PLIEGO DE PRESCRIPCIONES TÉCNICAS QUE HA DE REGIR LA LICITACIÓN, MEDIANTE PROCEDIMIENTO ABIERTO ARMONIZADO, PARA EL SUMINISTRO DE LOS DISTINTOS ELEMENTOS QUE CONFORMARÁN EL NUEVO SUPERCOMPUTADOR HYPERION DE FUNDACIÓN DONOSTIA INTERNATIONAL PHYSICS CENTER

1. OBJETO DEL CONTRATO

El objeto del contrato lo constituye el suministro de los distintos elementos que conformarán el nuevo supercomputador HYPERION de FUNDACIÓN DONOSTIA INTERNATIONAL PHYSICS CENTER, según lo previsto en el presente Pliego, en el Pliego de Cláusulas Administrativas Particulares y en su correspondiente Carátula.

2. CARACTERÍSTICAS MÍNIMAS DEL EQUIPO OBJETO DE LA ADQUISICIÓN

En el objeto del contrato, el suministro de los elementos que compondrán el nuevo supercomputador HYPERION, se pueden diferenciar cuatro partes: los racks en los que los nodos de cómputo irán instalados, el almacenamiento de scratch con el filesystem BeeGFS, las redes de comunicaciones internas de baja latencia, gestión e IPMI, y los propios nodos de cómputo. En líneas generales, el adjudicatario deberá proporcionar sólo el hardware, a no ser que se diga explícitamente lo contrario.

Los licitadores podrán solicitar una visita a la sala donde se construirá el supercomputador, y a la zona del exterior del edificio donde se instalarán las unidades exteriores de la refrigeración del supercomputador, enviando un mensaje a la dirección dipc-cc@dipc.org para convenir la fecha de la misma.

2.1. RACKS DE COMPUTACIÓN

El adjudicatario deberá suministrar íntegramente los racks que albergarán los nodos de computación (ya montados, o siendo montados dentro de la propia sala). Estos deberán ser de tipo estándar, de 42 Us, 800 cm de ancho y 1 metro de fondo. Además, el techo del rack debe ser accesible para poder introducir los cables de conexión a los switches Infiniband y Ethernet y las tomas de corriente de las PDUs (ver más adelante).

Deberá contar con ruedas lo suficientemente robustas como para poder moverlo aproximadamente unos 20 centímetros una vez cargado con los nodos de computación que

albergará en su interior, ya que para facilitar la carga de los nodos por la parte frontal o para facilitar el acceso al cableado por la parte posterior, es posible que se tenga que desplazar temporalmente en dirección perpendicular a las puertas. Debido a las dificultades de espacio con las que cuenta el DIPC en la sala que los contendrá, la refrigeración de los racks tiene que ser in situ, es decir, cada rack deberá contar con una puerta trasera refrigeradora basada en gas, que también suministrará e instalará el adjudicatario. Cada rack también incorporará las PDUs necesarias para satisfacer los requisitos que se detallarán a continuación.

2.1.1. Puertas refrigeradas

El suministro de las puertas refrigeradas, de las unidades exteriores asociadas (que se colocarán en el exterior del edificio), la instalación física de ambas, su configuración y puesta en marcha, incluida la instalación de cualquier tipo de conexión de las puertas refrigeradas a las correspondientes unidades exteriores, deberán estar incluidas en la oferta. La conexión a la red eléctrica de todos estos elementos la realizará la empresa que DIPC estime oportuna. Las puertas refrigeradas deberán cumplir con los siguientes requerimientos mínimos:

a) Las tuberías que conectan la puerta con su unidad exterior se colocarán por encima del rack, y serán lo suficientemente flexibles y tendrán la suficiente holgura como para poder mover el rack unos 20 cm, aproximadamente, en dirección perpendicular a la puerta refrigerada y en ambos sentidos, como ya se ha indicado anteriormente. Esto sirve para asegurar el espacio suficiente como para poder cargar cómodamente los nodos por la parte frontal del rack y para facilitar el acceso al cableado por la parte posterior del rack (ver figura 1).

