AWS re:Invent

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

CMP218-L

Building on 15 years of compute innovation

Dave Brown (he/him)
Vice President, Amazon EC2

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

1

re:Inventing Compute on a Global Scale

2

Performance, Security, and Cost

3

Accelerated Workloads

4

Compute Where You Need It

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



re:Inventing Compute on a Global Scale

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

2021



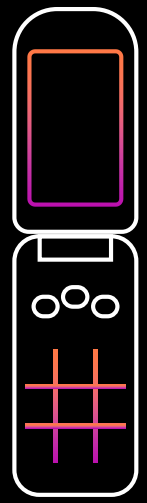
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

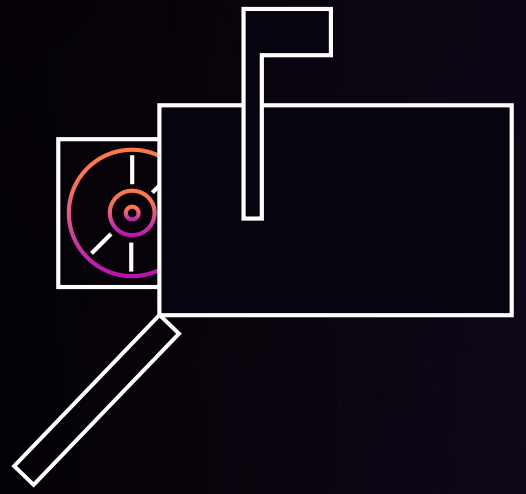
2006



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

2006

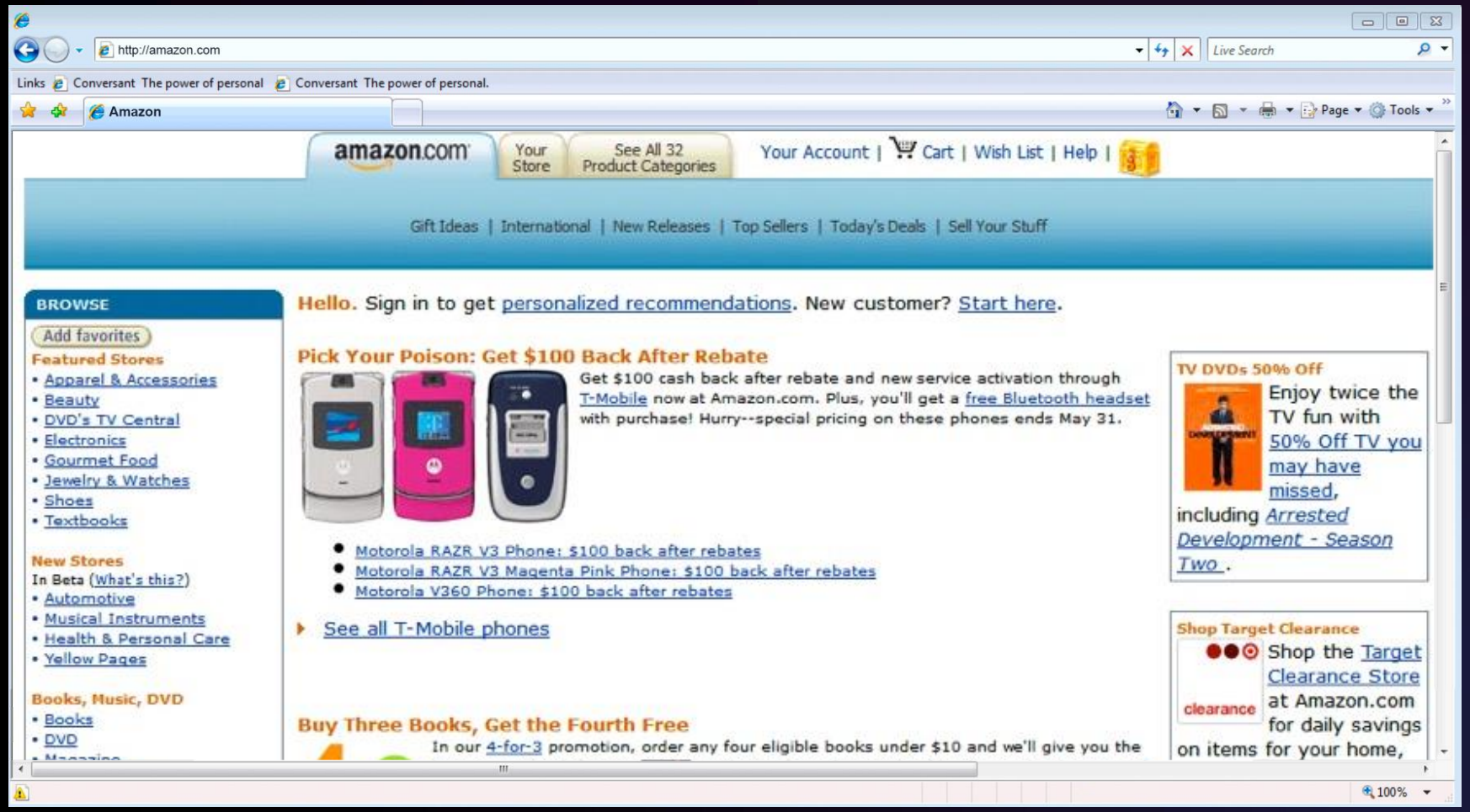




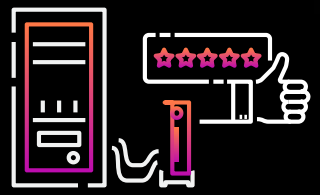


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

2006


http://amazon.com
Live Search
Links Conversant The power of personal Conversant The power of personal.
Amazon
Page Tools
amazon.com
Your Store
See All 32 Product Categories
Your Account | Cart | Wish List | Help |
Gift Ideas | International | New Releases | Top Sellers | Today's Deals | Sell Your Stuff
BROWSE
Add favorites
Featured Stores
• Apparel & Accessories
• Beauty
• DVD's TV Central
• Electronics
• Gourmet Food
• Jewelry & Watches
• Shoes
• Textbooks
New Stores
In Beta (What's this?)
• Automotive
• Musical Instruments
• Health & Personal Care
• Yellow Pages
Books, Music, DVD
• Books
• DVD
Hello. Sign in to get personalized recommendations. New customer? Start here.
Pick Your Poison: Get $100 Back After Rebate
Get $100 cash back after rebate and new service activation through T-Mobile now at Amazon.com. Plus, you'll get a free Bluetooth headset with purchase! Hurry--special pricing on these phones ends May 31.
• Motorola RAZR V3 Phone: $100 back after rebates
• Motorola RAZR V3 Magenta Pink Phone: $100 back after rebates
• Motorola V360 Phone: $100 back after rebates
See all T-Mobile phones
Buy Three Books, Get the Fourth Free
In our 4-for-3 promotion, order any four eligible books under $10 and we'll give you the
TV DVDs 50% Off
Enjoy twice the TV fun with 50% Off TV you may have missed, including Arrested Development - Season Two .
Shop Target Clearance
clearance
Shop the Target Clearance Store at Amazon.com for daily savings on items for your home,
100%

amazon.com

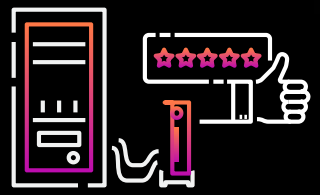


INFRASTRUCTURE
EXCELLENCE

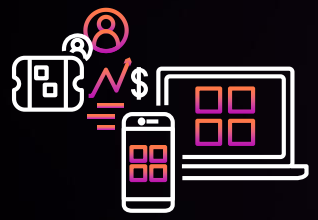
aws

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

amazon.com



INFRASTRUCTURE EXCELLENCE

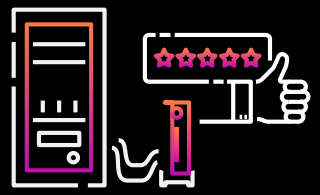


SERVICE-ORIENTED ARCHITECTURE

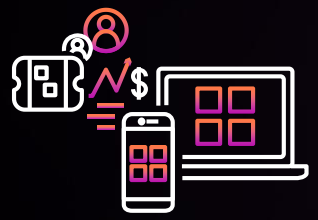
aws

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

amazon.com



INFRASTRUCTURE EXCELLENCE



SERVICE-ORIENTED ARCHITECTURE

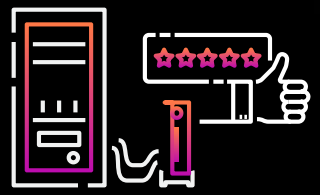


STREAMLINE DEVELOPMENT

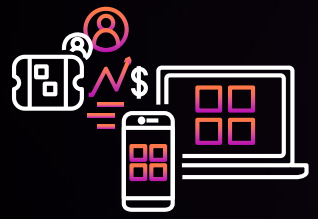
aws

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

amazon.com



INFRASTRUCTURE EXCELLENCE



SERVICE-ORIENTED ARCHITECTURE



STREAMLINE DEVELOPMENT



API ADOPTION

aws

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

Provide the tools and services to allow developers to build any application and run any workload

