PAPER • OPEN ACCESS

Evolution of the Building Management System in the INFN CNAF Tier-1 data center facility.

To cite this article: Pier Paolo Ricci *et al* 2017 *J. Phys.: Conf. Ser.* **898** 082005

Evolution of the Building Management System in the INFN CNAF Tier-1 data center facility.

Pier Paolo Ricci[1,4], Massimo Donatelli[1], Antonio Falabella[1], Andrea Mazza[1] and Michele Onofri[1].

[1] INFN CNAF Viale Berti Pichat 6/2 40127 Bologna, Italy.

E-mail: pierpaolo.ricci@cnaf.infn.it

Abstract. The INFN CNAF Tier-1 data center is composed by two different main rooms containing IT resources and four additional locations that hosts the necessary technology infrastructures providing the electrical power and cooling to the facility. The power supply and continuity are ensured by a dedicated room with three 15,000 to 400 V transformers in a separate part of the principal building and two redundant 1.4MW diesel rotary uninterruptible power supplies. The cooling is provided by six free cooling chillers of 320 kW each with a N+2 redundancy configuration. Clearly, considering the complex physical distribution of the technical plants, a detailed Building Management System (BMS) was designed and implemented as part of the original project in order to monitor and collect all the necessary information and for providing alarms in case of malfunctions or major failures. After almost 10 years of service, a revision of the BMS system was somewhat necessary. In addition, the increasing cost of electrical power is nowadays a strong motivation for improving the energy efficiency of the infrastructure. Therefore the exact calculation of the power usage effectiveness (PUE) metric has become one of the most important factors when aiming for the optimization of a modern data center. For these reasons, an evolution of the BMS system was designed using the Schneider StruxureWare infrastructure hardware and software products. This solution proves to be a natural and flexible development of the previous TAC Vista software with advantages in the ease of use and the possibility to customize the data collection and the graphical interfaces display. Moreover, the addition of protocols like open standard Web services gives the possibility to communicate with the BMS from custom user application and permits the exchange of data and information through the Web between different third-party systems. Specific Web services SOAP requests has been implemented in our Tier-1 monitoring system in order to collect historical trends of power demands and calculate the partial PUE (pPUE) of a specific part of the infrastructure. This would help in the identification of “spots” that may need further energy optimization. The StruxureWare system maintains compatibility with standard protocols like Modbus as well as native LonWorks, making possible reusing the existing network between physical locations as well as a considerable number of programmable controller and I/O modules that interact with the facility. The high increase of detailed statistical information about power consumption and the HVAC (heat, ventilation and air conditioning) parameters could prove to be a very valuable strategic choice for improving the overall PUE. This will bring remarkable benefits for the overall management costs, despite the limits of the non-optimal actual location of the facility, and it will help us in the process of making a more energy efficient data center that embraces the concept of green IT.

1. Introduction: the INFN CNAF infrastructure and IT resources.

Since 2005 the INFN CNAF Tier-1 has become the Italian national data center for the INFN computing activities [1]. In particular, all of the LHC experiments (ALICE, ATLAS, CMS and LHCb) use our site as a Tier-1 computing resource provider. In addition, about 20 other non-LHC collaborations including Astroparticle Physics (e.g. CDF, Kloe, AMS2, Argo, Auger, CTA, Magic, Pamela, Borexino, Darkside,

[4] Corresponding Author

Virgo, etc.) currently use the computing and storage resources of our centre with a guarantee 24/7 level of service support (24h every day non-stop service availability). At present the Tier-1 IT resources are represented by the following areas:

- Computing Power: a cluster of roughly 205,000 HEP-SPEC06 provided by 22,000 CPU cores
- Disk Storage: an available disk space capacity of ~20 PByte used space served by 80 disk servers (using 8 GPFS clusters with tape extension [2])
- Tape Storage: 1 Oracle SL8500 Tape library (with redundant robotic) with 10,000 slot capacity ~ 34 PByte tape used space (8.5 TByte cartridge)
- Network facility: 4 Core Switches with a LAN composed by a total of ~350 x 10 Gb/s network optical ports and ~468 x 1Gb/s ports. The WAN connections in production are 6 + 2 (general purpose) x 10 Gb/s connections.

