

Cool It: Team Tackles the Thermal Challenge Data Centers Face

A U.S. Department of Energy grant will help NVIDIA and seven partners build an advanced liquid-cooling system.

Author: Rick Merritt

Two years after he spoke at a conference detailing his ambitious vision for cooling tomorrow's data centers, Ali Heydari and his team won a \$5 million grant to go build it.

It was the largest of 15 awards in May from the U.S. Department of Energy. The DoE program, called COOLERCHIPS, received more than 100 applications from a who's who list of computer architects and researchers.

"This is another example of how we're rearchitecting the data center," said Ali Heydari, a distinguished engineer at NVIDIA who leads the project and helped deploy more than a million servers in previous roles at Baidu, Twitter and Facebook.

"We celebrated on Slack because the team is all over the U.S.," said Jeremy Rodriguez, who once built hyperscale liquid-cooling systems and now manages NVIDIA's data center engineering team.

The project is ambitious and comes at a critical moment in the history of computing.

Processors are expected to generate up to an order of magnitude more heat as Moore's law hits the limits of physics, but the demands on data centers continue to soar.

Soon, today's air-cooled systems won't be able to keep up. Current liquid-cooling techniques won't be able to handle the more than 40 watts per square centimeter researchers expect future silicon in data centers will need to dissipate.

So, Heydari's group defined an advanced liquid-cooling system.

Their approach promises to cool a data center packed into a mobile container, even when it's placed in an environment up to 40 degrees Celsius and is drawing 200kW — 25x the power of today's server racks.

It will cost at least 5% less and run 20% more efficiently than today's air-cooled approaches. It's much quieter and has a smaller carbon footprint, too.

"That's a great achievement for our engineers who are very smart folks," he said, noting part of their mission is to make people aware of the changes ahead.

The team's solution combines two technologies never before deployed in tandem.

First, chips will be cooled with cold plates whose coolant evaporates like sweat on the foreheads of hard-working processors, then cools to condense and re-form as liquid. Second, entire servers, with their lower power components, will be encased in hermetically sealed containers and immersed in coolant.

They will use a liquid common in refrigerators and car air conditioners, but not yet used in data centers.

The three-year project sets annual milestones — component tests next year, a partial rack test a year later, and a full system tested and delivered at the end.

Icing the cake, the team will create a full digital twin of the system using NVIDIA Omniverse, an open development platform for building and operating metaverse applications.

The NVIDIA team consists of about a dozen thermal, power, mechanical and systems engineers, some dedicated to creating the digital twin. They have help from seven partners:

Binghamton and Villanova universities in analysis, testing and simulation

BOYD Corp. for the cold plates

Durbin Group for the pumping system

Honeywell to help select the refrigerant

Sandia National Laboratory in reliability assessment, and

Vertiv Corp. in heat rejection

“We’re extending relationships we’ve built for years, and each group brings an array of engineers,” said Heydari.

Of course, it’s hard work, too.

For instance, Mohammed Tradat, a former Binghamton researcher who now heads an NVIDIA data center mechanical engineering group, “had a sleepless night working on the grant application, but it’s a labor of love for all of us,” he said.

Heydari said he never imagined the team would be bringing its ideas to life when he delivered a talk on them in late 2021.

“No other company would allow us to build an organization that could do this kind of work — we’re making history and that’s amazing,” said Rodriguez.

See how digital twins, built in Omniverse, help optimize the design of a data center in the video below.

Picture at top: Gathered recently at NVIDIA headquarters are (from left) Scott Wallace (NVIDIA), Greg Strover (Vertiv), Vivien Lecoustre (DoE), Vladimir Troy (NVIDIA), Peter Debock (COOLERCHIPS program director), Rakesh Radhakrishnan (DoE), Joseph Marsala (Durbin Group), Nigel Gore (Vertiv), and Jeremy Rodriguez, Bahareh Eslami, Manthos Economou, Harold Miyamura and Ali Heydari (all of NVIDIA).

Original URL: <https://blogs.nvidia.com/blog/2023/05/25/liquid-cooling-doe-challenge/>