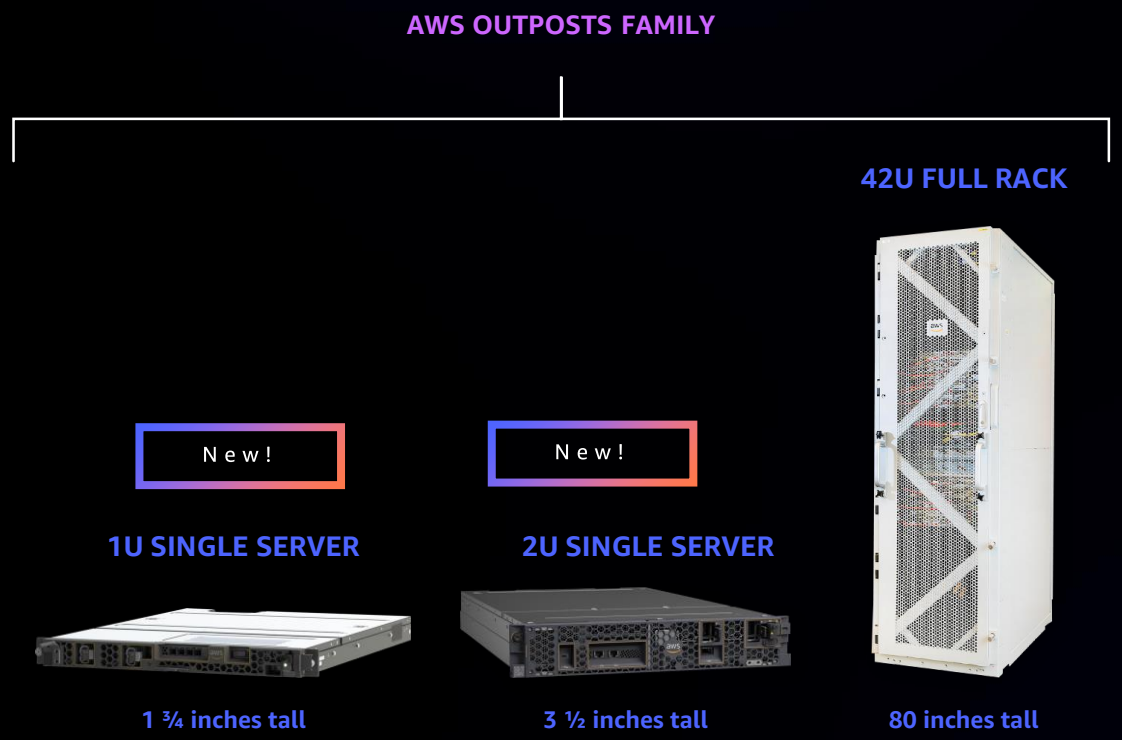


aws

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

Introducing: AWS Outposts servers

RUN AWS OUTPOSTS IN LOCATIONS WITH LIMITED SPACE OR CAPACITY REQUIREMENTS



Offers the same AWS infrastructure, services, APIs, and tools on-premises, now with a smaller form factor

Choose between a 1U form factor with an AWS Graviton2 processor or a 2U form factor with an Intel processor

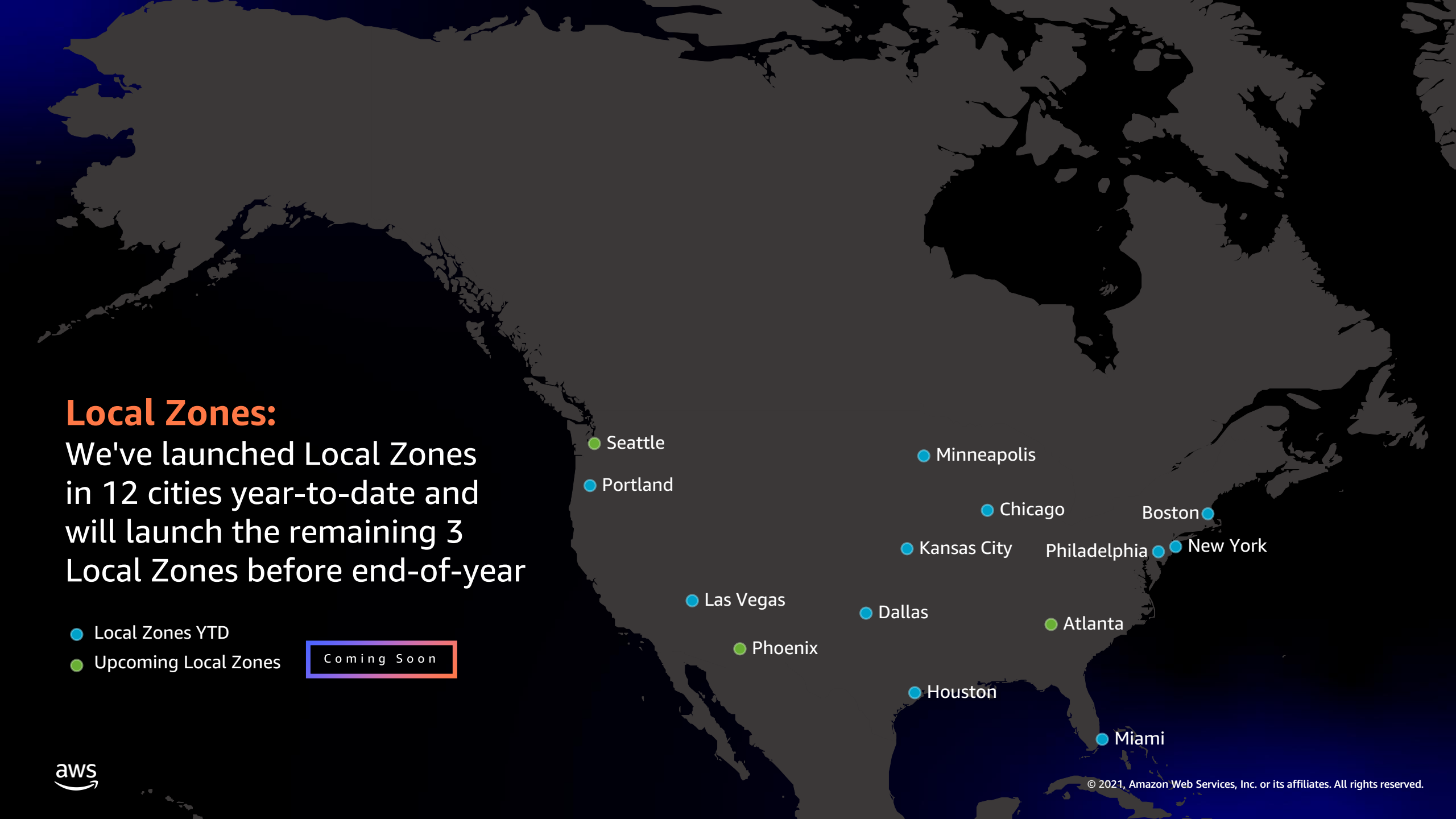
Run AWS services locally, like EC2, and edge services like AWS IoT Greengrass