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

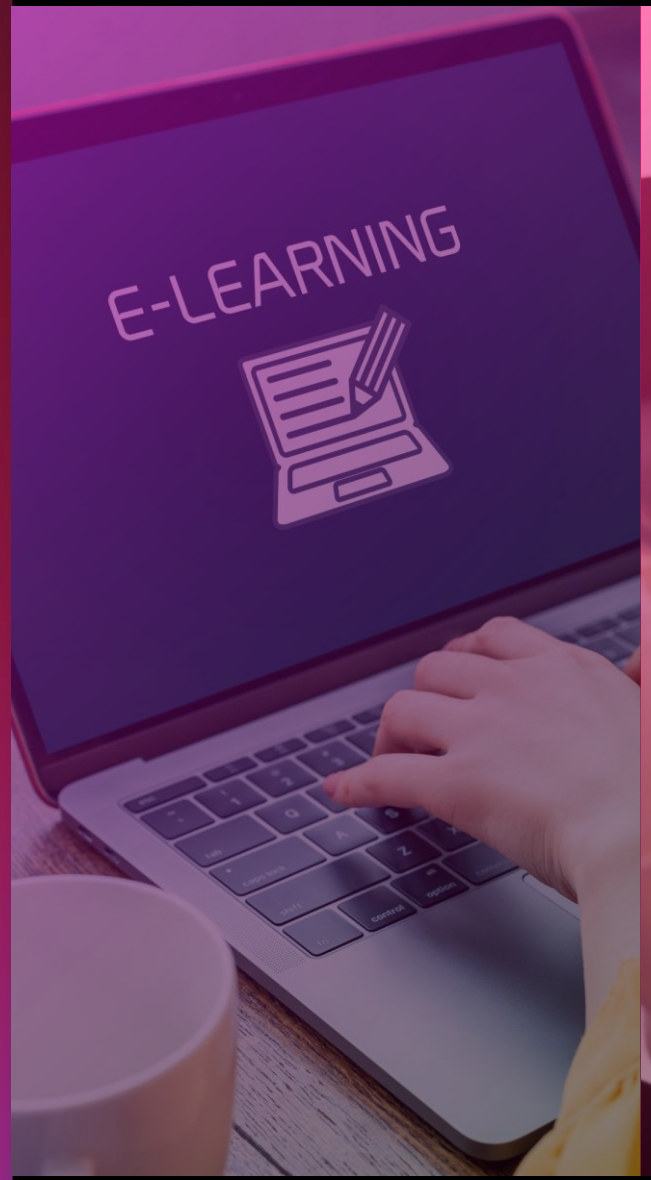


CAPE TOWN, SOUTH AFRICA



Today, over 60 million new instances are spun up every day on Amazon EC2





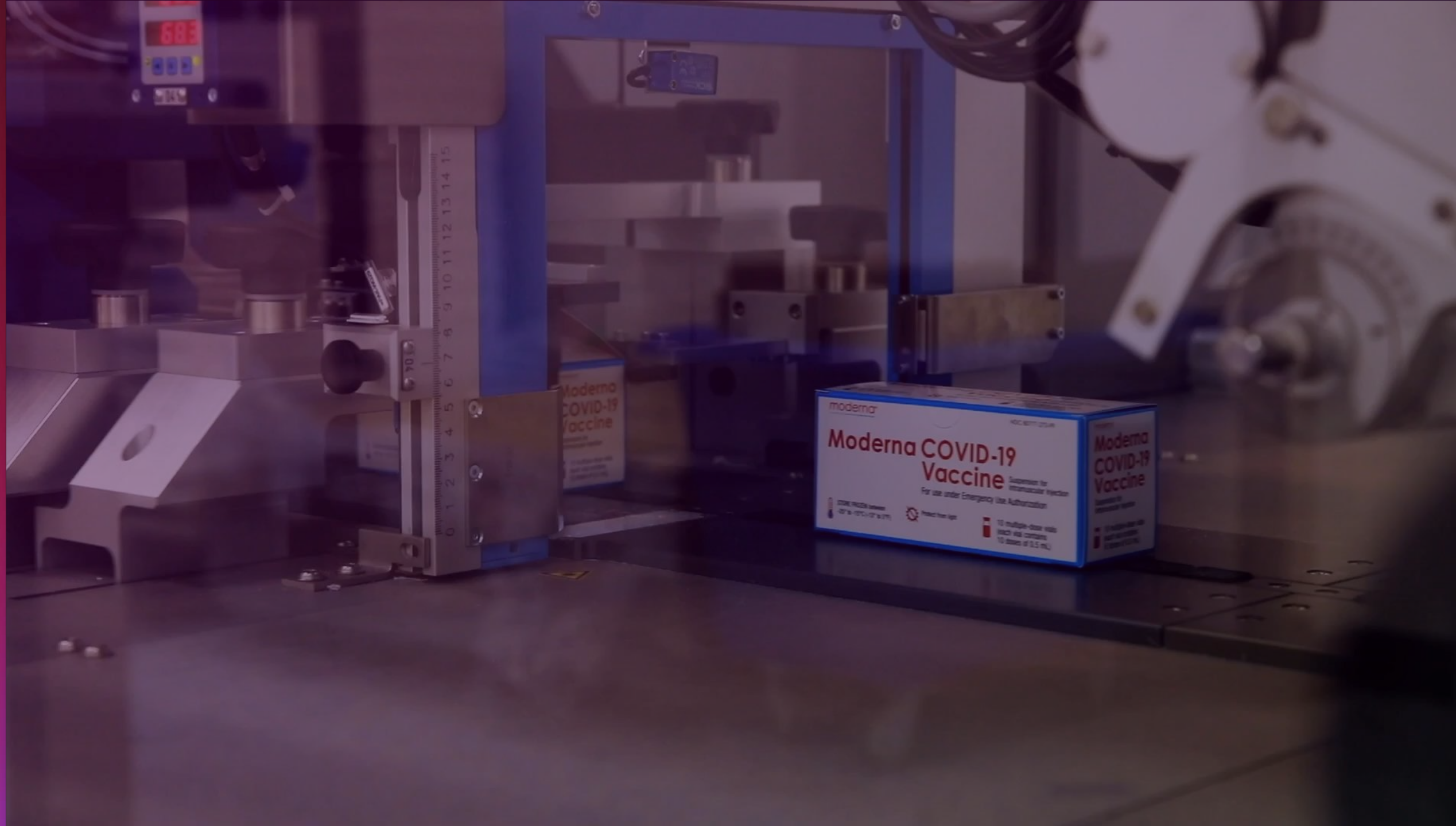




© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



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


moderna
Moderna COVID-19 Vaccine
For use under Emergency Use Authorization
Moderna COVID-19 Vaccine

High Performance Computing

OPTIMIZED FOR VIRTUALLY ANY HPC WORKLOAD WITH THE ELASTICITY TO SCALE



Elastic Fabric Adapter
+100 Gbps Networking



AWS
ParallelCluster

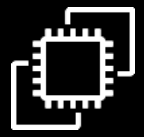


AWS
Batch



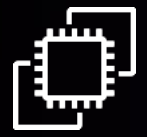
Amazon FSx
for Lustre

HPC OPTIMIZED INSTANCES



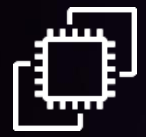
C7G

COMPUTE
OPTIMIZED



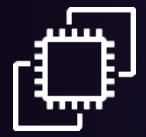
C6I

COMPUTE
OPTIMIZED



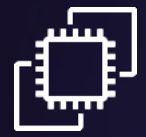
C5A(D)

COMPUTE
OPTIMIZED



C6GN

NETWORK
OPTIMIZED



C5N

NETWORK
OPTIMIZED

aws

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

High Performance Computing is a fast growing space and one that demands continuous innovation

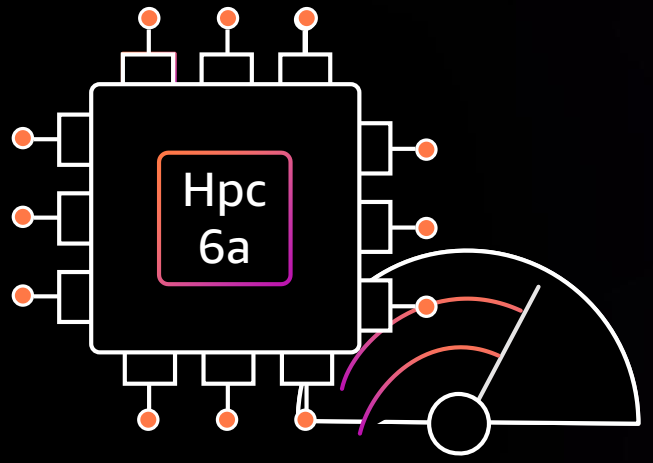
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

COMING SOON

AMD-based HPC Optimized instance

POWERED BY THIRD-GENERATION EPYC MILAN PROCESSOR



Available in select AWS Regions

Up to 96 CPU cores, 384 GiB memory, and 100 Gbps networking

Offer up to 65% better price/performance compared to similar x86, compute optimized instances

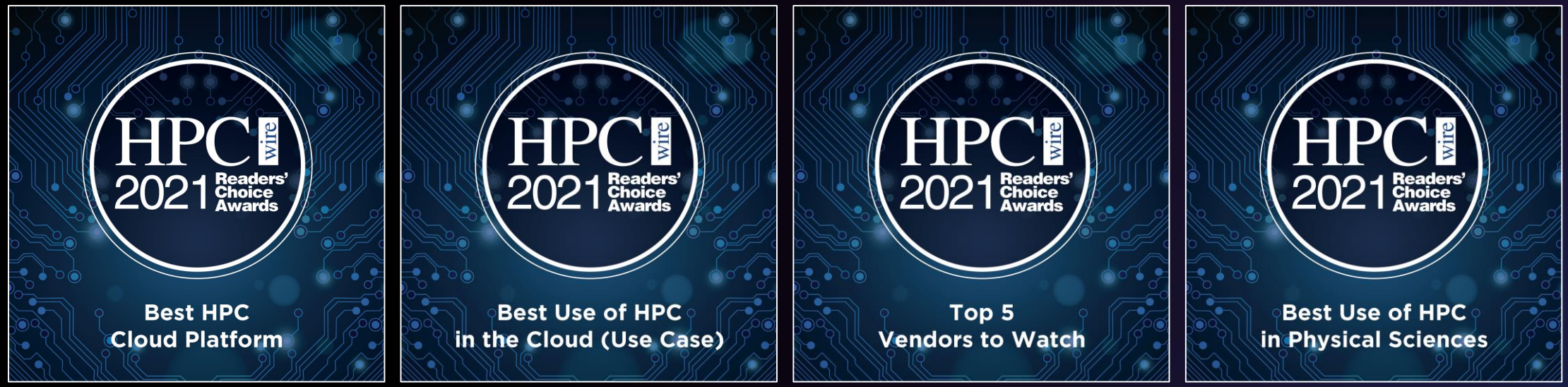
Ideal for large scale computational workloads across weather, financial, fluid dynamics, and healthcare & life sciences

aws

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

Recognized by the HPC community

BEST HPC CLOUD PLATFORM FOR THE 4TH CONSECUTIVE YEAR



aws

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

Recognized by the HPC community

BEST HPC CLOUD PLATFORM FOR THE 4TH CONSECUTIVE YEAR



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

1

re:Inventing Compute on a Global Scale

2

Performance, Security, and Cost

3

Accelerated Workloads

4

Compute Where You Need It

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

2

Performance, Security, and Cost

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

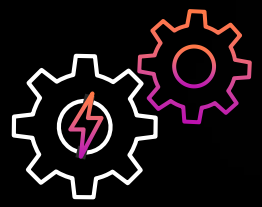
The AWS Nitro System architecture

OFFERING THE BEST SECURITY, PERFORMANCE, AND INNOVATION IN THE CLOUD



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

The AWS Nitro System enables …



PERFORMANCE

Better performance across CPU, networking, and storage



SECURITY

Enhanced security that continuously monitors, protects, and verifies the instance hardware and firmware



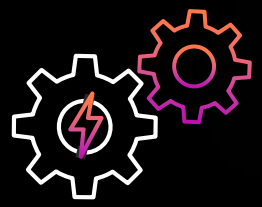
INNOVATION

Building blocks can be assembled in many different ways, giving us the flexibility to design and rapidly deliver EC2 instances

aws

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

The AWS Nitro System enables …



PERFORMANCE

Better performance across CPU, networking, and storage



SECURITY

Enhanced security that continuously monitors, protects, and verifies the instance hardware and firmware



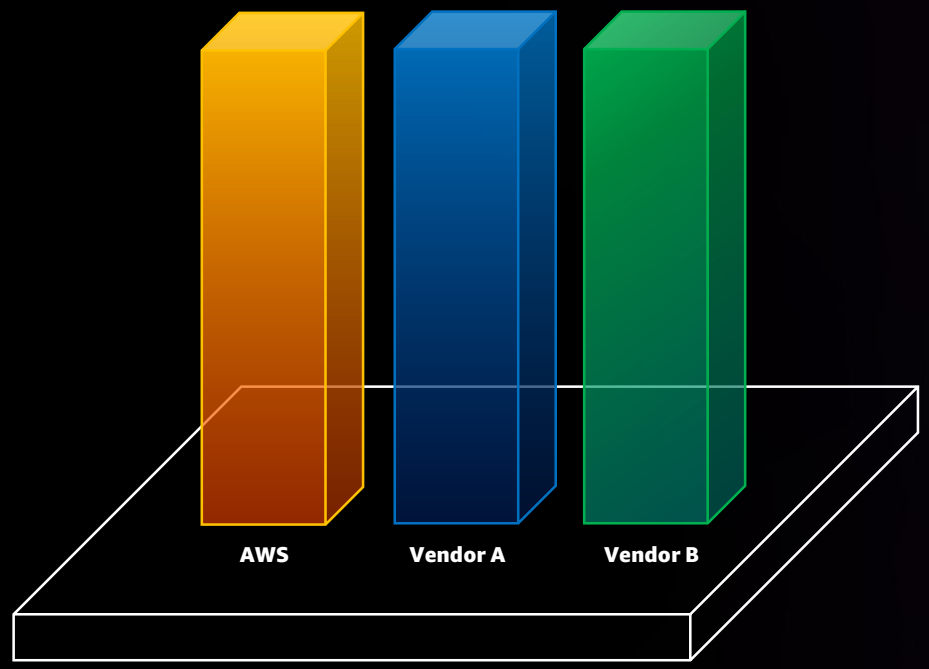
INNOVATION

Building blocks can be assembled in many different ways, giving us the flexibility to design and rapidly deliver EC2 instances

