QumulusAI

NEWS RELEASE

QumulusAI Signs $32 Million, Two-Year NVIDIA Blackwell B300 Agreement With AI Inference Platform Provider

2026-07-23

The customer, an inference platform serving production generative AI applications, will run image, video and other generative workloads on dedicated Blackwell B300 clusters from QumulusAI's hyperdistributed AI cloud

ATLANTA--(BUSINESS WIRE)-- QumulusAI (Nasdaq: QMLS), a neocloud infrastructure provider purpose-built for the AI computing era, today announced a signed two-year agreement, valued at more than $32 million, to supply NVIDIA Blackwell B300 capacity to an AI inference platform provider focused on generative AI applications. The agreement includes renewal options, with capacity expected to come online in the fall of 2026.

The customer's platform delivers fast, scalable infrastructure for generative AI applications, serving developers and enterprises that run image, video and other generative models where speed and reliability determine the user experience. Under the agreement, QumulusAI will provide dedicated GPU clusters that give the platform committed, high-performance capacity as demand from its customers scales.

Serving generative media workloads at scale typically calls for sustained, dedicated compute rather than opportunistic spot capacity — a dynamic reflected in multiyear commitments such as this agreement. Capacity will be served from QumulusAI's U.S. data center footprint. The company's demand-led deployment model places capacity into available pockets of power across a distributed network of colocation and owned facilities, enabling it to bring GPU capacity online in months, not years.

The agreement adds a two-year commitment of more than $32 million to QumulusAI's book of business.

“Generative media workloads put real pressure on inference infrastructure — images and video are compute-intensive to serve, and the user experience depends on speed,” said Mike Maniscalco, CEO of QumulusAI. “This agreement reflects a pattern we're seeing in our own business: inference customers want dedicated, committed capacity they can count on, and our model is built to put that capacity to work quickly."

About QumulusAI

QumulusAI is a distributed AI cloud platform that delivers accelerated access to high-performance GPU compute. Through an inference-first, demand-led deployment model across a network of data center sites, QumulusAI brings compute closer to customer demand, helping AI teams and enterprises scale production AI workloads with speed, flexibility and control. By combining rapid deployment with flexible private cloud infrastructure, QumulusAI gives customers a faster, more adaptable path beyond the capacity constraints of traditional centralized and hyperscale cloud models. Learn more at QumulusAI.com.

Follow us on LinkedIn and X @QumulusAI.

Forward-Looking Statements

This press release contains forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, including statements regarding the company's agreement, the anticipated value and term of the agreement, the expected timing of deployment of NVIDIA Blackwell B300 capacity, the company's ability to provide dedicated GPU clusters that give the platform committed, high-performance capacity, the company's ability to serve the agreement from its existing data center footprint, and the pace at which the company can deploy capacity and bring GPU compute online. Words such as "anticipate," "believe," "estimate," "expect," "guidance," "intend," "can," "may," "on track," "plan," "project," "target," "will" and similar expressions are intended to identify forward-looking statements. These statements are based on management's current expectations and assumptions as of the date of this release and are subject to risks and uncertainties that could cause actual results to differ materially, including, among others, the company's dependence on a limited number of large customers; the availability and cost of power, network connectivity and specialized hardware such as graphics processing units; the company's substantial capital requirements and access to financing; competition and rapid technological change in the high-performance computing and AI markets; the company's limited operating history and history of net losses; and those described in the "Risk Factors" section of the company's registration statement on Form S-1, as amended (File No. 333-292514), filed with the U.S. Securities and Exchange Commission (SEC), as such factors may be updated in the company's subsequent filings with the SEC. QumulusAI undertakes no obligation to update or revise any forward-looking statement, whether as a result of new information, future developments or otherwise, except as required by applicable law.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20260723459787/en/

Investor Contact
investors@qumulusai.com

Media Contact
media@qumulusai.com

Source: QumulusAI