FINAL REPORT

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: 11/20/2019

2. REPORT TYPE: ESTCP Final Report

3. DATES COVERED *(From - To)*: 9/21/2013 - 5/31/2019

4. TITLE AND SUBTITLE
Assessment of Data Center Liquid Cool for Energy Savings

5a. CONTRACT NUMBER: 13-C-0038

5b. GRANT NUMBER: N/A

5c. PROGRAM ELEMENT NUMBER:

5d. PROJECT NUMBER: EW-201332

5e. TASK NUMBER:

5f. WORK UNIT NUMBER:

6. AUTHOR(S)
Steve Branton
Otto VanGeet
James Elsworth

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Asetek Inc.
5285 Hellyer Ave #110
San Jose, CA
95138

8. PERFORMING ORGANIZATION REPORT NUMBER
EW-201332

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Environmental Security Technology Certification Program
4800 Mark Center Drive, Suite 16F16
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
ESTCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
EW-201332

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Data centers are the most energy-intensive DoD buildings. They consume more than 10% of all DoD electricity (40% for cooling) and produce 7.5 tril. 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 reused of the heat as on-site energy. It can also enable 2.5x data center consolidation with no additional infrastructure costs. 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. 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.

15. SUBJECT TERMS
Assessment, Data Center Liquid Cool, Energy Savings

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Not classifiec	Not classifiec	Not classifiec	Unlcass	57	Steve Branton
					19b. TELEPHONE NUMBER *(Include area code)*: (408) 823-4486

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

FINAL REPORT
Project: EW-201332

TABLE OF CONTENTS

Page

ABSTRACT IX
EXECUTIVE SUMMARY ES-1
1.0 INTRODUCTION 1
1.1 OBJECTIVE OF THE DEMONSTRATION 1
2.0 TECHNOLOGY DESCRIPTION 3
2.1 TECHNOLOGY OVERVIEW 3
2.2 TECHNICAL DETAILS OF THE PRODUCTS 3
2.3 TECHNOLOGY (PRODUCT) DEVELOPMENT 4
2.4 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY 4
2.4.1 Increased Power Density 5
2.4.2 Greater IT Equipment Performance 5
2.4.3 Increased Rack Power Density 6
2.4.4 Energy Efficiency 6
2.4.5 Limitations of Liquid Cooling 7
3.0 PERFORMANCE OBJECTIVES 8
4.0 FACILITY/SITE DESCRIPTION 9
4.1 FACILITY/SITE LOCATION AND OPERATIONS 9
4.2 FACILITY/SITE CONDITIONS 12
4.3 UPDATED FACILITY/SITE DESCRIPTION [PHASE 2 AND PHASE 3] 13
5.0 TEST DESIGN 15
5.1 CONCEPTUAL TEST DESIGN 15
5.2 BASELINE CHARACTERIZATION 17
5.3 DESIGN AND LAYOUT OF TECHNOLOGY COMPONENTS 17
5.4 SAMPLING RESULTS 19
5.4.1 Lessons Learned 21
6.0 PERFORMANCE ASSESSMENT 22
6.1 PERFORMANCE OBJECTIVE 1: COOLING ENERGY USAGE 22
6.1.1 Server Energy Used 22
6.1.2 Energy Removed 23
6.1.3 Efficiency 24

TABLE OF CONTENTS (Continued)

Page

6.1.4 Other Servers 25
6.1.5 Other Data 26
6.1.6 Chiller Plant Data 27
6.1.7 CRAC Air Handling Units 28
6.1.8 Overall Calculations 29
6.2 PERFORMANCE OBJECTIVE 2: SERVER ENERGY USAGE 30
6.3 PERFORMANCE OBJECTIVE 3: POWER USAGE EFFECTIVENESS (PUE) AND ENERGY REUSE EFFECTIVENESS (ERE) 30
6.4 PERFORMANCE OBJECTIVE 4: DATA CENTER PEAK LOAD 31
6.5 PERFORMANCE OBJECTIVE 5: SERVER UPTIME 31
6.6 PERFORMANCE OBJECTIVE 6: CAPACITY CONSOLIDATION 31
6.7 PERFORMANCE OBJECTIVE 7: EASE OF USE 31
6.8 PERFORMANCE OBJECTIVE 8: LIFECYCLE GHG EMISSIONS 32
6.9 PERFORMANCE OBJECTIVE 9: LIFECYCLE COST ANALYSIS 32
6.10 PERFORMANCE OBJECTIVE 10: END-USER ACCEPTANCE 32
7.0 COST ASSESSMENT 33
8.0 IMPLEMENTATION ISSUES 35
9.0 REFERENCES 36
APPENDIX A POINTS OF CONTACT A-1
APPENDIX B SUPPLEMENTAL DATA B-1

LIST OF FIGURES

Page

Figure 1. Transforming an Air-cooled Server to a Liquid-cooled Server 3
Figure 2. Direct-to-Chip Liquid Cooling Technology Types and Details 4
Figure 3. Comparison of Heat transfer air versus liquid volumes at 11°C (20°F) ΔT 6
Figure 4. Sparkman Complex Showing Location of Data Center and Cooling Towers 10
Figure 5. Data Center Floor at Redstone 11
Figure 6. Facilities Warm Water Loop Plumging One-line Schematic 11
Figure 7. Facilities Liquid Loop Equipment Sled 12
Figure 8. ARL 'White Cell' Data Hall Floor Plan 14
Figure 9. DL560 Server Before and After Liquid Cooing Retrofit 18
Figure 10. Rear View of Phase 1 Rack Installed at Redstone 19
Figure 11. Cooling Energy Savings Demonstrated in Phase 1 20
Figure 12. Server Energy Savings Demonstrated in Phase 1 20
Figure 13. Server Load into the 5 IBM Servers Post-retrofit 22
Figure 14. Energy removed by Direct-to-Chip Liquid Cooling Technology from 5 Racks of IBM 'Fob' Servers - Individual and Total 23
Figure 15. Energy Removed by Direct-to-Chip Liquid Cooling Technology System for the 5 Racks of IBM Servers 24
Figure 16. Percentage of Input Energy Removed by the Direct-to-Chip Liquid Cooling Technology for the 5 IBM Servers 24
Figure 17. Energy Removed by the Direct-to-Chip Liquid Cooling Technology System Across All 13 Retrofitted Server Racks 25
Figure 18. Energy Removed by the Direct-to-Chip Liquid Cooling Technology System Before and After the Retrofitting of the First 4 ervers. 26
Figure 19. Data Available for the Chiller Energy Use for the Whole Plant 27
Figure 20. COP of the Chiller Plant in Economizer Mode and with Chiller Running 28
Figure 21. Energy Use by the Air Handling Unit Cooling System 29
Figure 22. Energy Removed from Each Individual Server by the Direct-to-Chip Liquid Cooling Technology Systems B-1
Figure 23. Energy Removed by the Direct-to-Chip Liquid Cooling Technology Systems Installed on the 8 IBM and 4 Appro Racks of Servers Only. B-2
Figure 24. Total Energy Removed by the Direct-to-Chip Liquid Cooling Technology System over Another Time Period B-2
Figure 25. Meter Readings Showing an Unexplained Decrease in the Rate of Energy Removal by the Direct-to-Chip Liquid Cooling Technology System Over Several Days in late October 2017. .... B-3
Figure 26. Energy Removed from Select Servers for Which Energy Input Data Was Not Available B-4
Figure 27. Energy Removed from the 4 Appro Servers (based on data available) B-4

LIST OF TABLES

Page

Table 1. CPU performance and Power Dissipation Ranges 5
Table 2. Example of a Generic Liquid-cooled System Added to an Existing Condenser Water Loop (no chiller load) 7
Table 3. Data for white cell load from 3 select days in November 2018. 26
Table 4. Summary of Calculated Load Data for the Various Cooling System Components. 29
Table 5. Investment and Savings for Direct-to-Chip Liquid Cooling Technology vs. Air-cooling for the Proposed ESTCP Demonstration. 33
Table 5a. 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. 34

ACRONYMS AND ABBREVIATIONS

ARL	Army Research Laboratory Data Center
BLCCA	bridge life-cycle cost analysis
BTU/year	British thermal unit per year
°C	Celsius
C_p	Specific heat
CDU	cooling distribution unit
CFM	cubic feet per minute
CIO	Chief Information Officer's
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
DCIM	Data center infrastructure management
DCM	Data center management
DIMM	dual in-line memory module
DoD	Department of Defense
ΔT	delta-Temperature
ERE	energy reuse effectiveness
ESCTP	Environmental Security Technology Certification Program
°F	Fahrenheit
GHG	greenhouse gas
GPM	gallons per minute
GPU	graphics processing unit
HPC	high performance computing
HX	heat exchanger
ISO	International Organization for Standardization
IT	Information technology
$kgCO_2$/year	kilogram of carbon dioxide per year
kW	kilowatt
kWh	kilowatt-hour
l/s	liters per second
M	mass flow rate

MMBTU	1 Million British Thermal Unit
NEC-R	US Army Network Enterprise Center - Redstone
NIST	National Institute of Standards and Technology
NREL	National Renewable Energy Laboratory
OEM	original equipment manufacturer
PCIe	peripheral component interconnect express
PO	performance objective
PUE	power usage effectiveness
Q	energy flow rate
ROI	Radius of Influence
SNMP	Simple Network Management Protocol
SOW	statement of work
STG	Signature Technology Group
TCO	total cost of ownership
V	voltage
W	watt

ACKNOWLEDGEMENTS

This work was performed by the Asetek and supported by ESTCP, contract number EW-201332 Thanks to key partner Army Research Laboratory, 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.

ABSTRACT

Data centers are the most energy-intensive DoD buildings. They consume more than 10% of all DoD electricity[i] (40% for cooling) and produce 7.5 tril. 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 reused of the heat as on-site energy. It can also enable 2.5x data center consolidation with no additional infrastructure costs. 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.

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. 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 is rejects into the outdoors more efficiently via this all liquid path. The hottest components are a server are its processors (CPUs and GPGPUs) followed by memory modules.

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.

EXECUTIVE SUMMARY

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 Army Research Laboratory (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[ii] (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 reused 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.

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 is rejects into the outdoors more efficiently via this all-liquid path. The hottest components are 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-cold plate units) as shown in **Error! Reference source not found.** and tubes 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 peripheral component interconnect express (PCIe) slot.

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

Increased Power Density

Although the power density of air-cooled IT hardware has risen continuously with each new generation of equipment 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.

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 kilowatt (kW) per rack using conventional air-cooling methods. In contrast, liquid cooling systems can easily cool 20 kW per rack, and there are multiple systems currently available that can cool 100 kW per rack, and some that can accommodate 200 kW or more per rack.

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-T (ΔT) of 11°Celsius (°C) (20°Fahrenheit [°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 cooling distribution unit (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.

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 computer room air conditioning (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.

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 performance objective (POs) 1 and 2, plus sub-meter readings in location of waste-heat re-use.	Power usage effectiveness (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 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 International Organization for Standardization (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

Results

- **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**: The project team don't have this data. The project team 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 National Renewable Energy Laboratory's (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 computer room air conditioning (CRAC) power draw post-retrofit and coefficient of performance (COP) calculation, the project team 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 kilogram of carbon dioxide (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 statement of work (SOW), because far 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.

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 DoD buildings. They consume more than 10% of all DoD electricity[iii] (40% for cooling) and produce 7.5 tril. 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 reused 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, 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 GHG emissions by 5 million tons per year.

Through this project the project team validated the energy savings of direct-to-chip liquid cooling by showing that, compared to state-of-the-art air-cooling, hot-spot cooling can provide 60-80% reduction in cooling energy. This demonstration provided the data, awareness and confidence needed to drive adoption of direct-to-chip liquid cooling throughout DoD and beyond.

1.1 OBJECTIVE OF THE DEMONSTRATION

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 original plan called for the demonstration to be conducted in a single data center. The project team would begin by installing a facilities level warm water liquid cooling loop in that data center and then installing new liquid cooled servers (provided as part of the project) to demonstrate the viability and reliability of the technology with no operational risk to the data center. Once user acceptance of this deployment was achieved, the project team would demonstrate our ability to retrofit a small number of the data centers existing servers to satisfy the operators that retrofitting is viable and reliable. Finally, the project team would conclude the demonstration with retrofitting a large number of servers to provide an at scale demonstration of the technology. 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 TECHNOLOGY DESCRIPTION

2.1 TECHNOLOGY OVERVIEW

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 is rejected into the outdoors more efficiently via this all liquid path. The hottest components are a server are its processors (CPUs and GPUs) followed by memory modules.

2.2 TECHNICAL DETAILS OF THE PRODUCTS

Direct-to-Chip liquid cooling technology 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-cold plate units) as shown Figure 1 in and tubes 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.

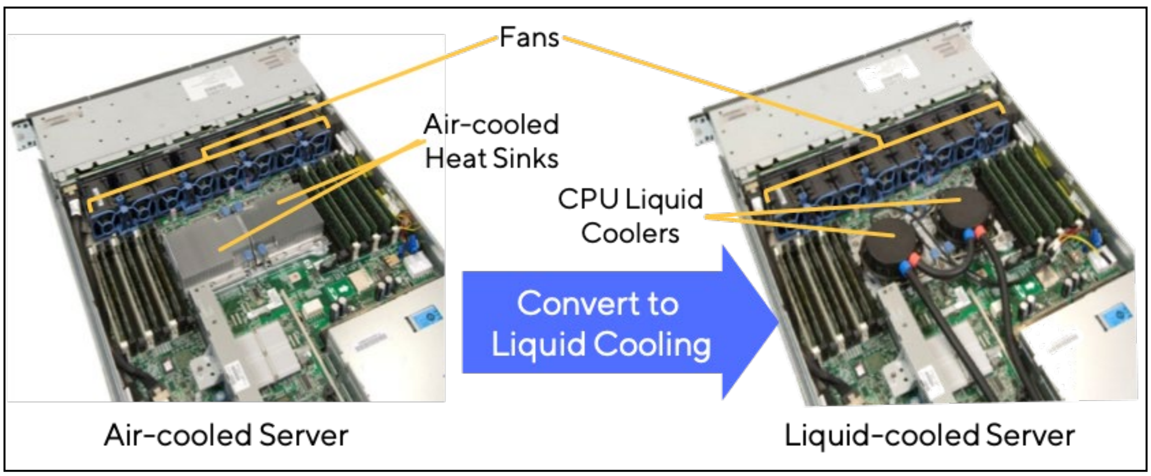

Figure 1. Transforming an Air-cooled Server to a Liquid-cooled Server

The Direct-to-Chip liquid cooling technology maybe fitted into an extension that mounts on the rear of the rack or the HEX cabinet maybe mounted in the top 4U of the rack and the manifolds mounted in the zero-U vertical space in the rear of the rack. The Direct-to-Chip liquid cooling technology manifolds connect to liquid cooled servers via tube sets—one hot and one cold tube—with quick connectors at each end. Direct-to-Chip liquid cooling technology is connected to facilities plumbing by means of flexible hoses. These facilities plumbing connection points may be above the racks or below a raised floor.

