INNOVATE

AI/ML EDITION

24 February 2022

Solving challenges with AI/ML powered by Intel

Akanksha Balani

AWS APJ Alliance Head
Global AI and HPC GTM Lead
Intel

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

Agenda

- AI/ML for industry today
- Why Intel & AWS
- Intel AI on AWS for your business
- Success stories on scale
- Learn | Engage | Innovate AI with Intel

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

AI/ML transformation across industries









Consumer	Health	Finance	Retail	Government	Energy	Transport	Industrial	Other
Smart Assistants	Enhanced Diagnostics	Algorithmic Trading	Support	Defense	Oil & Gas Exploration	Autonomous Cars	Factory Automation	Advertising
Chatbots	Drug Discovery	Fraud Detection	Experience	Data Insights	Smart Grid	Automated Trucking	Predictive Maintenance	Education
Search	Patient Care	Research	Marketing	Safety & Security	Operational Improvement	Aerospace	Precision Agriculture	Gaming
Personalization	Research	Personal Finance	Merchandising	Resident Engagement	Conservation	Shipping	Field Automation	Professional & IT Services
Augmented Reality	Sensory Aids	Risk Mitigation	Loyalty	Smarter Cities		Search & Rescue		Telco/Media
Robots			Supply Chain					Sports
			Security					