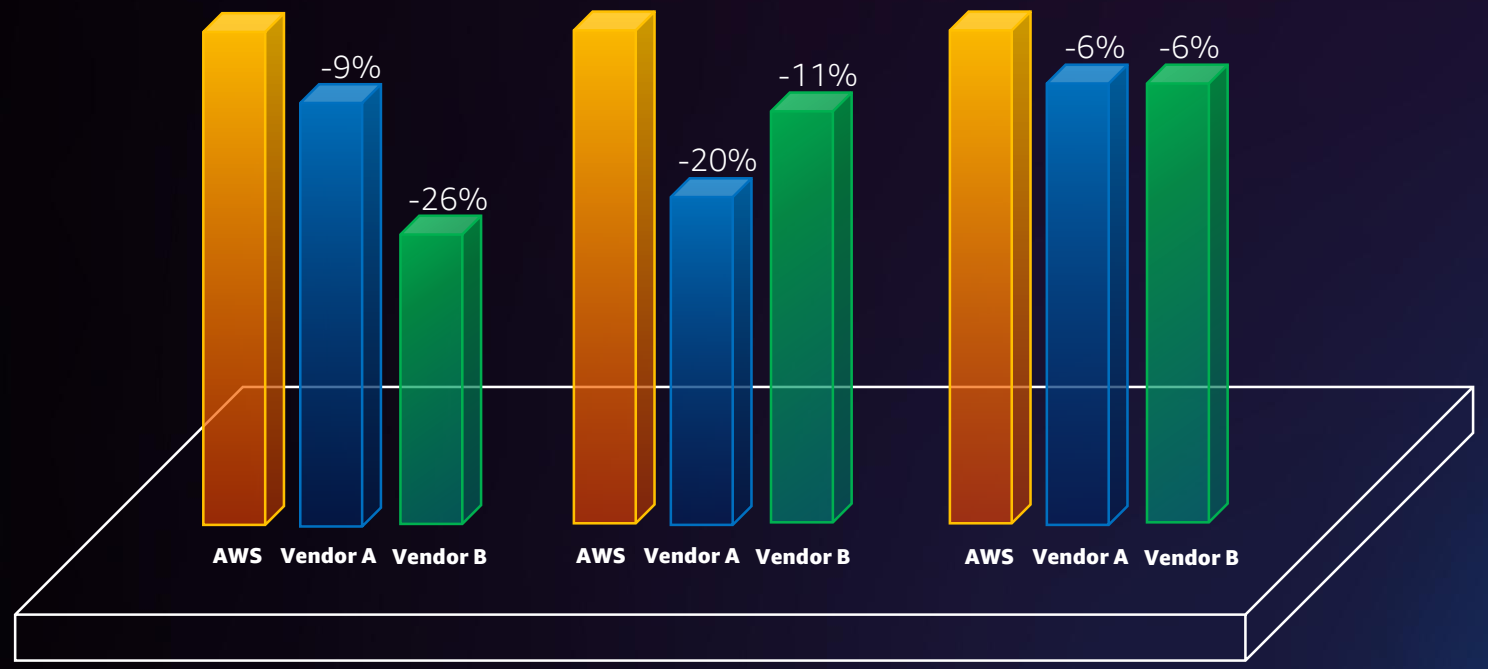
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro performance for real-world workloads

NITRO SYSTEM PROVIDES 14% BETTER PERFORMANCE ON AVERAGE



SPECINT2017



MEMCACHED

NGINX WEBSERVING

MYSQL SYSBENCH

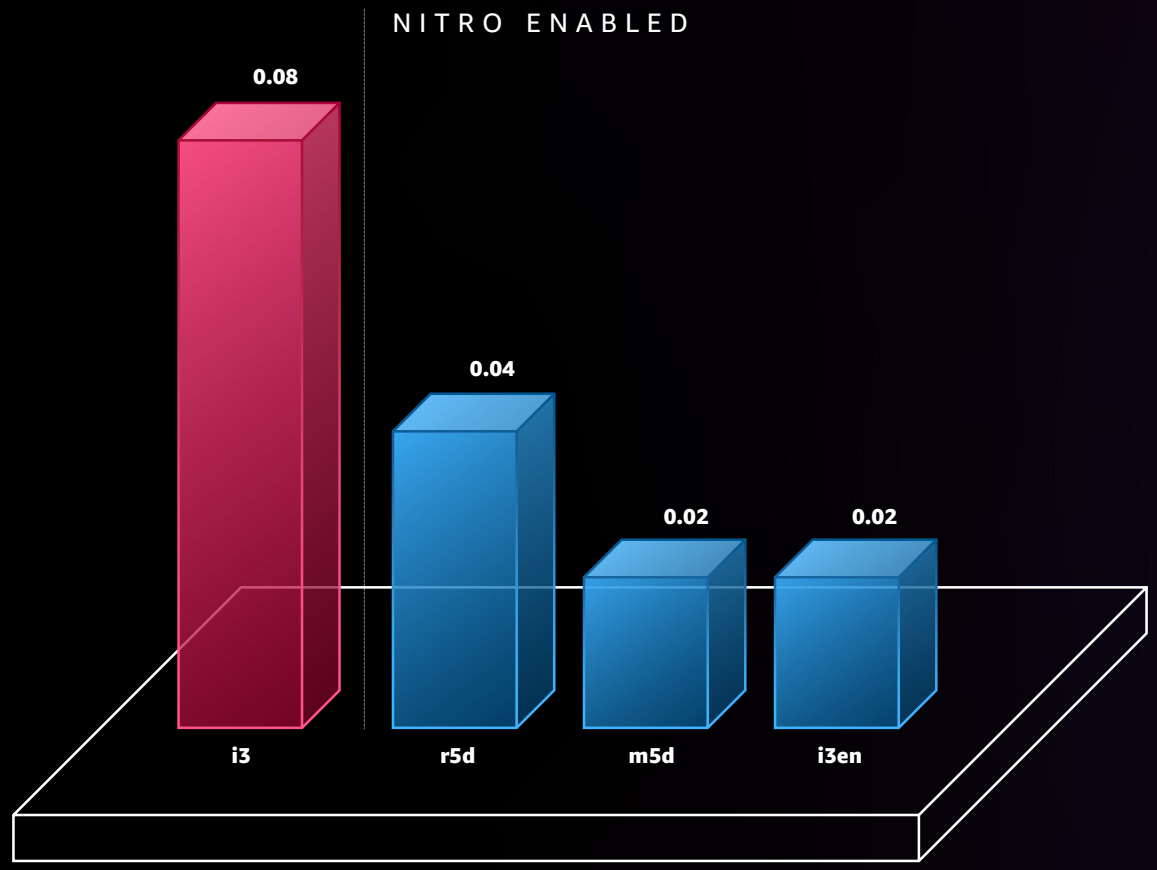
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

While the Nitro System provides an overall performance benefit, we also focus on component level performance

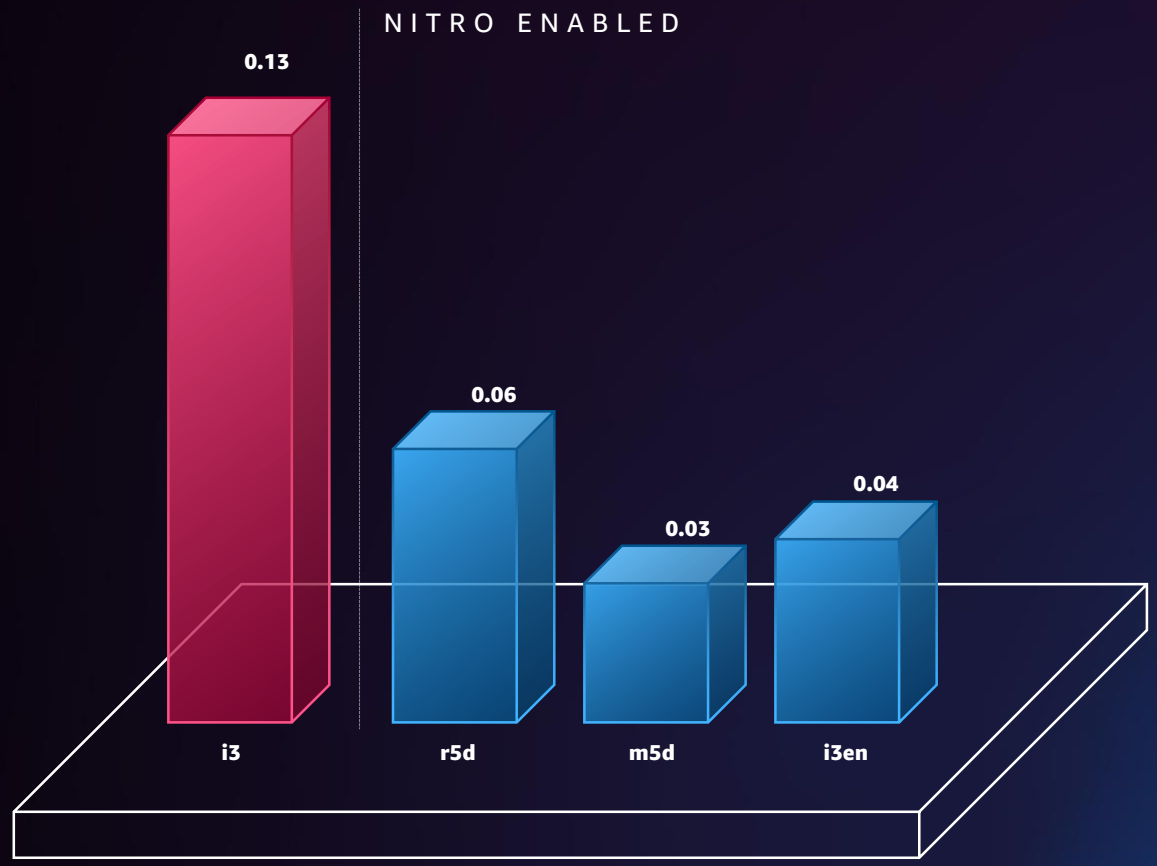
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro System storage performance

UP TO 4X IMPROVEMENT IN INSTANCE LATENCIES



AVERAGE LATENCY (P50)



TAIL LATENCY (P99)

aws

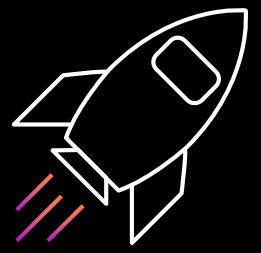
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

AVAILABLE NOW

Introducing AWS Nitro SSD

HIGH PERFORMANCE, LOW LATENCY SSD CUSTOM DESIGNED BY AWS



LOWER LATENCIES

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



IMPROVED RELIABILITY

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



NITRO SECURITY

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

aws

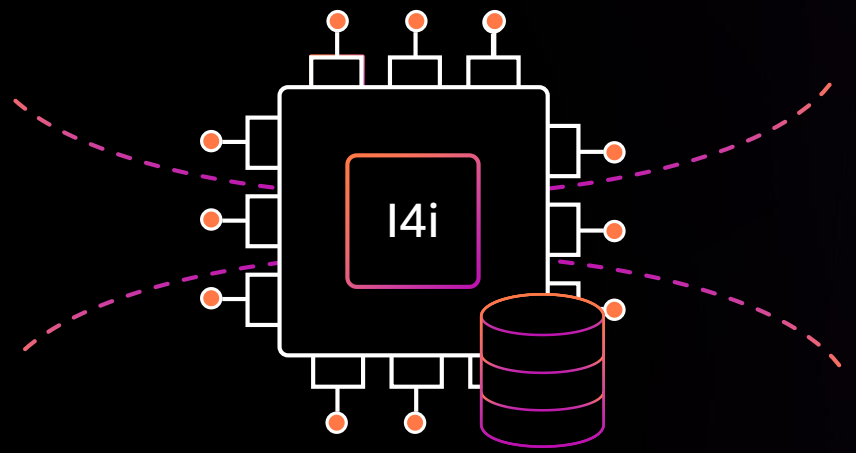
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

COMING SOON

Intel-based storage optimized instances

HIGHEST X86-BASED LOCAL STORAGE PERFORMANCE WITHIN AMAZON EC2



Powered by 3rd generation Intel Xeon Scalable processors (Ice Lake)

Up to 128 vCPUs, 1 TiB memory, 75 Gbps networking, and 30 TiB of high-performance AWS Nitro SSD NVMe storage

Offer 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 workloads

aws

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

HammerDB Benchmark

UP TO 68% PERFORMANCE IMPROVEMENT

HammerDB

HammerDB benchmark is used to test the performance of relational databases such as MySQL, PostgreSQL, Microsoft SQL Server and Oracle.

"I4i increases the performance of Microsoft SQL servers by up to 68% and reduces the transaction response time by up to 40% compared to I3. I4i also increases the performance of PostgreSQL by up to 32% compared to I3."

aws

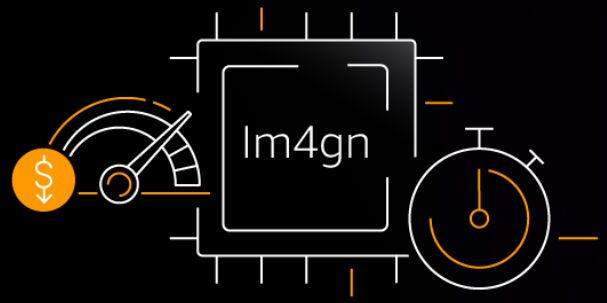


NEW!

COMING SOON

AWS Graviton2-based storage optimized instances

THE BEST PRICE PERFORMANCE FOR STORAGE INTENSIVE WORKLOADS IN AMAZON EC2



Up to 64 CPU cores, 256 GiB memory, and 100 Gbps networking

Provide up to 40% better price performance compared to I3 instances

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



Up to 32 CPU cores, 192 GiB memory, and 50 Gbps networking

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

Ideal for filesystems, steam processing and real-time databases, and log analytics

aws

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



Reducing storage costs

UP TO 50% DECREASE IN TCO

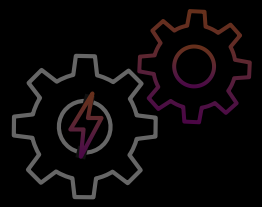
splunk>

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."

