International Conference on Renewable Energies and Power Quality (ICREPQ'13)
Bilbao (Spain), 20th to 22th March, 2013
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, No.11, March 2013

New trends in datacenter energy efficiency: beyond PUE

Redondo Gil, C. [1] [2]; **Ruiz-Falco, A.** [3] [4] and **Martínez, J.M.** [5]

[1] Managing Director and Scientific Director, **Castile and León Technological Center for Supercomputing (FCSCL)**
Edificio CRAI-TIC, Campus de Vegazana s/n.
University of León. E-24071-León (Spain).
carlos.redondo@fcsc.es http://www.fcsc.es

[2] Electrical Engineering & Systems Engineering and Automatic Control Department
Faculty of Industrial and Computer Engineering, **University of León**. E-24071-León (Spain).
carlos.redondo.gil@unileon.es http://www.unileon.es

[3] Technical Director, **Castile and León Technological Center for Supercomputing (FCSCL)**
Edificio CRAI-TIC, Campus de Vegazana s/n.
University of León. E-24071-León (Spain).
antonio.ruiz-falco@fcsc.es http://www.fcsc.es

[4] Catón, S. L. C/ Parque de Las Ciencias nº 1 Of. 2-A 18006 Granada (Spain).

[5] Systems Manager, **Castile and León Technological Center for Supercomputing (FCSCL)**
Edificio CRAI-TIC, Campus de Vegazana s/n.
University of León. E-24071-León (Spain).
jose.martinez@fcsc.es http://www.fcsc.es

Abstract

The energy efficiency is one of the great challenges of the IT industry. This challenge is even greater for supercomputing centers, since they are very intensive in energy consumption. The FCSCL is a center that was born in 2008 and its infrastructures were designed to achieve higher energy efficiency. In addition, energy efficiency is the priority line of research in FCSCL.

To solve the problem of energy efficiency it is necessary to establish metrics, and organizations as The Green Grid PUE has established the PUE as the primary measure of energy efficiency.

The FCSCL has developed the MONICA project in collaboration with the company Caton S. L. and the University Jaume I, and the first objective of the project has been active PUE monitoring. The PUE is the main indicator of the energy balance of a datacenter, and it is obvious that the first thing to do is monitor it to keep it under control

But the analysis of the data acquired in MONICA project has shown that taking measures to improve the PUE can worsen the overall efficiency of the datacenter and IT equipment, as will be demonstrated in the article.

KEY WORDS

Datacenter - *Smart Data Center* - *SMART IT Infrastructure*-, Energy Efficiency, PUE, Metric, DCIM.

3. Objectives

The objectives of the article are the following:

1. We will demonstrate the need for a control system for monitoring operating parameters of a datacenter including those related with energy.
2. We will show innovative systems to display monitoring results along time for easy understanding.
3. We will demonstrate the relationship between the different variables involved.
4. Particularly important are the relationships between power consumption, CPU utilization rates, outside temperature and density and density (Kw/rack).
5. We will demonstrate the correlation between high density and efficiency. That is, how much greater is the power/rack, the greater the efficiency.
6. We will show the natural evolution of energy efficiency in datacenters, from free cooling to DCIM (DataCenter Infrastructure Management)

2. *MONICA Project*: project of monitoring and control of HPC system[1][2]

Supercomputers are very energy intensive. Large clusters are composed of hundreds or thousands of processors working in parallel, as is easily seen, the processors are the system component that consumes the most. An indicative figure is that a standard server tripled their consumption when their processors work at 100% load.

Since power consumption of the system is converted to heat in a high percentage, the problem of consumption Supercomputing Centers is really important. It is therefore imperative to design the system more energy efficient way possible: from the supercomputer configuration with the processor model to optimize the performance / cost (it is clear

that consumption is one of the variables that impact on the cost) to some ancillary infrastructure (cooling, UPS, etc.). PUE [3] [4] that allow efficient (Power Usage Effectiveness) low.

