Global Experiences with HPC Operational Data Measurement, Collection and Analysis

Michael Ott*, Woong Shin†, Norman Bourassa‡, Torsten Wilde§, Stefan Ceballos†, Melissa Romanus‡, Natalie Bates¶,
*Leibniz Supercomputing Centre, michael.ott@lrz.de
†Oak Ridge National Laboratory, {shinw, ceballossl}@ornl.gov
‡Lawrence Berkeley National Laboratory, {njbourassa, mromanus}@lbl.gov
§Hewlett Packard Enterprise, wilde@hpe.com
¶Energy Efficient HPC Working Group, natalie.jean.bates@gmail.com

***Abstract*—As we move into the exascale era, supercomputers grow larger, denser, more heterogeneous, and ever more complex. Operating such machines reliably and efficiently requires deep insight into the operational parameters of the machine itself as well as its supporting infrastructure. To fulfill this need, early adopter sites have started the development and deployment of Operational Data Analytics (ODA) frameworks allowing the continuous monitoring, archiving, and analysis of near real-time performance data from the machine and infrastructure levels, providing immediately actionable information for multiple operational uses.**

To understand their ODA goals, requirements, and use cases, we have conducted a survey among eight early adopter sites from the US, Europe, and Japan that operate top 50 high-performance computing systems. We have assessed the technologies leveraged to build their ODA frameworks, identified use cases and other push and pull factors that drive the sites' ODA activities, and report on their operational lessons.

***Index Terms*—exascale, Top500, HPC operations, energy efficiency, site survey, operational data**

I. Introduction

Next-generation high-performance computing systems will achieve unprecedented compute and I/O performance. However, achieving extreme scales requires the orchestration of a correspondingly extreme number of components that comprise the system and push the limits of current silicon technology (i.e., power & energy, thermal, and physical area). As a result, emerging systems are more dynamic, heterogeneous, and complex than ever before. Understanding system behavior in this complex environment requires unorthodox cross-domain operational considerations that take into account not only the HPC system itself but also the surrounding data center as a whole [1].

Fulfilling the operational demand of these emerging systems and their data centers introduces new operational challenges. Historically, monitoring and alerting of HPC systems has been used to identify potential problems or unexpected behaviors. However, as the number of components comprising these systems increases toward exascale, the volume and velocity of sensor data as well as potential sources of failure and their ripple-effects among interconnected components makes traditional methods of monitoring infeasible. In addition, the impact of system issues on the data center as a whole is not captured. Even if this data is collected, translating it into actionable knowledge – let alone finding an optimal solution – in near real-time is a significant challenge. Addressing these challenges requires consideration of both the technical aspects of processing such an amount of data as well as resolving the cognitive burden in making use of the data.

Early efforts aimed at addressing these challenges have centered on the creation of operational frameworks that are capable of integrating loosely coupled technologies, resources, and data. To this end, early adopter HPC sites are developing and deploying new technologies that allow for collecting, consolidating, and analyzing operational data from both the facility and the HPC system into a unified framework [2]–[4].

This paper details the frameworks and experiences of sites with large-scale deployments of HPC operational data collection, consolidation, and analytics technology that span both the facility and compute systems, correspondingly providing a holistic operational view of the HPC data center. As members of the Energy Efficient HPC Working Group (EEHPCWG) with a special interest in such a multi-aspect operational data oriented activity, we as authors formed a team on **Operational Data Analytics (ODA)** to pursue the share of experience in the architecture, challenges, impact, and future of the supporting systems or frameworks. To achieve this goal, we have conducted a global survey performing qualitative interviews of multiple sites from the United States, Europe, and Japan, and performed a synthesis of the results with a cross-functional team of experts from these sites as well as the vendor and academic communities.

As a result, we identified a focus group of eight sites among the top 50 HPC sites [5] to provide a deep dive into the technical details and experiences of their ODA frameworks. Technical details revealed that large-scale deployments with high-volume and high-frequency data are technically feasible and are possible with readily available tools especially from the open source community. Sites that deployed such ODA frameworks were able to develop use cases that benefit their operations with cost savings, also leading to new use cases that were not anticipated before. Yet, there were still areas of exploration in increasing the level of the data aggregation between the facility and the HPC system. Sites with a high degree of aggregation, however, projected additional future value in such investments. Also, analysis of data still involves human inspection where many sites have expressed plans in employing machine learning (ML) approaches in their operations. Though, those ML approaches are mostly in the research phase at the moment.

II. Background

A. Complexity of Operating (pre-)Exascale HPC Systems

As machine densities continue to increase, air cooling is being phased out and replaced by (direct-)liquid cooling, leading to a tighter coupling between HPC systems and their cooling infrastructure. At the same time, many performance characteristics of the machines become more and more dynamic and highly dependent on the application workload running on the machine. Accounting for this dynamic behavior, HPC sites are starting to over-provision their machine installations beyond the limits that their cooling infrastructures and power distribution can sustain. Although this is possible because the average load will be significantly lower than the theoretical peak load, it requires high-frequency monitoring and control loops to keep the system within safe operating conditions. At the same time, the power draw of (pre-)exascale machines continues to grow, making energy efficiency a key issue for future HPC deployments.

B. Demand for High-fidelity Holistic Monitoring

With the increasing complexity, monitoring capability that gives a holistic view becomes ever more important for HPC operations. Status quo solutions that separate compute and infrastructure data streams may no longer be suitable for future deployments since they cannot provide a holistic overview of all operational parameters. This wide range of parameters — building infrastructure systems, HPC system hardware, operating system, and runtime environments to name a few — becomes much more useful and easier to analyze if it is unified in a single monitoring framework. Consequently, monitoring is becoming more and more complex and poses significant challenges in terms of scalability and near real-time data access requirements. Particularly, an often desired high sampling interval of sensors and a large number of monitored components found in current and future HPC installations are driving the need for development of scalable monitoring solutions.

