25 July, 2024:

Transcript:

During Nosana's July Community Call, Caroline shared updates on recent events like ETHCC and upcoming ones including Solana Breakpoint in Singapore and the World AI Summit in Amsterdam. She highlighted a partnership with Theoriq and ongoing initiatives such as a giveaway for tickets to World AI Summit. Jesse discussed advancements in AI models, including the efficient deployment of the Llama 3.1 and the upcoming Mistral Large 2. Jaïr provided technical updates on new node onboardings, a new node image caching system, and upcoming features like a node leaderboard and enhanced Explorer. The call concluded with community acknowledgments and a reminder to vote for the giveaway finalists.

Rochelle:
Hello, Nosana community! Thank you for joining our July Community Call. I'm Rochelle, your host for today. We have an exciting agenda lined up and some fantastic updates to share. Joining me today are Jesse, Jaïr, and Caroline. Let's get started with Caroline, who will cover our recent and upcoming events and updates.

Caroline:
Thank you, Rochelle. It's been a busy month for us at Nosana. Our team had an incredible time at ETHCC, connecting with future clients, partners, and community members. We also participated in Encode's event, where Sjoerd spoke on a DePIN-focused panel about Nosana's visions.

As summer winds down, we look forward to September, when event activities pick up again. The Nosana team will be heading to Solana Breakpoint in Singapore at the end of September. This is an exciting milestone as it's our first trip to Asia.

In October, we'll be at the World AI Summit and World AI Week in Amsterdam, our hometown. We'll be connecting with AI industry leaders, developers, and startups. We're also hosting our own event on October 8th with industry partners, clients, and our amazing community.

We're thrilled that our Nosana Insiders, who've been doing an excellent job with community-curated content on X, will join us at the World AI Summit. Expect some behind-the-scenes content from them.

We also have a giveaway for two more tickets to the World AI Summit. Today, we'll be voting on the three finalists, and the community will choose the final winner who will join us in Amsterdam.

Regarding partnerships, we've recently joined forces with Theoriq, a high-quality AI agent network. Jesse will share more about this partnership later in the call. We also hosted an AMA on this partnership, which is available on our X.

Lastly, there's exciting news about multi-GPU support. Although it's not yet released, stay tuned for a special partner announcement related to this on Tuesday.

We've participated in several podcasts this month, including DePIN Unleashed and Solana Uni Space. We're proud to be highlighted in Grayscale's recent report on decentralized AI. You can read the full report through links on our LinkedIn or X account.

We're also hiring! We have several open roles and prefer to hire directly from our community. If you're passionate about Nosana, engage with us on Discord and X. More open roles will be posted soon, so keep an eye on our LinkedIn.

Jesse:
Thanks, Caroline. Now, let's dive into what's new on the AI-inference side of the network. Our AI team has been experimenting with several new models on the grid, including the 70B Llama using frameworks like LMDeploy, vLLM, and Ollama. These experiments have provided valuable insights, and we're working on sharing these with the wider AI ecosystem soon.

We are compiling an extensive template directory to help users quickly set up common AI endpoints. Our community has also initiated some impressive testing—shoutout to cgn_1 for their contributions.

Client testing continues to show that the Nosana CLI is becoming a highly efficient tool for deploying AI endpoints. You can open WebUIs for Jupyter, Llama, and Stable Diffusion within a minute using any necessary GPUs.

We're excited about the Llama 3.1 release. The 8B model is almost as powerful as the 70B model, which is remarkable. Nosana efficiently runs both the 8B and 70B 3.1 models within hours of their release at high tokens per second. Additionally, the Mistral Large 2 model is coming out soon. Exciting times ahead!

I'll hand the mic over to Jaïr to talk about some Test Grid and technical updates.

Jaïr:
Thanks, Jesse. Now for the technical updates. We've onboarded several new nodes, including A5000, A100, 4090s, and H100 SXM. Some existing A100s were removed, and we've added 10 new H100 nodes, with plans to selectively onboard more.

For the A100s, we've onboarded 90 nodes, bringing the total to over 100 active A100 nodes. The A5000 market has doubled from 18 to 34 nodes, and we'll be onboarding 86 nodes for the 4090 market, with 200 on the waitlist.

We've created a new market for A100 40GB nodes and, in total, have onboarded an additional 200 nodes. Onboarding will be paused until the end of August.

All markets have been updated to use the new SFT collection, which allows atomic swaps and facilitates nodes in changing their GPU. We're also rolling out a new node image caching system to speed up the Nosana CLI's AI-endpoint setup.

We've put significant effort into benchmarking job stats, providing granular details of node performance and facilitating 'proof of GPU'. New nodes will need to execute a benchmark as their first job to join the network.

A node leaderboard is coming soon, allowing the community to see and compare performance statistics. We're also developing a new Explorer to resolve current performance issues.

Rochelle:

Thank you, Jaïr. Now for a quick, well-deserved shout-out to two of our community members: MachoDrone and ersugteralbaner. You two rock! Thanks for all the extra help and support you've given our community members in the server, especially regarding onboarding and Test Grid! Round of applause, everyone!

Thanks to all our speakers, and to all of you who joined us today. Don't forget to vote and choose a finalist who will win tickets to the World AI Summit, the post is on our X account. And, as always, tune into our next community call, which is the last Thursday of every month, so see you back here, same time, same place, on August 29th! Thanks, everyone!