But this, that the information technology and communications is considered at present as a point of arrival, in the FCSCL is considered a starting point: on the one hand, it is useless to have an engineering facility that allows a PUE on whether use of the supercomputer is not right. In the case of FCSCL, his scheme of use and service level accords signed with users forced to have excess computing power, ie necessarily exist unused supercomputer nodes. To solve the inefficiency of having burning and unused nodes, Calendula has been implemented in a system that turns on and off nodes automatically depending on demand.

A second aspect, and more importantly, is a system that monitors all the variables that affect the consumption of cooling (system load, weather, slogans cooling system, etc..) and calculates the most suitable configuration adapted to the situation.

Both elements are implemented in *MONICA*: (1) monitoring system and (2) control of Castile and Leon Supercomputing Center.

4. Introduction to the issue of energy efficiency [5]

The new computing technologies have brought about deep changes in the datacenter in recent years. In the mid 80's it was normal for every organization to have just a single computer, with a huge cost. The relationship between the costs of purchase of the computer (typically million of dollars) and the energy cost was very unevenly because these computers just consumed a few kilowatts and moreover, were cheap. But the increased processing power and decrease in size led, in the nineties, to a shift to client-server architectures and many host servers where replaced by little but powerful "pizza box" servers. The reality is that in just a couple of decades many data centers have grown from a single computer to hundreds or even thousands of little but very powerful servers.

The problem is that these little and cheap servers are very intensive in energy consumption. A modest modern server has 12 processing cores and just costs 2.500€ (without peripherals). But this server consumption is, depending on the configuration, between 400w and 700w. Obviously, this amount will be increased with the energy necessary for the cooling. This means that the cost of electricity consumption in three years exceeds the purchase cost of the server [6]-[9].

This process of change has been especially significant in supercomputing centers and facilities for HPC in general: a Cray J916, a very powerful supercomputer in 1996, consumed less than 4kW (the cabin of processors) and other peripherals 4kW each cabin. However, nowadays-usual large clusters are composed of thousands of nodes. Process cores operate at very high frequencies and thus, the power consumption is enormous. All this bring the problem of density: since demand causes the increasing of HPC cluster nodes the power density per square meter is also growing, and it is usual to have configurations with more than 100 servers and over 1000 processing cores per rack. In the case of Calendula, the supercomputer of the FCSCL, the configuration of MPI cluster nodes is 128 8-core servers per rack. These racks reach consumption ≥40kW/rack therefore represents a significant engineering problem. It is necessary not only to refrigerate it, but also to do it efficiently [10]- [12].

5. Measurement Systems

The first problem to solve is to provide a measuring and monitoring system [13]. This requires addressing two problems:

1. Organizational problem: in many organizations it is common that IT equipment and infrastructure that support them (datacenter, chillers, power, ups, security, etc.) have different hierarchical dependencies.

2. Technical problem: each computer (hardware, networking, UPS, cooling, generator, etc) is provided with its own system of measurement (if available), but it is unusual to have a system of monitoring, control and measurement unifying the entire system, and even more unusual that this system controls IT equipment and infrastructure simultaneously.

The difficulty in establishing measurement systems is large. Normally there is a measuring instrument for the total energy consumed. But it is not common to have measurement systems to obtain accurate consumption of IT load, especially in legacy datacenters. In some cases it is possible to provide partial data, and manually (for example, there are installations in which it is possible to see the load at the output of the UPS, but it is necessary to perform a manual process of data collection).

This means that, in many organizations, it becomes common to perform a static measure of the PUE: taking manually a reading of total energy consumed and the load of the UPS output (or simply an estimate of IT load power ratings using nominal consumptions, which has a huge error rate, as discussed below).

One of the most common mistakes is to consider that the IT load is flat over time. The difference in consumption with load and without it is about three times more in a modern server. Figure 1 show the consumption of a C7000 chassis with 32 BL2x220c servers (E5450 Xeon processors with two each). On the left side of the figure without load (server on and operating system loaded), and on the right side

running a Linpack test with np=256, N=245000, NB=160, P=16, Q = 16 parameters. As shown, the consumption goes from 4kw to over 10Kw.

