



Voltage Park Dedicated Reserve Cloud

Overview

Voltage Park's Dedicated Reserve Cloud product is a cost-effective and high-performance GPU leasing solution custom-designed for AI companies, research institutions, and enterprises that require dedicated, uninterrupted compute capacity for extended durations. Voltage Park ensures predictable access to cutting-edge NVIDIA H100 GPUs, mitigating the risks of cloud compute shortages and price volatility. This solution is ideal for organizations with consistent, long-running AI workloads, providing a reliable and scalable platform for machine learning model training, research, and development.

Key Benefits

- **Guaranteed Availability** – Secure your AI compute resources in advance, ensuring uninterrupted model training and experimentation, eliminating concerns about resource contention or sudden price surges.
- **High-Performance Infrastructure** – Access to 24,000+ NVIDIA H100 GPUs (80GB SXM5) interconnected with 3.2 Tb/s InfiniBand for industry-leading training speeds, enabling rapid iteration and faster time-to-market for AI models.
- **Cost Optimization** – Long-term commitments provide significant savings compared to on-demand pricing, offering a predictable cost structure that aligns with budget planning and resource allocation.
- **Scalability and Flexibility** – Choose from flexible reservation blocks (e.g., 3 months, 6 months, or annualized contracts) to fit project timelines, ensuring that compute resources match the specific needs and duration of AI initiatives.
- **Bare-Metal Performance** – Direct access to GPUs with no virtualization overhead, maximizing efficiency and performance for compute-intensive AI workloads.
- **Sustainable Compute** – Voltage Park's Seattle data center in Washington State runs entirely on 99% renewable hydro-electric power, reducing environmental impact and lowering operational costs, contributing to corporate sustainability goals.
- **24x7 on-site, anytime support** – Dedicated on-site customer service ensuring real-time assistance and troubleshooting, no matter the time zone, minimizing downtime and maximizing productivity.

Product Specifications

Hardware and Networking

- **GPU Model:** NVIDIA H100 (80GB, SXM5), the latest generation of high-performance GPUs designed for AI and deep learning workloads.
- **Cluster Size:** Configurable from tens to thousands of GPUs, providing the flexibility to scale compute resources to match the size and complexity of AI models.
- **Interconnect:** 3.2 Tb/s InfiniBand, ensuring ultra-fast multi-node training and efficient communication between GPUs, critical for large-scale distributed AI workloads.
- **CPU Complement:** High-core-count x86 processors for data preprocessing and other CPU-bound tasks, ensuring balanced system performance and efficient utilization of resources.
- **Storage:** High-speed NVMe and networked storage options available, providing ample capacity and fast data access for AI model training and inference.
- **Power and Cooling:** Optimized for high-density AI workloads, ensuring efficient power delivery and thermal management for maximum performance and reliability.

Software and Compatibility

- **Supported ML Frameworks:** PyTorch, TensorFlow, JAX, DeepSpeed, and other popular machine learning frameworks, providing flexibility and compatibility with a wide range of AI development tools and libraries.
- **Orchestration Tools:** Kubernetes, Slurm, MosaicML for efficient workload scheduling and resource management, simplifying the deployment and management of AI workloads at scale.
- **Multi-Node Training:** Optimized for distributed AI model training at scale, enabling efficient parallelization and faster training times for large and complex AI models.
- **Flexible Environment Support:** Ability to bring custom ML stacks or leverage pre-configured environments, providing flexibility and customization options to meet the specific needs of AI development teams.

Pricing and Reservation Model

- **Hourly Rate Discounts:** Pricing varies based on reservation length, with longer reservations offering greater discounts. This provides cost predictability and allows organizations to optimize their AI compute spending based on their project timelines.
- **Enterprise Contracts:** Custom agreements for long-term, high-volume compute requirements, offering tailored solutions for organizations with large-scale and ongoing AI initiatives.

Use Cases

- **AI Model Training** – Large-scale LLM and diffusion model training requiring high-performance GPUs, enabling the development of advanced AI models for natural language processing, computer vision, and other applications.
- **Enterprise AI Workloads** – Continuous AI-powered analytics, simulations, and automation, providing a dedicated and scalable platform for running AI applications in production environments.
- **Research and Academia** – Universities and labs needing dedicated AI clusters for multi-month experiments, supporting advanced research and development in AI and machine learning.
- **Cloud Compute Replacement** – Cost-efficient alternative to hyperscaler on-demand GPU instances, offering a more predictable and potentially cost-effective solution for organizations with consistent AI compute needs.

Why Voltage Park?

- **Tech Leader Adoption:** Already supporting Cursor, Imbue, Character.ai, CalTech, and Atomic AI, demonstrating reliability and market trust among leading AI companies.
- **Strategic Data Center Locations:** Deployments in Texas, Virginia, and Washington State, ensuring low-latency, high-availability compute across multiple regions.
- **Transparent, Fair Allocation:** No opaque bidding wars—competitive block leasing ensures equitable access to compute resources, promoting fairness and transparency in the allocation process.

Contact and More Information

For pricing, availability, and custom lease terms, visit voltagepark.com or email sales@voltagepark.com.