First operational experience from a compact, highly energy efficient Data Center Module

V. Acín¹, R. Cruz¹, M. Delfino², F. Martínez³, M. Rodríguez¹, P. Tallada²

Port d’Informació Científica (PIC), Universitat Autònoma de Barcelona, Edifici D, ES-08193 Bellaterra (Barcelona) Spain

E-mail: vacin@pic.es

Abstract. PIC, the Port d’Informació Científica in Barcelona, Spain has provisioned a compact, highly efficient Data Centre Module in order to expand its CPU servers at a minimal energy cost. The design aims are to build an enclosure of 30 square meters or less and equip it with commodity data centre components (for example, standard gas expansion air conditioners) which can host 80 KW of CPU servers with a PUE less than 1.7 (to be compared with PIC’s legacy computer room with an estimated PUE of 2.3). Forcing the use of commodity components has lead to an interesting design, where for example a raised floor is used more as an air duct rather than to install cables, resulting in an “air conditioner which computes”. The module is instrumented with many thermometers whose data will be used to compare to computer room simulation programs. Also, each electrical circuit has an electric meter, yielding detailed data on power consumption. The paper will present the first experience from operating the module. Although the module has a slightly different geometry from a “container”, the results can be directly applied to them.

1. Introduction

The Port d’Informació Científica (PIC) is a scientific data centre located in the campus of the Universitat Autònoma de Barcelona (UAB) in Spain. PIC is the Spanish Tier-1 centre for LHC data processing. PIC services also include the main data centre for the MAGIC-2 Telescope, a partial mirror of data from the Dark Energy Survey project, a data repository for cosmology and turbulent flow simulations and an application portal for research with magnetic resonance images of the brain.

PIC was founded in 2003 and since then its capacity has grown at a very rapid rate. In the period 2008-2010 the year-to-year growth rate was 37%. At the time of writing, the installed capacity was about 2200 cores, mostly implemented as HP Blade Centers with dual-processor, quad-core Intel “Nehalem” processors, and close to 4 Petabytes of disk servers, implemented on Sun, Data Direct and

¹Also at Institut de Física d’Altes Energies, IFAE, Edifici Cn, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain
²Also at Universitat Autònoma de Barcelona, Department of Physics, ES-08193 Bellaterra (Barcelona), Spain
³Also at Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

Published under licence by IOP Publishing Ltd
SGI hardware using disks ranging from 0.5 to 2 TB. Two tape robots from Storagetek and IBM provide tertiary storage. The cluster is linked via a redundant doublet of Arista 10 GbE wire speed fiber switches and a variety of 1 GbE copper switches (mostly used for remote management) and a Cisco 6500 central router. A variety of servers, implemented as 1U or 2U or blade machines from Dell, Fujitsu, HP and Sun, are used to implement cluster, database and Grid infrastructure services.

The PIC machine room consists of a 140 m$^2$ partition within a 300 m$^2$ traditional raised floor, high ceiling computer room built in the late 1980s. The rest of the computer room is used for general services of UAB. The initial installation of PIC in 2003 increased the IT power in the computer room from 40 to 240 KVA. Engineering studies showed that only minor modifications needed to be performed on the cooling system, as it had a large cold water generation capacity originally destined for cooling an IBM 3090 class mainframe. In the current configuration, the cold water from three chillers is fed into 4 Computer Room Air Conditioner (CRAC) units, three from the original installation and one added in 2003.

Given the ever growing power needs of the installation, a major upgrade of the power input and diesel backup was undertaken in 2008 by the UAB Architecture and Logistics Department. Two new transformers of 1 MVA each were installed for power input backed up by up to 4 500 KVA diesel generators.

2. IT Power increase: cooling and budget hit a wall

In late 2008, the UAB general services had grown to consume 80 KVA IT and PIC requested two 100 KVA IT power increases for 2009-2010 (total of 300 and 400 KVA IT in 2009 and 2010, respectively). Unfortunately, although these power increases could be provided by the upgraded power input and diesel backup, it was found that the cooling infrastructure was at its limit and that the electrical costs were sky-rocketing. Such problems were surfacing at many computing and data processing installations around the world, and continue to be one of the most important topics in the field, leading to a complete re-thinking of how to build these type of installations.

In the specific case of UAB and PIC, it was realized that the PUE$^4$ of the computer room was very high. The PUE could not be directly measured due to lack of segmentation in the power meters. A number of indirect measures lead to the value 2.3 ± 0.2. There is no single culprit for this high PUE value. Many factors contribute, from the non-optimal location of the chillers (basement rather than rooftop), the obsolete analog control system of the cooling system, the architecture of the building itself, etc.