The whole INFN CNAF Tier-1 is hosted in a university complex building which has shown to be a quite inadequate location. During 2009 the upgrade of the Tier-1 required all the available space and this corresponds to a total of 1950 m^2 space occupation for two IT resources rooms (250 m^2 + 350 m^2) and four additional locations for the remaining main infrastructure facilities. These four locations includes the transformers room, the UPS room (rotary UPS + one standard generator used for backup), the chillers room (including water pumps and piping system) and the power room with all the power switches and electrical measurement instruments and connections.

In Figure 1 a simplified schema of the Tier-1 infrastructure resources is shown. As illustrated in the figure the main UPS power supply for the IT resources is guaranteed by two Eurodiesel Rotary UPS with a nominal power of 1700 kVA each, while the cooling power is currently provided by six Emerson free cooling chillers with 300 kW cooling capacity in a N+2 redundancy configuration. All the power distribution is carried out using two separated physical lines (identified by the RED and GREEN label) therefore it is possible to provide a dual redundant power supply to all the IT equipment which supports it.

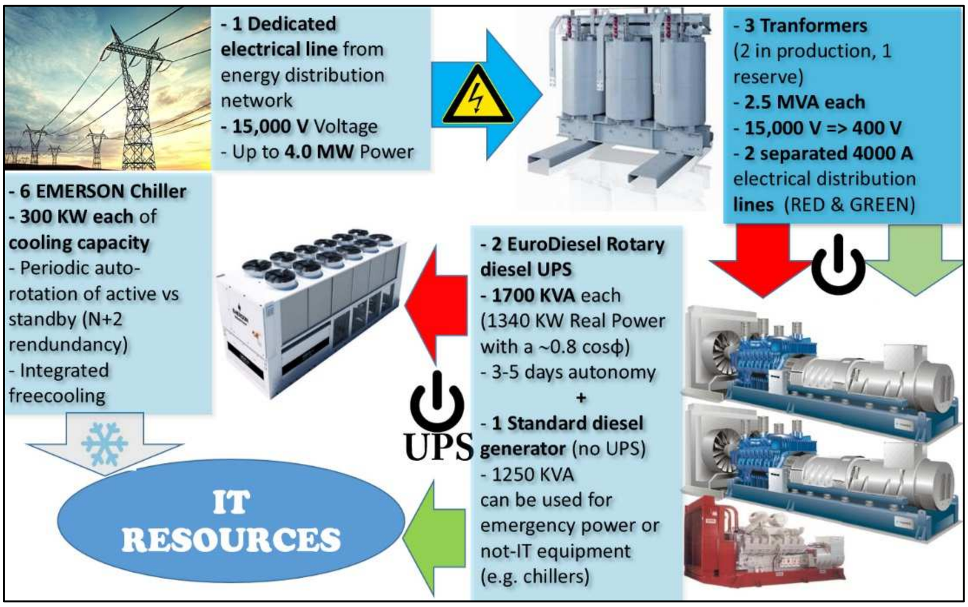

- 1 Dedicated electrical line from energy distribution network
- 15,000 V Voltage
- Up to 4.0 MW Power
- 3 Tranformers (2 in production, 1 reserve)
- 2.5 MVA each
- 15,000 V => 400 V
- 2 separated 4000 A electrical distribution lines (RED & GREEN)
- 6 EMERSON Chiller
- 300 KW each of cooling capacity
- Periodic auto-rotation of active vs standby (N+2 rendundancy)
- Integrated freecooling
- 2 EuroDiesel Rotary diesel UPS
- 1700 KVA each (1340 KW Real Power with a ~0.8 cosϕ)
- 3-5 days autonomy
+
- 1 Standard diesel generator (no UPS)
- 1250 KVA can be used for emergency power or not-IT equipment (e.g. chillers)
UPS
IT RESOURCES

Figure 1. Diagram representing the Tier-1 electrical and cooling power resources.

