

CASE STUDY



AssemblyAI Delivers AI-Powered Audio Intelligence with NVIDIA A100 GPUs and Cirrascale Cloud Services

AssemblyAI is a leading Speech Recognition and Audio Intelligence startup focused on making advanced Deep Learning technology accessible to everyday developers and global enterprises.

THE CUSTOMER

From startups to Fortune 500 companies, developers use AssemblyAI to build innovative applications with their Speech-to-Text and Audio Intelligence APIs. Backed by leading investors such as Insight Partners, Accel and Y Combinator, AssemblyAI powers highly accurate transcription and powerful audio analysis for hundreds of customers every day.

AssemblyAI is doing amazing things by enabling their customers to automatically convert audio/video files and live audio streams into text and build powerful applications with features like Summarization, Entity Detection, Content Moderation, Sentiment Analysis, PII Redaction, and more.

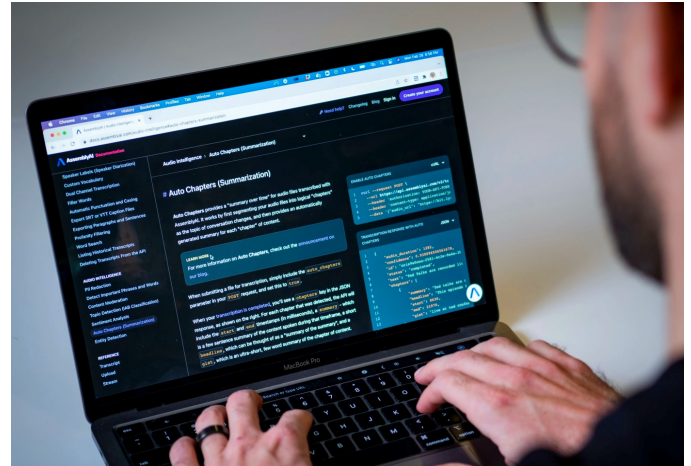


Cirrascale Cloud Services has enjoyed a long-term partnership with AssemblyAI over the past five years. The companies have worked closely together as AssemblyAI has grown from a small start-up to a company making waves in AI and now powered by \$30M in new Series B funding from its investment partners while taking the world of speech-to-text by storm.

To learn more about AssemblyAI, visit their website at www.assemblyai.com.

THE CHALLENGE

Over the course of the past five years, AssemblyAI has utilized a variety of specialized hardware configurations to meet the needs of their developers, and ultimately to accelerate the overall iteration speeds of their models. Earlier this year,



AssemblyAI realized it was becoming critical they begin to look at shifting their thinking. They needed to move from renting dedicated cloud resources to owning their own managed hardware with Cirrascale's help in order to better control their overall compute resource expenses. "We've tried running our own GPU servers on-premises in the past, but servers kept going offline due to a myriad of problems like disk issues, power surges, or broken GPU cards. This meant our team had to spend time debugging hardware versus focusing on our core competencies. We knew we needed something different. Basically, we needed the best of what cloud offered us, but we wanted to take advantage of our NVIDIA[®] Inception Member benefits as well and purchase NVIDIA HGX[™] A100 servers." said Dylan Fox, founder and CEO for AssemblyAI.

THE SOLUTION

As an NVIDIA Partner Network Elite CSP Partner and Solution Provider, Cirrascale has worked with customers to develop hybrid CapEx/OpEx solutions that make sense. In this case, Cirrascale provided AssemblyAI with the ability to purchase and deploy a large-scale NVIDIA HGX A100 cluster with NVIDIA high-speed networking and a flash-based, hot-tier storage solution.



The large-scale NVIDIA A100 cluster is owned by AssemblyAI but is deployed and managed by Cirrascale, and is interconnected with additional cloud-based resources

(continued on reverse side)

that AssemblyAI uses for ongoing development. This hybrid managed and hosted cloud service approach helped to give AssemblyAI the perfect hybrid solution. "We were really concerned that we were not going to be able to get the total solution we were looking for. Cirrascale's hybrid approach ensures our team isn't spending time maintaining and debugging infrastructure so they can stay focused and move faster, enabling us to improve the accuracy of our models faster," said Mr. Fox.

Additionally, the managed services Cirrascale provides enables AssemblyAI to utilize the company like an extension of its own IT team. "All of these options were great, and the Cirrascale team was very flexible and helped to keep our costs down, but a main reason we continue to partner with them is the overall support and customer experience we received from Cirrascale," states Mr. Fox. "They really care about our account, which makes us feel like we are working with a true partner who helps us solve our problems as we grow."

"I honestly can't recommend Cirrascale enough. The experience has been great... We plan to use Cirrascale long term for our increasing workloads as our company continues to grow."

Dylan Fox
CEO, AssemblyAI

THE FUTURE

AssemblyAI also utilizes Cirrascale's ability to provide various configurations to easily allow for benchmarking varying models. "Our customers span across the following major industries: contact centers, broadcast media, entertainment, medical, and legal. We provide speech recognition technology to companies in these industries in order to power the features within their products/services," states Mr. Fox.

"Our primary goal is to improve the accuracy of our speech recognition models. This requires a lot of model training, experimentation, and various NVIDIA GPUs to iterate fast enough. With Cirrascale, we're able to experiment and iterate

quickly with a large number of various GPUs at an affordable cost and without having to manage a fleet of GPU servers ourselves."

DISCOVER THE DIFFERENCE

Cirrascale Cloud Services worked closely with AssemblyAI, as it does all of its customers, to truly deliver a white glove experience. The company's ability to provide tailored, dedicated, multi-GPU cloud services has set it apart from other cloud providers along with the ability to support tiered storage and truly scalable network architecture supporting NVIDIA InfiniBand and 100GbE connectivity. With flat-rate billing, no ingress/egress fees on data, and superior support practices, Cirrascale invites you to discover the difference.