

MONICA, Sistema de monitorización y control inteligente de la eficiencia energética para Centros de Procesos de Datos

Eficiencia energética

La eficiencia energética es uno de los grandes retos de la industria TIC, y este reto es aún mayor en los centros de supercomputación, pues son muy intensivos en consumo de energía. Las infraestructuras de SCAYLE fueron diseñadas para alcanzar una alta eficiencia energética.

El objetivo del proyecto (TSI-080500-2011-79 de la convocatoria Avanza TIC Verdes del Ministerio de Industria, Energía y Turismo) es desarrollar un sistema de monitorización y control para el conjunto del centro de proceso de datos, que es contemplado como una planta industrial: todos sus componentes deben funcionar de una forma automatizada y con un único punto de control.

Esto obliga a romper las barreras tecnológicas actuales e integrar en un único sistema todos los componentes del centro de proceso de datos, desde la entrada de electricidad a la última aplicación. Además, MONICA no es un sistema de monitorización clásico que se limita a adquirir información en tiempo real, el verdadero objetivo es realizar un control dinámico inteligente. MONICA debe tomar decisiones de parámetros de funcionamiento para mejorar la eficiencia energética e implementarlos en tiempo real.



Sistema de Alimentación Ininterrumpida de Caléndula. © SCAYLE.

El problema de la eficiencia energética

Las nuevas tecnologías de computación han provocado cambios profundos en los centros de procesos de datos en los últimos años. A mediados de los años 80 era normal que cada organización tuviera un único computador y que éste tuviera un coste enorme. La relación entre los costes de compra del ordenador (típicamente millones de dólares) y su coste energético era muy desequilibrada, pues estos ordenadores apenas consumían unos pocos kW que, además, eran baratos.

Pero el incremento de potencia de los procesadores y su disminución de tamaño, provocaron en los años noventa un vuelco hacia arquitecturas cliente/servidor, con lo que muchos hosts (servidores) fueron sustituidos por servidores en formato "caja de pizza". La realidad es que, en apenas un par de décadas, muchos centros de proceso de datos han pasado de tener un único ordenador a tener centenares o incluso miles de servidores (el centro de proceso de datos de la SCAYLE aloja unos 500 servidores).

Un servidor modesto de hoy en día tiene 12 núcleos de proceso, cuesta apenas 2.000€ y, dependiendo de su configuración, consume entre 400 y 700W a plena carga. Esto quiere decir que el coste del consumo eléctrico en tres años supera el coste de compra del sistema.

Este proceso de cambio ha sido especialmente significativo en los centros de supercomputación e instalaciones dedicadas a HPC (High performance Computing) en general: un Cray J916 de 1996 consumía menos de 4kW (la cabina de procesadores), y otros 4kW cada cabina de periféricos. Sin embargo, en el día de hoy son habituales los grandes clústers compuestos por miles de nodos y decenas de miles de núcleos de proceso. Los núcleos de proceso funcionan a frecuencias muy elevadas y por tanto, el consumo eléctrico asociado es enorme. A todo esto se suma el problema de la densidad: a medida que la demanda provoca que los clústeres HPC tengan cada vez más nodos, la densidad eléctrica por metro cuadrado también crece, y ya es habitual tener configuraciones con más de 100 servidores y más de 1000 núcleos de proceso por rack. En el caso de

Caléndula, el superordenador de SCAYLE, la configuración del cluster MPI de nodos finos es de 128 servidores de 8 cores por rack, es decir, 1024 cores/rack. Estos racks alcanzan consumos >40Kw/rack, por lo que representan un problema de ingeniería considerable. Es necesario no sólo refrigerarlos, sino hacerlo eficientemente.

Objetivo del proyecto

El primer problema a solventar es disponer de un sistema de medida y monitorización. Esto a su vez exige abordar dos problemas:

- Problema organizativo: en muchas organizaciones es habitual que los equipos IT y las infraestructuras que soportan estos (datacenter, enfriadoras, electricidad, UPS, seguridad, etc.) tengan dependencias jerárquicas diferentes.
- Problema técnico: cada equipo (hardware, networking, UPS, enfriadora, generador, etc.) es entregado con su propio sistema de medida (si es que dispone de él), pero no es habitual disponer de un sistema de monitorización, control y medida unificado para toda la instalación, y menos aún que controle equipos IT e infraestructuras simultáneamente.

