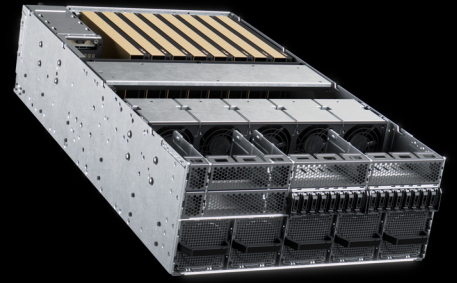




NVIDIA Aerial RAN Computer MGX 4U

High-density distributed AI infrastructure for centralized AI-RAN and AI-centric deployments.



NVIDIA Aerial RAN Computer MGX 4U

NVIDIA Aerial™ RAN Computer (ARC) MGX 4U is a modular, high-performance AI-RAN platform in a 4U form factor. Designed for AI-centric workloads, it supports centralized AI-RAN (C-RAN) and AI-only deployments at aggregated sites, where operators can concentrate compute capacity and serve compute demands across multiple locations.

With up to 8 NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs, ARC MGX 4U provides capacity for demanding AI inference workloads. In AI-RAN deployments, operators can run AI services alongside software-defined RAN functions. In AI-only deployments, they can dedicate the platform's compute resources to AI applications.

Built for High-Density AI

The NVIDIA ARC MGX 4U helps operators turn aggregation sites into AI compute locations. Its GPU capacity supports larger models and more concurrent inference workloads, creating opportunities to offer AI services through existing network infrastructure.

AI capabilities include:

- > **AI service delivery** by processing AI traffic from AI agents, generative-AI-based video and voice applications, AR glasses, and more
- > **Processing enterprise AI use cases** across robotics, smart cities, cybersecurity, video search and summarization, and computer vision solutions
- > **AI-on-RAN** for hosting edge AI applications for low-latency inference and analytics
- > **AI-for-RAN** for using advanced AI and machine learning (AI/ML) algorithms and neural network models for RAN optimization, including massive MIMO, beamforming, channel prediction, power management, scheduling, and AI-powered dApps, rApps, and xApps
- > **Sensor fusion** and integrated sensing and communication (ISAC) workloads

5G/6G RAN capabilities include:

- > **Support for current 5G standards**, including frequency-division duplexing (FDD) and time-division duplex (TDD) multiple input, multiple out (MIMO) (e.g., 4 transceiver, or TR, units) and TDD massive MIMO (e.g., 64TR)
- > **Integrated high-speed networking** with precise timing interfaces for fronthaul, midhaul, and backhaul connectivity

Key Features

- > NVIDIA MGX™ modular architecture in a 4U form factor
- > Up to 8 NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPU acceleration
- > Support for x86 CPUs
- > NVIDIA ConnectX™-8 networking
- > Deployable into COTS servers by OEMs
- > Designed for AI-centric, high-density deployments
- > 5G FDD/TDD and massive MIMO today; ultra MIMO with software upgrade to 6G
- > Advanced AI/ML algorithms, edge AI applications and services, and sensor fusion

- > Full software stack compatibility for virtualized RAN, Open RAN, and Cloud RAN architectures
- > 6G readiness, including support for future 6G features, including new spectrum bands, AI-assisted radio processing, and higher order MIMO

Create More Value at Aggregation Sites

By concentrating AI compute at aggregation sites, operators can serve demand across a wider network area and make AI capacity available for new services. ARC MGX 4U supports two deployment paths: Combine AI and RAN workloads on shared infrastructure, or allocate the platform to AI-only services.

Technical Specifications

Architecture	NVIDIA MGX
GPU	Up to 8 NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs with Multi-Instance GPU (MIG) technology; up to 4 MIG instances for each GPU
CPU	x86 CPUs
Ethernet	Up to 4 ConnectX-8 SuperNIC™ (1x 800GbE, 2x 400GbE) for East/West and fronthaul traffic; 1x BlueField™-3 (B3220) for North/South traffic
CUDA™ Cores	NVIDIA Blackwell generation
Tensor Cores	Fifth generation
Ray-Tracing Cores	Fourth generation
AI TOPS	4 PFLOPS of FP4 AI performance per GPU
Software Stack	NVIDIA AI Aerial libraries, CUDA, DOCA™, cuDNN™, TensorRT™ AI-RAN reference applications; DOCA dUPF Third-party vRAN/O-RAN-compliant software
RAN Technology	Multi-RAT, upgradable to 6G, both FDD and TDD
RAN Performance	Standard MIMO (e.g., 4TR), massive MIMO (e.g., 64TR), ultra MIMO (e.g., 256TR)
Power Consumption and Energy Efficiency	1.6KW–7KW based on workload and GPUs
RAN Software Architecture	GPU-accelerated and software-defined L1 and L2
Environmental	10–35°C (50–95°F) with no direct sunlight on the equipment (operational)
Form Factor	4U chassis

Ready to Get Started?

To learn more about NVIDIA AI Aerial, visit:
developer.nvidia.com/aerial

© 2026 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, Aerial, BlueField, ConnectX, CUDA, cuDNN, DOCA, MGX, RTX PRO, SuperNIC, and TensorRT are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective owners with which they are associated. 5776627. OCT26