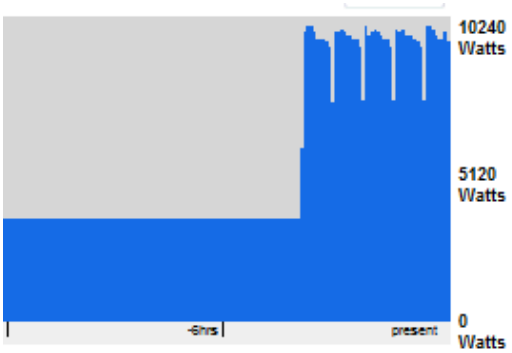

Figure 1. Power consumption with and without load

In order to achieve efficiency in a facility like the FCSCL it is essential to have a system of monitoring and dynamic control of the facility. The aim of the MONICA project is to develop a monitoring technology that integrates the whole CPD, all in one, resolving the fact that existing technologies are just specific and not all integrated. This means that technology provides monitoring of the various elements of a CPD but these tools provide a partial view of its operation. They have tools to monitor the status of power supply systems, cooling systems, fire systems, network communications and computing infrastructure and services, but each of these tools provides an overview part of CPD. The MONICA project has as one of its objectives to provide a vision of a CPD as a complete, integrated manufacturing plant, providing comprehensive information on all the parameters of its operation, and possible to establish metrics that include information collected from subsystems of different nature (outside temperature, chiller consumption, consumption per PDU, state of charge of processors, etc.).

But MONICA has not only been designed as a monitoring and data acquisition. MONICA must process the information gathered, make decisions about the operating system installation and deploy them in real time. MONICA has to make active decisions on the best possible configuration to improve efficiency (e.g., shutdown inactive nodes, moving loads between servers, configuration changes in the cooling system according to the Calendula operating, outside weather, etc.,).

To develop MONICA, it has been necessary to provide the FCSCL with the installation of the hardware elements necessary (especially data acquisition interfaces and electrical infrastructure), develop data acquisition interfaces for heterogeneous devices (SNMP, Modbus, etc.) and interpolate data not directly existing in the equipment. For example, power consumption data of heat exchangers depending on the working load.

Monica has two major differences compared to traditional monitoring systems:

First of all MONICA conceived the datacenter as an ”industrial plant” where everything is integrated. It is impossible to do the proper management of a server (for example, risk management as proposed by ISO 27001) if the monitoring system that takes no account of the elements on which it depends including the infrastructure. MONICA so not only monitors the IT hardware and software, but also controls all auxiliary infrastructures involved.

The data obtained from monitoring are not only for the management of events and alarms. MONICA should be able to make decisions and implement them automatically in real time according to predefined business rules.

One of these rules is the energy efficiency. MONICA must control all the parameters, calculating the PUE and decide the best configuration of the cooling system at a given time.

MONICA captures more than 3,000 parameters of all kind from datacenter every five minutes (state of the hardware, applications, etc.). From these parameters, it uses 150 for optimizing the PUE (power load at PDUs and UPS, main load, power load and working mode of the chillers, outlet temperature, water temperature, In- Row, parameters, etc.

With this data MONICA calculates the optimum operating model and how to set the setpoint of the key elements. The problem that exists today is that there are things that can be implemented automatically (for example, turn off or on servers) but others are not possible because there is no suitable transducer, so that in this case MONICA is unable to deploy it automatically (the most important example is the setpoint temperature water drive). It is necessary to implement these transducers to achieve full automation, which will result in a better optimization (feedback will be much faster by not needing human intervention in any case).

6. PUE is not constant

A common mistake is to think that the PUE is flat. Actually, due to increasing awareness, more organizations have been interested in knowing the PUE. The PUE, as defined, requires data collection for twelve months, and their optimization requires working proactively in times when it is too high. That is, acting in the instant PUE to avoid the peaks that raise it during the period of twelve months integration.

The first thing MONICA has to do is to calculate the instant PUE. To do this, MONICA collects all the necessary information: the total energy consumed by the FCSCL (cooling consumption, consumption of IT load, etc.) and displays this data.

The data collector takes measurements every five minutes. The following is an analysis of data collected during six months of all the parameters involved in energy efficiency, which are more than 50,000 measures of each parameter.