Ideal for workloads that require low-latency and local processing needs

Simple device installation by either your own on-premises personnel or an AWS preferred third-party vendor

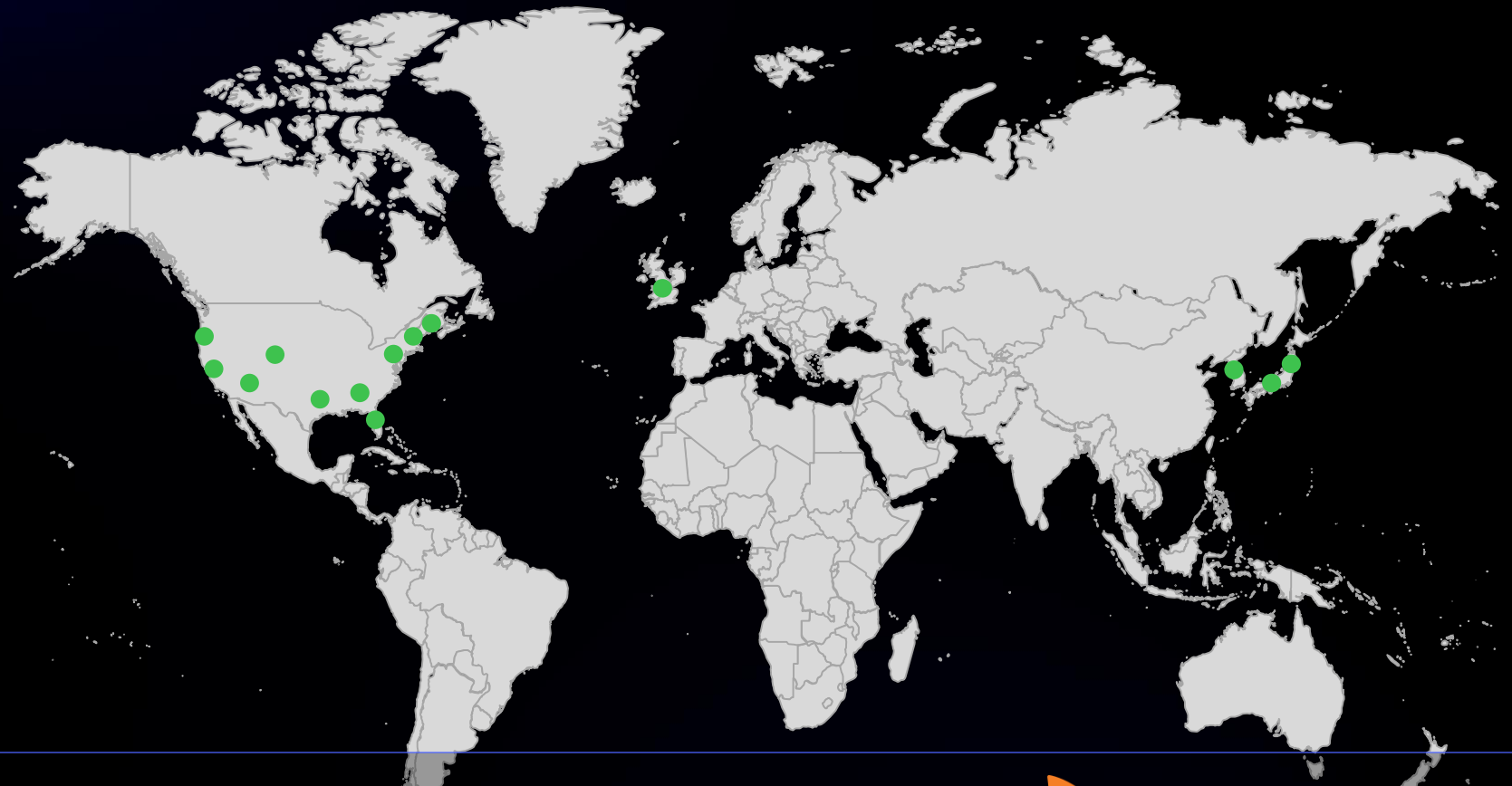
aws

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



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

AWS Wavelength global partnerships



5G partners

Bell

Coming Soon

verizon

KDDI

SK telecom

vodafone business

aws

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

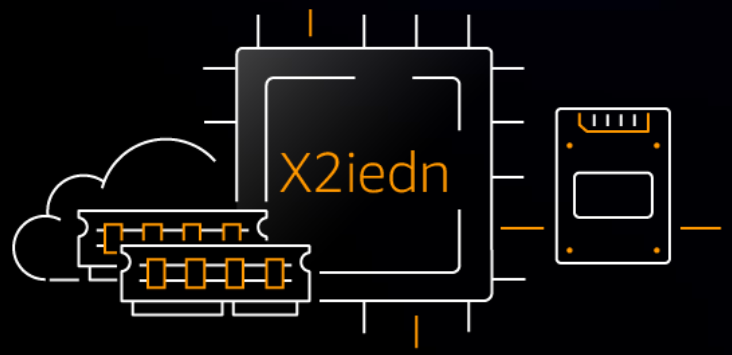
#7 – New Memory Optimized Instances

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

Amazon EC2 X2idn and X2iedn instances

X2idn/X2iedn instances

Coming Soon



Next gen memory optimized instances powered by 3rd gen Xeon Scalable processors

Up to 50% better price performance than previous generation instances for memory intensive workloads

