INNOVATE

DATA EDITION

23 August, 2022

Solving business challenges with your data - powered by Intel on AWS

Akanksha Balani

AWS APJ Alliance head @ Intel
Global AI GTM Lead

Agenda

- AI Today
- Why Intel & AWS
- Intel AI on AWS
- Introducing Habana – Gaudi instance
- Learn | Engage | Innovate AI with Intel

Transformation with AI



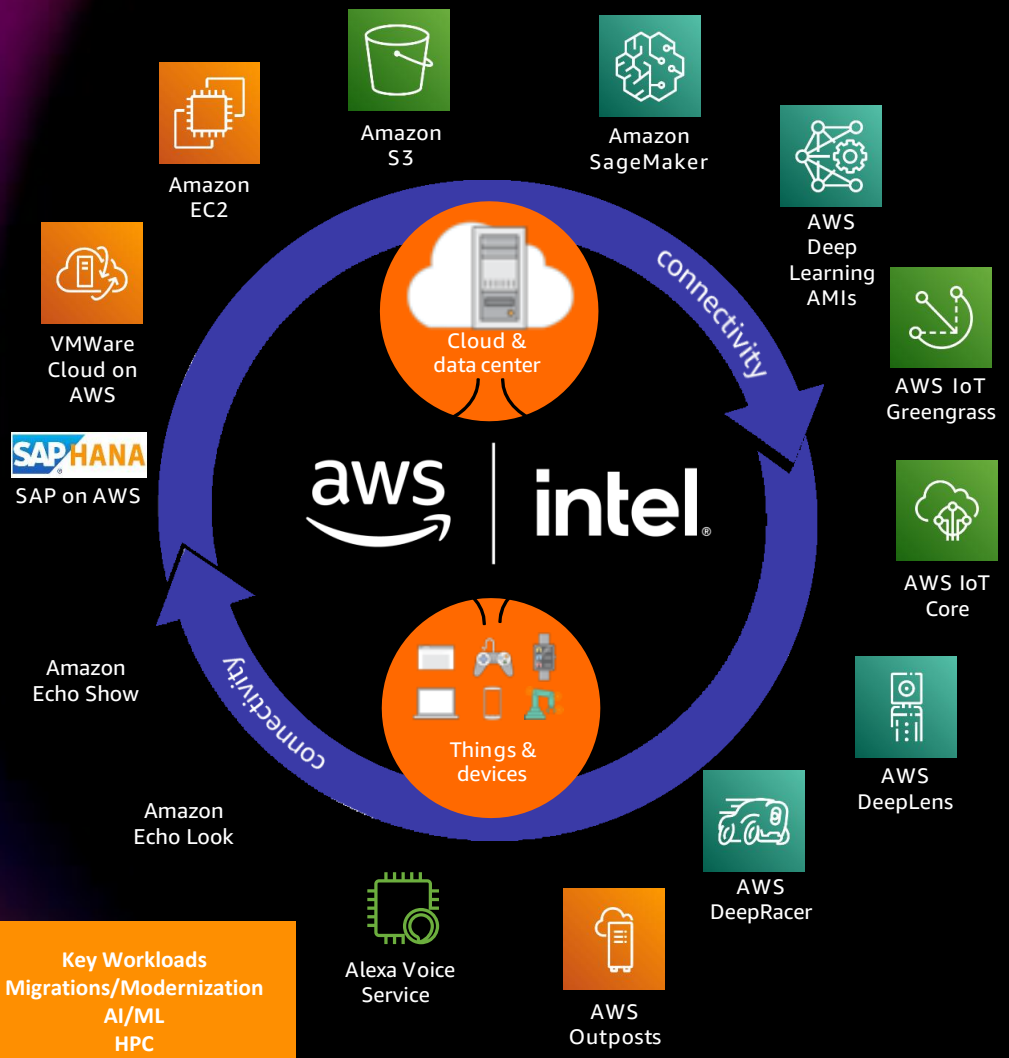






Consumer	Health	Finance	Retail	Government	Energy	Transport	Industrial	Other
Smart Assistants	Enhanced Diagnostics	Algorithmic Trading	Support	Defense	Oil & Gas Exploration	Autonomous Cars	Factory Automation	Advertising
Chatbots	Drug Discovery	Fraud Detection Research	Experience	Data Insights	Smart Grid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

What does Intel do with AWS?