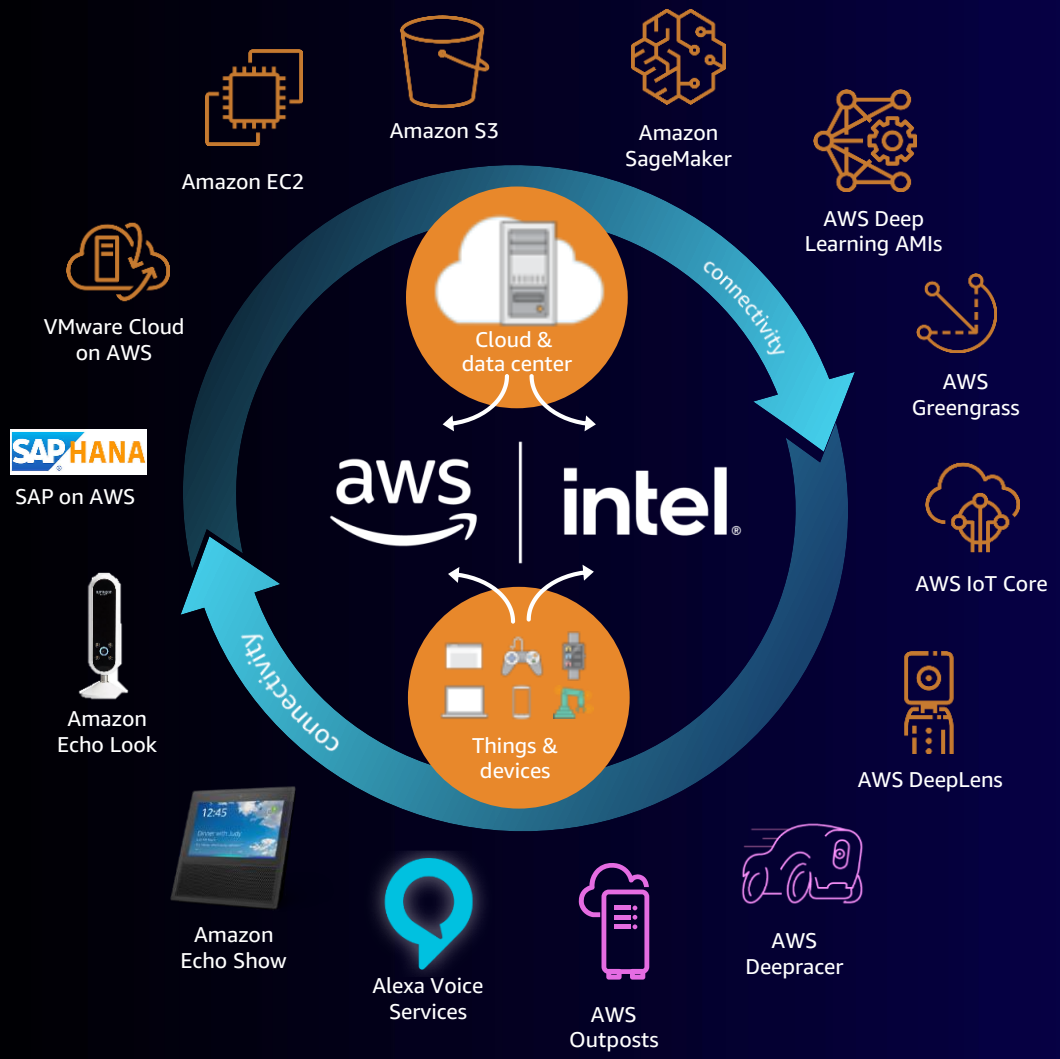
Source: Intel forecast

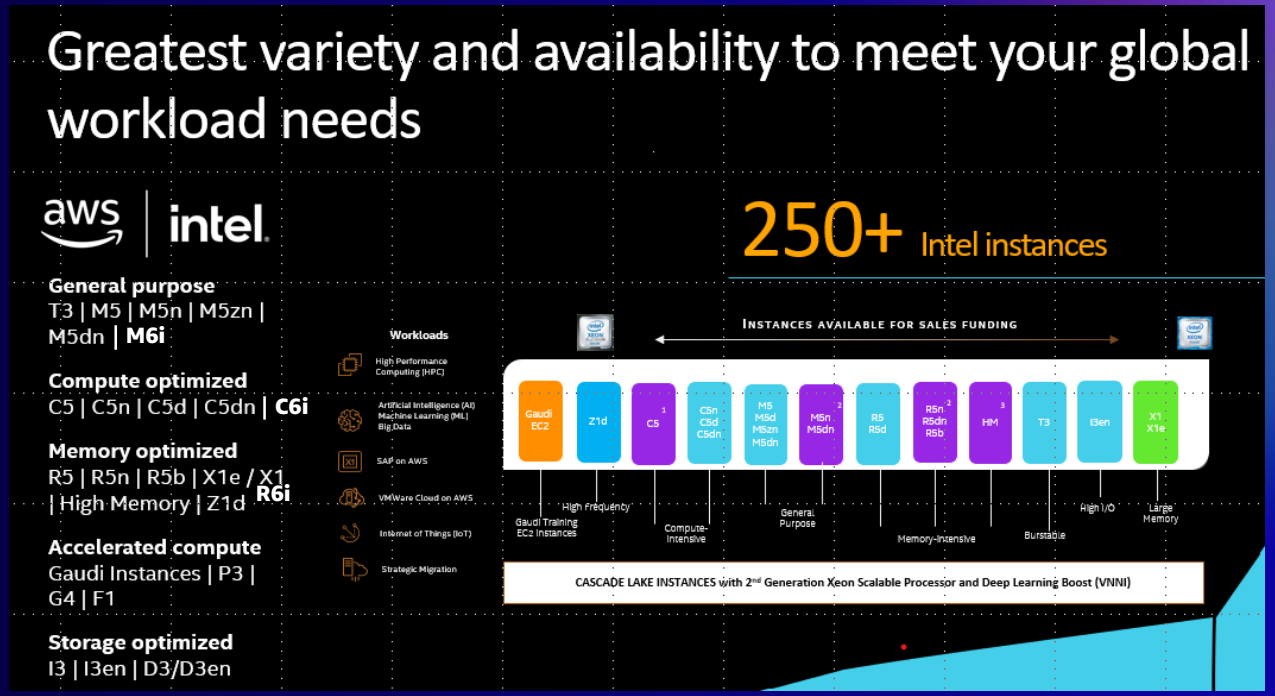
aws

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

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

What does Intel do with Amazon?





Key Workloads
Strategic migrations

- AI/ML
- HPC
- SAP
- Hybrid
- Edge

"Intel is a very deep partner of AWS and will be for a long time. That's not changing."

Andy Jassy
CEO, AWS

† Formerly the Intel® Computer Vision SDK
**Other names and brands may be claimed as the property of others.*
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.