Up to 75% faster memory to increase performance of memory intensive workloads

Up to 65% higher performance for SAP applications

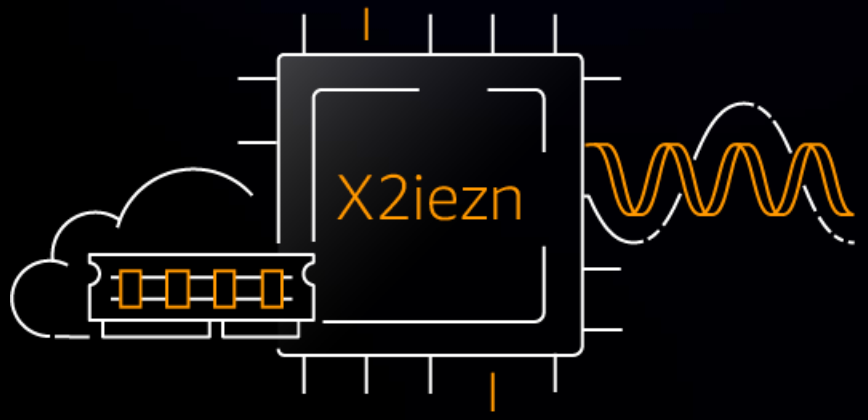
aws

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

Amazon EC2 X2iezn instances

X2iezn instances

Coming Soon



High frequency processing for electronic design automation

Up to 4.5 GHz all core turbo frequency, the fastest Intel Xeon Scalable processor in the cloud

2x more performance per vCPU compared to X1e to reduce license costs by up to 50% by consolidating workloads

Up to 55% better price performance and up to 40% faster runtimes compared to X1e instances

aws

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

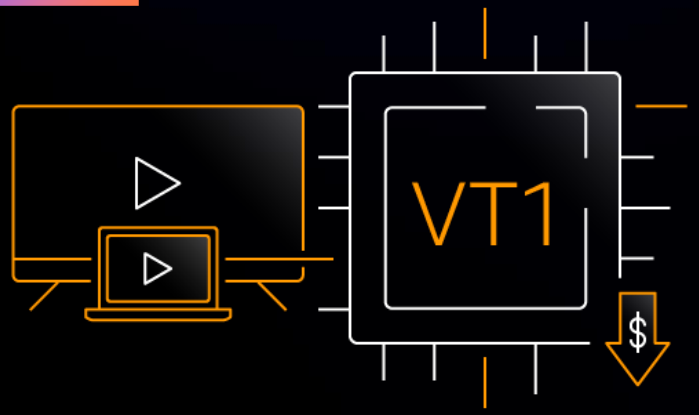
#6 – New Graphics Optimized Instances

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

Amazon EC2 VT1 instances

VT1 instances

New!



Best price per stream in the cloud for live video transcoding

Up to 30% better price/stream than comparable instances

Features Xilinx U30 cards for live video transcoding: H.264 & H.265 - integration with FFmpeg & GStreamer

Scale on demand in multiple EC2 Regions globally

On-premises and in-country deployment with AWS Outposts

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

Amazon EC2 G5 instances

G5 instances

New!



High performance GPU-based instances for graphics-intensive applications and ML inference

Up to 3x higher graphics performance and up to 40% better price performance than G4dn instances

Up to 3x higher performance and up to 40% better price performance for machine learning inference compared to G4dn instances

Up to 15% lower cost-to-train than Amazon EC2 P3 instances. They also deliver up to 3.3x higher performance for ML training compared to G4dn instances.

aws

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

Amazon EC2 G5g instances

G5g instances

New!



Most cost-effective GPU-based platform for graphics applications such as Android cloud gaming

Up to 30% lower cost per stream per hour for Android game streaming

First Arm-based instances in a major cloud to feature GPU acceleration, powered by AWS Graviton2 processors and NVIDIA Tesla T4 GPUs

Run Android games natively without need for Arm-based cross compilation or emulation

