Source: *The Business Times*, p1
Date: 29 April 2023

Source: *The Business Times,* p9
Date: 29 April 2023

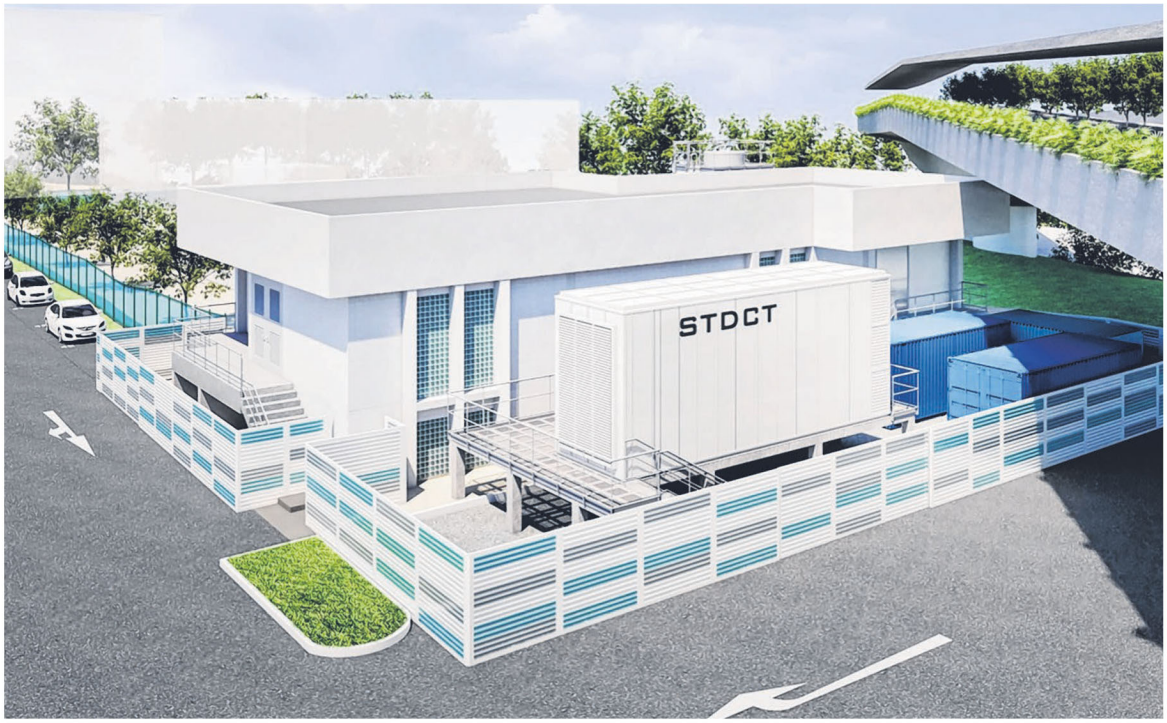


PHOTO: NUS

THE RACE TO KEEP DATA CENTRES COOL

As data centre workloads climb, scientists and data centre operators are pushing the limits of cooling technology to keep temperatures under control.

BY YONG JUN YUAN

A MICROWAVE oven typically puts out about 600 watts to 1,200 watts of heat to pop popcorn and reheat frozen pizzas. Now, imagine a similar amount of power condensed into an area about the size of a postage stamp.

With 80 billion transistors packed into just 814 square millimetres, Nvidia's latest H100 graphics processing unit (GPU) draws 700 watts of power running full-bore. Nearly all of that power is released as heat.

Rather than powering a gamer's desktop, large pools of these GPUs are connected together to verify payments on blockchains, render animations and crunch data for machine learning, among other use cases. >>>

Source: *The Business Times,* p10
Date: 29 April 2023

Some estimates suggest that the large language model ChatGPT would need to operate on the combined power of 30,000 of Nvidia's last-generation A100 GPUs.

