

Global GPU Inference Server Supply, Demand and Key Producers, 2026-2032

<https://marketpublishers.com/r/GF09B4D54626EN.html>

Date: August 2026

Pages: 175

Price: US\$ 4,480.00 (Single User License)

ID: GF09B4D54626EN

Abstracts

The global GPU Inference Server market size is expected to reach \$ 77247 million by 2032, rising at a market growth of 22.0% CAGR during the forecast period (2026-2032).

A GPU inference server is a high-performance server system or rack-scale computing platform built around data-center GPUs or AI accelerator cards for online deployment, batch inference, and real-time inference of trained models. It addresses the requirements of large language models, multimodal models, recommendation models, and vision models in enterprise private deployments, cloud services, and edge nodes, with a focus on low-latency response, high-concurrency throughput, long-context processing, compute utilization, energy efficiency, and operational reliability. These systems typically integrate multiple CPUs, multiple GPUs or AI accelerators, high-bandwidth memory, high-speed NVMe storage, RDMA networking, high-speed GPU interconnects, redundant power supplies, air or liquid cooling, and system management software, while using PCIe, NVLink, NVSwitch, HGX, MGX, OAM, or domestic accelerator platforms to deliver different levels of acceleration capability. Typical customers include cloud service providers, internet platforms, financial institutions, manufacturers, research organizations, telecom operators, governments, and data center operators. Common delivery formats include 2U to 10U rack servers, eight-GPU high-density nodes, liquid-cooled systems, rack-scale AI factories, private cloud appliances, and cloud-rented clusters. Their commercial value is concentrated in reducing cost per token, improving model serving stability, shortening AI application deployment cycles, and supporting localized enterprise data deployment.

GPU inference servers are evolving from ?servers configured with high-end GPUs? into system-level infrastructure for production AI services. Their evaluation criteria no longer focus only on peak compute performance, but also include low latency, high

concurrency, long-context processing, memory capacity, GPU interconnects, network throughput, storage access, cooling capacity, and operations software. Large language models and agentic applications are shifting inference requests from simple question answering to continuous tasks involving multiple steps, multiple models, and tool calls, increasing dependence on KV cache, context windows, and network communication. As a result, eight-GPU HGX nodes, NVLink and NVSwitch interconnects, RDMA networking, GPU Direct Storage, liquid cooling, and certified software stacks are becoming essential components of high-end systems. Benchmarks such as MLPerf Inference strengthen the comparability of system-level inference performance and encourage vendors to move from hardware stacking toward joint optimization across hardware, software, networking, and energy efficiency. Over the long term, the value of GPU inference servers will be reflected more in cost per token, service stability, cluster utilization, and deployment speed than in the purchase price of a single server.

The global supply landscape shows clear ecosystem layering. U.S. companies remain strong in GPUs, branded systems, AI software stacks, and high-end server platforms. NVIDIA defines much of the underlying architecture for inference servers through HGX, DGX, certified systems, and the AI Enterprise software ecosystem, while Dell, HPE, Cisco, and Supermicro convert these platforms into enterprise-grade systems, racks, and data center solutions. Mainland Chinese companies are building differentiated paths around government and enterprise private deployment, domestic AI accelerator adaptation, and local computing platforms, with Huawei Atlas, IEIT Systems? YuanNao, H3C UniServer, and xFusion product lines covering central inference, model development, vertical AI, and large-scale clusters. Taiwanese companies play a key role in ODM design, manufacturing, liquid cooling integration, and HGX and MGX platform adoption, with QCT, Wiwynn, Inventec, ASUS, GIGABYTE, and Pegatron supporting server supply for global cloud service providers and branded vendors. Japanese and Korean vendors are more focused on local delivery, industry system integration, and regional market services, forming a complementary competitive layer.

Demand-side growth is being driven by cloud service provider expansion, enterprise private deployment, vertical digital transformation, and edge real-time inference. Large North American cloud service providers remain the largest customers for high-end GPU inference servers, primarily for generative AI, recommendation systems, search advertising, coding assistants, and multimodal content services. In China, large-model applications, enterprise and government data localization, and domestic computing policies are jointly pushing inference servers into finance, telecom, government, manufacturing, healthcare, education, and research scenarios. Europe, Japan, Korea,

and the Middle East are forming demand around sovereign AI, local enterprise AI, manufacturing automation, and newly built AI data centers. As inference call volume grows faster than training workloads, compute deployment will expand from a small number of ultra-large training clusters toward more distributed inference nodes closer to business systems. Future competition will center on energy efficiency, delivery speed, liquid cooling readiness, software ecosystems, cross-GPU resource scheduling, and lifecycle services, and the industry outlook remains broadly positive.

This report studies the global GPU Inference Server production, demand, key manufacturers, and key regions.

This report is a detailed and comprehensive analysis of the world market for GPU Inference Server and provides market size (US\$ million) and Year-over-Year (YoY) Growth, considering 2025 as the base year. This report explores demand trends and competition, as well as details the characteristics of GPU Inference Server that contribute to its increasing demand across many markets.

Highlights and key features of the study

Global GPU Inference Server total production and demand, 2021-2032, (Units)

Global GPU Inference Server total production value, 2021-2032, (USD Million)

Global GPU Inference Server production by region & country, production, value, CAGR, 2021-2032, (USD Million) & (Units), (based on production site)

Global GPU Inference Server consumption by region & country, CAGR, 2021-2032 & (Units)

U.S. VS China: GPU Inference Server domestic production, consumption, key domestic manufacturers and share

Global GPU Inference Server production by manufacturer, production, price, value and market share 2021-2026, (USD Million) & (Units)

Global GPU Inference Server production by GPU Interconnect Architecture, production, value, CAGR, 2021-2032, (USD Million) & (Units)

Global GPU Inference Server production by Application, production, value, CAGR, 2021-2032, (USD Million) & (Units)

This report profiles key players in the global GPU Inference Server market based on the following parameters - company overview, production, value, price, gross margin, product portfolio, geographical presence, and key developments. Key companies covered as a part of this study include NVIDIA Corporation, Dell Technologies Inc., Hewlett Packard Enterprise Company, Super Micro Computer, Inc., Cisco Systems, Inc., Lenovo Group Limited, IEIT Systems Co., Ltd., New H3C Technologies Co., Ltd.,

Huawei Technologies Co., Ltd., xFusion International PTE. LTD., etc.

This report also provides key insights about market drivers, restraints, opportunities, new product launches or approvals.

Stakeholders would have ease in decision-making through various strategy matrices used in analyzing the World GPU Inference Server market

Detailed Segmentation:

Each section contains quantitative market data including market by value (US\$ Millions), volume (production, consumption) & (Units) and average price (K US\$/Unit) by manufacturer, by GPU Interconnect Architecture, and by Application. Data is given for the years 2021-2032 by year with 2025 as the base year, 2026 as the estimate year, and 2027-2032 as the forecast year.

