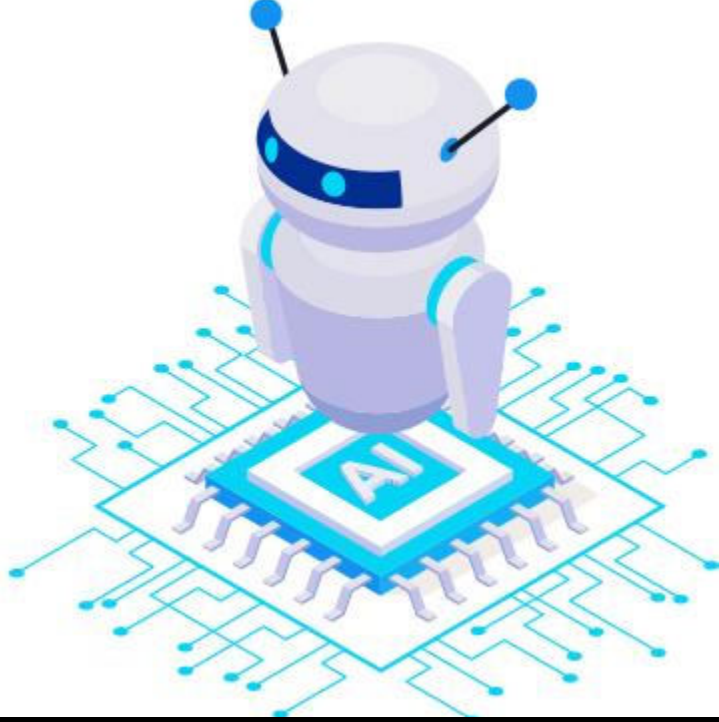**E2E Cloud**

Helped an **Artificial Intelligence** company from Insurance and Automotive Industry to run their mission critical **AI/ ML application** on **NVIDIA A100 GPU Cloud servers** on **E2E Cloud**

How E2E Cloud fulfilled the requirement to run AI/ ML Inference application with 24*7 availability
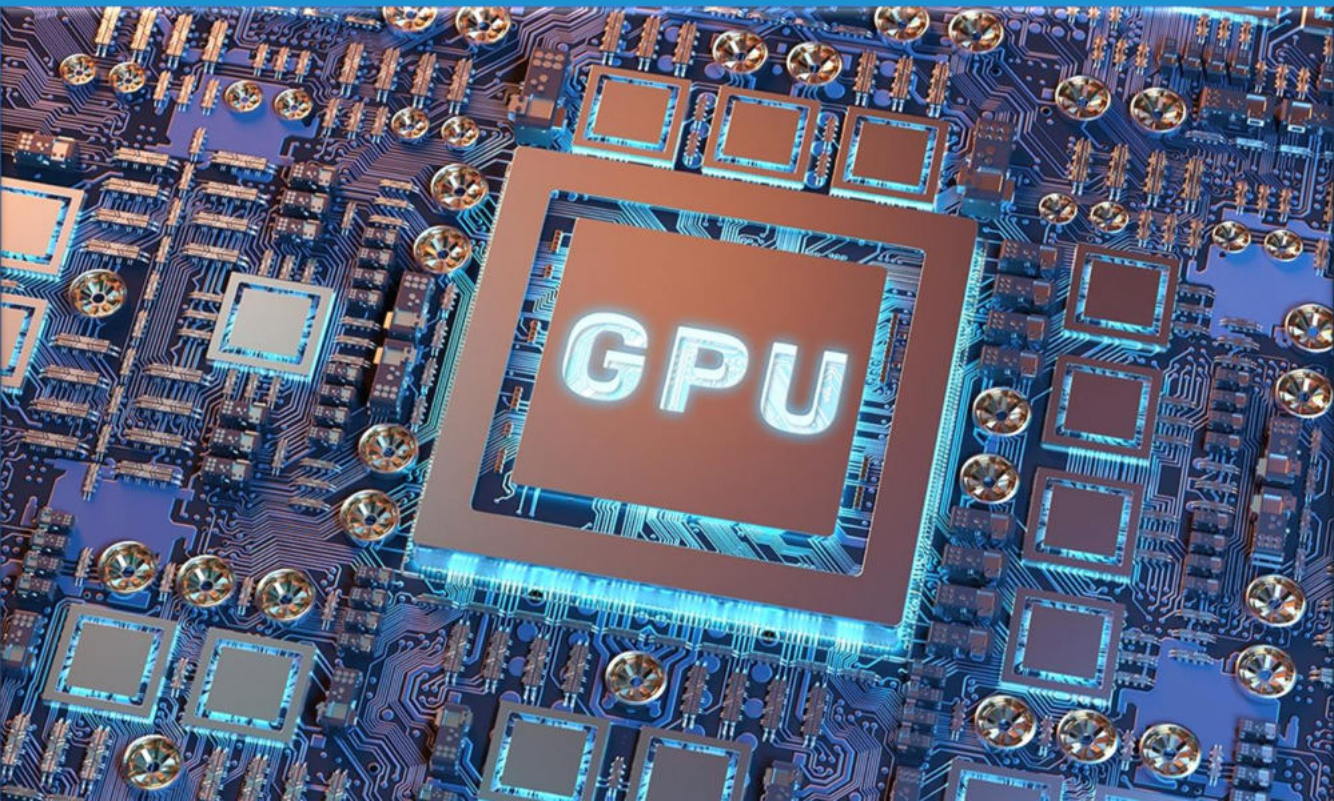
About Company

The company is one of the largest AI/ML companies in Bangalore targeting Insurance, Automobile and logistics industry and serving big MNCs of India like Bajaj Insurance, Mahindra First Choice and SBI General insurance. The company was recently awarded NASSCOM Emerge 50 award.

Pain Points

The company was using on premise GPU servers for training. When their product was ready for live deployment, public Cloud GPU servers were required.

The application required 24*7 uptime so on premise servers were out of question.

Requirements

This customer needed security of a dedicated firewall, options of saved images and dedicated GPU card public cloud servers wherein there are no noisy neighbor problems. This customer wanted 100% uptime SLA and was also expecting yearly committed nodes for further reducing their costs. The customer wanted GPU Cloud servers in the India region to give better latency to their end clients.

Solutions

This company tried various options on AWS but either the pricing was too high or configuration provided was too low. Finally, this company got to know that E2E Networks is the only Cloud service provider partnered with NVIDIA to sell GPU Cloud servers in India.

The price to performance ratio offered by **E2E Cloud for A100 GPU Cloud servers** were best what they can get in India and so they decided to deploy their application on E2E Cloud:

https://www.e2enetworks.com/nvidia-a100/

It's been **more than 2 years** and the client is up and running and growing their business with full confidence on **E2E Cloud**.

Customer with the help of NVIDIA NGC containers, migrated the application to E2E Cloud. Redundant architecture was set up so that a reliable, highly available application experience can be given to the end customers. Data was smoothly uploaded to E2E Object storage and from there to E2E GPU Cloud servers .

E2E Cloud

Contact Us:

sales@e2enetworks.com
+91-11-3001-8095
www.e2enetworks.com

Stay Connected