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

The AWS Nitro System enables …



PERFORMANCE

Better performance across CPU, networking, and storage



SECURITY

Enhanced security that continuously monitors, protects, and verifies the instance hardware and firmware



INNOVATION

Building blocks can be assembled in many different ways, giving us the flexibility to design and rapidly deliver EC2 instances

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

Abby Fuller

Principal Security Engineer, EC2

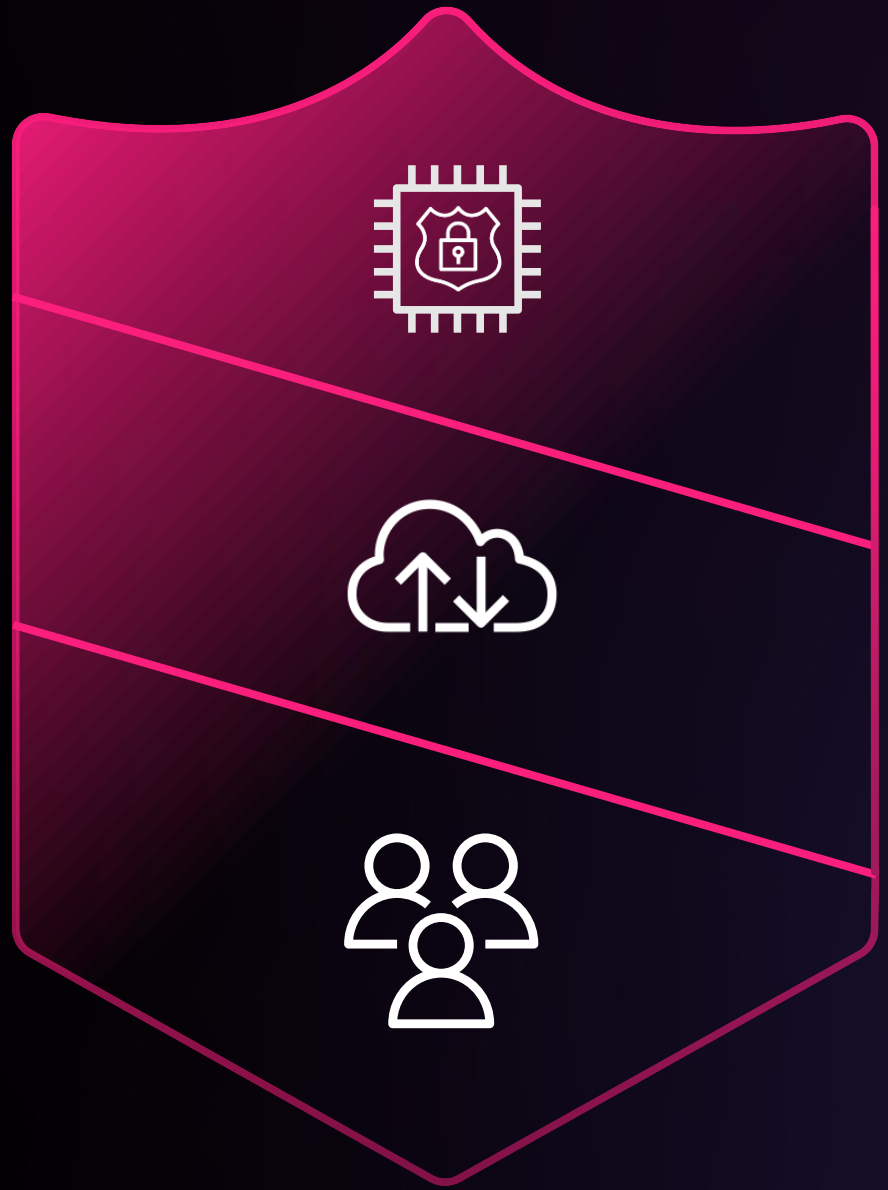


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Confidentiality & Privacy



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



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

Nitro System security

World-class security in the cloud

Engineered with a **hardware-based root of trust** using the Nitro Security Chip, allowing for the system to be cryptographically measured and validated

Dedicated **hardware-based virtualization** minimizes attack surfaces, reducing risk of vulnerabilities

No operator access. There is no mechanism for any system or person to log in to the underlying EC2 host. There is no root user, or interactive access.

VM tenancy protections. Instances are isolated from other tenants on the same EC2 host via VM, never containers.

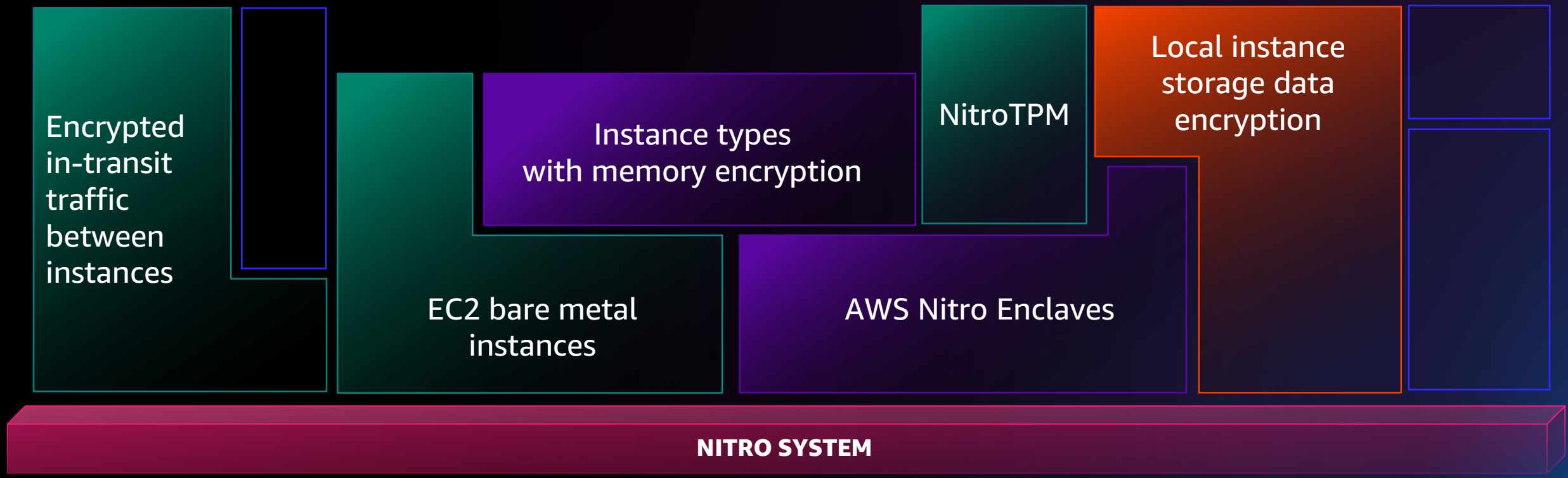


aws

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

Nitro System security

Nitro is the foundation for innovations in confidentiality and privacy



aws

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

NEW!

COMING SOON

Introducing AWS Nitro TPM

ADDITIONAL PROTECTION OF SENSITIVE INFORMATION IN EC2 INSTANCES



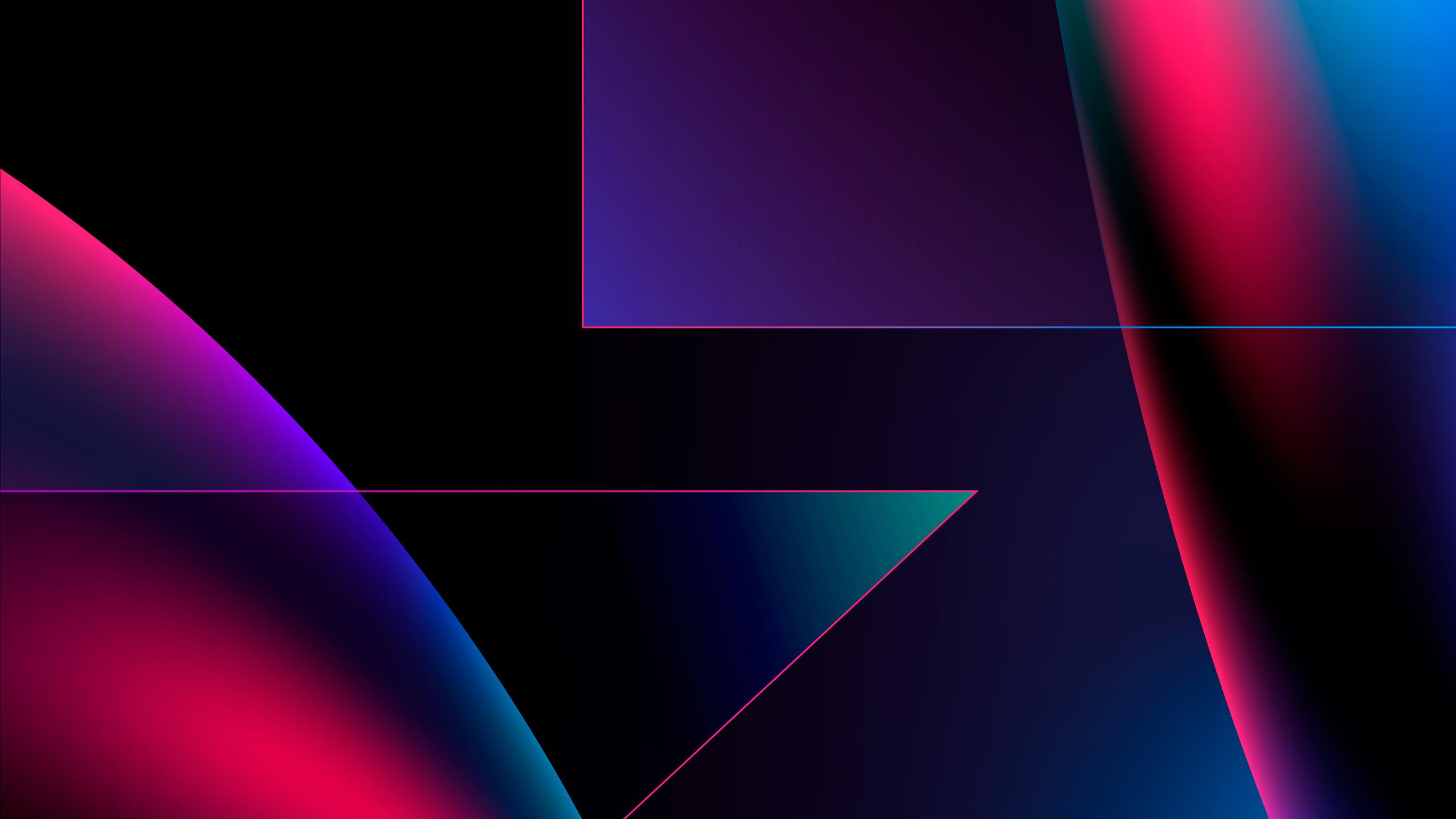
Allows attestation of the health of EC2 instances and provides a source of trust that is rooted in silicon

Easy to migrate on-premises applications that depend on a Trusted Platform Module (TPM)

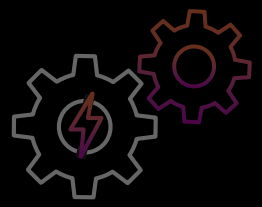
Conforms to the Trusted Platform Module 2.0 spec

aws

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



The AWS Nitro System enables …



PERFORMANCE

Better performance across CPU, networking, and storage



SECURITY

Enhanced security that continuously monitors, protects, and verifies the instance hardware and firmware



INNOVATION

Building blocks can be assembled in many different ways, giving us the flexibility to design and rapidly deliver EC2 instances

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

Broadest and deepest 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 4.5 GHz)

High memory footprint
(up to 24 TiB)

Instance storage
(HDD and NVMe)

Accelerated computing
(GPUs and FPGA)

Networking
(up to 800 Gbps)

Bare Metal

Size
(Nano to 112xlarge)

