A demonstration of monitoring and measuring data centers for energy efficiency using opensource tools

Jonas Gustafsson
RISE SICS North
Luleå, Sweden
jonas.gustafsson@ri.se

Sebastian Fredriksson
RISE SICS North
Luleå, Sweden
sebastian.fredriksson@ri.se

Magnus Nilsson-Mäki
RISE SICS North
Luleå, Sweden
magnus.nilsson-maki@ri.se

Daniel Olsson
RISE SICS North
Luleå, Sweden
daniel.x.olsson@ri.se

Jeffrey Sarkinen
RISE SICS North
Luleå, Sweden
jeffrey.sarkinen@ri.se

Henrik Niska*
Lition AB
Luleå, Sweden
hm@lition.se

Nicolas Seyvet†
OP5
Stockholm, Sweden
nseyvet@op5.com

Tor Björn Minde‡
RISE SICS North
Luleå, Sweden
tor.bjorn.minde@ri.se

Jonathan Summers§
RISE SICS North
Luleå, Sweden
jon.summers@ri.se

ABSTRACT

Data centers are complex systems that require sophisticated operational management approaches to provide the availability of digital services against the backdrop of cost and energy efficiency. To achieve this, data center telemetry data is required since, as is commonly said *it is not possible to manage what cannot be measured.* This paper details how it is possible to construct the key data center infrastructure management (DCIM) elements of monitoring and measuring by a combination of available opensource software tools that permit both scalability and an environment where analytics can be employed on the data center operation, which can offer relevant insight into energy efficient operational practices.

KEYWORDS

Data centers, monitoring, measuring, opensource software

ACM Reference Format:
Jonas Gustafsson, Sebastian Fredriksson, Magnus Nilsson-Mäki, Daniel Olsson, Jeffrey Sarkinen, Henrik Niska, Nicolas Seyvet, Tor Björn Minde, and Jonathan Summers. 2018. A demonstration of monitoring and measuring data centers for energy efficiency using opensource tools. In *e-Energy '18: The Ninth International Conference on Future Energy Systems, June 12–15, 2018, Karlsruhe, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3208903.3213522

*Former employee of RISE SICS North.
†Former employee of Ericsson
‡Position in Ericsson
§Senior Lecturer, School of Mechanical Engineering, University of Leeds, Leeds, UK

e-Energy '18, June 12–15, 2018, Karlsruhe, Germany

ACM ISBN 978-1-4503-5767-8/18/06…$15.00
https://doi.org/10.1145/3208903.3213522

1 INTRODUCTION

Monitoring and measuring the key hardware components within the data center have emerged as the critical elements of the Data Center Infrastructure Management (DCIM) tools that are commercially available. There is no agreed definition of DCIM, but it is perhaps understood as to what it should aspire to be, namely the integrated management of information technology and facility via a converged *monitoring*, *measuring* and capacity planning of critical systems [4]. There are many opensource monitoring tools available, GeekFlare have listed and compared some of them, including Nagios and Zabbix[1]. In the Data Center Knowledge article, [3], the importance of comprehensive information is highlighted as a means of improving data center energy efficiency and is also a strong argument for the return on investment in the employment of monitoring and measuring tools.

There are different models of the goal of a DCIM, but perhaps the one that is relevant to what is presented in this paper is that defined by Gartner, "integrating facets of system management with building management and energy management, with a focus on IT assets and the physical infrastructure needed to support them", [2]. This can be essentially summarised as having three primary steps, namely **input**, **process** and **output**. The importance of an end-to-end approach that includes monitoring both the IT and the supporting infrastructure has been clearly highlighted in the recent paper by Levy and Hallstrom [6]. The output is usually presented in graphical format on a *dashboard* and as a result the input parameters can be adjusted or even the information can be presented as a three dimensional gaming platform based on big data analysis approach [5].

This paper is organised into 3 main sections that covers a description of the data center, the software platform that has been adopted and configured within the data center followed by some operational analysis that demonstrates energy usage considerations.