The IT resources hosted in the INFN CNAF Tier-1 are contained in two main rooms (Room 1 and Room 2) where the architecture of the APC Schneider Electric “hot-aisle” containment system is utilized. In particular a total number of 44 APC InRow RP (IR-RP) precision cooling air handlers units [3] with 2-ways valves, three Electronic controlled (EC) fans and humidity control are used in six “hot-aisle” blocks over the 2 separated rooms. The cold water for the IR-RPs is provided by the six Emerson chillers with 300 kW each of cooling capacity. Referring to the nominal cooling power of each IR-RP (50 kW) it is possible to provide a total of 1600 KW with N+2 redundancy (in terms of IR-RP number) to the IT resources. Since the installation requirement of the two rooms were different, a total of 48 racks with 10 kW cooling capacity each is available in Room 1 (mainly designed for the disk storage resources) whereas the higher capacity is in Room 2 with 76 usable racks available and up to 20 kW cooling capacity each. The “hot-aisle” containment is realized without a floating floor, all the water pipelines use simple floor conduits and the network and the power cabling uses the front top of the racks. All the six “hot-aisle” blocks are carefully monitored with a network of temperature and humidity probes in addition to water leak sensors in the floor conduits. As a reference, the chillers set-point which proved to be ideal for our environment is 15°C for the cold water supply flow and 20°C for the return flow. In the two rooms the set-points for the temperature and relative humidity control are currently 24°C and roughly 45-60%. This translates into real measured average values of ~24°C for the cold corridor temperature and ~31°C in the hot one.

2. The new Building Management System (BMS) software

2.1. Migration of the BMS system.

The whole infrastructure of the INFN CNAF Tier-1 consists of a relatively complex distribution of technical plants over different buildings and floors. A very reliable and detailed Building Management System (BMS) was needed for an overall and clear view as well as for alarms and events notifications. In 2009 during the upgrade of our center, the TAC Vista software was implemented in our Tier-1 and it has proven to be useful for many years. However, the software has started the “phase out” during the last few years and, in addition, it proved to be rather obsolete, a disadvantage which could be hardly ignored. In particular the software proved to be very difficult to edit and therefore it was almost impossible to customize without the constant support of external expertize. It was also difficult to interconnect with other monitoring tools due to complete lack of compatibility with open communication protocols, in fact it uses only the TAC Vista proprietary protocol for data exchange and this limited the possibility of customize and interface the BMS with other Tier-1 software tools. In addition, the user GUI of the TAC Vista software uses Java which limits the interoperability, therefore it was very difficult to access the interface GUI from mobile devices.

The original idea behind a BMS software upgrade started in 2015 when we decided that a more flexible system, that could take advantage of mostly part of the TAC Vista network hardware (sensors and collectors), was needed. This was thought to limit the hardware cost for the migration and simultaneously to improve the user-friendly management of the whole software infrastructure.

The obvious and most natural choice was to migrate the system into the Schneider StruxureWare™ Building Operation software (SBO) architecture [4]. The SBO software is fully compatible with the Modbus [5] protocol (TCP/IP and serial) which was heavily present in the previous TAC Vista implementation and it is also fully compatible with the TAC Vista Lonworks [6] network, giving us the possibility to reuse the existing cabling and a great part of the hardware boxes. Furthermore, the web GUI user interface just needs a standard browser for working (HTM5 compatible) so it is directly accessible from mobile device and the SBO software suite is certified to be open to standard protocol like Web Services (*Serve* and *Consume* operation) and this gives us the possibility to interface the SBO software with other monitoring tools.

