PLIEGO DE PRESCRIPCIONES TÉCNICAS PARA LA ADQUISICIÓN DE UNA INFRAESTRUCTURA INFORMÁTICA DENOMINADA "STARLIFE" PARA EL SOPORTE DE APLICACIONES HPC (HIGH PERFORMANCE COMPUTING) Y BIGDATA PARA EL DESARROLLO DE UN CONJUNTO DE PROYECTOS ESTRATÉGICOS DENTRO DEL ÁREA DE CIENCIAS DE LA VIDA.

EXPEDIENTE CONSU02017010OP

Los requerimientos mínimos de los elementos básicos para esta infraestructura se pueden resumir en:

- Cluster de cómputo con un mínimo de 2160 cores que se puedan configurar de forma dinámica para su uso para cargas HPC o BigData
- Almacenamiento centralizado basado en filesystem paralelo con una capacidad mínima de 2 PB netos y un rendimiento de 20 GB/s
- Almacenamiento en cinta con una capacidad bruta mínima de 9PB sin compresión, junto un sistema de HSM (Hierarchical Storage Management) que realice la migración automática de datos entre el almacenamiento en disco y el sistema de almacenamiento en cinta.

Toda la infraestructura deberá ser entregada "llave en mano" tal como describe este pliego y con un mantenimiento mínimo incluido de 3 años, que corresponde con la vida útil estimada del mismo. El tiempo de entrega del hardware será de máximo un mes desde la firma del contrato y toda la infraestructura deberá estar en producción antes del 30 de enero del 2018.

El hardware entregado deberá ser nuevo y no se aceptará hardware refurbished o usado previamente.

La infraestructura antes descrita resulta de la unión de los conocimientos y recursos de tres centros de investigación punteros en Cataluña (BSC, CRG e IRB), la concentración en una única infraestructura informática para el almacenamiento y el cómputo de grandes volúmenes de datos biológicos permitirá la ejecución de proyectos estratégicos en el área de ciencias de la vida.

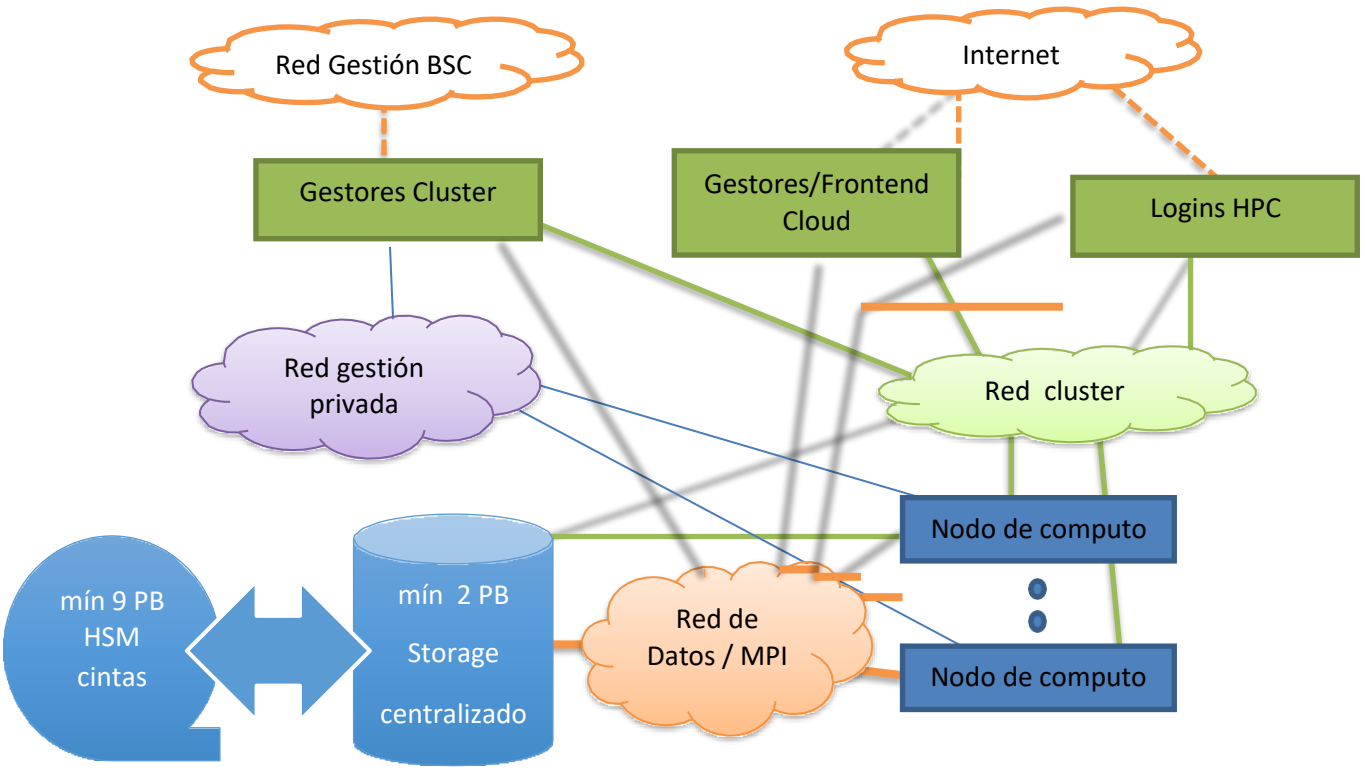
La existencia de una infraestructura así permitirá la participación en proyectos internacionales con altos requerimientos informáticos como son, EGA (European Genome Phenome Archive), MuG (Multidimensional Genomics) o ICGC/Pancancer.

Dado la diversidad de paradigmas de ingesta, almacenamiento y explotación de datos que el área de ciencias de la vida tiene, la infraestructura deberá poderse configurar y adaptar de forma dinámica para poder ejecutar y procesar eficientemente:

- Aplicaciones en un ámbito de HPC clásico (códigos basados en MPI/OpenMP con alto requerimiento de operaciones en coma flotante, red de baja latencia y filesystem paralelo de alto rendimiento)
- Aplicaciones con alta demanda de almacenamiento usando paradigmas de virtualización o containers y procesado de los mismos con aplicaciones de escalado horizontal o BigData.

En la siguiente figura se puede observar una imagen esquemática del diseño de la infraestructura con sus diversos componentes.

FEDER Fondo Europeo de Desarrollo Regional
UNIÓN EUROPEA
"Una manera de hacer Europa"
Obra Social "la Caixa"



A continuación, pasamos a realizar una breve introducción de cada uno de los 3 componentes que establece la infraestructura informática "StarLife".

Cluster de cómputo

El cluster de cómputo estará formado por un conjunto de nodos de computo con un mínimo de 2160 cores, los cuales serán de tres tipos:

- Nodos thin
  - Nodos con un mínimo de 4GB/core
- Nodos fat
  - Mínimo un 5% de los nodos totales del cluster propuestos
  - Nodos con un mínimo de 8GB/core
- Nodos hadoop
  - Un 33% de los nodos totales del cluster propuestos
  - Nodos con un mínimo de 4GB/core
  - Nodos con un almacenamiento local mínimo bruto para datos de 67 TB ampliables
  - El total de almacenamiento bruto local del conjunto de nodos hadoop deberá ser mínimo de 1000TB

Los nodos thin y fat deberán ser idénticos a excepción de los DIMMs de memoria con los que vengan configurados. Todos los nodos de cómputo de un mismo tipo deberán ser idénticos entre ellos.

El cluster deberá incluir todos los elementos administrativos necesarios para poder dar servicio a los diversos paradigmas de uso que se esperan de él:

- Servidores de administración, gestión y monitorización
- Servidores de Login y Frontend

Todos los nodos de cómputo y logins estarán conectados a través de 3 redes:

- *Red de administración privada*: Conectando todos los elementos de gestión del cluster y sólo accesible por los administradores de la infraestructura
- *Red del cluster*: Red basada en tecnología 10 Gbit Ethernet visible entre los nodos del cluster que permitirá el tráfico de los diversos servicios del cluster (NTP, DNS, tráfico control filesystem paralelo, etc.) y por otro lado donde se definirán las redes virtuales que se usarán para hablar entre las diversas máquinas virtuales o containers.
- *Red de datos y baja latencia para alto rendimiento*: Red con un ancho de banda de 100Gb por nodo para el paso del tráfico de datos desde/hacia el almacenamiento centralizado y para el paso de mensaje de las aplicaciones paralelas HPC a ejecutar

Almacenamiento centralizado

Sistema de almacenamiento centralizado con una capacidad mínima de 2 PB netos y un rendimiento mínimo de 20 GB/s, que deberá ser montado en todos los nodos de cómputo y logins de la infraestructura.

Dada la diversidad de proyectos que la infraestructura debe poder realizar y la confidencialidad de los datos médicos y biológicos que se van a albergar a ella, se propone un sistema de ficheros paralelo de alto rendimiento como por ejemplo GPFS, que ofrece las siguientes características necesarias:

- Definición de filesets (posibilidad de particionar un filesystem en varios elementos espacio de nombres administrativamente independientes, con sus propios usuarios, quotas, puntos de montaje, etc.)
- Sistema integrado de Information lifecycle management que permita la implementación de políticas y acciones según un sistema de reglas
- Sistema de metadatos distribuido y balanceado con datos en inodo
- Soporte de montaje nativo en máquinas virtuales mediante protocolo 9p

El sistema de almacenamiento centralizado no deberá tener ningún punto único de fallo.