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

Greatest variety and availability to meet your global workload needs

aws | intel

250+ Intel instances

16 years of partnership

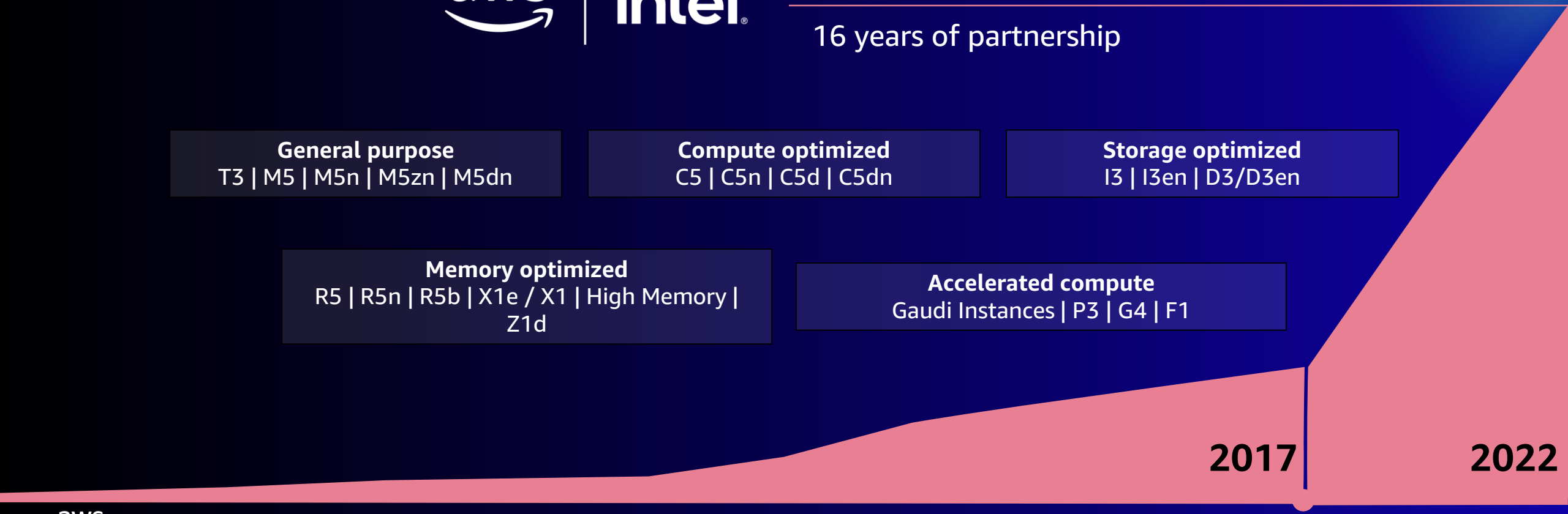
General purpose
T3 | M5 | M5n | M5zn | M5dn

Compute optimized
C5 | C5n | C5d | C5dn

Storage optimized
I3 | I3en | D3/D3en

Memory optimized
R5 | R5n | R5b | X1e / X1 | High Memory | Z1d

Accelerated compute
Gaudi Instances | P3 | G4 | F1



aws

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

AWS data acceleration with Intel

AWS C5 Deep Learning AMI Optimized for Intel CPU (Training)

7.4x

faster than training on the stock Tensorflow 1.6 binaries

AWS C5 Deep Learning AMI Optimized for Intel CPU (Inference)

12x

faster than default configuration for NMT Inference with MxNet

AWS SageMaker Machine Learning Optimized for Intel CPU (Training & Inference)

10x

Machine Learning Algorithms optimized for IA CPU

Amazon SageMaker

Amazon Rekognition

Amazon Forecast

Amazon Personalize

AWS Marketplace: Deep learning AMIs | Multiple containers with OneDNN-optimized frameworks

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

Highlights of the past year – AWS and Intel

Career Launcher Rapidly Scales Learning Portal during Pandemic
Intel & AWS partner to help serve >160,000 students in India within two months on AWS.[1]

65x more parallel wildfire simulations
Intel & AWS partner with RONIN to help increase fire fighting effectiveness in Australia. [4]

intel SELECT SOLUTIONS +

AWS ParallelCluster

AWS as first CSP with verified Intel Select Solution [2]

Amazon EC2 M5zn instance – fastest Intel Xeon Scalable CPU in the Cloud [5]
Highest all-core turbo CPU performance with a frequency up to 4.5 GHz.