b) Las puertas deben poder abrirse en un espacio que no comprometa la instalación futura de hasta tres filas de racks, como aparece en la figura 1 (la imagen de la figura 1 muestra lo que será el supercomputador completamente construido, dentro de dos tres años).

c) El estado de cada puerta debe ser consultable a distancia, preferentemente mediante página web, con la emisión de alarmas por correo electrónico u otro medio en caso de avería o sobrecalentamiento.

d) La potencia de las puertas refrigeradas debe ser la suficiente como para soportar la carga de los equipos que se mencionan en el apartado 2.4. El licitador tomará la decisión acerca del número de puertas que asociará a cada unidad exterior de refrigeración, y por tanto el número de unidades exteriores, de manera que se

cumpla este requisito. Todas las puertas refrigeradas han de ser capaces de soportar la misma potencia.

e) Debido a los problemas de espacio en el exterior del edificio, el adjudicatario deberá disponer las unidades exteriores en dos o más niveles. Deberá estar incluido en la oferta la construcción de la estructura física que implemente los niveles. La zona donde se ubicarán las unidades exteriores se encuentra justo al lado de la fachada del edificio que da a la sala donde se instalan los racks (a escasos metros de distancia). Se recuerda que los licitadores pueden visitar el centro para acceder al lugar donde se instalarán las unidades exteriores, tomar medidas, etc.

El licitador deberá incluir en la oferta solamente nueve de estos racks, que conformarán la fase inicial del supercomputador objeto de esta licitación, junto con las correspondientes nueve puertas traseras refrigeradas y tantas unidades exteriores como necesiten las puertas refrigeradas para poder satisfacer el apartado d) anteriormente descrito. El equipamiento del apartado 2.4 no es suficiente para llenar los nueve racks que se solicitan, pero aun así se proporcionará la puerta refrigeradora correspondiente para estos racks vacíos o semivacíos, teniendo en cuenta que todas las puertas refrigeradas deben ser capaces de soportar la misma potencia.

