Anomaly Detection in Data Center IT & Physical Infrastructure

Elisabetta Ronchieri[1,2,*], *Luca* Giommi[1,**], *Luigi Benedettto* Scarponi[1,***], *Luca* Torzi[1,****], *Alessandro* Costantini[1,†], *Doina Cristina* Duma[1,‡], and *Davide* Salomoni[1,§]

[1]INFN CNAF, Bologna, Italy
[2]Department of Statistical Sciences, University of Bologna, Bologna, Italy

Abstract. Anomaly detection in data center IT and physical infrastructure is challenging due to the amount of heterogeneous data to be analyzed. Defining a solution that early identifies unexpected anomalies is particularly important to prevent data losses, breakdown of the system, and any other event considered to be critical for the activity of the data center.
In the context of the INFN CNAF data center, one of the WLCG Tier-1s, we have performed a study based on monitored cooling, electrical, and IT hardware and software metrics to identify anomalies. In the present work, we aim to explore statistical approaches and machine learning solutions in the anomaly detection field for time series numerical metrics related to IT and physical infrastructure sensors.
With the usage of statistical Z-score and percentile approaches and clustering DBSCAN technique, we have been able to group and identify anomalous events. Using the presented approach, different relevance can be attributed to the likely anomalies.

1 Introduction

Data centers host IT and physical infrastructures to support researchers in transmitting, processing, and exchanging data. They use infrastructure observability platforms to have access to heterogeneous data that can be analyzed and used to detect and predict events of interest, particularly unexpected anomalies. Data include, among others, CPU and memory consumption, network traffic, cooling, and electrical states. Anomaly detection in data center IT and physical infrastructures is of great significance in order to prevent data losses, breakdown of the system, and any other event considered to be critical for the activity of the data center. This solution might support system managers and engineers to properly design the redundancy of IT apparatus and monitor the whole data center.

[*]e-mail: elisabetta.ronchieri@cnaf.infn.it
[**]e-mail: luca.giommi@cnaf.infn.it
[***]e-mail: luigi.scarponi@cnaf.infn.it
[****]e-mail: luca.torzi37@gmail.com
[†]e-mail: alessandro.costantini@cnaf.infn.it
[‡]e-mail: cristina.aiftimiei@cnaf.infn.it
[§]e-mail: davide.salomoni@cnaf.infn.it

The CNAF data center of the Italian Institute for Nuclear Physics (INFN), one of the WLCG Tier-1s, serves more than 40 international scientific collaborations in multiple scientific domains, including high-energy physics experiments running at the Large Hadron Collider in Geneva. INFN CNAF already provides a set of pillars, such as connectivity to the cloud and infrastructure for data transmission, and security for data and privacy protection. Within this context, we have performed a study, based on monitored cooling, electrical and IT hardware metrics, to add the intelligence data pillar to the INFN CNAF main functionalities. This new pillar can be implemented through infrastructure and algorithms in order to extract value from service and physical resources (i.e. logs and monitoring metrics), convert them into useful information (e.g. detecting anomalies), and properly intervene. This might allow for performing predictive maintenance with machine learning (ML) techniques and time series analysis.

In previous work, we have focused on the detection of anomaly patterns by considering service log files and data coming from IT monitoring measurements, and leveraging natural language processing (NLP) solutions juxtaposed with multivariate time series anomaly detection techniques [1]. This study has revealed thousands of anomalies that have been verified by a comparison with the same log messages derived from the different services considered for the analysis. It has also computed anomaly scores on monitoring data to identify the timeframe where we could overlap services and monitored data anomalies to perform predictive maintenance analysis.

This contribution brings work [1] further, exploring the statistical Z-score [2] and percentile [3] approaches, and Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (DBSCAN) [4] in the anomaly detection field for time series numerical metrics related to IT and physical infrastructure sensors. The paper describes models defined by considering critical scenarios and a wide range and type of monitoring data. Our study also takes advantage of the threshold-based alarming system, set for each monitored metric of IT and physical infrastructures, to label the recorded events and use them later within semi-supervised ML techniques. The relationship between the anomaly scores and the threshold-risk values can be assessed to be used for predictive maintenance management. The theoretical studies, based on real monitored data, and the related achievements are adopted and used to improve the existing monitoring platform to recognise and prevent anomaly detection within the national data center of INFN, CNAF.