AWS announces DL1, M6i, C6i
DL1 - AI instances with 40% better price/perf built on Habana Gaudi [3]
Intel's Habana & AWS co-engineer solution using up to 8 Gaudi accelerators

https://aws.amazon.com/intel/

[1] https://www.intel.com/content/www/us/en/customer-spotlight/stories/career-launcher-customer-story.html
[2] https://docs.aws.amazon.com/parallelcluster/latest/ug/intel-select-solutions.html
[3] https://aws.amazon.com/ec2/instance-types/habana-gaudi/
[4] https://dpgresources.intel.com/asset-library/intel-aws-the-csiro-spark-intel-poc-summary/
[5] https://aws.amazon.com/blogs/aws/new-ec2-m5zn-instances-fastest-intel-xeon-scalable-cpu-in-the-cloud/

aws

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

Data analytics portfolio

Solutions
Solution Architects

Platforms · *Finance* · *Healthcare* · *Energy* · *Industrial* · *Transport* · *Retail* · *Home* · *More…*

Toolkits
App Developers

OpenVINO™ Toolkit
OpenVINO Toolkit for inference deployment on CPU, processor graphics, FPGA & VPU using TF, Caffe & MXNet**

Deep Learning Developer Toolkit
Optimized inference deployment for all Intel® Movidius™ VPUs using TensorFlow & Caffe**

Libraries
Data Scientists

MACHINE LEARNING LIBRARIES

Python	R	Distributed
• Scikit-learn	• Cart	• MlLib (on Spark)
• Pandas	• RandomForest	• Mahout
• NumPy	• E1071	

DEEP LEARNING FRAMEWORKS

TensorFlow* MXNet* Caffe* BigDL/Spark* Caffe2* PyTorch* PaddlePaddle*

Foundation
Library Developers

ANALYTICS, MACHINE & DEEP LEARNING PRIMITIVES

Python
Intel distribution optimized for machine learning

DAAL
Intel® Data Analytics Acceleration Library (for machine learning)

MKL-DNN
Open-source deep neural network functions for CPU, processor graphics

clDNN

Hardware
IT System Architects

FOUNDATION

ACCELERATORS

habana An Intel Company

Inference

† Formerly the Intel® Computer Vision SDK
**Other names and brands may be claimed as the property of others.*
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.

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

New 3rd generation Intel Xeon scalable processor

- Higher workload performance
- Designed for reliability at scale
- New crypto acceleration
- Advanced security capabilities
- Total Memory Encryption (TME)

Up to 1.58x Improvement in web microservices performance

Up to 40% Performance improvement (Specrate2017_int_base)
on new Ice Lake SKU offerings vs. Cascade Lake

Up to 1.42X More cores per processor
40-core Ice Lake vs. 28-core Cascade Lake

aws

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

Habana Gaudi-based Instances – DL1

ML Training Powered by New Habana Gaudi Processors from Intel

habana®
An Intel Company

New EC2 instances built specifically for ML training and powered by up to 8 new Habana Gaudi processors from Intel

Will deliver up to 40% lower cost-to-train deep learning models over GPU-based instances

Will allow customers to iterate and train models more frequently

Benefit from full stack of Amazon EC2 services—DL AMIs, DLC for containerized applications, ultimately SageMaker

Developers can implement Gaudi-based instances via AWS ECS and EKS for containerized applications

Will support common frameworks like TensorFlow and PyTorch

Wide range of ML workloads for applications including, NLP, image classification, object detection, recommendation systems

For efficient scaling across multiple Gaudi-based EC2 Instances, support for AWS Elastic Fabric Adapter

aws

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

Customer programs with Intel

HOW TO ENGAGE

- Participate in an Intel/AWS workshops at the AWS events
- Explore 16 Intel-optimized software libraries and frameworks on AWS Marketplace
- Learn from other customers like FootAsylum, Thorn, Zignal Labs, Krispy Kreme, GE, Thomson Reuters, ASIC: https://aws.amazon.com/solutions/case-studies/
- Connect with a partner – DataRobot, Data Bricks, IntellHQ, Intellify, Peak.AI, C3.AI, H2O.AI, Slalom

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

Success stories

THE UNIVERSITY of ADELAIDE

Analyzed 3 TB of Plant Breeding and Acclimatization Genomics Data in Hours to study the diversity of wheat.