Sistema de HSM

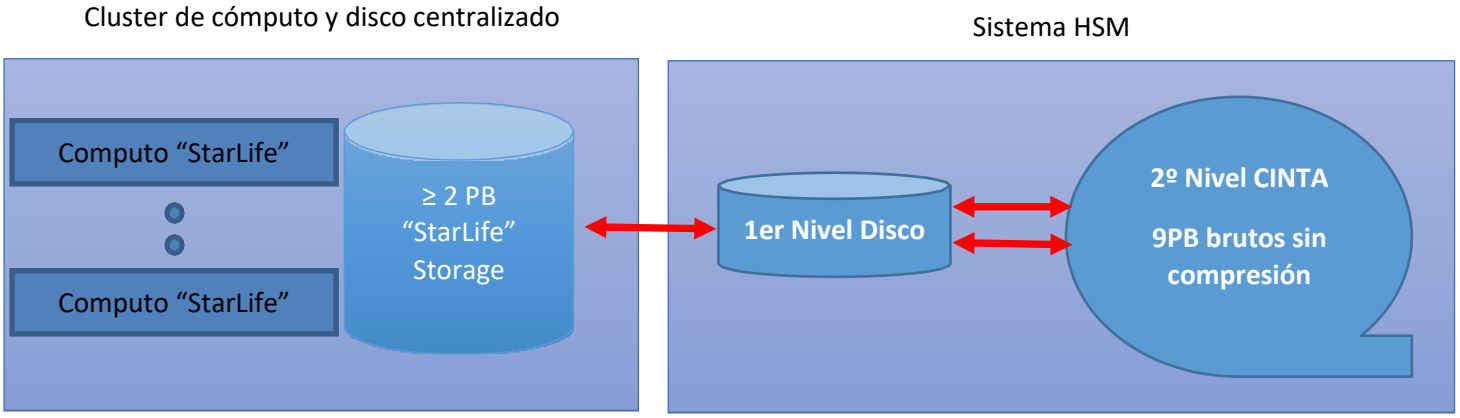
El sistema de HSM estará compuesto de 2 niveles de almacenamiento un basado en disco; y un segundo nivel basado en cinta con una capacidad mínima de 9PB bruto sin compresión.

El sistema HSM deberá mostrar toda su capacidad como un único espacio de filesystem independientemente si el dato esté en disco o cinta. Y deberá migrar datos entre las dos tecnologías de almacenamiento según las políticas a definir de acceso y reclamo.

El sistema de HSM a proveer deberá integrarse con el almacenamiento centralizado y el cluster de cómputo propuestos para el proyecto “StarLife” que este pliego describe.

Para la implementación de este sistema HSM, se propone la reutilización de parte de la infraestructura informática que se encarga de las funcionalidades de backup del BSC-CNS.

A continuación, se muestra un esquema de la solución HSM propuesta:



La adquisición de toda esta infraestructura se establece en 2 lotes:

Lote 1: Adquisición Cluster de Computo y Almacenamiento centralizado

Lote 2: Ampliacion infraestructura Robot BSC para la implementación de un sistema HSM para “StarLife”

A continuación, pasamos a describir en detalle los requerimientos mínimos y los deseables técnicos para la infraestructura informática de “StarLife”.

En las siguientes tablas, los campos se identifican por las letras R y D, cuyo significado es:

R- Representa que lo anunciado es un requerimiento que se debe cumplir en la solución presentada, en el caso de no hacerlo la oferta quedará descalificada.

D- Representa un requerimiento deseable a tener y que se valorará positivamente aquellas soluciones que lo incorporen.

Lote 1 .- Cluster de computo y almacenamiento centralizado

1.- Hardware

A continuación, pasamos a describir en detalle los requerimientos mínimos sobre el hardware que se demanda para la plataforma “StarLife” en el lote referente al cluster de cómputo y el almacenamiento centralizado.

1.1.- Hardware Cluster cómputo

Ref	Descripción
R1	Cluster de cómputo de propósito general formado por el número de nodos de cómputo necesarios para proporcionar un mínimo de 2160 cores con una potencia de cálculo mínima de 120 TFlops pico (doble precisión). Los procesadores ofertados deberán ser capaz de ejecutar códigos de alto rendimiento HPC, como soportar aceleración en virtualización y ejecución de códigos de escalado horizontal ó BigData, como por ejemplo puede ser las plataformas x86_64 o Power. No se considerará de propósito general aceleradores gráficos (GPU) o la familia de procesadores Xeon Phi (KNC,KNL, …).
R4	Se consideran 3 tipos de nodo de cómputo que podrá haber en el cluster de cómputo: - Nodo cómputo thin - Nodo cómputo fat - Nodo cómputo hadoop Cada nodo de cómputo sólo podrá pertenecer a uno de los 3 tipos. Los nodos thin y fat deberán ser idénticos a excepción de los DIMMs de memoria con los que vengan configurados. Todos los nodos de cómputo de un mismo tipo deberán ser idénticos entre ellos.
R5	Los nodos de computo de tipo thin deberán cumplir con los siguientes requisitos: - 2 Chips o sockets de propósito general por nodo - Mínimo 4 GB de memoria principal por core - La configuración de memoria presentada deberá ser equilibrada desde/hacia todos los cores de un mismo socket a la memoria (DIMMs misma velocidad y

	tamaño) y la frecuencia de acceso a memoria deberá ser la más alta que la familia de los procesadores ofertados permita. - Los buses que interconectan los sockets de un nodo deberán ser equilibrados y tener el máximo ancho de banda que la familia de los procesadores ofertados permita, la cantidad de estos buses será evaluado. - SSD con capacidad mínima de 120GB para el sistema operativo - Tarjeta para conexión a una red de baja latencia para el tráfico de datos al sistema fichero compartido y para aplicaciones MPI de un ancho de banda mínimo teórico de 100Gbits/nodo. - Una interfaz 1 Gbit Ethernet para la gestión out-of-line - Dos Interfaces 10 Gbit Ethernet conectadas a la red interna del cluster
R6	Un mínimo del 5% de nodos totales del cluster serán de tipo fat, dichos nodos serán idénticos a los nodos thin a excepción que contarán con un mínimo de 8 GB/core. A destacar que, deberán seguir cumpliendo que la configuración de memoria presentada deberá ser equilibrada desde/hacia cada socket y la frecuencia de acceso a memoria deberá ser la más alta que la familia de los procesadores ofertados permita.
R8	Un 33% de los nodos ofertados serán considerados de tipo Hadoop, dichos nodos serán igual que los nodos thin a excepción de: - Deberán albergar un mínimo de 67 TB brutos en almacenamiento local tipo SATA o NL-SAS. - Estos nodos deberán incorporar las controladoras de disco hardware internas necesarias para acceder al almacenamiento local de forma adecuada. - La ocupación en Us de cada nodo hadoop en el rack puede ser diferente a la de los nodos thin Todos los nodos hadoop deberán tener el mismo número de discos por nodo y todos los discos locales deben ser del mismo tamaño y tipo. En conjunto, el total de almacenamiento bruto local de todos los nodos hadoop deberá ser mínimo de 1000TB
R10	En caso de que los nodos estén empaquetados en un chasis: - La interfaz de gestión out-of-line podrá ser compartida por todos los nodos del chasis - En el caso de tener switch interno para alguna de las otras redes deberá cumplir el número de interfaces por nodo.
R11	Todo nodo de cómputo deberá de ofrecer los buses independientes suficientes para poder soportar las conexiones a las diversas redes que se describen anteriormente,

	sin ser ningún factor limitante.
R12	Se requiere un esquema de bloques de los nodos de cómputo ofertados con los anchos de banda entre los diferentes componentes de un nodo (máximo y útiles expresados en GB/s): procesadores, memoria, diversos buses PCI-Express, cualquier componente I/O..
R13	Se deberán proveer 4 nodos extra, a usarse 2 como login nodes para el cluster HPC, y los otros dos como nodos frontend para los nodos que funcionen como cloud. Estos 4 logins deberán ser idénticos que los nodos de cómputo thin, a excepción que deberán contar con una interfaz extra de 10Gbit ethernet para permitir la conexión al exterior del cluster. (En total 3 interfaces activas 10 Gbit Ethernet). Los cores de los login nodes no se podrán contabilizar como cores del cluster de cómputo.
R14	Todos los nodos del cluster y logins deberán disponer de un sistema de administración remoto (out-of-band), el cual debería permitir como mínimo: poder realizar el power on/off, coger la consola, monitorización del entorno (Temperatura, consumo, …), generación de alarmas, detección de problemas hardware/firmware, led de identificación, etc.
R15	Se deberán incluir todos aquellos servidores para la gestión del cluster. Entre otras cosas este hardware deberá hacerse responsable de la gestión de imágenes de sistema operativo, servicios básicos para el cluster como head servers, DHCP, NTP, DNS, …; el sistema de colas, monitorización, etc. Estos servidores deberán de disponer del hardware necesario para realizar las tareas asignadas a nivel de: cpu, memoria, almacenamiento, interfaces de red, rendimiento, etc. Se deberá rellenar la tabla número 2 para cada tipo de server de gestión ofrecido, especificando la función que realizará de las antes anunciadas. Como mínimo se propone: - 2 servidores de administración central (head nodes) - Almacenamiento compartido en una controladora externa basada en tecnología flash/SSD de 4TB mínimo neto accesible por todos los servers de administración - 2 servidores que hagan las funciones de monitorización y scheduler Todos estos servidores deberán tener una interfaz Gigabit Ethernet extra para la conexión a la red del BSC.
R17	Se deberá enrackar un Keyboard-Video-Mouse de 1U con acceso a la consola gráfica

	de todos los servers de administración mediante un switch de consolas. Todos los servidores de gestión deberán disponer de un sistema de administración remoto (out-of-band), el cual debería permitir como mínimo: poder realizar el power on/off, coger la consola gráfica en remoto via web, monitorización del entorno (Temperatura, consumo, …), generación de alarmas, detección de problemas hardware/firmware, etc.
R18	Todos los servidores y hardware que conformen la administración del cluster deberán estar completamente redundados en modo de alta disponibilidad, no deberá de existir elementos que sean un único punto de fallo.
R19	Se requiere que se rellene la siguiente tabla (Tabla 1- Descripción hardware Cluster Cómputo StarLife), en ella se especifican los valores mínimos a cumplir, y se deberá indicar los valores ofertados.