The migration from the TAC Vista to the SBO software of the whole infrastructure was easily split into three phases over an eight week period. The first phase was a four-week period during which the two systems were capable of co-working, monitoring together the TCP Modbus devices as well as APC “hot-aisles” and Programmable Logic Controller (PLCs). Then, two separated fortnight phases when the most critical migrations were realized, but due to the limits of the physical connection (Serial Modbus Network and Lonwork Network), only the new system could access to the infrastructure and retrieve the field data. Nevertheless, during the last phase the new SBO system proved to be complete and adequately reliable, allowing the disposal of the TAC Vista server from production.

2.2. The Schneider SBO architecture description and implementation.

The Schneider SBO software relies on two separated server for providing the core of the software management, the web user interface server and the long term archiving backend:

- **Enterprise software server (ES):** it runs the core software services for the management, configuration and backup operation of the system.
- **Report server:** it is used for archiving the long term trends of the collected variables and it also includes advanced reporting options. It uses a Microsoft SQL Server database for storing all the information.

The two servers runs on virtual Microsoft Windows machines and are used for managing the real "engines" of the systems. The real "engines" are actually three Schneider Automation Server (AS) [4] devices which are located in three strategic physical location of our buildings (i.e. the Transformers Room, the Chiller Room and the Power Room). The AS runs as a stand-alone device and can be accessed directly with a web interface in addition to the overall interface present in the Enterprise software server (ES). The AS collects data directly from the Lonwork network, from the Modbus network and provide control logics for the whole system. Figure 2 reports a schema of the logic interconnection of the three AS. As reported in the figure the Lonwork (LON) network is essentially composed of 30 programmable modules of the old TAC Vista "Xenta Family" hardware [7] connected under a separated LON network bus. These Xenta boxes provide ~300 variables or states (defined "points" in the SBO software terminology) and they supervise all the strategic elements of the LON network. The AS also uses the TCP and serial connection bus to the Modbus network providing access and control logic to all the devices that uses this protocol (e.g. the Emerson chiller and the six redundant enterprise PLCs that are used in our center) for a total of additional ~4500 SBO "points".

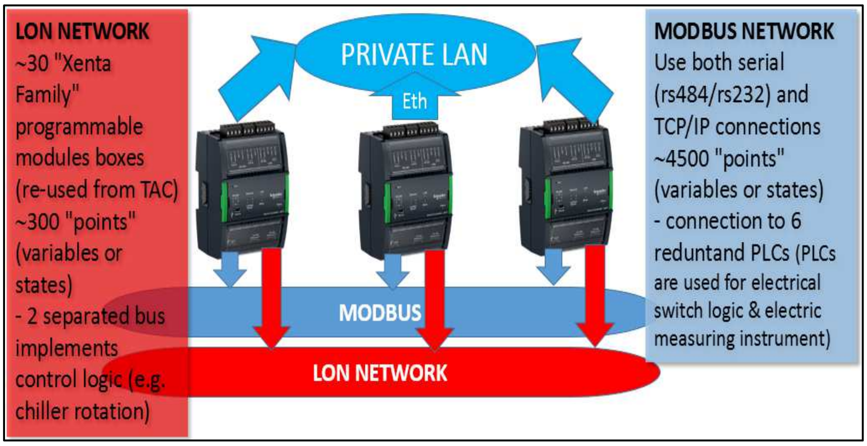

LON NETWORK
~30 "Xenta Family" programmable modules boxes (re-used from TAC)
~300 "points" (variables or states)
- 2 separated bus implements control logic (e.g. chiller rotation)
PRIVATE LAN
Eth
MODBUS
LON NETWORK
MODBUS NETWORK
Use both serial (rs484/rs232) and TCP/IP connections
~4500 "points" (variables or states)
- connection to 6 reduntand PLCs (PLCs are used for electrical switch logic & electric measuring instrument)

Figure 2: Schema of the two communication busses interconnection (Modbus and LON) of the three AS of our INFN CNAF Tier-1.