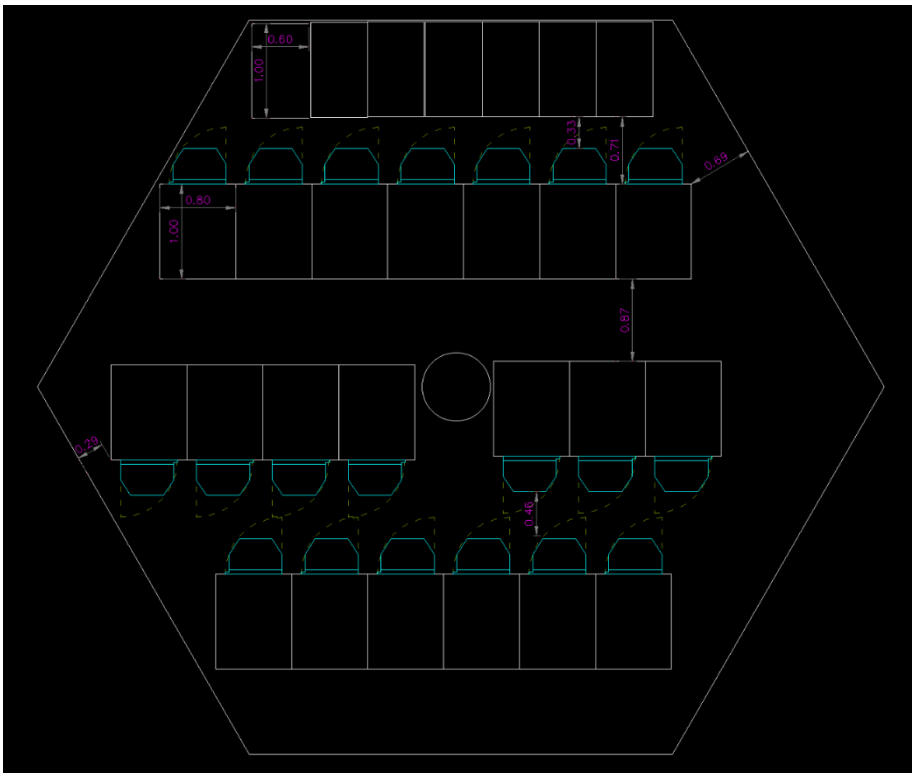


Figura 1. La figura redonda que se encuentra en la mitad de la imagen es una columna que va del suelo al techo del CPD, atravesando un falso techo que, de ser necesario, se puede eliminar.

2.1.2. PDUs

El licitador deberá ofertar, así mismo, dos PDUs verticales (0 U) por rack. Como se solicitan, en esta licitación, nueve racks, el licitador deberá ofertar 18 PDUs verticales con las siguientes características unitarias (por PDU, no por rack):

a) Ser capaz de soportar cargas de entre 0 y 12 kW. A este respecto hay que aclarar que, aunque las tomas de las dos PDUs se conectarán a otras tantas fuentes de alimentación, se trata de fuentes de alimentación redundantes. Cada nodo de computación tiene dos fuentes de alimentación y lo normal es que solo se use una de manera simultánea, por lo que cada armario no tendrá un consumo máximo de 32 kW, sino que no pasará de 12 kW. Además, el valor de 12 kW es un consumo máximo: el consumo típico por rack estará normalmente muy por debajo de ese valor.

b) Deberán ser trifásicas, conectables a la red eléctrica con toma CETAC roja.

c) Deberán ser monitorizables mediante página web, con capacidad para enviar alarmas mediante correo electrónico por averías o consumo excesivo.

d) Deberán proporcionar un mínimo de 24 tomas IEC 60320 C13/C14 ó, en cualquier caso, el tipo de tomas que permitan conectar los equipos del apartado 2.4, a los que las PDUs van a dar servicio.

El conexionado de las PDUs a la red eléctrica, las tomas CETACs rojas hembra, no forma parte del objeto de esta licitación porque dichas CETACs estarán ya disponibles en una zona cercana a la parte superior del rack.

2.2. ALMACENAMIENTO BeeGFS

Se incluirá un sistema de exportación de disco global a todos los nodos (scratch). Por compatibilidad con la administración de los otros sistemas de exportación de disco global con los que ya cuenta el DIPC, este sistema habrá de ser un filesystem paralelo BeeGFS que sea conectable a la red de producción Infiniband HDR y a la red de gestión de 100 Gbps (ver apartado 2.3). Tendrá una capacidad de almacenamiento mínima de 500 TB netos y

tendrá que ser implementado en un total de 16 servidores. Deberá incluirse la instalación del software BeeGFS sobre la distribución Rocky Linux. Los servidores que implementarán el filesystem paralelo BeeGFS deberán cumplir con los siguientes requerimientos mínimos:

- Doble fuente de alimentación.
- Dos procesadores Intel Xeon Gold 6140.
- Disco SSD para el sistema operativo de 1 TB.
- Memoria RAM de 128 GB.
- Interface Infiniband HDR.
- Interface Ethernet a 100 Gbps.
- Interface IPMI a 1 Gbps.

2.3. REDES DE COMUNICACIONES INTERNAS

El licitador deberá implementar tres redes de comunicaciones internas para el supercomputador. Por un lado, una red Infiniband HDR a 200 Gbps de baja latencia para el intercambio de mensajes MPI entre los distintos nodos que conforman el supercomputador (ver apartado 2.4) y para el acceso al servidor de almacenamiento BeeGFS del apartado 2.2. Por otro lado, deberá implementar una red de gestión Ethernet a 100 Gbps para el gestionado de todos los elementos del supercomputador, tanto los de computación (los nodos), como los de infraestructura (las puertas refrigeradas y las PDUs). Finalmente, deberá implementar una red Ethernet para IPMI a 1 Gbps para poder acceder por IPMI a todos los nodos de computación del apartado 2.4 y a los servidores que conforman el almacenamiento BeeGFS del apartado 2.2.