Tabla1 – Descripción hardware Cluster Computo StarLife

Concepto	Valor mínimo	Valor ofertado
Características nodo de cómputo (comunes)		
Número chips o sockets por nodo	2	
Modelo procesador		
Ancho de banda (GB/s) entre procesadores		
Cores por procesador ofertado		
Frecuencia nominal de cada core		
Frecuencia turboboost de cada core		
Frecuencia (modo vectorial) de cada core		
FLOPs pico por ciclo de cada core del procesador		
GFLOP pico por procesador		
GFLOP pico por nodo		
Consumo típico por procesador (max TDP)		
GFlop pico por procesador / max TDP		
Tecnología y frecuencia memoria RAM		
Frecuencia real funcionamiento memoria		
Capacidad almacenamiento local para SO	120	
Tecnología almacenamiento local para SO	SSD	
IOPS almacenamiento local para SO		
Interfaces 10GE incorporadas por nodo	2	
Interfaces 1 GE incorporadas por servidor	1	
Número de interfaces red baja latencia	1	
Tecnología interfaces red baja latencia		
Ancho banda a red de baja latencia	100 Gbit/s	

Nodo cómputo thin		
Número de nodos de cómputo thin		
Memoria RAM por core ofertada	4	
Memoria RAM ofrecida por nodo		
Número DIMMs y tamaño por DIMM		
Nodo cómputo fat		
Número de nodos de cómputo fat		
% número de nodos fat respecto al total	5%	
Memoria RAM por core ofertada	8	
Memoria RAM ofrecida por nodo		
Número DIMMs y tamaño por DIMM		
Nodo cómputo Hadoop		
Número de nodos de cómputo Hadoop		
Memoria RAM por core ofertada	4	
Memoria RAM ofrecida por nodo		
Número DIMMs y tamaño por DIMM		
Almacenamiento local por nodo	67 TB	
Número de discos ofertados por nodo		
Capacidad por disco ofertado		
Interfaz y tipo disco (NL-SAS/SATA) (2,5”/3,5”)		
Bahías libres por nodo Hadoop		
Modelo controladora interna RAID		
Rendimiento controladora RAID interna (GB/s)		
Características globales cluster cómputo (sin considerar los logins)		
Número nodos totales computo (thin+fat+Hadoop)		
TB Memoria RAM total (thin+fat+Hadoop)		
Almacenamiento interno para SO (thin+fat+Hadoop)		
Almacenamiento interno para datos (Hadoop)		
TFlop pico nodos computo total (thin+fat+Hadoop)	120	

Tabla2 – Descripción hardware por Servidor de gestión

Concepto	Valor mínimo	Valor ofertado
Características servidores de gestión		
Número de servidores gestión tipo A		
Servicio proporcionado servidor tipo A		
Número y modelo de procesador		
Memoria RAM		
Configuración DIMMs por servidor		
Almacenamiento compartido para la gestión del cluster (si aplica)		
Almacenamiento interno por servidor (# discos, tamaño y tecnología)		
Controladora RAID (si aplica)		
Interfaces 1 Gbit Ethernet por servidor		
Interfaces 10 Gbit Ethernet por servidor		
Interfaces 40 Gbit Ethernet por servidor		

1.2.- Hardware Almacenamiento compartido

Ref	Descripción
R1	Se deberá de proporcionar un sistema de almacenamiento centralizado del proyecto “StarLife” que debe ofrecer una capacidad mínima neta de 2 PB (Petabytes) y un rendimiento mínimo en lecturas y escrituras de 20 GB/s.
R3	Se propone un filesystem paralelo basado en tecnología GPFS o similar que tenga las diversas funcionalidades avanzadas: - Definición de filesets para poder particionar un filesystem en varios elementos y espacio de nombres administrativamente independeintes, con sus propios usuarios quotas, y puntos de montaje - Definición de storage pools - Filesystem con sistema integrado de ILM (Information Lifecycle Management) que permita definir políticas de migrado o gestión de los datos - Sistema de metadatos distribuido y balanceado con data in-inode - Soporte de montaje del filesystem o de directorios del mismo en modo pass-through mediante protocolo 9p - Incorporación módulos de integración con Hadoop como object store - Posibilidad de conexión de filesystems de diferentes clusters (montaje multi-cluster)
R4	La solución propuesta deberá ser modular, todos los módulos de datos presentados deberán ser idénticos. De tal manera que pueda ampliar el almacenamiento centralizado añadiendo módulos, ofreciendo de forma proporcional más capacidad y más rendimiento. Un módulo deberá estar formado por: - Un mínimo de 2 servidores que funcionarán como NSD servers, con las conexiones necesarias hacia las redes del cluster de computo (40 GbE y baja latencia) como las conexiones hacia el almacenamiento que esos servers sirvan. - Cabina/s de almacenamiento que se conectarán a los servidores del módulo, y que se puedan configurar con algún sistema de protección hardware como puede ser: DDP,Declustered RAID, RAID6.

Ref	Descripción
R7	Todos los elementos presentados deberán ser redundantes, y no deberán tener un único punto de fallo. Concretamente todos los elementos deberán llevar fuente redundantes pudiendo funcionar con la mitad de ellas sin pérdida de rendimiento.
R8	Se deberá (si fuera necesario) ampliar los servidores de gestión del cluster de cómputo para proveer todos aquellos servidores y elementos necesarios (hardware/software) para la gestión del almacenamiento centralizado. En el caso, de usar alguno de los servidores de gestión del cluster de cómputo, indicar cuales y que función desarrollarían.
R9	Los servidores proporcionados deberán disponer de una metodología de administración out-of-line: consola remota, power on y off, monitorización de alarmas y temperaturas, etc. que deberá conectarse a la red de gestión privada del cluster de cómputo.

Tabla3 – Descripción hardware Almacenamiento centralizado

Concepto	Valor mínimo	Valor ofertado
Servidor módulo datos (metadatos, si se ofrece)		
Número servidores por módulo datos	2	
Modelo servidor		
Número chips o sockets por nodo		
Modelo procesador		
Ancho de banda (GB/s) entre procesadores		
Cores por procesador ofertado		
Frecuencia nominal de cada core		
Memoria RAM por servidor		
Tecnología y frecuencia memoria RAM		
Frecuencia real funcionamiento memoria		
Número y configuración Discos locales para SO	2 en RAID1	
Capacidad almacenamiento local para SO		
Tecnología almacenamiento local para SO	SAS/SSD	
Número de puertos PCI-Express x16 por servidor		
Número de puertos PCI-Express x8 por servidor		
Tipo interfaz de conexión a expansiones	SAS/FC	
Bandwith por link de conexión a expansiones (Gbits)		
Número de tarjetas de conexión a expansiones		
Núero de puertos por tarjeta de conexión a exp.		
Interfaces 40GE incorporadas por server	1	
Interfaces 10GE incorporadas por nodo		
Interfaces 1 GE incorporadas por servidor		
Número de interfaces red baja latencia	1	
Tecnología interfaces red baja latencia		
Ancho banda a red de baja latencia	100 Gbit/s	
Almacenamiento módulo datos		
Número de expansiones/controladora por modulo		
Modelo expansión/controladora		

Rendimiento por controladora esperado (GB/s)		
Número puertos a host por controladora/expansión		
Número total de discos ofertados por módulo		
Tecnología discos ofertados	NL-SAS/SATA	
Capacidad por disco duro		
Tecnología de protección hardware propuesta		
Configuración de protección hardware propuesta		
Espacio neto por módulo		
Almacenamiento módulo metadatos (si se ofrece mejora)		
Número de expansiones/controladora por modulo		
Modelo expansión/controladora		
Rendimiento por controladora esperado (GB/s)		
IOPS lectura 4KB		
IOPS escritura 4KB		
Número puertos a host por controladora/exapansión		
Número total de discos ofertados por módulo		
Tecnología discos ofertados	SSD/Flash	
Capacidad por disco ofertado		
Tecnología de protección hardware propuesta		
Configuración de protección hardware propuesta		
Espacio neto por módulo metadatos		
Características globales almacenamiento centralizado		
Número módulos ofertados datos	2	
Número módulos ofertados metadatos	0	
Capacidad total neta almacenamiento (PB)	2	
Rendimiento total esperado almacenamiento (GB/s)	20	
Ocupacion en rack U’s por módulo datos		

1.3.- Switches y redes

Todos los nodos de cómputo y almacenamiento centralizado estarán conectados a través de 3 redes disjuntas físicamente, con lo que se deberá proveer de todos los elementos para su formación:

- *Red de administración privada*: Conectando todos los elementos de gestión del cluster y sólo accesible por los administradores de la infraestructura
- *Red del cluster*: Red basada en tecnología 10 Gbit Ethernet visible entre los nodos del cluster que permitirá el tráfico de los diversos servicios del cluster (NTP, DNS, tráfico control filesystem paralelo, etc.) y por otro lado donde se definirán las redes virtuales que se usarán para hablar entre las diversas máquinas virtuales o containers.
- *Red de datos y baja latencia para alto rendimiento*: Red con un ancho de banda de 100Gb por nodo para el paso del tráfico de datos desde/hacia el almacenamiento centralizado y para el paso de mensaje de las aplicaciones paralelas HPC a ejecutar

A continuación, se detallan los requisitos comunes para todas las redes del cluster de cómputo y almacenamiento centralizado "StarLife" y en las consiguientes tablas los requisitos específicos para cada una de las redes.