CAREER LAUNCHER

Worked with educators to address the needs of Delhi's schools during Covid-19. Within two months a solution was was scaled to serve over 160,000 students.

CSIRO DATA 61

Australia's CSIRO research agency was able to increase by 60x the number of parallel wildfire simulations with 98% utilization of large Amazon EC2 C5-based instances.

THE UNIVERSITY OF SYDNEY

Researchers democratized data and drove ML models on genomic sequencing for endangered species.

fractal

Delivered more than 100 dashboards using anonymized government and public COVID-19 data points to prevent disease transmission.

GRAYMATICS

AI solution with the Intel® Distribution of OpenVINO™ toolkit Helps identify, detect, and respond in real time to hazards along Singapore's coastline.

aws

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

AWS Customer References

THORN and MEMSQL

Critical time saved

65%

Powered by Intel® Xeon® Scalable processors, Amazon EC2 C5 instances with Amazon S3 and Amazon Rekognition help law enforcement fight child trafficking.

18,119
Victims identified

5,791
Children identified

6,553
Traffickers identified

AI/ML
Image Analysis

"Spotlight helps officers identify child sex-trafficking ads much faster than the old paper-and-pencil methods."

Brooke Istook
Director of Strategy and Operations, Thorn

Thorn Finds More Human Trafficking Victims Faster

Empowering law enforcement to collaborate beyond jurisdictions.

Need: Abusers use advanced technology to facilitate their exploitation of children, and law enforcement needed to turn the tables and find these children faster.

Solution: Spotlight software ingests >100,000 online escort ads/day, storing them in Amazon S3. Amazon Rekognition in combination with MemSQL, a scalable database for operational analytics, helps find photos that have been edited to defeat image-search engines. Spotlight's ML models run on Amazon EC2 C5 instances, powered by Intel® Xeon® Scalable processors.

Value: Spotlight cut the time it takes by 65% to help the U.S. and Canada identify 18,119 victims—with 5,791 children and 6,553 traffickers identified.

Refer to https://aws.amazon.com/solutions/case-studies/thorn/ for details. Results may vary.

aws

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

AWS Customer References

AI/ML, Big Data
Online Marketing

8,400%
return on ad spend

The AI system runs on Amazon EC2 C5 instances, powered by Intel® Xeon® Scalable processors.

"Intel gives us the ability to scale, so we can give our AI solutions the computational power they need. This is critical for our business."

Mylo Portas,
Head of Retail, Peak AI

FOOTASYLUM

PEAK

Boost sales with AI-based customer personalization

Footasylum and Peak leverage AWS to dramatically increase return on ad spending.

Peak, an AWS Advanced Consulting Partner, implements AWS-powered solutions that enabled Footasylum to deliver advanced personalization in marketing communications via AI-powered customer segmentation.

Need: Deliver a personalized, connected retail experience to increase return on advertising and revenue via email marketing.

Solution: Customer purchasing data is stored in Amazon S3 and processed using Apache Spark running on Amazon EMR. The AI system, powered by Intel® Xeon® Scalable processors, runs on Amazon EC2 C5 instances to train its ML models and run predictions based on the models.

Value: Predictions and prescriptions target groups of customers in a way to achieve an 8,400% return on ad spend (ROAS), and a 28% revenue boost via email marketing.

Refer to https://aws.amazon.com/partners/success/footasylum/ for details. Results may vary.

aws

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

AWS Customer References

Public Sector + Machine Learning

Scalable and powerful

Amazon EC2 instances powered by Intel® Xeon® processors.

"It is indeed great to collaborate with tech giants such as Intel and AWS to help education stay uninterrupted in this period of crisis. We hope to take this partnership to a hundred such large projects globally and to be seen as practitioners of the do-good-do-well philosophy."

