



# Diseño de un Datacenter Ultra Denso para HPC con muy alta eficiencia energética

## Design of an Ultra HPC Datacenter with very high energetic efficiency

◆ Antonio Ruiz-Falcó Rojas

### Resumen

En el presente artículo se describe el proceso de diseño del Datacenter que aloja a Caléndula, el superordenador de la Fundación Centro de Supercomputación de Castilla y León (FCSCCL) y sus infraestructuras asociadas.

Pero la FCSC no es un Centro de Supercomputación al uso. Dentro de sus objetivos está la prestación de servicios en el ámbito empresarial, lo que obliga a disponer de unas infraestructuras que permitan el cumplimiento de acuerdos de nivel de servicio (SLA's) muy exigentes.

Pero el requisito más importante del diseño del datacenter es la altísima densidad que debía alojar, así como la alta eficiencia energética alcanzada.

**Palabras clave:** Datacenter, Caléndula, superordenador, FCSCCL.

### Summary

Design of an Ultra Dense HPC Datacenter with very high energetic efficiency

In the present article there is described the design process of a Datacenter that includes Calendula, the supercomputer of the Supercomputing Center Foundation of Castilla y León (FCSCCL) and its associated infrastructures.

But FCSC is not a common Supercomputing Center. One of its objectives is offering services at business level, which forces to arrange infrastructures that allow the fulfilment of highly demanding Service Level Agreements (SLA's).

But the most important requirement of the datacenter design is the very high density that it has to host,, as well as the high energetic efficiency reached.

**Keywords:** Datacenter, Calendula, supercomputer, FCSCCL.

## 1. Introducción

El Centro de Supercomputación de Castilla y León es una Fundación Pública creada en 2008 cuyo objetivo es la prestación de servicios HPC (High Performance Computing). Partiendo desde cero ha sido necesario diseñar todo: el superordenador Caléndula, el datacenter que lo aloja y todos los sistemas auxiliares (comunicaciones, servidores de la FCSCCL, almacenamiento, etc.).

Además, como condición de partida está el espacio disponible por la FCSCCL. Como Centro de nueva creación, carece de edificio propio, y sólo cuenta con los espacios cedidos por la Universidad de León en el edificio CRAI-TIC del Campus de Vegazana. Son dos aulas de aproximadamente 90m<sup>2</sup> cada una en la planta baja del citado edificio. Por tanto, uno de los requisitos de partida es la creación de un datacenter para un gran superordenador en un espacio tan reducido. Y este espacio incluye el área de trabajo necesaria para el equipo técnico de la FCSCCL.

Al reto de crear un datacenter para un superordenador hay que sumar el hacerlo en un espacio tan reducido. Como el superordenador Caléndula iba a estar compuesto por varios clusters que totalizarían centenares de servidores y miles de cores, el siguiente reto es el de la densidad. Era necesario crear el datacenter que alojara un sistema de semejantes características en tan sólo 90m<sup>2</sup>.

◆  
En el artículo se describe el proceso de diseño del Datacenter que aloja a Caléndula, el superordenador de la FCSCCL

◆  
El Centro de Supercomputación de Castilla y León es una Fundación Pública creada en 2008 cuyo objetivo es la prestación de servicios HPC

Esto requiere trabajar con densidades desconocidas hasta el momento.

Pero a diferencia de otros centros de supercomputación públicos muy orientados a proyectos de investigación en ciencia básica, la vocación de la FCSCCL es la realización de proyectos en el ámbito empresarial, lo que exige el cumplimiento de SLA's muy exigentes. Esto condiciona enormemente las infraestructuras: aunque es evidente que lo deseable es que los superordenadores tengan un régimen de funcionamiento 24x7x365, la realidad es que ni los sistemas, ni los CPD's que los alojan, están diseñados para un funcionamiento tolerante a fallos. Sin embargo, en el caso de la FCSCCL era necesario diseñar un sistema que permitiera un funcionamiento continuo 24x7 y completamente tolerante a fallos.