Los switches centrales de las tres redes (ver los siguientes apartados) se ubicarán en los dos racks que están pegados a la columna de la figura 1, por lo que estos racks necesitarán seguramente un acceso más grande por la parte superior del rack para que puedan salir cómodamente todos los cables hacia los switches hoja.

2.3.1. Red de baja latencia

El licitador deberá implementar una red Infiniband HDR a 200 Gbps. La oferta tendrá las siguientes características mínimas:

a) La topología deberá ser Fat-tree, con un factor de bloqueo mínimo de 1:2.

b) Deberá incluir tantos switches centrales (spine) como sean necesarios para conectar con 20 switches hoja (si es más de uno, se repartirán entre los dos racks centrales

de la figura 1, que son los que se encuentran pegados a la columna) y mantener el factor de bloqueo antedicho. El número 20 se deduce de todos los switches hoja que se pretenden conectar en el futuro (uno por rack).

c) Deberá incluir un switch hoja por rack suministrado, haciendo por tanto un total de nueve switches hoja, ya que en esta licitación se suministran sólo nueve racks. Aunque se tienen que suministrar todos los switches spine necesarios para poder conectar a la Fat-tree los veinte racks que formarán en el futuro el supercomputador, en esta licitación sólo se suministrarán nueve de los switches hoja, tantos como racks suministrados. Los switches hoja, que serán todos iguales, tendrán tantos puertos como sea necesario para conectar todos los elementos de los apartados 2.2 y 2.4.

d) Se incluirán todos los cables Infiniband necesarios para conectar a los switches hoja todos los elementos de los apartados 2.2 y 2.4. Aproximadamente, el 5% de los cables tendrá una longitud de 1 metro y el 95% restante de 2 metros.

e) Se incluirán todos los cables Infiniband necesarios para conectar a los switches centrales los switches hoja que se suministran (los nueve), manteniendo el factor de bloqueo de 1:2 antes comentado. Para calcular las longitudes de los cables, hay que tener en cuenta que los switches centrales se instalarán, como se ha comentado anteriormente, en los dos racks que se encuentran en contacto con la columna que aparece en la figura 1, y el resto de los racks se dispondrán en la primera fila de la figura 1 (parte superior de la imagen), completándola. Se recuerda que puede ser necesario en algún momento mover los racks en la dirección perpendicular a las puertas en ambos sentidos, por lo que la longitud de estos cables Infiniband debe ser lo suficientemente grande como para facilitar dicho movimiento.

f) La configuración de los switches, de manera que puedan interconectar todos los equipos de los apartados 2.2 y 2.4, deberá estar incluida en la oferta.

2.3.2. Red de gestión

El licitador deberá implementar una red Ethernet a 100 Gbps de gestión para interconectar los equipos de los apartados 2.1, 2.2 y 2.4. Sólo se desea interconectar por Ethernet todo el equipamiento, por lo que, si el switch Ethernet proporciona otras funcionalidades adicionales, éstas no serán tenidas en cuenta en la valoración subjetiva. La oferta tendrá las siguientes características mínimas:

a) La topología se implementará con switches en cascada tipo estrella.

b) Deberá incluir un switch central que pueda proporcionar un mínimo de 20 puertos Ethernet a 100 Gpbs hacia los switches hoja. El switch central se encontrará instalado en uno de los racks que se encuentran pegados a la columna (ver figura 1). El número 20 se deduce de todos los switches hoja que se pretenden conectar en el futuro (uno por rack).