C. Dealing with the Data: Operational Data Analytics (ODA)

Holistic monitoring is only a prerequisite for another important challenge in state-of-the-art HPC operations: Operational Data Analytics (ODA). The goal of ODA is to find optimal operational parameters for the HPC system and its supporting infrastructure that allow for the most computationally and energy-efficient operation [6]. While collecting all operational parameters is mainly a technical challenge, making use of all the data is a cognitive challenge. The sheer amount of different parameters found in large-scale HPC systems make for an almost unsolvable optimization problem. The latest trends in the big data and machine learning research areas may offer solutions to make use of such a wealth of data [7], [8]. In the meantime, having all data in a single unified framework allows for much easier data visualization and correlation tasking, which rapidly helps in understanding the dependencies of parameters and hence makes operational optimization easier.

TABLE I
Participant sites labeled with country of residence

Group[1]	Participant sites	Country
Seed	Los Alamos National Laboratory (LANL)	United States
Seed	Lawrence Berkeley National Laboratory (LBNL)	United States
Seed	Lawrence Livermore National Laboratory (LLNL)	United States
Seed	National Renewable Energy Laboratory (NREL)	United States
Respondent	Interuniversity Consortium of High Performance Systems (CINECA)	Italy
Respondent	Leibniz Supercomputing Centre (LRZ)	Germany
Respondent	National Center for Atmospheric Research (NCAR)	United States
Respondent	Oak Ridge National Laboratory (ORNL)	United States
Respondent	The Institute of Physical and Chemical Research (RIKEN)	Japan

[1] Seed indicates participation through prior contact for the purpose of developing the survey. Respondent indicates the participation through reaching out via the developed survey. Both groups were treated equally during the survey and analysis.

III. Methodology

This paper explores the experiences of sites with large-scale deployments of HPC operational data collection, consolidation, and analytics technology that spans both the facility and the HPC system. The methodology for collecting the data that captures these experiences was qualitative interviews of multiple sites from the United States, Europe, and Japan. The synthesis and analysis of this data were done by a cross-functional team of experts from these sites as well as the vendor and academic community. This section describes the objectives of the survey, the methodology used for the selection of the sites, the development of the questions, the mechanics of the interviews, and the process used for synthesis and analysis of the data collected.

A. Objectives

The primary objective of this survey is to help understand why some large-scale supercomputing centers [5] are interested in collecting, merging, and analyzing data that comes from two domains: both the facility and the HPC system. This survey attempts to understand the motivation behind these efforts and any use cases or benefits that might have been derived. There are two other objectives for the survey: to address potential issues of scalability and data source interoperability.

It is outside of the scope of this survey to understand the motivation and use cases for those sites that collect and analyze these domains independently, even though they may use them very effectively.

B. Survey and Execution

Based on these objectives, a global survey was designed with an initial seed group of four sites the authors had contact with through the EEHPCWG (Table I). The seed sites had prior experience in deploying systems that aggregate facility and HPC system data. Three of the seed sites were top 10 [5] sized supercomputing centers in the United States (Lawrence Livermore National Laboratory (LLNL), Lawrence Berkeley National Laboratory (LBNL), Los Alamos National Laboratory (LANL)) and the fourth was the National Renewable Energy Laboratory (NREL). NREL has been an innovator with ODA in their HPC data center. Their deployment, however, is more for research and demonstration than in support of large scale production computing. Therefore, NREL was a critical seed site but was not included in the global site survey.

With the result of this initial design phase, the survey was conducted during a one year period from September 2018 to September 2019 in two steps: first with a brief cursory survey to identify a focus group, and second, an in-depth survey on the focus group.

To identify the focus group among the sites, the survey employed a set of cursory questions. The questions were designed to reflect our focus on the aggregation of the facility and the HPC system data. The sites were asked whether they collect and integrate facility and HPC system data, and also would be interested in participating in an in-depth survey. From the Top500 [5] list, we have reached out with the cursory survey to 20 sites we had contact with that are ranked top 50. 14 sites responded to the cursory survey out of which five sites expressed intent to go further. As a result, we were able to identify a focus group of total eight sites with three sites from the seed group and the additional five sites that expressed interest in going further (Table I). The sites in the focus group participated in the survey by making responses to the questionnaire, and by responding to a teleconference interview which was later transcribed by the survey team. The sites also gave consent to the team to publish the result. Later, the results were analyzed by the team.

1) In-depth Survey Questions: For the in-depth survey, 15 questions were prepared (Table II) that are aimed at further exploring how the sites execute data collection and aggregation. The questions were designed to gain understanding about the tools in use, the data being collected and how they are aggregated and used, and additional challenges and plans. The in-depth questionnaire was originally developed during the interactions with the seed group. Later during the survey phase, the responses of the initial seed group were distributed as examples to help the participants make better responses. Though in analysis, all responses were treated equally.

First, there were two questions with a specific focus on understanding the tools in use. We asked what tools were used for their data aggregation systems (Q1). On top of this, we also asked about the scalability of these tools (Q14).

TABLE II
IN-DEPTH SURVEY QUESTIONS

Q1)	What tools do you use for your data aggregation system? (database systems, graphing tools, data analytics)
Q2)	How many buildings or site-wide equipment have sensors where you are collecting data for this aggregated data collection system?
Q3)	Can you describe cases where site level data was used and led to a value-add action?
Q4)	What kinds of data do you collect from the HPC data center?
Q5)	Can you describe cases where data collected at the HPC data center level was used and led to a value-add action?
Q6)	What are you planned next steps for collecting data at the HPC center level?
Q7)	How many HPC platforms have sensors where you are collecting data for this aggregated data collection system?
Q8)	a) What are the platform-level functions for which data is collected? b) In what ways has your site achieved integration of HPC platform data with HPC data center data?
Q9)	Can you describe usage cases where data collected at the HPC platform level was used and led to a value-add action?
Q10)	What are your planned next steps for collecting data at the HPC platform level?
Q11)	How would you describe the accuracy of your sensors?
Q12)	a) Does your site face security or organizational constraints? b) If so, do you have advice for how to address those challenges?
Q13)	Do you have any additional constraints or challenges that you would like to describe?
Q14)	Do you know how scalable are the tools that you use? b) Do you anticipate any issues with scalability, especially in the future with exascale class systems?
Q15)	Is there anything else you'd like to tell us?