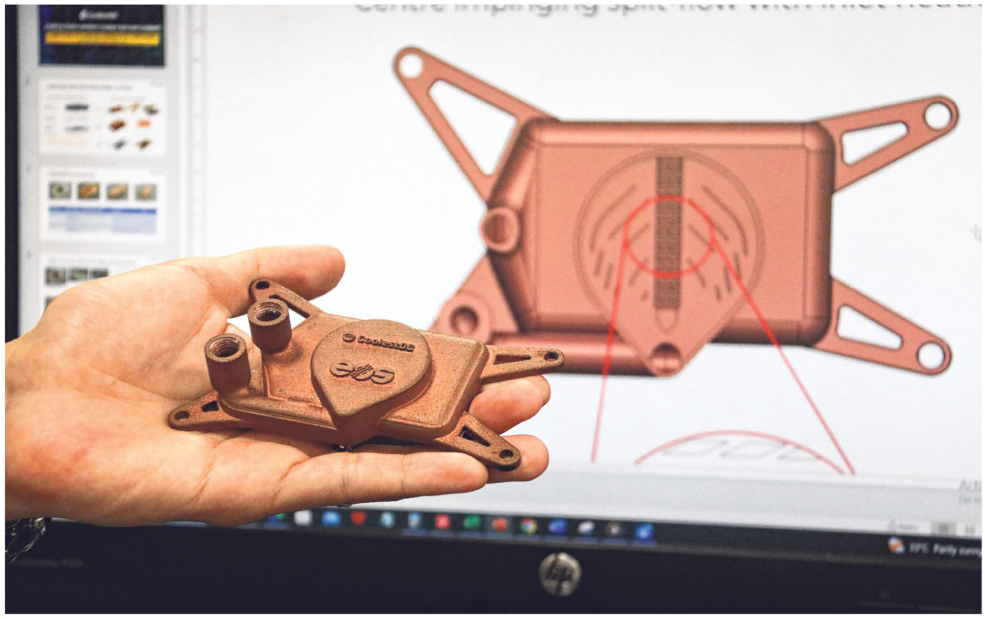
Experts believe that OpenAI, the company behind ChatGPT, could adopt Nvidia's H100 GPUs, which are expected to be up to nine times faster than its predecessor in artificial intelligence training.

The way National University of Singapore (NUS) associate professor Lee Poh Seng sees it, there is a need for higher efficiency cooling solutions to be used as high-compute operations such as ChatGPT produce more heat.

"Associated with all the stringent criteria set by the Economic Development Board (EDB) and Infocomm Media Development Authority (IMDA), I think Singapore will need to be very selective in terms of the kinds of data centres that we host," Prof Lee said.

"In all likelihood, we will only be looking at very high-end data centres. For example, those that actually support artificial intelligence (AI), blockchain, online gaming, (use cases) that require high compute," he added.

The race is heating up for data centres to keep cool while remaining sustainable.



A 3D-printed copper unibody leakless cold plate by NUS' Prof Lee Poh Seng, which reduces the risk of leaks when deployed in a server.
PHOTO: YONG JUN YUAN, BT

Keeping it cool

After a moratorium was placed on the approval of new data centre builds in Singapore, EDB and IMDA released a data centre call for application (DC-CFA) pilot in July 2022.

One of the conditions that applicants have to meet is to achieve a power usage effectiveness (PUE) of 1.3 or better at 100 per cent IT load. PUE is derived by taking the total power consumed by a data centre, divided by the power consumed by IT equipment.

As server manufacturers pack more power-hungry hardware into less space, Prof Lee noted that server rack densities have trended upwards, from less than 10 kilowatts (KW) to about 25 KW. Such high densities exceed traditional air cooling techniques' ability to manage temperatures, he said.

"From the latest announcements by the chipmakers over the past 18 months, (the power trend) is actually increasing sharply again," Prof Lee said.

He added that this has led to increased interest in liquid cooling technologies. As liquid conducts heat better than air, it can net significant efficiency gains.

While liquid cooling has been popular among niche enthusiast communities to squeeze more performance from consumer hardware, implementing such techniques at scale in data centres is not as straightforward.

For instance, the cold plates that gamers use typically have a copper cold plate and acrylic top plate that are secured together with a mechanical gasket. Such designs could potentially leak and spill water over numerous computer systems in a dense server, which would be very costly for its owner.

Instead, Prof Lee has researched and produced copper unibody leakless cold plates through additive manufacturing, commonly known as 3D printing.

By building the cold plates a layer at a time, Prof Lee has also been able to print the oblique fins of the heat sink instead of wire-cutting them, which saves on manufacturing time.

These oblique fins allow for better heat transfer between the heat sink and the water flowing through the cold plates as well.

