EXECUTIVE SUMMARY

Assessment of Data Center Liquid Cool for Energy Savings

ESTCP Project EW-201332

NOVEMBER 2019

Steve Branton
Otto VanGeet
James Elsworth
Asetek, Inc.

Distribution Statement A
This document has been cleared for public release

ESTCP

EXECUTIVE SUMMARY

Project: EW-201332

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

ACRONYMS AND ABBREVIATIONS

ARL	Army Research Laboratory
BLCCA	bridge life-cycle cost analysis
BTU/year	British thermal unit per year
CFM	cubic feet per minute
CIO	Chief Information Officer
COP	coefficient of performance
CPU	central processing unit
CRAC	computer room air conditioning
CRAH	computer room air handler
D2C	Direct-to-Chip liquid cooling technology
DoD	Department of Defense
ΔT	delta-Temperature
ERE	energy reuse effectiveness
ESTCP	Environmental Security Technology Certification Program
GHG	greenhouse gas
GPM	gallons per minute
GPU	graphics processing unit
HPC	high performance computing
HX	heat exchange
IT	Information Technology
kW	kilowatt
kWh	kilowatt-hour
l/s	liters per second
NIST	National Institute of Standards and Technology
PO	performance objectiveness
PUE	power usage effectiveness
SOW	statement of work
TCO	total cost of ownership

ACKNOWLEDGMENTS

This work was performed by the Asetek and supported by ESTCP, contract number EW-201332 Thanks to key partner Army Research Laboratory (ARL), and Redstone Arsenal for the High-Performance Computing clusters used in the study, for the facilities that housed and cooled them, and for the metering and data acquisition used for gathering the data. Specific appreciation to James Elsworth, and Otto VanGeet at the National Renewable Energy Laboratory for the data analysis and providing narrative.

1.0 INTRODUCTION

The purpose of report is to outline the technical outcomes of the ESCTP EW-201332 project executed by Asetek (hereafter "project team" or "product provider") at Redstone (Phase I), and ARL (Phase II and III). This report summarizes findings and assessment of direct-to-chip liquid cooling for the purpose of reducing total energy and peak energy consumption for cooling data center equipment. The product used for the test was the RackCDU (hereafter referred as Direct-to-Chip liquid cooling technology) unit.

Data centers are the most energy-intensive Department of Defense (DoD) buildings. They consume more than 10% of all DoD electricity[i] (40% for cooling) and produce 7.5 tril. British thermal units per year (BTUs/year) of unused heat. Direct-to-chip liquid cooling is a unique data center efficiency technology that brings high-performance liquid-cooling directly to the hottest elements inside each server ("hot-spot cooling"), with the potential to cut cooling energy by 60-80%, and to allow for reuse of the heat as on-site energy. It can also enable 2.5x data center consolidation with no additional infrastructure costs. In the right data center described below, the specific innovation design used by product provider can be retrofitted into existing servers and data centers with a payback of <1 year, enabling rapid adoption across all DoD installations, saving the DoD $200M per year in energy costs and cut greenhouse gas (GHG) emissions by 5 million tons per year.

The purpose of the demonstration is to document performance of the equipment about energy savings, reliability, and life cycle cost that can be achieved in the real-world environment of a DoD data center. The data and insights gained in the demonstration has been used to create awareness and acceptance of the technology to facilitate future technology transfer across all DoD data centers. The necessary data and insight would be gained by pursuing these objectives:

1. Retrofit existing equipment without disrupting operations;
2. Document energy savings and peak-load energy reductions;
3. Document waste-heat recovery opportunities and viability;
4. Document potential for increasing server density within existing foot-print and cooling infrastructure;
5. Document total system reliability relative to pre-retrofit levels;
6. Document GHG- savings;
7. Document return on investment potential for the existing DoD data center stock;
8. Address perceived barriers to broad adoption of the technology in new data centers.