Second, we asked about the data aggregation activities (Q2 to Q11). From the perspective of a unified ODA framework each site deployed, the sites were asked to identify facility-related data sources from the site level (Q2) and the HPC data center level (Q4), and also the HPC system data sources from the HPC platform level (Q7, Q8). These questions were followed by the future extensions planned by each site (Q6, Q10). On each level, we asked about the value-add actions or use cases made possible by employing a data source from a particular level (Q3, Q5, Q9).

There were a few additional questions. We asked about the accuracy of the sensors involved (Q11) and about challenges and difficulties that are not necessarily technical but important from an operational perspective (Q13). We concluded the questionnaire with a rather open question (Q15) to facilitate an open discussion during the teleconference interviews.

2) Synthesis: As a result of this outreach, two artifacts were captured from each of the participating sites. One is a questionnaire response which was a written response directly from the site, and the other was a transcript of the in-depth interview conducted by the team. With these artifacts, synthesis and analysis were performed by a cross-functional team of experts from the focus group sites as well as the vendor and academic community. Reflecting the goal in Section III-A, the focus was on understanding the details and the value of aggregating facility and HPC system data. With this focus,

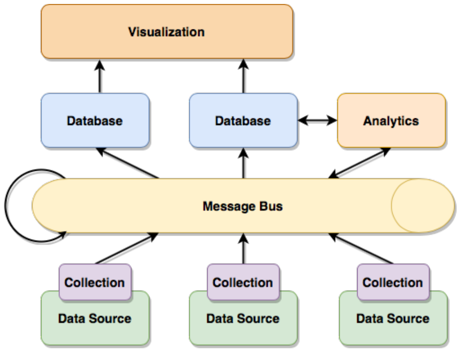

Fig. 1. Schematic representation of common architecture components

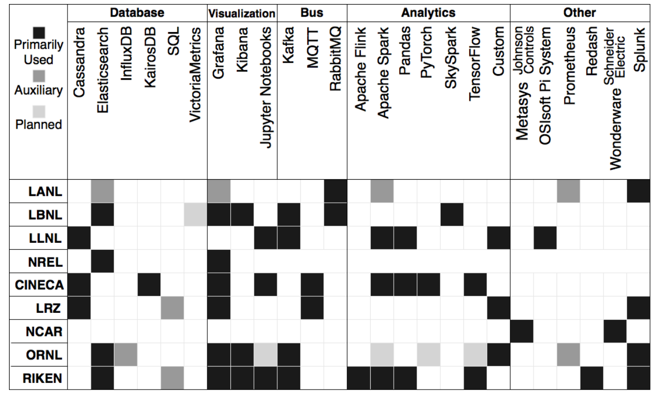

Fig. 2. Matrix of tools in use grouped based on the features and usage each site reported to use (site order based on Table I)

the team has performed qualitative analysis on the data based on the framework embedded in the in-depth questionnaire (Section III-B1).

With this framework, members of the team tackled synthesis from the angles of technical details, use cases, and reflection & projections. On the technical details, the team was able to find a certain architectural trend across the sites (Section IV). For the use cases, several value-added actions and usage scenarios that are based on data aggregation were identified (Section V). Further, the team established key trends from the open questions made in the questionnaire on reflecting the process, experience, and discussing the future projections (Section VI).

IV. Architecture

The tools and frameworks used to implement an ODA infrastructure are diverse among the individual sites surveyed. However, analysis of the functionalities and features of these infrastructures identified a number of similarities in their overall architectures. A generic representation of such an architecture is depicted in Figure 1: data collection agents gather telemetry data at their source and stream it to some sort of bus or message broker that distributes it to consumers. Consumed data could be forwarded to a database or be transformed and produced back into the bus for further consumption. In all implementations, one of the consumers is a — typically NoSQL — database that persistently stores the telemetry data. Others are tools for visualization and analytics. The latter two do not necessarily need to hook into the bus directly to acquire the data but might as well query it from the database. This mostly depends on the use case whether live or historical data is required.

It would be beyond the scope of this paper to describe the actual implementations of this generic architecture in more detail. However, many of the surveyed sites have published scientific papers or given presentations on their implementations we would like to reference here: CINECA's Exascale Monitoring Framework for HPC (ExaMon) [2], LBNL's Operations Monitoring and Notification Infrastructure (OMNI) [3], LLNL's SONAR [9], LRZ's Data Center Data Base (DCDB) [10], and RIKEN [11].

A. Data Collection

In many cases, the data collection agents are proprietary or self-developed software that are tailor-made to the specific requirements of the site. Depending on the sources of telemetry data, they might need to provide the glue code to connect to a multitude of different devices via various protocols and APIs. Implementing those data collection agents can be labor intensive, potentially an area of improvement. The other components in the architecture can usually be picked off the shelf and only need to be configured and deployed. Occasionally, they will have to be adapted to meet the particular needs of a site, but as most of them are open source, this can be easily done.

While there is quite a variety of off the shelf tools to pick from for particular tasks, some are more popular than others among the surveyed sites. Figure 2 gives an overview of the most commonly used tools for the individual components of the architectures among the different sites. The classification of the tools was not always straight-forward as many tools provide features from different categories. In those cases, we have grouped the respective tool based on the features a particular site reported to use. A more detailed discussion on each component follows in the next subsections.

B. Database

There is a clear trend in ODA away from relational databases such as MySQL, MariaDB, and PostgreSQL, toward NoSQL, time series databases, such as InfluxDB and Cassandra, as well as document stores, such as Elasticsearch, which is often also used as a time series database. This makes sense as the semantics of telemetry data do not require relational databases: typically, the data to be stored or queried is time series data. Each entry for a sensor is a tuple of a timestamp and the actual sensor reading. Data is mostly appended and hardly inserted somewhere in the middle of a time series; it is usually consumed in a streaming manner where all data points for a set of sensors over a particular time frame are requested simultaneously. While such data access patterns can also be handled by relational databases, most NoSQL