In trials with data centre player Digital Realty, Prof Lee found that the servers saw efficiency improve from a PUE of 1.4 to about 1.1 when they moved from air cooling to water cooling.

"It can be very substantial. I think you're looking at, easily, 50 per cent and above in reduction (of power consumption)," he said.

Another area of research that Prof Lee's Cooling Efficiency Lab is working on is the use of evaporative cooling, in preparation for future generations of computer chips that consume even more power.

Just as humans perspire to remove heat from our bodies, the evaporative cooling loop that the lab is researching boils a special 3M-patented fluid to remove heat from computer chips more efficiently.

However, NUS senior research fellow Matthew Law said 3M is discontinuing the use of the product due to potential health harms and its impact on the environment.

The search is now on for a suitable substitute. Such a liquid would need to have a boiling point that matches the ideal operating temperature of a chip, conduct heat well and be environmentally friendly. Any fluid that is also a good insulator of electricity would be a bonus.

These experiments as well as others in the lab, will eventually be tested in the Sustainable Tropical Data Centre Testbed (STDCT). Launched by NUS and Nanyang Technological University (NTU) in 2021, the S$23 million research programme will be tested and commissioned in June this year.

The STDCT data centre is housed in a decommissioned power substation in NUS' Kent Ridge campus. It will have a relatively small power capacity of 500 KW, to serve the high-performance compute needs of the universities as various cooling solutions are tested.

Cool collaborations

Seraya Partners managing director Christopher Han said that liquid cooling technologies may also change the operating environment of data centres. Seraya Partners acquired Dodid, a data centre with a total power capacity of 12.5 megawatts at Tagore Lane, in December 2021.

He noted that the entire data hall, where the data centre's servers run, have to be kept at about 23 degrees Celsius now.

"It will actually get hotter, because we don't have to cool the data halls like this anymore," Han said. In effect, cooling can be directed at the components of a server that need it most, without having to cool all the air in a data hall.

ST Telemedia Global Data Centres (STT GDC) group chief technology officer Dan Pointon noted that such technologies require deeper collaboration between data centres and their tenants.

Instead of simply hosting their tenants' equipment, these data centres may now have to work more closely with tenants to use equipment that is compatible with more advanced cooling methods.

As many of these tenants are large multinational companies that have to meet environmental, social and governance (ESG) targets,

NUS professor Praveen Linga is looking at ways to harness cold energy from liquefied natural gas regasification to cool data centres, through the use of a water-based phase-change fluid that could carry four to five times more cold energy per unit volume than chilled water.

Source: *The Business Times,* p11
Date: 29 April 2023

Pointon noted that there had been increased tenant interest in driving further efficiencies in data centre operations.

In addition, tenants are also seeking space and energy savings that will ultimately drive costs down, he said.

STT GDC announced in April 2022 that it had partnered with cooling specialist Iceotope and Schneider Electric to test different liquid cooling systems.

For such solutions to be adopted, Pointon said tenants would have to take cues from existing standards bodies. Technical bodies such as the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) have started developing such standards, he noted.

"One can think of us like an IT hotel where a customer walks through the door – any customer, any server – and we know we can service them in an air-cooled environment," Pointon said.

"There's 30 years of built-up industry experience in terms of how to cool servers (with air), but does that stack up in a liquid-cooled environment? Maybe not," he added, noting that he had only just started seeing server manufacturers create product ranges that are certified for liquid cooling in the last six months.

As tenants catch up with the latest in cooling technologies, data centre operator Equinix is going one step further to offer bare metal servers that customers can access remotely through an application programming interface.

The company has also published information about its trials with liquid cooling at its facilities in New York in January this year.

This represents a shift from the company's traditional co-location model that most data centres adopt, where it charges tenants to place servers in its data centres.

Equinix executive vice-president and general manager of data centre services Jon Lin noted that even though 99 per cent of the company's customers still own their own infrastructure, some are willing to use the new service for its convenience.

"Customers value that tremendously, to be able to just over the course of a weekend, stand up an entirely new geographic zone for their IT infrastructure," he said.

Waste not, want not

Researchers are also looking to reduce energy consumption through the use of waste energy from various sources.

An example of this is NUS professor Praveen Linga's efforts to harness cold energy from the Singapore LNG Terminal to cool various data centres.

In 2019, NUS, Singapore LNG Corporation and Keppel Data Centres announced that they would develop a way to harvest cold energy from the regasification of liquefied natural gas (LNG).