[1]https://geekflare.com/best-open-source-monitoring-software/

2 THE DATA CENTER PODS

The so-called ICE, Infrastructure and Cloud research & test Environment, is the data center facility that has been developed for a number of privately and publicly funded projects at RISE SICS North. The initiative has more than 60 industry partners with interests from big data/cloud platforms all the way through the data center stack down to energy demand and building physics. The first module became operational by February 2016. Measurement data from equipment and sensors is collected for modeling, simulation and optimization. The server infrastructure of module 1 is based on Dell SmartEdge RX730xd and is enhanced with GPU acceleration. In addition Openstack and Hadoop data storage and processing platforms are offered.

The second module, running since May 2017, is more flexible and lab-like featuring fast connectors for power, water and networking to enable easier exchange of equipment. The server infrastructure is made up primarily with Dell and HP blade servers.

The two main DC modules have both been built by PriorIT System 42 based on their room-in-room system. The two modules are located in a warehouse, and are hence physically a room within a room.

Table 1: List of hardware equipment in module 1.

Equipment	Brand	Type	N.o. units
Racks	Minkels		10
Spine Switch	Dell	s6000	1
T.o.R. Switch	Dell	s4048	9
Management Switch	Dell	1548	10
PDUs	Schleifenbauer	hPDU	20
UPS	Riello	Multi Power	1
Servers	Dell	R730xd	150
Servers	OCP	Windmill	42
Cooling units	SEE Cooling	HDZ-2	4
Switchboard	Schneider		1
Electricity meters	ABB	A44 213	4
Heat meters	Kamstrup	Multical 6M2	2
Env. monitoring	Raritan	EMX2-888	4
Temp. and humidity	Raritan	DPX2-T3H1	20
Air flow sensors	Raritan	DPX-AF1	10

2.1 Module 1

Module 1 is primarily used for IT-related research, such as big data analysis, cloud technology development and as a reference model for e.g. CFD model validation [8]. The general equipment consists of 10 IT racks organized in two rows, see Figure 1, four SEE Cooling High density cooling devices, HDZ-2[2] and a Riello Uninterruptible Power Supply (UPS). It is configured with slab floor cooling, where the cold air from the cooling units is directed towards the racks. The module has dimensions are 5 x 6,5 metres and currently has a door separating the hot and cold aisles, which is not shown in Figure 1.

2.1.1 IT Equipment. Most servers installed in Module 1 are of type R730xd[3] from Dell. The majority are configured with 4 x 4TB 7200 rpm & 2 x 600 GB 10K rpm hard drives, 256GB RAM and two Intel Xeon E5-2620 v3 2.4GHz CPUs. Racks 1-4 & 7-10 are filled with 18 730xd servers. Rack six contains 6 730xd servers, which are used for database and DC management purposes. Rack 5 contains 42 open compute project (OCP) windmill servers[4]. Each rack is equipped with one Dell S4048 Top of Rack (ToR) switch which all connect to a Dell S6000 spine switch creating a 10/40Gbps network architecture, see also Table 1 for a list of equipment found in module 1.

2.1.2 IT Measurements. Data is collected from 144 Dell R730xd servers through the Dell iDRAC interface twice a minute. These servers are mounted in racks 1-4 & 7-10, with each rack containing 18 servers. The power delivered to the IT equipment is monitored through the Schleifenbauer Hybrid Power Distribution Units (hPDU)[5]. These PDUs measure current, voltage, power factor and cumulative wattage on all three input phases and on all the outlets to the servers. All network switches and network traffic are monitored and stored.

2.1.3 Environment and Facility Equipment. As indicated earlier, the DC module is cooled using four SEE cooling HDZ-2 units. These CRAHs are liquid to air heat exchangers that receive their cooling liquid through a liquid to liquid heat exchanger located outside the module. This heat exchanger is in turn connected to a building central cooling loop. The heat rejection and cooling supply is performed by compressor based chillers, which are part of the warehouse and therefore not accessible, see also 2 for a graphical representation of the cooling system. To have full control of the environment in the module, a separate humidifier is installed, capable of keeping the humidity at a constant level independent of the air temperature. The UPS system from Riello is installed inside module 1. The batteries, located outside of module 1, attached through the UPS enable up to 19 minutes of black-out operation (at a full 80 kW critical load). The module energy usage is monitored using ABB electrical power meters and Kamstrup thermal fluid flow meters. The environmental conditions (temperature, humidity and air flowrates) are monitored using Raritan EMX2-888 units with connected sensors.