Satya Narayanan R
Chairman, Career Launcher

Career Launcher

Career Launcher Rapidly Scales Learning Portal in Pandemic

Expands to serve >160,000 students within two months on AWS.

Career Launcher (CL) is Asia's leading online education service provider and is led by a team of IIT-IIM alumni, who share a passion for education.
Need: A fast-scaling solution to ensure educational continuity during the COVID-19 lock-down.

Solution: A host of Amazon services for media processing and content delivery – Amaon Elemental, Cloudfront and more – were employed along with Amazon EC2 C and R family instances powered by Intel® Xeon® Scalable processors.
Value: Within two months, Project Aspiration 2020 was scaled to serve >160,000 students with powerful learning tools and the potential to accommodate many more.

Refer to link for more details.

aws

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

intel

3DAT

32% better price performance with Habana Gaudi DL1 instances

Advancing the Understanding of Human Motion

Products and Solutions

Discover Intel 3DAT Video

3DAT re:Invent track session

3DAT + AWS Sagemaker Blog

In preparation for the Olympic Games, Intel®, an American multinational corporation and one of the world's largest technology companies, developed a concept around 3D Athlete Tracking (3DAT). 3DAT is a machine learning (ML) solution to create real-time digital models of athletes in competition in order to increase fan engagement during broadcasts. Intel was looking to leverage this technology for the purpose of coaching and training elite athletes.

Classical computer vision methods for 3D pose reconstruction have proven to be cumbersome for most scientists, given that these models mostly rely on embedding additional sensors on an athlete and the lack of 3D labels and models. Although we can put seamless data collection mechanisms in place using regular mobile phones, developing 3D models using 2D video data is a challenge, given the lack of depth of information in 2D videos. Intel's 3DAT team partnered with the Amazon ML Solutions Lab (MLSL) to develop 3D human pose estimation techniques on 2D videos in order to create a lightweight solution for coaches to extract biomechanics and other metrics of their athletes' performance.

3DAT leverages Amazon's EC2 M6i instances based in 3rd Gen Intel Xeon Scalable processors to achieve an 18% better compute price performance. Using EC2 DL1 instances based on Habana Gaudi AI processors to improve training speed and efficiency of deep learning models. DL1 instances provided up to 32% improvement in price/performance.

[1]For more complete information about performance and benchmark results, visit https://www.intel.com/content/www/us/en/customer-spotlight/stories/berlin-institute-health-customer-story.html

aws

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

habana®

Additional Habana DL1 References:

leidos

"Given Leidos' and its customers need for quick, easy, and cost-effective training for deep learning models, we are excited to have begun this journey with Intel to use Amazon EC2 DL1 instances based on Habana Gaudi AI processors. Using DL1 instances, we expect an increase in model training speed and efficiency, with a subsequent reduction in risk and cost of research and development."

Chetan Paul, CTO Health and Human Sciences at Leidos."

fractal

"AI and deep learning are at the core of our Machine Vision capability, enabling customers to make better decisions across industries we serve. In order to improve accuracy, data sets are becoming larger and more complex, requiring larger and more complex models. This is driving the need for improved compute price-performance. The new Amazon EC2 DL1 instances promise significantly lower cost training than GPU-based EC2 instances. We expect this to make training of AI models on cloud much more cost competitive and accessible than before for a broad array of clients."

Srikanth Velamakanni, Group CEO of Fractal.

SEAGATE

"We expect the significant price/performance advantage of Amazon EC2 DL1 instances, powered by Habana Gaudi accelerators, could make a compelling future addition to AWS compute clusters. As Habana Labs continues to evolve and enables broader coverage of operators, there is potential for expanding to additional enterprise use cases, and thereby harnessing additional cost savings."