OPTIONS

Amazon EBS

Amazon Elastic Inference

MORE THAN

475

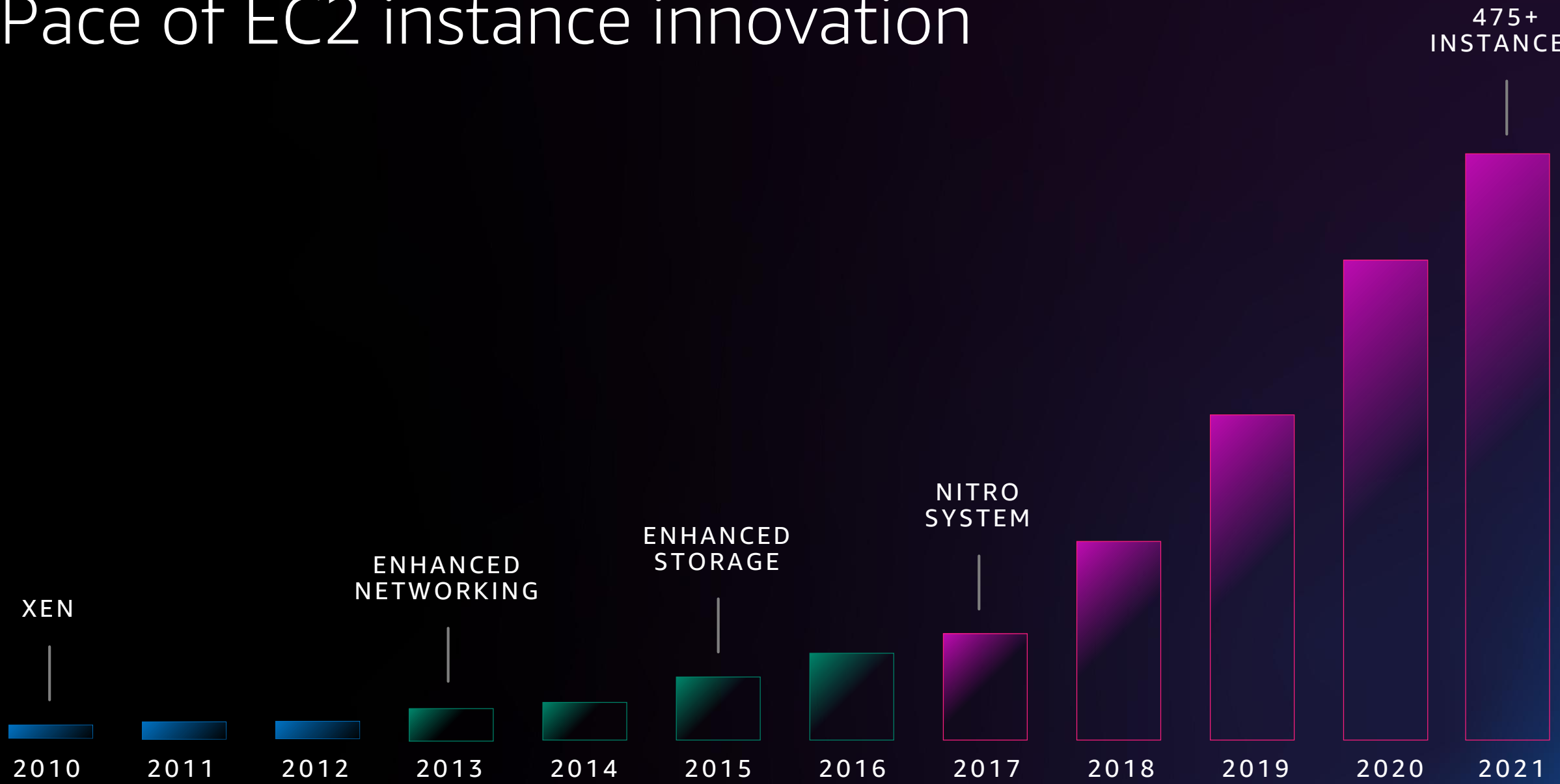
INSTANCE TYPES

for virtually every workload and business need

aws

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

Pace of EC2 instance innovation



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

Is there more we could do to ensure that customers are able to use their selected instance without the risk of deprecation?

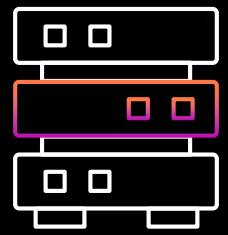
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

ANNOUNCING

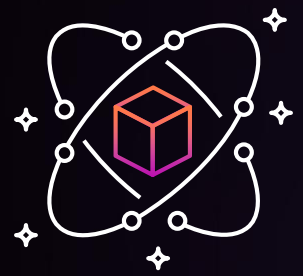
Long-term support for previous generation instances

SUPPORT FOR XEN-BASED EC2 INSTANCES ON THE NITRO SYSTEM



EXTENDED LIFETIME

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



NITRO BENEFITS

Enhanced security, scalability, and elasticity provided by the AWS Nitro System



SEAMLESS MIGRATION

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.

Broadest choice of processors and architectures

RIGHT COMPUTE FOR THE WORKLOAD



Intel Xeon Scalable processors



AMD EPYC processors



AWS Graviton processors

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

Innovating with Intel

15 YEARS OF COLLABORATION AND INNOVATION WITH AWS



COLLABORATION

Deep engineering collaboration across AWS portfolio



EXTENSIVE INTEGRATION

Over 275 EC2 instances are powered by Intel Xeon processors



FASTEST PROCESSOR

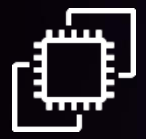
Fastest processor in the cloud and widest selection of Ice lake instances

INTEL-BASED INSTANCES



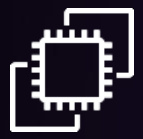
I4I

STORAGE OPTIMIZED



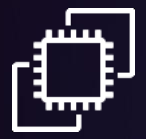
C6I

COMPUTE OPTIMIZED



R6I

MEMORY OPTIMIZED



M6I

GENERAL PURPOSE

aws

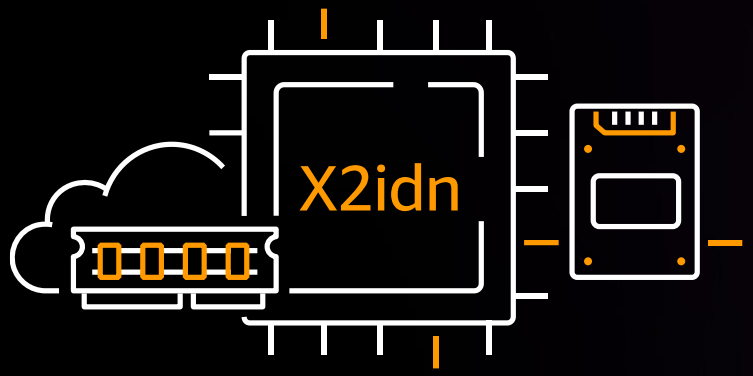
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

COMING SOON

Intel-based memory optimized instances

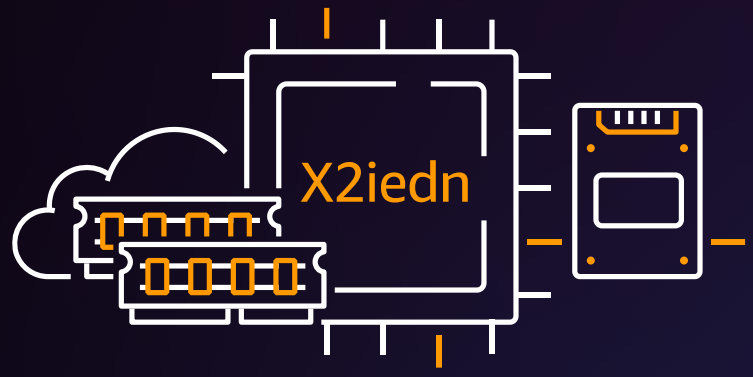
NEXT GENERATION MEMORY OPTIMIZED INSTANCES



Up to 64 CPU cores, 2 TiB memory, 80 Gbps Amazon EBS, and 100 Gbps networking

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

Ideal for in-memory databases, big data processing engines, and electronic design automation (EDA)



Up to 64 CPU cores, 4 TiB memory, 80 Gbps Amazon EBS, and 100 Gbps networking

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

Ideal for large in-memory databases (SAP HANA), in-memory analytics, and high performance databases

aws

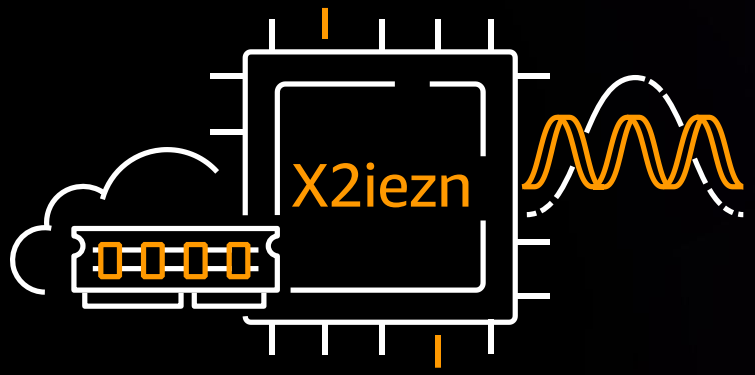
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

COMING SOON

High frequency memory optimized instance

POWERED BY THIRD-GENERATION EPYC MILAN PROCESSOR



Available in select AWS regions

Up to 96 CPU cores, 384 GiB memory, and 100 Gbps networking

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

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.

Electronic Design Automation

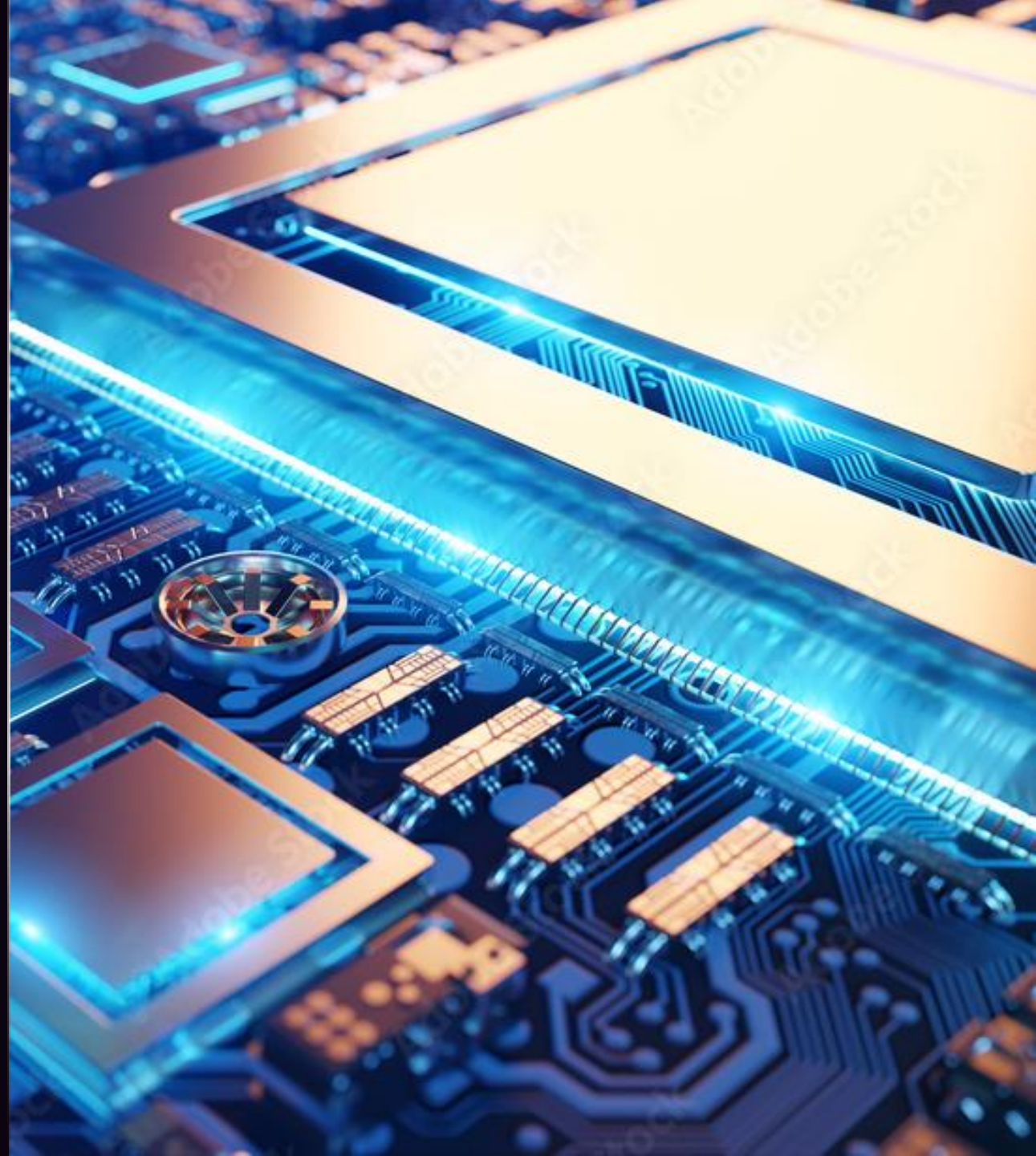
UP TO 25% PERFORMANCE IMPROVEMENT

AsteraLabs

Astera Labs, a fabless semiconductor company headquartered in California's Silicon Valley, develops purpose-built connectivity solutions for data-centric systems.