2.1.4 Environment and Facility Measurement. To measure, monitor and be able to analyze the operational characteristics of the DC, all hardware equipment is monitored. The majority of the facility equipment is monitored using MODBUS RTU (serial based) or MODBUS TCP communication. All MODBUS RTU signals are converted into MODBUS TCP to easily get them on to the management network, and into the common data collection system (described later in section 3).

With the ambition of having a holistic view of the DC, the same management network is employed for both IT and facility monitoring.

[2]SEE HDZ, https://www.seecooling.com/files/2016-02/1455785046_see-cooler-hdz-w.pdf
[3]http://www.dell.com/en-us/work/shop/povw/poweredge-r730xd
[4]Open Compute Windmill http://files.opencompute.org/oc/public.php?service=files&t=935ae31805d3e7a120bf5b0d08db819c
[5]Schleifenbauer Hybrid PFU https://schleifenbauer.eu/en/hybrid-pdu

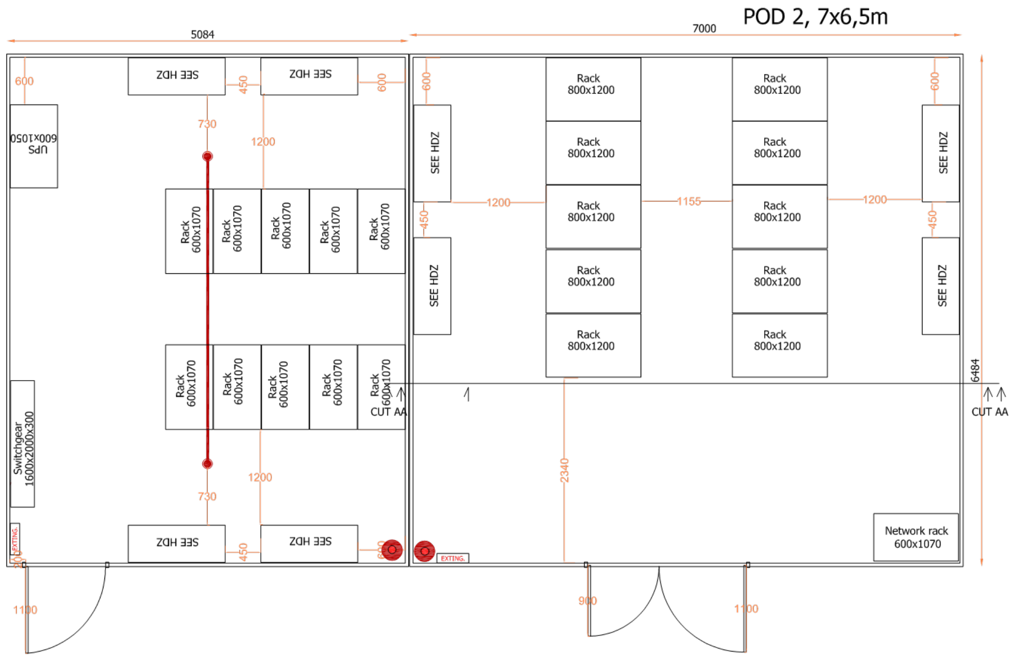

Figure 1: SICS ICE Module 1 to the left and Module 2 to the right.

2.2 Module 2

SICS ICE Module 2 is also a room-in-room module by PriorIT System 42 and contains 10 IT racks and 4 SEE HDZ-2 CRAHs. A major difference is that Module 2 is equipped with a Bergvik raised floor, which facilitates experiments that can employ slab floor cooling, with a hot central aisle, or a raised floor cooling with a central cold aisle and floor tiles. The module dimensions are 6,5 x 7 metres and the layout can be seen in Figure 1.

The module hosts simpler IT hardware compared to module 1, and is primarily used for hardware testing, which can involve both IT and Facility equipment. The facility has been used to compare airflow arrangements, see [7], but also assessment of different cooling approaches or analysis of IT load balancing schemes, listing a few other possible experiments.

2.2.1 IT Equipment. The IT equipment is easily replaceable, but usually the module is fitted with ordinary 1U and 2U servers from Dell, R430 and R530 series servers. The network architecture in this module is limited to 1Gbps ethernet, since this module is not currently configured for advanced IT-research. Both management and production networks are based on the Dell N1500 switch series.