aws

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

#5 – New Machine Learning Instances

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

Supporting ML workloads

Largest portfolio of ML instances in the cloud

TRAINING



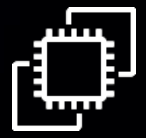
DL1

HABANA GAUDI



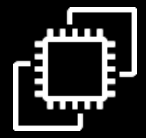
G5

NVIDIA GPU



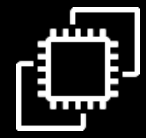
R5

MEMORY OPTIMIZED



P3

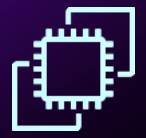
NVIDIA V100 TENSOR CORE



P4D

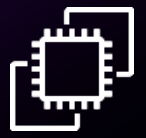
NVIDIA A100 TENSOR CORE

INFERENCE



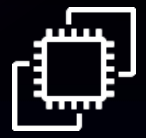
G5G

AWS GRAVITON 2



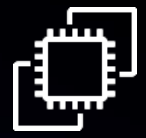
G5

NVIDIA GPU



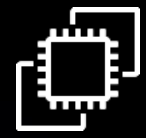
INF1

AWS INFERENTIA



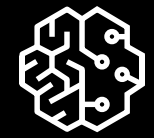
C5

COMPUTE OPTIMIZED



G4

NVIDI T4 GPU



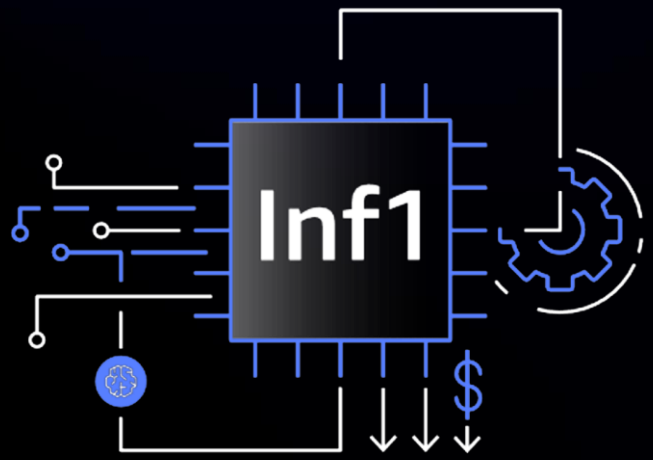
EI

ELASTIC INFERENCE

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

Amazon EC2 Inf1 instances

Inf1 instances



High performance and the lowest cost ML inference in the cloud

Up to 2.3x higher throughput and up to 70% lower cost-per-inference than GPU instance NEW!

Powered by AWS Inferentia and AWS Neuron SDK

Integrated with popular ML frameworks TensorFlow, PyTorch, and MXNet. New support for HuggingFace model repository. NEW!

Many Amazon services utilize Inf1 instances, including Amazon advertising, Amazon Alexa, Amazon Rekognition, and Amazon robotics

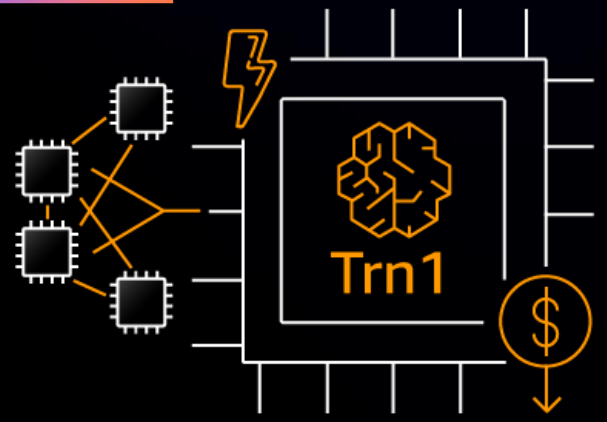
aws

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

Amazon EC2 Trn1 Instances

Trn1 Instances

Preview



Highest performance and lowest cost instance for training deep learning models

Best price performance in the cloud for ML learning training. Lower time-to-train when compared to the latest EC2 GPU-based instances

Will leverage Neuron SDK that is integrated with popular ML frameworks such as TensorFlow and PyTorch

Deployable in EC2 UltraClusters, a super computer in the cloud with tens of thousands of Trainium accelerators with petabit scale networking bandwidth

aws

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

#4 – Amazon Linux 2022 & EC2 Serial Console

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

Introducing Amazon Linux 2022 (AL2022)

Amazon Linux 2022

Preview



Next generation of Amazon Linux from AWS, optimized for the cloud, with a focus on stability, security, and performance

New Amazon Linux major versions will be available every 2 years and each major version, including AL2022, will come with 5 years of LTS

Based on Fedora, AL2022 provides frequent and flexible quarterly updates

SELinux enabled and enforced by default to help you meet your compliance needs

Version locking feature gives you control over how and when you absorb updates

aws

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

Introducing Amazon EC2 serial console

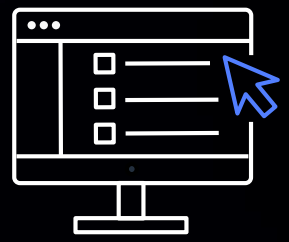
NEW!

TROUBLESHOOT BOOT AND CONNECTIVITY ISSUES WITH A SINGLE CLICK



SECURE

Secure and dedicated access governed by IAM and AWS Organizations policies



AWS CLI, API & CONSOLE

Connect to the serial port of your Linux or Windows machine



TROUBLESHOOT

Attempt grub rescue, fix misconfigured firewall, iptables, disk failures, install missing drivers

aws

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

#3 – Nitro System Updates

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

Journey from then to now

2006 “Instance”