Ref	Descripción
R1	Se deberán de proveer de esquemas de conexionado físico de cada una de las redes que conforman el cluster y el almacenamiento centralizado de la infraestructura "StarLife". También se debe describir el ancho de banda disponible a cada nivel de las redes y la latencia introducida por cada elemento hardware. Cada una de las redes descritas debe ser completamente disjunta a nivel físico.
R2	Todos los switches de cualquier red deberán tener doble fuente de alimentación, y redundancia a nivel de ventiladores. Todos estos componentes deberán ser modulares y poderse cambiar en caliente, sin la parada del switch en cuestión. Si dentro de una misma red se proveen switches de fabricantes diferentes y se detecta cualquier incompatibilidad a la hora de conectarlos entre ellos (GBIC, fibra, limitación funcionalidades, rendimiento), el licitante deberá sustituir los switches necesarios para que todos sean del mismo fabricante para eliminar la incompatibilidad.
R3	Para cada una de las redes y una vez todos los componentes estén conectados deberá existir un 10% de puertos libres físicos (sin break-outs) en cualquiera de los switches.
R4	Se requiere que se rellene la tabla (Tabla 4- Descripción hardware Switches y redes), en ella se especifican los valores mínimos a cumplir. En el caso de proporcionar más de un tipo de switch por red, se deberán rellenar los datos de la tabla 4 por cada tipo de switch proporcionado.
R6	Todos los switches de segundo nivel de cualquiera de las redes (en el caso que sea necesario) deberán ser redundantes entre ellos, pudiendo evitar cualquier punto único de fallo. Debería poder caer un equipo y realizar su sustitución sin ningún tipo de corte o afectación.

R7	En todas las redes que hayan conexiones de velocidades diferentes, los switches deberán incorporar los buffers necesarios para ofrecer los rendimientos line-rate entre las diferentes velocidades.

Ref	Descripción
Red cluster	
R8	Se deberá proveer del hardware necesario (switches, cables, etc.) para poder establecer la red interna del cluster con tecnología 10 Gigabit Ethernet. Todos los cables y fibras de esta red física que vayan a la misma velocidad deberán ser del mismo color y de un color diferente a las otras redes de la máquina, de tal manera que puedan distinguirse visualmente.
R9	Todos los puertos de cada tipo de un mismo switch deberán ser line-rate entre ellos sin ningún tipo de sobre-suscripción.
R10	Requerimientos de funcionalidades de los switches de esta red: - Soporte Jumbo Frames (MTU > 9000) - Line-rate Nivel 2 switching - Definición de Access-list a nivel 2 y nivel 3 - Spaning-tree (MSTP y RSTP) - Capacidad para filtrar los paquetes BPDU a nivel de puerto físico del equipo - Port mirroring - Broadcast storm control - QoS - Snmp - SSH - Minimum 256 VLANs - LACP (Soporte hash LACP L3 + L4) - Flowcontrol - Soporte de más de 10000 MACs en la tabla de forwarding - 802.1Q - Fuentes redundantes y hot-swap - Ventiladores redundantes y hot-swap - MC-LAG (Multi-Chassis Link Aggregation Group) ó VLT (Virtual Link Trunking) en todos los switches de esta red - Soporte de tecnologías de red cloud: VxLAN
R11	En esta red física se configurarán un mínimo de 2 dominios de broadcast diferentes (2 VLANs): - 1 VLAN => Red Interna cluster (DHCP,Boot, …) - 1 VLAN => Red datos control filesystem paralelo para almacenamiento centralizado A parte de estas dos mínimas, se podrán llegar a definir varias más provenientes de los nodos que estén configurados en modo cloud.
R12	Se requerirán los siguientes GBICs y puertos libres en los switches de esta red para las redes externas del cluster: - 2 GBIC Long Range SFP+ para conectar el cluster a Internet a través de las redes del BSC - 2 GBIC Long Range SFP+ para la conexión a la red del IRB - 4 GBIC Short Range SFP+ de spare para conexiones futuras

Ref	Descripción
	Se requerirán las fibras asociadas para estas conexiones públicas (longitud 20 metros), acabadas en LC y por otro lado en SC.
R13	En esta red física se conectará: - Cada nodo de cómputo con dos interfaces 10 Gbit Ethernet - Cada login con sus interfaces 10 Gbit Ethernet - Dos interfaces por cada servidor de gestión del cluster de 10 y/ó 40 Gbit Ethernet - Todas las Interficies 40 GbitEthernet de los servidores del almacenamiento centralizado - Conexión hacia la red del IRB - Conexión hacia la red del BSC
R14	Se propone en el esquema número 1 al final de esta tabla el diseño propuesto para formar la red de cluster de la solución, que asegure que la sobresuscripción máxima a soportar desde cualquier nodo de cómputo a los switches core de esta red sea de 2:1
Red gestión privada	
R16	Se deberá proveer del hardware necesario (switches, cables, etc.) para poder establecer la red para la gestión de todos los componentes hardware que componen el cluster de cómputo y el almacenamiento centralizado de “StarLife” basado en tecnología 1 Gigabit Ethernet. Todos los cables y fibras de esta red física que vayan a la misma velocidad deberán ser del mismo color y de un color diferente a las otras redes de la máquina, de tal manera que puedan distinguirse visualmente.
R17	Todos los puertos de cada tipo de un mismo switch deberán ser line-rate entre ellos sin ningún tipo de sobre-suscripción.
R18	Requerimientos de funcionalidades de los switches de esta red: - Soporte Jumbo Frames (MTU > 9000) - Line-rate Nivel 2 switching - Line-rate Nivel 3 routing - Definición de Access-list - Routing (dinámico y estático) - Spaning-tree (MSTP y RSTP) - Capacidad para filtrar los paquetes BPDU a nivel de puerto físico del equipo - Port mirroring - Broadcast storm control - QoS - Snmp - SSH - Minimum 256 VLANs - LACP (Soporte hash LACP L3 + L4) - Flowcontrol - Soporte de más de 10000 MACs en la tabla de forwarding - 802.1Q - Fuentes redundantes y hot-swap - Ventiladores redundantes y hot-swap - MC-LAG (Multi-Chassis Link Aggregation Group) ó VLT (Virtual Link Trunking) (requerimiento para switches de segundo nivel)

Ref	Descripción
R19	En esta red física se configurará 1 dominio de broadcast (1 VLAN): Red de gestión privada
R20	En esta red física se conectará: - Cualquier interfaz de gestión de cualquiera de los componentes del cluster (racks, IPMI servers, puertas frías, PDU, switches, etc.) - Interfaz de gestión out-of-line de los nodos de cómputo y logins - Interfaz de gestión out-of-line de los servidores del almacenamiento compartido - Interfaces controladoras de disco del almacenamiento compartido - Interfaz de gestión out-of-line de cualquier servidor de gestión - Interfaz 1 Gbit Ethernet de cada servidor de administración central para acceder a esta red
Red Interconexión MPI / GPFS datos RDMA	
R21	Se deberá proveer del hardware necesario (switches, cables, etc., su esquema y etiquetado) para poder establecer la red interna de alto rendimiento y muy baja latencia sobre la cual se va enviar: - Comunicaciones MPI - Tráfico de datos filesystem paralelo RDMA Esta red deberá ofrecer un mínimo de 100 Gbits por link Todos los cables y fibras de esta red física deberán ser del mismo color y de un color diferente a cualquier otra red de la máquina. La única excepción puede ser los cables de cobre que sólo se fabriquen en color negro. Para el resto de cableado se deberá cumplir ese requisito intra o inter rack.
R22	Todos los nodos de cómputo y logins deberán estar conectados a esta red de interconexión, como los servidores de monitorización/gestión de esta red (en el caso de ser necesarios). También se deberían conectar todos los servidores de datos del almacenamiento centralizado.
R23	Dicha red no deberá ser bloqueante.
R25	Todos los switches de la red de baja latencia deberán poder ser gestionables desde la red ethernet gestión privada del cluster.