In addition to the Dell R430 and R530 servers, seven HP c7000 HPC enclosures[6] are also available in module 2. These blade systems draw considerably more power per rack than the Dell servers, making it possible to test IT loads from 2.6kW to 19.2kW per rack, which enables experiments of different power densities within the same module.

[6]https://www.hpe.com/us/en/product-catalog/storage/disk-enclosures/pip.hpe-bladesystem-c7000-enclosures.1844065.html

2.2.2 IT Measurement. Metrics from the IT-equipment are collected using the identical setup as in module 1.

2.2.3 Environment and Facility Equipment. The facility equipment installation is very similar to the one found in module 1. A major difference to module 1 is that it is possible to operate module 2 using raised floor cooling. The cool air can either be forced to flow under the raised floor, or be directed towards the racks directly. The default racks installed in module 2 are also wider compared to those in module 1, namely 800mm compared to 600mm. The power distribution is also different in module 2 as it is based on busbars from AP Netherlands. Tap-off boxes with internal Jantiza power meters are also available in module 2.

2.2.4 Environment and Facility Measurement. What distinguishes the facility measurement from module 1 is the metered tap-of boxes, other than this, the facility measurement capability is the same as in module 1.

3 THE SOFTWARE COMPONENTS

The ambition of the research program has been to develop a holistic view of the DC, where both facility and IT measurement data can easily be accessed from a single system, and to achieve this holistic analysis by employing appropriate opensource tools as much as possible. There are numerous DC monitoring tools available, and many opensource tools, e.g. Icinga, Monit and Zabbix[7]. For the SICS ICE installation, Zabbix has been adopted, since it has support for the different protocols, such as MODBUS, SNMP and IPMI.

Zabbix is configured to collect data from PDUs, power switches, servers, network switches, environmental sensors, UPS, coolers,

[7]https://www.zabbix.com/

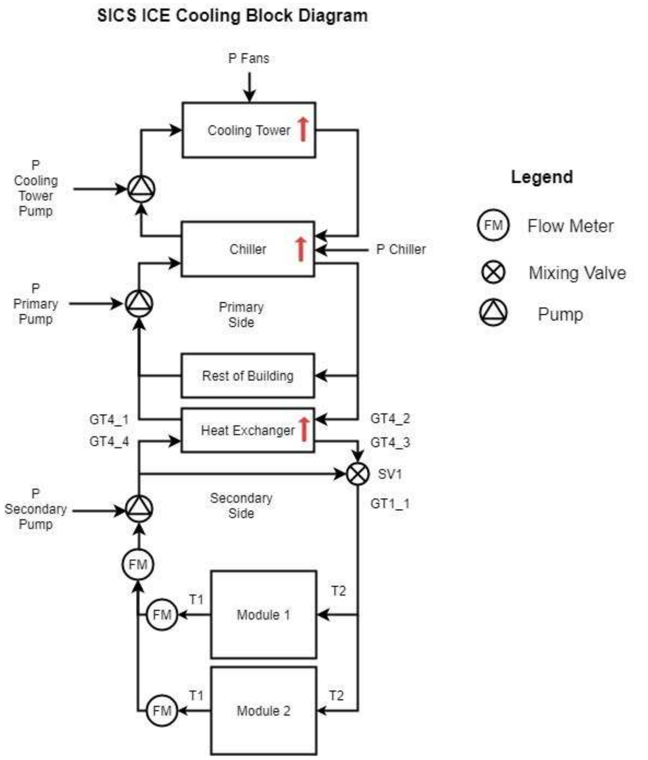

Figure 2: SICS ICE Cooling system arrangement. The DC heat is rejected through a heat exchanger, which transfers heat to the warehouse internal cooling circuit. This circuit it cooled by compressor based chillers with evaporators located outside on the roof of the warehouse.

power meters, heat meters and corrosion sensors, also Zabbix offers the possibility to install operating system resident agents to monitor individual server IT component usage. Zabbix collects and stores data, and offers a user friendly interface to monitor changes in device statuses and operation in real-time, different alarms and notifications can also be configured in Zabbix to notify operators if something is operating out of specified ranges.

However, as the Zabbix database grows, the interactive environment becomes slower and it is clear that the Zabbix database arrangement is not intended to handle very large amounts of data.