Both Direct-to-Chip liquid cooling technology types are passive with the exception that each contains a monitoring system that reports vital statistics to DCIM (data center infrastructure management) or DCM (data center management) software via Simple Network Management Protocol (SNMP).

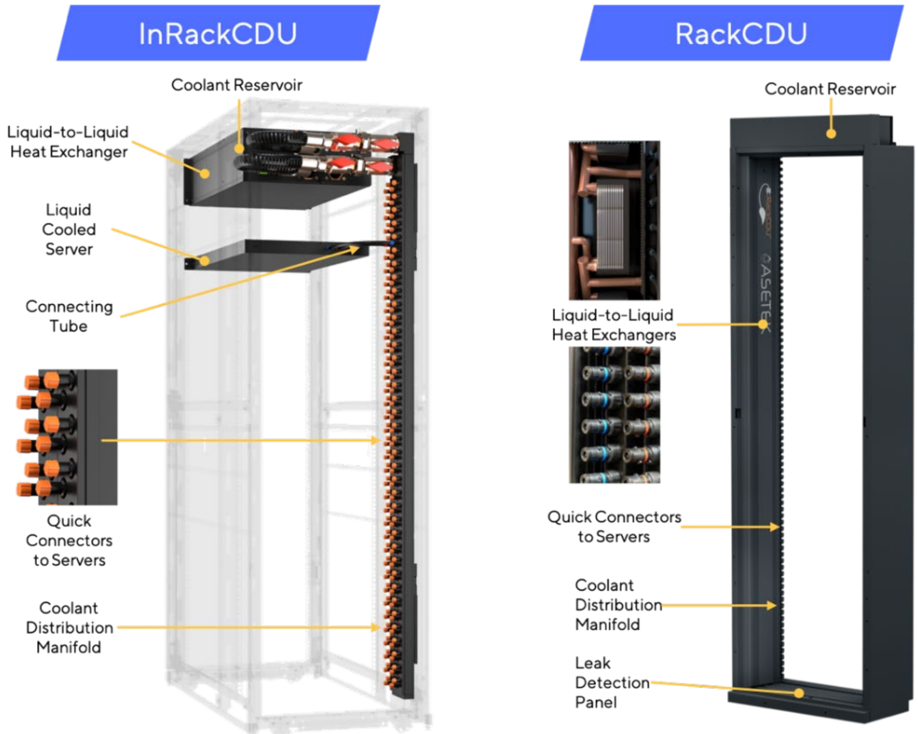

Figure 2. Direct-to-Chip Liquid Cooling Technology Types and Details

Facilities liquid cooling systems including plumbing fittings and appliances, circulating pumps, heat rejection devices (cooling towers and dry coolers) are sold and installed by third parties. DCIM and DCM software is also sold and installed by third parties.

2.3 TECHNOLOGY (PRODUCT) DEVELOPMENT

At the time the demonstration project began, Direct-to-Chip liquid cooling technology was a developed, pre-commercial product. Laboratory and proof-of-concept work was complete. No further technical development was required. Design work was limited to configuring the cooling loops for specific server models (largely adjusting tube lengths) and Direct-to-Chip liquid cooling technologies for different models of racks (matching Direct-to-Chip liquid cooling technology frames to rack frames).

2.4 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

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.

2.4.1 Increased Power Density

Although the power density of air-cooled IT hardware has risen continuously with each new generation of equipment 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.

Table 1. CPU performance and Power Dissipation Ranges

CPU Performance	Low-power processors Light workloads	General purpose Economy	High performance High efficiency	Maximum performance, scalability, memory, cache capacity	HPC applications Existing and future processors
Power Dissipation	6-8 watts	17-87 watts	60-130 watts	130-150 watts	150-450 watts

Due to the fundamental thermal transfer property of liquid cooling, it provides 4000X higher of heat transfer capacity for heat-generating components when compared with traditional air cooling. Liquid cooling allows more powerful processors and other components to be contained in close proximity within IT equipment, while providing greater heat removal capacity and using less space and energy. In addition to the more obvious benefits of substantially fewer racks and much less floor space, liquid cooling allows improved IT performance, as detailed in the next section.

Table 1 illustrates two trends appearing from generation to generation. The number of cores in the CPU and GPU processors is increasing; at the same time, so is processor power from 6 watts (W) to 450 W. During all these processor developments, one parameter is kept almost constant: the silicon die temperature operating limit, which is important from a reliability perspective. So, if the power goes up and so does the flux density of the dies, it becomes increasingly challenging to transfer the amount of heat necessary to maintain processor die temperatures.

2.4.2 Greater IT Equipment Performance

Higher component density provides multiple IT performance benefits. These are derived from lower latency, which is attributable both to shorter intersystem connectivity distances within the rack and to improved overall multisystem throughput due to shorter distances for rack-to-rack and rack-to-core connectivity.

In addition, as the CPU case temperature moves upward toward its maximum safe operating temperature, CPU performance needs to be throttled back to avoid thermal runaway. High performance computing in particular requires that CPU and memory temperatures remain low in order to operate at maximum speed. Liquid cooling allows densely packed systems to operate continuously at their maximum voltage and clock frequency while avoiding overheating.

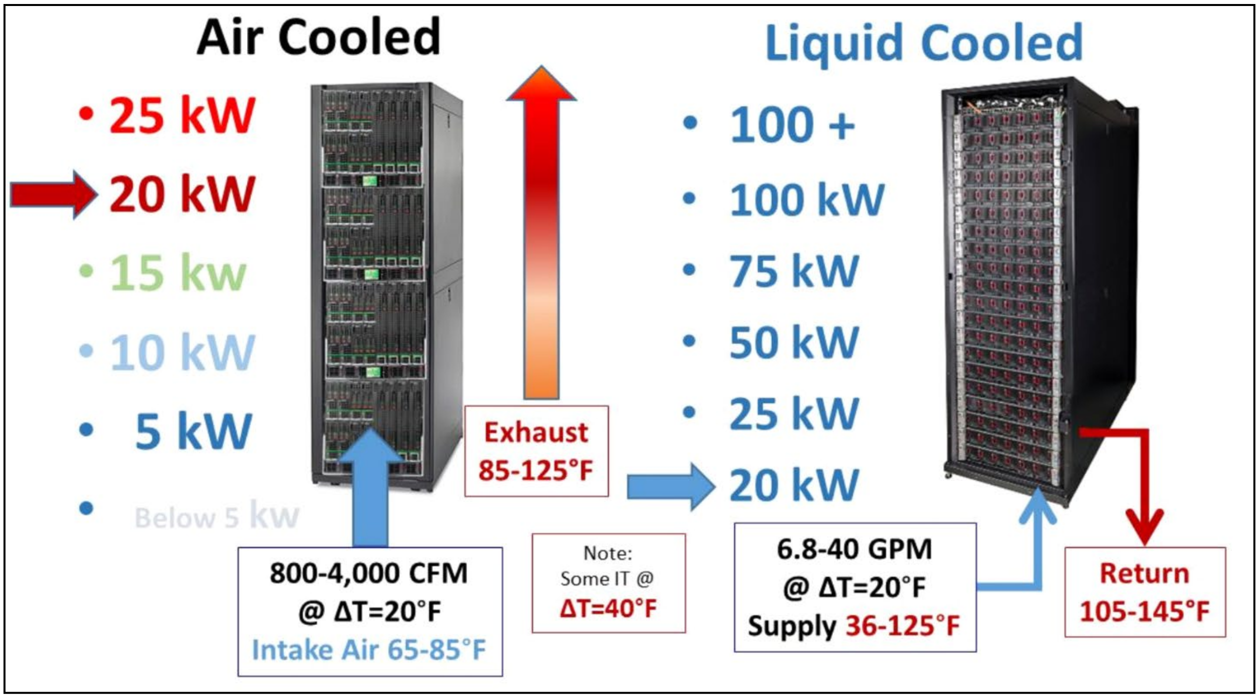

Figure 3. Comparison of Heat transfer air versus liquid volumes at 11°C (20°F) ΔT

2.4.3 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, and there are multiple systems currently available that can cool 100 kW per rack, and some that can accommodate 200 kW or more per rack.

2.4.4 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-T (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. (See Table 2.)

Table 2. Example of a Generic Liquid-cooled System Added to an Existing Condenser Water Loop (no chiller load)

Summarized		Facility CRAH Fans		IT Internal Fans	Fans and Pumps
Type	Heat load	Flow rate	Power	IT Air Intake @25 °C (77 °F)	Total (power)
Air-cooled (fan)	20 kW	3200 CFM (90.6 m^3/min)	2.4 kW	1.5 kW	3.9 kW
Liquid-cooled (pump)		6.8 GPM (2.57 l/min)	0.1 kW	0 kW	0.1 kW

2.4.5 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. In many implementations, this is built as a complete rack-scale system, as in the Direct-to-Chip liquid cooling technology. Overcoming hydrophobia in IT and facilities managers requires education and operating experience to prove the safety, reliability, and performance of liquid cooling systems. It is therefore important for data center designers, owners, and operators, as well as IT system architects, to discard fears based on myths and misperceptions about the risks, such as leaks, of using liquid-cooled systems. [ii]

3.0 PERFORMANCE 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 ($kWh_{cooling}/kWh_{server}$)	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 re-use.	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 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 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

4.0 FACILITY/SITE DESCRIPTION

The Army's 7th Signal Command, Network Enterprise Center, Redstone (NEC-R) was selected to host the demonstration and provide the servers being retrofit in phases 2 and 3 and handle delivering data gather by monitoring equipment to product provider and NREL.

The data center is located on the basement level of the Sparkman Center at Redstone Arsenal near Huntsville, Alabama. At the time, the data center had ~80,000 sq.ft. of 'white space' (floor space usable for deploying servers), had power to support ~350kW of IT load, and was home to 1000+ servers in 250 racks.

Primary drivers for selecting this site included:

1. Perceived high "need" – the site was a target Army consolidation site with rapid expansion plans.
2. Perceived high "value" – opportunity for significant capital avoidance in expanding the data centers cooling capacity
3. Perceived high "enthusiasm" from both IT staff and facilities team
4. High "visibility" – one of 7 data centers at Redstone Arsenal
5. Selected by the Army CIO

4.1 FACILITY/SITE LOCATION AND OPERATIONS

Phase 1 was conducted at the Network Enterprise Date Center, Redstone outside Huntsville, Alabama. The data center is located on the basement level of the Sparkman Center, a large office complex with an attached "central plant" consisting of a mechanical room and yard supporting cooling for the entire complex including the data center. The data center was cooled with CRAH units that received chilled water from Chillers located in the mechanical room. These chillers rejected their heat via cooling towers in the mechanical yard.

Discussion with the facilities team at Redstone determined that the best approach for implementing a facilities warm water liquid cooling loop is tapping into the condenser water loop. This loop moves water between the condensers of the central plant's chillers and cooling towers. The water is heated by the chiller condensers and is "free cooled" to near the ambient wet bulb air temperature by the cooling towers. The pumps and filters and other plumbing equipment associated with circulating the warm water to the data center is located in the mechanical room. A provision for mixing chilled water from the CRAH chilled water loop into the warm water loop was included in case on rare occasion the wet bulb temperature exceeds the temperature needed to keep the warm water loop below 104°F. In that case chilled water would provide the necessary additional cooling. The warm water loop terminated in the ceiling of the data center above the intended rack locations. This decision necessitated a piping run of approximately 1000 ft between the cooling towers and the data center.

Figure 4 shows the Sparkman complex with the location of the data center, cooling towers, chillers and piping run identified. Figure 5 shows a map of the data center floor with the planned location for installation of Direct-to-Chip liquid cooling technology equipped servers and the path back to the central plant identified. Figure 6 provides a plumbing schematic for the facilities warm water loop that was installed at Redstone. Figure 7 shows the equipment sled for the facilities loop including, from left, pairs of circulation pumps to data center, liquid-to-liquid heat exchangers, self-cleaning filters for tower water and tower circulation pumps.

This facility's liquid loop was built with sufficient capacity to support the full demonstration plus some capacity for expanding the installation of liquid cooled servers on project completion.