1.7 GHz Xeon processor

1.75 GB of RAM

160 GB of local disk

250 Mbps network bandwidth

2021

4.5 GHz Xeon processor

M5zn instance

24 TiB of RAM

High Memory instances

60 TB of NVMe local storage

I3en.metal instances

336 TB of local disk

D3en instances

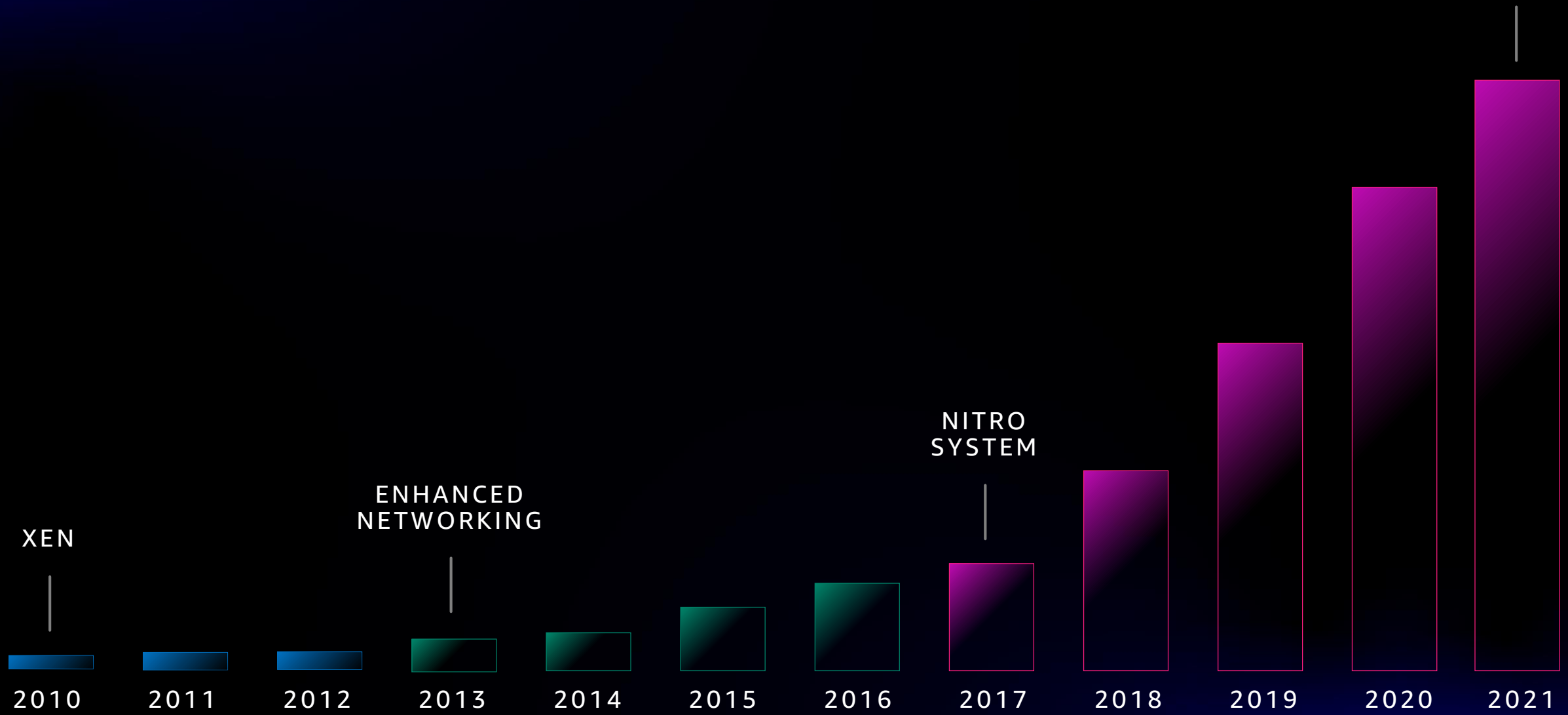
400 Gbps network bandwidth

CPUs, GPUs, FPGAs, Accelerators

aws

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

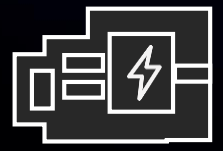
Pace of EC2 Instance Innovation



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

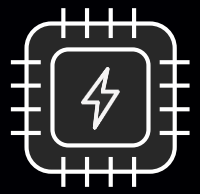
Nitro is the engine that powers EC2

Nitro Card



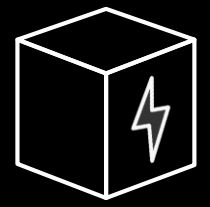
Local NVMe storage
Amazon Elastic Block Storage
Networking, monitoring, and security

Nitro Security Chip



Integrated into motherboard
Protects hardware resources

Nitro Hypervisor

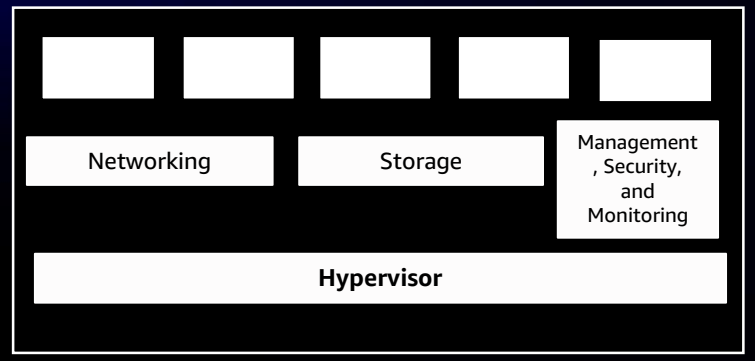


Lightweight hypervisor
Memory and CPU allocation
Bare Metal-like performance

Modular building blocks for rapid design and delivery of EC2 instances

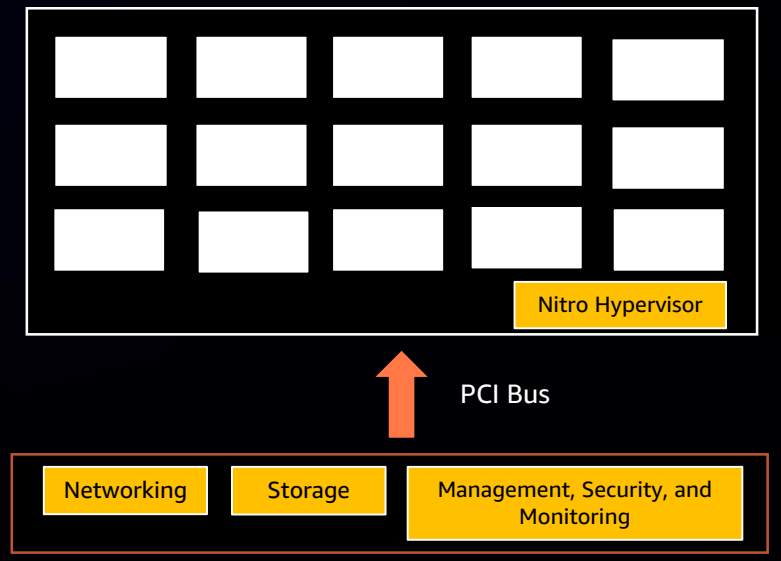
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Reinventing virtualization for the cloud