TABLE III
USE CASES IDENTIFIED IN SITE RESPONSES

	Facility infrastructure optimization	Power capping	Electrical power prediction	Electrical power cost reduction	Design next data center	Use of ML or AI for data analysis
Count	5	1	6	3	4	2
Sites[2]	**LBNL**, LLNL, LRZ, **NCAR**, **ORNL**	**RIKEN**	CINECA, LANL, **LLNL**, LRZ, NREL, ORNL	LLNL, **LRZ**, RIKEN	**CINECA**, **LBNL**, LANL, ORNL	CINECA, LRZ

[2] Bold site names indicate that a use case in the corresponding category from the respective site is presented in more detail in this SectionV

databases are much more efficient at it as they are designed for such access patterns. They are typically more efficient at storing the data on disk in terms of storage overhead as well. On the other hand, NoSQL databases do not provide such a rich query language as their SQL counterparts and some more sophisticated operations such as joins have to be implemented in the application logic or at the API level. Another benefit of NoSQL databases is that they can be easily scaled with a distributed design to accomodate the required storage space needs as well as data ingestion or retrieval rates. While relational databases can show performance decreases with hundreds of gigabytes worth of data, NoSQL databases can scale to petabytes, as seen from some of the survey participants, such as ORNL's Elasticsearch instance.

C. Visualization

When it comes to visualizing the collected data, there is a clear preference towards Grafana. Although Kibana is tightly integrated with the Elastic Stack (Elasticsearch)which many sites are using in their ODA framework, many of them prefer Grafana for visualization, either additionally to Kibana or exclusively. For sites that use a different database back end, Grafana seems to be the natural choice due to its ability to connect to a multitude of data sources thanks to its plugin architecture. While both, Kibana and Grafana, are dedicated visualization tools, Jupyter Notebook is a very universal tool to work with data. Among its many features it allows for plotting data series and hence is often used for visualization tasks; it also provides a rich programming interface that allows for interactively analyzing the data. Some sites are also leveraging Splunk, a commercial tool mostly for log file analysis which offers a full data platform, including interactive visualizations, for this task.

D. Bus

With an already limited amount of message buses available to use in ODA, the available options are implemented in different ways and the use cases needed may dictate which message bus implementation is selected. Current message buses are typically deployed in a cluster; this distributed approach allows for fault tolerance and scalability. The two prevalent implementations for general-purpose message systems either use the "traditional" smart broker/dumb consumer approach, as implemented in RabbitMQ and MQTT, or the dumb broker/smart consumer approach, as in Kafka. In a smart broker/dumb consumer system, the broker keeps track of messages delivered to consumers and dequeues after acknowledgement from consumers that the message has been received. In a dumb broker/smart consumer system, the messages are retained for a set period and the consumer is responsible for maintaining their state; this allows for the ability to consume past messages from an earlier state.

E. Analytics

While the purpose of building an ODA framework is to analyze the operational data, we noticed in the responses to the questionnaire and interviews from many sites that the "analytics" capabilities are currently mainly performed by visual inspection. Sites that are working with data analytics tools such as Apache Spark, Apache Flink, or TensorFlow, to name but a few, are currently in the evaluation or prototyping stage and not using such tools in production yet. The driving factors in selecting which data analytics tools to use include integration with existing tools in the ODA framework, as well as community support.

V. USE CASES

All interviewed HPC sites had made a committed investment in ODA and could describe value from specific use cases. We asked the sites to describe cases where data led to a value-add action, but we did not ask them to make a comprehensive list of all of their cases. As such, the cases we heard about were probably most noteworthy to the site. They probably reflect cases where the value-add was recognized and clear. Most of them are work in progress. The reported use cases can be grouped into three main categories:

- Optimizing facility infrastructure
- Managing power and energy
- Improving strategic planning

The majority of use cases focus on managing power and energy. In this category we were able to identify three further sub-categories on power capping, power prediction, and cost reduction. A detailed breakdown of the number of use cases in each (sub-) category as well as the sites reporting them can be found in Table III. Two of those use cases use artificial intelligence for the analysis of the data. Although they are still in an early development stage, we point them out in a separate column.

The integration of facility and HPC system data is in very early stages of deployment and the use cases are limited. Though, all of the sites had plans to expand the ODA capability. Currently applied use cases mainly looked into reducing

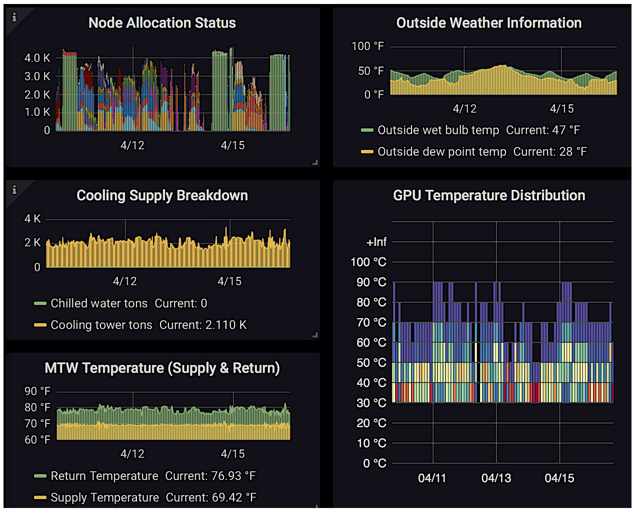

Fig. 3. Near real-time datacenter and HPC platform integrated dashboard for medium temperature water (MTW) cooling operations (ORNL)

OPEX by, either increasing the infrastructure efficiency or adhering to power contract requirements. It is interesting to note that all use cases were site-specific; no two use cases were identical across the sites.

Forward looking, we believe that connecting real-time IT information with the facility data will make the facility infrastructure more efficient.

The following sub-sections will discuss some representative use cases for each of the three main categories.

A. Optimizing Facility Infrastructure

1) NCAR: The NCAR-Wyoming Supercomputing Center (NWSC) at National Center for Atmospheric Research (NCAR) has aggregated data from three different systems: site-wide electrical systems, facility mechanical systems, and HPC system Cooling Distribution Units (CDUs).

They found during small brownout conditions, that the HPC system would not go offline, but the Variable Frequency Drive (VFD) of the CDU pumps would fail and trigger a turnover between the two redundant pumps in the CDUs. This failure would occasionally cause over-temperature conditions in the HPC system's secondary cooling loop.