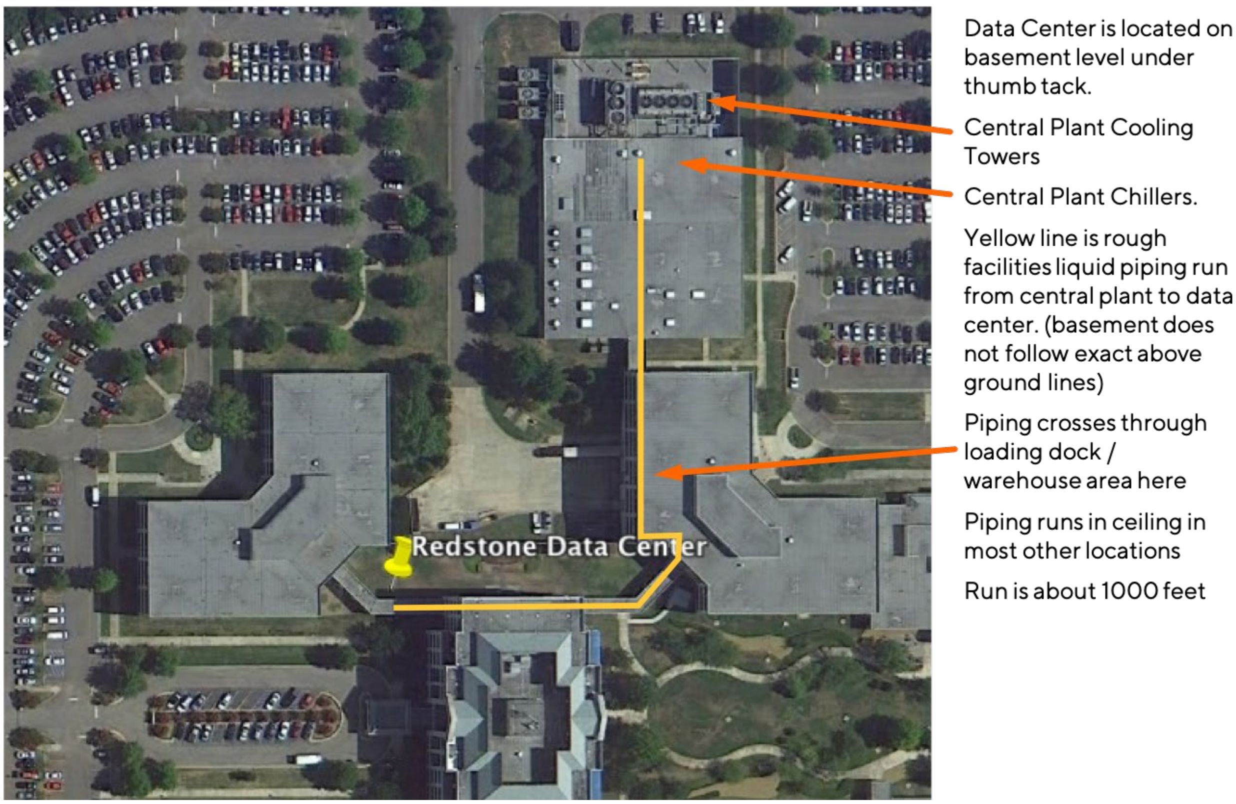

Figure 4. Sparkman Complex Showing Location of Data Center and Cooling Towers

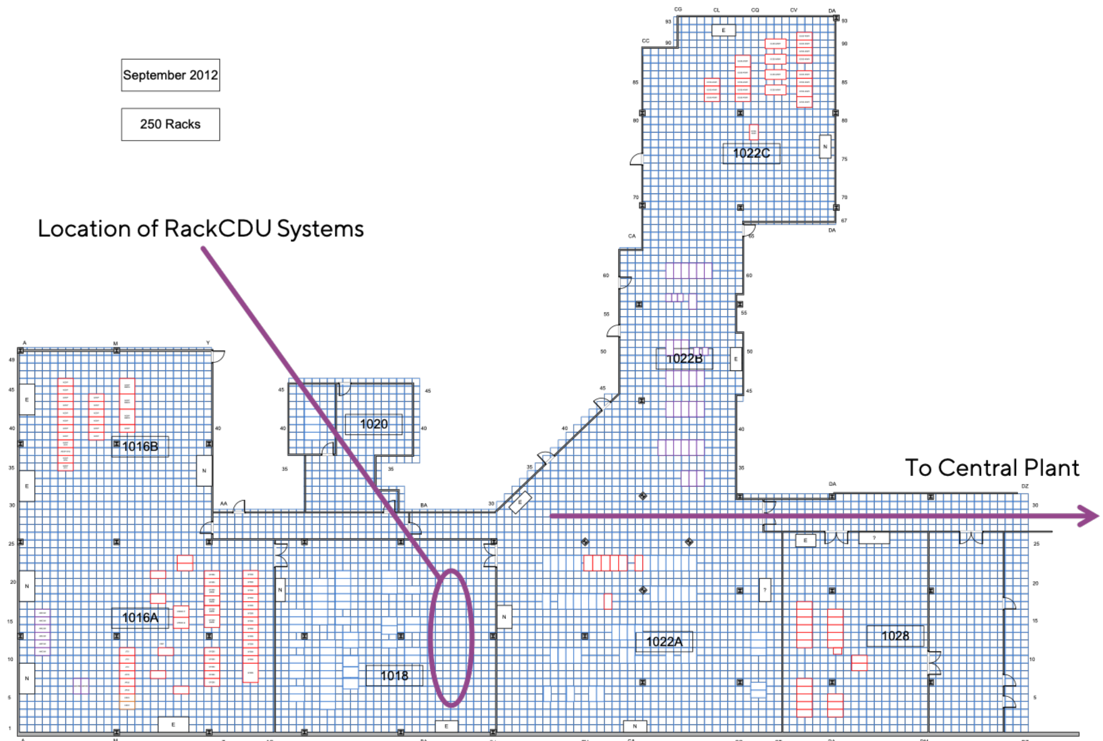

Figure 5. Data Center Floor at Redstone

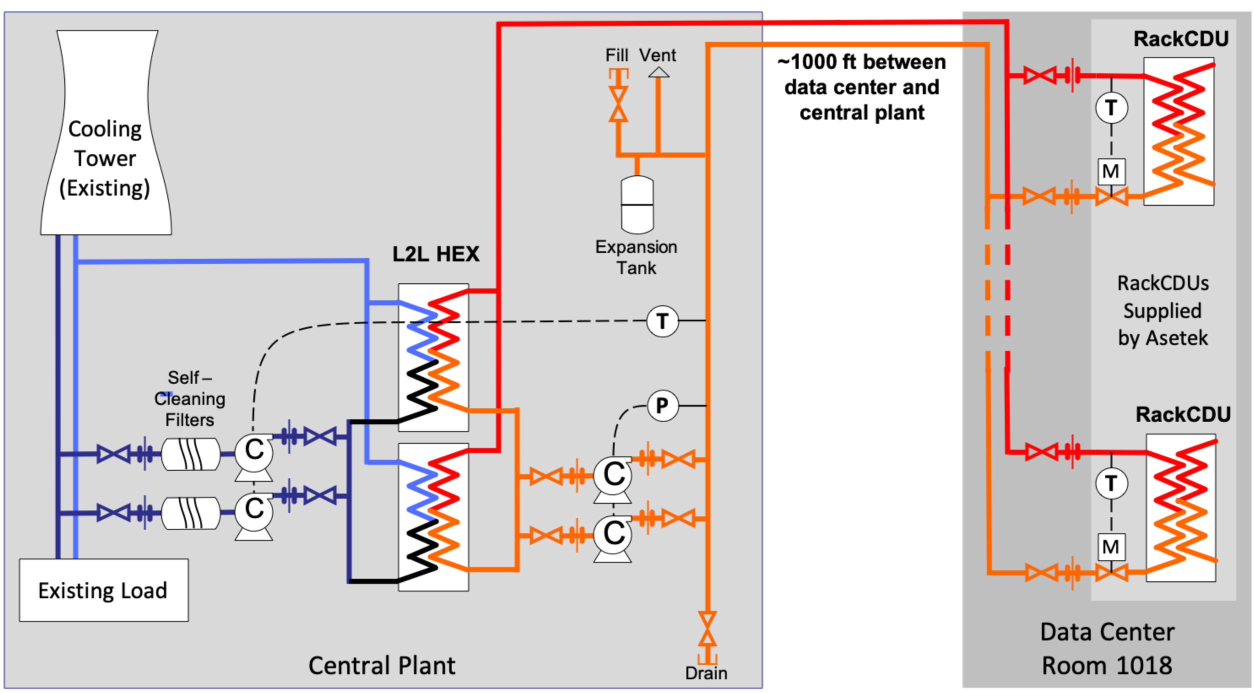

Figure 6. Facilities Warm Water Loop Plumging One-line Schematic

Figure 7. Facilities Liquid Loop Equipment Sled

This piping run contributed significantly to the cost of installing the warm water facilities loop. Another cost factor was planning for expansion.

4.2 FACILITY/SITE CONDITIONS

At the beginning of the program, Redstone was selected as a demonstration site information provided by the site indicated that up to 500 servers were potentially available for retrofit under the project. While there were nearly 500 servers at the site US Army Network Enterprise Center - Redstone (NEC-R) was basically running a co-location data center. In the colocation 'business' model the data center owner (NEC-R here) provides floor space, power, network communications, security and cooling to external groups. Those external groups ('clients') 'rent' space with these services in the data center wherein these clients operate their own servers. The ownership of the servers remains with the clients. The data center owner ('host') is not free to do anything with these servers without the client's permission. Thus, when the time came to identify servers that could be used in phases 2 and 3 of the project, the number of truly available servers was quite small. Considerable effort was required to get a list of the servers available for phase 2.

The list of server candidates for retrofit in phase 2 was provided in February 2015. It included a total of 91 individual servers. These servers ranged in age from models produced in 1998 to 2012. The majority, 66, were 2007 or older models 8 years or older at the time. The list includes 29 different models of servers. Most models were represented in single digit quantities. Two models contributed over 10 servers to the list 12 Dell PowerEdge 2650 from model year 2004 and 17 HP ProLiant DL380 G5 from model year. 2007. These servers used 12 different CPU socket types. The socket type is important because the socket design governs how coolers are mounted on the CPU.

The project team reviewed all the available servers. The project team determined that it was only possible to retrofit 53 of the 91 individuals. Of these, 3 were available in enough quantity to consider building coolers for them in the DK office. For the remainder the project team determined that the best approach would be to deliver kits of ingredients to Redstone and to configure cooling loops on site for the other units. This approach represents a significant increase in the labor to retrofit serves when compared to using mass produced liquid coolers.

As the logistics were being worked to move all of the needed parts to Redstone for the phase 2 retrofit, the project team learned the many of the servers on our list were no longer running production work loads and many of the others were slated for retirement. In fairness to the NEC-R team it should also be noted that in this period the Redstone data center was removed from the list of potential consolidation sites. This created significant uncertainty about the future for the site. None the less, retrofitting retired servers would not provide the demonstration with evidence of the performance of liquid cooling on servers in normal operation. As understanding how liquid cooling performs in a production environment was a key part of the demonstration, a painful decision was taken to change the demonstration site from Redstone to the Army Research Laboratory at Aberdeen Proving Grounds in Maryland.

Attempting to move into Phase 2 uncovered useful information for long-term deployment with the DoD and other data centers:

1. Understanding the ownership of the servers is critical to success. Both 'clients' and 'hosts' need to appreciate the value of liquid cooling.
2. Implementation of liquid cooling is best done when new servers are purchased, either to upgrade existing servers or to support new missions.
3. Server owners (clients) must understand the value of liquid cooling to their mission (beyond energy savings) to embrace the technology.

The change of demonstration sites introduced considerable delays into the project. In an effort to claw back some of this delay and armed with commercial success for the technology outside of the DoD a decision was taken to perform phases 2 and 3 simultaneously at ARL, the new demonstration site.

4.3 UPDATED FACILITY/SITE DESCRIPTION [PHASE 2 AND PHASE 3]

The ARL at Aberdeen, Maryland was chosen to host Phases 2 & 3 of the demonstration. The ARL data center was exempt from the consolidation effort on-going elsewhere in the DoD. The site owned and operated all the servers within the data center. The site is a HPC site rather than an enterprise data center meaning its servers run at higher utilization and energy savings are more apparent than in sites with lower utilization. The server population within the site was much more homogeneous than had been experienced at Redstone. This is because HPC data centers tend procure servers in 'clusters' where a large number of matched performance computers work together to solve a problem. The site also had experience with deploying liquid cooling from previously deployed supercomputers.

The funds available for facilities work at ARL were too limited to support installation of a warm water-cooling loop. Instead, the decision was taken to tap into the existing building chilled water system. This decision limits the ability to demonstrate the full energy savings potential of the technology as the expense of chilling water to cool the servers were still be incurred, despite this chilling being un-necessary for Direct-to-Chip liquid cooling technology cooled servers.

The data center at ARL is split into multiple rooms or halls. The 'white cell' room was chosen for the demonstration. Distribution plumbing was installed and instrumented to bring chilled water from the existing facilities plumbing to the racks in the white cell room. At the time of selection, the white cell room contained 34 rack of IT equipment, not all of them servers. Of interest were 5 racks of Appro servers and 8 racks of IBM servers. The room was cooled by 3 CRAC units with outdoor air-cooled condensers.

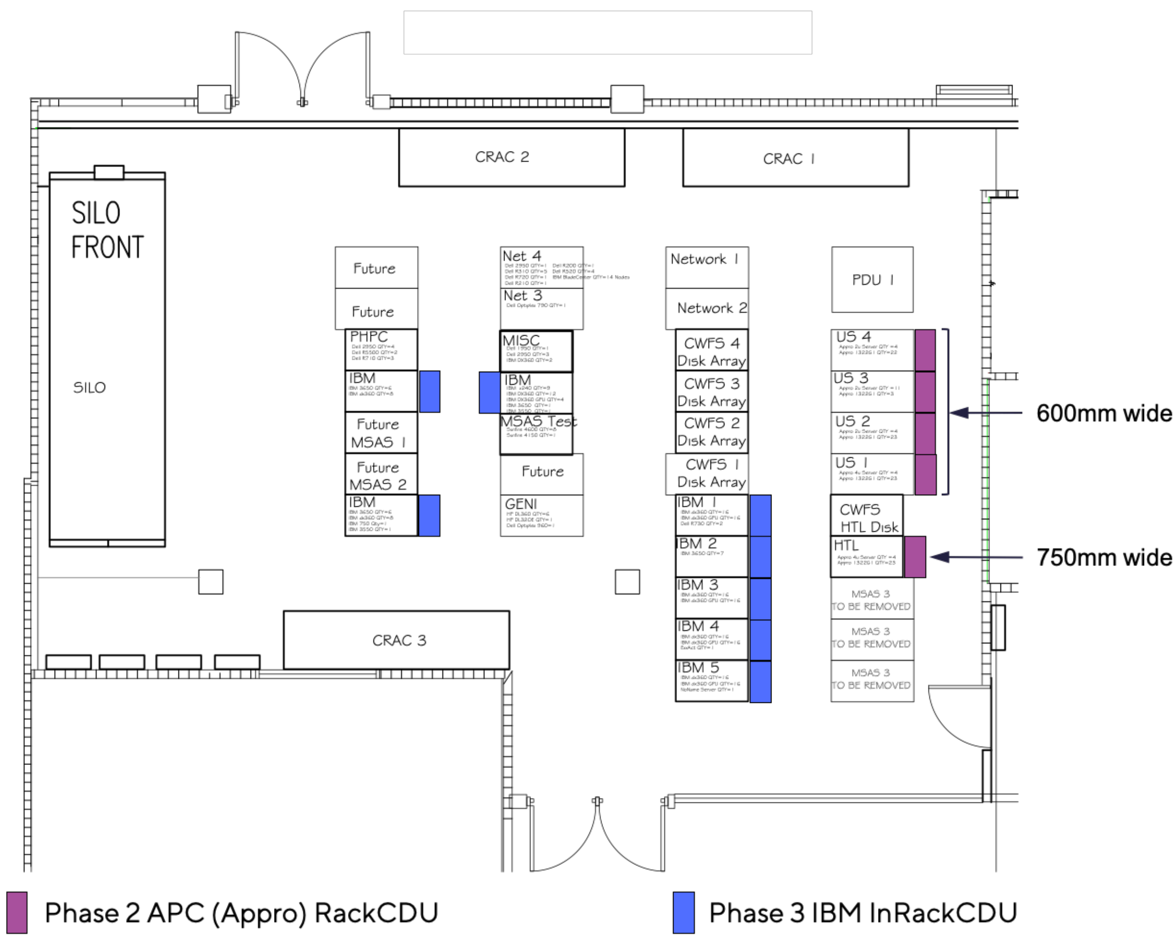

Figure 8. ARL 'White Cell' Data Hall Floor Plan

Figure 8 illustrates the floor plan of the 'White Cell' data hall at ARL. The colored rectangles indicate the positions of the Direct-to-Chip liquid cooling technology and the text in the adjacent white rectangle indicates what servers where original in these racks. As noted, the Appro servers were designated as Phase 2 servers and the IBM servers were designated as phase 3 servers. The intent was to execute both phases simultaneously. An interesting feature of this hall is the viewing window to the right of the door at the bottom of the floor plan.