2 Sensors Data

The INFN CNAF data center has a total space occupation near to 2000 m^2 distributed in different floors. The area dedicated to the IT devices is 500 m^2, the rest is occupied by the facility plants such as the transformers rooms and the cooling central. The data center is designed to satisfy an IT power consumption equal to 1.2 MW.

Figure 1 shows a logical scheme of the data center facility, highlighting the typologies of data sources considered in this study.

2.1 The electrical system

The INFN CNAF data center has a dedicated electrical power supply at a 15000 [V] medium voltage. There is a transformation room with N=3 transformers (TRs) of 2.500 [kV] in N+1 redundancy: two are always in operation, while the third one is in reserve. The system is designed to switch transformers by simply flipping the switches. From the transformation rooms, there are two power lines (green and red lines in Figure 1) that supply power to all the IT devices of the data center. In this way, a full dual redundant power supply is guaranteed.

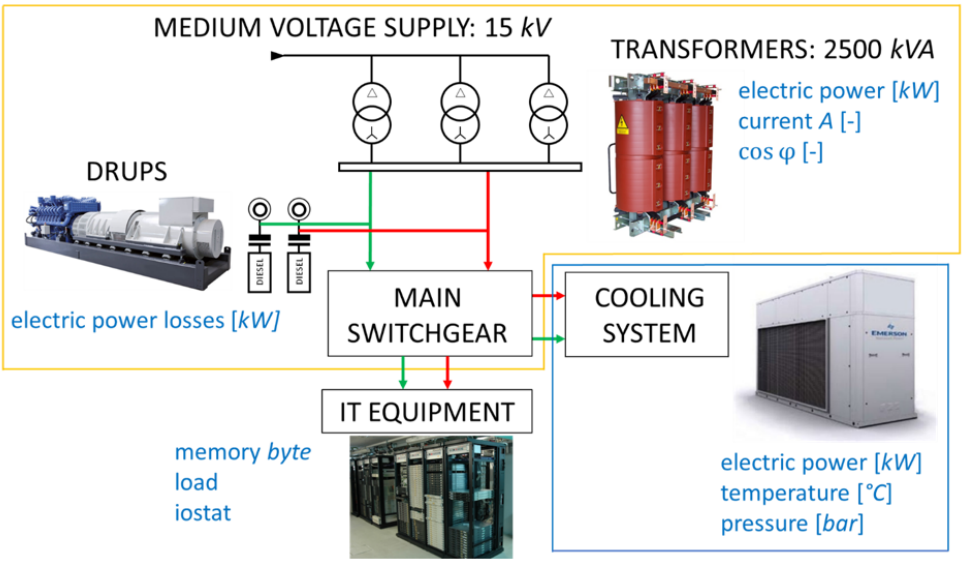

Figure 1. The INFN CNAF Tier1 Data Sources: the electrical system in the orange colour; the cooling system in the blue colour; and the IT equipment.

To ensure electrical continuity in case of a national grid blackout, there are two rotative continuity groups (DRUPS) of 1.750 [kV] (one for each power line), which have the dual function of UPS and electrical power generator. These DRUPS are dedicated to powering the IT equipment while another generator supplies power to the cooling central.

The electrical system provides the following sensor data:

- P_{TR_i} is the absorbed electric power [kW] for each i-th transformer TR;
- I_{TR_i} is the phase current [A] for each i-th transformer TR;
- $\cos(\phi_{TR_i})$ is the power factor [-], that is the ratio between active and apparent power for each i-th transformer TR, which value is $\in [0,1]$;
- P_{DRUPS} is the electric power loss of the DRUPS system [kW] expressed by the difference between the input and outlet power of the DRUPS system.

2.2 The cooling system

The cooling system is composed of M=6 chillers (CHs) with a cooling power equal to 300 [kW] for each, designed in redundancy M+2. The pumping system has a redundancy of 2 chillers.

The IT area is designed with the technology of hot aisle containment and the cooling of IT rack is guaranteed by an in-row cooling system. The chillers cool water down to 15°C, which is sent to the in-rows coolers and then returns to the chiller at 20°C. The set point temperature of the air supplied to IT devices is 24°C. The facilities are monitored and adjusted by a Building Management System (BMS), which records every variable of interest, such as currents, temperatures, and pressures. The BMS sends alarms to warn of any malfunctions in the system: e.g. it warns when the outlet water of the chiller is more than 18°C.