The Tier-1 SBO is integrated by another Schneider (formerly APC) software package: The StruXureware Data Center Expert (DCE). The DCE permits a fine monitoring, tuning and notification of the data center "hot-aisle" components and the metered Power Distribution Units (PDUs) that are used for the electrical distribution of many elements of the datacenter. The DCE is a stand-alone software with a proprietary client interface and with a specific Modbus TCP optional module that can export all the relevant variable to the Modbus network. It is therefore possible for the SBO software to access directly to the software variables and information and it can consequently integrate them in a unique and central entry point.

In Figure 3 three window screenshots examples of the SBO software interface are reported. The GUI interface is intuitive and provide all the necessary information, also all the schemas are fully customizable and offer interface and pop-up selector that helps the user to navigate through the different windows. User-friendly animated graphic pages are provided for all the HVAC (Heat, Ventilation and Air Conditioning) infrastructure elements, for the electrical circuit diagrams and mechanical pages, for the fire prevention system and for the flooding and water leaking sensors. The states, values and trends of power switches, electrical quantities, temperature and humidity sensors and generic devices (e.g. water pumps, air conditioners, etc...) are all clearly reported in real-time in the SBO software. The alarm notification system is also fully customizable and can provide a careful monitoring of an eventual error condition or variables threshold crossing.

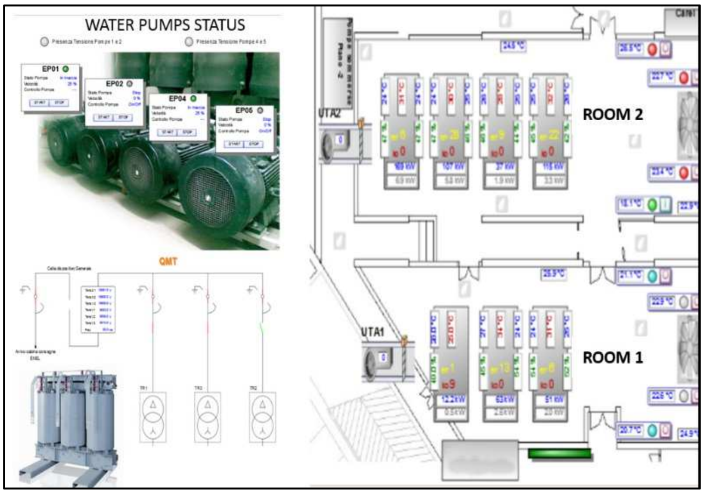


Figure 3: On the left side of the picture two separated screenshots of the SBO software windows are reported. The upper left is the pump status schema while the lower left represents the electrical sketch of the transformer room with the 3 transformers switches status reported in the electric diagram. In the right side of the picture a screenshot of the schematic diagram layout of the two IT rooms is reported. Some relevant dynamic value like the mean temperature of the hot and cold corridor are present.

3. BMS metric and measurement collection

In the SBO software the "*metrics*" represents the value of states or variable points. In general the metrics are one of the most important features in an optimal BMS since they represent the physical quantities and the actual status of devices that are present in the technical plants and in the data center. In the SBO terminology the "*logs*" and "*trends*" are the historical values of the metrics that can be collected by the software thus allowing the archiving of long term data collection of relevant values. A good metric collection is essential for different purpose, in particular it can help:

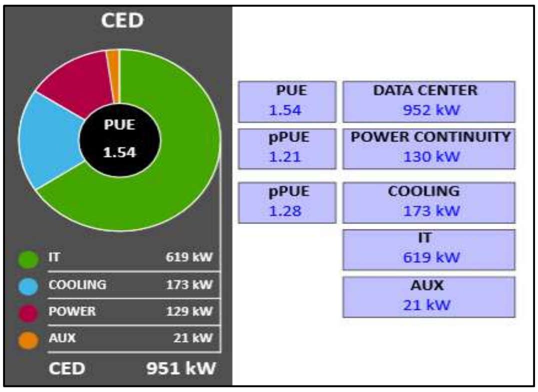
- the precise identification of periodic "hot spots" or critical area in the "hot-aisle" layout
- the data center HVAC and electrical power distribution optimization
- the data center HVAC or electrical power outage and major failure analysis and reverse engineering