Global GPU Inference Server Market, By Region:

United States

China

Europe

Japan

South Korea

ASEAN

India

Rest of World

Global GPU Inference Server Market, Segmentation by GPU Interconnect Architecture:

PCIe Direct-Attached GPU Inference Server

NVLink-Interconnected GPU Inference Server

NVSwitch Fully Interconnected GPU Inference Server

Supernode Unified-Interconnect GPU Inference Server

Other

Global GPU Inference Server Market, Segmentation by GPU Count Density:

Dual-GPU Low-Density GPU Inference Server

Four-GPU Mid-Density GPU Inference Server

Eight-GPU High-Density GPU Inference Server

Rack-Level Ultra-High-Density GPU Inference System

Global GPU Inference Server Market, Segmentation by Delivery Form:

Bare-Metal Complete GPU Inference Server

Private Cloud Integrated GPU Inference Server

Rack-Scale AI Factory GPU Inference System

Cloud-Rented GPU Inference Server

OEM Custom GPU Inference Server

Global GPU Inference Server Market, Segmentation by Application:

Large Language Model Online Inference

Retrieval-Augmented Generation Inference

Multimodal Content Generation

Recommendation and Advertising Ranking Inference

Vision and Video Analytics Inference

Enterprise Private Knowledge Assistant

Scientific Computing and Simulation Post-Processing

Edge Real-Time Decision-Making

Other

Companies Profiled:

NVIDIA Corporation

Dell Technologies Inc.

Hewlett Packard Enterprise Company

Super Micro Computer, Inc.

Cisco Systems, Inc.

Lenovo Group Limited

IEIT Systems Co., Ltd.

New H3C Technologies Co., Ltd.

Huawei Technologies Co., Ltd.

xFusion International PTE. LTD.

ASUSTeK Computer Inc.

GIGA-BYTE Technology Co., Ltd.

Quanta Cloud Technology Inc.

Wiiwynn Corporation

Inventec Corporation

ASRock Rack Inc.

Pegatron Corporation

KTNF Co., Ltd.

Fujitsu Limited

NEC Corporation

Lambda Labs, Inc.

Exxact Corporation

Penguin Solutions, Inc.

Key Questions Answered:

1. How big is the global GPU Inference Server market?
2. What is the demand of the global GPU Inference Server market?
3. What is the year over year growth of the global GPU Inference Server market?
4. What is the production and production value of the global GPU Inference Server market?
5. Who are the key producers in the global GPU Inference Server market?
6. What are the growth factors driving the market demand?

Contents

1 SUPPLY SUMMARY

- 1.1 GPU Inference Server Introduction
- 1.2 World GPU Inference Server Supply & Forecast
 - 1.2.1 World GPU Inference Server Production Value (2021 & 2025 & 2032)
 - 1.2.2 World GPU Inference Server Production (2021-2032)
 - 1.2.3 World GPU Inference Server Pricing Trends (2021-2032)
- 1.3 World GPU Inference Server Production by Region (Based on Production Site)
 - 1.3.1 World GPU Inference Server Production Value by Region (2021-2032)
 - 1.3.2 World GPU Inference Server Production by Region (2021-2032)
 - 1.3.3 World GPU Inference Server Average Price by Region (2021-2032)
 - 1.3.4 North America GPU Inference Server Production (2021-2032)
 - 1.3.5 China Taiwan GPU Inference Server Production (2021-2032)
 - 1.3.6 China GPU Inference Server Production (2021-2032)
 - 1.3.7 Japan GPU Inference Server Production (2021-2032)
 - 1.3.8 South Korea GPU Inference Server Production (2021-2032)
 - 1.3.9 Southeast Asia GPU Inference Server Production (2021-2032)
- 1.4 Market Drivers, Restraints and Trends
 - 1.4.1 GPU Inference Server Market Drivers
 - 1.4.2 Factors Affecting Demand
 - 1.4.3 GPU Inference Server Major Market Trends

2 DEMAND SUMMARY

- 2.1 World GPU Inference Server Demand (2021-2032)
- 2.2 World GPU Inference Server Consumption by Region
 - 2.2.1 World GPU Inference Server Consumption by Region (2021-2026)
 - 2.2.2 World GPU Inference Server Consumption Forecast by Region (2027-2032)
- 2.3 United States GPU Inference Server Consumption (2021-2032)
- 2.4 China GPU Inference Server Consumption (2021-2032)
- 2.5 Europe GPU Inference Server Consumption (2021-2032)
- 2.6 Japan GPU Inference Server Consumption (2021-2032)
- 2.7 South Korea GPU Inference Server Consumption (2021-2032)
- 2.8 ASEAN GPU Inference Server Consumption (2021-2032)
- 2.9 India GPU Inference Server Consumption (2021-2032)

3 WORLD MANUFACTURERS COMPETITIVE ANALYSIS

- 3.1 World GPU Inference Server Production Value by Manufacturer (2021-2026)
- 3.2 World GPU Inference Server Production by Manufacturer (2021-2026)
- 3.3 World GPU Inference Server Average Price by Manufacturer (2021-2026)
- 3.4 GPU Inference Server Company Evaluation Quadrant
- 3.5 Industry Rank and Concentration Rate (CR)
 - 3.5.1 Global GPU Inference Server Industry Rank of Major Manufacturers
 - 3.5.2 Global Concentration Ratios (CR4) for GPU Inference Server in 2025
 - 3.5.3 Global Concentration Ratios (CR8) for GPU Inference Server in 2025
- 3.6 GPU Inference Server Market: Overall Company Footprint Analysis
 - 3.6.1 GPU Inference Server Market: Region Footprint
 - 3.6.2 GPU Inference Server Market: Company Product Type Footprint
 - 3.6.3 GPU Inference Server Market: Company Product Application Footprint
- 3.7 Competitive Environment
 - 3.7.1 Historical Structure of the Industry
 - 3.7.2 Barriers of Market Entry
 - 3.7.3 Factors of Competition
- 3.8 New Entrant and Capacity Expansion Plans
- 3.9 Mergers, Acquisition, Agreements, and Collaborations

4 UNITED STATES VS CHINA VS REST OF THE WORLD

- 4.1 United States VS China: GPU Inference Server Production Value Comparison
 - 4.1.1 United States VS China: GPU Inference Server Production Value Comparison (2021 & 2025 & 2032)
 - 4.1.2 United States VS China: GPU Inference Server Production Value Market Share Comparison (2021 & 2025 & 2032)
- 4.2 United States VS China: GPU Inference Server Production Comparison
 - 4.2.1 United States VS China: GPU Inference Server Production Comparison (2021 & 2025 & 2032)
 - 4.2.2 United States VS China: GPU Inference Server Production Market Share Comparison (2021 & 2025 & 2032)
- 4.3 United States VS China: GPU Inference Server Consumption Comparison
 - 4.3.1 United States VS China: GPU Inference Server Consumption Comparison (2021 & 2025 & 2032)
 - 4.3.2 United States VS China: GPU Inference Server Consumption Market Share Comparison (2021 & 2025 & 2032)
- 4.4 United States Based GPU Inference Server Manufacturers and Market Share, 2021-2026