The cooling central system provides the following sensor data:

- Pr_{CH_i} is the electric power [kW] for each i-th chiller CH;
- Tin_{CH_i} and $Tout_{CH_i}$ are inlet and outlet temperatures of the water [°C] for each i-th chiller CH;
- p_{CH_i} is the pressure of the refrigerant [bar] before entering the condenser, for each i-th chiller CH;
- $Tenv$ is the environment temperature [°C];
- $Pump$ is the electric power [kW] of the pumping system.

2.3 The IT equipment

The IT equipment is characterized by a large amount of resources for which monitoring metrics have been collected to get information about the health status of machines. In this study, we have considered a subset of INFN CNAF Tier1 resources: 1000 different worker nodes, 20 user interfaces, 130 PB on tape, and 70 PB on disk.

Monitoring metrics are obtained by using Linux commands that provide statistics belonging to three categories: the actual *load* of the machine, and its average (*load_avg*) over multiple time frames (i.e., 1, 5 or 15 minutes); the *memory usage*; the *central processing unit* (CPU) usage and *input/output* (I/O) (i.e., *io_stat*) rate of data. Load averages are often provided at the last minute, the last five minutes, and the last fifteen minutes. There are many metrics related to the memory usage of a system on which processes are running (see Table 1). IoStat is another command, which allows checking more general metrics about the current status of the system (see Table 1). In our study, 18 metrics have been considered.

Table 1. A subset of memory usage and IoStat metrics.

Category	Metric	Description
Memory usage	total	Total installed memory
Memory usage	used	Memory currently in use by running processes
Memory usage	free	Unused memory
Memory usage	shared	Memory shared by multiple processes
Memory usage	buffers	Memory reserved by the operating system to be allocated as buffers when processes need them
Memory usage	cached	Recently used files stored in RAM
Memory usage	buff/cache	Buffers + Cache
Memory usage	available	Estimation of available memory for starting new applications
IoStat	user	CPU% used executed at user level
IoStat	nice	CPU% used executed at the user level with nice priority
IoStat	system	CPU% used while executing at the system (kernel) level
IoStat	iowait	time% of the CPU(s) in idle when system pending requests
IoStat	steal	time% spent on involuntary wait due to another virtual processor
IoStat	idle	time% of the CPU(s) in idle with no system pending requests

Figure 2 shows default information of the Linux *top* command output in kibibytes: total, used, free, and buffers measures are provided for the memory, while total, used, free, and cached are given for the swapped memory.

```
Mem:  1695028k total, 1677560k used,   17468k free, 131036k buffers
Swap:   499704k total,     18164k used, 481540k free, 218792k cached
```

Figure 2. Statistics collected with the Linux *top* command.

Table 2 shows an example of the *memory.used* monitoring metric turned into the csv format. The variables are: *time* expressing when the metric has been registered in the local time zone; *tags* and *domain* providing information about the collected machine values; *name* and *metric* providing the metric name and category; *value* containing the metric value.

Table 2. The memory.used monitoring metric turned into the csv format.

name	*tags*	*time*	*domain*	*duration*	*metric*	*value*
memory.used	host=api-int.cnsa.cr.cnaf.infn.it	XXXX	cnsa.cr.cnaf.infn.it	0.22133	metrics-memory	1,37128E+09
memory.used	host=api-int.cnsa.cr.cnaf.infn.it	XXXX	cnsa.cr.cnaf.infn.it	0.23058	metrics-memory	1,39689E+16
memory.used	host=api-int.cnsa.cr.cnaf.infn.it	XXXX	cnsa.cr.cnaf.infn.it	0.22633	metrics-memory	1,46820E+16
memory.used	host=api-int.cnsa.cr.cnaf.infn.it	XXXX	cnsa.cr.cnaf.infn.it	0.21558	metrics-memory	1,49160E+16
memory.used	host=api-int.cnsa.cr.cnaf.infn.it	XXXX	cnsa.cr.cnaf.infn.it	0.21825	metrics-memory	1,49583E+16

3 Pre-processing phase

Data have been properly pre-processed before feeding predictive models.

The majority of the collected data has a time window that begins on January 6, 2022; however, the DRUPS system data starts on May 31, 2022. Some data that sum together the values of different sensors starts on May 17, 2023 and have less than 5,000 values. All the collected data time windows end on July 7, 2023.