Key Workloads
Migrations/Modernization
AI/ML
HPC
Hybrid/Edge

COMMON HISTORY AND VALUES

17 years of engineering partnership

Digital transformation

Shared customer passion

High performance + low costs

World-class supply chain

> *"Intel is a very deep partner of AWS and will be for a long time. That's not changing."*
> **Andy Jassy,** CEO, AWS

Greatest variety and availability to meet your global workload needs

aws | intel

275+ Intel instances

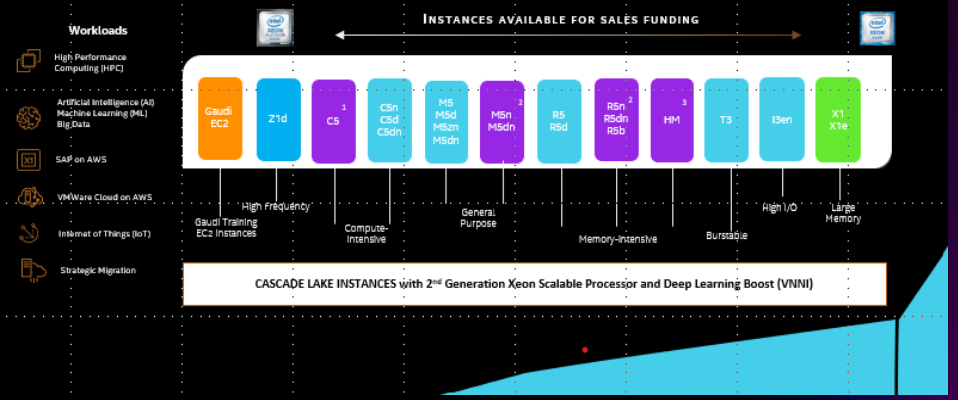
General purpose
T3 | M5 | M5n | M5zn | M5dn | **M6i**

Compute optimized
C5 | C5n | C5d | C5dn | **C6i**

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

Accelerated compute
Gaudi Instances | P3 | G4 | F1

Storage optimized
I3 | I3en | D3/D3en



aws

Instance Types on Intel

275+ Intel instances

	General Purpose	Compute Optimized	Memory Optimized	Accelerated Compute	Storage Optimized
	General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads.	Compute optimized instances are ideal for compute bound applications that benefit from high performance processors.	Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.	Accelerated computing instances use hardware accelerators, or co-processors, to perform functions more efficiently.	Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage.
Habana Gaudi				DL1	
Ice Lake	M6i	C6i	R6i		
Cascade Lake	M5n, M5dn, M5zn		R5n, Rdn, R5b	G4dn, P4d	D3, D3en
Skylake & Cascade Lake	T2	C5, C5d	R5, R5d, HMI		
Skylake	T3, M5, M5d	C5n	z1d	P3dn	I3en
Broadwell	M4		X1, R4	P2, P3, G3, F1	H1, I3
Haswell		C4	X1e		D2

Highlights of the past year

Career Launcher Rapidly Scales Learning Portal during Pandemic

Intel & AWS collaborate to help serve >160,000 students in India within two months on AWS.[1]

65x more parallel wildfire simulations

Intel & AWS work with RONIN to help increase fire fighting effectiveness in Australia. [4]

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/performance built on Habana Gaudi [3]