"Astera Labs was able to see performance gains of up to 25% compared to similar EDA workloads running on EC2 R5 instances."

aws



Innovating with Intel

15 YEARS OF COLLABORATION AND INNOVATION WITH AWS



COLLABORATION

Deep engineering collaboration across AWS portfolio



EXTENSIVE INTEGRATION

Over 275 EC2 instances are powered by Intel Xeon processors



FASTEST

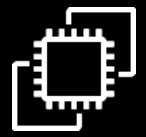
Fastest processor in the cloud and widest selection of Ice lake instances

INTEL-BASED INSTANCES



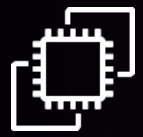
I4I

STORAGE OPTIMIZED



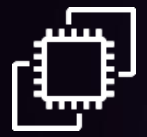
C6I

COMPUTE OPTIMIZED



R6I

MEMORY OPTIMIZED



M6I

GENERAL PURPOSE



X2IDN

MEMORY OPTIMIZED



X2IEDN

MEMORY OPTIMIZED



X2IEZN

HIGH FREQUENCY

aws

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

Broadest choice of processors and architectures

RIGHT COMPUTE FOR THE WORKLOAD



Intel Xeon Scalable processors



AMD EPYC processors



AWS Graviton processors

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

Innovating with AMD

10% LOWER COST VERSUS COMPARABLE X86 INSTANCES



FLEXIBILITY AND CHOICE

Help you optimize both cost and performance for your workloads



BETTER ECONOMICS

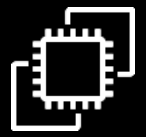
Deliver up to 10% lower cost versus comparable instances.



SEAMLESS MIGRATION

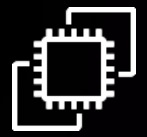
Easily migrate applications currently running on existing EC2 instances to the new AMD-based variants with little to no modification.

AMD-BASED INSTANCES



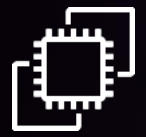
M5A

GENERAL PURPOSE



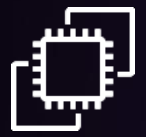
R5A

MEMORY OPTIMIZED



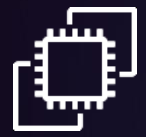
C5A

COMPUTE OPTIMIZED



T3A

BURST OPTIMIZED



G4AD

GPU OPTIMIZED



HPC6A

HPC OPTIMIZED

aws

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

AMD for microservices

10% REDUCTION ON COMPUTE COSTS

Expedia®

Expedia has migrated a number of microservices to AMD, where they were underutilizing compute

10% lower cost, application compatibility, and familiar instances were the reasons they decided to use AMD instances

aws

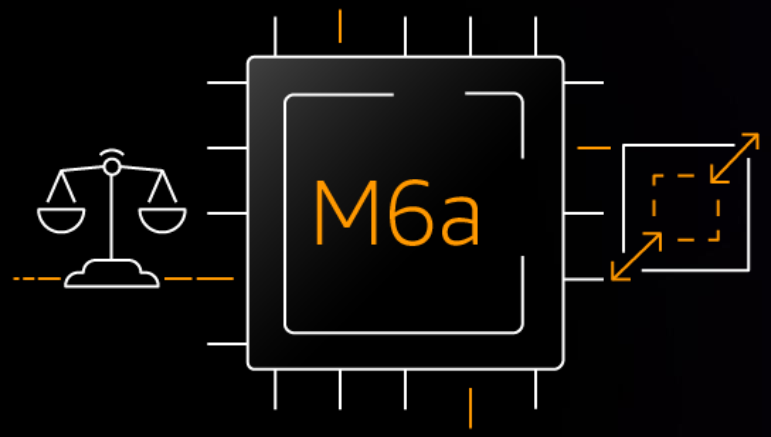


NEW!

AVAILABLE NOW

AMD-based General Purpose instance

GENERAL PURPOSE INSTANCES POWERED BY 3RD GENERATION AMD EPYC PROCESSOR



Up to 192 vCPUs, 768 GB memory, 50 Gbps networking, and 40 Gbps EBS bandwidth

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

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

aws

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

Innovating with AMD

10% LOWER COST VERSUS COMPARABLE X86 INSTANCES



FLEXIBILITY AND CHOICE

Help you optimize both cost and performance for your workloads



BETTER ECONOMICS

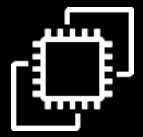
Deliver up to 10% lower cost versus comparable instances.



SEAMLESS MIGRATION

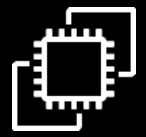
Easily migrate applications currently running on existing EC2 instances to the new AMD-based variants with little to no modification.

AMD-BASED INSTANCES



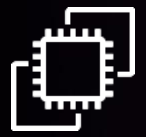
M5A

GENERAL PURPOSE



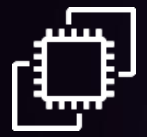
R5A

MEMORY OPTIMIZED



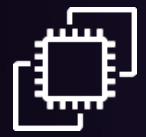
C5A

COMPUTE OPTIMIZED



T3A

BURST OPTIMIZED



G4AD

GPU OPTIMIZED



HPC6A

HPC OPTIMIZED



M6A

GENERAL PURPOSE

aws

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

Broadest choice of processors and architectures

RIGHT COMPUTE FOR THE WORKLOAD



Intel Xeon Scalable processors



AMD EPYC processors



AWS Graviton processors

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


powered by
annapurnalabs
an amazon company

AWS Graviton2

40% BETTER PRICE PERFORMANCE FOR BROAD RANGE OF WORKLOADS



BEST PRICE PERFORMANCE

Delivers up to 40% better price performance over comparable x86-based instances



EXTENSIVE ECOSYSTEM

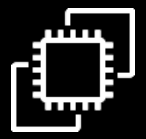
Supported by popular Linux operating systems along with many popular application and services from AWS and ISVs



ENHANCED SECURITY

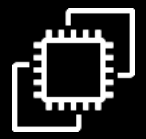
Provide key capabilities for application security, including always on 256-bit DRAM encryption

GRAVITON2 INSTANCES



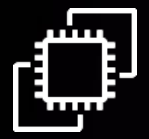
C6G(D)

COMPUTE OPTIMIZED



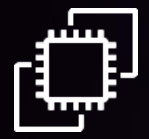
M6G(D)

GENERAL PURPOSE



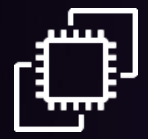
R6G(D)

MEMORY OPTIMIZED



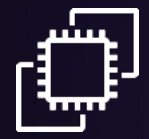
C6GN

ENHANCED NETWORKING



X2GD

ENHANCED MEMORY



T4G

BURSTABLE



IM4GN

STORAGE OPTIMIZED



IS4GEN

STORAGE OPTIMIZED

aws

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

SAP Hana Cloud to Graviton

SAP HANA is a world's leading in-memory database that serves as the foundation of the SAP Business Technology Platform. "Over the past decade, SAP HANA has helped customers manage their most mission critical transactional and analytics workloads," said Irfan Khan, President of HANA Database & Analytics, SAP. "AWS investments and innovations on ARM-based AWS Graviton processors and SAP HANA Cloud are a great match with potential to deliver step-wise operation and performance improvement benefits to our enterprise customers, and to SAP's cloud analytics and data management solutions powered by SAP HANA Cloud."



Igniting the Arm ecosystem

OPERATING SYSTEMS

- Amazon Linux 2
- Ubuntu 16.04, 18.04 and newer
- RedHat Enterprise Linux 7.6 and 8.0
- SUSE Linux Enterprise Server for Arm 15
- Docker Desktop Community and Docker Enterprise Engine

CONTAINERS

Most Docker official images support arm64

Amazon ECS
AVAILABLE TODAY

Amazon EKS
PUBLIC PREVIEW

INTEGRATED SERVICES

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

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

Migration is now simple and fast, in some cases we have seen customers move in as little as 4 days!

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

Supporting Prime Day

FEWER INSTANCES FOR 12.5% MORE LOAD

amazon

This year we increased our number of normalized instances by 12.5% using a three-Region cluster composed of over 53,000 C6g instances.

That sounds like a lot, but it's actually 6,000 fewer physical servers than we used on Cyber Monday 2020, and it was because of the performance and efficiency of Graviton2 that allowed us to do that.

aws

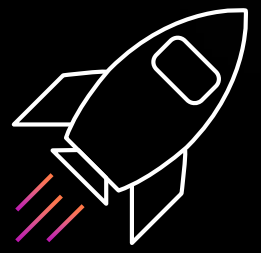


NEW!

AVAILABLE IN PREVIEW

Introducing AWS Graviton3

PROVIDE A 25% PERFORMANCE IMPROVEMENT OVER GRAVITON2



IMPROVED PERFORMANCE

Up to 25% higher compute performance and 2x higher floating-point to accelerate compute-intensive workloads



FASTER MEMORY

Supports DDR5 memory to provide 50% more memory bandwidth over DDR4 memory



ENERGY EFFICIENT

Consumes 60% less power for the same performance compared to other CPUs

aws

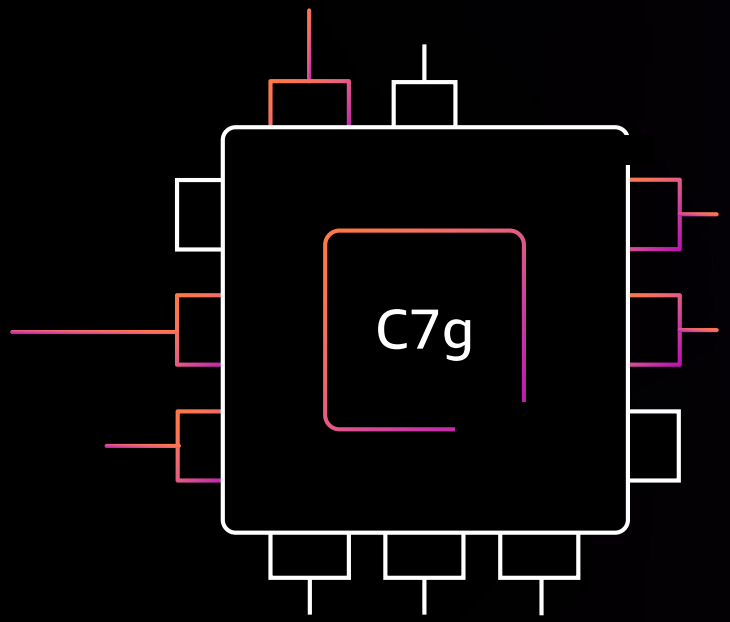
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