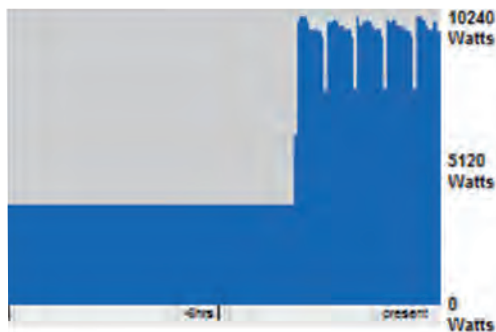
La dificultad para establecer sistemas de medida es grande. Normalmente se dispone de un instrumento de medida para el total de la energía consumida. Pero no es habitual disponer de sistemas de medida que permitan obtener con precisión el consumo de la carga IT. En algunos casos es posible disponer de datos parciales, y de forma manual (por ejemplo, existen instalaciones en las que es factible ver la carga en la salida de la UPS, pero es necesario realizar un proceso manual de toma de datos).

Esto provoca que, en muchas organizaciones, sea habitual realizar una medida estática del PUE (Power Usage Effectiveness): se toma manualmente la lectura del total de la energía consumida y de la carga a la salida UPS (Uninterruptible Power Supply), o simplemente se realiza una estimación de carga IT utilizando consumos nominales lo que tiene una enorme tasa de error, como veremos a continuación).

Uno de los errores más habituales es considerar que la carga IT es plana en el tiempo. La diferencia de

consumo con carga y sin ella es aproximadamente el triple en un servidor moderno. En la figura siguiente se puede apreciar el consumo de un chasis C7000 con 32 servidores BL2x220c (con dos procesadores Xeon E5450 cada uno). En el primer caso (izquierda de la gráfica) sin carga (servidor encendido y sistema operativo cargado), y en el segundo (derecha de la gráfica) ejecutando un test de linpack con parámetros $np=256$, $N=245.000$, $NB=160$, $P=16$, $Q=16$.

Como puede apreciarse, el consumo pasa de 4Kw a más de 10Kw.



Evolución del consumo. Proyecto MONICA, SCAYLE.

Como cualquier otro datacenter, SCAYLE necesita monitorizar sus instalaciones, para lo que abordó el desarrollo del proyecto MONICA (MONitorización Integral de CALéndula, Caléndula Integrated Monitoring System), En colaboración con la empresa Catón y el grupo HPCA de la Universidad Jaume I

MONICA tiene dos diferencias fundamentales respecto a los sistemas de monitorización tradicionales:

- La concepción tradicional del datacenter ha sido que los equipos IT dependen del departamento IT (y son monitorizados por él), mientras que las infraestructuras auxiliares (electricidad, alimentación ininterrumpida, generadores, refrigeración, seguridad, etc.) dependen del departamento de infraestructuras y mantenimiento. Sin embargo, MONICA concibe el datacenter como una "Planta Industrial" en el que todo está integrado. No se puede hacer una adecuada gestión de un servidor (por ejemplo, gestión del riesgo como la propuesta por la norma ISO 27001) si el sistema que lo monitoriza no tiene en cuenta los elementos de los que depende. Por tanto MONICA no solo

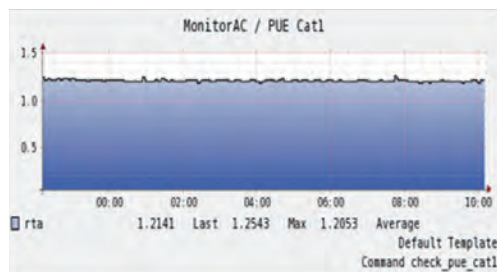
monitoriza el hardware y software IT: también controla todas las infraestructuras auxiliares.

- Los datos obtenidos de la monitorización no son únicamente para la gestión de eventos y alarmas. MONICA debe poder tomar decisiones e implantar las mismas de forma automática en tiempo real conforme a unas reglas de negocio predefinidas.

Una de estas reglas es la eficiencia energética. MONICA debe controlar todos los parámetros, calcular el PUE y decidir la mejor configuración del sistema de refrigeración en un momento dado.

Un error muy común es pensar que el PUE es plano. En la actualidad, debido a una cada vez mayor concienciación, son más las organizaciones que se han ocupado de conocer el PUE. El PUE, tal como está definido, requiere la toma de datos durante doce meses, y su optimización requiere actuar dinámicamente en los momentos en los que es alto. Es decir, actuar en el "PUE instantáneo" para evitar los picos que suben el PUE en la integración de un periodo de doce meses.