c) Deberá incluir un switch hoja por rack suministrado, para un total de nueve switches hoja, ya que en esta licitación se suministran sólo nueve racks. Los switches hoja, que serán todos iguales, tendrán tantos puertos como sea necesario para conectar por Ethernet todo el equipamiento de los apartados 2.2 y 2.4, la puerta trasera refrigerada (ver apartado 2.1.1) y las dos PDUs (ver apartado 2.1.2). Si es necesario, esta funcionalidad se podrá dar con dos o más switches apilados.

d) Se incluirán todos los cables Ethernet necesarios para conectar a los switches hoja todo el equipamiento descrito en c). Aproximadamente, el 5% de los cables tendrá una longitud de 1 metro y el 95% restante de 2 metros.

e) Se incluirán todos los cables Ethernet necesarios para conectar al switch central los switches hoja que se suministran (los nueve). Para calcular las longitudes de los cables, hay que tener en cuenta que el switch central se encontrará, como se ha comentado anteriormente, en uno de los dos racks que se encuentran en contacto con la columna que aparece en la figura 1, y el resto de los switches se dispondrán en la primera fila de racks la figura 1 (parte superior de la imagen), uno por rack. Se recuerda que puede ser necesario en algún momento mover los racks en la dirección perpendicular a las puertas en ambos sentidos, por lo que la longitud de estos cables Ethernet debe ser lo suficientemente grande como para facilitar dicho movimiento.

f) La configuración de los switches, de manera que puedan interconectar todo el equipamiento descrito en c), deberá estar incluida en la oferta.

2.3.3. Red IPMI

El licitador deberá implementar una red Ethernet a 1 Gbps para interconectar los interfaces IPMI de los equipos de los apartados 2.2 y 2.4. La oferta tendrá las siguientes características mínimas:

a) La topología se implementará con switches en cascada tipo estrella.

b) Deberá incluir un switch central que pueda proporcionar un mínimo de 20 puertos Ethernet a 1 Gpbs hacia los switches hoja. El switch central se encontrará instalado en uno de los racks que se encuentran pegados a la columna (ver figura 1). El número 20 se deduce de todos los switches hoja que se pretenden conectar en el futuro (uno por rack).

c) Deberá incluir un switch hoja por rack suministrado, para un total de nueve switches hoja, ya que en esta licitación se suministran sólo nueve racks. Los switches hoja, que serán todos iguales, tendrán tantos puertos como sea necesario para conectar por Ethernet todo el equipamiento de los apartados 2.2 y 2.4. Si es necesario, esta funcionalidad se podrá dar con dos o más switches apilados.

d) Se incluirán todos los cables Ethernet necesarios para conectar a los switches hoja todo el equipamiento descrito en c). Aproximadamente, el 5% de los cables tendrá una longitud de 1 metro y el 95% restante de 2 metros.

e) Se incluirán todos los cables Ethernet necesarios para conectar al switch central los switches hoja que se suministran (los nueve). Para calcular las longitudes de los cables, hay que tener en cuenta que el switch central se encontrará, como se ha comentado anteriormente, en uno de los dos racks que se encuentran en contacto con la columna que aparece en la figura 1, y el resto de los switches se dispondrán en la primera fila de racks de la figura 1 (parte superior de la imagen), uno por rack. Se recuerda que puede ser necesario en algún momento mover los racks en la dirección perpendicular a las puertas en ambos sentidos, por lo que la longitud de estos cables Ethernet debe ser lo suficientemente grande como para facilitar dicho movimiento.

f) La configuración de los switches, de manera que puedan conectar todo el equipamiento descrito en c), deberá estar incluida en la oferta.

2.4. NODOS DE CÓMPUTO

En cuanto a los nodos de cómputo, serán de dos clases, los nodos ya existentes con los que cuenta ya el DIPC, que no deben ser ofertados por tanto por los licitadores, y que han de ser instalables, por parte de los técnicos del DIPC, en la infraestructura proporcionada por el adjudicatario en los apartados 2.1 y 2.3, y por otro lado los nodos nuevos, que serán proporcionados por el adjudicatario y por tanto también tendrán que ser ofertados por los licitadores.