2.0 OBJECTIVES

The following performance objectives were established for the project:

Performance Objective	Metric	Data Requirements	Success Criteria
1. Cooling Energy Usage	Average Cooling Energy Intensity (kilowatt-hour $[kWh]_{cooling}/kWh_{server}$)	Cooling Sub-Meter Readings, Direct-to-Chip liquid cooling technology Sub-Meter Readings (pumps and dry cooler load) and Data Load (% of server processing capacity in-use).	60-80% reduction in net annual cooling energy consumption
2. Server Energy Usage	Average server Energy Intensity ($kWh_{Direct\text{-}to\text{-}Chip\ liquid\ cooling\ technology}/kWh_{air\text{-}cooled}$)	Server Sub-Meter Readings and Data Load	5-10% reduced server annual energy consumption
3. Power Usage Effectiveness (PUE) and Energy Reuse Effectiveness (ERE)	Calculated PUE and ERE (unit-less): See figure 3, below, for details	Data from PO's 1 and 2, plus sub-meter readings in location of waste-heat reuse.	PUE reduction from 2 to 1.5; ERE reduction of 10% from PUE
4. Data Center Peak Load	Peak Power Ratio ($kW_{Direct\text{-}to\text{-}Chip\ liquid\ cooling\ technology}/kW_{air\text{-}cooled}$)	Sub-Meter Readings (cooling + IT+ Balance-of-System) and Data Load.	20-30% reduction in Peak Load
5. Server Up-Time	Percentage of server capacity available to do work (% availability)	Service logs and up-time meter readings	At or above the pre-retrofit levels for the site
6. Capacity Consolidation	Total Processing Intensity (flops/sqft and $flops/BTU_{of\ air\text{-}cooling}$)	Data from performance objective (PO) 1 and Square Footage Measurements, Nameplate Processing Capacity	2-3x increase in processing intensity
7. Ease of Use	Degree of perceived usability/ complexity of operations	Likert-type survey performed before and after Direct-to-Chip liquid cooling technology installation	No statistical change in usability/ serviceability
8. Lifecycle GHG Emissions	Total lifecycle GHG emissions (metric tons)	Standard ISO 14044-compliant lifecycle modeling based on demonstration data	20-30% reduction in lifecycle GHG emissions compared to air cooling
9. Lifecycle System Economics	Dollars spent	Calculations of projected lifetime energy cost-savings, and capital/maintenance savings	Simple Payback < 1 year; Lifecycle Radius of Influence (ROI) > 1000%
10. End-User Acceptance	Degree of acceptance at Host site and elsewhere in DoD	Responses to 5-point Likert-type surveys	Greater than 70% acceptance of Direct-to-Chip liquid cooling technology

3.0 TECHNOLOGY DESCRIPTION

Direct-to-Chip liquid cooling gathers heat from the hottest components in a server and removes it from the data center in an all-liquid path. This heat load bypasses the air conditioning system that normally cools the data center and radiates into the outdoors more efficiently via this all-liquid path. The hottest components to a server are its processors (central processing units [CPUs] and graphics processing units [GPUs]) followed by memory modules.

Direct-to-Chip liquid cooling technology (D2C) products are built from a series of standard building blocks. These building blocks include processor coolers (pump and cold plate units) for various processors, tubing, quick connectors, the Direct-to-Chip liquid cooling technology unit and their frames. Tubing connects the other building blocks and routes the coolant through the system. When an existing air-cooled server is adapted for liquid cooling, the air heat sinks are replaced with CPU liquid coolers (pump and cold plate units) as shown in Figure 3 and tubes are run from the CPU liquid coolers to quick connects at the rear of the chassis (not visible). The CPU liquid coolers use the same mounting points and hardware as the air heat sinks making them drop-in replacements for the heat sinks. Quick connectors are typically mounted in an unused PCIe slot.

The need for liquid cooling is being driven by several factors, including the increased demand for greater power density, coupled with higher IT performance for HPC and some hyper-scale computing, and the overall industry focus on energy efficiency.

3.1 INCREASED POWER DENSITY

Although the power density of air-cooled IT hardware has risen continuously with each new generation of equipment and is approaching an asymptote. This asymptote is due to the inherent thermal transfer limitation of air, that makes it such that significant volumes of air are required to absorb and transfer the heat away from the highest heat–producing components—such as the CPU, GPU, and memory—as well as from the internal power supplies.

3.2 INCREASED RACK POWER DENSITY

In theory, air-cooled IT equipment in standard industry racks has no formal power density limit. However, it becomes increasingly difficult (and fan energy intensive) to cool racks much beyond 20 kW per rack using conventional air-cooling methods. In contrast, liquid cooling systems can easily cool 20 kW per rack. There are multiple systems currently available that can cool 100 kW per rack, and some that can accommodate 200 kW or more per rack.