5.0 TEST DESIGN

The objective of the technical approach was to introduce this new technology into the data center gradually, first with a step that minimized risk to the operation of the data center, then as confidence was gained increasing the size of the demonstration. The project team chose a 3-phase plan to do implement this approach. The project team also wanted to demonstrate two deployment modes: installation of servers with pre-installed liquid coolers and retrofit of liquid coolers into existing systems.

In phase 1 the project team would install new servers equipped with pre-installed liquid cooler and operate them initially with test loads. This demonstrates that liquid cooled servers run reliably. Eventually, the site would be able to migrate real workloads to these servers and the operators would gain experience and confidence in working with liquid cooled servers.

In phase 2 the project team would retrofit a small number of existing servers with liquid cooling. Expectation was these servers would occupy 3 to 5 racks when air cooled, and the project team would consolidate them into 2 racks when liquid cooling them. In addition, this phase was is designed to show that retrofitting can be accomplished with minimal disruption to data center operations, show resolution of warranty issues and further enhance user confidence. These servers were to be performing non-mission critical tasks, keeping the risk low. At the end of phase 2 the project team planned a check point to determine if the site was comfortable moving to phase 3.

In phase 3 the project team planned to do a full-scale retrofit of a large number of servers. Here would retrofit some 500 servers distributed across roughly 33 racks and consolidate them into ~13 racks. The servers were expected to consume ~250W of power and liquid cooling them would have a measurable impact on the air-cooled load in the data center. A simulation of waste-heat recovery was also planned. Phase 3 is where all of the performance objectives were to be achieved.

The principle performers for the project were identified as: Asetek, Inc. the inventor of the technology and the company intending to configure, manufacture and market products embodying Direct-to-Chip liquid cooling technology, D2C technology. Johnson Controls is the third-party that would implement the necessary facilities level liquid cooling loop, including the design, provision of needed materials, fixtures and appliances as well as much of the instruments and monitoring equipment. Signature Technology Group (STG) would provide the labor necessary to retrofit servers and install Direct-to-Chip liquid cooling technology's in phases 2 and 3 (project team did this in phase 1). STG would also make third-party warranty service available should this not continue to be provided by the server original equipment manufacturer following the retrofit with liquid cooling. The Department of Energy' NREL would provide data analysis and reporting for the project.

5.1 CONCEPTUAL TEST DESIGN

Fundamental Problem: This project demonstrated the energy/cost savings, consolidation potential and reliability of Direct-to-Chip liquid cooling technology in the two main deployment modes (retrofit in-place and pre-installed).

<u>Demonstration Questions</u>: 1) Annual and seasonal energy savings compared to air-cooling; 2) annual and seasonal energy generation from waste-heat; 3) impact on server reliability/up-time; 4) quantify the potential for consolidation for future capacity expansion; 5) impacts on non-energy operating costs, occupant productivity and comfort; 6) cost impact, if any, of retrofit warranty issues; 7) simple payback rate and lifetime cost benefits; 8) cost/energy/GHG impact across all DoD building stock, and 9) general "acceptance" of Direct-to-Chip liquid cooling technology as a long-term DoD data center solution. *The assumptions for costing purposes* are included in the Economic Benefits section, below.

The project team collected quantitative and qualitative data addressing each PO.

- *Hypothesis*: Direct-to-Chip liquid cooling technology will provide significant lifecycle energy savings, energy security, cost-savings and infrastructure capital avoidance relative to air-cooling across most DoD data centers, will not negatively impact operations, reliability or warranties, and will be broadly accepted by data center operators.
- *Test Design*: Installation and testing occurred in three stages to maximize the probability of success and mitigate any possible operational risk to the data center host (while also building confidence and acceptance over time): Stage 1: New Server Installation and Testing (no risk; new servers are installed on-site, but are not part of the day-to-day workload of the data center). Stage 2: Small-Scale Retrofit Testing (minor risk; retrofit of 3-5 racks of servers that handle "non-critical" workloads). Stage 3: Full-Scale Retrofit and Testing (higher risk; retrofit to ~400kW of server capacity in the facility, including critical systems).

This three-stage testing design gave the Project team an opportunity to demonstrate the ability to install and operate Direct-to-Chip liquid cooling technology without impacting server up-time. It also gave the DoD operating team a chance to become familiar with the system and its operation at a small scale before deploying across the entire data center. This ensured that the team is ready to operate the full system and provide confidence that operations will not be impacted in any negative way.

<u>Stage 1</u>: One rack of new servers was installed, half air-cooled and half with Direct-to-Chip liquid cooling technology internal cooling loops pre-installed. Side-by-side measurements was be made. This model represents the "pre-installed" mode of data center retrofit. In a second step, the air-cooled servers were retrofitted to liquid-cooling and the rack re-measured, demonstrating "retrofit-in-place". The NEDC technical team ran all servers as part of their day-to-day operations, to become familiar with the system and ensure reliability under typical loads.

<u>Stage 2:</u> Five racks of existing servers were retrofit ("retrofit-in-place"), but without server consolidation. Data was collected before and after the retrofit. In a second step, the servers were moved from the original 5 racks into 2 racks and re-measured, demonstrating consolidation potential.

User surveys and interviews during Stage 1 and 2 ensured no undue burden was placed on the data center team when deployed broadly. Stage 1 and 2 monitoring also validated energy savings and reliability in a full operating environment (but at small scale).

Stage 3 deployed Direct-to-Chip liquid cooling technology across the rest of the data center through a combination of retrofit-in-place and pre-installed (leveraging DoD's planned server refresh and expansions). Data was collected before and after installation.

- *Independent variables*: Cooling type (air vs Direct-to-Chip liquid cooling technology); install mode (retrofit-in-place vs pre-installed).
- *Dependent variables*: See Performance Objectives section, above.
- *Controlled variables*: Side-by-side and before-and-after measurements, as described "Test Design" above. Local weather and Data Load were used to normalize for variations in operational and environmental conditions throughout the project.
- *Test Phases*: See Test Design section, above.

5.2 BASELINE CHARACTERIZATION

The goal of this demonstration effort is to demonstrate direct-to-chip liquid cooling using direct-to-Chip liquid cooling technology in a full-scale operational data center deployed through both installation modes (in-place and pre-installed). The project team installed sub-metering equipment and instrumentation to monitor the following variables: 1) Server load; 2) CRAC load; 3) Direct-to-Chip liquid cooling technology input and output temperatures and internal pump loads; 4) facilities loop pumps and warm water plant; 5) waste heat flow-rate and temperature difference; 6) server "up-time"; 7) other miscellaneous equipment loads; and 8) weather.

The project team then collected a baseline of the existing data center and surrounding building for 12 weeks prior to installation of the Direct-to-Chip liquid cooling technology. Direct-to-Chip liquid cooling technology was the retrofitted throughout the data center using retrofit-in-place and pre-installed modes and monitored for 6 months.

5.3 DESIGN AND LAYOUT OF TECHNOLOGY COMPONENTS

A single rack containing 20 HPE DL560 2U servers with an attached OnDirect-to-Chip liquid cooling technology was deployed in phase 1. All of this equipment was new including the rack. Ten of these servers were liquid cooled and 10 were air cooled. Each DL560 contains 4 processors (CPUs). All 4 processors were liquid cooled using two separate cooling loops in the liquid cooled servers. The liquid cooling in these servers was pre-installed off-site. The Direct-to-Chip liquid cooling technology was also installed at off-site and the full setup was tested and then shipped to Redstone. The product provider with the assistance of Johnson Controls and NEC-R personnel installed the rack, Direct-to-Chip liquid cooling technology and servers once the equipment arrived at Redstone. Figure 9 shows a DL560 server before and after being retrofit with liquid cooling. The retrofit took approximately 40 minutes during which time 2 liquid cooling loops were installed. A more typical 2 processor server would only require 1 loop.

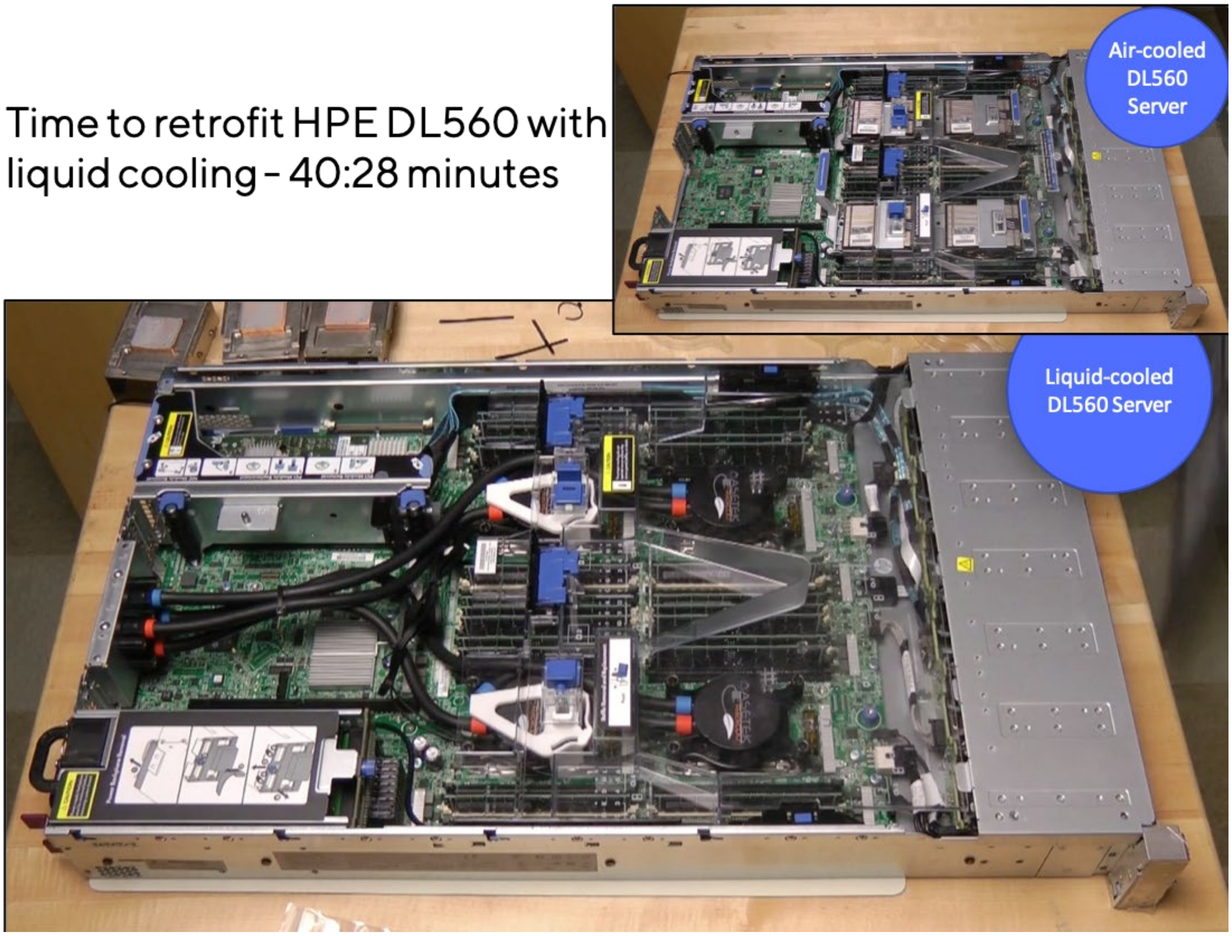

Figure 9. DL560 Server Before and After Liquid Cooing Retrofit

Figure 10 shows the phase 1 rack installed at Redstone. This view is from the rear or hot aisle side of the rack. The rear doors are mounted on the rear of the OnDirect-to-Chip liquid cooling technology to preserve physical security. The OnDirect-to-Chip liquid cooling technology extension is mounted on the rear of the rack and extends a bit above the rack. Liquid tubes are visible in the bottom half of the rack connecting the Direct-to-Chip liquid cooling technology manifolds (left) to the servers at right. The orange device at upper left is a flow control valve and the gray lines behind it are the facilities liquid hoses. The other lines above the rack provide power to the servers.

Figure 10. Rear View of Phase 1 Rack Installed at Redstone

This section should provide a thorough description of all technology components, as applicable. Provide a subsection for each significant technology component describing its design and location. Provide detailed maps or drawings showing the location(s) of all technology components. Schematic diagrams of equipment showing details should be provided when such equipment is an integral part of what was demonstrated. If details are available in other published documents, a brief description should be provided, and reference made to those documents. This section also should provide a thorough description of any controls used.

