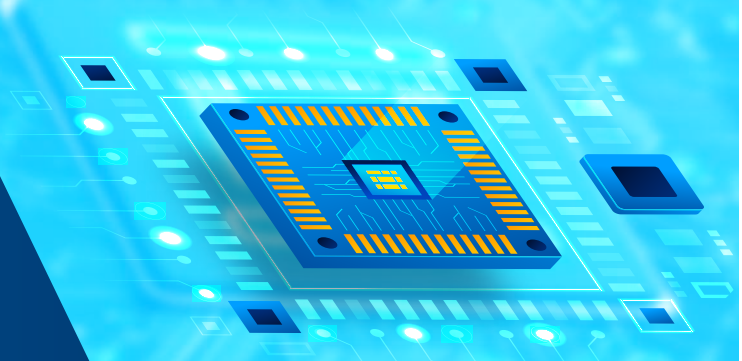
JioCloud

GPU Virtual Machine AMD



Overview

JioCloud GPU Virtual Machine AMD offers powerful, scalable virtual machines built on AMD MI300x GPUs. Designed for deep learning, HPC, and data processing, these VMs help you train large models, process unstructured data, and run simulations faster—without the setup burden of physical infrastructure. Each VM includes pre-installed RoCM drivers and AI frameworks like PyTorch and TensorFlow, so teams can get started immediately. You can choose from flexible pricing options, manage the entire VM lifecycle on your own, and monitor performance in real time—all while scaling from one to eight GPUs as your needs grow.

Key Features

- **AMD MI300x GPU power**
  High memory bandwidth and compute performance optimised for AI/ML and HPC workloads.
- **Preinstalled RoCM drivers**
  Ready-to-use GPU environment supporting popular deep learning libraries.
- **AI-optimized VM setup**
  Ideal for LLMs, computer vision, NLP models, and training large datasets.
- **HPC and simulation support**
  Accelerate scientific workloads, engineering models, and climate simulations.
- **Flexible pricing plans**
  Available as on-demand or reserved instances - monthly or long-term.
- **End-to-end monitoring**
  Track GPU usage, memory bandwidth, temperature, and system metrics in real time.

JioCloud

Benefits

- Train deep learning and LLM models faster, reducing time to production.
- Accelerate simulations and data processing for research and analytics.
- Deploy AI environments without time-consuming setup.
- Scale up or down between 1 and 8 GPUs based on project needs.
- Manage usage and costs through flexible billing options.
- Optimize performance with built-in observability and metrics.

Technical Specifications

Category	Details
GPU Architecture	AMD CDNA 3
Memory	128 GB HBM3
Compute Units	Upto 304
Memory Bandwidth	Up to 3.2 TB/s
GPU Interconnect	Infinity Fabric, XGMI
Compute Performance	TFLOPS (FP64/FP32/FP16)
Supported OS	Ubuntu
AI Frameworks	PyTorch, TensorFlow
Storage	100 GB

Use Cases

- **Retail personalization**
  Train recommendation engines quickly using customer behavior data for real-time, personalized shopping experiences.
- **Medical research**
  Speed up drug discovery and genomic analysis with GPU-accelerated AI models for simulation and prediction.
- **Climate modeling**
  Run large-scale environmental simulations with faster training and inference cycles for better forecasting.
- **Log chatbots for RCA**
  Use AI chatbots to analyze large volumes of logs in plain English, reducing RCA time and speeding up issue resolution.

JioCloud

Architecture Diagram