Para hacer esto, la primera cosa que tiene que hacer MONICA es calcular el "PUE instantáneo". Para hacerlo, MONICA captura toda la información necesaria (total de energía consumida por la FCSCCL, (consumo de refrigeración, consumo de la carga IT, etc.) y muestra estos datos. En la gráfica siguiente puede verse la evolución del PUE en SCAYLE durante un periodo de 12 horas:



Evolución del PUE de SCAYLE en un periodo de 12h. Proyecto MONICA, SCAYLE.

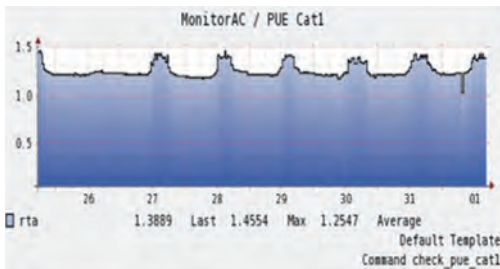
Aunque pueda parecer que el PUE no varía, es evidente que lo hace en función de los parámetros de explotación:

- Estado de carga del sistema: hemos visto cómo son las diferencias de consumo entre los servidores calculando y sin calcular. Es decir, el

denominador de la ecuación (ITP) puede variar enormemente en función del tiempo.

- Meteorología: en una instalación donde hay free cooling es obvio que el consumo variará en función de si se está haciendo uso del mismo.
- Parámetros de configuración del sistema: configuración de los intercambiadores de calor en la sala (en el caso de SCAYLE 16 unidades APC InRow RC), temperatura de consigna del agua, temperatura de consigna del aire de entrada en los racks, Δt , etc.

En la gráfica siguiente se muestra una evolución del PUE durante una semana:



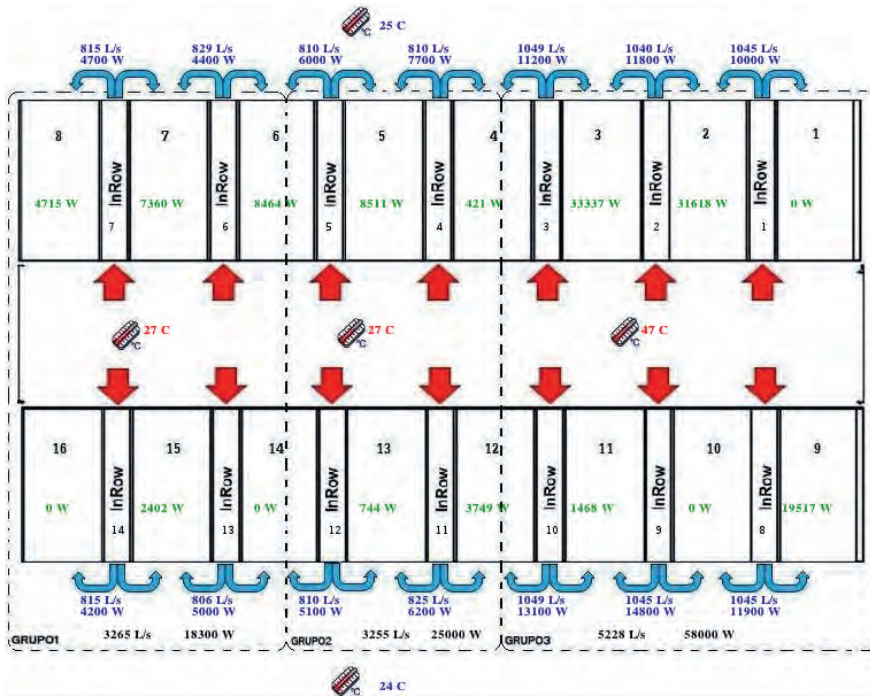
Evolución del PUE de SCAYLE en una semana.
Proyecto MONICA, SCAYLE.

En esta gráfica puede verse que en la mayoría del tiempo, el PUE instantáneo es $\approx 1,2$, mientras que en un porcentaje reducido del tiempo el PUE sube, pero estos picos $\leq 1,5$.

SCAYLE tiene una instalación moderna y energéticamente eficiente. Pero gracias a MONICA, SCAYLE ha conseguido reducir el PUE implantando políticas activas de control.

El primer requisito es la obtención de información. Para ello, SCAYLE ha instalado los transductores necesarios para la recogida de datos, que son mostrados en una página web. El primer dato mostrado es el PUE instantáneo (medidos en la UPS -PUE cat 1- y en las PDU's -PUE cat-2). El PUE se muestra en color verde si es $<1,3$, Naranja si $1,3 < PUE \leq 1,5$ y rojo si $PUE > 1,5$.