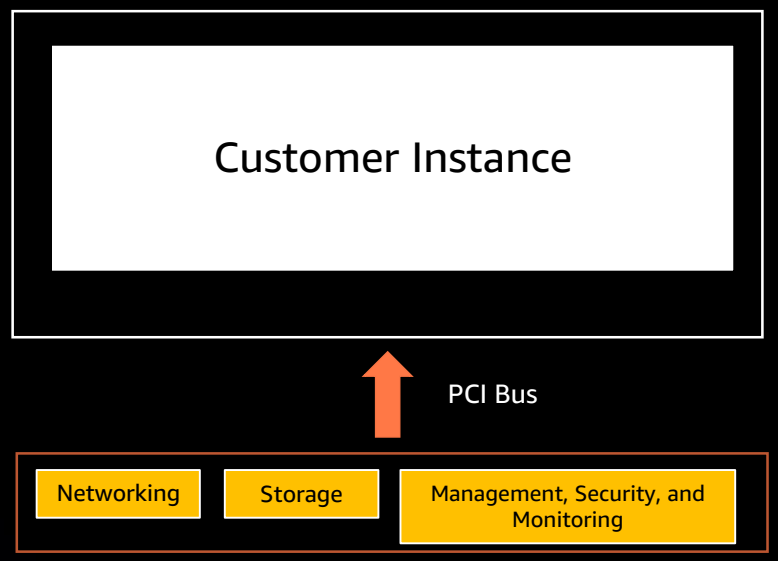
Classical virtualization

- Large Trusted computing base (TCB) with networking, storage, and monitoring are all running on the same hardware that runs customers' VMs.
- Reduce performance and resources for VMs



EC2 virtual machines

- Trusted computing base significantly reduced.
- All networking, storage, and monitoring runs on a separate, isolated, and secured hardware.
- Prohibits all administrative access, including those of Amazon employees.
- Only a thin hypervisor on the host.



EC2 bare metal instances

- All benefits of Nitro with the hypervisor removed.
- Customers can deploy applications that use physical hardware resources directly on AWS infrastructure.

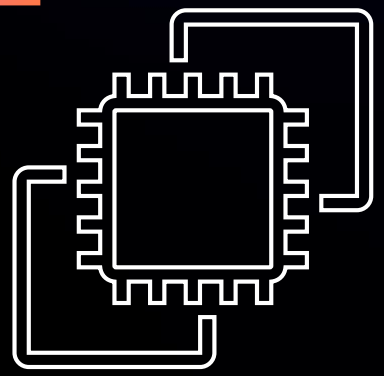
aws

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

Nitro support for previous generation instances

Additional support for early generation EC2 instances

Coming Soon



Support for Xen-based EC2 instances on AWS Nitro System

Run your workloads on the instances they were built on for years to come

Nitro Security benefits now for previous generation instances

Seamless transition to the AWS Nitro System with instance specifications, AMIs, and workflows remaining the same

aws

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

#2 – Intel & AMD Instances with Next-Gen Processors

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

intel®

15 years of collaboration and innovation

KEY HIGHLIGHTS

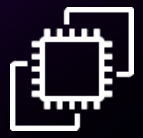
- Deep engineering collaboration across AWS portfolio
- Over 275 EC2 instances are powered by Intel Xeon processors
- Fastest processor in the cloud
- First cloud to launch Ice Lake

RECENT AMAZON EC2 INSTANCES



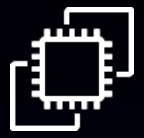
I4I

STORAGE OPTMIZED



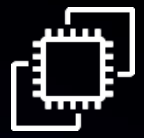
C6I

COMPUTE OPTIMIZED



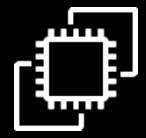
R6I

MEMORY OPTIMIZED



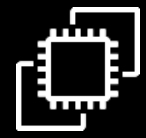
M6I

GENERAL PURPOSE



X2I(E)DN

MEMORY OPTIMIZED



X2IEZN

HIGH FREQUENCY

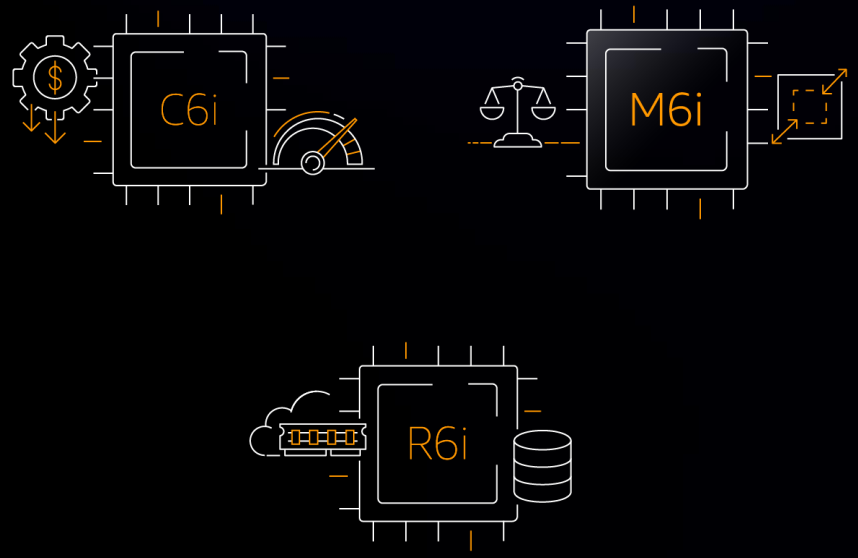
aws

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

Amazon EC2 C6i, M6i, and R6i instances

NEW!