To create a more scalable solution, and enable analytics tools access to the data collected by Zabbix, more efficient and suitable long-term data storage is required. To meet the requirements of different researchers and data scientists, a dual storage system was setup, both a Hadoop file system (HDFS) and a time series database (KairosDB) where installed. The data is extracted from the Zabbix database and fed into a Kafka streaming platform, where multiple input and output data streams can be connected, see Figure 3. In this setup two Kafka output modules were created, one to store the data in the KairosDB and the other for the HDFS system. Kafka therefore replicates data streaming to both KairosDB and HDFS. Depending on the data scientists preferred tools, datasets can easily be accessed via both KairosDB and the HDFS.

A challenging part in the data transition between the Zabbix system and the long term storage was how the naming and data format should be designed to be easily searchable and accessible. Data was extracted from Zabbix in a format suitable for a time series database.

The SICS ICE storage schema design. The following standard time series database format was chosen:

```
<metric name> <time stamp> <value> <tag> <tag>...
```

Where, in this case, <metric name> is one word. <time stamps> is time in seconds since Jan 1, 1970 (unix epoch). <value> holds the measured value (in long or double format). <tag> is of the form key=value.

Metric. The metric should represent a specific "thing" or measurement, like Ethernet packets or temperature, but not broken out into a particular instance of a "thing". The definition of what the metric represents is defined with the <tag>s.

Tag(s). The <tag> field applicable in the SICS ICE installation can take the following keys:

dc Defines what specific data center in the case of several data centers being monitored.
pod Data center module (POD) number, 1, 2, 3 etc.
rack The rack number in the pod. (This is left empty for equipment not mounted in rack.)
host Defines which host the values are read from, servers, weather station, CRAH, etc. (This data will in some cases be redundant.)
source What is the source of the measurement on the current host. For temperature metrics this could be for example cpu, ram, top-front, etc.
unit The unit the metric is represented in.
id If one host has multiple sources of the same type, id is used to distinguish between them, e.g. there can be multiple cpus on the same host.
opt Optional key, very rarely used.

Zabbix item configuration requirements to support time series data extraction. In Zabbix, each Item[8] needed to be marked with a metric and enough tags to be able to make the combination of metric and tags unique.

The following tags: **pod**, **rack**, **host** could be extracted from the hostname associated with the given item. This is possible because of the SICS ICE naming convention of hosts. For example "p01r02srv10" will give rise to the tags pod=1, rack=02, host=p01r02srv10. The tag **unit** is taken directly from the *Units*-field (if set) and tag **dc** is set depending on the DC geographic location, e.g. dc=lulea.

The metric and additional tags are entered into either the *Key*- or *Description*-field of the Zabbix item. The reason to support the *Description*-field was because on occasions the preferred *Key*-field was already populated.

The *key* or *description* field was formatted according to the following format examples.

[8] An Item represents one point of measurement, e.g. the voltage at outlet number 2, in PDU2 in Module 2

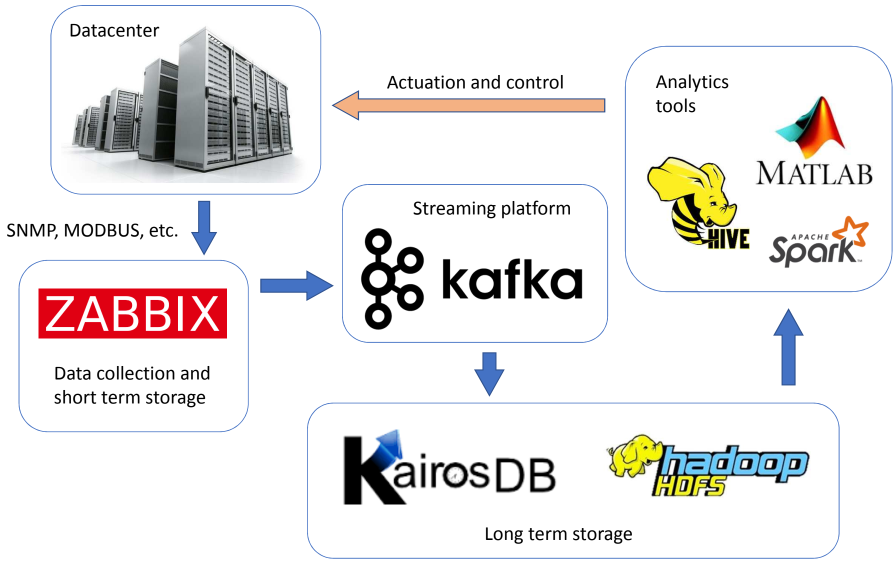

Figure 3: Data measurement flow. All metrics are collected using Zabbix. Relevant data is fed into a Kafka streaming platforms, which in turn outputs data to KairosDB and HDFS platforms for long term storage. The data can be accessed by various analysis tools, to create a better understanding of the DC operation.