Dado que Caléndula está compuesto por clusters, es factible adaptar el sistema informático de Cálculo a un entorno fault-tolerant. Pero hacer lo mismo con el Centro de Proceso de Datos implica conseguir al menos un nivel Tier-III según la norma TIA-942, siendo deseable el nivel Tier-IV.

## 2. Requisitos

Los requisitos técnicos de la instalación son ciertamente exigentes. En primer lugar, hay que tener en cuenta las características básicas del diseño de caléndula:

- Potencia de cálculo superior a 30TeraFLOPS.
- Compuesto por cuatro clusters: (MPI, Memoria Compartida, Visualización y GPU).
- Compuesto por centenares de servidores y miles de cores.
- Dos redes GbE/10Gbe (servicios auxiliares de cálculo y administración del sistema) y una red Infiniband DDR (Cálculo).

Los requisitos técnicos para el diseño del datacenter eran enormemente exigentes:

- El CPD debía ocupar un espacio máximo de 90m<sup>2</sup>.
- Jaula de seguridad con aislamiento electromagnético, aislamiento frente a agua y protección RF-120.
- Alimentación a red y grupo.
- Dos ramas de alimentación eléctrica por rack, cada una de ellas con alimentación ininterrumpida separada.
- Refrigeración preparada para densidades >40kw/rack.
- El espacio disponible impide la maniobra de racks en el CPD. Por tanto, estos se iban a dotar inicialmente, por lo que cualquier sistema que se instale en el mismo debe hacerlo en los racks existentes.

Pero las condiciones de contorno imponían unos requisitos formales que convertían el proyecto casi en misión imposible:

- Cumplimiento de estándares.
- Cumplimiento del nivel Tier-III de la norma TIA-942, es decir, debe permitir un funcionamiento tolerante a fallos y permitir cualquier actividad sobre cualquier componente de la infraestructura sin que esto suponga interrupción de ningún tipo, así como redundancia de componentes.
- Coste cerrado.
- Cumplimiento de plazos. La creación del datacenter debía estar perfectamente sincronizada con el suministro e instalación de Caléndula.
- Interlocutor único.
- Muy alta densidad.



La vocación de la FCSCCL es la realización de proyectos en el ámbito empresarial, lo que exige el cumplimiento de SLA's muy exigentes



Las condiciones de contorno imponían unos requisitos formales que convertían el proyecto casi en misión imposible



- Espacio libre para ampliar capacidad (al menos cuadruplicar la capacidad de cálculo con la tecnología existente).
- Minimalismo.

## 2.1. Alta Densidad

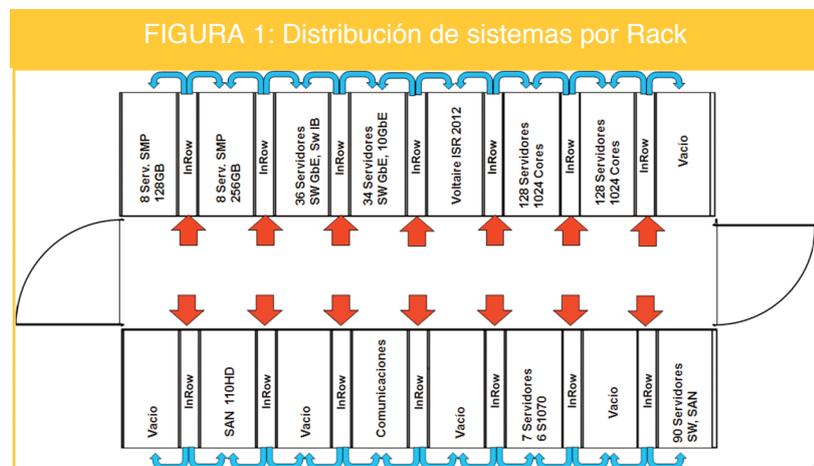
Es evidente que albergar un superordenador en un espacio tan reducido obliga forzosamente a que el sistema tenga una alta densidad. Además, y tal como se ha indicado, es evidente que no debe utilizarse todo el espacio disponible, y que hay que tener capacidad de ampliación. Por ello, y dado que Caléndula está compuesto por clusters, hay que elegir una tecnología con un alto grado de integración. Por ello, en la FCSCl existen armarios que tienen en su interior nada más y nada menos que:

◆  
Albergar un superordenador en un espacio tan reducido obliga forzosamente a que el sistema tenga una alta densidad

- 128 Servidores.
- 1024 cores.
- 2TeraBytes de Memoria RAM.
- 128 discos duros.
- 16 Switches Gigabit de 20 puertos.
- 24 Fuentes de alimentación.
- 24 Ventiladores.

El simple encendido de todos los componentes provoca un consumo aproximado en el rack de 12Kw. Pero si se envía a todos los servidores un programa que ponga al sistema al 100% de carga (por ejemplo, un test de Linpack configurado para que "sature" memoria), podemos ver en los amperímetros que el consumo en el rack sube a 44Kw. Es evidente que Linpack es un programa que, por sus especiales características, pone al sistema en condiciones extremas. Los programas de cálculo habituales, dependiendo de sus características (estados de espera, sincronización entre procesos, I/O a disco, etc.) alcanzarán niveles medio inferiores. Pero las mediciones realizadas indican tasas de consumo sostenidas superiores a 35Kw en estos racks.

◆  
Linpack es un programa que, por sus especiales características, pone al sistema en condiciones extremas



## 2.2. Eficiencia Energética

Y el requisito más importante se queda para el final, y no puede ser otro que conseguir una altísima eficiencia energética. No se trata sólo de conseguir un superordenador con una gran potencia de cálculo. Se trata de que sea muy eficiente desde el punto de vista energético.

Cada vez existe una mayor concienciación por la sociedad en general del problema de la eficiencia energética en las TIC. De hecho, basta una cifra para visualizar el alcance el problema: en 2007, el consumo de los Centros de Proceso de Datos de Estados Unidos representaba un 2% del total de la energía eléctrica consumida en el país. Esto convierte a las TI en la segunda industria más contaminante, solo por detrás de la aviación. Y de mantenerse la tendencia actual, pronto será la primera.

Por ello, el gran logro que había de conseguirse en el diseño era cumplir con todos los requisitos expuestos, es decir, lograr una infraestructura tolerante a fallos para un superordenador en un espacio reducidísimo, con una altísima densidad... y que además de muy fiable tuviera una gran eficiencia energética.

## 3. Nuevas tecnologías de Centro de Proceso de Datos

Es evidente que, después de un largo periodo en el que se han mantenido prácticamente inalterables, la tecnología para centro de procesos de datos está sufriendo una rápida transformación, que viene provocada por los sistemas de hoy en día. Como es obvio, el primer lugar en el que se manifiestan estos problemas es en los sistemas dedicados a HPC, pues tienen unos requisitos muy exigentes: tienen gran cantidad de procesadores funcionando con elevados niveles de carga. Y los procesadores son, precisamente, los componentes del sistema que más electricidad consumen, y como es obvio esta es proporcional a su nivel de carga. Un centro de supercomputación, que debería funcionar a regímenes cercanos al 100% de carga, tendrá a sus componentes en un nivel de estrés que en instalaciones normales no se produce.