Through integrated data and analysis, they were able to correlate site-wide power quality issues with VFD failures. The NWSC staff then transitioned the CDU power from the direct utility to UPS backed up power. This enhancement gave the CDUs clean power, so they were no longer exposed to the brownout conditions as before. NWSC could avoid multiple incidents (around two per year) where the brownout would have affected the CDUs. The cost per incident is roughly estimated to be \$20k.

2) ORNL: The Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL) extends data aggregation to the HPC system component (GPU, CPU)

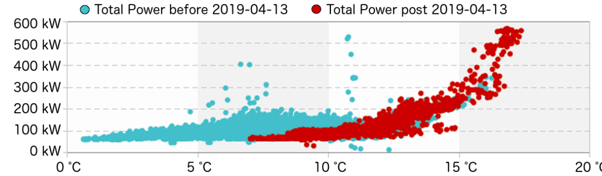

Fig. 4. Scatter plot from LBNL's OMNI system showing the power draw of the cooling plant at different outside air temperatures before (turquoise) and after (red) optimization

level to increase efficiency of their medium temperature water cooling controls for the Summit [3] system.

To maximize cooling efficiency, OLCF leverages medium temperature water (MTW) in the secondary loop for the Summit system. MTW minimizes the use of chilled water by enabling the use of water-side economizers based on evaporative cooling when the weather conditions are advantageous. To further reduce the use of chilled water, especially when the economizers are not efficient, OLCF uses the aggregated data to optimize the cooling water temperatures and flow rates. GPU and CPU temperatures are kept just under the threshold of throttling to help ORNL to increase the amount of free cooling they can utilize even on days with adverse wet bulb temperatures and hence improve the site's energy efficiency and OPEX.

To enable such optimization, OLCF uses a holistic visualization of the data center depicted in Figure 3. Data center level electrical and mechanical data is aggregated with power and temperature data emitted from individual nodes and is processed, summarized, and rendered in near real-time. Given the node allocation and outside weather conditions, notable data center parameters such as MTW supply & return temperature and MTW flow are cross-checked with the histogram-based component-wise temperature distribution summary of the HPC platform (27,756 GPUs and 9,252 CPUs).

3) LBNL: In 2015, the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory (LBNL) moved into a new facility that does not use compressor-based cooling systems. It became apparent that a different level of systems and environmental instrumentation would be needed to optimize operations. As a result, NERSC developed the Operations Monitoring and Notification Infrastructure (OMNI) [3] system, a versatile platform of applications that combines a vast amount of HPC and IT systems data with comprehensive cooling and facility systems performance data.

NERSC leveraged the OMNI system to optimize the efficiency of their cooling water plant at lower outside air temperature conditions by increasing the minimum tower water supply temperature setpoint. Figure 4 shows an example of an automated and real-time generated scatter plot analysis from OMNI. The graph shows how the total power reduced significantly at lower outside air temperatures but remained

[3] https://www.olcf.ornl.gov/olcf-resources/compute-systems/summit

similar at higher temperatures after changing the setpoint. This change allowed the cooling tower fans to ramp down at low temperatures without making an impact on pump energy. The real-time feedback of the plot showed the ODA team that increasing the minimum setpoint was a step in the right direction. NERSC was then able to increase the setpoint until no further improvements were detected and thereby improving the energy efficiency of the cooling infrastructure. This individual optimization task is among several other cooling plant settings improvements that used the same process, producing a total estimated $23k of annual energy cost savings for cooling plant operations.

B. Managing Power and Energy

1) RIKEN: For the Institute of Physical and Chemical Research (RIKEN), effective management of power usage is of utmost importance due to a contractual agreement between the supercomputing facility and the power company. Exceeding the upper limit of this agreement results in costly penalty fees. To mitigate this possibility, RIKEN must monitor the power usage of the facility as a whole together with the estimated power consumption of upcoming scheduled jobs on the supercomputer.

Users wishing to run large-scale jobs must first execute a small version of their job (~10% of the system) to obtain a power consumption profile. Aggregating this application data together with facility data, RIKEN can estimate the power consumption of the full-scale job and determine whether to (1) schedule it at a time when other consumption is low, (2) lower the power draw from the grid by starting a backup gas-powered generator, or (3) determine that it cannot be executed within the facility's current constraints. If a running job overruns its estimate and the facility power draw approaches the upper limit of their allowed power consumption, RIKEN's monitoring system automatically detects and kills the job, thereby avoiding utility penalties.

2) LRZ: The Leibniz Supercomputing Centre (LRZ) has deployed a scheme called "Energy Aware Scheduling" (EAS) to reduce an application's power draw without hurting application performance or runtime. It leverages dynamic voltage and frequency scaling (DVFS) to reduce the operating frequency of the CPUs the application is running on and consequently their power draw. EAS employs CPU performance counters to determine the characteristics of the application and to asses whether it is memory or compute bound. Applications that are memory bound run at a lower frequency as they would not be able to benefit from higher CPU frequencies since their performance is limited by the performance of the memory subsystem. Applications that are compute bound and hence can benefit from higher CPU frequencies, on the other hand may run at sticker frequency. Taking into account the application characteristics ensures that application performance does not degrade due to lower CPU frequencies which would have a detrimental effect on the application's energy to solution, i.e., the product of its runtime and the average power consumption. LRZ estimates the cost savings due to EAS at 1.8M€ for lifetime of its previous flagship system SuperMUC.

3) LLNL: The Lawrence Livermore National Lab's power draw shows fluctuations and their electricity provider asked them to provide forecasts when the site's total power consumption goes up or down by 750 KW over a 15 minutes window. As LLNL monitors and archives not only the power consumption of its HPC systems but also of the data center and the whole site, they were able to analyze three years worth of data to identify power spike patterns. Using Fourier transformation on the monitoring data, LLNL determined that over 50% of large power spikes occur between two hours in the morning and one hour in the evening. These are attributable to employee schedules. They also found seasonal and weekday patterns. The remaining spikes were not periodic and can be attributed to both scheduled and random events. Interestingly, only 8% of the spikes were due to load changes on the HPC system, and another 8% were caused by experiments of the National Ignition Facility on LLNL's site.