Key:

```
Format:  -<metricname>(-<tag>.<value>)*
Example: -temperature-source.cpu-id.1
```

Description:

```
Format:  |key:<metricname>(-<tag>.<value>)*|
Example: |key:temperature-source.cpu-id.1|
```

Based on the above information, it is possible to export the data following the time series database format specified earlier. Examples of KairosDB posts are highlighted in Figure 4.

The KairosDB database can operate with an open HTTP api, which enables easy export of data to analysis tools, such as Mathworks MATLAB. On request by partners in the Sendate project, an implementation of a export script with a complementary GUI and advanced search functionalities have been provided.

As mentioned earlier, all data is also stored on a HDFS file system to enable Hadoop based big data search and analytic tools. The data is stored using the JSON (JavaScript Object Notation) data interchange format, which is easily readable by humans and easy for computers to generate and parse. A new JSON file is generated each hour to store all the data generated during the previous hour. In Figure 5 an example for a server exhaust temperature reading post is demonstrated.

4 OPERATIONAL RESULTS

4.1 Data center energy metrics

Module 2 enables full control of the IT equipment where each individual server can be stressed to produce synthetic workloads based on Stress-ng[9] and more specifically, using the so-called matrixprod CPU stress method. In addition to the ability to change the IT load as a function of time, the thermal environment can be fully controlled. Cooler fan speeds, supply water temperature to the module, water flow and air temperature in front of the servers can all be controlled.

One of the most common DC metrics is the Power Usage Effectiveness (PUE) [1], which can be instantaneously monitored in Zabbix but also evaluated based on data from the KairosDB long term storage. A specific part of the DC can similarly be analyzed through the partial PUE (pPUE). Any previous experiment and all corresponding data can easily be investigated a posteriori by using the opensource data collection and storage system described in this work.

4.2 Examples of IT system energy performance

As an example of a DC energy related experiment to demonstrate how the controlled environment and IT equipment in combination with the data collection setup can generate useful data, the liquid supply temperature setpoint was simply changed instantly from 15.5°C to 17.5°C at 5h, while maintaining all other DC parameters constant. Figure 6 demonstrates how the increase in cooling liquid temperature (solid line) affects the DC PUE (dotted line). This shows how the CRAH fans slow down in combination with increased speeds of internal server fans due to higher server inlet air temperatures. The cause of the system's behaviour and drop in PUE can be analyzed in more detail if data from all other environmental sensors and individual components were included. However, the prime focus of this work has been on the adoption of opensource tools in monitoring a live DC. Measured and collected data from the

[9] https://wiki.ubuntu.com/Kernel/Reference/stress-ng

```
temperature <time> 23.1 dc=lulea pod=1 rack=3 host=p01r02emx888 unit=C source=front-top
temperature <time> 45.8 dc=lulea pod=1 rack=5 host=p01r05srv18 unit=C source=cpu id=1
temperature <time> -1.2 dc=lulea host=weather_station unit=C source=air
current <time> 1.1 dc=lulea pod=1 rack=5 host=p01r05hpdu18 unit=A source=outlet id=1
voltage <time> 231.1 dc=lulea pod=1 rack=5 host=p01r05hpdu18 unit=V source=outlet id=1
speed_rpm <time> 2280.0000 dc=lulea pod=2 rack=8 host=p02r08srv05 unit=RPM source=system id=3
```

Figure 4: Examples of database entries.

```
{
    "host":"zabbix",
    "metric":"temperature",
    "value":"33.0000",
    "timestamp":1521454425392,
    "tags":
    {
        "dc":"lulea",
        "host":"p02r07srv21",
        "pod":"2",
        "rack":"7",
        "unit":"C"
        "source":"exhaust"
    },
}
```

Figure 5: JSON formatted data example as stored on HDFS.