In addition, newer equipment with higher power densities were causing hot spots at random locations in the computer room. Inlet air temperatures had to be kept rather low, around 17°C, and inlet air flow velocities high in order to remove enough energy from the room. Both of these settings, however, lower the efficiency and increase the PUE. The cooling system runs in an enthalpy range with lowered efficiency and the IT equipment spends additional energy running its fans at full speed in order to horizontally draw into the cabinets cold air with high vertical velocity coming from the floor vents.

This increased energy consumption has important budget impact. Roughly, 1 Watt-Year costs 1 Euro, so with a 400 KVA IT power and a PUE of 2.3 the electricity bill would be close to a million euros per year, a figure comparable to the yearly LHC Tier-1 equipment investment cost at PIC.

---

$^4$Power Usage Effectiveness (PUE) is a measure of how efficiently a computer data centre uses its power. A first approximation to PUE is the ratio of the total power (including cooling and other overhead) to the IT power. A lower value of PUE is better, and the minimum value is 1. The term is used rather loosely in this article, as the instrumentation and methodologies for measuring PUE in our installations are limited.
In summary, by early 2009 it became clear that implementing the PIC power upgrade in the existing computer room was impossible in the short term, and that alternatives had to be explored.

3. Solution based on commodity-component Data Centre Module
A worldwide search for alternatives lead us to understand that a new industry sector was developing around modular data centers. Development has been very fast, thanks in part to the involvement of specialized bodies such as ASHRAE, TÜV, etc. which were generating new installation and safety codes, best practices manuals, etc.

In spite of this, we realized that the emphasis of many purchases of modular data centers was either in portability (for field support of oil exploration, for example) or building re-use (for creating a standard computer room by installing a module in an underused warehouse, for example). The portability applications were dominated by the ISO shipping container format, while the building re-use applications had more flexible shapes. Both, however, had rather low power densities.

In parallel, we had identified two possible locations for a PIC Data Centre Module. One possibility was to use part of the floor of the IFAE mechanical shop where the ATLAS Tile Calorimeter had been built. The other was to use part of a basement storage room in the same building where PIC is located. The second location was chosen, as it had obvious advantages which outweighed the three main disadvantages: a slight possibility of flooding, tighter constraints on size and lack of access to install a container format solution.

In summary, the solution chosen was to build a “building-within-a-building” Data Centre Module located in the basement, with a minimal size and medium-high power density (23 kW per rack). Given the growth requirements of PIC, it was required that the whole chain of designing, building, commissioning and turning to production the module be done in 6 months. This dictated the use of many commodity components. In the following sections, we give more details.

3.1. Basic design parameters of the PIC Data Centre Module
The basic design parameters are given in Table 1. They respond to our need to quickly deploy a module which could host a large amount of CPU servers in a minimal space with much better PUE than our standard computer room. No provision was made for the module to be “people friendly”, as all equipment would be remotely controlled. In fact, as opposed to our standard computer room, the module is not classified as a work-place, as people are only present in the module for very short interventions. Given the location in a basement storage room, the module was specified to have extremely good fire containment characteristics.
| Parameter                                                      | Conditions                              | Comments                                                                 |
|---------------------------------------------------------------|-----------------------------------------|--------------------------------------------------------------------------|
| Outside dimensions, including clearances and equipment entry  | Maximum footprint 7.5m x 7.5 m Maximum height 2.5 meters | Set by basement storage room location                                     |
| False floor                                                   | Not required                            |                                                                          |
| IT electrical capacity                                        | Nominal 100 KVA protected by UPS up to 10 minutes | For micro-cut protection. Additional diesel generator backup provided at building level |
| Cooling electrical capacity                                   | Nominal 2 x 40 KW unprotected           | Short interruption can be tolerated. Additional diesel generator backup provided at building level |
| Cooling type                                                  | Commodity gas expansion air conditioners with outside units located on rooftop (4 stories up) | Outside unit location forced by UAB campus architectural considerations |
| Cooling operating range                                       | Inlet temperature 15-25°C delta-T 15-20°C | Reserve possibility to operate at higher temperatures than is customary |
| PUE                                                           | Maximum 1.7 After tuning 1.5            |                                                                          |

Table 1: Basic design parameters for the PIC Data Centre Module

3.2. Overall layout of the PIC Data Centre Module
The overall layout of the module resulting from the engineering studies is shown in Figure 1. As can be seen, the solution is in a sense an “Air Conditioning that Computes”. Figure 2 shows the location of the module within the building basement, and the 200 KVA power branch going to the main building power panel (labelled DERIVACIÓN CUADRO). The standard computer room is located above the area labelled “Area Instal·lacions”. It is well worth noting that the PIC Data Centre Module hosts 50% of the IT power hosted in the 140 m2 standard computer room in an area comparable to two public restrooms.