Most of the sensors values are sampled every 15 minutes, however, some sensors (like the DRUPS system) are sampled every 10 minutes. The *load* average is provided over multiple time frames of 1, 5, or 15 minutes. When merging together more sensors data, using the timestamp as key, a tolerance of 15 minutes has been inserted to obtain rows where every timestamp has the values in each sensor.

The cooling system contains some data, that came from sensors that measure the temperature of the incoming and outgoing water to the Refrigerator Unit and the evaporator temperature, which have values ten times bigger than the real values: they have been scaled by a factor 10.

3.1 Correlating data

Considering the amount of data and outcomes, to identify if there are sensors to omit in the results, all data in the various categories have been correlated.

The various IT equipment metrics are correlated both positively and negatively, each one belonging to a metric category: *iostat.avg-cpu.pct_idle* that represents the average usage of CPU in the idle state of the machine; *load_avg.fifteen* that represents the average load of work on 15 minutes timespan; *memory-usage* that is the ratio between the memory used and the total memory space.

In the vast majority of every correlation matrix of the chiller system (see Figure 3), there is a semi-strong correlation between the power consumption $Pump$ and the pressure p_{CH} of the cells of the unit itself. Furthermore, in most of the units, there is a high correlation between the environmental temperature $Tenv$ and the pressure of the cells.

In every transformer of the electrical system, there is a strong correlation between the current phase 1, 2, 3 (I_{TR_1}, I_{TR_2}, I_{TR_3}) and the apparent power, as shown in Figure 4. Furthermore, the trends of the neutral power and the cosine of phi $\cos(\phi)$ are similar.

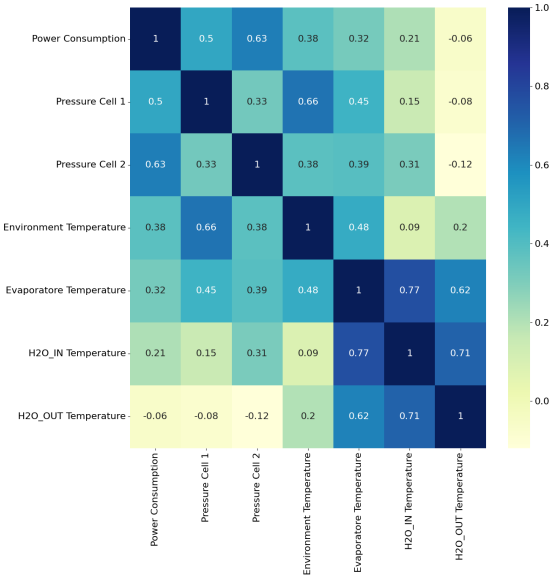

Figure 3. Correlation matrices of the chiller (CH) 1.

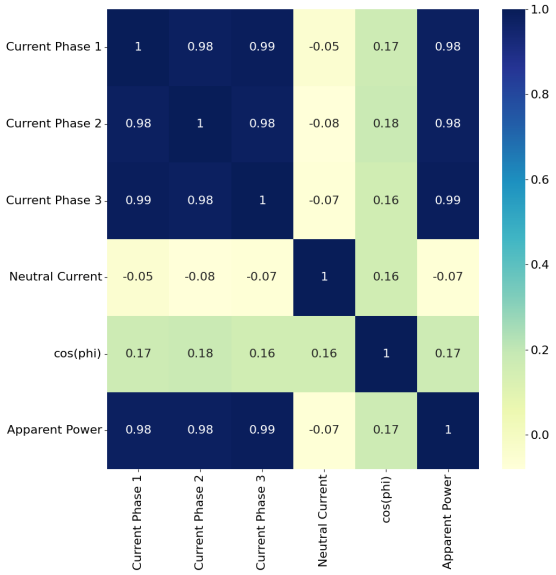

Figure 4. Correlation matrix of the Transformer (TR) 1.

3.2 Alarm signals

Each system sends alarms if the sensors values overcome their corresponding thresholds. For example, the chiller system sends alarms if the outgoing temperature overcome 15°C or if the incoming temperature overcame 20°C. Figure 5 and Figure 6 show the trends of the temperatures with respect to the limits.