Esquema número 1 – Red Cluster "StarLife"

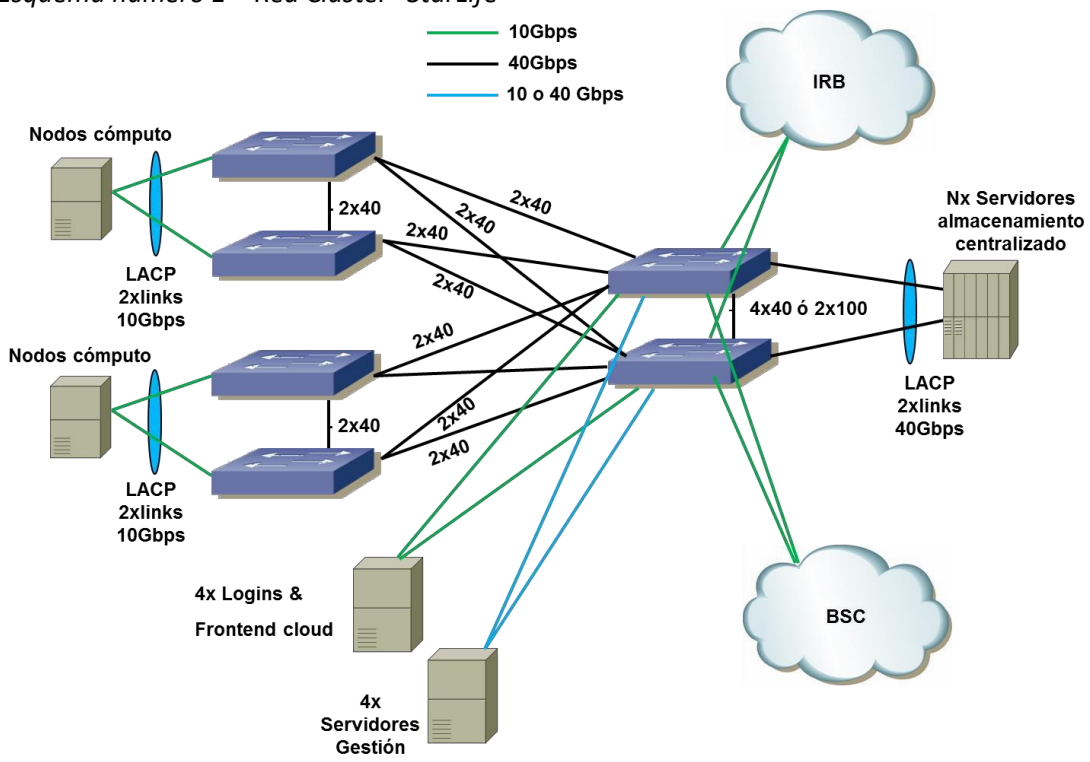


Tabla 4 – Descripción hardware switches y redes "StarLife"

Concepto	Valor mínimo	Valor ofertado
Red cluster		
Número de switches proporcionados		
Marca switch		
Modelo switch		
Número de puertos 1GE por switch		
Número de puertos 10GE por switch		
Número de puertos 40GE por switch		
Número de puertos libres		
Latencia introducida por el switch		
Red gestión privada		
Número de switches proporcionados		
Marca switch		
Modelo switch		
Número de puertos 1GE por switch		
Número de puertos 10GE por switch		
Número de puertos libres		
Latencia introducida por el switch		
Red MPI / Filesystem paralelo datos RDMA		
Número de switches proporcionados		
Marca switch		
Modelo switch		
Número de puertos por switch		
Tecnología de conexión		
Ancho de banda por puerto		
Número de puertos libres		
Latencia introducida por el switch		

2.- Operacional

En este apartado se describen los requerimientos operacionales relacionados con los componentes del cluster de cómputo y almacenamiento centralizado del proyecto “StarLife”. Los componentes del Lote 1 de “StarLife” se instalarán en un CPD a determinar que está en un radio de 1 Km del campus nord de la UPC.

Ref	Descripción
Requerimientos operacionales	
R1	Los racks deberán venir incluidos con la solución y deberán incorporar las PDU’s adecuadas para conectar todos los equipos de la solución y proporcionar redundancia en la circunstancia de la caída del 50% de las PDUs de cada rack, redundancia N+N, sin ninguna pérdida de rendimiento. Las PDUs han de ser PDUs monitorizables ó gestionables se tendrá de integrar dentro de la red privada de gestión del cluster VLAN de gestión.
R2	Como máximo todos los componentes del lote 1 deberán ocupar 5 racks standard 42U .
R3	Los racks deberán incorporar refrigeración dentro del rack por puerta trasera que elimine como mínimo el 95% del calor generado. En caso de puertas traseras gestionables deberá de conectarse a la red privada gestión del cluster.
R4	Si para la puesta en marcha, se necesitara una modificación o adaptación de la infraestructura de la sala para adaptarse a los nuevos racks, a nivel eléctrico como de instalación hidráulica estos cambios deberían estar incluidos y no deberán representar ningún coste extra.
R5	El peso de cada rack no deberá nunca superar más de 1200 Kg. x m2.
R6	Se deberá presentar en la documentación un esquema frontal con la ocupación de los racks de los diversos equipos presentados en la solución. En ella se deberá claramente especificar el hardware ofertado, como las U’s que ocupa cada componente de la solución. También se deberán de especificar por cada tipo de rack el número de cables/fibras que sale de cada rack para cada una de las redes definidas.
R7	Todos los nodos de cómputo y resto de componentes de la solución deben disponer de fuentes de alimentación redundadas N+N. Que se conectaran a las PDU correspondientes redundadas y a su vez a cuadros eléctricos diferentes.
R8	Se deberá presentar esquema de conexionado eléctrico interno de cada tipo de rack de la solución. Mostrando la redundancia de todos los elementos a nivel de alimentación a través de diferentes grupos de PDU. Se deberá realizar la conexión de elementos a PDUs para que haya una repartición uniforme entre las diversas fases eléctricas.
R9	Los elementos del almacenamiento centralizado no podrán coexistir en el mismo rack con nodos de cómputo. Se propone la siguiente distribución de elementos en racks: - Racks de cómputo: Conteniendo nodos de cómputo y los switches de primer nivel de las diversas redes correspondientes - Racks de management/storage: Componentes del almacenamiento

Ref	Descripción
	compartido, logins, switches de segundo nivel, servidores de administración
R11	Los racks deberán soportar el paso de cableado eléctrico y de datos en bandeja por encima de los racks como en una configuración de suelo técnico.
R12	Se deberá incluir todas las tareas de cableado de los compontentes según el diseño acordado entre el BSC y la empresa licitadora.
R13	Los servidores deberán disponer de un sistema para poder consultar su temperatura de ambiente de trabajo, para poder definir alertas y avisos. De la misma manera, debe soportar la parada automática en caso de detectar temperatura muy alta.
R14	La sala dispone de la alimentación de suministro eléctrico por cada rack, con 2 salidas redundantes por rack a 2 cuadros eléctricos diferentes. La sala dispondrá de 4 conectores CETAC eléctricos por rack, si la solución presentada requiere más conectores o modificación de los existentes, dichos cambios correrán por cuenta del licitador sin significar coste extra.
R15	Se deberán conectar todos los rack´s a la red de equipotencial existente. O la modificación de dicha red si hiciera falta.
R16	De la misma manera, la sala dispondrá de un sistema de refrigeración por agua para poder conectar las puertas traseras de los racks presentados, y se deberá adaptar a la tubería existente. Cualquier adaptación de dicho sistema para poder conectar las puertas traseras de los racks para su funcionamiento óptimo según tabla de especificaciones deberá ser a cargo del licitante. Para poder realizar una puesta en marcha de la conexión hidráulica adecuada se deberá purgar la puerta trasera al realizar su llenado y revisar los caudales y presiones con el equipo apropiado cualquier modificación de la instalación o requerimiento deberá hacerse cargo la empresa adjudicataría.
R17	Los racks ofertados deben disponer de puerta trasera fría con una bandeja para condensados en acero inoxidable para puerta RDHx que refrigere un mínimo del 85% del calor generado , conexiones por la parte inferior.
R18	La sala dispone de un punto de muestreo para la detección de humos por aspiración, que debe ser instalado en el interior de los racks, cualquier modificación o requerimiento deberá hacerse cargo la empresa adjudicataría.
R19	Se deberá incluir todos los elementos y componentes necesarios para el transporte y ubicación de los rack totalmente instalados en dicha sala CPD.
R20	Se exigirá en la instalación cableado (fibra, eléctrico, cobre Ethernet, etc.) ordenado, elegante y vistoso debido a que quedará a la vista. Todos los racks deberán instalarse con puerta frontal y con llave, tanto delantera como trasera.

Ref	Descripción
R21	Todo cable o fibra que forme parte de la misma red y tecnología deberá ser del mismo color en toda la máquina y entre cualquiera de los diferentes componentes hardware que formen esa red. Cada red física deberá usar un color diferente entre ellas.
R22	Todo componente de la solución (rack, server, switch, cable, fibra, …) deberá ir debidamente etiquetado, para ser identificado físicamente de forma única según nomenclatura que se establezca entre el BSC y la empresa instaladora. En los cables y fibras se deberá indicar origen y destino de la conexión usando la misma nomenclatura.
R23	La solución deberá incluir el montaje en racks de toda la solución, además de la recogida de todos los materiales sobrantes de la instalación.
R24	Cada uno de los racks se deberá entregar con todos sus componentes enrackados y con el cableado intra-rack completamente realizado y completamente optimizado para la refrigeración de todos los componentes y el fácil acceso a los diversos componentes para su sustitución. Todos los nodos de cómputo deberán haber pasado un burn- in test en fábrica para evitar los DOA (Dead on Arrival).
R25	Cualquier U vacía en cualquiera de los racks deberá taparse frontalmente con tapas ciegas.
R26	Se deberá presentar un diagrama de Gantt con la indicación de las diversas fases de la instalación del lote1 desde el inicio del contrato hasta la finalización de la instalación del mismo.
R28	Se deberá de realizar limpieza diaria del material sobrante a medida que se vaya instalando, la empresa licitadora se deberá hacer cargo de la limpieza del material sobrante y de la limpieza final de obra, la recogida de residuos se deberá realizar a un vertedero oficial para su correcto reciclaje.
R29	Con la finalización del proyecto se deberá presentar la siguiente documentación en formato Autocad 2000 o superior y PDF: - Planos Asbuilt con todos los elementos instalados. Fichas técnicas de todos los elementos instalados.

3.- Software

En este apartado se describe el software a proporcionar en el cluster de cómputo y el almacenamiento compartido el proyecto “StarLife”.