On the other hand, however, we must be aware that collecting a huge amount of metrics can easily overload the system and increase the response time of the AS infrastructure. This condition should be avoided in order to provide a useful and prompt interface that takes only a couple of seconds for showing real-time data. Compared to the old TAC Vista system the SBO architecture has increased the number of collected metrics and the overall archiving duration (thanks to the separated Report server and database layout). At present over 2500 metrics are being collected with a 15 minutes granularity which gives us the opportunity to store over 10 years of trend history. The optimization of variables collection has also been implemented in order to reduce load (e.g. a power switch condition is logged only when a change occurs) and an intuitive system GUI has been directly implemented in the graphical schemas, providing a simple and fast access to the end user.

In addition to the standard variable metrics, the Power Usage Effectiveness (PUE) is calculated and collected. The PUE is a measure of how efficiently a data center uses energy; specifically, how much energy is used by the computing equipment in contrast to cooling and other overhead. We have also introduced the partialPUE (pPUE) metric in order to monitor the power demand of a specific area of the infrastructure and try to optimize it. For example and indicator of the rotary diesel power continuity energy loss can be calculated with the following formula:

$$POWER\ CONTINUITY\ pPUE = \frac{UPS\ Energy + IT\ Energy}{IT\ Energy}$$

where *UPS Energy* is the loss due to the UPS rotary and the *IT Energy* is the total amount of energy provided to the IT equipment in the 2 data center rooms. In Figure 4 an extract of the PUE report layout designed in the SBO interface is reported. The pPUE values and the total power pie chart give an immediate visual overview of the power distribution of our facility and all the relevant value can be easily extracted.



IT 619 kW
COOLING 173 kW
POWER 129 kW
AUX 21 kW
CED 951 kW

PUE 1.54
DATA CENTER 952 kW
pPUE 1.21
POWER CONTINUITY 130 kW
pPUE 1.28
COOLING 173 kW
IT 619 kW
AUX 21 kW

Figure 4: The total power and PUE graphical page of the INFN CNAF Tier-1 SBO system. In the left side of the figure a pie chart reports the ratio between the principal power distribution areas of our center: IT, COOLING (including chillers, pumps and air-handler supply), POWER (including UPS and distribution losses) and AUX (all the auxiliary electrical system that are not specifically IT equipment).

4. The SBO Webservices interface and the integration with the Tier-1 monitoring interface

One of the most important advantages of the migration to the Schneider SBO software platform is the possibility to introduce open protocols for communication and data exchange. The compatibility of SBO with the open standard Web Services was particularly interesting and we have focused our attention on a possible implementation in our software environment. The SBO is capable of both *Serve* and *Consume* Web Services operations which means that is possible to introduce external variable values in the SBO software layout (*Consume*) and conversely export variable “points” that are stored in the SBO infrastructure to some web software (*Serve*). This is accomplished using SOAP requests mechanism and standard coding. For example it will be possible in a future implementation to use the SBO *Consume* operation in order to access an external weather forecast website and subsequently adjust automatically HVAC scheduling and parameters (i.e. an anticipation of an immediate future temperature drastic change). At present, however, the most interesting feature that we decided to use was the *Serve* modality of strategic infrastructure variable in order to export them to our general Tier-1 monitoring infrastructure.