To do so effectively, Prof Linga has adopted semiclathrate hydrates slurries, a water-based phase-change fluid that could carry four to five times more cold energy per unit volume than chilled water.

"You could have data centre clusters built around the island that can be supplied through a pipeline (with) this cold energy that is extracted from the LNG terminals," he said.

Because these semiclathrates are more effective than chilled water at a lower volume, new data centres could also adopt this compound in place of water and reduce the size of the chiller equipment they require. This would reduce the amount of energy needed to pump cold coolant throughout a data centre.

Notably, STT GDC Thailand had earlier also signed a memorandum of understanding in November last year with PTT Digital, an information and communication technology service provider under PTT Group, to harness cold energy from the regasification of liquefied natural gas.

In addition, Pointon said that the company is looking into the use of hydrogen fuel cells to generate power on-site.

As the process generates waste heat, he noted that the company is looking to direct the heat towards a thermal absorption chiller. Such a chiller would be able to take heat as an energy source and drive a compression machine to feed cold water in a data centre.

"That coolant that you produce is going to offset the amount of cooling that you would otherwise need in the data centre," Pointon said.

On Apr 25, Keppel Data Centres announced that it had signed heads of agreement with energy supplier Woodside Energy to potentially purchase approximately 1,000 tonnes of liquid hydrogen per day, and power its data centre facilities as early as 2030.

AI to the rescue

Data centres could also turn to AI to further improve efficiency.

While the technology has been used for some time to optimise data centre operations, NTU professor Wen Yonggang believes that his AI model works better than other simulation models available on the market.

The company he co-founded, Red Dot Analytics (RDA), improves on existing simulation models by building digital twins of data centres with information provided by operators.

Using information such as the power consumption and data hall layouts, RDA can then run simulations for data centres which are more accurate than the industry standard. For instance, temperatures can be predicted to 0.5 deg Celsius of accuracy.

Unlike past AI models, which analyse a data centre's past records to improve its operations, Prof Wen is proud of the fact that his AI model can virtually simulate and generate data for "free".

"If you ask to turn up the temperature (of a data centre) to 27 deg Celsius, I can simulate the environment and... tell you the risk, costs and benefits," he said.

RDA co-founder Angelina Terlaki noted that the company has seen year-to-date demand more than double that of what it saw last year.

"In the past, customers viewed AI as a high-risk technology with unclear benefits," she said. "However, with recent advancements in AI and its increasing prevalence in our daily lives, customers are now more aware of the potential benefits of AI and have become more comfortable using it."

Similarly, Prof Wen noted that the company has seen more demand from companies that would like more accurate data about the carbon emission profiles of their data centres.

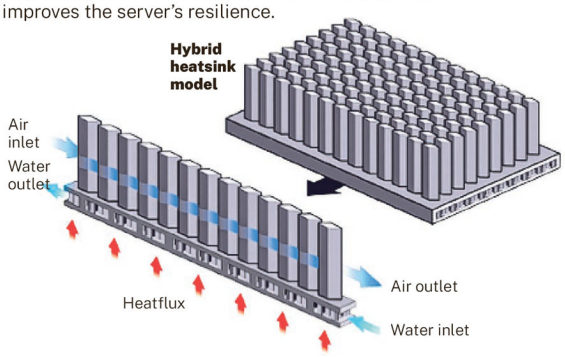
Currently, such data is typically generated by hiring consultants to do on-site assessments and measurements, which is a process that could introduce errors.

"If I use the digital twin, I can compute all this data for you and I know the accuracy... which is much better than just human (consultants)," he said.

junyuany@sph.com.sg

How it works

The hybrid heatsink, which is placed over a computer chip, helps remove heat from the chip with a combination of air and water. The heatsink removes more heat via both water and air through a series of fins, improving cooling efficiency. Even when the water loop supplying water to the heatsink undergoes maintenance, the chip can continue to run at a reduced capacity, which improves the server's resilience.



SOURCE: NUS GRAPHICS: BTVISUAL

PHOTO: RED DOT ANALYTICS

Calvin Sun (far left), Wen Yonggang (centre) and Angelina Terlaki, co-founders of Red Dot Analytics, leverage artificial intelligence to better model the performance of data centres by building digital twins with information provided by operators.