Ref	Descripción
R1	El sistema operativo deberá ser UNIX like y compatible con el X/Open Standard POSIX 1003 (IS/IEC 9945). El sistema operativo deberá ser Linux, todos los componentes deberán llevar la misma versión de sistema operativo. Dicho sistema operativo deberá proporcionar soporte Enterprise y estar soportado por cualquiera del resto de componentes del software stack de la máquina: Sistema de clustering, sistema de colas, sistema de ficheros, compiladores, drivers, etc. El sistema operativo ofertado deberá estar soportado por la última versión del filesystem paralelo ofertado y soportar la instalación de virtualización y containers.
R2	Se deberá aportar también todo el software necesario para la gestión de todos los componentes que formen la solución: Switches, Controladoras disco, etc.
R3	Cualquier elemento hardware (servidores de administración, nodos de computo, logins) deberá soportar el standard IPMIv2, y deberá soportar estar gestionado por el sistema de clustering xCAT.
R4	Con el sistema operativo se debe incluir todo el entorno de programación para la arquitectura de la máquina, como mínimo deberá incluir C, C++, Java, Fortran. A parte del entorno de programación Open-Source proporcionado por el sistema operativo, se deberá de proporcionar el entorno de programación específico para la arquitectura del procesador proporcionada.
R5	Los compiladores de los diversos procesadores ofertados deberán venir con licencias ilimitadas por cada uno de los logins HPC ofertados.
R6	Se deberán de proporcionar las librerías numéricas (secuenciales y paralelas) proporcionadas por el fabricante de los procesadores debidamente optimizadas para cada arquitectura. Como pueden ser MKL ó ESSL/pESSL. Se deberán aportar actualizaciones de librerías con nuevos sistemas.
R7	También se deberá proporcionar los compiladores, librerías y/o las herramientas necesarias para el uso paralelo de la arquitectura mediante paradigmas estándares como OpenMP ó MPI. Para OpenMP deberá soportar la versión 3.1, y para MPI se deberá soportar completamente el estándar MPI versión 3.0. A parte, de la versión open-source se deberá proveer de una implementación especializada en la arquitectura propuesta en el caso que exista, como por ejemplo Intel MPI, Spectrum MPI o similares.
R8	Las librerías paralelas para el uso de MPI deben ser optimizadas para el uso de la red de baja latencia MPI ofertada para el cluster de cómputo.
R9	Se deberán incluir el software y las licencias del sistema de ficheros paralelo (como por ejemplo GPFS) de tipo servidor y cliente para todos los nodos de cómputo, logins y servidores de datos ofertados para la infraestructura del lote 1 de “StarLife”.
R10	Se deberá incluir software de gestión y monitorización de la red de interconexión MPI que permita de forma centralizada: - Localización de “soft failures”. - Links con failure rates por encima de lo deseado - Alarmas y detección de errores graves dentro de la red

Ref	Descripción
	- Mostrar la carga de tráfico a nivel real por cada link y a nivel global - Poder seleccionar un subconjunto de nodos y realizar una monitorización de los mismos
R11	Para ayudar a definir las características ténicas de los diversos servidores de administración del lote 1 de “StarLife”, comentar que el equipo de sistemas del BSC planifica instalar el siguiente software stack: - Software Clustering: xCAT - Monitorización: Ganglia, Nagios - Sistema de colas: Slurm

4.- Mantenimiento y soporte

Ref	Descripción
R1	Se deberá proporcionar un mantenimiento de una duración mínima de 3 años desde el momento de aceptación del lote 1. Delante de fallos hardware se deberán reparar con una respuesta en 4 horas dentro de las horas de oficina (08:00 – 17:00) y con un servicio de soporte de Next Business Day. En caso de incidencias muy críticas que impliquen una afectación global de la producción de los clusters, se deberá proveer un seguimiento continuo 24x7 hasta la resolución de la incidencia.
R2	La empresa licitadora se hará cargo de la reparación y sustitución durante el periodo del proyecto “StarLife” de cualquier componente hardware de los clusters de cómputo y almacenamiento centralizado.
R4	El proyecto de instalación incluirá las siguientes tareas: - Instalación del hardware “llave en mano”, completamente cableado y labelado, listo para instalar el software. - La comprobación del buen funcionamiento inicial del hardware como firmware de los componentes ofertados. La empresa de licitación deberá recomendar los niveles de firmware necesarios y deberá instalarlos. - Proporcionar los valores recomendados a configurar en el firmware/BIOS para el máximo rendimiento de los nodos de computo, adecuados para cada uno de los dos tipos de ejecuciones que se van a realizar (HPC y BigData/Hadoop/Virtualización). - Configuración de IPs de las interficies de gestión de todos los elementos de la solución según las indicaciones del equipo de operaciones del BSC. - Instalación completa del sistema de almacenamiento centralizado hardware y software por técnicos de laboratorio especializados en el sistema de ficheros paralelo ofertado. - En el proyecto de instalación se deberá incluir la instalación de la red de baja latencia (IB/OPA) en el cluster de cómputo. Configuración switches, instalación del software stack en los nodos de cómputo, posible adaptación para ofrecerlo a los nodos configurados de tipo cloud, configuración de gestores o fabric managers. El resto de tareas de instalación del cluster de cómputo será responsabilidad del equipo de sistemas del BSC, pero se deberá poder conectar con el equipo que diseñó la solución presentada en caso de malfuncionamiento o dudas sobre su configuración técnica (hardware/firmware).
R5	Se exigirá un trabajo en equipo con el departamento de operaciones del BSC, para la coordinación del diseño hardware de la solución. Cualquier plan o toma de decisión se deberá verificar con el departamento de operaciones del BSC antes de llevarla a cabo.

R6	Se proporcionará (dentro de período del proyecto “StarLife”): - Acceso a todo el software upgrade (incluyendo sistemas operativos, clientes filesystem paralelo y firmware) de todos los componentes de la solución - Punto único de suporte para el aviso de problemas e incidencias de cualquier componente que componga la solución
R7	Se exigirá soporte pro-activo, notificando y recomendado subidas de versión tanto de software como de firmware de cualquier componente de la solución.
R8	Se deberá entregar al final de la instalación una documentación digital en formato office en la que se describa: - Descripción general de los componentes de la solución - Esquemas físicos de los conexionados finales realizados - Versiones de firmware instalados y recomendaciones de configuración BIOS/firmware - Explicación procedimientos hardware para: Puesta en marcha, parada y diagnostico delante de problemas hardware - Proyecto de instalación y configuración del almacenamiento centralizado
R9	Toda la instalación y desarrollo del proyecto se deberá hacer on-site en las instalaciones del BSC bajo la supervisión del grupo de sistemas del BSC. En ningún caso se permitirá el acceso externo o remoto para la configuración o instalación de la solución presentada. Durante el mantenimiento no se permite el acceso remoto y toda modificación se debe hacer on-site.
R10	En la implantación de la solución presentada se exigirá la participación activa y presencial (si se requiere) de los expertos de cada uno de los componentes que forman la solución: - Responsables de hardware/ desarrolladores de firmware - Desarrolladores o responsables técnicos de redes o switches ethernet - Desarrolladores o responsables técnicos de red MPI ofertada Teniendo la posibilidad el personal del BSC poder intercambiar emails de forma directa con dichas personas con el fin de solucionar cualquier problema que surja durante el desarrollo e instalación de la máquina.
R11	El equipo técnico encargado de la instalación hardware y software deberá disponer de la formación y capacidades técnicas para la realización de este tipo de instalaciones, ya que es imprescindible para la correcta ejecución del contrato.
R12	De la misma manera, para las tareas de mantenimiento hardware del cluster y almacenamiento una vez en producción, la empresa licitadora deberá disponer de un equipo de personas suficiente para el desarrollo y asistencia en la cercanía de Barcelona, con posibilidad de personarse en las dependencias del BSC en menos de 2 horas.
R13	Una vez acabada la instalación física se deberá de comprobar que todos los requerimientos operativos establecidos en este pliego se cumplen. Como, por

	ejemplo, sin estar limitados a: - Redundancia eléctrica allí donde se requiera - Adecuación y optimización de la infraestructura - Repartición de las fases eléctricas - Cableado óptimo para el flujo de aire y refrigeración de todos los componentes - Etc.