The INFN CNAF Tier-1 monitoring infrastructure has been evolving since its former implementation [8] and nowadays it consists of several open source components and home-made solution in order to keep the system flexible while reducing the maintenance effort and cost effectiveness. It is designed to give an overall view of all the resources status of our center using an immediate and “easy to navigate” interface. The design of the monitoring system can be broadly divided in backend and frontend components. The backend part is responsible for the collection of the monitoring data received from several information sources and persist them in time series databases. The data feeding relies partly on open source and partly on custom developed python scripts. The frontend consists of JavaScript web dashboard. The time series chosen for the data persistency are Whisper that is part of the Graphite software project [9] implementing an HTTP API. It is possible to interface Whisper using simple python scripts in order to send the monitoring data in a key/value format (where the key is the quantity to be measured). For the infrastructure monitoring the relevant data to be measured are the overall power consumption, the PUE, the specific “hot aisle” blocks power consumption and the cooling power as long as the room temperatures of different areas. This information is retrieved from the SBO enterprise server via a customized SOAP request which queries the SBO Web Service interface. The monitoring of the IT services is performed using the Sensu [9] monitoring platform, which provides several plugins for the most used metrics (e.g. CPU load, memory usage). The Sensu platform implements also a notification system that supports various services such as email, online chat and others. For the monitoring handled with Sensu the backend database is InfluxDB[9] which offers complementary features to Whisper and it easily integrates with Sensu. Both the infrastructures use the same frontend software layer which is Grafana[9], a powerful web-based dashboard application with a rich and intuitive web interface and fully supporting both Whisper and InfluxDB. It allows the creation of custom dashboards providing graphs or pie chart and can be interactively modified. An example of the plots of the infrastructure dashboard monitoring system web site is shown in Figure 5. The dashboard is accessible from the web from external users and is completely separated from the SBO software network so there is no possibility of interference. In such a way, some relevant parameters that are collected and calculated from the SBO software are also freely accessible from the Tier-1 dashboard web site.

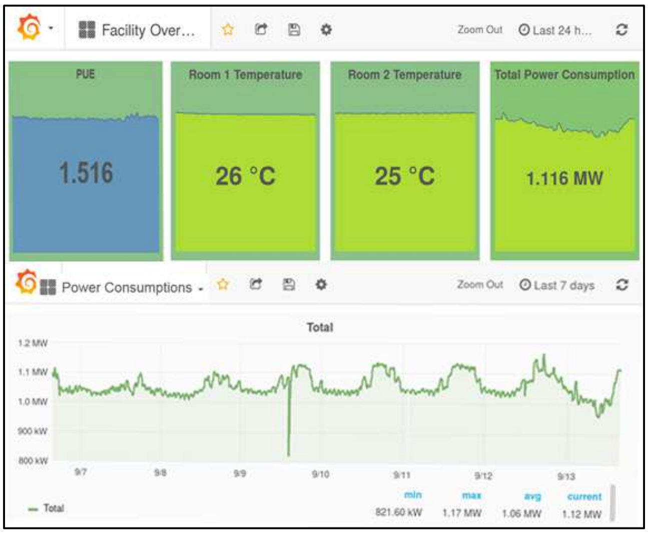


Figure 5: The INFN CNAF Tier-1 Grafana monitoring dashboard. In the upper part of the figure the actual PUE, room temperatures and total power consumption value are displayed. In the lower part of the figure the total power consumption trend over a 7 day period is reported.

5. Future development and conclusion

The compatibility of our SBO software implementation with protocols like Modbus TCP has given us the possibility to introduce new probes and elements in our Tier-1, provided that they uses standard Modbus implementation. In other words it will be granted the use of a unique system of infrastructure monitoring and a single central entry point (the SBO interface) but with the opportunity of using compatible and low-cost platforms in the process of integrating the monitoring infrastructure of our data center. At present a study for using Arduino with specific Modbus and ethernet modules (which exports data through cable or wireless TCP) as a low-cost monitoring device with compatible sensors collectors is under way. This solution could be used for additional or redundant sensor monitoring network (e.g light sensors or additional fire and water leaks detectors) with a minimum economical effort. It will also be possible to include some simple actuators (e.g. electrical relay, indicator light, acoustic buzzers etc…) that could integrate the existent PLCs logic in the data center. This could also open up to the design of "custom" sensors designed for specific needs whenever they are too expansive or currently not available on the market. An example could be the implementation of the "home made" dust sensor project, developed at CERN [10], in order to monitor the air quality within the tape library with a custom integration of the Modbus interface in the SBO software. The dust sensor currently under construction will be capable of detecting the amount of suspended dust particles in the room, taking into account moisture air and simulating the behavior of tape drive cooling fans. Some preliminary tests in this direction are actually under way. Also the Arduino temperature, humidity and light sensors are capable of recording these values in different areas of our data centre and expose them via Modbus to our SBO system. Some preliminary comparison with the values obtained with the "official" monitoring probes are currently under analysis. If the reliability and accuracy of the low-cost sensors network will prove to be satisfactory a new layer of sensor using this technology and wireless TCP Modbus connection could be realized as redundant measurement probe (that could be used in case of failure of the principal probes) and for monitoring particular areas of the data centre (e.g. service and technical alleys) that are not fully covered by our present sensors layer.