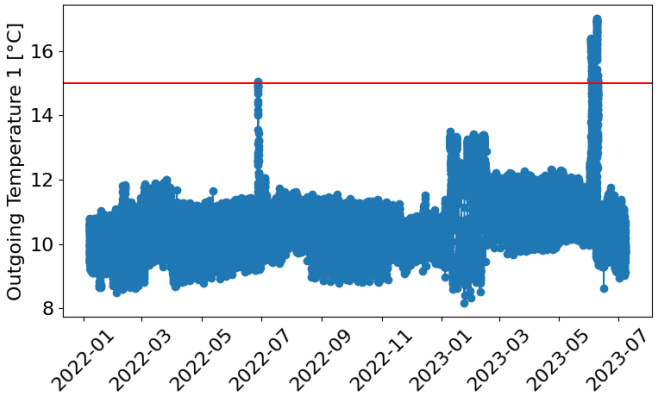

Figure 5. Outgoing temperature for chiller (CH) 1 ($Tout_{CH_1}$). The red line represents the limit of 15°C.

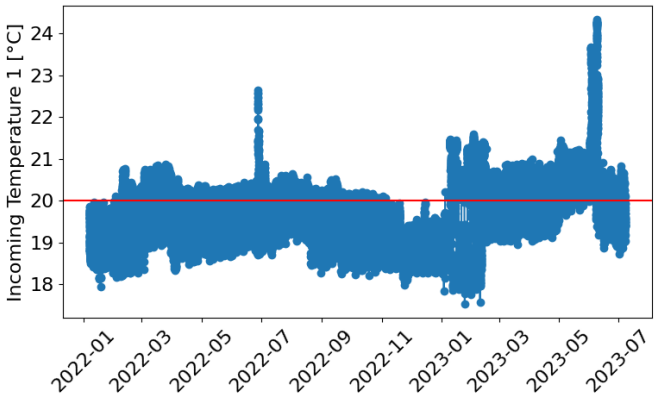

Figure 6. Incoming temperature for chiller (CH) 1 (Tin_{CH_1}). The red line represents the limit of 20°C.

4 Models

The collected variables have been scaled and used to feed the proposed models, based on two statistical approaches and a clustering algorithm, to group and identify anomalous events.

The first method is a statistical one based on the mean and the standard deviation of the values: all the values that have an absolute value of the Z-score (equation 1) bigger than 4 are

classified as anomalous.

$$Zscore = \frac{x - \mu}{\sigma} \tag{1}$$

The second method is based on the percentile: every value that is smaller than the 0.2-percentile or bigger than the 99.8-percentile is detected as anomalous. This type of measurement is more robust against the outliers because we are not considering values at the borders.

The third method used is a ML algorithm called DBSCAN: it creates clusters of data points and the points that do not belong to any cluster are detected as anomalous. Before using this last method, the data values are normalized over both the time and the values: this operation is performed using the *StandardScaler* class.

5 Results

To compute the anomalies with the Z-score and the percentiles, only the values are used, instead to compute the anomalies using DBSCAN both the values and temporal information are considered, such that it is possible to create clusters of data points that have similar behavior in a time period.

The parameters used in DBSCAN are $\epsilon = [0.5, 0.7, 0.8]$ and $min\_sample = [5, 120, 250]$. Furthermore, all the clusters that do not respect equation 2 are considered as anomalous:

$$num\_sample\_in\_cluster > \frac{num\_entries}{S} \tag{2}$$

where $S = [30, 100, 50]$ if sensors belong to the chiller system, electrical system, and IT equipment.

To capture the rapid evolution of some sensors (like those that measure the pressures inside the cells of each refrigerator unit), after the standard scaling, the timestamps have been scaled up by a factor of 15 in the sensors of the chiller system. This is done so that values that change a lot (with respect to the normal behavior in a time window) are detected as anomalous by DBSCAN, instead without this operation those values are included in the clusters.

The percentiles method is able to detect as anomalous only points values that are outside of the range considered. The Z-score can be able to detect anomalous points according to the variation of the sensors values: when it is really big and most values are considered normal, Z-score is not able to detect any point.

6 Conclusions

Using the presented approach, we can attribute different relevance to the likely anomalies, contributing to turn them in proper alarm signals to trigger different actions in the data center and to refine policies aimed to address and solve arising problems.

References

[1] L. Viola, E. Ronchieri, C. Cavallaro, Computers **11** (2022)

[2] C.H. Brase, C.P. Brase, *Understanding Basic Statistics* (Cengage Learning, 2018)

[3] R. Hyndman, Y. Fan, The American Statistician **50**, 361 (1996)

[4] M. Ester, H.P. Kriegel, J. Sander, X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, in *Proc. of 2nd International Conference on Knowledge Discovery and* (1996), pp. 226–231