Lote 2: Ampliación infraestructura Robot BSC para la implementación de un sistema HSM para "StarLife"

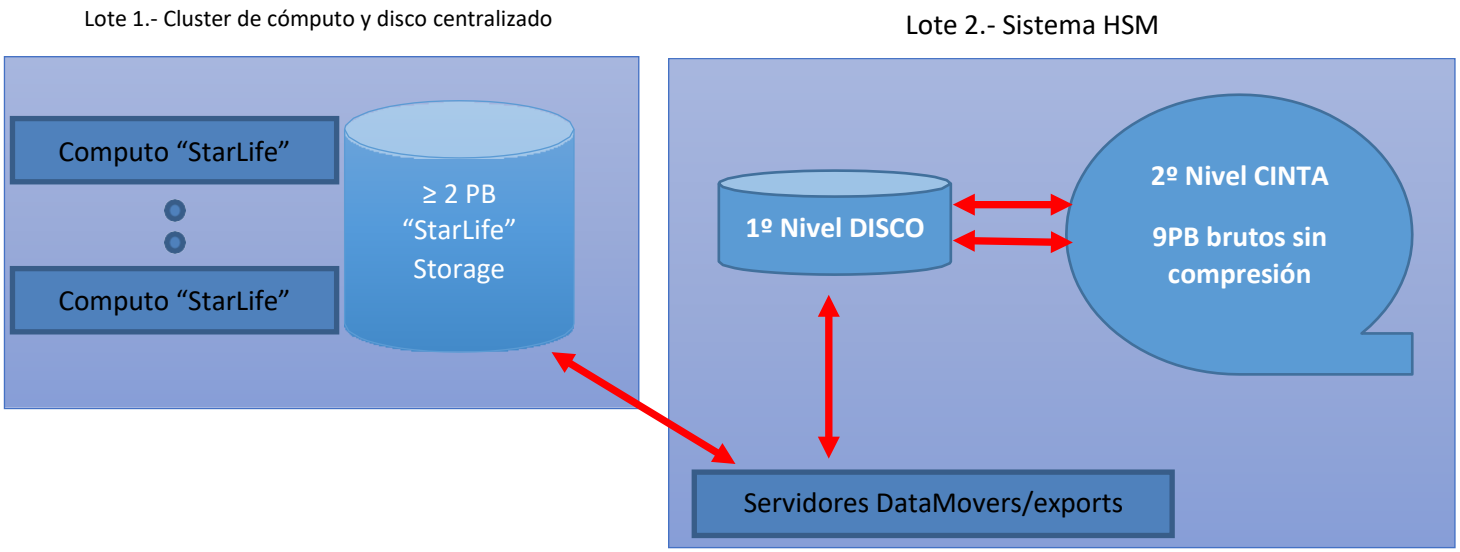
En este segundo lote, se describe las prescripciones técnicas para la adquisición de un almacenamiento en cinta con una capacidad bruta mínima de 9PB sin compresión, junto un sistema de HSM (Hierarchical Storage Management).

El sistema de HSM propuesto estará compuesto de 2 niveles de almacenamiento uno basado en disco; y un segundo nivel basado en cinta con una capacidad mínima de 9PB bruto sin compresión.

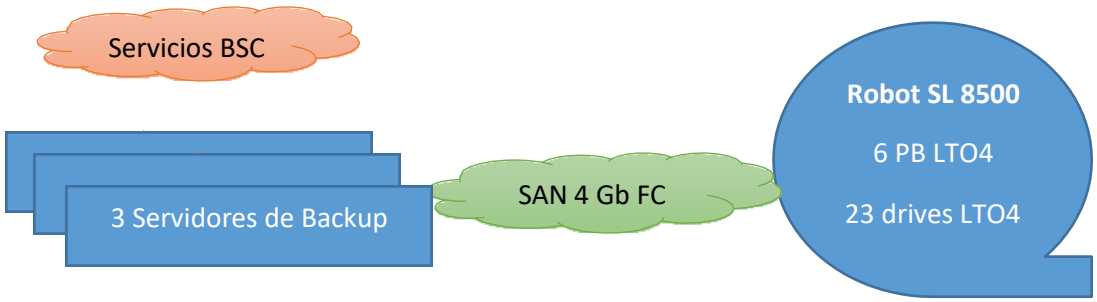
El sistema HSM deberá mostrar toda su capacidad como un único espacio de filesystem independientemente si el dato esté en disco o cinta. Y deberá migrar datos entre las dos tecnologías de almacenamiento según las políticas a definir de acceso y reclamo. También deberá soportar el reclamo automático cuando un fichero se acceda y esté en cinta.

El sistema de HSM a proveer deberá integrarse con el almacenamiento centralizado y el cluster de cómputo propuestos para el proyecto "StarLife" que el lote 1 describe, en concreto con la inclusión de 2 servidores "Data Movers" que serán los encargados del movimiento y acceso de datos entre los dos almacenamientos.

A continuación, se muestra un esquema de la solución HSM propuesta:



Para la implementación de este sistema HSM, se propone la reutilización y ampliación de la infraestructura informática que se encarga de las funcionalidades de backup del BSC-CNS. A continuación, se muestra un esquema de la infraestructura de backup del BSC actual:



La adaptación propuesta se diseñará de tal manera que sea fácilmente ampliable y que aparte de dar el servicio de HSM para "StarLife", permita la futura implementación al BSC de su propio sistema de archivado HSM.

En resumen, los cambios a realizar serían:

- Ampliación del robot del BSC con un mínimo de 4 drives lectores/escritores
- Compra e instalación de las cintas necesarias para proporcionar 9 PB brutos sin comprimir
- Sustitución de la SAN actual por una de 16Gbits que interconecte los nuevos y los viejos componentes (Drives, Sistema de metadatos, Servidores backup y Servidores HSM).
- Creación de un almacenamiento (servidores y controladoras de discos) basado en GPFS o similar que funcionará como el primer nivel de almacenamiento del HSM
- Adquisición de los servidores necesarios para realizar las funciones propiamente de HSM entre los dos almacenamientos (discos y cintas)
- Adquisición de los servidores con los servicios de DataMovers y Exports hacia el almacenamiento compartido de "StarLife" del Lote 1
- Instalación y configuración del software del sistema de HSM con el hardware ofertado

Para el primer nivel de almacenamiento de HSM, se requiere un filesystem paralelo como por ejemplo GPFS con un almacenamiento en datos de 3 PB neto y con un almacenamiento en flash para metadatos de mínimo de 100 TB netos. Para configurar dicho filesystem se podrá usar material spare que dispone el BSC (696 discos de 6 TB, referencia IBM FRU 38L6721).

El sistema HSM se deberá entregar "llaves en mano" completamente configurado y funcionando según los requerimientos aquí descritos. Los componentes hardware y software adquiridos en este lote (incluyendo el material proporcionado por el BSC (discos 6TB spare) deberán estar cubiertos con mantenimiento de 3 años desde la aceptación de este lote.

1.- Hardware

A continuación, pasamos a describir en detalle los requerimientos mínimos sobre el hardware que se demanda para la plataforma “StarLife” en el lote referente al sistema HSM para el proyecto “StarLife”.

1.1.- Hardware Robótica

Ref	Descripción
R1	Se deberá proporcionar la cantidad de cintas necesarias (compatibles con el robot SL8500 del BSC), para proporcionar una capacidad mínima de 9PB netos sin compresión. Todas las cintas deberán ser de la misma tecnología y tamaño. La tecnología de cintas a proporcionar deberá ser la más actual del mercado que existirá en la fecha de entrada a producción de “StarLife” (30 enero 2018). Cada cinta deberá poder albergar un mínimo de 12 TB sin compresión.
R2	Todas las cintas deberán llegar etiquetadas según las indicaciones del BSC, con el formato standard para que sean reconocidas por el robot SL8500 del BSC.
R4	Se deberán proveer un mínimo de 4 unidades lectoras/escritoras para las cintas ofrecidas, compatible con el robot SL8500 del BSC. Cada lectora deberá proporcionar un rendimiento mínimo de 350 MB/s
R6	Se deberán proveer de unas 30 cintas limpiadoras LTO de drives adecuadamente etiquetadas, según indicaciones del BSC.
R7	Se deberá incluir cualquier tarea (hardware o software) necesaria para la instalación y buen funcionamiento en el robot SL8500 del BSC de las cintas y drives antes mencionados, sin disruptir el funcionamiento de los componentes ya existentes. Se prevén, sin estar limitado a ellas, las siguientes tareas como ejemplo: - Upgrade de firmware robot - Actualización y configuración ACSLS (ownership de cintas, nuevos drives) - Retirada de los drives y cintas en desuso con tecnología LTO3 - Introducción cintas y chequeo de auditoría - Instalación drives en el robot, tirada fibras nuevas desde el robot hacia los nuevos switches FC - Etc.
R8	Actualización de la SAN actual basada en tecnología 4 Gb FiberChannel a una SAN de tecnología 16Gb FiberChannel. Esta SAN deberá estar formada por un mínimo de 2 switches redundantes. Esta SAN deberá disponer de un mínimo de 96 puertos para poder usar, los puertos deberán poder funcionar a FC4,FC8 ó FC16, y todos los puertos deberán ser line-rate. La nueva SAN aparte de conectar los nuevos drives con los servidores HSM, deberá conectar la infraestructura actual de backup del BSC, formada por:

Ref	Descripción
	- 23 drives LTO4 - 18 links provenientes de los servidores de backup del BSC La nueva SAN debe venir licenciada para definir un mínimo de 6 zonas. Todos los puertos de las SAN deberán ser line-rate, y la SAN no podrá disponer de ningún punto de fallo único.
R9	Se deberá incluir las tareas de diseño, configuración y migración de la infraestrucutra de backup del BSC para que use la nueva SAN.
R10	Se deberán de proveer de todos los GBIC FiberChannel para los nuevos componentes. Para los antiguos componentes, se podrán usar los actuales (Picolight Sun PLRXPLVCSG324N) siempre y cuando sean compatibles con los switches propuestos a proporcionar, sino se deberán de incluir esos GBICs también. Se deberán de proveer de un 5% extra de GBICs FC8/16 para futuras ampliaciones.
R11	Se incluirá en las tareas la tirada de nuevas fibras por el falso suelo de la sala entre el robot SL8500 y el/los nuevos switches FiberChannel. Se deberán tirar 30 pares de fibra en total para que sirva para futuras ampliaciones o reparaciones de fibras.
R12	Si la tecnología de cintas y drives propuesta no estuviera disponible a la hora de la firma del contrato, se deberá de proveer no más tarde de un mes desde la "General Availability" de dicha tecnología.
R13	Se deberá realizar un estudio durante el proyecto de instalación sobre la colocación física de los drives y cintas en el robot del BSC para optimizar el rendimiento del sistema HSM como el sistema en producción de backup.
R15	Los nuevos drives se deberán integrar de forma nativa con los sistemas de monitorización y gestión del robot. (SLConsole, ACSLS, etc.) Realizando cualquier tarea hardware/software que eso implique.