As a conclusion we can outline that the choice of migrating our BMS system to the Schneider SBO software has proven to be successful since it has showed a good reliability and great compatibility with the previous hardware and software installation. Our PUE and pPUE analysis clearly shows that we need some improvement. The seasonal PUE value of about 1.5 is a clear indication of the low energy-cost effectiveness. The pPUE analysis has shown that a more consistent improvement would be gained increasing the chillers efficiency and therefore a new project concerning a chiller technology refresh has already been developed and will be fulfilled in the next months. Also a finer granularity of rack power consumption measurement could help the optimization of electric and cooling power distribution and for this reason an increase in the number of metered PDUs with TCP Modbus support will be considered among the future hardware installations. Eventually the SBO compatibility with open standards like Web Services improved our BMS integration and "open-mindedness", and the alignment of communication protocols to not-proprietary ones, in particular Modbus, will permit the integration of different platforms (e.g. Arduino custom sensor probe) under a single BMS system.

References

[1] G. Bortolotti et al. 2012 The INFN Tier-1 J*ournal of Physics Conference Series (JPCS) IOP Publishing* Volume **396** (2012) 042016 DOI: 10.1088/1742-6596/396/4/042016

[2] P.P. Ricci et al. 2014 The INFN-CNAF Tier-1 GEMSS Mass Storage System and database facility activity *JPCS IOP Publishing* Volume **608** (2015) 012013 DOI:10.1088/1742-6596/608/1/012013

[3] N. Rasmussen and V. Avelar 2011 Deploying High-Density Pods in a Low-Density Data Center *Schneider Electric Data Center Science Center White Paper 134 Rev 2*

[4] Schneider Electric 2015 StruxureWare Building Operation Technical Reference Guide *04-16006-04-en* avaliable online https://ecobuilding.schneider-electric.com

[5] Modicon Inc. 1996 Modicon Modbus Protocol Reference Guide avaliable online www.modbus.org

[6] More info about Lonworks avaliable online under Echelon Corporation website www.echelon.com and www.lonmark.org

[7] TAC Xenta and LonMaker® release 3 Manual *0-004-7775-1 (GB) 2001-02-15* avaliable online https://ecobuilding.schneider-electric.com

[8] Stefano Antonelli et.al., 2010 INFN-CNAF Monitor and Control System JPCS IOP Publishing Volume **331** (2011) 042032 DOI:10.1088/1742-6596/331/4/042032

[9] More info avaliable online https://github.com/graphite-project/carbon https://sensuapp.org https://www.influxdata.com and http://grafana.org/

[10] More info avaliable online http://www.ohwr.org/projects/dces-dtrhf-ser1ch-v1 and CERN site shttp://cds.cern.ch/journal/CERNBulletin/2016/09/News%20Articles/2133814?ln=en

Acknowledgments

Authors wish to thank all the Schneider Electric Italy SBO and sales manager staff for the help with the planning and the Web Service access support. We also express our sincere gratitude to the Schneider certified partner "*VEM sistemi*" for all the effort they put in the migration process.