**Hewlett Packard Enterprise**

# QUALCOMM CLOUD AI 100 ACCELERATOR FOR HPE (R9Q09A)

## Computational and Graphics Accelerators for Servers



## WHAT'S NEW

- Compact, single-slot, low-profile, and low-power PCIe accelerators is optimized for AI interference infrastructure at the edge.

- Based on Qualcomm AI Core, Inference Optimized Architecture leverages over a decade of power-efficient deep learning technology for inference computing.

- Achieves impressive peak TOPS(trillion operations per second) performance of 350 Int8, 175 FP16.

- High performance can dramatically lower infrastructure acquisition costs.

## OVERVIEW

Does your organization require optimized infrastructure for deploying AI/ML trained models at the edge? The Qualcomm Cloud AI 100 Accelerator for HPE dramatically improves server performance on AI/ML inference workloads for faster insights and higher accuracy data transformations on computer vision, natural language, and other state-of-the-industry AI/ML use cases. Based on the Qualcomm AI Core, this server accelerator is tailored for inference, supporting model customization and quantization to lower precision datatypes for optimized performance and industry-leading power efficiency. With low power consumption, the Qualcomm Cloud AI 100 Accelerator for HPE can increase performance while lowering operating cost.

## FEATURES

### High Performance

Purpose-designed for AI/ML workloads, the Qualcomm Cloud AI 100 Accelerator for HPE is a market leader in performance and power efficiency based on industry-standard benchmarks including MLPerf.

Up to eight of these low-profile PCIe accelerators can be configured within an HPE Edgeline EL8000.

Each standard card delivers 350 TOPS while consuming only 75 watts of power.

Using ResNet-50 benchmark, performance translates to inferencing up to over 100,000 images per second in a single HPE Edgeline EL8000 system.

### Compact and Energy Efficient

Each Qualcomm Cloud AI 100 Accelerator for HPE is rated at 75 Watts TDP.

Qualcomm Cloud AI 100 Accelerator for HPE design is a PCIe Gen4 (x8) connector in a low-profile form factor (both half-height and half-length) for compact use at the edge.

### Easy to Use

The Qualcomm Cloud AI 100 Accelerator for HPE leverages Qualcomm's 6th generation of AI/ML technologies and includes a comprehensive SDK and extensive suite of developer tools (compiler, performance optimization, and quantization tools) and code samples.

Features simplified support for popular AI/ML frameworks.

## Technical specifications     Qualcomm Cloud AI 100 Accelerator for HPE

| | |
|---|---|
| **Product Number** | R9Q09A |
| **Peak Single Precision Performance** | Peak Integer (INT8) / Floating Point (FP16) Ops.  Up to 350 TOPS / 175 TFLOPS |
| **Number of accelerators per card** | 14 AI cores per PCIe Gen4 Accelerator card. HPE Edgeline e920 supports 1xPCIe card. HPE Edgeline e920d supports 4xPCIe cards.  HPE Edgeline e920t supports 3xPCIe cards. |
| **Cores** | 14 Qualcomm AI Cores supported on each AI 100 PCIe card |
| **Memory size per board** | 16 GB |
| **Accelerator applications** | Supports industry-standard AI/ML trained models |
| **Architecture features** | High throughput, low-latency and low-energy AI/ML inference for deriving actionable insights from computer vision and natural language processing<br>low-profile, PCIe Gen4 card, half-height, half-length.  Operates at up to 75 Watt TDP. |
| **System** | HPE Edgeline e920, e920d, e920t |
| **Product Dimensions (metric)** | Half-height (68.9 mm), half-length (167.65 mm), single-width (18.11 mm) PCIe card |
| **Weight** | 384 g with standard I/O bracket |
| **Warranty** | 3/1/2001 |

# HPE POINTNEXT SERVICES

HPE Pointnext Services brings together technology and expertise to help you drive your business forward and prepare for whatever is next.

## Operational Services from HPE Pointnext Services

HPE Pointnext Tech Care provides fast access to product-specific experts, an AI-driven digital experience, and general technical guidance to help enable constant innovation. We have reimagined IT support from the ground up to deliver faster answers and greater value. By continuously searching for better ways to do things—as opposed to just fixing things that break—HPE Pointnext Tech Care helps you focus on achieving your business goals.

HPE Pointnext Complete Care is a modular, edge-to-cloud IT environment service that provides a holistic approach to optimizing your entire IT environment, and achieving agreed upon IT outcomes and business goals through a personalized and customer-centric experience. All delivered by an assigned team of HPE Pointnext Services experts.

**HPE Integration and Performance Services** help you customize your experience at any stage of your product lifecycle with a menu of services based on individual needs, workloads, and technologies.

- Advise, design, and transform
- Deploy
- Integrate and migrate
- Operate and improve

- Financial Services
- Greenlake Management Services
- Retire and sanitize
- IT Training and personal development

## **Other related services**

HPE Education Services delivers a comprehensive range of services to support your people as they expand their skills required for a digital transformation. Consult your HPE Sales Representative or Authorized Channel Partner of choice for any additional questions and support options.

**Defective Media Retention** is optional and allows you to retain Disk or eligible SSD/Flash Drives replaced by HPE due to malfunction.

# HPE GREENLAKE

HPE Greenlake is HPE's market-leading IT as-a-Service offering that brings the cloud experience to apps and data everywhere – data centers, multi-clouds, and edges – with one unified operating model. HPE GreenLake delivers public cloud services and infrastructure for workloads on premises, fully managed in a pay per use model.

If you are looking for more services, like **IT financing solutions**, please explore them here.

**Hewlett Packard** Enterprise