5.4 SAMPLING RESULTS

In Phase 1 of the demonstration, the top half of the rack was installed with air-cooled servers and the bottom half installed with liquid-cooled servers. This allowed precise side-by-side performance comparisons to be made. Figure 11 shows initial energy measurements from this system, comparing cooling energy for air-cooled vs. liquid-cooled servers. Overall, the initial system provides an average savings of over 60% in cooling energy (with memory dual in-line memory module (DIMM) coolers, not included in this installation, efficiency is expected to be 67%).

In addition to cooling savings, the project team see almost 11% energy reduction from the servers themselves, due to reduced internal fan speed (Figure 12).

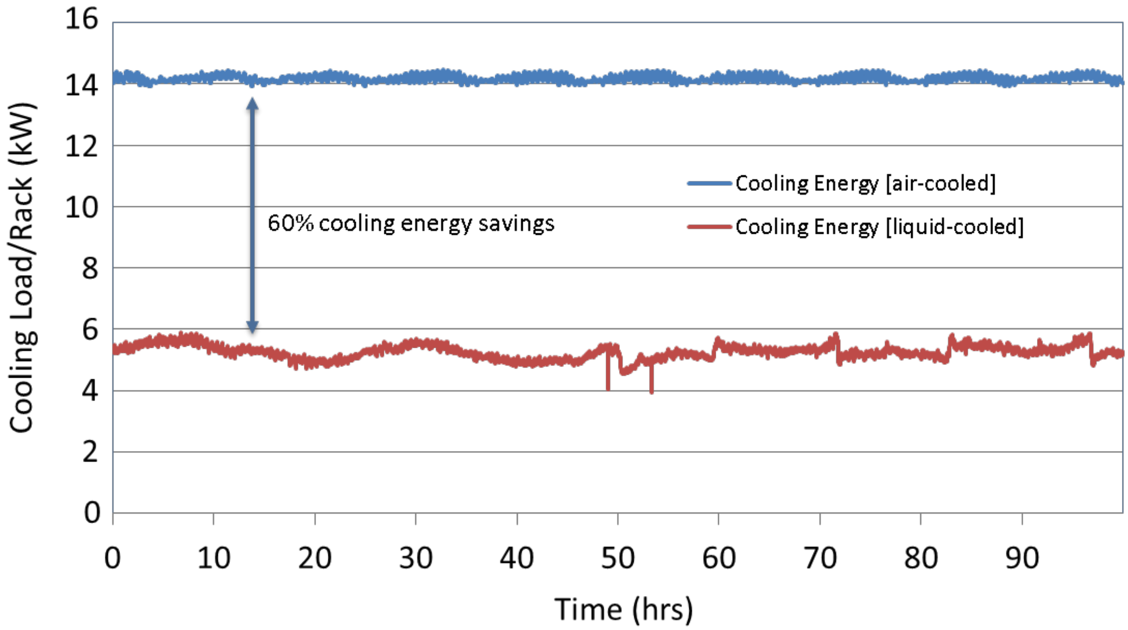

Figure 11. Cooling Energy Savings Demonstrated in Phase 1

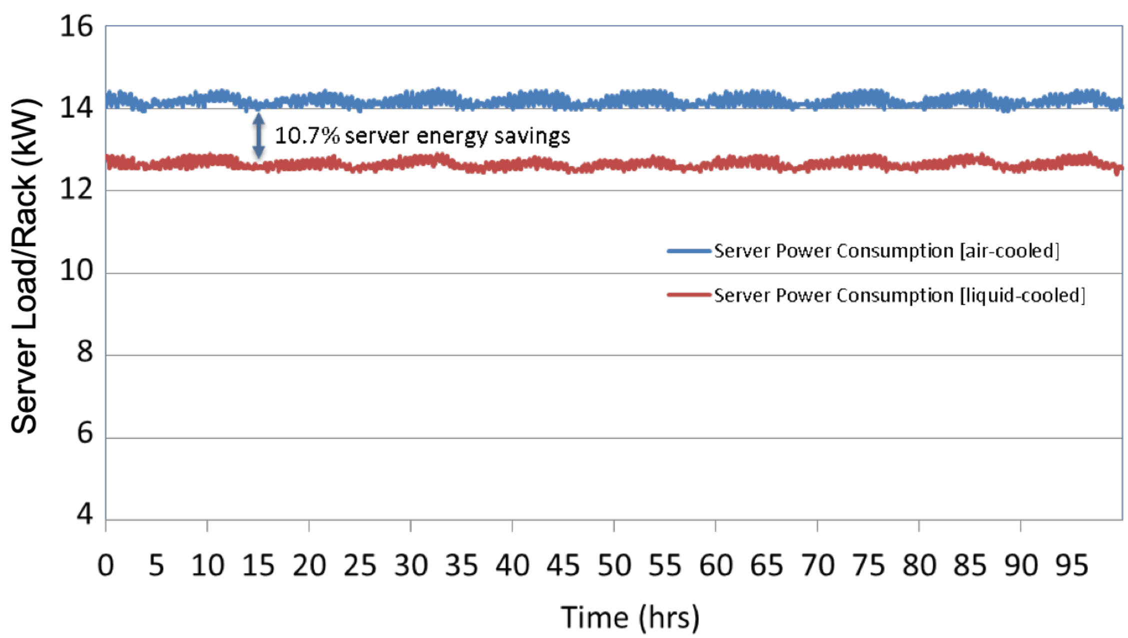

Figure 12. Server Energy Savings Demonstrated in Phase 1

5.4.1 Lessons Learned

The above results were obtained while all 20 servers were running high utilization test loads. The DL560 serves were specifically selected for the demonstration in consultation with NEC-R staff for their capability to run highly virtualized workload. Virtualization in enterprise environments improves server utilization. The expectation was that once the test loads had run for several months, that NEC-R would move virtualized production jobs on to these servers. However, this never happened. Initially there was a security concern about the network connections used monitoring system. When it was explained to the staff raising this concern that monitoring systems were on an out-of-band network that could not communicate with NEC-R's production networks this concern seemed to be resolved. However, for reasons that remain unclear, production work was never transferred to these servers. Perhaps more frequent contact between the product team and the NEC-R team to keep focus on this plan would have resolved this but that did not occur. Thus, energy savings while running production workloads were not quantified.

6.0 PERFORMANCE ASSESSMENT

6.1 PERFORMANCE OBJECTIVE 1: COOLING ENERGY USAGE

The objective of the Direct-to-Chip liquid cooling technology system is to improve the energy efficiency of data center cooling. This is accomplished by transferring cooling load from an existing cooling system (CRAC air-handling units in this demonstration) to a more efficient liquid cooling system, the overall energy directed towards the cooling systems decreases. This is largely because the liquid used in the Direct-to-Chip liquid cooling technology system is three orders of magnitude more efficient at removing heat as compared to air (Sickinger, 2014). For this demonstration, meter readings give the rate of heat removal from the servers by the Direct-to-Chip liquid cooling technology system. This is heat load that the in-place cooling system (the CRAC units) no longer have to remove from the space. By transferring this heat load from the CRACs to the more efficient liquid system, the overall energy required to cool the server room decreases. Meter readings should show an increase in chiller load and a decrease in CRAC load. Data for energy into the servers can give the total server load, too, which yields the percentage of total load that is removed by the Direct-to-Chip liquid cooling technology liquid cooling system.

6.1.1 Server Energy Used

By analyzing the electrical load into the servers and the rate of energy removal from the servers by the Direct-to-Chip liquid cooling technology system, the project team can calculate the amount of heat removed by the liquid cooling system (and thus diverted from the CRAC cooling system) and also the percentage of load removed by the technology.

The only time window for which the project team obtained both server energy in and energy removed by the Direct-to-Chip liquid cooling technologys was 4/18/2018-9/5/2018. At this point, 13 racks of servers (8 IBMs, 4 Appros, and 1 Penguin rack) in the “White Cell” server room had already been retrofitted with the Direct-to-Chip liquid cooling technology. These data points, however, were only available for five racks of the IBM servers (labeled fob1-fob5). The following graphs visualize this:

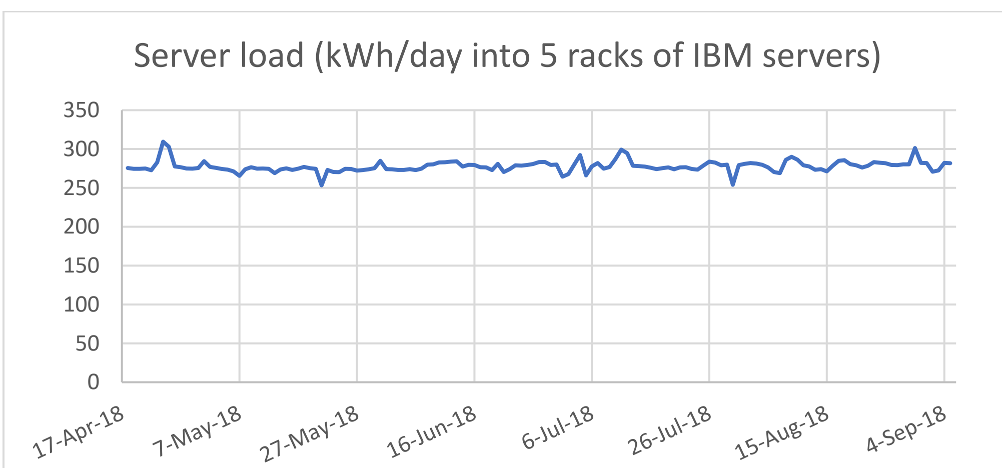

Figure 13. Server Load into the 5 IBM Servers Post-retrofit

6.1.2 Energy Removed

Meter readings for the Direct-to-Chip liquid cooling technology system provided data for flow rate (GPM) and cooling liquid temperature in and out (°C) on a 15 second interval. As the heat exchange fluid flows through the servers, it removes heat which raises the temperature of the working fluid. This difference in temperature and the flow rate of the fluid can give the rate of energy removal from the servers. These fifteen second readings were used to average a total energy removed over those fifteen seconds using the formula:

(a) $\dot{Q} = C_p * \dot{m} * \Delta T$,

where $\dot{Q}$*= energy flow rate (Btu/hr),* C_p *= specific heat (Btu/lbm * °F),* $\dot{m}$ *= mass flow rate (lbm/hr),* ΔT *= temperature difference (°F).*

After applying the appropriate conversion factors and taking C_p *= 0.998 Btu/lbm-°F, ρ = 8.33 lbm/gal, and converting the Celsius temperature differences to Fahrenheit by multiplying by 1.8, the above equation becomes*

(b) $\dot{Q} = 500 * gpm * \Delta T$,

where $\dot{Q}$ *is measured in Btu/hr* (Sickinger, 2014).

After converting from Btu/hr to Watts and using this instantaneous meter reading as a representation for the 15 second interval, total kWh draw over those 15 seconds can be calculated as follows:

$$\frac{\boldsymbol{X}\,Watt * 15\,sec}{1} * \frac{1\,min}{60\,sec} * \frac{1\,hour}{60\,min} * \frac{1\,kW}{1000\,W} = \boldsymbol{Y}\,kWh$$

These 15-second energy outputs throughout an entire day can be summed to give the energy removed per day, as shown below:

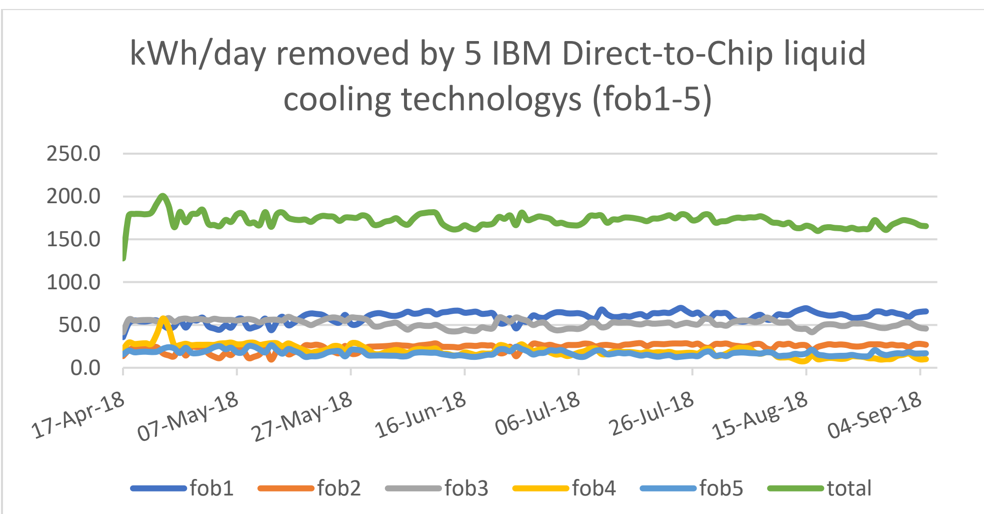

Figure 14. Energy removed by Direct-to-Chip Liquid Cooling Technology from 5 Racks of IBM 'Fob' Servers - Individual and Total

(The above graph includes a total, the below graph does not).

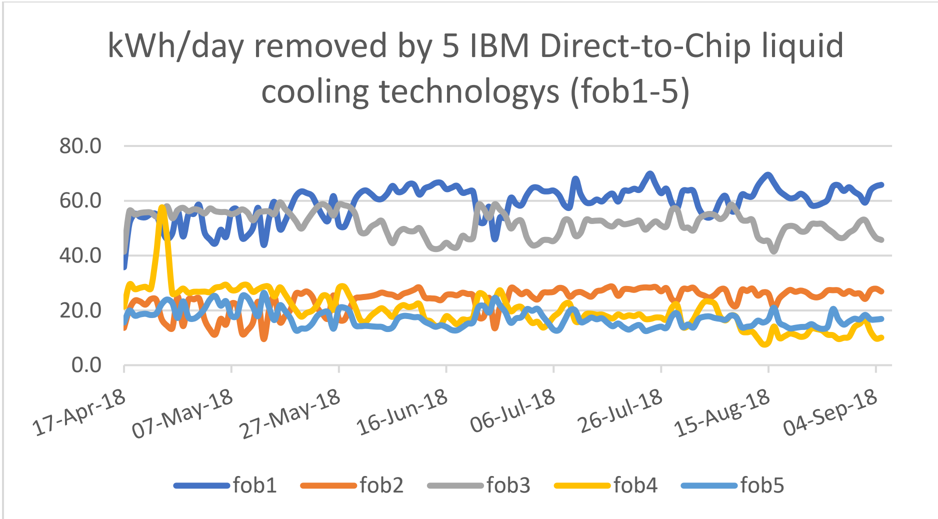

Figure 15. Energy Removed by Direct-to-Chip Liquid Cooling Technology System for the 5 Racks of IBM Servers

The charts in Figure 14 and Figure 15 show that, while some servers are more active than others, the Direct-to-Chip liquid cooling technology system removes a relatively steady ~175 kWh/day (average of 172.5 kWh/day) from these five racks of IBM servers over the post total retrofit time frame from April 2018 through September 2018.