3.3. The AST Smart Shelter enclosure
The enclosure of the Data Centre Module was built using an AST [1] Smart Shelter [2] panel based structure, as shown in Figure 3. This structure provides several advantages: modularity, physical security, thermal and acoustic insulation, fire protected walls, and fast assembly. The structure complies with the most recent and stringent regulations. All cables and fibers are introduced into the Smart Shelter module using flexible self-sealing cable ducting devices that maintain thermal and fire isolation.
Figure 1: Overall layout of the PIC Data Centre Module holding 6 standard 19” racks. The UPS and CRAC units are labelled SAI and CLIMA, respectively. Labels L1 refer to ceiling lights. The actual module has part of the UPS located near the labels L7L3 and a curtain extends from this location to RACK6 in order to split the hot and cold aisles.

Figure 2: Layout of the PIC Data Centre Module within the building basement
Since it is assembled from panels, the Smart Shelter could be adapted to the limited space that was available, including the relatively low external height of 2.5 meters. An additional benefit is the possibility to relocate the structure if needed by disassembling and reassembling it. The whole 26 m$^2$ x 2.5 meter module was built in only one week after a material provisioning time of one month.

3.4. Environmental Monitoring and Control System
Great importance was given to automating environmental data collection and centralizing data access in order to correlate information on electricity, temperature, humidity and monitoring information from the computing equipment itself in order to maximize our understanding of the behavior of the module. In addition, we developed an automated shutdown procedure for the computers in the module in case of temperature or power problems.

The centralized monitoring console is deployed with Centreon, an open source application based on Nagios. A variety of data collection methods were used to integrate the various components, as described in the following sections.

3.5. Instrumented electrical distribution panel and Uninterruptible Power Supply (UPS)
Detailed information about power usage is obtained from a number of devices. A Circutor [3] power quality analyzer module was installed on the input and Circutor individual power meters were installed on each branch of the electrical distribution panel. As provided by the manufacturer, the analyzer and power meters display their measurements on an LCD screen or via a PC application. We developed a program that captures information from the Modbus interfaces, allowing the reading of the power analyzer and branch power meters using the TCP/IP protocol. This in turn allowed integration into the Centreon console.

3.6. Commodity cooling system
We consider to use water cooled racks in order to achieve a better efficiency but it was not possible to expand the water collector due to technical and economical issues. Because of that cooling system is composed of four Liebert HPA Condenser units located in the building rooftop (4 stories up) which dissipate the heat coming from two air-cooled direct expansion CRACs inside the module. The cooling capacity is 40 KW per unit. The system is equipped with the iCOM Control System.

The indoor units are equipped with Electronically Commutated fans and located in the hot aisle. Under the floor cold air delivery using standard CRACs was the most standard and cheapest cooling solution that could be implemented given the project time schedule, the current infrastructure in the building and the limitations of our responsibilities as tenants. Given this, a standard computer room raised floor was deployed in the module in order to form a giant air duct connecting the hot and cold aisles.

3.7. Temperature and Humidity monitoring
Temperature and humidity measurements are collected from different locations and sources in order to analyze the behavior of the module. One difficulty encountered was the lack of equipment in the market for environmental monitoring which could host a large number of sensors and report the values over SNMP, which is a common standard for data center monitoring. Typical commercial systems host one temperature, one humidity and one additional sensor hooked up to a PC which shows the information on the screen. Our colleagues at the IFAE physics institute solved this by developing a micro-processor based “slow control” module which hosts 8 industrial sensors (temperature, humidity, pressure, air flow, etc.) and reports their data over SNMP and http protocols.

In addition, most modern IT equipment comes with several temperature sensors built-in. In our case, we measure air temperatures entering and exiting each HP Blade Center using their sensors.

The CRAC units also report air temperatures and humidities through http protocols (SNMP is available as a rather expensive add-on). A simple gateway program was developed to bridge the CRAC information from the http to the SNMP protocol.