Although if it may be seem that the PUE remains unchanged, it is clear that it does in terms of operating parameters:

State of system load: we have seen how the differences in consumption are between unloaded servers and fully loaded (100%) servers. That is, the denominator of the equation (ITP) can vary greatly depending on time.

Figure 2 shows de IT loads distribution.

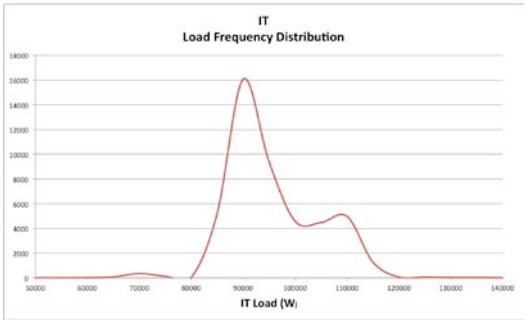

Figure 2. IT Load Distribution (mean 93kW, σ=9kW)

Weather: In an installation where there is free cooling is obvious that consumption will vary depending if it is working with free cooling or normal cooling. In the case of the FCSCL, the free cooling works when outdoor temp is below 10°C.

System Configuration Parameters: the configuration of the coolers in the computer room (in the case of FCSCL, 16 units APC InRow RC), water temperature setpoint, cold air temperature, etc.

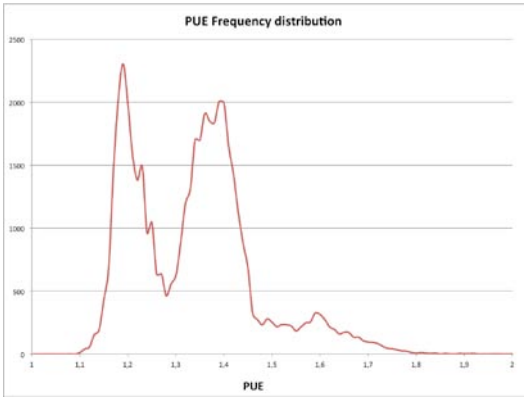

Figure 3. PUE Distribution

The Figure 3 shows the PUE Frequency distribution. The graph shows two local maximums frequencies:

1. The first one is centred in PUE=1.19. This corresponds with the entire system working on free cooling, and the PUE can vary from 1.10 to 1.28 depending on the system load. As the consumption of the cooling system is constant (pumps, fans and heat exchangers), it is optimal that the supercomputer it works at the maximum load. Thus the denominator is increased to the maximum and, therefore, the PUE improves.
2. The second one is centred at 1.39. In this case, the cooling system is working with compressors. In the case of the FCSCL installations, the chillers have four compressors each. In case of the FCSCL installation, the chillers (Emerson SBH030) have not variable speed compressors. Because of this, consumption of each compressor is fixed: 26kW. Following reasoning similar to that performed for the case of free cooling, the efficiency is better when the thermal load is close to the maximum compressor capacity.

In the case of the FCSCL[2], the minimum IT loads is, approximately, 60kW. This represents all compute nodes powered on and without load, storage systems, communications, frontends, etc. The following table shows the average PUE in different loading conditions.

Table 1. - Average PUE depending on IT load

IT Load	Average PUE	Measures
<70kW	1,459	441
>70kW <80kW	1,439	158
>80kW <90kW	1,313	21303
>90kW <100kW	1,356	13956
>100kW <110kW	1,336	9439
>110kW <120kW	1,321	1333
>120kW <130kW	1,241	87
>130kW <140kW	1,212	36

Clearly, the under loaded system is not efficient. If the IT load is below 80kW the PUE medium is above 1.4, even in the case of cooling by free cooling. At the opposite end, with the system operating at full load (over 130kW) efficiency is very high. The correlation coefficient between PUE and temperature is 0,721.