6.1.3 Efficiency

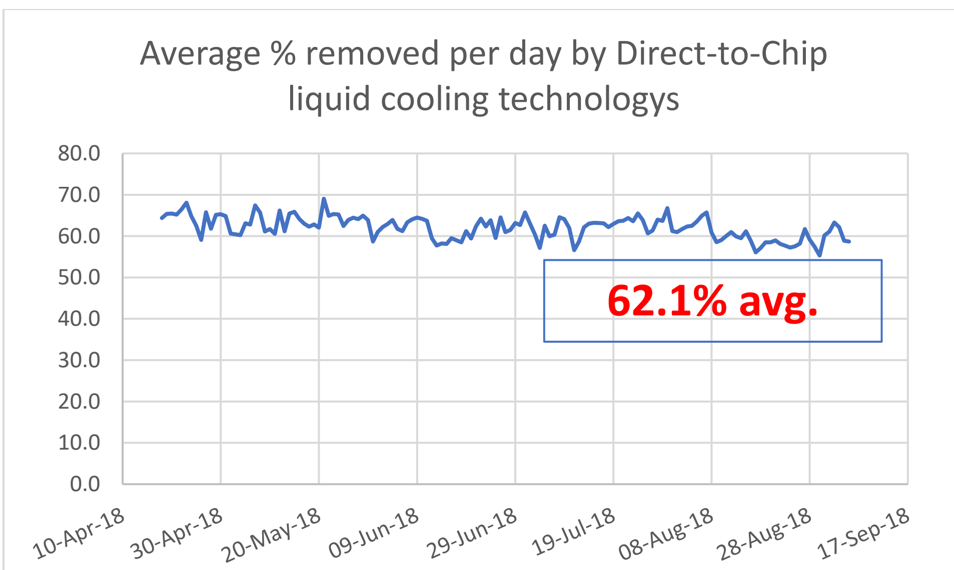

Figure 16. Percentage of Input Energy Removed by the Direct-to-Chip Liquid Cooling Technology for the 5 IBM Servers

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% from April 17, 2018 through September 4, 2018.

6.1.4 Other Servers

For the remaining retrofitted servers (3 IBM racks named 'pitch,' 'yaw,' and 'cst,' 4 Appro server racks, and the penguin server rack) data were only available for the energy removed from the servers by the Direct-to-Chip liquid cooling technology and not for the energy into these servers. These data for flow rate, temperature in and temperature out were provided at 15 second intervals for the same time period as for the IBM fob servers detailed above. The figures below depict the energy (kWh/day) removed from these servers by the Direct-to-Chip liquid cooling technology system.

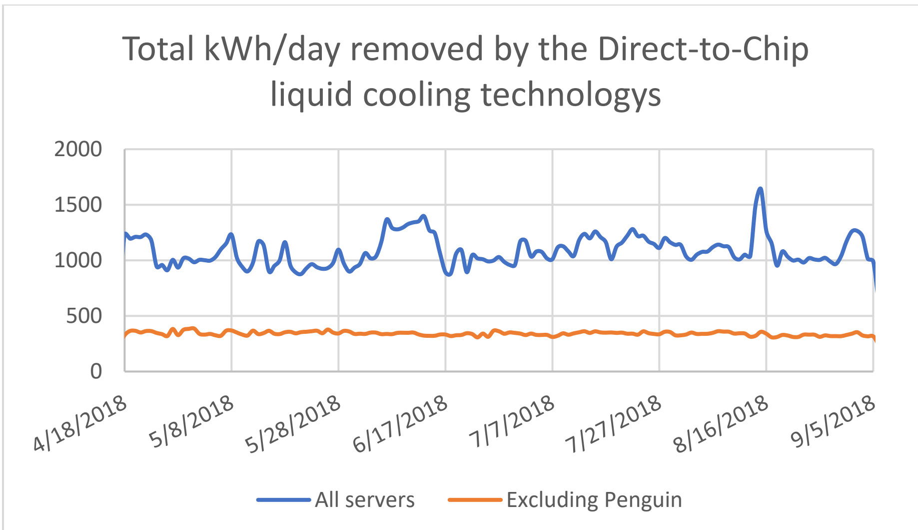

Figure 17. Energy Removed by the Direct-to-Chip Liquid Cooling Technology System Across All 13 Retrofitted Server Racks

The average daily kWh removed by the Direct-to-Chip liquid cooling technologies from all the retrofit servers was 1087 kWh/day. While there are some data dropouts and inconsistencies, the Direct-to-Chip liquid cooling technology units on the 8 IBMs and 4 Appro servers remove a relatively steady amount, averaging 341 kWh/day. (See Appendix B)

The Direct-to-Chip liquid cooling technology system removed 62.1 % of the server load from the IBM servers. If the project team maintain the assumption that the Direct-to-Chip liquid cooling technology system removes 62.1% of incoming server load from the other servers as well and knowing the amount of energy removed from the 13 servers, the project team can estimate the amount of energy that goes into these 13 servers. The average (from above) of energy removed by the system was 1087 kWh/day, which corresponds to **1693 kWh/day** of electrical power delivered to those servers equipped with the Direct-to-Chip liquid cooling technology system. If these servers represent an approximated half of the load of the entire room, the project team estimate **the total room load to be ~ 3386 kWh/day.** Actual server room load data was gathered for three days in November 2018, and is given below:

Table 3. Data for white cell load from 3 select days in November 2018.

Panel	Day 1	Day 2	Day 3	<<- White Cell Load
Panel 1	75	70	68	
Panel 2	77	76	76	Voltage (V)
Panel 3	45	44	44	208
Panel 4	5	5	5	
PDU	220	221	222	Averages
Sum (amps)	422	416	415	418
kW	152	150	150	150
kWh/day	3649	3597	3588	3611

Taking these as a daily average values and averaging them puts the **draw of the server room at around 3611 kWh/day.** This is of the same order as the 3386 kWh/day approximated above, for reference. Taking the 3611 kWh as the actual load implies that, in reality, approximately 47% of the White Cell room server load is retrofitted with the Direct-to-Chip liquid cooling technology system. Should all of the servers be retrofitted, and the system continue to remove 62.1% from all of the servers, the Rack CDU system would remove an average of 2242 kWh/day from the entire room.

6.1.5 Other Data

There is also sporadic data for energy removed from the servers from earlier in the project lifecycle which may be of use to the report. This first figure shows that the Direct-to-Chip liquid cooling technology begin expelling heat from the servers after the system is installed on the initial four Appro servers. After this retrofit, the Direct-to-Chip liquid cooling technology liquid cooling system removes between **225 – 300 kWh/day**.

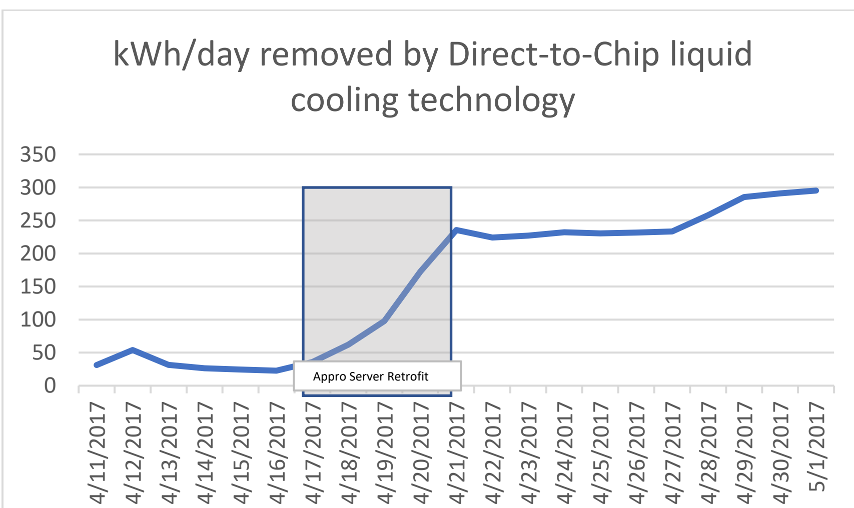

Figure 18. Energy Removed by the Direct-to-Chip Liquid Cooling Technology System Before and After the Retrofitting of the First 4 ervers.

6.1.6 Chiller Plant Data

The 1800-ton chiller plant bears the added cooling burden from the Direct-to-Chip liquid cooling technology units. This is a boon to the overall energy efficiency of the site, as the chiller cooling system is more efficient at cooling the space than the air handling system, as liquid cooling as a heat-exchange medium is three orders of magnitude more efficient than air (Sickinger, 2014). The project team only have a limited amount of energy data from the chiller plant, as is shown below:

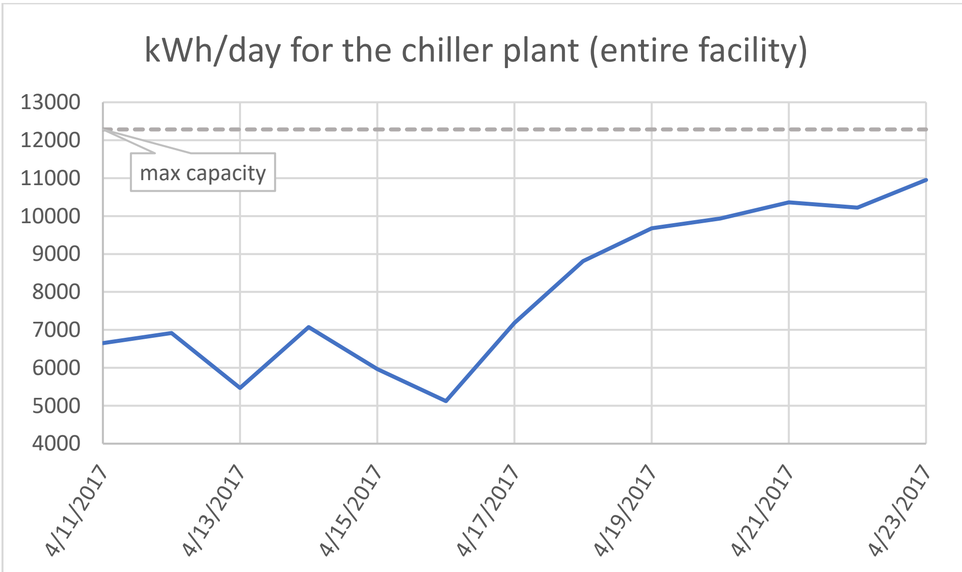

Figure 19. Data Available for the Chiller Energy Use for the Whole Plant.

Theoretically, there is an increase in chiller energy use as the Direct-to-Chip liquid cooling technologys are connected. The scale of this plant, however, is on the order of 10,000 kWh/day, while the increased load from the Direct-to-Chip liquid cooling technology liquid cooling system ranges from 400-1,200 kWh/day (see Figure 18). So, given the limited chiller data, and the disparity in cooling loads, the project team cannot glean anything significant from the chiller plant data. *Note: If the 1800-ton chiller operated at full capacity for a full day, the load would be 12,284 kWh/day*

The project team also have chiller plant load data for 10/16/18. Taking the sum of the three-chiller shelter AC electrical panels and dividing by the instantaneous load meter reading yields a kW/ton efficiency. Taking 3.517 and dividing by this number will convert to COP. COPs in economizer and non-economizer modes are given in the figure below:

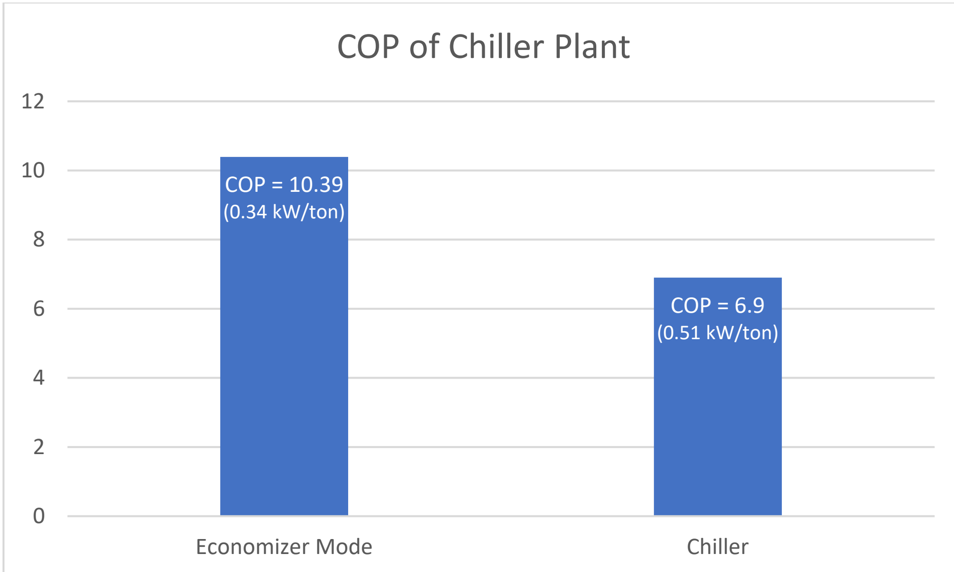

Figure 20. COP of the Chiller Plant in Economizer Mode and with Chiller Running.

Estimating that the chiller runs in economizer mode for 3000 hours a year (*based on communication with the site operator*), or 34.2% of the year, then a weighted average of the two COPs above would give a yearly **average COP of 8.09.**

6.1.7 CRAC Air Handling Units

- The CRAC data did not show clear enough trends over the study period to confidently quantify any savings from reductions on the air-cooling side of the project. Ideally, the Direct-to-Chip liquid cooling technologys would absorb some (typically ~50%) of the cooling load of the space's previous cooling system, the three 30-ton CRAC units in this case. The air-cooled system power could then be reduced, resulting in savings. There are several possible reasons the data did not show a reduction in CRAC use:

  a. Until the IBM CPU retrofit date (late Aug, early Sept 2017), the CRACs were set to a constant cooling setting – a constant kW draw. Thus, no savings were seen after the Appro Installation and no clear trend is seen after the IBM CPU install (See Figure 21).

  b. Throughout the study period, servers were added and removed from the room (as is typical in server rooms), as equipment was added and upgraded. Thus, the actual heat load of the room that needed be removed was regularly changing. Without more precise data on the dates of these changes and the amount of server load that was added and removed, the project team cannot draw a conclusion on the infield realized savings via a reduction in CRAC cooling output.

