Cirrascale stands at the forefront of cloud innovation, providing purpose-built infrastructure solutions designed to meet the demanding requirements of modern AI, machine learning, inference, and high performance computing workloads.

Our AI Innovation Cloud offers cost-effective, high-speed cloud solutions with the features you need. As a specialized cloud services provider, we deliver unmatched performance, scalability, and expertise to organizations pushing the boundaries of what's possible in AI and computational technology.

Core Offerings and Services

The Cirrascale AI Innovation Cloud

Our flagship AI Innovation Cloud is a leading platform for AI development and deployment. Unlike traditional cloud providers, Cirrascale Cloud Services offers:

- The only cloud service providing access to every leading AI accelerator, including NVIDIA, AMD, Cerebras, Qualcomm, and more.
- Purpose-built infrastructure optimized for Generative AI, LLMs, and complex AI workloads.
- Industry-leading power capacity supporting the highest-density GPU configurations.
- Dedicated, bare-metal servers ensuring maximum performance without virtualization.

Cirrascale Inference Platform for Enterprise

Optimized for real-time AI inferencing, our platform ensures low-latency, high-efficiency performance at enterprise scale.

- NVIDIA Blackwell B200, B40, and H200 GPUs at debut.
- Dynamically balances workloads across regions.
- Serverless deployments that provide an instant AI Model Pipeline.
- Pre-compiled foundational models optimized for underlying accelerators.
- Web console-based deployment and configuration.
- API Access to the AI model utilizing OpenAI API's.
- Supports fine-tuned models and custom weights.
- Priced by token volume for the given AI Model used in the pipeline.

Advanced Cloud Infrastructure

Cirrascale's AI Innovation Cloud infrastructure is engineered for performance:

- 25Gb/50Gb/100Gb Ethernet or 200Gb HDR and 400Gb NDR NVIDIA Quantum-2 InfiniBand.
- Lightning-fast NVMe hot-tier storage solutions.
- Interconnect to your existing infrastructure, either on-premises or with a hyperscale cloud provider like AWS, GCP, Azure or Oracle.
- Kubernetes and cluster management expertise.
- Comprehensive network management including DNS and out-of-band management.

No-Surprises Cloud Services Billing

Transparent pricing that lets you focus on innovation:

- Simple, flat-rate billing model.
- No hidden ingress or egress fees.
- Predictable costs for long-term planning.
- Discounts for extended and large commitments.

Premium Managed Services

For enterprises requiring end-to-end cloud management, our white-glove services delivers comprehensive support.

- 2N power redundancy with A+B fully managed service.
- Enterprise-grade security with access control.
- SSAE-16 / ISAE, SOC1 Type 2 and SOC2 Type 2 certified secure environments.
- Expert installation, monitoring, and break/fix support.
- Streamlined deployment with automated provisioning capabilities.

Strategic Technology Partnerships with Today's Latest AI Accelerators

The Cirrascale AI Innovation Cloud stands apart as the only cloud service where customers can access and deploy workloads across every leading AI accelerator technology in a single environment. Through strategic partnerships with industry leaders, we deliver a comprehensive ecosystem of accelerator options including the NVIDIA B200 and H200 TensorCore accelerators, the AMD Instinct MI300x and MI350x accelerators, the Cerebras CS-2 WSE, and the Qualcomm Cloud AI 100 Ultra. This diverse portfolio ensures customers can test, optimize and select the best architecture for their specific workloads — from training to fine-tuning to inference — without being locked in to a single vendor's ecosystem. That's the Cirrascale advantage.

Industry Solutions

Generative AI and Large Language Models

- Purpose-built infrastructure for training and inferencing generative AI models.
- Specialized solutions for pre-training, fine-tuning, and inference.
- Support for leading frameworks including PyTorch, TensorFlow, JAX, ONNX and more.
- Optimized for popular models like GPT, BERT, and T5 variants and any other models available on Hugging Face.

Computer Vision and Autonomous Systems

- Optimized platforms for image classification, object detection, and tracking.
- High-performance infrastructure for robotics and autonomous system development or validaton.
- Scalable solutions for training and deploying computer vision models.

High Performance Computing

- Advanced computational resources for scientific research and simulation.
- Seamless integration of traditional HPC with modern AI capabilities.
- Support for complex data analytics and visualization workflows.

NVIDIA Inception Members

Cirrascale Cloud Services helps NVIDIA Inception members get the most out of their program benefits. Whether you're a community or premier member, we can assist in extending your benefits.

Cirrascale has been helping NVIDIA Inception program members evolve faster by providing them with unique benefits. By partnering with Cirrascale, approved Inception members can purchase hardware at preferred pricing levels, receive discounts on cloud or managed services provided by Cirrascale, and partner with us to receive referral discounts for other startups you refer.

Cirrascale Cloud Services works closely with NVIDIA and your assigned dedicated relationship manager to provide you with the best experience as an Inception member.

Need Specialized Services?

At Cirrascale, we take pride in being flexible and receptive to our customer's individualized needs. We understand that not every customer has the same requirements. If you need some additional services that are not discussed above, please reach out to us.

Cirrascale®

CSM001 - REV M - 3/2025