OpenCloud ARM Compute

Modern Compute for AI Inference and High-Performance

OpenCloud ARM Compute extends the US Signal OpenCloud platform with high-efficiency Ampere® ARM processors built for today's AI-driven and CPU-intensive workloads. Designed for AI inference, Small Language Models (SLMs), Retrieval-Augmented Generation (RAG), APIs, and scale-out applications, it delivers predictable performance, exceptional energy efficiency, and cloud flexibility—without the cost and complexity of GPU infrastructure.

As AI moves from experimentation to production, many organizations are discovering that not every workload requires GPU acceleration. While GPUs remain essential for training large language models, production inference, application serving, and other CPU-driven workloads demand consistent performance, high core density, and cost efficiency. Powered by the AmpereOne® platform and delivered through OpenCloud's flexible consumption model, OpenCloud ARM Compute enables organizations to accelerate AI adoption, optimize infrastructure, and scale confidently while maintaining enterprise-grade security, transparency, and operational control.

Platform Highlights

Feature	Benefit
AmpereOne® ARM Processors	High-performance ARM architecture optimized for cloud workloads
Up to 192 CPU Cores	High-density compute for parallel processing
AI Inference Ready	Optimized for production AI workloads without GPU overhead
Consumption-Based Billing	Pay only for the compute resources you use
Enterprise OpenCloud Platform	Secure, reliable, and scalable infrastructure
US-Based Support	Expert assistance from US Signal cloud specialists
100% Uptime SLA	Enterprise-grade reliability and availability

Idea Workloads

Artificial Intelligence

- AI inference
- Small Language Models (SLMs)
- Retrieval-Augmented Generation (RAG)
- Vector search
- Recommendation systems

Cloud Applications

- High-throughput APIs
- Containerized applications
- Microservices
- Web services
- Application serving

Compute-Intensive Workloads

- CPU-bound processing
- Scale-out applications
- Parallel workloads
- Data processing services

info@ussignal.com **ussignal.com** **866.2.signal**

High Core Density

Powered by AmpereOne® processors featuring up to 192 cores, OpenCloud ARM Compute supports demanding multi-threaded and scale-out workloads while offering flexible instance sizing for a variety of applications.

Flexible Cloud Consumption

OpenCloud ARM Compute enables organizations to:

- Scale resources on demand
- Pay only for consumed resources
- Eliminate over-provisioning
- Deploy workloads quickly

Optimized for AI Inference

Purpose-built for inference workloads including:

- Question answering
- Text summarization
- Recommendation engines
- Semantic search
- Retrieval-Augmented Generation (RAG)
- AI-powered APIs and microservices

Enterprise Cloud Platform

Every ARM Compute instance is backed by the same trusted OpenCloud foundation, including:

- Enterprise-grade security
- Compliance-ready infrastructure
- US-based technical support
- 100% uptime SLA

Exceptional Performance per Watt

Industry-leading energy efficiency helps organizations:

- Reduce power consumption
- Lower operational costs
- Improve infrastructure efficiency
- Support sustainability initiatives

Predictable Performance

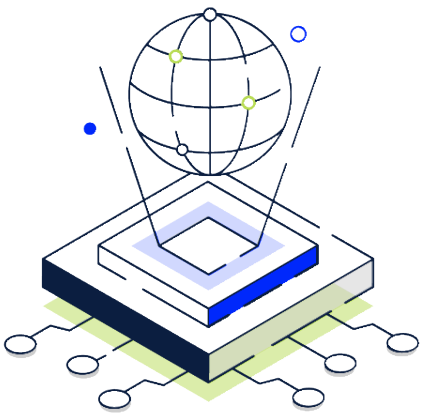
Ampere's high-throughput architecture delivers:

- Consistent latency
- Dedicated CPU resources
- Reliable application performance
- Efficient scaling under sustained utilization

Build Smarter AI Infrastructure

OpenCloud ARM Compute provides an efficient foundation for modern AI and cloud applications. With high-performance Ampere processors, flexible cloud consumption, predictable scaling, and enterprise-grade reliability, organizations can deploy AI inference and application workloads confidently while optimizing cost and efficiency.

Speak with a US Signal cloud expert today.



info@ussignal.com **ussignal.com** **866.2.signal**