C. Improving Strategic Planning

1) CINECA: For the Interuniversity Consortium of High Performance Systems (CINECA), the investment in developing power management and monitoring technology has been beneficial for designing and planning power and cooling upgrades. The use of historical power measurement data changed their decision-making process when it comes to defining new power and cooling requirements changing their traditional approach. In the past, CINECA designed and planned the power distribution and cooling infrastructure based on the datasheet of the HPC system. Now the decision process instead leverages historical data collected by the University of Bologna. This data is used to make informed decisions based on a cost-benefit analysis. This process was used for the last two partitions of the current flagship system GALILEO. By analyzing the collected data, models of the facility cooling system under different possible designs were developed. Based on those models they decided to increase the free-cooling capacity in the data center rather than adopting an approach based on rear door heat exchangers (see [12] for more details). The decisions are now more focused on the specific data center conditions such as climate data or non-uniformity inside the data room.

2) LBNL: In preparation for the delivery of a new pre-exascale sized system in late 2020 named Perlmutter [4], NERSC leveraged ODA capabilities of their OMNI system to plan for an expansion to the electrical supply for infrastructure systems.

During planning for the new mechanical systems that would need to support Perlmutter and the current system Cori, a routine capacity study was conducted calculating the theoretical peak load with all components working at full load. The study recommended the addition of a second mechanical systems electrical substation in order to support Perlmutter. Fortunately, the LBNL Facilities Master Specification permits

[4] https://www.nersc.gov/systems/perlmutter

a secondary calculation method for mechanical load planning, if at least one full year of operational power metering data is available. The NERSC OMNI archives contained more than the needed data and the subsequent analysis determined that the operational power draw did not exceed 60% of the total power rating of the existing substation. Therefore, the analysis showed an additional substation would not be needed.

Since the elimination of an entire electrical substation is a bold step for the large Perlmutter project, further analysis looked at the number of expected peak cooling hours under warm conditions which could push mechanical power demand above the 1 MW maximum rating of the current substation. The data showed that the mild Berkeley climate had a very small number of days annually where the max substation rating would be stressed. Ultimately, this ODA analysis enabled the LBNL Facilities team to obtain confidence that the increased Perlmutter electrical load would not require a new mechanical infrastructure substation, thereby saving NERSC about $2M of the project cost.

VI. Global Experiences

A. Pull & Push Factors

There are many reasons why sites are developing ODA capabilities: most often such activities go along with the objective to improve the energy efficiency of their HPC operations and cooling infrastructures, either because of environmental or budgetary reasons. Most interestingly, sites that did not respond to the survey because they do not see the need for data aggregation between their facility and HPC systems also seem to not have energy efficiency related constraints or mandates. Some sites have power limits imposed by either their utility or their own infrastructure and require such holistic and detailed monitoring and analysis capabilities to safely operate their HPC systems. Other sites develop those frameworks because they have a genuine interest in pushing science and want to get a deeper insight into the operational parameters of their HPC systems and their facility. Many of them derive such research goals directly from their mission statement.

At many sites, the activities to aggregate the data collection of the facility and the HPC system seem to be driven by facility engineers and/or researchers, while only a few sites have system administrators leading the efforts at their sites.

B. Operational Lessons & System Adjustments

One very interesting observation from the questionnaire and the interviews was a "build it and they will come" experience among many sites. Although they had certain use cases and users for the collected data in mind when they deployed it, they found further use cases as well as users once the data became available.

Another common theme among the site respondents that had deployed NoSQL databases for their monitoring was the fact that they see additional benefit in being able to scale those databases horizontally and add almost unlimited storage capacity. This allows for higher frequency data, more monitored components, and longer data retention time. Particularly the ability to maintain long term monitoring history (over many seasons and years) is highly useful for optimizing cooling infrastructures.

While most sites recognize that going back to historical data may be useful, not all sites keep all the data on-line forever. Some sites archive the data so they can retrieve it at a later point in time if need be. Others thin out their monitoring data as it gets older, for example, by storing only averages over a certain period of time instead of the raw data points.

C. Future Plans

ODA is in an early stage. The early adopter sites continue to develop their capabilities and constantly add new features and data sources to their frameworks. We already observed this dynamic when we conducted the interviews with the sites approximately half a year after the questionnaire and noticed that they reported additional features and capabilities in their frameworks.

Most development and research at the sites centers around the use of machine learning to analyze the data, automatic control feedback loops, and tighter integration of application performance data.

1) Machine Learning: As pointed out in Section IV-E, data analytics in ODA currently mostly takes place via visual inspection. However, most of the sites in the survey mentioned that they were working on machine learning in particular. This makes sense given the amount of data those sites are collecting and the inter-dependencies between different operational parameters that are not always obvious and hard to discover manually.

For example, LRZ is working on the "Infrastructure Data Analyzer and Forecaster (IDAF)", a machine learning based framework, that allows for forecasting various data center energy/power consumption relevant key performance indicators (KPIs) [13]. It is capable of assessing the modification-impact of relevant operational knobs (e.g., change in the number of currently active cooling towers, load of the deployed HPC system, etc.) on various data-center level KPIs ranging from cooling efficiency to the delivered system inlet temperature.

2) Automatic Control Feedback Loop: The holy grail of ODA appears to be a feedback loop that automatically analyzes the data and uses this information to set operational parameters. The most obvious application of such an approach seems to be the control loops of the cooling infrastructure, where setpoints could be modified depending on system load and weather conditions. But some sites are also thinking in the opposite direction, such as influencing batch scheduling decisions based on data from the supporting infrastructure such as available cooling or electrical capacity.

For example, ORNL has expressed plans to extend their ODA deployment to implement predictive power analysis to automatically adjust cooling parameters. Such automation will enable the cooling plant to closely follow the power consumption envelope, eventually leading to additional operational cost savings. The approach is based on passively deriving the near-future power consumption from the current set of jobs in the batch scheduler queue. The predictions are expressed to the central energy plant as recommendations, eventually relieving the human operator in the feedback loop making setpoint decisions.

