



AWS powered by Intel. Co-engineered at the processor and firmware layer, optimized for the AWS Nitro Hypervisor — Intel Xeon 6 instances (C8i, M8i, R8i, X8i) built exclusively for AWS, with certified benchmarks and production references across every major workload.

2.5x

MEMORY BANDWIDTH
VS PRIOR
GENERATION

+60%

NGINX THROUGHPUT
ON C8i

+52%

SLM INFERENCE
THROUGHPUT

+45%

TOKENS PER DOLLAR

142,100

SAP ASAPS —
HIGHEST AT LAUNCH

17%

LOWER SAP
MONTHLY COST ON
R8i

SAP Modernization

R8i and M8i deliver **17% lower monthly compute cost and 27% faster transaction throughput** versus 4th Gen Intel. R8i achieved **142,100 aSAPS** — the highest SAP HANA certification at launch — with the broadest BTP integration on any cloud.

Live reference: Intel's own RISE migration, with SAP-certified benchmarks and sizing guidance.

VMware Transformation & Hybrid Cloud

Move from VMware or supply-constrained on-premises to Intel-based AWS **without refactoring**. Nutanix NC2 on AWS runs exclusively on Intel bare metal. AWS Outposts delivers **Intel Xeon 6 on-premises**.

Live reference: AT&T modernized IT and OSS/BSS on Intel Outposts SKUs. OEM displacement closes in weeks.

Generative AI & Enterprise Inference — CPU-First

C8i and R8i deliver **52% higher SLM inference throughput and 45% better tokens per dollar**. Intel AMX accelerates RAG pipelines, vector database queries, and enterprise assistants — **without GPU dependency**.

Live reference: Intel's iGPT migrated to Bedrock in five weeks; it now processes 8.8B tokens monthly.

EDA Modernization

R8i for memory-bound signoff, C8i for controllers, X2iezn for high-clock flows. **2.5x memory bandwidth reduces instance-hours per tape-out job**.

Live reference: Intel migrated 1.2M EDA cores to AWS — the largest EDA cloud transformation on record. Siemens EDA Calibre on R8i is the primary ISV reference.

HPC & Engineering Simulation

R8i and X8i for memory-intensive solves, C8i for meshing, X2iezn for high-clock flows. **elastic burst to the massive capacity of AWS eliminates queue time**.

Live references: JetZero runs blended-wing-body CFD with Altair on R8i targeting 50% fuel reduction. Hyundai Steel runs smart factory simulation on R8i and M8i.

Data & Analytics Modernization

R8i for ETL and analytics, C8i for orchestration. **2.5x memory bandwidth compresses batch windows** with no application changes.

Live references: BMW Cloud Data Hub, J&J genomics (35% faster analysis), JPMorgan Chase data mesh, CrowdStrike ML security.

Web, API & Cloud-Native Microservices

C8i delivers **60% higher NGINX throughput, driving 10–25% frontend run-rate reduction** on web and API tiers.

Live references: Disney+ runs realtime analytics and personalization on C8i and R8i. AT&T achieved 10–20% service-tier run-rate reduction.

GPU-Adjacent AI Training

P4d, P5, P5e, and P5en pair NVIDIA GPUs with Intel Xeon Scalable host CPUs. P5en delivers **35% lower collective communications latency** through Granite Rapids with PCIe Gen5 — reducing GPU starvation at scale.

Where it matters: data staging and preprocessing for distributed AI training at scale.

✓ AWS-Exclusive Intel Xeon 6

✓ Certified Benchmarks per Workload

✓ Production References Across Industries

✓ No Refactoring Required

TECHNICAL DIFFERENTIATORS — MEASURED, CERTIFIED, IN PRODUCTION

CAPABILITY	WHAT THE DATA SHOWS
Custom Silicon	R8i, M8i, C8i run Intel Xeon 6 SKUs exclusive to AWS — co-engineered at processor, firmware, Nitro, and hypervisor layers. Not available in any other cloud or on-premises catalog.
Memory Bandwidth	2.5× greater than prior generation. Reduces instance-hours for SAP HANA, EDA signoff, large analytics, and inference preprocessing — no application changes.
Intel AMX (GenAI)	52% higher SLM inference throughput, 45% better tokens per dollar. C8i and R8i run RAG, vector DB, and AI assistants without GPU dependency.
SAP HANA	R8i: 17% lower monthly cost, 27% faster throughput vs 4th Gen Intel. 142,100 aSAPS. Broadest BTP integration on any cloud.
NGINX	+60% throughput on C8i vs 4th Gen Intel — drives 10–25% frontend compute run-rate reduction on web and API tiers.
PostgreSQL	+30% throughput on 8th Gen Intel Xeon 6 vs prior generation. No schema or application changes required.
DL Recommender	+40% deep learning recommender model throughput on 8th Gen vs prior generation Intel Xeon.
PCIe Gen5 (P5en)	35% lower collective communications latency. Granite Rapids host CPU reduces GPU starvation in distributed AI training.
Z1d Clock Speed	Up to 4.0 GHz all-core sustained turbo — for EDA signoff, financial modeling, and single-threaded latency-critical workloads.
AWS Outposts	2nd Gen racks ship Intel Xeon 6 on-premises (C8i, M8i, R8i) — the same silicon, the same instances, in your data center.

What Intel brings to the engagement

- › **Dedicated Intel engineering support** and joint account planning
- › **SAP-certified benchmarks** and HANA sizing guidance
- › **EDA ISV relationships:** Synopsys, Cadence, Siemens EDA
- › **HPC ISV ecosystem:** Altair, Ansys, Siemens Simcenter
- › **GenAI reference architecture** for CPU-first RAG and vector DB
- › **VMware / Nutanix NC2** bare metal TCO modeling
- › **iGPT on Bedrock:** 8.8B tokens/month, migrated in 5 weeks
- › **1.2M EDA core migration** as live proof point

Intel x86 AWS services by category

- EC2 COMPUTE & GENERAL PURPOSE**
C6i, C7i, C8i, C8i-flex · M6i, M7i, M8i, M8i-flex · EC2 Linux, Windows, RHEL, Spot, Auto Scaling
- MEMORY, STORAGE & HPC**
R6i, R7i, R8i, R8i-flex · X2idn, X2iedn, X2iezn, X8i, Z1d · I7i, I7ie · Hpc6id · P4d, P5, P5e, P5en
- DATABASES**
RDS on M6i/M7i, R6i/R7i, X2idn/Z1d · DocumentDB · ElastiCache (m5/m6i/r5/r6i)
- CONTAINERS & KUBERNETES**
ECS · EKS · EKS Auto Mode · EKS Anywhere · ECR
- ANALYTICS & STREAMING**
EMR · AWS Glue · MSK · Redshift RA3 · OpenSearch · CodeBuild
- STORAGE**
EBS · EFS · FSx for Lustre, NetApp ONTAP, OpenZFS, Windows File Server
- HYBRID & HPC**
AWS Outposts (2nd Gen: C8i/M8i/R8i) · ParallelCluster · MQ

Every workload listed ships with a certified benchmark, sizing guidance, and a named production reference. Ask your AWS account team to bring Intel into the conversation. Intel funds benchmark validation and TCO modeling on your actual workloads before any commitment.