

Driving AI Innovation:

Hypertec Cloud, Together AI and VAST Data



together.ai

Hypertec Cloud, a division of the Hypertec Group, is a multifaceted business with deep expertise in high-performance computing, AI infrastructure, and data center construction. Hypertec Cloud leverages its 40 years of experience in high-performance hardware and data center innovation to create bespoke AI infrastructure solutions. Hypertec Cloud is known for its innovative solutions in manufacturing high-performance hardware and developing cutting-edge data centers. Their offerings range from Infrastructure-as-a-Service (IaaS) to custom-tailored clusters for AI workloads. Hypertec Cloud's commitment to sustainability and efficiency is evident in their pioneering efforts in direct-to-chip liquid cooling, immersion cooling, and other energy-efficient technologies.

Hypertec Cloud's customer base is equally diverse, including Enterprise, Media and Entertainment (M&E), AI Platforms (PaaS), and developers of Frontier Large Language Models (LLMs) looking for high-performance, scalable infrastructure. Their ability to build clusters on demand and deliver them rapidly has made them a preferred partner for organizations with demanding workloads.

According to David Bitton, Vice President of AI and Product Strategy at Hypertec Cloud, the company's approach is rooted in innovation, flexibility, and sustainability. "Because of the strength of combining data center construction and hardware manufacturing, we have the luxury to build clusters on demand. This creates tailor-made solutions with the latest technologies," Bitton explained.

Bitton highlighted, "Innovation is key in whatever we do." This commitment to innovation led Hypertec Cloud to utilize the VAST Data Platform as part of the Hypertec Cloud offering. David underscored the strategic nature of the VAST Data collaboration: "Since day one when we decided to jump into AI, VAST Data was very involved with us. We got a lot of help in understanding and getting up to speed very quickly on AI. From the CEO to the engineering team, the level of support was exceptional." The partnership allowed Hypertec Cloud to deliver high-performance, tailor-made infrastructure solutions to its clients rapidly, even under exceptionally tight deadlines.



Together AI: Empowering AI with Scalability and Performance

Hypertec Cloud's ability to deliver large-scale clusters in record time has been pivotal for its customers. For instance, Hypertec Cloud retrofitted and deployed a data center for Together AI in under 60 days, a feat that underscored the company's capability to execute complex projects with speed and precision. "This was a revelation moment for both Together AI and Hypertec Cloud," Bitton said, reflecting on how the achievement solidified the partnership.

Together AI stands at the forefront of AI acceleration, offering an end-to-end platform designed to streamline the entire AI lifecycle. This includes robust tools for training, fine-tuning, and deploying generative AI models at scale. Their services extend to offering access to over 200 open-source models, custom fine-tuning options, and a cutting-edge inference platform that leverages proprietary optimizations like the Together Kernel Collection (TKC) to maximize performance.

Additionally, Together AI provides seamless integration with hardware orchestration, enabling clients to deploy models in various environments—whether on-premises, via API, or on dedicated GPU clusters. Deveaux Barron, Advisor, at Together AI, emphasized the collaboration's importance: "This partnership combines Together AI's expertise with Hypertec Clouds infrastructure and the VAST Data Platform to deliver transformative solutions for AI"



"This partnership combines Together AI's expertise with Hypertec Clouds infrastructure and the VAST Data Platform to deliver transformative solutions for AI,"

— **Deveaux Barron**, Advisor, at Together AI

TKC, which integrates advanced kernel-level optimizations to maximize GPU performance, improves GPU efficiency by up to 70% for inference and 10-20% for training, enabling significant reductions in computational overhead. Together AI's platform also supports seamless transitions across the AI pipeline, from model training and fine-tuning to deployment and inference, with a particular focus on speculative decoding, model distillation, and auto fine-tuning capabilities. Barron emphasized the tangible benefits: "Our customers experience lower operational costs, accelerated innovation, faster time-to-market, and unparalleled performance integration, all while leveraging the best of modern AI technology."

The partnership with Hypertec Cloud has also enabled Together AI to address the growing demand for scalable AI clusters. As Barron pointed out, "We're building more scalable, high-performance, cost-efficient clusters globally in 2025."

AI Transformation

The VAST Data Platform is a cornerstone of this collaborative effort. By providing reliable, fast, and AI-oriented solutions, VAST ensures that Hypertec Cloud and Together AI can meet their clients' needs effectively. "VAST Data's commitment to continuous innovation aligns with our mission to drive AI adoption across industries," stated Barron .

Bitton echoed this sentiment, highlighting the unique value that VAST brings to the partnership: "From day one, VAST has been deeply involved in supporting us. Their engineering team has helped us optimize performance at every level, ensuring uptime and reliability for our AI training and inferencing workloads."

A Case of Collaborative Innovation

Success lies in the collaborative approach of all three companies. Hypertec Cloud 's ability to rapidly deploy clusters, Together AI's cutting-edge AI solutions, and the VAST Data Platform collectively create a seamless workflow that drives innovation. This synergy has enabled Together AI to develop and deploy LLMs at scale, benefiting their diverse client base, which ranges from startups to large enterprises.

Barron described the experience as transformative: "Working with Hypertec Cloud and VAST has enabled us to deliver unparalleled performance, reliability, and support to our customers. This collaboration has been critical in realizing our vision of building the leading computational AI platform."

About VAST Data

VAST Data is the data platform company built for the AI era. As the new standard for enterprise AI infrastructure, organizations trust the VAST Data Platform to serve their most data-intensive computing needs. VAST Data empowers enterprises to unlock the full potential of their data by providing AI infrastructure that is simple, scalable, and architected from the ground up to power deep learning and GPU-accelerated data centers and clouds. Launched in 2019, VAST Data is the fastest growing data infrastructure company in history. For more information, please visit and follow VAST Data on [X \(formerly Twitter\)](#) and [LinkedIn](#).



Scan [here](#) to learn more.