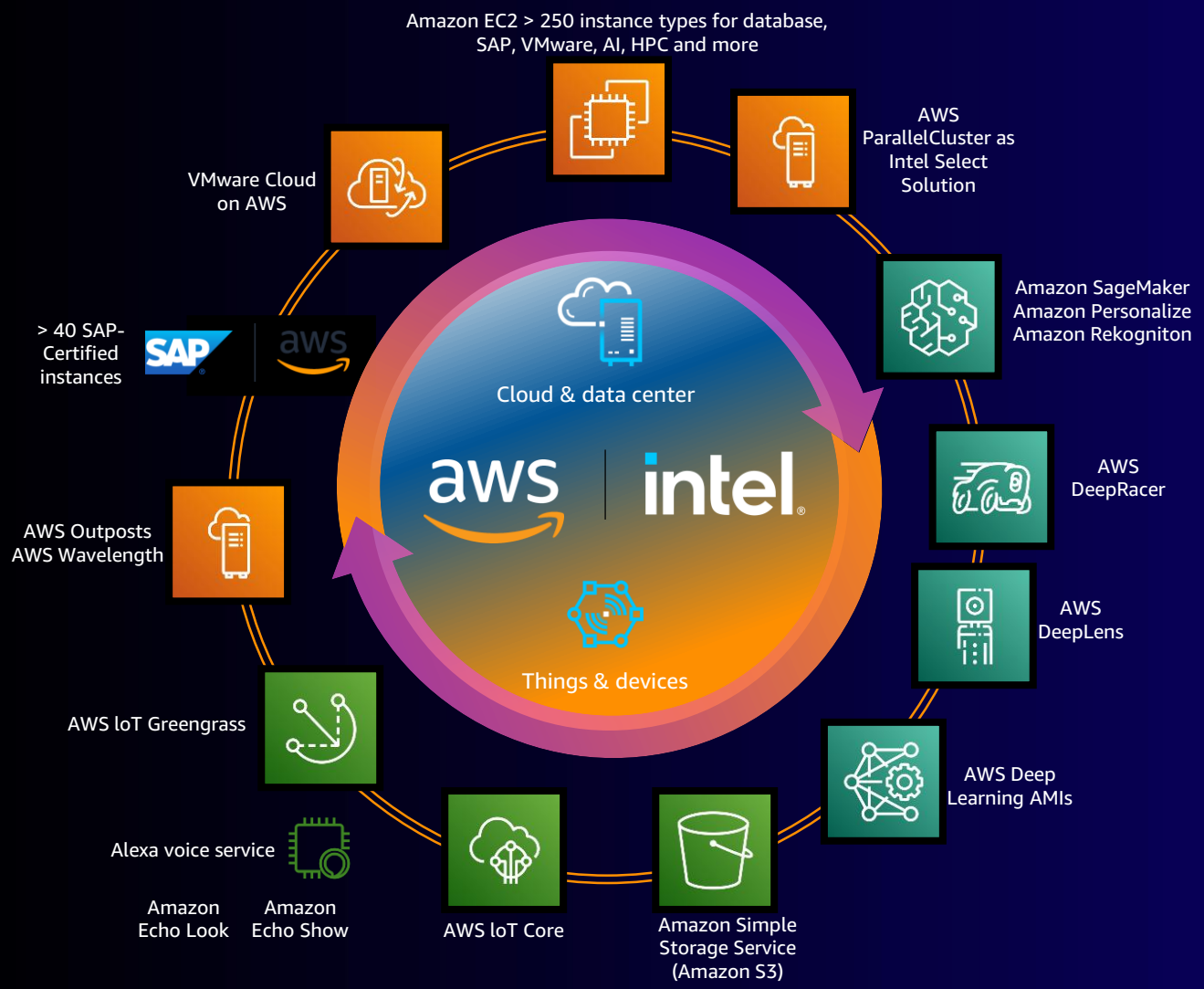
Darrell Louder, Seagate's Senior Engineering Director of Operations and Technology, Advanced Analytics, Darrell Louder. "

Habana DL1 Instance References

aws

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

Summary



- Close collaboration between Intel and AWS has resulted in best-in-class end-user experience and customer successes.
- Instance types with the best TCO on Intel to accelerate your customers' applications across a variety or workloads.
- Existing solutions for deployment with many successful outcomes delivering both high performance and cost savings.

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

Thank you for attending AWS Innovate – AI/ML Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apj-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws

aws

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

Thank you!

Akanksha Balani

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