3.3 ENERGY EFFICIENCY

It takes 158 cubic feet per minute (CFM) or 75 liters per second (l/s) of air to cool a 1 kW air-cooled server with a Delta-Temperature (ΔT) of 11°C (20°F). Conversely, when using a liquid such as water, cooling a 1 kW server only requires approximately 0.34 gallons per minute (GPM) per kW at a ΔT of 11°C. Due to the higher thermal transfer characteristics, most liquid cooled IT can produce a ΔT of 11°C or even higher. (See Figure 2.) However, using a heat exchanger (HX) allows for different flow rates and ΔTs on each side. Many liquid cooling systems have a dedicated CDU to allow compatibility and easy integration with existing chilled water systems. In many cases, the CDU can also use condenser water, instead of chilled water, to avoid adding load to the chiller, thus saving chiller system capacity and energy.

3.4 LIMITATIONS OF LIQUID COOLING

Despite all the technical advantages of liquid cooling, "hydrophobia" continues despite the fact that water is already commonly used in many data centers, typically in conjunction with CRAHs. In some cases, liquid cooling manufacturers use pumped refrigerant or other fluids instead of water both to address this fear and for their efficient thermal transfer characteristics.

Nonetheless, one of the major reasons that airflow-based cooling remains the predominant choice—in spite of the heat density limitations and energy efficiency issues—is its relative ease of installation and removal of IT equipment. This is because for many years, air cooling was the de facto standard, which makes it even more difficult to change. Several liquid cooling manufacturers, e.g., Asetek Inc., have addressed this by making the IT hardware easily removable and installed. They do this by utilizing dripless quick-connect fluid couplers to allow for servicing and system upgrades.

4.0 PERFORMANCE ASSESSMENT

- **Performance Objective 1**: The average percentage of the energy into these five racks of IBM servers that was removed by the Direct-to-Chip liquid cooling technology system was 62.1%
- **Performance Objective 2**: We don't have this data. We can assume no server power savings from the retrofit. Normally would expect power savings at server level of 5 to 10% as indicated in Phase 1 result of the demonstration.
- **Performance Objective 3**: There is no ERE calculation as this site is not engaging in any energy reuse. Some implementations (such as at NREL's Energy Systems Integration Facility high computing center) reuse the warm return liquid for low-quality heat applications.
- **Performance Objective 4**: Based on the max CRAC power draw post-retrofit and COP calculation, we see a peak load reduction of 7.8% in total room load.
- **Performance Objective 5**: This was not measured during this project due to insufficient data.
- **Performance Objective 6**: Due to the change in site, and hence the change in leadership of the data centers, consolation was not studied, or actualized.
- **Performance Objective 7:** While the retrofitting process can be challenging, the project did prove itself worthwhile and energy efficient. The process of having the factory complete the retrofits proved to be easily installed without any complications.
- **Performance Objective 8:** Due to electricity reduction of 1424 kWh/day, or 520,116 kWh/year. equates to 387,075 kgCO_2/year, or 427 tons. Over a 20-year lifetime, this amounts to 8540 tons of CO_2 emissions avoided. This is substantially lower than the estimate outlined in the SOW, because a much smaller server capacity was retrofitted at the new site than was originally planned.
- **Performance Objective 9:** Due to the various changes and complications during project execution, a lifecycle cost analysis was not performed for this project. See the cost assessment for more details.
- **Performance Objective 10:** The project team did not perform a Likert survey as it has been initially planned, due to the logistics around project execution. Instead, the project team had email, and verbal confirmation from host sites that once installed, user satisfaction is high.

5.0 COST ASSESSMENT

Before the project we performed TCO analysis, however we did not perform a cost assessment after project completion. Below is the result of the calculations from pre-project analysis.

Economic Benefit of this Project: Based on preliminary modeling for this project, using the NIST BLCCA process, we find a simple payback of less than 9 months, a lifetime cost-savings of more than $8 million and a savings-to-investment ratio of almost 12. Other BLCCA outputs can be found in Table 1.

Table 1. Investment and Savings for Direct-to-Chip Liquid Cooling Technology vs. Air-cooling for the Proposed ESTCP Demonstration.

RackCDU vs Air Cooling (Current Project)	Years			
	1	5	10	20
Initial Investment (RackCDU) ($)	**$423,493**	**$423,493**	**$509,618**	**$681,868**
Other Capex Savings ($)	$246,702	$246,702	$246,702	$246,702
Energy Savings ($)	$233,041	$1,237,246	$2,671,553	$6,261,896
Maintenance Savings ($)	$58,260	$309,311	$667,888	$1,565,474
Total Savings ($)	**$538,004**	**$1,793,260**	**$3,586,143**	**$8,074,073**
Savings/Investment	1.27	4.23	7.04	11.84

This represents an extraordinary cost- and energy-savings compared to the current state-of-the-art technology. These benefits are significant, reasonable, and consistent with the expectations, preliminary models, and preliminary experimental results from our small-scale demonstration project in San Jose, CA. The cost of this project is commensurate with the relatively low-risk and potentially extraordinary payoff for DoD if this technology is broadly adopted.