4.4.1 United States Based GPU Inference Server Manufacturers, Headquarters and Production Site (States, Country)

4.4.2 United States Based Manufacturers GPU Inference Server Production Value (2021-2026)

4.4.3 United States Based Manufacturers GPU Inference Server Production (2021-2026)

4.5 China Based GPU Inference Server Manufacturers and Market Share

4.5.1 China Based GPU Inference Server Manufacturers, Headquarters and Production Site (Province, Country)

4.5.2 China Based Manufacturers GPU Inference Server Production Value (2021-2026)

4.5.3 China Based Manufacturers GPU Inference Server Production (2021-2026)

4.6 Rest of World Based GPU Inference Server Manufacturers and Market Share, 2021-2026

4.6.1 Rest of World Based GPU Inference Server Manufacturers, Headquarters and Production Site (State, Country)

4.6.2 Rest of World Based Manufacturers GPU Inference Server Production Value (2021-2026)

4.6.3 Rest of World Based Manufacturers GPU Inference Server Production (2021-2026)

5 MARKET ANALYSIS BY GPU INTERCONNECT ARCHITECTURE

5.1 World GPU Inference Server Market Size Overview by GPU Interconnect Architecture: 2021 VS 2025 VS 2032

5.2 Segment Introduction by GPU Interconnect Architecture

5.2.1 PCIe Direct-Attached GPU Inference Server

5.2.2 NVLink-Interconnected GPU Inference Server

5.2.3 NVSwitch Fully Interconnected GPU Inference Server

5.2.4 Supernode Unified-Interconnect GPU Inference Server

5.2.5 Other

5.3 Market Segment by GPU Interconnect Architecture

5.3.1 World GPU Inference Server Production by GPU Interconnect Architecture (2021-2032)

5.3.2 World GPU Inference Server Production Value by GPU Interconnect Architecture (2021-2032)

5.3.3 World GPU Inference Server Average Price by GPU Interconnect Architecture (2021-2032)

6 MARKET ANALYSIS BY GPU COUNT DENSITY

6.1 World GPU Inference Server Market Size Overview by GPU Count Density: 2021 VS 2025 VS 2032

6.2 Segment Introduction by GPU Count Density

6.2.1 Dual-GPU Low-Density GPU Inference Server

6.2.2 Four-GPU Mid-Density GPU Inference Server

6.2.3 Eight-GPU High-Density GPU Inference Server

6.2.4 Rack-Level Ultra-High-Density GPU Inference System

6.3 Market Segment by GPU Count Density

6.3.1 World GPU Inference Server Production by GPU Count Density (2021-2032)

6.3.2 World GPU Inference Server Production Value by GPU Count Density (2021-2032)

6.3.3 World GPU Inference Server Average Price by GPU Count Density (2021-2032)

7 MARKET ANALYSIS BY DELIVERY FORM

7.1 World GPU Inference Server Market Size Overview by Delivery Form: 2021 VS 2025 VS 2032

7.2 Segment Introduction by Delivery Form

7.2.1 Bare-Metal Complete GPU Inference Server

7.2.2 Private Cloud Integrated GPU Inference Server

7.2.3 Rack-Scale AI Factory GPU Inference System

7.2.4 Cloud-Rented GPU Inference Server

7.2.5 OEM Custom GPU Inference Server

7.3 Market Segment by Delivery Form

7.3.1 World GPU Inference Server Production by Delivery Form (2021-2032)

7.3.2 World GPU Inference Server Production Value by Delivery Form (2021-2032)

7.3.3 World GPU Inference Server Average Price by Delivery Form (2021-2032)

8 MARKET ANALYSIS BY APPLICATION

8.1 World GPU Inference Server Market Size Overview by Application: 2021 VS 2025 VS 2032

8.2 Segment Introduction by Application

8.2.1 Large Language Model Online Inference

8.2.2 Retrieval-Augmented Generation Inference

8.2.3 Multimodal Content Generation

8.2.4 Recommendation and Advertising Ranking Inference

- 8.2.5 Vision and Video Analytics Inference
- 8.2.6 Enterprise Private Knowledge Assistant
- 8.2.7 Scientific Computing and Simulation Post-Processing
- 8.2.8 Edge Real-Time Decision-Making
- 8.2.9 Other

8.3 Market Segment by Application

- 8.3.1 World GPU Inference Server Production by Application (2021-2032)
- 8.3.2 World GPU Inference Server Production Value by Application (2021-2032)
- 8.3.3 World GPU Inference Server Average Price by Application (2021-2032)

9 COMPANY PROFILES

9.1 NVIDIA Corporation

- 9.1.1 NVIDIA Corporation Details
- 9.1.2 NVIDIA Corporation Major Business
- 9.1.3 NVIDIA Corporation GPU Inference Server Product and Services
- 9.1.4 NVIDIA Corporation GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.1.5 NVIDIA Corporation Recent Developments/Updates
- 9.1.6 NVIDIA Corporation Competitive Strengths & Weaknesses

9.2 Dell Technologies Inc.

- 9.2.1 Dell Technologies Inc. Details
- 9.2.2 Dell Technologies Inc. Major Business
- 9.2.3 Dell Technologies Inc. GPU Inference Server Product and Services
- 9.2.4 Dell Technologies Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.2.5 Dell Technologies Inc. Recent Developments/Updates
- 9.2.6 Dell Technologies Inc. Competitive Strengths & Weaknesses

9.3 Hewlett Packard Enterprise Company

- 9.3.1 Hewlett Packard Enterprise Company Details
- 9.3.2 Hewlett Packard Enterprise Company Major Business
- 9.3.3 Hewlett Packard Enterprise Company GPU Inference Server Product and Services
- 9.3.4 Hewlett Packard Enterprise Company GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.3.5 Hewlett Packard Enterprise Company Recent Developments/Updates
- 9.3.6 Hewlett Packard Enterprise Company Competitive Strengths & Weaknesses

