



Giving a Voice to Advanced Speech Recognition Technology

AssemblyAI is an applied artificial intelligence company specializing in Advanced Speech-to-Text models customized to its customer's use case for higher accuracy.

The Customer

AssemblyAI® is a team of researchers, engineers, and designers that have worked at some of the largest technology companies all over the world. Funded by Y-Combinator, the company's core offering is speech recognition technology offering higher accuracy by building custom speech recognition models tailored to the customer's specific use case. This unique solution enables customers to transcribe dozens of audio files in parallel in batch mode, or transcribe audio streams in real-time for interactive applications.

AssemblyAI offers its customers an on-premises engine that customers can license and run on their own servers for privacy and latency sensitive applications. Additionally, they offer a secure cloud API for fast transcription.

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

The Challenge

AssemblyAI needed their team to have access to reliable, cost-effective hardware for the development of their Speech-to-Text application. The company was faced with the common dilemma of purchasing their own hardware and running it on-premises, hosting it off-site, or utilizing costly virtualized cloud services through a third-party big tech cloud provider.

Dylan Fox, chief executive officer for AssemblyAI states, "We've tried running our own GPU servers on-premises, 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. Using other cloud services like AWS or GCP was very costly and the support was not great."



AssemblyAI needed to ensure whatever solution they chose would give their team the ability to focus on their core competencies of model building and training. Mr. Fox continued, "We needed to ensure our team wasn't spending time maintaining and debugging GPU infrastructure so they could stay focused and move faster, enabling us to improve the accuracy of our models faster."

(continued on reverse side)

The Solution

Cirrascale Cloud Services® worked closely with AssemblyAI in creating a dedicated, multi-GPU solution that fulfilled all of the requirements set forth by the customer. “For large scale, continuous training we found the big tech clouds to be expensive and inflexible,” stated Mr. Fox. “However, with Cirrascale we have a variety of hardware configurations and GPU cards we can configure to best suit our needs and budget. Plus, the level of support and customer experience is much better working with Cirrascale – it feels more like a complete managed solution.”

Ultimately, AssemblyAI selected Cirrascale’s cloud platform due to Cirrascale’s ability to provide a dedicated multi-GPU platform that could scale and meet AssemblyAI’s needs. “The hardware options are great and very flexible, the cost is very affordable, but a main reason was the support and customer experience we received from Cirrascale,” states

“I honestly can’t recommend Cirrascale enough. The experience has been great... we plan to use Cirrascale long term and for our increasing work loads as our company grows!”

*Dylan Fox
CEO, AssemblyAI*

Mr. Fox. “They really cared about our account, which made us feel like we were working with a true partner who would help us solve our problems as we grow.”

AssemblyAI also utilized 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 GPUs in order to iterate fast enough. With Cirrascale GPU servers, we’re able

to experiment and iterate quickly with a large number of 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 multi-tiered storage and truly scalable network architecture supporting 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.

Cirrascale Cloud Services 5775 Kearny Villa Road, San Diego, CA 92123 USA **Phone** 888-942-3800 **Web** www.cirrascale.com

©2019 Cirrascale Cloud Services. All Rights Reserved. Cirrascale and the Cirrascale logo are registered trademarks of Cirrascale Cloud Services. ZeroEyes is a trademark or registered trademarks of ZeroEyes LLC. All other names or marks are property of their respective owners. No part of this document may be reproduced without consent from Cirrascale Cloud Services. Technical specifications and pricing subject to change without notice.



CSM102 - REV A - 06/2019