7. Energy Efficiency vs. IT Efficiency

The installation of the FCSCL is very modern and has been designed with energy efficiency criteria. Since the beginning, the FCSCL has had a low PUE, and Monica has served as a better understanding of the installation and its optimization. But it has also served to locate the problem of efficiency in its true dimension. As an example, we will consider two specific days to show the PUE evolution graphs. The two specific days are March 19th and March 29th 2012. The counters show that the PUE of March 19th was 1.250, and March 29th was 1.328. So, at a first glance, the efficiency of day 19th was greater than day 29th. Let us consider the evolution over these two days:

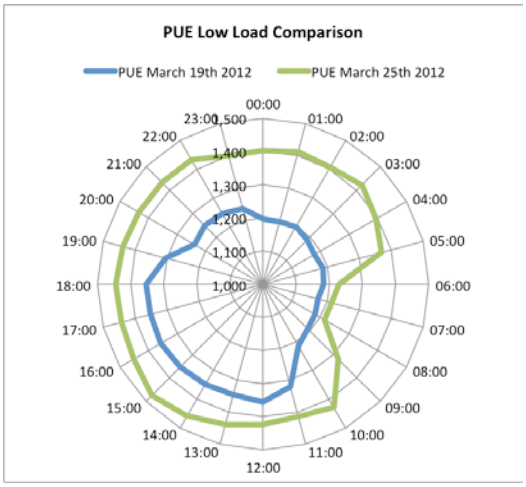

Figure 4. PUE Low load comparison

The data shows that on March 19th there were twelve hours with temperatures below six degrees, so there were many operating hours of free cooling. From these data anybody could think that day 19th has a better PUE than 25th. With this information, it seems that the efficiency on 19th was better than 29th. But there is another difference between these two days: Day 19th is a day without charge. That means, in order to perform the data collection necessary to show the idea presented in this article, days without almost any charge have been chosen. So, there have been elected the days when the workflow and SLA's allowed to delay executions, and let the supercomputer just with a minimum load (auxiliary systems, core processes, etc.). This minimum load (true bias of the system) is nearly constant in time. The rest of the systems and nodes of the clusters were on, operating system loaded and idle state.

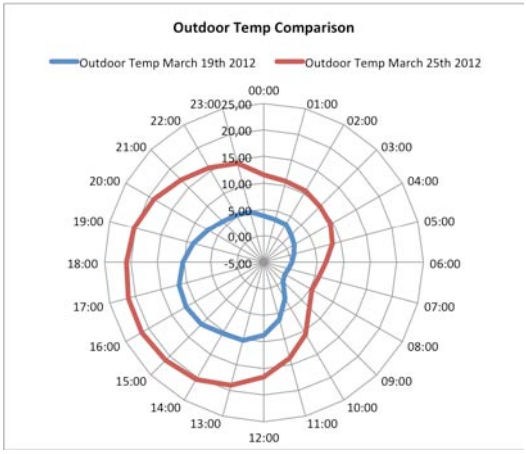

Figure 5. Outdoor temp comparison

However, the MPI cluster partition on which tests have been executed used over 33,000 hours of CPU. That is, the PUE was better on the 19th than on day 25th, but the system was more efficient on day 25th since it executed a much higher workload (50 times more) than on day 19th.

Energy efficiency is very important, and getting a low PUE is a good starting point. But do not forget that it is a measure of data center energy balance, and gives a precise idea of the quality of the engineering de- sign of it. We can neither forget that computers were designed to execute instructions. It is pointless for the data center to have a very low PUE if the computers that host do not run programs. The simile is very easy: the consumption figures of a car give an idea of its performance, but not of its efficiency because, what is more efficient, a car that consumes 4.5l of fuel to transport one person at a distance of 100km in an hour, or one that consumes 6.5 l carrying four people at a distance of 100km in 50 minutes?

8. Conclusions

The main conclusions are the following:

1. In the era of cloud computing data centers have many servers, in some cases thousands or tens of and its load varies greatly over time depending on various factors. It is therefore imperative to have monitoring systems are able to show the model in real time the status of all systems and energy efficiency.
2. The monitoring system should enable us to understand the efficiency model of each installation and its peculiarities.
3. Energy is one of the biggest costs of a data processing center, so it is very important to achieve high levels of efficiency.
4. For a datacenter reach high levels of efficiency is not enough that the auxiliary systems and especially the cooling system is efficient. It should be also an efficient IT system design.