9.4 Super Micro Computer, Inc.

- 9.4.1 Super Micro Computer, Inc. Details

- 9.4.2 Super Micro Computer, Inc. Major Business
- 9.4.3 Super Micro Computer, Inc. GPU Inference Server Product and Services
- 9.4.4 Super Micro Computer, Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.4.5 Super Micro Computer, Inc. Recent Developments/Updates
- 9.4.6 Super Micro Computer, Inc. Competitive Strengths & Weaknesses
- 9.5 Cisco Systems, Inc.
 - 9.5.1 Cisco Systems, Inc. Details
 - 9.5.2 Cisco Systems, Inc. Major Business
 - 9.5.3 Cisco Systems, Inc. GPU Inference Server Product and Services
 - 9.5.4 Cisco Systems, Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.5.5 Cisco Systems, Inc. Recent Developments/Updates
 - 9.5.6 Cisco Systems, Inc. Competitive Strengths & Weaknesses
- 9.6 Lenovo Group Limited
 - 9.6.1 Lenovo Group Limited Details
 - 9.6.2 Lenovo Group Limited Major Business
 - 9.6.3 Lenovo Group Limited GPU Inference Server Product and Services
 - 9.6.4 Lenovo Group Limited GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.6.5 Lenovo Group Limited Recent Developments/Updates
 - 9.6.6 Lenovo Group Limited Competitive Strengths & Weaknesses
- 9.7 IEIT Systems Co., Ltd.
 - 9.7.1 IEIT Systems Co., Ltd. Details
 - 9.7.2 IEIT Systems Co., Ltd. Major Business
 - 9.7.3 IEIT Systems Co., Ltd. GPU Inference Server Product and Services
 - 9.7.4 IEIT Systems Co., Ltd. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.7.5 IEIT Systems Co., Ltd. Recent Developments/Updates
 - 9.7.6 IEIT Systems Co., Ltd. Competitive Strengths & Weaknesses
- 9.8 New H3C Technologies Co., Ltd.
 - 9.8.1 New H3C Technologies Co., Ltd. Details
 - 9.8.2 New H3C Technologies Co., Ltd. Major Business
 - 9.8.3 New H3C Technologies Co., Ltd. GPU Inference Server Product and Services
 - 9.8.4 New H3C Technologies Co., Ltd. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.8.5 New H3C Technologies Co., Ltd. Recent Developments/Updates
 - 9.8.6 New H3C Technologies Co., Ltd. Competitive Strengths & Weaknesses
- 9.9 Huawei Technologies Co., Ltd.

- 9.9.1 Huawei Technologies Co., Ltd. Details
- 9.9.2 Huawei Technologies Co., Ltd. Major Business
- 9.9.3 Huawei Technologies Co., Ltd. GPU Inference Server Product and Services
- 9.9.4 Huawei Technologies Co., Ltd. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.9.5 Huawei Technologies Co., Ltd. Recent Developments/Updates
- 9.9.6 Huawei Technologies Co., Ltd. Competitive Strengths & Weaknesses
- 9.10 xFusion International PTE. LTD.
- 9.10.1 xFusion International PTE. LTD. Details
- 9.10.2 xFusion International PTE. LTD. Major Business
- 9.10.3 xFusion International PTE. LTD. GPU Inference Server Product and Services
- 9.10.4 xFusion International PTE. LTD. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.10.5 xFusion International PTE. LTD. Recent Developments/Updates
- 9.10.6 xFusion International PTE. LTD. Competitive Strengths & Weaknesses
- 9.11 ASUSTeK Computer Inc.
- 9.11.1 ASUSTeK Computer Inc. Details
- 9.11.2 ASUSTeK Computer Inc. Major Business
- 9.11.3 ASUSTeK Computer Inc. GPU Inference Server Product and Services
- 9.11.4 ASUSTeK Computer Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.11.5 ASUSTeK Computer Inc. Recent Developments/Updates
- 9.11.6 ASUSTeK Computer Inc. Competitive Strengths & Weaknesses
- 9.12 GIGA-BYTE Technology Co., Ltd.
- 9.12.1 GIGA-BYTE Technology Co., Ltd. Details
- 9.12.2 GIGA-BYTE Technology Co., Ltd. Major Business
- 9.12.3 GIGA-BYTE Technology Co., Ltd. GPU Inference Server Product and Services
- 9.12.4 GIGA-BYTE Technology Co., Ltd. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.12.5 GIGA-BYTE Technology Co., Ltd. Recent Developments/Updates
- 9.12.6 GIGA-BYTE Technology Co., Ltd. Competitive Strengths & Weaknesses
- 9.13 Quanta Cloud Technology Inc.
- 9.13.1 Quanta Cloud Technology Inc. Details
- 9.13.2 Quanta Cloud Technology Inc. Major Business
- 9.13.3 Quanta Cloud Technology Inc. GPU Inference Server Product and Services
- 9.13.4 Quanta Cloud Technology Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
- 9.13.5 Quanta Cloud Technology Inc. Recent Developments/Updates
- 9.13.6 Quanta Cloud Technology Inc. Competitive Strengths & Weaknesses