3) Tight Integration with Application Performance Data: Another area of ongoing development is the tighter integration of application performance data within the ODA frameworks. RIKEN and LRZ, are already leveraging such data for their use cases. Others are just adding it because it is rather easy to add additional data sources to a data aggregation system once the basic architecture for data collection is in place. The immediate benefit of also integrating application data as well may not always be obvious but they are basically following the ”build it and they will come” doctrine.

On one of their smaller production HPC systems, CINECA is planning to have their ODA system provide the ability to actively control frequency settings with phases of the application, based on whether or not the phases are compute or communication bound. They have demonstrated this ability under controlled conditions with Quantum Espresso, a scientific application code for electronic-structure calculations and materials modeling.

VII. Related Work

A. Operational Data Analytics

In this work, we adopted the term “operational data analytics” (ODA) as defined by Bourassa et al. [6] due to the proximity of the work with HPC site operational efficiency. Yet, the underlying concept of ODA around analyzing operational data to achieve an operational advantage or efficiency has been discussed and practiced throughout multiple disciplines and domains such as military [14], manufacturing [15], business [16] and information technology [17]. Such concept was loosely referred in the industry as operational intelligence [18], operational analytics [19], [20], IT operational analytics [21] and business intelligence [22].

In recent years, the technological advancements in data science and engineering has increased the ability to extract value from data [7], [23]. Many organizations and businesses are investing in such abilities to enhance the efficiency of their operations [24]. Investments in such capabilities are extended towards artificial intelligence with the recent hype and are starting to gain results [8], [23].

In HPC site operations, an effort to achieve operational efficiency utilizing operational data has been discussed extensively mainly in the context of energy efficiency, driven by the increasing power and energy costs induced by modern data centers [25], [26]. KPIs are defined and measured to enable quantifiable enhancements [27] where holistic data-driven optimizations are identified critical in the process [1]. To fulfill the needs in such data center operations, ODA systems such as Examon [2], OMNI [3], and DCDB Wintermute [4] have been implemented and deployed in HPC site operations.

B. HPC Site Surveys

This work shares the same spirit with the surveys sharing operational experiences within the HPC community. Coles et al. [28], surveyed environmental conditions for all of the United States Department of Energy sites in order to assess the potential for eliminating chillers in HPC data centers. Bates et al. [29], Patki et al. [30], and Clausen et al. [31] also surveys aspects of electrical grid integration for HPC centers in the United States and Europe.

Topically related to ODA, Maiterth et al. [32], [33] conducted a global survey of major supercomputing center’s experiences with energy and power-aware job scheduling and resource management (EPA JSRM). EPA JSRM is a potential use case for ODA systems that merge data streams from both the HPC system and the facility.

VIII. Conclusion

ODA is becoming increasingly important for the operation of flagship supercomputer systems. The complexity, size, heterogeneity, and dynamic load characteristics of current and future (pre-)exascale systems will require deep insight into the operational parameters of those systems. Such insight can only be gained from comprehensive and holistic monitoring and data aggregation frameworks that combine the various data streams from the facility and the HPC system.

Early adopter sites are in the process of developing and operating such holistic monitoring frameworks. Some of them participated in a global survey among top 50 HPC sites and shared the motives, technical details, and experiences with those activities with the HPC community.

The key lessons to be learned from the survey for operators of data centers that plan to deploy their own ODA frameworks can be summarized as follows:

- Software tools to build such telemetry collection, aggregation, and analysis tool chains are readily available, in particular from the open source community.
- Large-scale deployments with high volume and high-frequency data are technically feasible. Scalability and storage space is no longer a limiting factor for such endeavors.
- All sites have use cases where the availability of the monitoring data was beneficial for their operations and led to cost savings.
- The availability of holistic data often leads to further use cases and users of the data that were not anticipated before deployment.

For the ODA community, i.e. sites that are already engaging in ODA, there are two very interesting findings in the survey:

- Analysis of data is still mostly performed by manual inspection. ML approaches are being worked on by many sites but are mostly research projects and not used in operation, yet.
- Use cases that actually exploit data aggregation from the facility and the HPC system are still rare. However, most

sites that invest in such systems strongly believe that they provide additional value.

Those findings suggest that the development of ODA capabilities is still very dynamic. Many sites are performing active research on methods to (automatically) leverage the data they are collecting from their HPC operations. Research in this field is mostly driven by advances in the field of artificial intelligence. Holistic data collection capabilities, that are a prerequisite for ODA, are mostly state-of-the-art by now.

ACKNOWLEDGMENT

We thank the many individuals from the participant sites for their contribution to our survey. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. Also, this work was supported by, and used the resources of, the Oak Ridge Leadership Computing Facility, located in the National Center for Computational Sciences at ORNL, which is managed by UT Battelle, LLC for the U.S. DOE (under the contract No. DE-AC05-00OR22725).

REFERENCES

[1] T. Wilde, A. Auweter, and H. Shoukourian, "The 4 Pillar Framework for energy efficient HPC data centers," http://dx.doi.org/10.1007/s00450-013-0244-6, 2014.

[2] F. Beneventi, A. Bartolini, C. Cavazzoni, and L. Benini, "Continuous learning of HPC infrastructure models using big data analytics and in-memory processing tools," in *Design, Automation Test in Europe Conference Exhibition (DATE'17)*, 2017, pp. 1038–1043.

[3] E. Bautista, T. Davis, C. Whitney, M. Romanus, N. Bourassa, N. Bates, G. Koenig, and T. Kubaska, "Collecting, monitoring, and analyzing facility and systems data at the national energy research scientific computing center," in *Energy Efficient HPC State of the Practice Workshop, International Conference on Parallel Processing (ICPP'19)*, 2019.

[4] A. Netti, M. Mueller, C. Guillen, M. Ott, D. Tafani, G. Ozer, and M. Schulz, "DCDB Wintermute: Enabling Online and Holistic Operational Data Analytics on HPC Systems," in *Proceedings of the International Symposium on High-Performance Parallel and Distributed Computing (HPDC)*, 2020.

[5] J. J. Dongarra, H. W. Meuer, and E. Strohmaier. Top500. [Online]. Available: https://www.top500.org/

[6] N. Bourassa, W. Johnson, J. Broughton, D. M. Carter, S. Joy, R. Vitti, and P. Seto, "Operational Data Analytics: Optimizing the national energy research scientific computing center cooling systems," in *48th International Conference on Parallel Processing (ICPP Workshops'19)*, 2019.