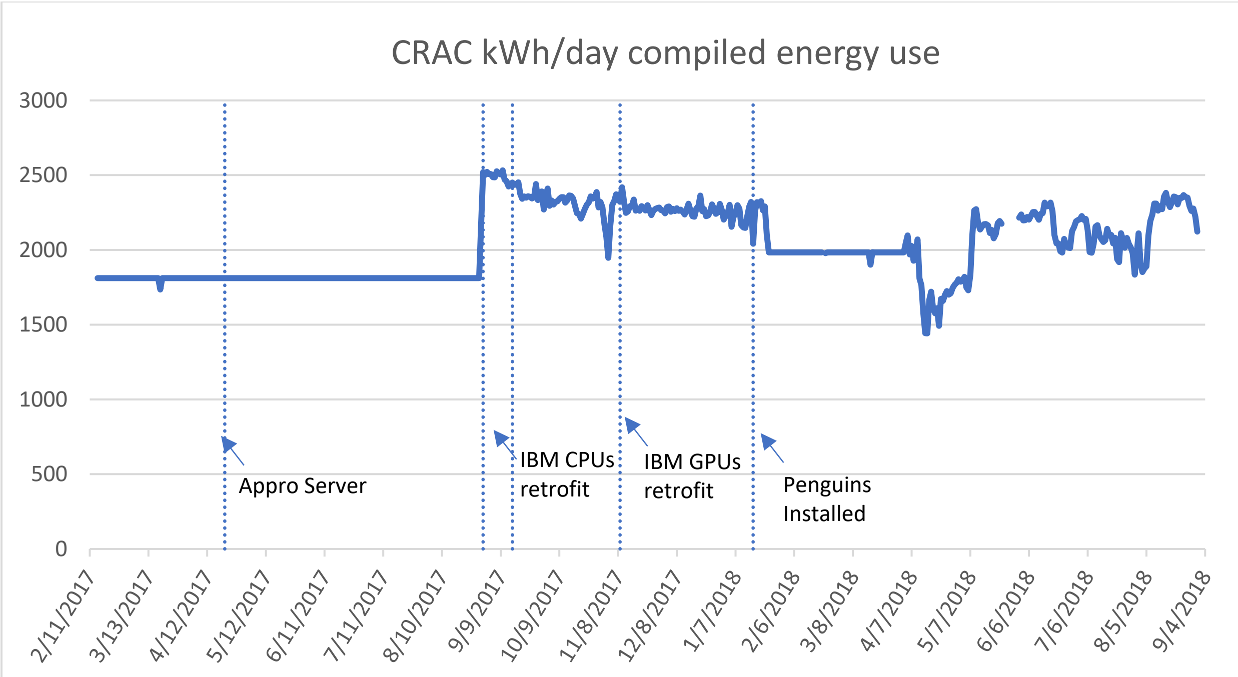

Figure 21. Energy Use by the Air Handling Unit Cooling System

The average for the post-retrofit period is 2043 kWh/day. Pre-installation was 1811 kWh/day.

6.1.8 Overall Calculations

All of the data and calculations below are based on data from 2018, after all the servers had been retrofit.

The data below is compiled from the previous sections of this summary:

Table 4. Summary of Calculated Load Data for the Various Cooling System Components.

	Average Load	COP
White Cell Power Draw	3611 kWh/day	8.09 (6.9 Chiller, 10.4 Economizer)
Direct-to-Chip liquid cooling technology Removed	1087 kWh/day	
CRAC Electricity Use	2043 kWh/day	

Assuming all the electricity into the room must be rejected as heat, the project team can calculate the efficiency savings realized by shifting the cooling load from the CRAC units to the chiller plant, as follows:

1. Of the 3611 kWh/day delivered to the white cell server room, 1087 kWh/day is removed by the Direct-to-Chip liquid cooling technology system. This leaves **2524 kWh/day** for the CRAC air handling units to remove.

2. The CRAC units use an average of 2043 kWh/day to remove this 2524 kWh/day of load. This means the CRAC units have a **COP** of 2524/2043 = **1.24.**
3. With this demonstration, the Direct-to-Chip liquid cooling technology system has shifted 1087 kWh/day from the CRAC system (COP 1.24) to the Chiller (COP 8.09). The CRAC would have used 1087/1.24 = 877 kWh/day to remove this load. The chiller system only needs to use 1087/8.09 = 134 kWh/day, effectively **saving 743 kWh/day, or 271,381 kWh/year.** This is comparable to the energy use of 19 average American residential houses (EIA, 2018).
4. The total cooling energy use from both the CRAC units and the chiller system averages to 2043 + 134 = 2177 kWh/day. This equates to an overall system COP of 3611/2177 = 1.66.
5. The CRAC units need to remove 2524 kWh/day from the space. This equates to a constant draw of 105 kW, or 30 tons of average load. There are currently three 30-ton CRAC units in the space, so one of them is redundant and could be hibernated and saved for use as a backup. As an average, the 30-ton load will certainly be exceeded, so two of the CRAC units would still be necessary, and certainly adequate to cover the peak load.
6. The average cooling energy intensity (kWhcooling/kWhserver) is
   - (134 kWh from chiller + 2043 kWh from CRAC) / (3611 kWh from server) =
   - 0.603

Unfortunately, the project team cannot perform a pre-retrofit analysis for comparison because the project team don't know the before white cell load (and CRAC load is lower-likely due to changes in load of the room, which would also lead to an indirect comparison).

6.2 PERFORMANCE OBJECTIVE 2: SERVER ENERGY USAGE

The project team don't have this data. The project team only have partial data on the IBM servers very small sample pre-retrofit, large sample post-retrofit. **The project team 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.

6.3 PERFORMANCE OBJECTIVE 3: POWER USAGE EFFECTIVENESS (PUE) AND ENERGY REUSE EFFECTIVENESS (ERE)

In order to calculate PUE, the project team would need more data on total energy loads in the server room, including lights & plugs, pumps. See https://www.nrel.gov/hpc/cool.html. 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. The project team believe that the plan for the original site included energy reuse.

6.4 PERFORMANCE OBJECTIVE 4: DATA CENTER PEAK LOAD

The max CRAC power draw post-retrofit was **131.27 kW** on 7/3/2018 at 4:10 PM. At a COP of **1.24, this removes 131.27 * 1.24 = 162.77** kW. At this time the Rack CDU removed 35.20 kW from the system. This load was diverted to the chiller plant, with a COP of 8.65, requiring 35.20/8.65 = 4.07 kW. The total energy draw then would be 131.27 + 4.07 = 135.34 kW. So, the CRAC units removed 162.77 kW at peak load and the Rack CDU Removed 35.20 kW. If all of this were removed by the CRAC (if the white cell had no Direct-to-Chip liquid cooling technology installed), the CRACs would have needed to remove the total 162.77 + 35.20 = 197.97 kW from the space. At a COP of 1.24, this would require 197.97/1.24 = 159.65 kW.

So, there is a total peak load savings of 159.65 – 135.34 = 24.31 kW. This amounts to a 24.31 / 159.65 = 15.2% reduction in peak energy cooling power.

While the project team do not have total data on all the loads of the white cell room, the project team can incorporate the power into the servers for a better picture of room performance. On Day 1 of the 3-day sample in November 2018, the servers drew 152 kW of power. Taking this as a constant load, the calculation becomes:

The peak load without Direct-to-Chip liquid cooling technology would be 159.65 + 152 = 311.65 kW.

The peak load with the Direct-to-Chip liquid cooling technology = 135.34 + 152 = 287.34 kW.

This leads to a reduction of 7.8% in total room load.

6.5 PERFORMANCE OBJECTIVE 5: SERVER UPTIME

This was not measured during this project due to insufficient data.

6.6 PERFORMANCE OBJECTIVE 6: CAPACITY CONSOLIDATION

Due to the change in site, and hence the change in leadership of the data centers, consolation was not studied, or actualized.

6.7 PERFORMANCE OBJECTIVE 7: EASE OF USE

Since the time of the project, the product provider has deployed Direct-to-Chip liquid cooling technology systems, both for as a retrofit to existing data centers, as well as deploying pre-Installed servers. The main different around the ease of use came down to the difficulties associated with retrofit installations rather than just adding servers that came with liquid cooling pre-installed. It was far more difficult to retrofit existing servers and came with additional risk and downtime, which makes that option unfeasible on a large scale. In contrast, the installation that of factory installed liquid cooling can provide a frictionless installation that is easy to use. Overall acceptance with the users was much higher on factory OEM liquid cooling. Based on the experience of this project and the discussions with stakeholders, retrofitting air cooled servers with liquid cooling is not feasible to scale going forward. However, the retroffting effort still got a nod of approval from users. One feedback received was that 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.

6.8 PERFORMANCE OBJECTIVE 8: LIFECYCLE GHG EMISSIONS

The project team assume GHG emissions reductions by reducing overall electricity usage of the facility which result from transferring the cooling load from air to water (4000x more efficient heat transfer medium). The direct electricity savings from this process (as noted above in Performance Objective 1, *Overall Calculations),* the chiller plant uses 743 kWh/day less than the CRAC units. An additional savings would come should one of the three CRAC units be shut down and kept as backup, reducing the CRAC load by an estimated 1/3, or 681 kWh/day. Together, these represent an electricity reduction of 1424 kWh/day, or 520,116 kWh/year. Using a national average equivalence factor of 0.744209 kgCO2/kWh (EPA, 2016), this 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 far smaller server capacity was retrofit at the new site than was originally planned. While the original SOW called for 475 kW of server load to be equipped with the Direct-to-Chip liquid cooling technology system, this installation retrofit around 70 kW only. There also were no realized heat savings as there was no energy recovery aspect of this installation. (52.79 kgCO2/MMBTU of heat conversion factor).

6.9 PERFORMANCE OBJECTIVE 9: LIFECYCLE COST ANALYSIS

Due to the various changes and complications during project execution, a lifecycle cost analysis was not performed for this project. See the cost assessment section below.

6.10 PERFORMANCE OBJECTIVE 10: END-USER ACCEPTANCE

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 Once installed, user satisfaction is high. Liquid cooling can offer a lot of benefits to data centers, including but not limited to:

- Decreased overall and server energy use
- Decreased processor temperatures
- Reduction in air cooling
- Uniform processor, server, and cluster temperatures
- Increased run time
- Increased performance
- Decreased greenhouse gas emissions

7.0 COST ASSESSMENT

Before the project the project team performed TCO analysis, however the project team 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 National Institute of Standards and Technology (NIST) bridge life-cycle cost analysis (BLCCA) process, the project team 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 5.

Table 5. 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. They are also reasonable and consistent with the expectations, the preliminary models, and the preliminary experimental results from the 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, the project team find it necessary to pay for a 3rd-party warranty to cover the servers that are retrofit, the economic implications will change. Based on the discussions with STG, a leading 3rd-party warranty supplier that will cover liquid-cooling retrofits, a modified BLCCA is provided below, in Table 5a. While the project team 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 5a. 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[iv].
7. For payback and ROI calculations, the project team 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.

8.0 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 the 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. The project team worked extensively with the Redstone team, the Army's data center consolidation team and the Army Chief Information Officer's (CIO's) office to try to address the situation, but the project team were ultimately unsuccessful.

In collaboration with the Army CIO's office and data center consolidation team, the project team 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 the 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).

9.0 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

APPENDIX A POINTS OF CONTACT

POINT OF CONTACT Name	ORGANIZATION Name Address	Phone Fax E-mail	Role in Project
Steve Branton	Asetek	Sbr@asetek.com	Principal Investigator
Tom Kendall	Army Research Lab	N/A	Host Site Lead
Derek Zak	Signature Technology Group	N/A	System Integration Lead
Vincent Cowger	Johnson Controls Inc	N/A	Site Installation
Otto Van Geet	NREL	Otto.VanGeet@nrel.gov	Site Modelling and Data Analysis Lead

APPENDIX B SUPPLEMENTAL DATA

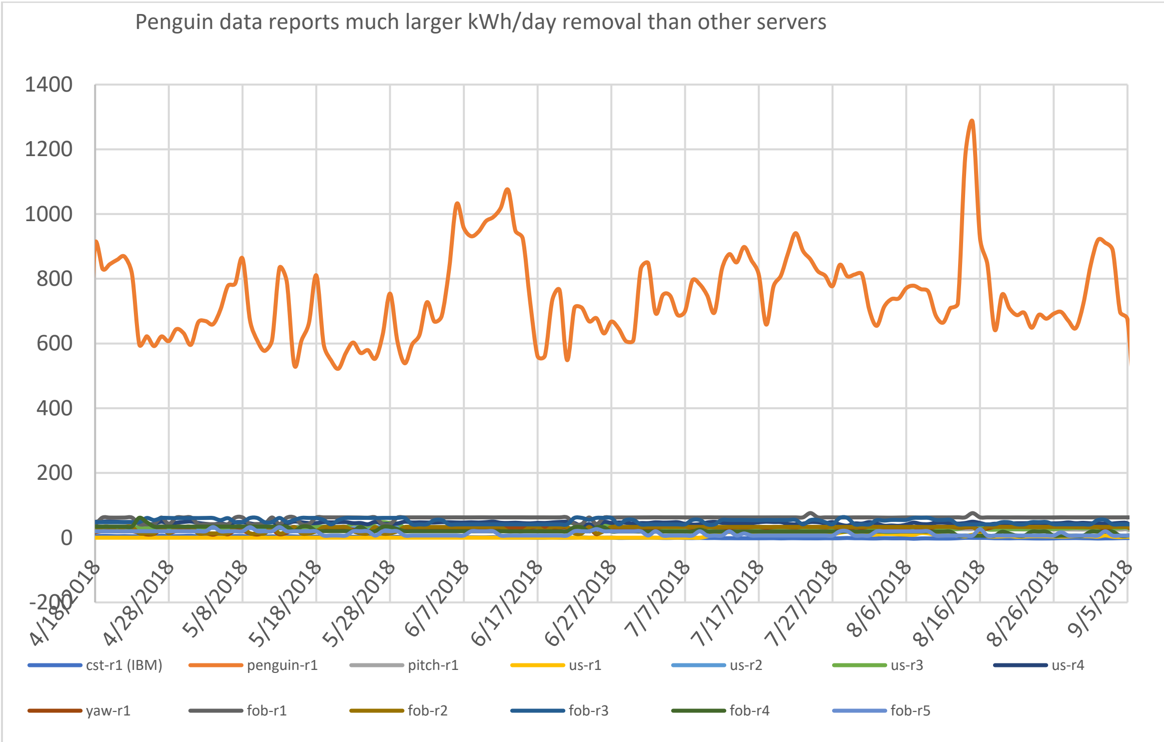

Figure 22. Energy Removed from Each Individual Server by the Direct-to-Chip Liquid Cooling Technology Systems