9.14 Wiwynn Corporation

9.14.1 Wiwynn Corporation Details

9.14.2 Wiwynn Corporation Major Business

9.14.3 Wiwynn Corporation GPU Inference Server Product and Services

9.14.4 Wiwynn Corporation GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)

9.14.5 Wiwynn Corporation Recent Developments/Updates

9.14.6 Wiwynn Corporation Competitive Strengths & Weaknesses

9.15 Inventec Corporation

9.15.1 Inventec Corporation Details

9.15.2 Inventec Corporation Major Business

9.15.3 Inventec Corporation GPU Inference Server Product and Services

9.15.4 Inventec Corporation GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)

9.15.5 Inventec Corporation Recent Developments/Updates

9.15.6 Inventec Corporation Competitive Strengths & Weaknesses

9.16 ASRock Rack Inc.

9.16.1 ASRock Rack Inc. Details

9.16.2 ASRock Rack Inc. Major Business

9.16.3 ASRock Rack Inc. GPU Inference Server Product and Services

9.16.4 ASRock Rack Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)

9.16.5 ASRock Rack Inc. Recent Developments/Updates

9.16.6 ASRock Rack Inc. Competitive Strengths & Weaknesses

9.17 Pegatron Corporation

9.17.1 Pegatron Corporation Details

9.17.2 Pegatron Corporation Major Business

9.17.3 Pegatron Corporation GPU Inference Server Product and Services

9.17.4 Pegatron Corporation GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)

9.17.5 Pegatron Corporation Recent Developments/Updates

9.17.6 Pegatron Corporation Competitive Strengths & Weaknesses

9.18 KTNF Co., Ltd.

9.18.1 KTNF Co., Ltd. Details

9.18.2 KTNF Co., Ltd. Major Business

9.18.3 KTNF Co., Ltd. GPU Inference Server Product and Services

9.18.4 KTNF Co., Ltd. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)

9.18.5 KTNF Co., Ltd. Recent Developments/Updates

- 9.18.6 KTNF Co., Ltd. Competitive Strengths & Weaknesses
- 9.19 Fujitsu Limited
 - 9.19.1 Fujitsu Limited Details
 - 9.19.2 Fujitsu Limited Major Business
 - 9.19.3 Fujitsu Limited GPU Inference Server Product and Services
 - 9.19.4 Fujitsu Limited GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.19.5 Fujitsu Limited Recent Developments/Updates
 - 9.19.6 Fujitsu Limited Competitive Strengths & Weaknesses
- 9.20 NEC Corporation
 - 9.20.1 NEC Corporation Details
 - 9.20.2 NEC Corporation Major Business
 - 9.20.3 NEC Corporation GPU Inference Server Product and Services
 - 9.20.4 NEC Corporation GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.20.5 NEC Corporation Recent Developments/Updates
 - 9.20.6 NEC Corporation Competitive Strengths & Weaknesses
- 9.21 Lambda Labs, Inc.
 - 9.21.1 Lambda Labs, Inc. Details
 - 9.21.2 Lambda Labs, Inc. Major Business
 - 9.21.3 Lambda Labs, Inc. GPU Inference Server Product and Services
 - 9.21.4 Lambda Labs, Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.21.5 Lambda Labs, Inc. Recent Developments/Updates
 - 9.21.6 Lambda Labs, Inc. Competitive Strengths & Weaknesses
- 9.22 Exxact Corporation
 - 9.22.1 Exxact Corporation Details
 - 9.22.2 Exxact Corporation Major Business
 - 9.22.3 Exxact Corporation GPU Inference Server Product and Services
 - 9.22.4 Exxact Corporation GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)
 - 9.22.5 Exxact Corporation Recent Developments/Updates
 - 9.22.6 Exxact Corporation Competitive Strengths & Weaknesses
- 9.23 Penguin Solutions, Inc.
 - 9.23.1 Penguin Solutions, Inc. Details
 - 9.23.2 Penguin Solutions, Inc. Major Business
 - 9.23.3 Penguin Solutions, Inc. GPU Inference Server Product and Services
 - 9.23.4 Penguin Solutions, Inc. GPU Inference Server Production, Price, Value, Gross Margin and Market Share (2021-2026)

9.23.5 Penguin Solutions, Inc. Recent Developments/Updates

9.23.6 Penguin Solutions, Inc. Competitive Strengths & Weaknesses

10 INDUSTRY CHAIN ANALYSIS

10.1 GPU Inference Server Industry Chain

10.2 GPU Inference Server Upstream Analysis

10.2.1 GPU Inference Server Core Raw Materials

10.2.2 Main Manufacturers of GPU Inference Server Core Raw Materials

10.3 Midstream Analysis

10.4 Downstream Analysis

10.5 GPU Inference Server Production Mode

10.6 GPU Inference Server Procurement Model

10.7 GPU Inference Server Industry Sales Model and Sales Channels

10.7.1 GPU Inference Server Sales Model

10.7.2 GPU Inference Server Typical Distributors

11 RESEARCH FINDINGS AND CONCLUSION

12 APPENDIX

12.1 Methodology

12.2 Research Process and Data Source

12.3 Disclaimer

List Of Tables

LIST OF TABLES

Table 1. World GPU Inference Server Production Value by Region (2021, 2025 and 2032) & (USD Million)

Table 2. World GPU Inference Server Production Value by Region (2021-2026) & (USD Million)

Table 3. World GPU Inference Server Production Value by Region (2027-2032) & (USD Million)

Table 4. World GPU Inference Server Production Value Market Share by Region (2021-2026)

Table 5. World GPU Inference Server Production Value Market Share by Region (2027-2032)

Table 6. World GPU Inference Server Production by Region (2021-2026) & (Units)

Table 7. World GPU Inference Server Production by Region (2027-2032) & (Units)

Table 8. World GPU Inference Server Production Market Share by Region (2021-2026)

Table 9. World GPU Inference Server Production Market Share by Region (2027-2032)

Table 10. World GPU Inference Server Average Price by Region (2021-2026) & (K US\$/Unit)

Table 11. World GPU Inference Server Average Price by Region (2027-2032) & (K US\$/Unit)

Table 12. GPU Inference Server Major Market Trends

Table 13. World GPU Inference Server Consumption Growth Rate Forecast by Region (2021 & 2025 & 2032) & (Units)

Table 14. World GPU Inference Server Consumption by Region (2021-2026) & (Units)

Table 15. World GPU Inference Server Consumption Forecast by Region (2027-2032) & (Units)

Table 16. World GPU Inference Server Production Value by Manufacturer (2021-2026) & (USD Million)

Table 17. Production Value Market Share of Key GPU Inference Server Producers in 2025

Table 18. World GPU Inference Server Production by Manufacturer (2021-2026) & (Units)

Table 19. Production Market Share of Key GPU Inference Server Producers in 2025

Table 20. World GPU Inference Server Average Price by Manufacturer (2021-2026) & (K US\$/Unit)

Table 21. Global GPU Inference Server Company Evaluation Quadrant

Table 22. World GPU Inference Server Industry Rank of Major Manufacturers, Based on

Production Value in 2025

Table 23. Head Office and GPU Inference Server Production Site of Key Manufacturer

Table 24. GPU Inference Server Market: Company Product Type Footprint

Table 25. GPU Inference Server Market: Company Product Application Footprint

Table 26. GPU Inference Server Competitive Factors

Table 27. GPU Inference Server New Entrant and Capacity Expansion Plans

Table 28. GPU Inference Server Mergers & Acquisitions Activity

Table 29. United States VS China GPU Inference Server Production Value Comparison, (2021 & 2025 & 2032) & (USD Million)

Table 30. United States VS China GPU Inference Server Production Comparison, (2021 & 2025 & 2032) & (Units)

Table 31. United States VS China GPU Inference Server Consumption Comparison, (2021 & 2025 & 2032) & (Units)

Table 32. United States Based GPU Inference Server Manufacturers, Headquarters and Production Site (States, Country)

Table 33. United States Based Manufacturers GPU Inference Server Production Value, (2021-2026) & (USD Million)

Table 34. United States Based Manufacturers GPU Inference Server Production Value Market Share (2021-2026)

Table 35. United States Based Manufacturers GPU Inference Server Production (2021-2026) & (Units)

Table 36. United States Based Manufacturers GPU Inference Server Production Market Share (2021-2026)

Table 37. China Based GPU Inference Server Manufacturers, Headquarters and Production Site (Province, Country)

Table 38. China Based Manufacturers GPU Inference Server Production Value, (2021-2026) & (USD Million)

Table 39. China Based Manufacturers GPU Inference Server Production Value Market Share (2021-2026)

Table 40. China Based Manufacturers GPU Inference Server Production, (2021-2026) & (Units)

Table 41. China Based Manufacturers GPU Inference Server Production Market Share (2021-2026)

Table 42. Rest of World Based GPU Inference Server Manufacturers, Headquarters and Production Site (State, Country)