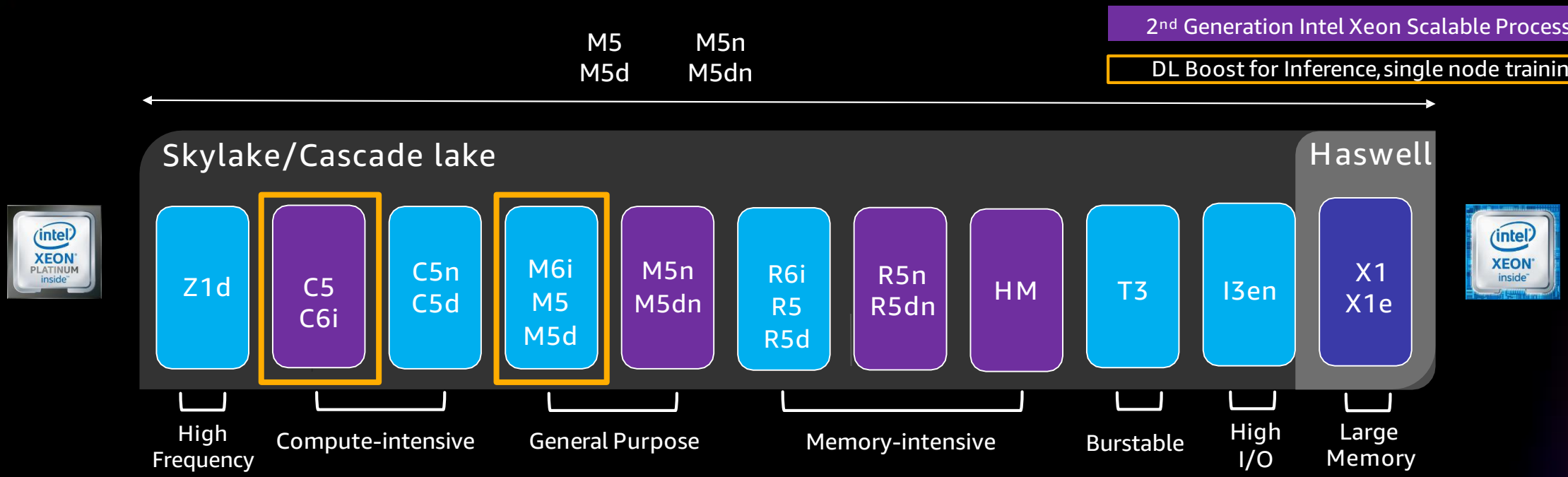
Intel's Habana & AWS co-engineered solution using up to 8 Gaudi accelerators

[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 | intel

aws

Intel-based Amazon EC2 Instances for ML



M5, C5 instances are suitable for all Computer Visions, ML and DL inference workloads.

R5 instances are for memory intensive workloads which use 3D-CNN/BERT-large/T5 topologies with memory requirement more than 192GB.

T3 instances are better suitable ML applications and low compute DL inference applications.

C5n instances are suitable for Distributed Deep learning training due to high NW performance required for inter-node communication.

Baremetal instances are preferred for large topologies as HVM based instances adds ~10% performance overhead.

More info: https://aws.amazon.com/ec2/instance-types/

New 3rd Gen 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

Introducing Habana Gaudi-based Instances – DL1

ML TRAINING POWERED BY NEW HABANA GAUDI PROCESSORS FROM INTEL

habana® An Intel Company

New Amazon 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 - AWS Deep Learning AMIs, Deep Learning Containers for containerized applications, ultimately Amazon SageMaker

Developers can implement Gaudi-based instances via Amazon ECS and Amazon 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 Amazon EC2 instances, support for AWS Elastic Fabric Adapter

aws

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
- Scikit-learn
- Pandas
- NumPy

R
- Cart
- RandomForest
- E1071

Distributed
- MLlib (on Spark)
- Mahout

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

[1] *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.a

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

Amazon 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

AI / ML customer programs with Intel

HOW TO ENGAGE

- Participate in an Intel/AWS Jam at the AWS events – NLP based challenge on Intel/AWS
- 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

AI Success Stories

THE UNIVERSITY of ADELAIDE

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

EnglishHelper

ReadToMe® software in digitally equipped government schools. Equipped with AI enabled multi-sensory technology that makes learning interactive for the students and enhances teacher effectiveness for grades 1 through 12.

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

Habana DL1 customer references

habana®

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

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 Amazon 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

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."