Note that the economic assessment above assumes that the server vendors agree to maintain their existing equipment warranty (except in instances of water leakage, which will be covered by the product provider). If, however, we find it necessary to pay for a 3rd-party warranty to cover the servers that are retrofit, the economic implications will change. Based on our discussions with STG, a leading 3rd-party warranty supplier that will cover liquid-cooling retrofits, a modified BLCCA is provided below, in Table 2. While we do not believe this will be necessary, even if it is, this is still a compelling rate of return. All other retrofit costs have been included in the cost/benefit analysis.

Table 2. Investment and Savings for Direct-to-Chip Liquid Cooling Technology vs. Air-cooling for the Proposed ESTCP Demonstration, Including Cost of 3rd-Party Warranty Coverage, if Necessary.

RackCDU vs Air Cooling (Current Project)	Years			
	1	5	10	20
Initial Investment (RackCDU) ($)	**$720,293**	**$720,293**	**$1,103,218**	**$1,869,068**
Other Capex Savings ($)	$246,702	$246,702	$246,702	$246,702
Energy Savings ($)	$233,041	$1,237,246	$2,671,553	$6,261,896
Maintenance Savings ($)	$58,260	$309,311	$667,888	$1,565,474
Total Savings ($)	**$538,004**	**$1,793,260**	**$3,586,143**	**$8,074,073**
Savings/Investment	0.75	2.49	3.25	4.32

Assumptions for costing purposes:

1. Servers will be "refreshed" (replaced with newer models) every 5 years
2. Future server refreshes will be pre-installed with new internal loops; all other Direct-to-Chip liquid cooling technology infrastructure will be reused.
3. Total data center remaining operable lifetime is 20 years
4. The site will expand server capacity according to plan, from 250kW to 600kW by the end of this demonstration project, realizing the full capital avoidance value enabled by Rack CDU.
5. Price of electricity and natural gas for NEDC is $0.079/kWh and 8.75/MMBTU, respectively.
6. 3% year-over-year increase in energy costs[ii].
7. For payback and ROI calculations, we have included only those costs that would be incurred for a typical Direct-to-Chip liquid cooling technology installation (i.e. excluding costs related only to running the demonstration).
8. Payback and ROI calculations begin at the time that Stage 3 installation starts.

6.0 IMPLEMENTATION ISSUES

At the launch of this project, the initial site of demonstration, Redstone, made a firm, written commitment to provide all the necessary servers for this project. Unfortunately, while our Phase 1 technical program has been extremely successful at Redstone, Phases 2 and 3 was in a stand-still for more than 18 months, due to factors outside the control of the product provider, Redstone or ESTCP. In particular, as part of the Army's data center consolidation initiative, Redstone Arsenal's data center was selected to be decommissioned.

As a result of Redstone's status, the servers that were committed to this project for Phases 2 and 3 were never made available to the project. For phase 2, Redstone was only able to provide a half of a rack of servers (4.5 racks short of the requirement); and these were servers that had been decommissioned long ago, and which were completely irrelevant for this project. Redstone was not able to provide any servers for phase 3. We worked extensively with the Redstone team, the Army's data center consolidation team and the Army CIO's office to try to address the situation, but we were ultimately unsuccessful.

In collaboration with the Army CIO's office and data center consolidation team, we met with dozens of new host sites, and were ultimately able to identify a high-quality host-site that will NOT be consolidated. This is the data center at the Army Research Labs, located at Aberdeen Proving Ground, where this program now being completed.

Making the move for this project to ARL required additional time and funding to complete the project. Additional funding is required to:

1. Recreate our data center models to meet the parameters of ARL's data center (Task 3)
2. Design and install a new facilities liquid loop at ARL (Task 4)
3. Relocate Phase 1 servers from Redstone to ARL (Task 5).

7.0 REFERENCES

1. "Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431", U.S. Environmental Protection Agency, ENERGY STAR Program (August 2, 2007)
2. Green Grid: "Liquid Cooling Technology Update", 2017.
3. National Institute of Standards and Technologies, energy forecast

---

[i] "Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431", U.S.

[ii] National Institute of Standards and Technologies, energy forecast