C6i, M6i, and R6i Instances



EC2 instances powered by third-generation Intel Xeon Scalable processors

Deliver up to 15% better price performance compared to previous gen C5, M5, and R5 instances

New larger sizes with up to 128 vCPUs and 512 GiB of memory, enabling you to consolidate workloads on fewer instances

Faster networking and higher memory bandwidth

Support for Total Memory Encryption (TME)

aws

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

AMD

10% cost savings for x86 workloads

AMD EPYC CPUs



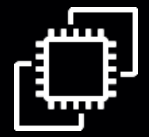
C5A(D)

COMPUTE OPTIMIZED



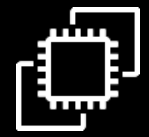
M5A(D)

GENERAL PURPOSE



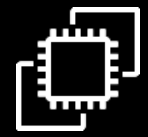
R5A(D)

MEMORY OPTIMIZED



T3A

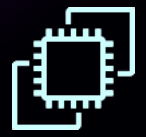
BURSTABLE



M6A

GENERAL PURPOSE

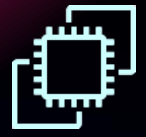
COMING SOON



HPC6A

HPC OPTIMIZED

RADEON PRO GPUs



G4AD

GRAPHICS OPTIMIZED

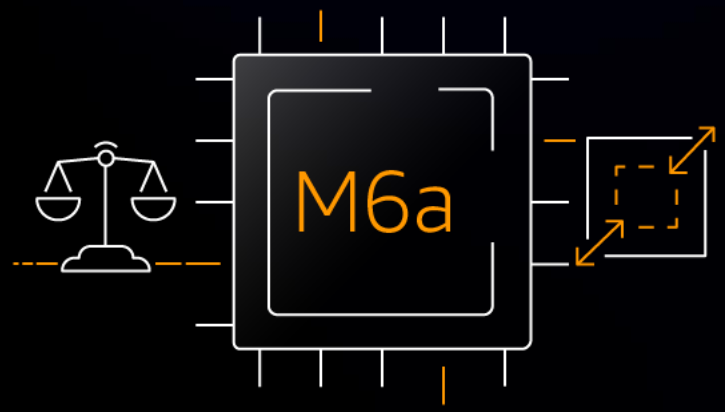
aws

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

Amazon EC2 M6a instances

M6a instances

New!



General purpose instances powered by 3rd gen AMD EPYC processors

Up to 35% better price performance compared to M5a instances and 10% lower cost than comparable x86-based EC2 instances

First EC2 instance with 3rd gen AMD EPYC processors with an all-core turbo frequency of 3.6 GHz. Support for always-on memory encryption

Ideal for enterprise applications, multi-player gaming servers, caching fleets, and app dev environments

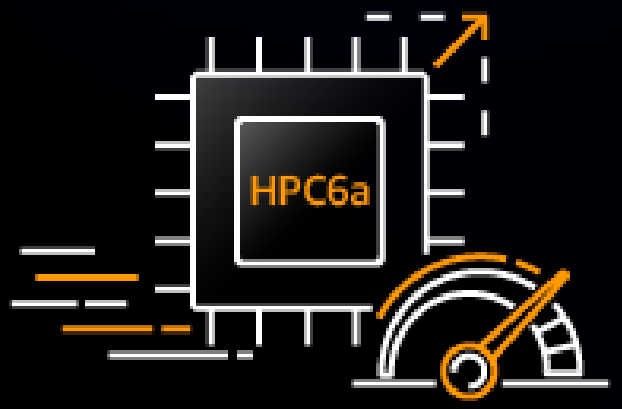
aws

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

Amazon EC2 Hpc6a instances

HPC6a instances

Coming Soon



Best price performance for compute-intensive, HPC workloads in Amazon EC2

Up to 65% better price performance over comparable Amazon EC2 x86 based, compute-optimized instances

3rd Gen AMD EPYC Milan processors with up to 96 cores, up to 3.6 GHz frequency, and 384GB of RAM

Optimized for HPC workloads with Elastic Fabric Adapter enabled by default for 100 Gbps networking

aws

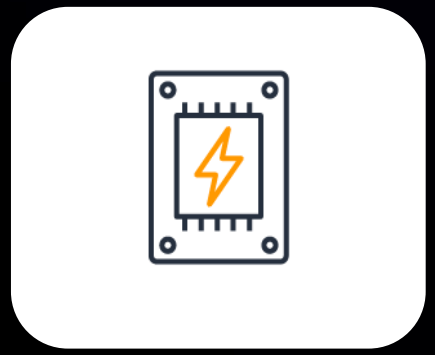
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

#1 – New AWS-designed Custom Silicon

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

AWS Nitro SSD

NEW!



High performance, low latency SSD custom designed by AWS

AWS Nitro SSD

Tightly integrated with the AWS Nitro System to provide 60% lower I/O latency compared to commercial SSDs

Quicker firmware updates to improve reliability without any downtime to the instance

All data stored on the disks is encrypted at rest with AES-256 ephemeral keys

aws

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

Amazon EC2 I4i instances

I4i instances

Coming Soon



Highest x86-based local storage performance within Amazon EC2

Powered by 3rd generation Intel Xeon Scalable processors and featuring AWS Nitro SSDs

Up to 55% better compute price performance and 60% lower storage I/O latency than I3 instances

Ideal for Microsoft SQL Server, Oracle databases, NoSQL databases, and data analytics

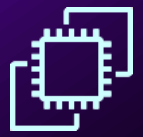
aws

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

Graviton2

40% better price performance for broad range of workloads

RECENT AMAZON EC2 INSTANCES



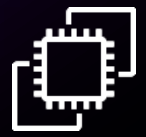
C6G(D)

COMPUTE OPTIMIZED



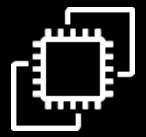
M6G(D)

GENERAL PURPOSE



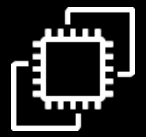
R6G(D)