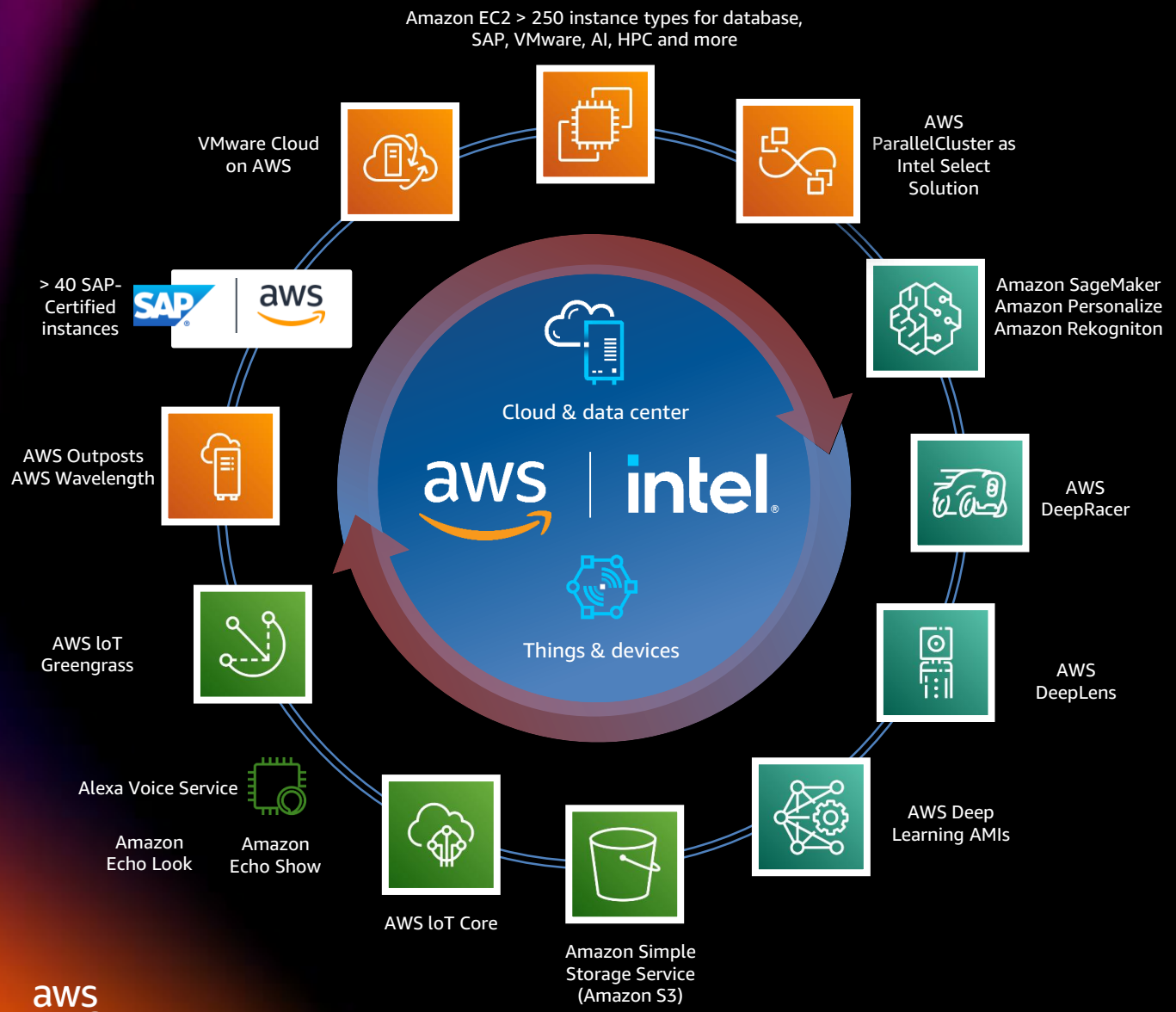
Chetan Paul, CTO Health and Human Sciences at Leidos

Learn more - Habana DL1 Instance references

aws

AWS and Intel - Better together

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.

Learn | Explore | Engage AI on Intel

Learn

More information at https://aws.amazon.com/intel/ on AWS & Intel

Explore

New instances based on Intel Xeon Scalable on AWS (M6i,C6i, R6i, M5, C5, C5n, R5, T3)

Engage

Contact your Intel/AWS representative for access to Intel AI and POC opportunities/case studies

CREATE WORLD CHANGING TECHNOLOGY THAT ENRICHES THE LIVES OF EVERY PERSON ON EARTH

aws

Thank you for attending AWS Innovate – Data 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

Thank you!

Akanksha Balani

AWS APJ Alliance head @ Intel
Global AI GTM Lead