Table 43. Rest of World Based Manufacturers GPU Inference Server Production Value, (2021-2026) & (USD Million)

Table 44. Rest of World Based Manufacturers GPU Inference Server Production Value Market Share (2021-2026)

Table 45. Rest of World Based Manufacturers GPU Inference Server Production, (2021-2026) & (Units)

Table 46. Rest of World Based Manufacturers GPU Inference Server Production Market Share (2021-2026)

Table 47. World GPU Inference Server Production Value by GPU Interconnect Architecture, (USD Million), 2021 & 2025 & 2032

Table 48. World GPU Inference Server Production by GPU Interconnect Architecture (2021-2026) & (Units)

Table 49. World GPU Inference Server Production by GPU Interconnect Architecture (2027-2032) & (Units)

Table 50. World GPU Inference Server Production Value by GPU Interconnect Architecture (2021-2026) & (USD Million)

Table 51. World GPU Inference Server Production Value by GPU Interconnect Architecture (2027-2032) & (USD Million)

Table 52. World GPU Inference Server Average Price by GPU Interconnect Architecture (2021-2026) & (K US\$/Unit)

Table 53. World GPU Inference Server Average Price by GPU Interconnect Architecture (2027-2032) & (K US\$/Unit)

Table 54. World GPU Inference Server Production Value by GPU Count Density, (USD Million), 2021 & 2025 & 2032

Table 55. World GPU Inference Server Production by GPU Count Density (2021-2026) & (Units)

Table 56. World GPU Inference Server Production by GPU Count Density (2027-2032) & (Units)

Table 57. World GPU Inference Server Production Value by GPU Count Density (2021-2026) & (USD Million)

Table 58. World GPU Inference Server Production Value by GPU Count Density (2027-2032) & (USD Million)

Table 59. World GPU Inference Server Average Price by GPU Count Density (2021-2026) & (K US\$/Unit)

Table 60. World GPU Inference Server Average Price by GPU Count Density (2027-2032) & (K US\$/Unit)

Table 61. World GPU Inference Server Production Value by Delivery Form, (USD Million), 2021 & 2025 & 2032

Table 62. World GPU Inference Server Production by Delivery Form (2021-2026) & (Units)

Table 63. World GPU Inference Server Production by Delivery Form (2027-2032) & (Units)

Table 64. World GPU Inference Server Production Value by Delivery Form (2021-2026)

& (USD Million)

Table 65. World GPU Inference Server Production Value by Delivery Form (2027-2032)

& (USD Million)

Table 66. World GPU Inference Server Average Price by Delivery Form (2021-2026) & (K US\$/Unit)

Table 67. World GPU Inference Server Average Price by Delivery Form (2027-2032) & (K US\$/Unit)

Table 68. World GPU Inference Server Production Value by Application, (USD Million), 2021 & 2025 & 2032

Table 69. World GPU Inference Server Production by Application (2021-2026) & (Units)

Table 70. World GPU Inference Server Production by Application (2027-2032) & (Units)

Table 71. World GPU Inference Server Production Value by Application (2021-2026) & (USD Million)

Table 72. World GPU Inference Server Production Value by Application (2027-2032) & (USD Million)

Table 73. World GPU Inference Server Average Price by Application (2021-2026) & (K US\$/Unit)

Table 74. World GPU Inference Server Average Price by Application (2027-2032) & (K US\$/Unit)

Table 75. NVIDIA Corporation Basic Information, Manufacturing Base and Competitors

Table 76. NVIDIA Corporation Major Business

Table 77. NVIDIA Corporation GPU Inference Server Product and Services

Table 78. NVIDIA Corporation GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 79. NVIDIA Corporation Recent Developments/Updates

Table 80. NVIDIA Corporation Competitive Strengths & Weaknesses

Table 81. Dell Technologies Inc. Basic Information, Manufacturing Base and Competitors

Table 82. Dell Technologies Inc. Major Business

Table 83. Dell Technologies Inc. GPU Inference Server Product and Services

Table 84. Dell Technologies Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 85. Dell Technologies Inc. Recent Developments/Updates

Table 86. Dell Technologies Inc. Competitive Strengths & Weaknesses

Table 87. Hewlett Packard Enterprise Company Basic Information, Manufacturing Base and Competitors

Table 88. Hewlett Packard Enterprise Company Major Business

Table 89. Hewlett Packard Enterprise Company GPU Inference Server Product and Services

Table 90. Hewlett Packard Enterprise Company GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 91. Hewlett Packard Enterprise Company Recent Developments/Updates

Table 92. Hewlett Packard Enterprise Company Competitive Strengths & Weaknesses

Table 93. Super Micro Computer, Inc. Basic Information, Manufacturing Base and Competitors

Table 94. Super Micro Computer, Inc. Major Business

Table 95. Super Micro Computer, Inc. GPU Inference Server Product and Services

Table 96. Super Micro Computer, Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 97. Super Micro Computer, Inc. Recent Developments/Updates

Table 98. Super Micro Computer, Inc. Competitive Strengths & Weaknesses

Table 99. Cisco Systems, Inc. Basic Information, Manufacturing Base and Competitors

Table 100. Cisco Systems, Inc. Major Business

Table 101. Cisco Systems, Inc. GPU Inference Server Product and Services

Table 102. Cisco Systems, Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 103. Cisco Systems, Inc. Recent Developments/Updates

Table 104. Cisco Systems, Inc. Competitive Strengths & Weaknesses

Table 105. Lenovo Group Limited Basic Information, Manufacturing Base and Competitors

Table 106. Lenovo Group Limited Major Business

Table 107. Lenovo Group Limited GPU Inference Server Product and Services

Table 108. Lenovo Group Limited GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 109. Lenovo Group Limited Recent Developments/Updates

Table 110. Lenovo Group Limited Competitive Strengths & Weaknesses

Table 111. IEIT Systems Co., Ltd. Basic Information, Manufacturing Base and Competitors

Table 112. IEIT Systems Co., Ltd. Major Business

Table 113. IEIT Systems Co., Ltd. GPU Inference Server Product and Services

Table 114. IEIT Systems Co., Ltd. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share

(2021-2026)

Table 115. IEIT Systems Co., Ltd. Recent Developments/Updates

Table 116. IEIT Systems Co., Ltd. Competitive Strengths & Weaknesses

Table 117. New H3C Technologies Co., Ltd. Basic Information, Manufacturing Base and Competitors

Table 118. New H3C Technologies Co., Ltd. Major Business

Table 119. New H3C Technologies Co., Ltd. GPU Inference Server Product and Services

Table 120. New H3C Technologies Co., Ltd. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 121. New H3C Technologies Co., Ltd. Recent Developments/Updates

Table 122. New H3C Technologies Co., Ltd. Competitive Strengths & Weaknesses

Table 123. Huawei Technologies Co., Ltd. Basic Information, Manufacturing Base and Competitors