AVAILABLE IN PREVIEW

Graviton3-based compute optimized instance

BEST PRICE PERFORMANCE FOR COMPUTE-INTENSIVE WORKLOADS ON AMAZON EC2



Available in 8 instance sizes

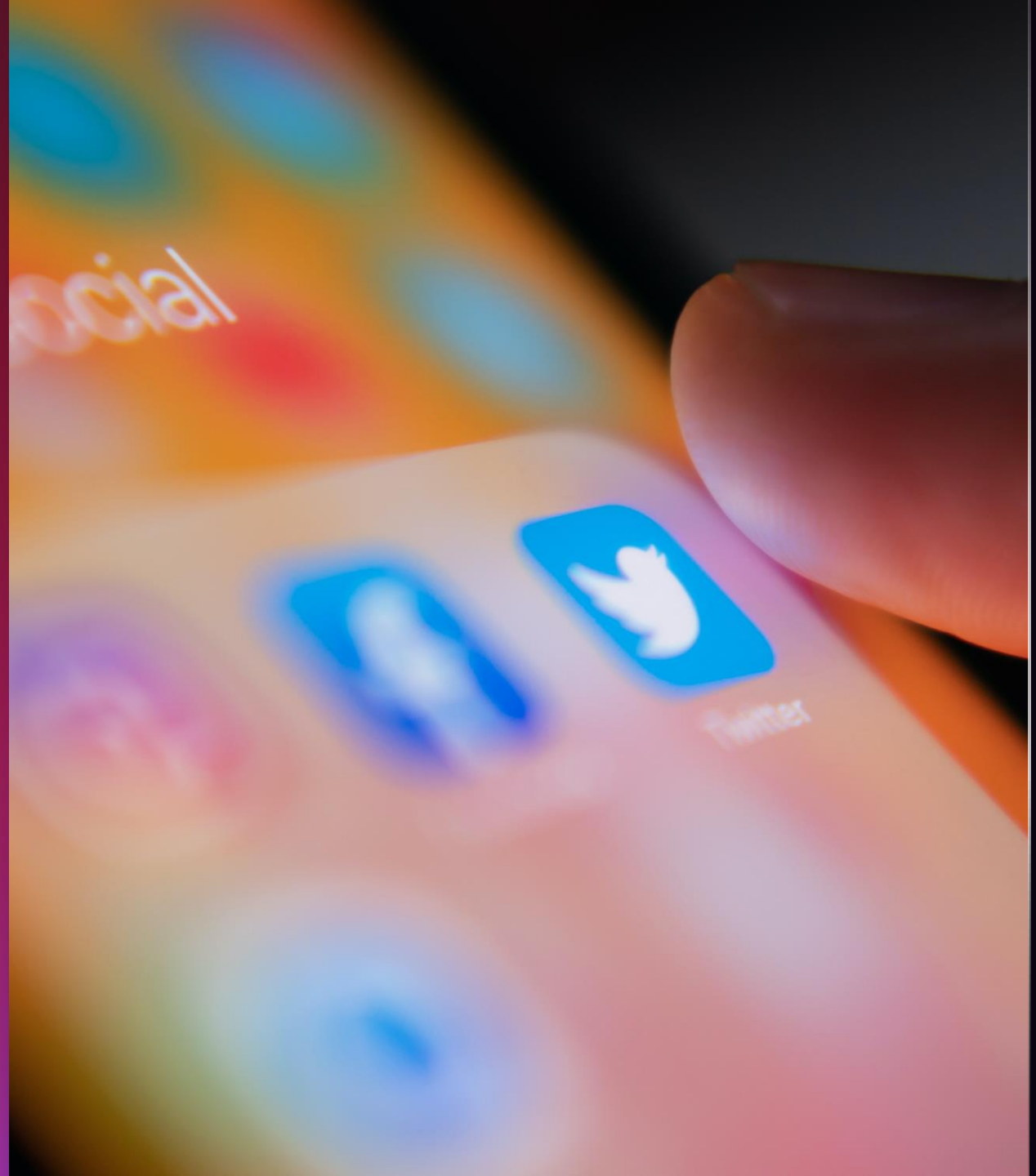
Up to 64 vCPUs, 128 GiB of memory, 30 Gbps of network performance, and 25 Gbps of EBS performance

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

aws

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



Twitter & Graviton3

UP TO 80% BETTER PERFORMANCE

Twitter is working on a multi-year project to leverage AWS Graviton-based EC2 instances to deliver Twitter timelines.

“Across a number of benchmarks that we've found to be representative of the performance of Twitter workloads, we found Graviton3-based C7g instances deliver 20-80% higher performance vs. Graviton2 based C6g instances, while also reducing tail latencies by as much as 35%."

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

Broadest choice of processors and architectures

RIGHT COMPUTE FOR THE WORKLOAD



Intel Xeon Scalable processors



AMD EPYC processors



AWS Graviton processors

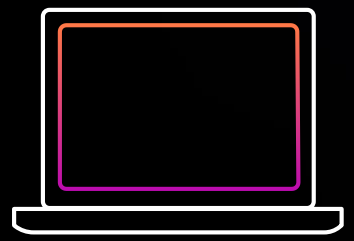
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



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

Apple Mac

RUN MAC WORKLOADS ON EC2



MACOS

Run macOS in the cloud, removing the need for dedicated hardware for Apple development



DEVELOPMENT

Speed up development lifecycle for iOS, macOS, tvOS, watchOS, iPadOS, and Safari apps



SCALE ON-DEMAND

Scale on-demand, pay-as-you go, and benefit from access to over 175 AWS services

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

Apple Development on AWS

4X REDUCTION IN BUILD TIMES

FLIPBOARD

"The setup has been very reliable, and we also enjoy the flexibility to use either command line or screen sharing. Build times have reduced from 20 minutes to 5 minutes. UI automation tests that used to take 3 hours now take less than 1 hour. Download speeds are amazing, and the VNC sessions are as smooth as butter. Needless to say, this project was a huge success and we are super happy with the finished results."

Colin Caufield, Lead iOS Engineer

aws



1

re:Inventing Compute on a Global Scale

2

Performance, Security, and Cost

3

Accelerated Workloads

4

Compute Where You Need It

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

3

Accelerated Workloads

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

Broadest choice of accelerators

RIGHT ACCELERATOR FOR THE WORKLOAD



XILINX FPGAs and Video Transcoding



NVIDIA A100 and T4 Tensor Core GPUs



AWS Inferentia and AWS Trainium

aws

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

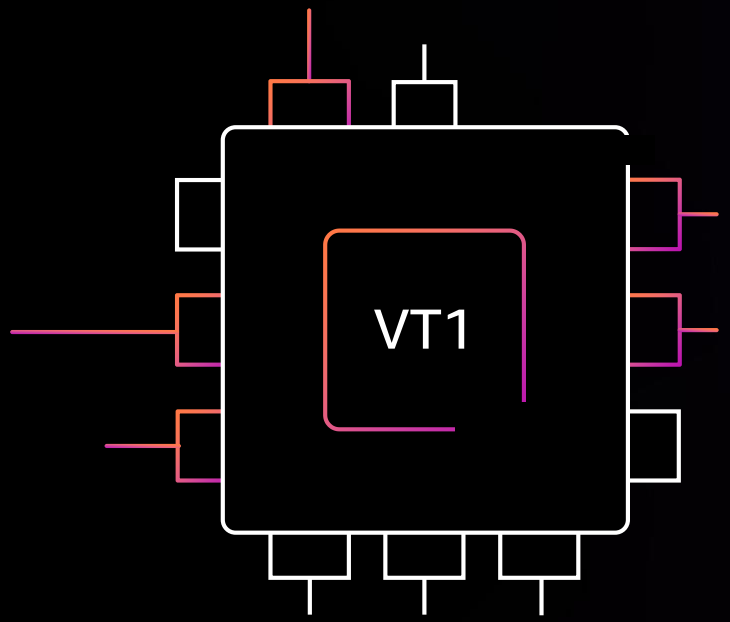

Live streaming surged during the pandemic and remains at levels previously unseen.

NEW!

RECENTLY LAUNCHED

EC2 instance for Video Transcoding

BEST PRICE PER STREAM IN THE CLOUD FOR LIVE VIDEO TRANSCODING



Available in 3 instance sizes

Up to 96 vCPUs, 192 GiB of memory, 25 Gbps of network performance

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

Support for up to 64 simultaneous 1080p60 streams in real time to reduce the cost per stream by up to 60% versus comparable EC2 instances

aws

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

Video Transcoding

Twitch is a live video streaming service that enables millions of people to interact and make their own entertainment every day.

“We are adopting Amazon EC2 VT1 instances to cost-efficiently transcode millions of live streams to deliver high streaming quality for viewers all over the world. We chose EC2 VT1 instances because they deliver the stream density and low latency we need without compromising on video compression or visual quality."





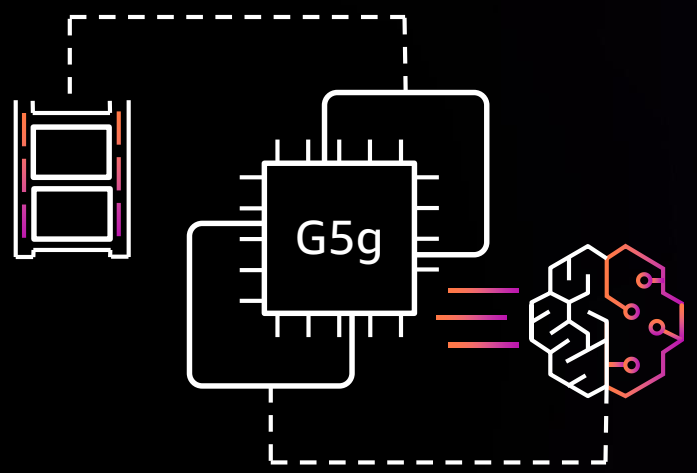
Mobile gaming has also seen a sharp increase with many game companies looking to stream games from the cloud.

NEW!

AVAILABLE NOW

Amazon EC2 G5g instance

MOST COST-EFFECTIVE GPU-BASED INSTANCE FOR GAME STREAMING



Up to 64 Graviton2 CPU cores, two NVIDIA T4g GPUs with 16 GB of GPU memory each, and 128 GB of system memory

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

Full support for Android, OpenGL, TensorFlow, and PyTorch with Graviton2 processors and NVIDIA T4G Tensor Core GPUs

Run Android games natively without need for expensive cross compilation or emulation

aws

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



Android game streaming

REDUCING THE COST OF STREAMING

BlueStacks

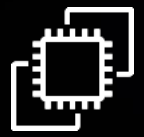
"There are two challenges in offering a cloud-based mobile gaming service. First is technical - for mobile gaming you need powerful GPUs and ARM CPUs working together. The second challenge is to be able to offer mobile cloud at a price point which makes the business viable. G5g is a major leap forward in both these dimensions, and we have been working closely with AWS to run NowCloudOS on top of G5g instances to bring these benefits to the mobile cloud."

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

Supporting ML workloads

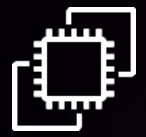
LARGEST SELECTION OF ML INSTANCES IN THE CLOUD

TRAINING



P3DN

NVIDIA V100
TENSOR CORE



P4D

NVIDIA A100
TENSOR CORE



G5

NVIDIA
GPU



DL1

HABANA
GAUDI



R6I

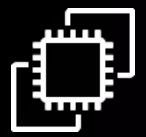
MEMORY
OPTIMIZED

INFERENCE



G5G

AWS
GRAVITON 2



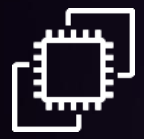
INF1

AWS
INFERENTIA



G5

NVIDIA
GPU



G4

NVIDIA
T4 GPU



EI

ELASTIC
INFERENCE



C6I

COMPUTE
OPTIMIZED

aws

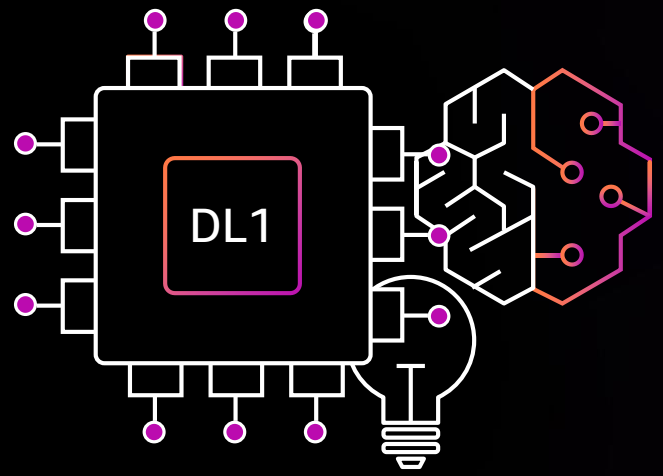
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

RECENTLY LAUNCHED

Amazon EC2 DL1 instances

LOW COST-TO-TRAIN FOR DEEP LEARNING MODELS



Intel-based with up to 8 Gaudi chips with 32 GB of high bandwidth memory (HBM) per accelerator, 768 GiB of memory, 400 Gbps networking, and 4 TB of local NVMe storage.

DL1 instances deliver up to 40% better price performance for training deep learning models compared to our latest GPU-based EC2 instances.

DL1 instances support leading ML frameworks such as TensorFlow and PyTorch, enabling you to continue using your preferred ML workflows.

aws

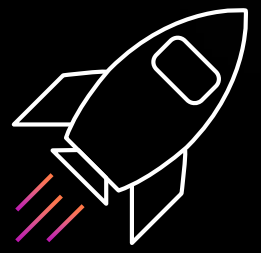
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

AVAILABLE NOW

Introducing AWS Trainium

HIGH PERFORMANCE MACHINE LEARNING TRAINING CHIP, CUSTOM DESIGNED BY AWS



IMPROVED PERFORMANCE

Offers the highest performance in Amazon EC2 for ML training



LOWER COST

Will provide the best price performance for training ML models in the cloud



FRAMEWORK SUPPORT

Integrated with leading ML frameworks, such as TensorFlow and PyTorch

aws

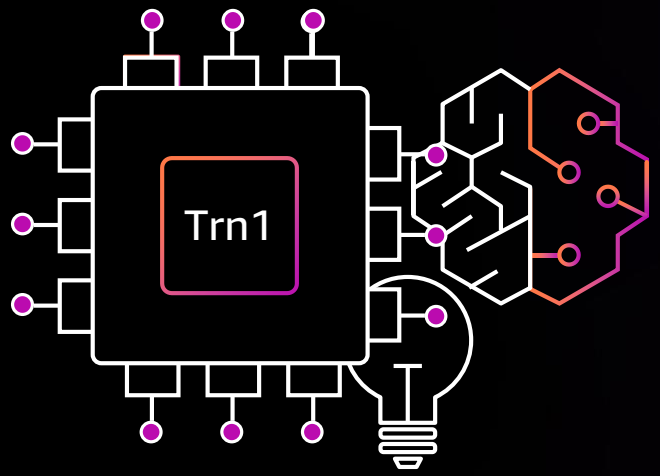
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW!