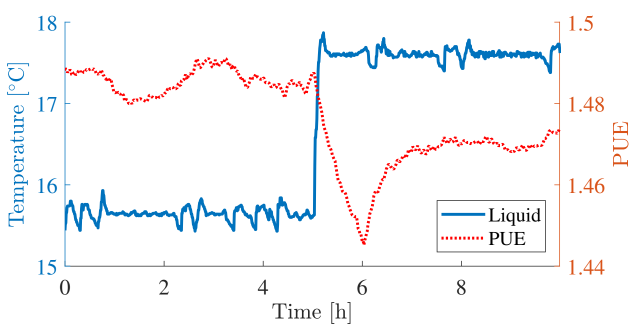

Figure 6: Example of what can be analyzed using the data monitoring system. The blue line represents the cooling medium supply temperature to the cooling units in module 2, and the red dotted line shows the corresponding PUE in the same module.

system presented here can offer greater understanding of transient effects, such as thermal inertia seen in Figure 6.

4.3 Historical analysis capabilities

All data stored in HDFS and KairosDB, can easily be exported for post analysis. It is also possible to use analytical real-time analysis tools that could aid early stage detection of possible failures, or degradation of the DC performance. The KairosDB include APIs for data export and a web-based GUI for viewing shorter periods of data directly in the browser. The data search is done by specifying the metric of interest, setting the time period, and defining which tags are of interest.

For example, to plot the blue line in Figure 6 the search parameters required would be: Metric: `temperature`, to identify what temperature, a number of tags are required to be specified to single out the parameter of interest, in this case the Tags: `dc=lulea pod=1 source=dc_supply` are required, additional tags are in this example redundant.

5 CONCLUSIONS

This paper demonstrates the possibility of setting up and configuring a powerful data collection platform, and a brief introduction of how it can be used to provide data for analytical research on DC performance, using a combination of established opensource tools.

The methodology of using the combined software tools offers stable and scalable monitoring and measuring tools that could be elements of a DCIM environment, where it can be employed to perform historical reporting of the DC performance and offer an excellent platform to augment relevant research of the whole DC as an integrated system.

6 FUTURE WORK

The arrangement of tools described here can be developed to include control infrastructure that can manage the facility, where the data, through the application of analytics tools, can feedback information to the control system to improve the system control. This information would naturally be communicated using the same Kafka streaming platform, enabling multiple users (or external systems) to access relevant information.

ACKNOWLEDGMENTS

The authors would like to thank Celtic+ and Vinnova for funding project Sendate EXTEND, Project-ID: C2015/3-1. The authors would also like to thank all partners in Sendate EXTEND for their valuable and fruitful discussions around DC monitoring and control. A special thanks also goes to the technical staff at RISE SICS North, for supporting the development and implementation of this system.

REFERENCES

[1] Victor Avelar, Dan Azevedo, and Alan French. 2012. PUE: A comprehensive Examination of the Metric. The Green Grid.

[2] David J Cappuccio. 2010. DCIM: Going Beyond IT. *Gartner ID G* 174769 (2010).

[3] Dave Cole. 2012. Data center infrastructure management. *Data Center Knowledge* (2012).

[4] Mark Harris and Hwaiyu Geng. 2015. Data center infrastructure management. *Data center handbook, 1st edn. Wiley, Hoboken, NJ* (2015), 601–618.

[5] Matthew Hubbell, Andrew Moran, William Arcand, David Bestor, Bill Bergeron, Chansup Byun, Vijay Gadepally, Peter Michaleas, Julie Mullen, Andrew Prout, et al. 2015. Big Data strategies for Data Center Infrastructure management using a 3D gaming platform. In *High Performance Extreme Computing Conference (HPEC), 2015 IEEE*. IEEE, 1–6.

[6] Moises Levy and Jason O Hallstrom. 2017. A new approach to data center infrastructure monitoring and management (DCIMM). In *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual*. IEEE, 1–6.

[7] Emelie Wibron. 2018. *A Numerical and Experimental Study of Airflow in Data Centers*. Ph.D. Dissertation. Luleå University of Technology.

[8] Emelie Wibron, Anna-Lena Ljung, and T Staffan Lundström. 2018. Computational Fluid Dynamics Modeling and Validating Experiments of Airflow in a Data Center. *Energies* 11, 3 (2018), 644.