En los últimos años hemos asistido a una carrera en la que se han aumentado enormemente los niveles de integración. Hace apenas cinco años eran un máximo de 40 servidores con dos procesadores de un solo núcleo los que entraban en un rack. Con los procesadores de entonces, esto suponía un consumo máximo (al 100% de carga) de unos 8Kw/rack. Pero, según el apartado anterior, se ha visto que en la FCSC los niveles de carga eléctrica por armario son muy elevados. Esto requiere solventar dos problemas:

- Por un lado, conseguir una instalación eléctrica capaz de suministrar la potencia requerida (en el caso del cluster MPI, los 44Kw/rack se distribuyen en 24 fuentes de alimentación con conector C19). Además, y para cumplir con el objetivo de alimentación redundante, hay que instalar PDU's (Power Distribution Units) trifásicas. En el caso del cluster de memoria compartida (24Kw/rack) son 32 fuentes de alimentación con conector C13.
- Por otro, el sistema de refrigeración tiene que conseguir evacuar la gran cantidad de calor generado... y además hacerlo de forma eficiente.

◆  
El requisito más importante es conseguir un superordenador muy eficiente desde el punto de vista energético

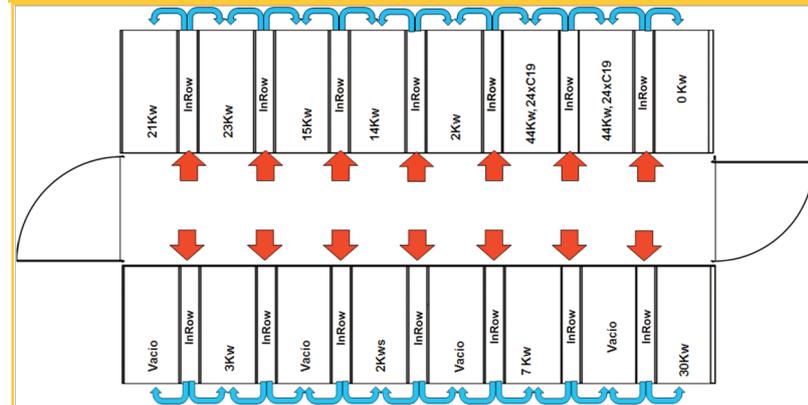
◆  
En los últimos años hemos asistido a una carrera en la que se han aumentado enormemente los niveles de integración



La instalación de la FCSCl reúne todas las innovaciones en sistemas de refrigeración de precisión y eficiencia energética

Las dos enfriadoras funcionan como un cluster activo-pasivo: en régimen normal funciona una, y la otra entra automáticamente si la otra falla

FIGURA 2: Potencias por armario al 100% de carga



### 3.1. Sistema de Refrigeración

Los sistemas de refrigeración clásicos de impulsión de aire por falso suelo, además de poco eficientes, tienen severos problemas con densidades superiores a 7Kw/rack. Además, es fácil comprobar que la refrigeración no es uniforme y que los equipos en la parte superior del rack no se refrigeran de la misma forma que los de la parte inferior. Como en el caso de un sistema de cálculo la utilización será uniforme entre todos los servidores, es lógico pensar que los peor refrigerados tendrán más estrés y, por tanto, una mayor tasa de fallos.

También es lógico pensar que el sistema será más eficiente cuanto más se aproxime el punto de intercambio de calor a la fuente, es decir, acercar los intercambiadores a los armarios.

La instalación de la FCSCl atiende a todos estos criterios y reúne todas las innovaciones en sistemas de refrigeración de precisión y eficiencia energética. El elemento refrigerante utilizado es agua. Tiene dos ventajas sobre el gas: mayor eficiencia y temperatura de trabajo (10°C) es fácil de obtener. La generación de frío se confía a dos enfriadoras (Super Chillers) Liebert SBH030. Cada una de ellas tiene una potencia frigorífica de 308Kw, es decir, una de ellas puede sobradamente con la carga del sistema. Cada enfriadora está dotada de cuatro compresores, que arrancan secuencialmente conforme a la demanda de frío, y dos bombas de circulación de agua.

El modelo SBS030 tiene además una batería de free cooling. Es decir, cuando la temperatura ambiente es inferior a 10°C (lo que en León sucede un gran número de horas al año), se paran los compresores y los ventiladores hacen circular aire a través de la batería de intercambio, enfriando el agua sin utilizar apenas electricidad.

