Energy impacts of liquid cooling servers in Minnesota data centers

Data centers house energy-intensive computing equipment that requires significant additional energy to reject excess heat. Most data centers have dedicated computer room air conditioner (CRAC) units to cool the ambient air in the space and then pass it over heat sinks with high-speed fans mounted near or directly on each server. Air cooling strategies consume a significant portion of total energy consumption at data center sites. According to one source, cooling accounts for 40% of all power consumed by data centers as a group,\(^1\) but energy use attributable to cooling at individual sites may vary significantly depending on data center size, application, and usage rate.

Using liquid as a cooling medium in data centers rather than air shows good potential to reduce the energy required for cooling and improve the performance of data center servers. Liquid has a much higher heat capacity than air, providing the same cooling with much lower volume. Liquid can be applied to high power density components directly rather than cooling the ambient temperature of the entire data center, which reduces the overall cooling capacity required. Waste heat from liquid coolant can be rejected to an existing cooling loop, or even recovered for useful purposes, more easily than with air-cooled systems. And there is the potential for data centers designed for liquid cooling to eliminate dedicated air cooling units and local high-speed fans completely.

The potential for liquid cooling was investigated in a CARD grant recently completed by GDS Associates. The technologies assessed in the project were developed by two local companies.

**LiquidCool Solutions (LCS),** based in Rochester, Minn., developed an innovative rack-mounted, total-liquid-submersion-cooled server. Their server is completely submerged in CoreCoolant™, which is a non-hazardous, dielectric fluid. The fluid is continuously circulated from the server, where it absorbs heat from the processor to a cooling unit where the heat is rejected.

**Ebullient Cooling,** based in Madison, Wis., developed a direct-to-chip retrofit technology that can be applied to existing servers by replacing typical air-cooled heat sinks. Coolant is circulated via hose directly to the heat sink on the processor of each individual server. Server heat is absorbed through vaporization of the coolant, which is recirculated back to the fluid distribution unit and then can be rejected to a facility cooling loop or dedicated heat rejection system.

The goals of the study were to:

1) Confirm that the two technologies function effectively in small- and medium-sized data centers without disrupting operations.
2) Characterize the energy impacts to inform consumers and CIP program administrators of the potential value compared to typical air cooling strategies.
3) Evaluate the cost-effectiveness of the new technologies.

The original research plan was to identify three mid-sized (4–10 server rack) data centers for participation—one site for installation and assessment of the LCS technology and two sites for the
Ebullient technology. However, the research plan changed several times over the course of this project because many issues arose with the technologies, site selection, and cost estimates.

The Ebullient product was never installed or evaluated. GDS was ultimately unsuccessful in recruiting a participating test site, in most part due to unforeseen costs associated with the installation of the product. From the results in this study, it appears as though this direct-to-chip cooling technology is not yet mature and that the current projected savings cannot yet justify the costs, which varied so widely as to be unpredictable to use for customer planning purposes.

**Figure 1: Data center test site for LCS product**

The LCS product also encountered application difficulties, but solutions were adopted at an appropriate test site where the product was assessed under four test cases. The original methodology for the study called for measuring energy consumption at a functioning air-cooled data center, installing the liquid-cooled product, and comparing energy consumption afterward. In practice, several unforeseen, uncontrolled variables made the analysis more difficult. First, the existing servers are used for a highly specialized deep learning medical application, which meant that comparisons to the new servers may undercount potential energy savings compared to a typical application.

A second deviation from the plan arose because it was not possible to interrupt the operation of the existing servers to run benchmarking server test programs. Instead, naturally occurring job loads were used to compare the systems. This produces useful results, but with slightly less control of the inputs and under a limited range of operating conditions.

The final major change from the test plan came from the fact that the existing data center houses more servers than the ones being monitored. The power consumption of the entire data center was logged, which allowed a reasonably accurate method of extracting the unaffected loads from the analysis. However, the existence of the unmonitored loads adds noise to the data and could lower the accuracy of the findings.
Overall, the constraints outlined above produced results that are not precise energy savings calculations. However, very clear patterns were observed that align with testing performed by the manufacturer in laboratory test conditions.

The following can be stated with confidence:

- It has been demonstrated that the full submersion technology can be safely installed in a medium-sized data center and that the liquid-cooled servers operate reliably to meet processing load.
- The liquid-cooled servers do not add any noticeable heat to the surrounding ambient air, which suggests that the system could be used without the need for air cooling equipment at the site. This dramatically improves the cost/benefit potential for the technology.
- There is much greater energy efficiency opportunity as the server process load increases. The more the servers are used and the harder they are driven, the more beneficial liquid cooling becomes in terms of energy conservation potential.

At lower confidence, the following findings were also observed:

- In an ideal case where servers are running at very high loads, up to 75% of the required cooling energy (10% of total energy consumed) can be conserved using full submersion liquid cooling.
- In more typical operational conditions, approximately 38% of cooling energy (4.6% total) was conserved.
- By optimizing the system design (specifically balancing the cooling distribution unit capacity with the server cooling load) and by choosing optimal applications, it is possible to achieve savings greater than the typical case, but the ideal case is unlikely to be seen in the field.
- Failing to consider the server application and system design from an energy perspective could result in a liquid-cooled installation that does not deliver any energy savings. A worst-case scenario was tested that showed the possibility of -1% cooling (-0.1% total energy) conserved.
- At current prices, installing liquid-cooled servers in existing, air-cooled data centers likely is not cost-effective. Prices are likely to fall in the future, making the technology more likely to become cost-effective eventually.
- New construction data centers that can eliminate the cost of upfront air cooling equipment by installing liquid-cooled servers from the onset are possibly cost-effective opportunities for deployment at current prices. This should be verified on a case-by-case basis.
- Sites where recovering waste heat is possible can benefit further from the technology and are more likely to be cost-effective opportunities at current prices.
- The technology is undergoing additional testing to prove out non-energy benefits such as improved server lifetime, lower maintenance costs, and methods of rewriting software to optimize equipment performance enabled by liquid cooling.

For more information, register for the upcoming webinar on December 19, 2017 at 11:00 a.m. A final report detailing the results is also forthcoming; check the CARD project page on the Department’s website. If you have further questions, contact project manager Mark Garofano or CARD program administrator Mary Sue Lobenstein.

Reference