AVAILABLE NOW

Amazon EC2 Trn1 instances

HIGHEST PERFORMANCE AND LOWEST COST FOR TRAINING DEEP LEARNING MODELS



16 Trainium accelerators with 512 GB of HBM2 high speed memory, 800 Gbps of networking throughput and 8 TB of local NVMe storage

Lower time-to-train when compared to the latest comparable EC2 instances

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.



Deep Learning Training

REDUCING THE COST OF TRAINING

ANTHROP\C

"A major key to our success is access to modern infrastructure that allows us to spin up very large fleets of high performance deep learning accelerators. We are looking forward to using Amazon EC2 Trn1 instances powered by AWS Trainium, as their unprecedented ability to scale to tens of thousands of nodes and higher network bandwidth will enable us to iterate faster while keeping our costs under control."

Tom Brown, Cofounder

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

Supporting ML workloads

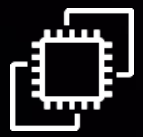
LARGEST SELECTION OF ML INSTANCES IN THE CLOUD

TRAINING



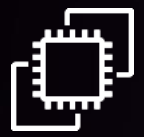
TRN1

AWS TRAINIUM



P3DN

NVIDIA V100 TENSOR CORE



P4D

NVIDIA A100 TENSOR CORE



G5

NVIDIA GPU



DL1

HABANA GAUDI



R6I

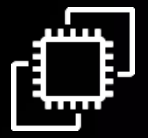
MEMORY OPTIMIZED

INFERENCE



G5G

AWS GRAVITON 2



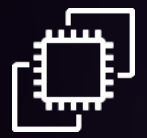
INF1

AWS INFERENTIA



G5

NVIDIA GPU



G4

NVIDIA T4 GPU



EI

ELASTIC INFERENCE



C6I

COMPUTE OPTIMIZED

aws

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

1

re:Inventing Compute on a Global Scale

2

Performance, Security, and Cost

3

Accelerated Workloads

4

Compute Where You Need It

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



Compute Where You Need It

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

Extending the cloud

DELIVERING THE CLOUD ANYWHERE CUSTOMERS NEED IT

GLOBAL



AWS Regions

METRO CENTERS



AWS Local Zones

5G NETWORKS



AWS Wavelength

ON-PREMISES



AWS Outposts

SMART DEVICES



Internet of Things

RUGGED EDGE



AWS Snow Family

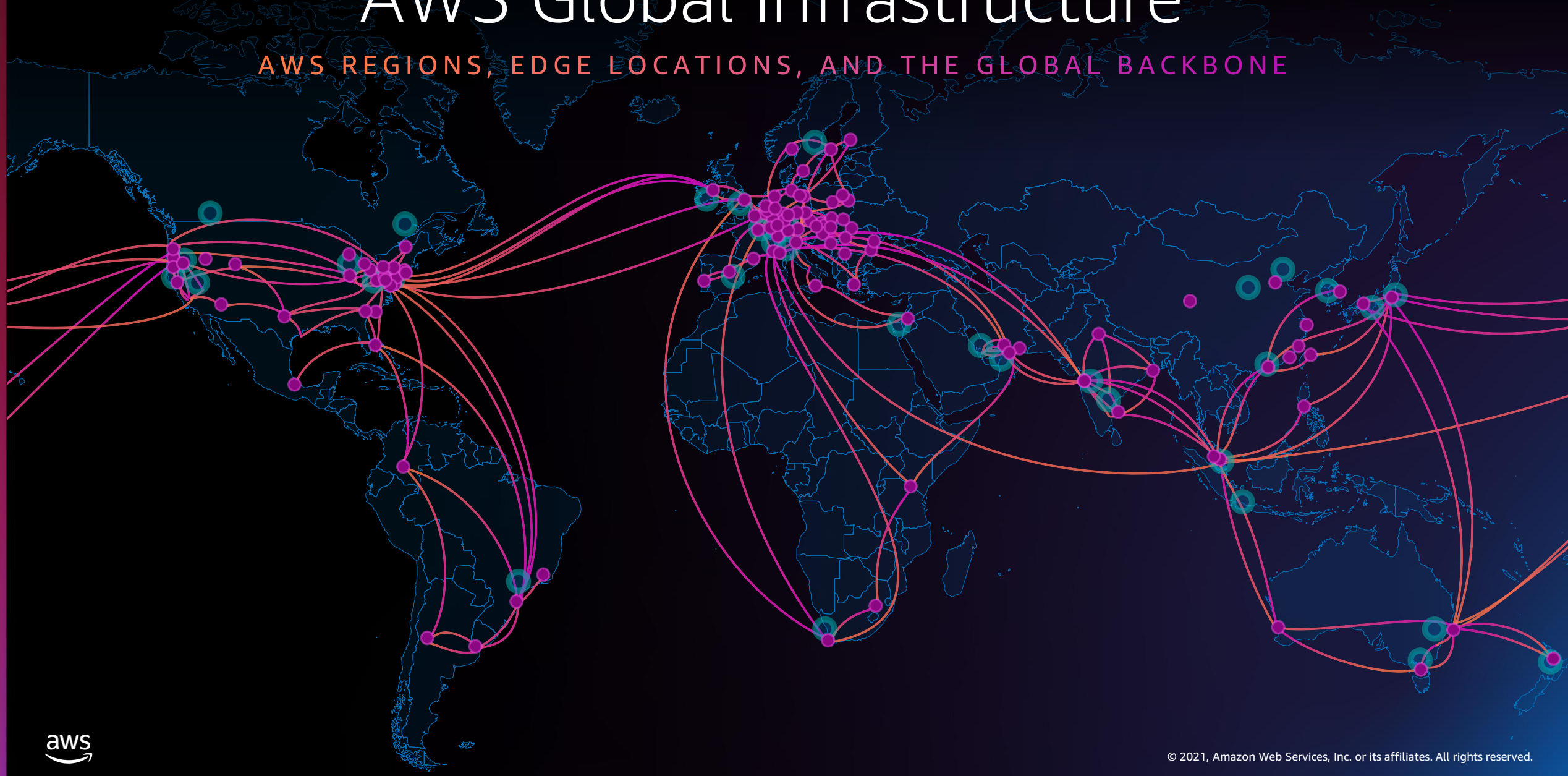
SAME INFRASTRUCTURE, SERVICES, APIS, AND TOOLS FOR A CONSISTENT EXPERIENCE

aws

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

AWS Global Infrastructure

AWS REGIONS, EDGE LOCATIONS, AND THE GLOBAL BACKBONE



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

NEW!

AWS Local Zones

RUN LATENCY-SENSITIVE APPLICATIONS AT THE EDGE USING AWS INFRASTRUCTURE AND SERVICES



LOW LATENCY

Extends AWS infrastructure services, APIs, and tools to where customers need it to support low-latency applications



FULLY MANAGED

Fully owned, managed, and supported by AWS



NEW AWS INFRASTRUCTURE

New type of AWS infrastructure that places AWS compute, storage, networking, and select AWS services closer to where your end users are located

aws

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

AWS Local Zones

RUN LATENCY SENSITIVE APPLICATIONS CLOSER TO END USERS



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

AWS Local Zones

RUN LATENCY SENSITIVE APPLICATIONS CLOSER TO END USERS



NEW LOCAL ZONES LAUNCHING IN SEATTLE, ATLANTA, AND PHOENIX BEFORE THE END OF 2021

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

Make sure you watch Werner's keynote on Thursday at 8 a.m. for more Local Zones announcements

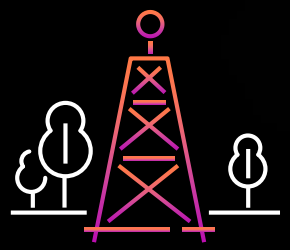
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.


Emergence of 5G

NEW!

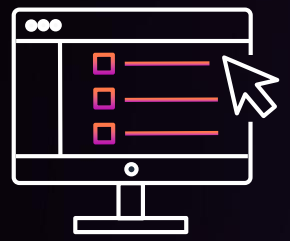
AWS Wavelength

RUN LATENCY-SENSITIVE APPLICATIONS AT THE 5G EDGE USING AWS INFRASTRUCTURE AND SERVICES



EMBEDDED IN 5G NETWORKS

Extends AWS infrastructure, services, APIs, and tools to 5G networks



AWS API & CONSOLE

Single pane of management in the cloud providing the same APIs and tools as in AWS Regions



BUILD ONCE, DEPLOY ANYWHERE

Simple to deploy application from AWS regions to 5G-enabled edge locations

aws

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

AWS Wavelength

RUN LATENCY SENSITIVE APPLICATIONS AT THE 5G EDGE



5G Partners

Bell

verizon

KDDI

SK telecom

vodafone business

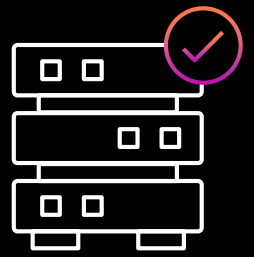
aws

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



AWS Outposts

AWS INFRASTRUCTURE AND SERVICES IN YOUR ON-PREMISES LOCATION



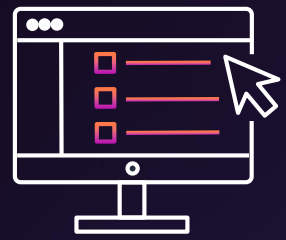
AWS DESIGNED

Same AWS-designed infrastructure as in AWS data centers (built on AWS Nitro System)



FULLY MANAGED

Fully managed, monitored, and operated by AWS as if in AWS Regions



AWS API & CONSOLE

Single pane of management in the cloud providing the same APIs and tools as in AWS Regions

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

AWS Outposts

AWS INFRASTRUCTURE AND SERVICES IN YOUR ON-PREMISES LOCATION





























Booz | Allen | Hamilton

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

NEW!

AVAILABLE NOW

Introducing AWS Outpost Servers

AWS INFRASTRUCTURE AND SERVICES IN YOUR ON-PREMISES LOCATION



1U SINGLE SERVER

Graviton2 processors

Up to 64 vCPU, 128 GiB RAM, 2x 1900 GiB NVMe SSD



2U SINGLE SERVER

Intel Xeon Scalable Processors (Ice Lake)

Up to 128 vCPU, 256 GiB RAM, 4x 1900 GiB NVMe SSD

aws

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

Extending the cloud

DELIVERING THE CLOUD ANYWHERE CUSTOMERS NEED IT

GLOBAL



AWS Regions

METRO CENTERS



AWS Local Zones

5G NETWORKS



AWS Wavelength

ON-PREMISES



AWS Outposts

SMART DEVICES



Internet of Things

RUGGED EDGE



AWS Snow Family

SAME INFRASTRUCTURE, SERVICES, APIS, AND TOOLS FOR A CONSISTENT EXPERIENCE

aws

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

Exchange in the Cloud

HYBRID COMPUTE FOR A GLOBAL MARKET

Nasdaq

Leveraging the on-premises capabilities of AWS Outposts, which extends AWS infrastructure, AWS services, APIs, and tools to virtually any datacenter, co-location space, or on-premises facility, Nasdaq plans to configure its network to deliver ultra-low-latency edge compute capabilities directly in its primary data center, Equinix NY11, in Carteret, NJ, to establish the first ever Private Local Zone in the capital markets industry at Carteret.

aws



1

re:Inventing Compute on a Global Scale

2

Performance, Security, and Cost

3

Accelerated Workloads

4

Compute Where You Need It

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



Gartner recognizes **AWS as a Leader** for the 11th straight year

Magic Quadrant for Cloud Infrastructure and Platform Services

Gartner, Magic Quadrant for Cloud Infrastructure & Platform Services, Raj Bala, Bob Gill, Dennis Smith, Kevin Ji, David Wright, 27 July 2021. Gartner and Magic Quadrant are registered trademarks of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose. This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from AWS.

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

Thank you!

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