Como es evidente, las dos enfriadoras funcionan como un cluster activo-pasivo: en régimen normal funciona una, y la otra entra automáticamente si la otra falla (rotan una vez a la semana para repartir las horas de trabajo). Pero si la temperatura ambiente es baja, antes de arrancar compresores, pueden funcionar las dos en free-cooling (cluster activo-activo). De esta forma, con temperaturas inferiores a 8°C se puede refrigerar Caléndula al 100% de carga... sin arrancar ni un solo compresor.

La evacuación de calor de la sala incluye dos innovaciones: intercambiadores de calor de columna y técnica de pasillo caliente cerrado (HACS, Hot Aisle Containment System). Anteriormente se ha mencionado que los sistemas de refrigeración por falso suelo no son válidos para altas densidades. Por tanto, entre cada dos racks se ha instalado un intercambiador de calor InRow RC (existen sistemas

similares en diversos fabricantes). La topología de distribución de agua es en estrella, por lo que en caso de fuga se aísla automáticamente al intercambiador afectado y la instalación puede seguir funcionando sin alteración.

El intercambiador de columna reúne dos grandes ventajas: por un lado tiene la capacidad suficiente para estas densidades y por otra una alta eficiencia al acercar el punto de intercambio. Además, están diseñados para trabajar en toda la altura del rack, por lo que solventa los problemas de falta de uniformidad de la solución por falso suelo.

Por otra parte, la técnica de pasillo caliente cerrado (HACS) incrementa notablemente la eficiencia. La separación de ambientes y la confinación del aire caliente, mejoran sustancialmente el funcionamiento del sistema. Basta abrir las puertas del pasillo caliente durante un rato para comprobar como se generan gran cantidad de turbulencias en el CPD y sube notablemente el consumo.

## 3.2. Instalación Eléctrica

La acometida eléctrica de la FCSC parte directamente del transformador del edificio y, con las debidas protecciones, alimentan en paralelo a dos UPS Symmetra de 160KVA. Estas UPS tienen gran redundancia (en realidad son diez módulos de 16KVA cada una) y gran eficiencia, pues desconecta los módulos no utilizados para mantener altas tasas de eficiencia. Los cuadros de distribución disponen de interruptores magnetodiferenciales superinmunizados, cuya característica principal es que permiten su conexión y reasignación de fase en caliente. Es decir, los cuadros permiten ampliación y/o manipulación sin desconexión, evidentemente manteniendo la seguridad.

## 3.3. Otros elementos

Como es fácil adivinar el CPD incluye gran cantidad de detalles que, aunque parezcan triviales, es frecuente que sean obviados en proyectos similares: pasacables antifuego, monitorización integral tridimensional, posibilidad de monitorización y actuación remota en todos los elementos, etc.

## 4. Conclusiones

Los dos grandes objetivos planteados al principio están más que conseguidos. Por un lado, la fiabilidad y tolerancia a fallos. La disponibilidad del sistema desde su instalación es superior al 99,98%, que es el objetivo de una instalación Tier-III. Por otro la eficiencia energética está más que lograda, y la FCSC se ha convertido en una instalación de referencia. Baste citar que en una auditoría externa recientemente realizada, el PUE medido ha sido 1,26 lo que representa una marca sin precedentes en un sistema de este tipo. De la misma forma, el DCIE ha sido el 79%.



La separación de ambientes y la confinación del aire caliente, mejoran sustancialmente el funcionamiento del sistema

La disponibilidad del sistema desde su instalación es superior al 99,98%, que es el objetivo de una instalación Tier-III

**Antonio Ruiz-Falcó Rojas**  
 (antonio.ruizfalco@fcsc.es)  
 Fundación Centro de Supercomputación de Castilla y León (FCSC)