MEMORY OPTIMIZED



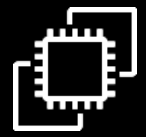
C6GN

ENHANCED NETWORKING



X2GD

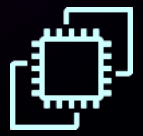
ENHANCED MEMORY



T4G

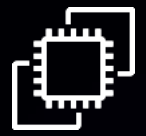
BURSTABLE

FREE TRIAL



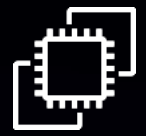
IM4GN

STORAGE OPTIMIZED



IS4GEN

STORAGE OPTIMIZED



G5G

ACCELERATED COMPUTING

aws

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

The path to Graviton2

AWS Graviton processor

Released at re:Invent 2018

First Arm processor in AWS

16nm

~5 Billion transistors

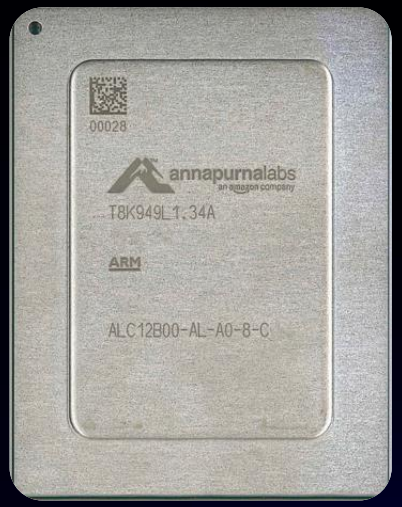


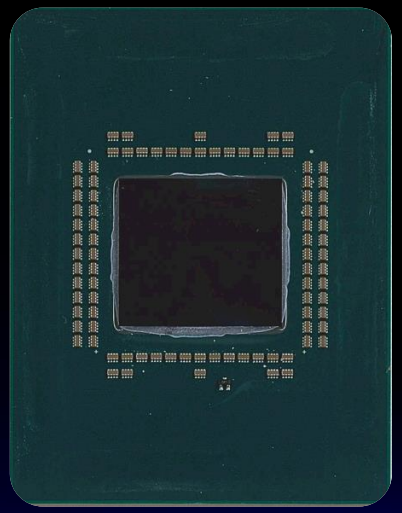
AWS Graviton2 processor

4x vCPUs with **7x** CPU performance

7nm *(First 7nm chip in EC2)*

~30 Billion transistors





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

Customers adopting Graviton2











































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

NEW!

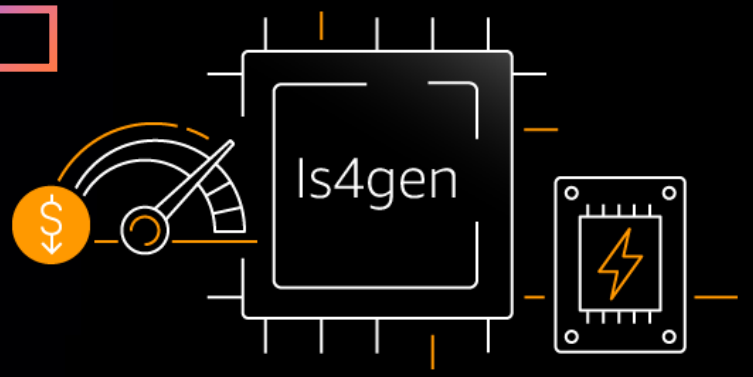


Best Price/Performance

Up to 40% better price performance compared to I3 instances

Ideal for NoSQL databases, data indexing and search, and data analytics

NEW!



Lowest cost per TB

Up to 15% lower cost per TB of storage and up to 48% better compute performance compared to I3en instances

Ideal for stream processing, real-time databases, and log analytics

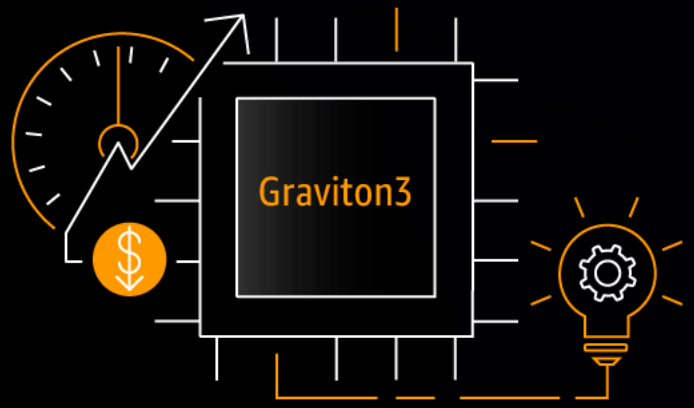
Powered by AWS Graviton2 processors and featuring AWS Nitro SSDs

aws

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

AWS Graviton3

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
- Up to 3x better machine learning performance

First in the cloud to feature DDR5 memory

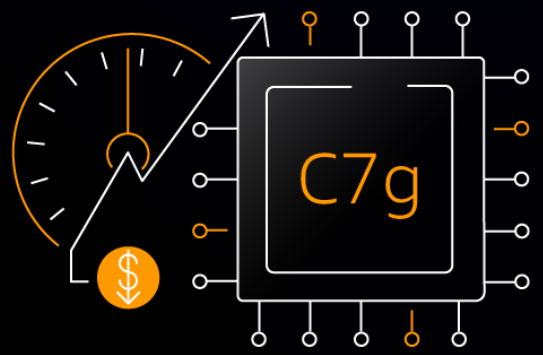
60% more energy efficient over comparable EC2 instances

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

Amazon EC2 C7g instances

C7g Instances

Coming Soon



Best price performance for compute-intensive workloads in Amazon EC2

Up to 25% better compute and 2x higher floating point performance over Graviton2 to accelerate compute workloads

First in the cloud to feature the cutting edge DDR5 memory, that provides 50% more bandwidth compared to DDR4 memory

20% higher enhanced networking bandwidth and 30% higher storage bandwidth compared to C6g instances for network intensive applications, such as network appliances and data analytics

aws

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

Everything new!

- Intel “Ice Lake” & AMD “Milan” instance updates
- New instance types
- Next generation Graviton3 Processor
- AWS Nitro SSDs
- New instance capabilities and variants
- AWS Wavelength, Outposts, and Local Zone updates

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

Thank you!

Adam Boeglin

boeglina@amazon.com

linkedin.com/in/adamboeglin



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