[7] McKinsey Global, "The age of analytics: competing in a data-driven world," *McKinsey & Company*, December 2016.

[8] A. Cam, M. Chui, and B. Hall, "Global AI Survey: AI proves its worth, but few scale impact," *McKinsey Analytics (McKinsey & Company)*, November 2019.

[9] K. Shoga, "Monitoring and Data Integration at LLNL," Presentation at the Ninth Annual Workshop for the Energy Efficient HPC Working Group, 2018.

[10] A. Netti, M. Mueller, A. Auweter, C. Guillen, M. Ott, D. Tafani, and M. Schulz, "From facility to application sensor data: Modular, continuous and holistic monitoring with DCDB," in *International Conference for High Performance Computing, Networking and Storage (SC'19)*, 2019.

[11] K. Yamamoto, "Operational Data Processing Pipeline," Presentation at the BoF on Operational Data Analytics at SC'19, 2019.

[12] C. Conficoni, A. Bartolini, A. Tilli, C. Cavazzoni, and L. Benini, "Integrated Energy-Aware Management of Supercomputer Hybrid Cooling Systems," *IEEE Transactions on Industrial Informatics*, vol. 12, p. 1299–1311, 2016.

[13] H. Shoukourian, T. Wilde, D. Labrenz, and A. Bode, "Using machine learning for data center cooling infrastructure efficiency prediction," in *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2017, pp. 954–963.

[14] "Tactical operational intelligence," in *Department of Defense Dictionary of Military and Associated Terms (as amended through 9 June 2004) (1-02, US DoD Joint Publication)*. Department of Defense, 2001, p. 528.

[15] J. A. Harding, M. Shahbaz, Srinivas, and A. Kusiak, "Data Mining in Manufacturing: A Review," *Journal of Manufacturing Science and Engineering*, vol. 128, no. 4, pp. 969–976, nov 2006.

[16] S. Radeschütz, B. Mitschang, and F. Leymann, "Matching of process data and operational data for a deep business analysis," in *Enterprise Interoperability III*, K. Mertins, R. Ruggaber, K. Popplewell, and X. Xu, Eds. Springer London, 2008, pp. 171–182.

[17] R. Veneberg, M. Iacob, M. van Sinderen, and L. Bodenstaff, "Enterprise Architecture Intelligence: Combining Enterprise Architecture and Operational Data," in *2014 IEEE 18th International Enterprise Distributed Object Computing Conference*. IEEE, sep 2014, pp. 22–31.

[18] "Operational Intelligence Enters the Spotlight," in *Information Week*, 2006. [Online]. Available: https://www.informationweek.com/software/information-management/operational-intelligence-enters-the-spotlight/d/d-id/1044564?

[19] R. Malluk, "Insight Analytics, The Ultimate Guide to Operational Analytics," in *Insight Analytics*, 2019. [Online]. Available: https://insightsanalytics.com/2019/05/operational-analytics/

[20] A. Böhm, J. Dittrich, N. Mukherjee, I. Pandis, and R. Sen, "Operational analytics data management systems," *Proceedings of the VLDB Endowment PVLDB*, vol. 9, no. 13, pp. 1601–1604, 2016.

[21] "The ITOA Landscape." [Online]. Available: http://www.itoa-landscape.org

[22] "What is Business Intelligence (BI)?" [Online]. Available: https://olap.com/learn-bi-olap/olap-bi-definitions/business-intelligence/

[23] McKinsey Analytics, "Analytics comes of age," *McKinsey & Company*, January 2018.

[24] M. Buluswar, V. Campisi, A. Gupta, Z. Karu, V. Nilson, and R. Sigala, "How companies are using big data and analytics: Interview," *McKinsey Analytics (McKinsey & Company)*, April 2016.

[25] J. Loper and S. Parr, "Alliance to save energy, energy efficiency in data centers: A new policy frontier," http://www.ase.org/resources/energy-efficiency-data-centers-new-policy-frontier, 2007.

[26] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *34th annual international symposium on Computer architecture (ISCA'07)*, 2007, p. 13.

[27] J. N. B. I. P. Council, "DPPE: Holistic framework for data center energy efficiency - kpis for infrastructure, it equipment, operation (and renewable energy)," http://home.jeita.or.jp/greenit-pc/topics/release/pdf/dppe_e_20120824.pdf, Aug 2012.

[28] H. Coles, M. Ellsworth, and D. J. Martinez, "Hot for warm water cooling," in *International Conference for High Performance Computing, Networking and Storage (SC'11)*, 2011.

[29] N. Bates, G. Ghatikar, A. Ghaleb, G. A. Koenig, S. Bhalachandra, M. Sheikhalishahi, T. Patki, B. Rountree, and S. Poole, "The electrical grid and supercomputing centers: An investigative analysis of emerging opportunities and challenges," *Energiinformatik*, vol. 38, pp. 111–127, 2014.

[30] T. Patki, N. Bates, G. Ghatikar, A. Clausen, S. Klingert, G. Abdulla, and M. Sheikhalishah, "Supercomputing centers and electricity service providers: A geographically distributed perspective on demand management in europe and the united states," in *International Conference on High Performance Computing (ISC'16)*, 2016.

[31] A. Clausen, G. A. Koenig, S. Klingert, G. Ghatikar, P. M. Schwartz, and N. Bates, "Supercomputing centers and electricity service providers," in *The 48th International Conference on Parallel Processing (ICPP Workshops'19)*, 2019.

[32] M. Maiterth, G. Koenig, K. Pedretti, S. Jana, N. Bates, A. Borghesi, D. Montoya, A. Bartolini, and M. Puzovic, "Energy and power aware job scheduling and resource management: Global survey initial analysis," in *IEEE International Parallel and Distributed Processing Symposium (IPDPS Workshops'18)*, 2018.

[33] ——, "Energy and power aware job scheduling and resource management: An in-depth analysis," in *Workshop on Data-center Automation, Analytics, and Control (DAAC'18)*, 2018.