5. To get a high levels of efficiency in a datacenter is not enough that the auxiliary systems and especially the cooling system be efficient. It should be also an efficient IT system design.
6. Besides the design is essential to have a system of monitoring and dynamic management depending on the IT load.
7. In systems equipped with free cooling cooling there is a close correlation between the outside temperature and the PUE. In the case of the FCSCL, the correlation coefficient of the measures taken (over 50,000) is 0.712.
8. The PUE also correlates with the level of IT load and density.
9. IT load greatly affects energy efficiency. But the PUE is a unit of measurement of energy efficiency. Therefore, to know if a datacenter is efficient metric must include controlling IT efficiency.
10. It is essential to standardize metrics to not only measure, but also to compare between different systems.
11. The PUE is a measure of the energy balance of a datacenter, but not of the efficiency of its usefulness.

9. Main Contributions

The main contribution of the article is the demonstration of two important facts:

1. It is necessary high density to achieve higher energy efficiency.
2. Make decisions based solely on the PUE can act against real efficiency.

This is very important, and it is logical: a datacenter is an infrastructure to house IT equipment. A measure of efficiency that ignores the way in which the IT equipment is used and its results are not a complete measure, because ignores the fundamental objective of the installation.

The conclusion is that PUE has proven to be a powerful tool to determine the energy balance of a datacenter infrastructure, but it is essential to establish new metrics to determine the utilization efficiency in IT equipment. Finally we show the proposal FCSCL for these metrics.

The consequence of this study is that modern DCIM systems must control both aspects together: datacenter infrastructures and IT equipment.

Acknowledgement

The company Caton S. L. and the University Jaume I.

References

[1] C. Redondo Gil, A. Ruiz-Falcó and J. M. Martínez, Optimization of energy consumption in HPC centers: Energy Efficiency Project of Castile and Leon Supercomputing Center FCSCL, International Conference on Renewable energies and Power Quality, (2012).

[2] Ruiz-Falcó, A., "Design of an Ultra Dense HPC Datacenter with very high Energy Efficiency" Boletín RedIris nº 90, Madrid (2011), pp. 26-31.

[3] D. Gaffney, The Business Value of PUE and Beyond, The Green Grid Forum 2012, (2012)

[4] --. "The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE". The Green Data Center, 2007. http://www.thegreengrid.org.

[5] The Green Grid, Breaking New Ground on Data Center Efficiency, The Green Grid, EEUU (2012)

[6] --. "Abordar el Reto de la Eficiencia Energética mediante las Tecnologías de la Información y la Comunicación". COM(2008) 241 final. Comunicación de la Comisión al Consejo, al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones.

[7 Brad Allenby, Darian Unger. "Information Technology Impacts on the U.S. Energy Demand Profile". E-Vision 2000 Conference.

[8] Brad Allenby. "Creating Economic, Social and Environmental Value: an Information Infrastructure Perspective". Int. Journal of Environmental Technology and Management. Vol. 7 (5/6) pp. 618–631, 2007.

[9] --. "Results of the U.S. Department of Energy and U.S. Environmental Protection Agency's National Data Center". Energy Efficiency Strategy Workshop, 2008.

http://www.energystar.gov/ia/partners/prod_development/downloads/energy_eff_data_centers_rec.pdf.

[10] Neil Rasmussen. "Implementing Energy Efficient Data Center". American Power Conversion White Paper 114, 2006. http://www.apc.com.

[11] --. "Creating the Green Data Center". http://www.adc.com.

[12] --. "Creating a Green Data Center to Help Reduce Energy Costs and Gain a Competitive Advantage". IBM Report, 2008.

http://www-935.ibm.com/services/us/cio/outsourcing/gtw03020-usen-01.pdf.

[13] Matthew L. Massie, Brent N. Chun, David E. Culler. "The Ganglia Distributed Monitoring System: Design, Implementation, and Experience". Parallel Computing, vol 30(7), pp. 817–840, 2004.