All environmental information is centralized in the Centreon system, where it can be aggregated, graphed and analyzed. An example can be seen in Figure 4.
4. Results from first operational experience

Although as of this writing the PIC Data Center Module has only been operational for a few months, there are already some interesting operational results. The results quoted correspond to the module hosting 12 fully loaded HP Blade Centers, corresponding to 1312 cores.

Power monitoring of the individual electrical branches can be easily aggregated and charted with the Centreon system, as shown in Figure 5. The way in which the HP Blade Centers perform power management is clearly seen, responding to the varying CPU load of jobs when, for example, they wait for input-output. This correlates with the fine-grained control of the CRACs which reacts to the varying heat load of the Blade Centers.
Figure 5: Centreon chart of power consumption in the PIC Data Centre Module. The orange line is the power consumed by Blade Centres while the red and purple lines are the power consumed by CRACs.

The power monitoring of all elements enables us to calculate in real time the PUE of the module, as shown in Figure 6. As can be seen in the figure, the design value of a maximum PUE of 1.7 has been achieved. Real-time PUE monitoring allows keeping a close check on the efficiency of the infrastructure and to quickly detect any hidden faults. In future, it will also allow us to perform environmental tuning of the module, for example varying air floor vents and CRAC temperature settings. In this way, we hope to lower the PUE further, hopefully reaching a value of 1.5.

Figure 6: Chart of instantaneous PUE measurements (green), total power including cooling (red) and IT power being sourced by the UPS (blue).
5. Lessons learned and outlook
The PIC Data Center Module is performing with a much better PUE than our main computer room. We have achieved our main objective, which was to increase IT power in a more economical manner, using minimal floor space and with minimal interference with the main computer room.

Some problems have arisen and we have learned some lessons for the future. Control of the CRACs has been a challenge, and it represents the most fragile part of the module. Although work is still in progress, we can already conclude that operating two CRACs in a coordinated manner in such a small volume is very difficult. One larger CRAC would be easier to control, but at the expense of operational downtime to perform maintenance.

We expect that the situation with the control of the CRACs will improve for several reasons. First, the module is only at 60% of its IT power capacity, which causes the CRACs to periodically switch off completely – as we install more IT equipment, both CRACs will always be on under the fine-grained digital control. Second, we hope to reach an agreement with the CRAC manufacturer and be able to raise the operating temperature while retaining maintenance and warranty – this will make the cooling system operate in a more efficient and easier to control enthalpy range. Finally, better isolation between the hot and cold aisles will be made. The air volume in the module is just 55 m$^3$, so high velocities are needed in order to move the necessary volumes of air. This in turn has caused problems with the curtain between hot and cold aisle (see Figure 1), which will be changed to a rigid partition with a sliding door. In addition, many small spaces and holes in the rack space need to be filled in order to prevent cold air leakage back to the hot aisle.

PIC will definitely consider building additional modules based on the experience acquired. Emphasis will be given to better designs of the overall cooling system and of cold and hot air isolation. One interesting possibility is to use chimney racks [4] and installing hot or cold air ducting to move the air directly to/from the CRACs. In this case, the CRACs could be located outside the module and a 26 m$^2$ module could host as many as 12 racks in two rows.

6. Acknowledgments
The Port d’Informació Científica (PIC) is maintained through a collaboration between the Generalitat de Catalunya, CIEMAT, IFAE and the Universitat Autònoma de Barcelona. This work was supported in part by grants FPA2007-66152-C02-00 and FPA2007-66708-C03-02 from the Ministerio de Educación y Ciencia, Spain, grants FPA2009-07474, AYA2009-13936-C06-05 and PAU CSD2007-00060 Consolider-Ingenio 2010 from the Ministerio de Ciencia e Innovación, Spain. Additional support was provided by the EU 7th Framework Programme INFRA-2007-1.2.3: e-Science Grid infrastructures Grant Agreement Number 222667, Enabling Grids for e-Science (EGEE) project.
Figure 5: Conceptual design for a future data center module using chimney racks

References
[1] AST Modular, S.L. is a Spanish company with headquarters in Barcelona which is a worldwide leader in modular data centres. See http://www.astmodular.com.

[2] The Smart Shelter is a product of AST Modular, S.L.. For more details, refer to http://www.astmodular.com/solutions/4.html

[3] Circutor is a Spanish company which manufactures electrical metering devices. See http://www.circutor.com

[4] Chimney racks are relatively new on the market. See for example http://www.42u.com/cooling/hot-aisle-containment/chatsworth-passive-cooling.htm