2.4.1. Nodos de computación ya existentes

El DIPC cuenta ya con 111 nodos de cómputo con tarjeta Infiniband HDR modelo DELL Poweredge R740, de formato 2 U. El licitador deberá realizar una simulación de dónde colocar cada nodo en los nueve racks que se suministran, y tenerlo en cuenta para poder elaborar la oferta para los apartados 2.1 y 2.3. En la simulación, se ha de empezar a llenar el primer rack de la izquierda de la primera fila de racks de la figura 1 (en la parte superior) y seguir llenando el resto de racks consecutivamente hacia la derecha, intentando llenar un rack lo máximo posible antes de empezar a utilizar el siguiente, teniendo en cuenta que las PDUS son verticales (0 U) y que hay que dejar espacio para poder colocar los switches (1 switch Infiniband, un switch de gestión y un switch para IPMI por rack).

2.4.2. Nodos de computación nuevos

Los nodos de computación nuevos deberán ser enracables en los racks del apartado 2.1 y contar con las siguientes características mínimas:

- Dos procesadores Intel Xeon Gold 6248R o superior.
- 96 GB de RAM balanceados y optimizados, es decir, implementados con 12 módulos de 8 GB, de manera que todos los canales de todos los controladores de todos los sockets tengan al menos un módulo insertado (si se ofrece más RAM, ésta deberá estar también balanceada y optimizada).
- Discos duros locales de estado sólido con una capacidad neta de 1 TB. Esta capacidad neta puede estar implementada en varios discos individuales con RAID 0 o en uno solo.
- Interface Infiniband HDR a 200 Gbps.
- Interface Ethernet a 100 Gbps.
- Interface IPMI a 1 Gbps.
- La garantía de todo el equipamiento hardware relativo a los nodos de computación nuevos debe ser de un mínimo de cinco años.

El licitador deberá asegurar que los cables eléctricos suministrados con los nodos de cálculo son compatibles con las PDUs ofertadas.

En relación a qué puede considerarse un procesador superior al solicitado como mínimo, se considerará como tal aquel procesador Intel Xeon que tenga mayor número de cores (y cuyo juego de instrucciones incluya el del Xeon Gold 6248R) y/o un juego de instrucciones más avanzado, manteniéndose una velocidad de reloj y una cache iguales o superiores a las del Xeon Gold 6248R. Tampoco podrá ofertarse un procesador distinto al Intel Xeon, ya

que las aplicaciones científicas que van a ejecutar estos servidores están ya compiladas para el juego de instrucciones de un Xeon Gold 6248R.
El licitador tendrá que maximizar el número de núcleos y minimizar el tamaño de cada nodo. Así, se puntuarán positivamente el número de núcleos total ofertado y negativamente el número de Us ocupadas por cada nodo.

Como ambas cantidades son tenidas en cuenta en las puntuaciones objetivas, el número exacto de núcleos ofertado en los nodos nuevos y el número de Us que ocupa cada nodo **debe aparecer en la documentación del sobre 3 y sólo en la documentación del sobre 3**, y **nunca** en la documentación del sobre 2. Tampoco debe aparecer en el sobre 2 el número de nodos ofertado. Aunque no será necesario suministrar la instalación física ni el sistema operativo, el adjudicatario tendrá que instalar el sistema operativo Rocky Linux para verificar que el hardware lo soporta y comunicará todos los pasos realizados a los técnicos del DIPC, que replicarán esta instalación en el resto de los nodos de computación.

Importante: no se admitirán soluciones que tengan más de cuatro nodos por chasis (doble twin). De esta manera se minimiza el número de nodos afectados por la avería de un chasis.