El control de MONICA no termina aquí. El datacenter de SCAYLE está constituido por un cubo con pasillo caliente cerrado compuesto por dos filas de racks que tienen intercalado un intercambiador InRow entre ellos. La siguiente página de información es el esquema de racks que puede verse a continuación:



Este esquema muestra datos importantes de funcionamiento como potencia consumida por rack (nótese que en el momento de captura de la imagen el consumo de los racks 2, 3 y 9 era de 31,6, 33,3 y 19,5Kw respectivamente), caudal de aire de los intercambiadores InRow, potencia frigorífica aportada por los grupos de InRow, etc. Los datos son actualizados cada 5 minutos.
Proyecto MONICA, SCAYLE.

MONICA captura más de 3.000 parámetros de todo tipo de Caléndula (estado del hardware, aplicaciones, etc). De estos parámetros, utiliza 100 para la optimización del PUE: calcula el modelo de funcionamiento óptimo y cómo deben configurarse los puntos de consigna de los elementos fundamentales.

El inconveniente que existe en la actualidad es que hay cosas que no se pueden implantar automáticamente, el ejemplo más importante es la temperatura de consigna de impulsión de agua.

Es necesario implantar transductores para conseguir una automatización total, lo que se traducirá en una mejor optimización (la retroalimentación será mucho más rápida).

Componentes de MONICA

Con el fin de aportar soluciones a los retos mencionados anteriormente, los objetivos científicos y tecnológicos del proyecto son los siguientes:

- Desde el punto de vista de la Monitorización.
 - Integración de todos los subsistemas que conforman un gran Centro de Proceso de (CPD) en un único sistema de monitorización.
 - Gestión jerárquica de alarmas y eventos.
 - Gestión proactiva de eventos.
 - Gestión multidimensional de los transductores.
 - Visualización tridimensional de datos.
 - Sistema experto de detección de problemas.
- Desde el punto de vista de la Eficiencia Energética:
 - MONICA debe ser capaz de una gestión eficiente de la eficiencia energética para, dado un nivel de carga concreto, optimizar el consumo energético del CPD.
 - Para poder llevar a cabo lo anterior, es evidente que MONICA debe ser capaz de llevar a cabo una distribución de carga. Es decir, MONICA debe ser capaz de tomar un control total sobre el software y el hardware para poder poner en producción la decisión tomada.
 - Por último y no menos importante, además del sentido de temperaturas a lo largo del CPD, MONICA tiene que tener un módulo de planificación consciente de los puntos calientes, que permita modelizar correctamente el comportamiento térmico del CPD.

Los objetivos científico técnicos detallados se plasmarán en un software modular. Cada módulo es autosuficiente y tiene entidad en sí mismo, por lo que puede ser utilizado por aquellos usuarios que

sólo necesiten la funcionalidad de un módulo concreto. No obstante, el proyecto se plantea como una suite con una clara complementariedad entre módulos, que:

- Interfaz de Usuario basado en un interfaz Web. El uso de MONICA está basado en web. Es el interfaz sobre el que se apoyan los diferentes componentes. A su vez, el interfaz de usuario se apoya sobre tres patas: almacenamiento de datos, módulo de configuración del sistema y módulo de mapas.
- Desarrollo de un sistema de *plugins*. Es la forma más eficiente de incorporar transductores de todo tipo y fabricante, MONICA se apoya sobre un sistema de *plugins* fácil de desarrollar para poder añadir dispositivos fácilmente al sistema. Hay tres tipos de *plugins*: pasivos (se ejecutan en el sistema principal de MONICA), activos (se ejecutan en terceros y exigen un receptor de datos en MONICA) y proxy.
- Operador Automático. Es el que incorpora la gestión proactiva, el gestor de jerarquías y el configurador de reglas de negocio.
- Visualizador 3D: Es el módulo que permite la visualización de la película 3D de los datos seleccionados.
- Módulo de Inteligencia de Negocio. Es el gestor encargado de la redistribución de la carga conforme a las reglas, gestión de la eficiencia energética, etc., son actualizados cada 5 minutos.

Líder del proyecto

SCAYLE, a asumido las funciones de coordinación y desarrollo dentro del marco del proyecto MONICA.

Financiación

El proyecto MONICA es liderado por la empresa Catón y ha sido financiado por la convocatoria Avanza TIC Verdes del Ministerio de Industria, Energía y Turismo.