Table 124. Huawei Technologies Co., Ltd. Major Business

Table 125. Huawei Technologies Co., Ltd. GPU Inference Server Product and Services

Table 126. Huawei Technologies Co., Ltd. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 127. Huawei Technologies Co., Ltd. Recent Developments/Updates

Table 128. Huawei Technologies Co., Ltd. Competitive Strengths & Weaknesses

Table 129. xFusion International PTE. LTD. Basic Information, Manufacturing Base and Competitors

Table 130. xFusion International PTE. LTD. Major Business

Table 131. xFusion International PTE. LTD. GPU Inference Server Product and Services

Table 132. xFusion International PTE. LTD. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 133. xFusion International PTE. LTD. Recent Developments/Updates

Table 134. xFusion International PTE. LTD. Competitive Strengths & Weaknesses

Table 135. ASUSTeK Computer Inc. Basic Information, Manufacturing Base and Competitors

Table 136. ASUSTeK Computer Inc. Major Business

Table 137. ASUSTeK Computer Inc. GPU Inference Server Product and Services

Table 138. ASUSTeK Computer Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 139. ASUSTeK Computer Inc. Recent Developments/Updates

Table 140. ASUSTeK Computer Inc. Competitive Strengths & Weaknesses

Table 141. GIGA-BYTE Technology Co., Ltd. Basic Information, Manufacturing Base and Competitors

Table 142. GIGA-BYTE Technology Co., Ltd. Major Business

Table 143. GIGA-BYTE Technology Co., Ltd. GPU Inference Server Product and Services

Table 144. GIGA-BYTE Technology Co., Ltd. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 145. GIGA-BYTE Technology Co., Ltd. Recent Developments/Updates

Table 146. GIGA-BYTE Technology Co., Ltd. Competitive Strengths & Weaknesses

Table 147. Quanta Cloud Technology Inc. Basic Information, Manufacturing Base and Competitors

Table 148. Quanta Cloud Technology Inc. Major Business

Table 149. Quanta Cloud Technology Inc. GPU Inference Server Product and Services

Table 150. Quanta Cloud Technology Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 151. Quanta Cloud Technology Inc. Recent Developments/Updates

Table 152. Quanta Cloud Technology Inc. Competitive Strengths & Weaknesses

Table 153. Wiwynn Corporation Basic Information, Manufacturing Base and Competitors

Table 154. Wiwynn Corporation Major Business

Table 155. Wiwynn Corporation GPU Inference Server Product and Services

Table 156. Wiwynn Corporation GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 157. Wiwynn Corporation Recent Developments/Updates

Table 158. Wiwynn Corporation Competitive Strengths & Weaknesses

Table 159. Inventec Corporation Basic Information, Manufacturing Base and Competitors

Table 160. Inventec Corporation Major Business

Table 161. Inventec Corporation GPU Inference Server Product and Services

Table 162. Inventec Corporation GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 163. Inventec Corporation Recent Developments/Updates

Table 164. Inventec Corporation Competitive Strengths & Weaknesses

Table 165. ASRock Rack Inc. Basic Information, Manufacturing Base and Competitors

Table 166. ASRock Rack Inc. Major Business

Table 167. ASRock Rack Inc. GPU Inference Server Product and Services

Table 168. ASRock Rack Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 169. ASRock Rack Inc. Recent Developments/Updates

Table 170. ASRock Rack Inc. Competitive Strengths & Weaknesses

Table 171. Pegatron Corporation Basic Information, Manufacturing Base and Competitors

Table 172. Pegatron Corporation Major Business

Table 173. Pegatron Corporation GPU Inference Server Product and Services

Table 174. Pegatron Corporation GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 175. Pegatron Corporation Recent Developments/Updates

Table 176. Pegatron Corporation Competitive Strengths & Weaknesses

Table 177. KTNF Co., Ltd. Basic Information, Manufacturing Base and Competitors

Table 178. KTNF Co., Ltd. Major Business

Table 179. KTNF Co., Ltd. GPU Inference Server Product and Services

Table 180. KTNF Co., Ltd. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 181. KTNF Co., Ltd. Recent Developments/Updates

Table 182. KTNF Co., Ltd. Competitive Strengths & Weaknesses

Table 183. Fujitsu Limited Basic Information, Manufacturing Base and Competitors

Table 184. Fujitsu Limited Major Business

Table 185. Fujitsu Limited GPU Inference Server Product and Services

Table 186. Fujitsu Limited GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 187. Fujitsu Limited Recent Developments/Updates

Table 188. Fujitsu Limited Competitive Strengths & Weaknesses

Table 189. NEC Corporation Basic Information, Manufacturing Base and Competitors

Table 190. NEC Corporation Major Business

Table 191. NEC Corporation GPU Inference Server Product and Services

Table 192. NEC Corporation GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 193. NEC Corporation Recent Developments/Updates

Table 194. NEC Corporation Competitive Strengths & Weaknesses

Table 195. Lambda Labs, Inc. Basic Information, Manufacturing Base and Competitors

Table 196. Lambda Labs, Inc. Major Business

Table 197. Lambda Labs, Inc. GPU Inference Server Product and Services

Table 198. Lambda Labs, Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 199. Lambda Labs, Inc. Recent Developments/Updates

Table 200. Lambda Labs, Inc. Competitive Strengths & Weaknesses

Table 201. Exxact Corporation Basic Information, Manufacturing Base and Competitors

Table 202. Exxact Corporation Major Business

Table 203. Exxact Corporation GPU Inference Server Product and Services

Table 204. Exxact Corporation GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 205. Exxact Corporation Recent Developments/Updates

Table 206. Exxact Corporation Competitive Strengths & Weaknesses

Table 207. Penguin Solutions, Inc. Basic Information, Manufacturing Base and Competitors

Table 208. Penguin Solutions, Inc. Major Business

Table 209. Penguin Solutions, Inc. GPU Inference Server Product and Services

Table 210. Penguin Solutions, Inc. GPU Inference Server Production (Units), Price (K US\$/Unit), Production Value (USD Million), Gross Margin and Market Share (2021-2026)

Table 211. Penguin Solutions, Inc. Recent Developments/Updates

Table 212. Penguin Solutions, Inc. Competitive Strengths & Weaknesses

Table 213. Global Key Players of GPU Inference Server Upstream (Raw Materials)

Table 214. Global GPU Inference Server Typical Customers

Table 215. GPU Inference Server Typical Distributors

List Of Figures

LIST OF FIGURES

Figure 1. GPU Inference Server Picture

Figure 2. World GPU Inference Server Production Value: 2021 & 2025 & 2032, (USD Million)

Figure 3. World GPU Inference Server Production Value and Forecast (2021-2032) & (USD Million)

Figure 4. World GPU Inference Server Production (2021-2032) & (Units)