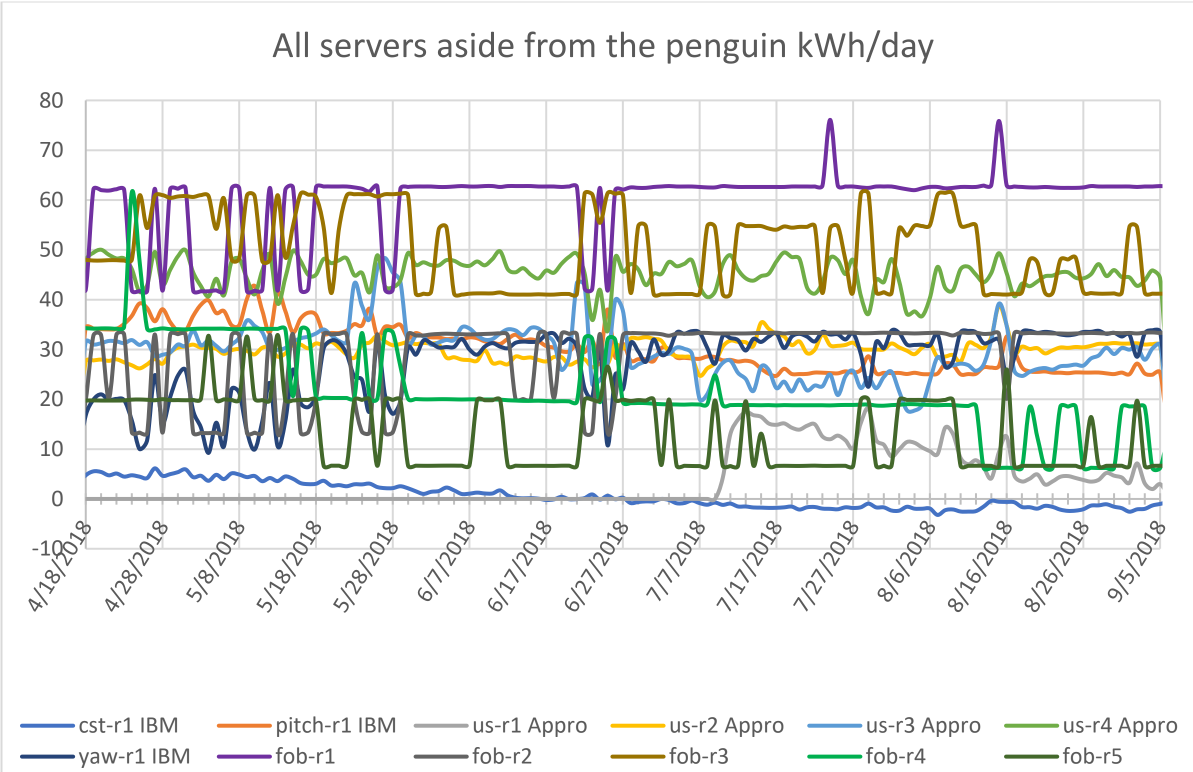

Figure 23. Energy Removed by the Direct-to-Chip Liquid Cooling Technology Systems Installed on the 8 IBM and 4 Appro Racks of Servers Only.

fob-r1 through fob-r5 are IBM server racks

Later in the year, after the technology is installed on eight additional IBM servers, around 400-500 kWh/day of load is rejected by the cooling technology. (The IBM CPUs were all retrofit around 8/29/2017, before these data):

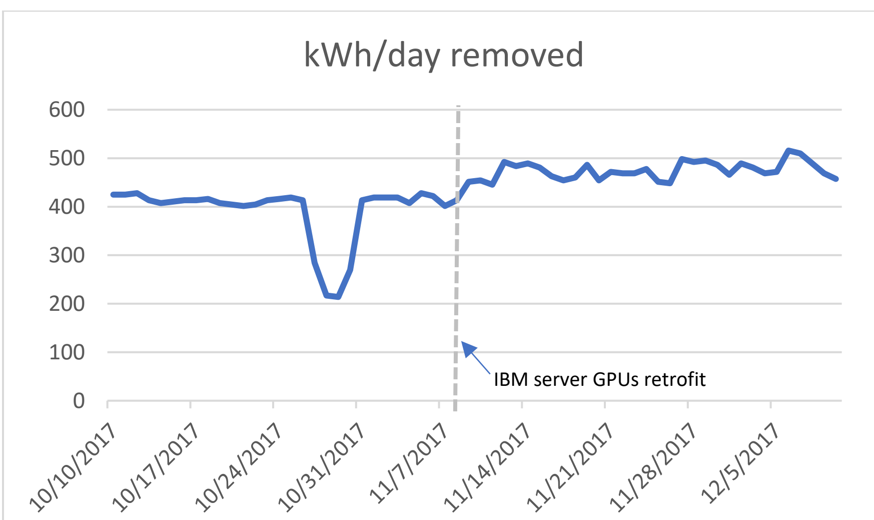

Figure 24. Total Energy Removed by the Direct-to-Chip Liquid Cooling Technology System over Another Time Period

Note that the approximate energy removed (~475 kWh/day) by the Direct-to-Chip liquid cooling technology system post IBM GPU retrofit shown here corresponds well with the amount removed later in the study period as shown in Figure 17 on the data series that excludes the penguin server. The penguin server was not yet installed during the dates shown on the above graph.

The dip in the graph is not a data dropout; there is enough data at every 15-minute interval throughout that time period. The decrease occurred between 10/27-10/30/2017, which was a Friday through Monday. It appears that energy usage just decreased over these days. The graph below provides evidence for this—there were no jumps or dips or gaps in the energy use data (as read from the BTU Meter), it just appears that the rate of use, as represented by the slope of the graph, declines for that period.

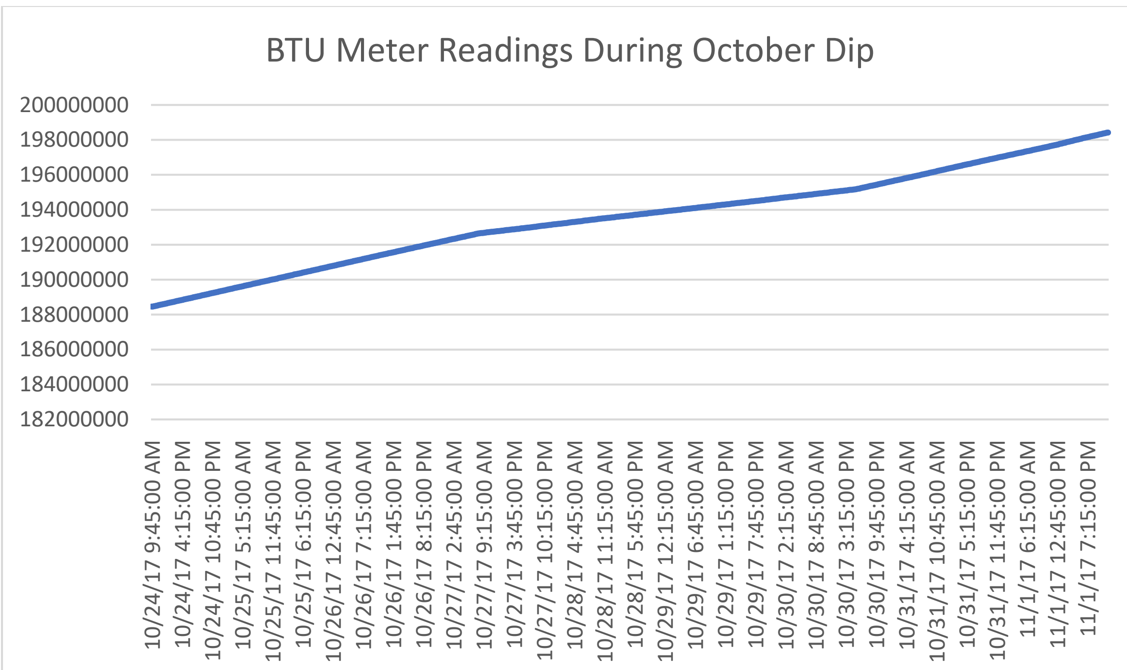

Figure 25. Meter Readings Showing an Unexplained Decrease in the Rate of Energy Removal by the Direct-to-Chip Liquid Cooling Technology System Over Several Days in late October 2017.

Note: The two days (2/19, 2/20) of post penguin retrofit data the project team have use 381 kWh and 392 kWh (lower than the trend in Figure 24).

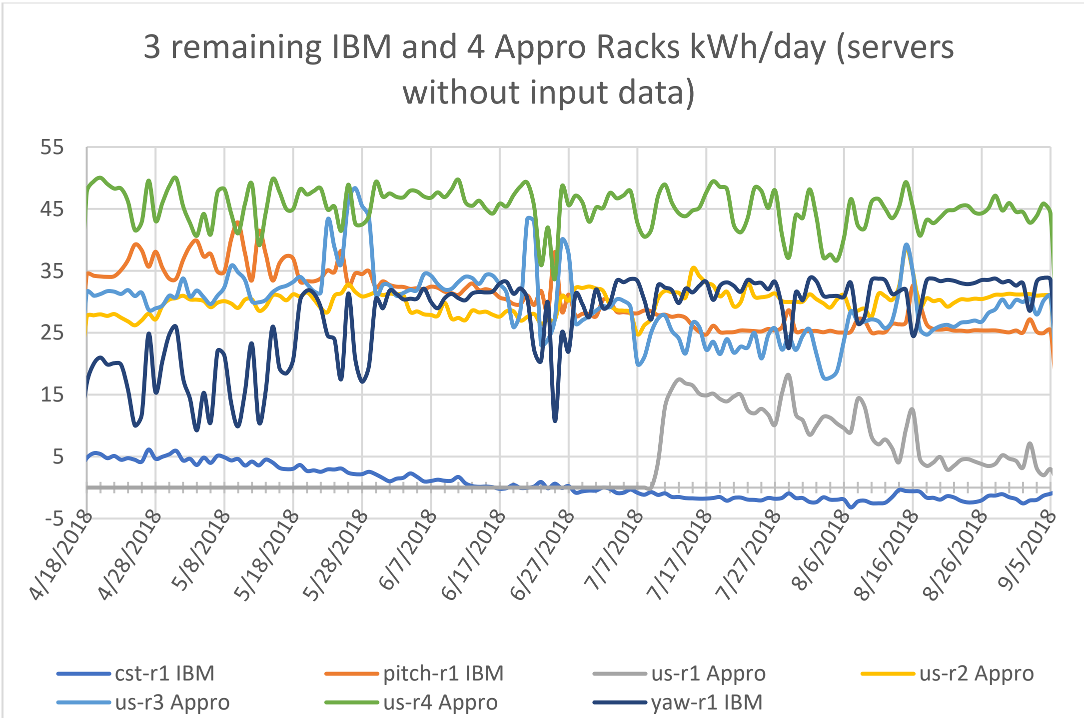

Figure 26. **Energy Removed from Select Servers for Which Energy Input Data Was Not Available**

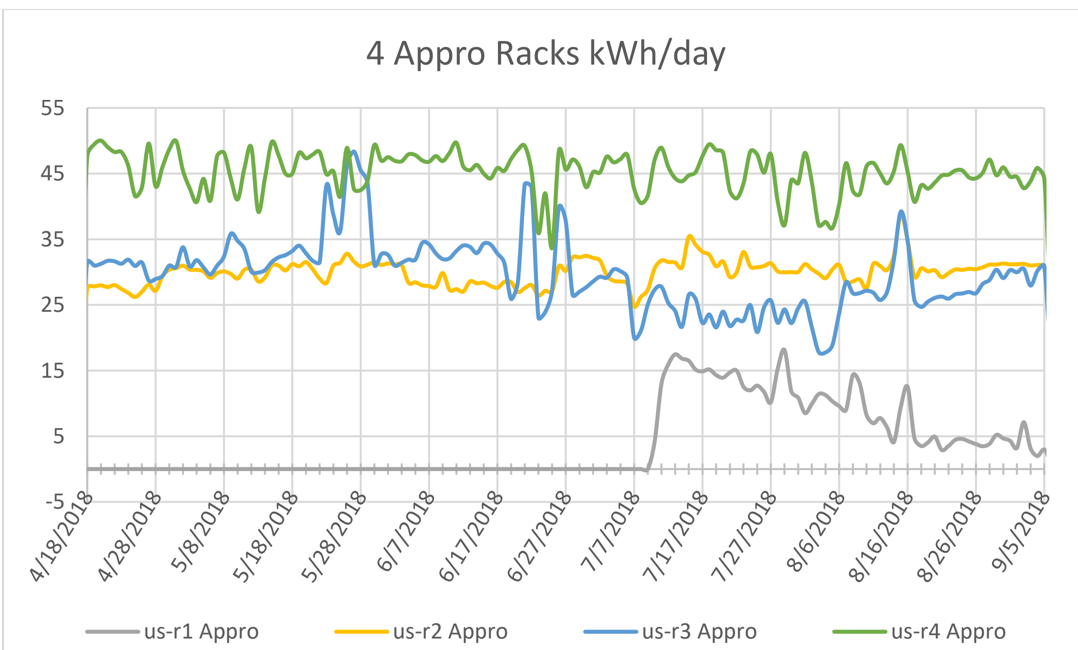

Figure 27. **Energy Removed from the 4 Appro Servers (based on data available)**

[i] "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)

[ii] "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)

[iii] "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)

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