Figure 5. World GPU Inference Server Average Price (2021-2032) & (K US\$/Unit)

Figure 6. World GPU Inference Server Production Value Market Share by Region (2021-2032)

Figure 7. World GPU Inference Server Production Market Share by Region (2021-2032)

Figure 8. North America GPU Inference Server Production (2021-2032) & (Units)

Figure 9. China Taiwan GPU Inference Server Production (2021-2032) & (Units)

Figure 10. China GPU Inference Server Production (2021-2032) & (Units)

Figure 11. Japan GPU Inference Server Production (2021-2032) & (Units)

Figure 12. South Korea GPU Inference Server Production (2021-2032) & (Units)

Figure 13. Southeast Asia GPU Inference Server Production (2021-2032) & (Units)

Figure 14. GPU Inference Server Market Drivers

Figure 15. Factors Affecting Demand

Figure 16. World GPU Inference Server Consumption (2021-2032) & (Units)

Figure 17. World GPU Inference Server Consumption Market Share by Region (2021-2032)

Figure 18. United States GPU Inference Server Consumption (2021-2032) & (Units)

Figure 19. China GPU Inference Server Consumption (2021-2032) & (Units)

Figure 20. Europe GPU Inference Server Consumption (2021-2032) & (Units)

Figure 21. Japan GPU Inference Server Consumption (2021-2032) & (Units)

Figure 22. South Korea GPU Inference Server Consumption (2021-2032) & (Units)

Figure 23. ASEAN GPU Inference Server Consumption (2021-2032) & (Units)

Figure 24. India GPU Inference Server Consumption (2021-2032) & (Units)

Figure 25. Producer Shipments of GPU Inference Server by Manufacturer Revenue (\$MM) and Market Share (%): 2025

Figure 26. Global Four-firm Concentration Ratios (CR4) for GPU Inference Server Markets in 2025

Figure 27. Global Four-firm Concentration Ratios (CR8) for GPU Inference Server Markets in 2025

Figure 28. United States VS China: GPU Inference Server Production Value Market

Share Comparison (2021 & 2025 & 2032)

Figure 29. United States VS China: GPU Inference Server Production Market Share Comparison (2021 & 2025 & 2032)

Figure 30. United States VS China: GPU Inference Server Consumption Market Share Comparison (2021 & 2025 & 2032)

Figure 31. United States Based Manufacturers GPU Inference Server Production Market Share 2025

Figure 32. China Based Manufacturers GPU Inference Server Production Market Share 2025

Figure 33. Rest of World Based Manufacturers GPU Inference Server Production Market Share 2025

Figure 34. World GPU Inference Server Production Value by GPU Interconnect Architecture, (USD Million), 2021 & 2025 & 2032

Figure 35. World GPU Inference Server Production Value Market Share by GPU Interconnect Architecture in 2025

Figure 36. PCIe Direct-Attached GPU Inference Server

Figure 37. NVLink-Interconnected GPU Inference Server

Figure 38. NVSwitch Fully Interconnected GPU Inference Server

Figure 39. Supernode Unified-Interconnect GPU Inference Server

Figure 40. Other

Figure 41. World GPU Inference Server Production Market Share by GPU Interconnect Architecture (2021-2032)

Figure 42. World GPU Inference Server Production Value Market Share by GPU Interconnect Architecture (2021-2032)

Figure 43. World GPU Inference Server Average Price by GPU Interconnect Architecture (2021-2032) & (K US\$/Unit)

Figure 44. World GPU Inference Server Production Value by GPU Count Density, (USD Million), 2021 & 2025 & 2032

Figure 45. World GPU Inference Server Production Value Market Share by GPU Count Density in 2025

Figure 46. Dual-GPU Low-Density GPU Inference Server

Figure 47. Four-GPU Mid-Density GPU Inference Server

Figure 48. Eight-GPU High-Density GPU Inference Server

Figure 49. Rack-Level Ultra-High-Density GPU Inference System

Figure 50. World GPU Inference Server Production Market Share by GPU Count Density (2021-2032)

Figure 51. World GPU Inference Server Production Value Market Share by GPU Count Density (2021-2032)

Figure 52. World GPU Inference Server Average Price by GPU Count Density

(2021-2032) & (K US\$/Unit)

Figure 53. World GPU Inference Server Production Value by Delivery Form, (USD Million), 2021 & 2025 & 2032

Figure 54. World GPU Inference Server Production Value Market Share by Delivery Form in 2025

Figure 55. Bare-Metal Complete GPU Inference Server

Figure 56. Private Cloud Integrated GPU Inference Server

Figure 57. Rack-Scale AI Factory GPU Inference System

Figure 58. Cloud-Rented GPU Inference Server

Figure 59. OEM Custom GPU Inference Server

Figure 60. World GPU Inference Server Production Market Share by Delivery Form (2021-2032)

Figure 61. World GPU Inference Server Production Value Market Share by Delivery Form (2021-2032)

Figure 62. World GPU Inference Server Average Price by Delivery Form (2021-2032) & (K US\$/Unit)

Figure 63. World GPU Inference Server Production Value by Application, (USD Million), 2021 & 2025 & 2032

Figure 64. World GPU Inference Server Production Value Market Share by Application in 2025

Figure 65. Large Language Model Online Inference

Figure 66. Retrieval-Augmented Generation Inference

Figure 67. Multimodal Content Generation

Figure 68. Recommendation and Advertising Ranking Inference

Figure 69. Vision and Video Analytics Inference

Figure 70. Enterprise Private Knowledge Assistant

Figure 71. Scientific Computing and Simulation Post-Processing

Figure 72. Edge Real-Time Decision-Making

Figure 73. Edge Real-Time Decision-Making

Figure 74. World GPU Inference Server Production Market Share by Application (2021-2032)

Figure 75. World GPU Inference Server Production Value Market Share by Application (2021-2032)

Figure 76. World GPU Inference Server Average Price by Application (2021-2032) & (K US\$/Unit)

Figure 77. GPU Inference Server Industry Chain

Figure 78. GPU Inference Server Procurement Model

Figure 79. GPU Inference Server Sales Model

Figure 80. GPU Inference Server Sales Channels, Direct Sales, and Distribution

Figure 81. Methodology

Figure 82. Research Process and Data Source

I would like to order

Product name: Global GPU Inference Server Supply, Demand and Key Producers, 2026-2032

Product link: <https://marketpublishers.com/r/GF09B4D54626EN.html>

Price: US\$ 4,480.00 (Single User License / Electronic Delivery)

If you want to order Corporate License or Hard Copy, please, contact our Customer Service:

info@marketpublishers.com

Payment

To pay by Credit Card (Visa, MasterCard, American Express, PayPal), please, click button on product page <https://marketpublishers.com/r/GF09B4D54626EN.html>