1.2.- Hardware Primer nivel almacenamiento HSM

El BSC dispone en su propiedad el siguiente hardware sin mantenimiento:

- 696 discos de 6 TB 7.2Krpm modelo: IBM FRU 38L6721

En este apartado se resumen los requerimientos técnicos para que usando ese hardware se cree un sistema de ficheros paralelo que sirva como primer nivel de HSM.

Ref	Descripción
R1	Se deberá proveer de todo el hardware necesario para establecer un sistema de ficheros paralelo (como por ejemplo GPFS) configurado de forma modular, para ese montaje se podrán usar los discos duros en spare que BSC dispone en stock. El

Ref	Descripción
	filesystem resultante debe proveer de una capacidad mínima neta de 3 PB para ser usado como primer nivel de HSM.
R3	Se deberán configurar 2 módulos de datos de filesystem paralelo (pudiendo usar los 696 discos en stock del BSC), los cuales podrán ser configurados de dos maneras posibles: - Módulos basados en tecnologías de declustered RAID con 6 expansiones de discos. - Módulos basados en controladoras RAID: Controladoras dobles hardware RAID6 con expansiones de alta densidad (en caso de usar los discos spare BSC deberán ser completamente compatibles con dichos discos).
R4	El filesystem resultante deberá proporcionar un mínimo de 20 GB/s de rendimiento tanto en lecturas como escrituras.
R6	Cada módulo de datos (ya sea declustered RAID o controladoras RAID6) deberá venir configurado con un mínimo de 2 servidores idénticos que funcionen como servidores de datos del filesystem paralelo. Dichos servers deberán tener las siguientes conexiones: - Enlaces SAS6Gbit/12Gbit ó FC16Gbit para conectarse directamente (sin SAN) al almacenamiento - 2 enlaces 40 Gbit Ethernet para servir los datos (en 2 tarjetas independientes, no se aceptará una única tarjeta dual-port) - 1 Enlace Gbit (gestión SSH) - Conexión out-of-line para su gestión (shared o dedicated) Dichos servidores deberán estar adecuadamente configurados para funcionar como servidores de datos del filesystem paralelo y primer nivel de HSM. Se deberá presentar un esquema del conexionado interno de cada módulo de datos, con las tarjetas previstas a usar y los puertos PCI-Express asociados.
R8	Por otro lado, se deberá proporcionar un módulo de almacenamiento de metadatos del filesystem paralelo basado en flash con dos servidores idénticos redundantes. Dicho almacenamiento deberá de disponer de un sistema de protección hardware a fallos (RAID o similar), soportar un mínimo de 200K IOPS por segundo, una capacidad neta mínima de 100 TB y un rendimiento mínimo de 10 GB/s. El sistema de ficheros deberá soportar tener un blocksize diferente para datos y para metadatos.
R10	Cada servidor de metadatos deberá de estar adecuadamente configurado para realizar sus funciones asignadas. Cada servidor dispondrá de las siguientes conexiones:

Ref	Descripción
	- Mínimo 8 enlaces Fiberchannel FC16 para conectarse mediante SAN antes descrita al almacenamiento flash - 2 enlaces 40 Gbit Ethernet para servir los datos (en 2 tarjetas independientes, no se aceptará una única tarjeta dual-port) - 1 enlace Gbit (gestión SSH) - 1 enlace out-of-line para su gestión(shared o dedicated)
R11	El módulo de metadatos deberá cumplir 3 funciones: - Almacenar y servir los metadatos del filesystem del primer nivel de HSM - Los servidores de metadatos deberán realizar las tareas de ILM (scan y generación de listas) sobre las reglas definidas de migración y recall de HSM - Creación de un filesystem de alto rendimiento en IOPS donde se almacenarán los datos administrativos del sistema HSM (bases de datos, recovery logs, etc.) - Creación de un segundo filesystem de alto rendimiento en IOPS donde se almacenarán los datos administrativos de los servidores Tivoli del BSC (bases de datos, recovery logs, etc.)
R12	Se deberá presentar un esquema de conexionado del módulo de metadatos, indicando los puertos fiberchannel a conectar de cada uno de los componentes y que ancho de banda disponen.
R14	Todos los servidores proporcionados deberán disponer de una metodología de administración out-of-line: consola remota, power on y off, monitorización de alarmas y temperaturas, etc.
R15	No deberá existir ningún punto único de fallo en cualquier componente que forme parte del almacenamiento de primer nivel HSM aquí descrito, y todo debería funcionar en modo de alta disponibilidad activo-activo.
R16	El hardware de primer nivel de HSM se deberá configurar con un filesystem paralelo basado en tecnología GPFS o similar que tenga las siguientes funcionalidades avanzadas: - Soporte del protocolo Data Management API (DMAPI) - Definición de filesets para poder particionar un filesystem en varios elementos y espacio de nombres administrativamente independientes, con sus puntos de montaje, pudiendo definir reglas de migración y quotas por fileset - Definición de storage pools (por ejemplo: data, metadata con blocksizes diferentes) todo dentro de un mismo filesystem - Filesystem con sistema integrado de ILM (Information Lifecycle Management) que permita definir políticas de migrado o gestión de los datos con un lenguaje “sql-like” pudiendo distribuir estas tareas entre diversos nodos (ejecución paralela). - Sistema de metadatos distribuido y balanceado con data in-inode - Soporte de montaje del filesystem o de directorios del mismo en modo pass-through mediante protocolo 9p o exportación via NFS/SAMBA con alta

Ref	Descripción
	disponibilidad activo-activo - Incorporación módulos de integración con Hadoop y/o object store - Posibilidad de conexión de filesystems de otros clusters (montaje multi-cluster)
R17	La solución de almacenamiento propuesta deberá soportar que los servidores HSM como los servidores Tivoli del BSC, puedan acceder a los filesystems paralelos de alto rendimiento de IOPS definidos en los flash mediante sus propios enlaces FiberChannel para minimizar la latencia de acceso, aún utilizando el protocolo del filesystem paralelo.
R18	Se requiere que se rellene la siguiente tabla (Tabla 5- Descripción hardware Primer nivel almacenamiento HSM), en ella se especifican los valores mínimos a cumplir, y se deberá indicar los valores ofertados.

Tabla5 – Descripción hardware Primer nivel almacenamiento HSM

Concepto	Valor mínimo	Valor ofertado
Módulo datos Filesystem paralelo		
Número servidores por módulo datos	2	
Modelo servidor		
Número chips o sockets por nodo		
Modelo procesador		
Ancho de banda (GB/s) entre procesadores		
Cores por procesador ofertado		
Frecuencia nominal de cada core		
Memoria RAM por servidor		
Tecnología y frecuencia memoria RAM		
Frecuencia real funcionamiento memoria		
Bandwith total Socket CPUs ←→ Memoria (GB/s)		
Número y configuración Discos locales para SO	2 en RAID1	
Capacidad almacenamiento local para SO		
Tecnología almacenamiento local para SO	SAS/SSD	
Número de puertos PCI-Express x16 por servidor		
Número de puertos PCI-Express x8 por servidor		
Tipo interfaz de conexión a expansiones	SAS/FC	
Bandwith por link de conexión a expansiones (Gbits)		
Número de tarjetas de conexión a expansiones		
Núero de puertos por tarjeta de conexión a exp.		
Interfaces 40GE incorporadas por servidor	2	
Interfaces 10GE incorporadas por servidor		

Interfaces 1 GE incorporadas por servidor		
Almacenamiento módulo datos		
Número de expansiones y/o controladora por modulo		
Modelo expansión y/o controladora		
Rendimiento por controladora y/o módulo esperado (GB/s)		
Número puertos a host por controladora y/o expansión		
Tecnología de protección hardware propuesta		
Configuración de protección hardware propuesta		
Espacio neto por módulo		
Módulo metadatos		
Servidores por módulo metadatos	2	
Modelo servidor		
Número chips o sockets por nodo		
Modelo procesador		
Ancho de banda (GB/s) entre procesadores		
Cores por procesador ofertado		
Frecuencia nominal de cada core		
Memoria RAM por servidor		
Tecnología y frecuencia memoria RAM		
Frecuencia real funcionamiento memoria		
Número y configuración Discos locales para SO	2	
Capacidad almacenamiento local para SO		
Tecnología almacenamiento local para SO	SSD	
Número de puertos PCI-Express x16 por servidor		
Número de puertos PCI-Express x8 por servidor		
Tipo interfaz de conexión a almacenamiento	FC	
Bandwith por link de conexión a expansiones (Gbits)		
Número de tarjetas de conexión a expansiones		
Número de puertos por tarjeta de conexión a exp.		
Interfaces 40GE incorporadas por servidor	2	
Interfaces 10GE incorporadas por servidor		
Interfaces 1 GE incorporadas por servidor		
Almacenamiento módulo metadatos		
Número de expansiones y/o controladora por modulo		
Modelo expansión y/o controladora		
Rendimiento por controladora esperado (GB/s)		
IOPS lectura 4KB		
IOPS escritura 4KB		
Número puertos a host por controladora y/o expansión		
Número total de discos ofertados por módulo		
Tecnología discos ofertados	Flash	
Capacidad por disco ofertado		
Tecnología de protección hardware propuesta		
Configuración de protección hardware propuesta		
Espacio neto por módulo metadatos		
Características globales almacenamiento primer nivel HSM		
Número módulos ofertados datos	2	
Número módulos ofertados metadatos	1	

Rendimiento total esperado almacenamiento (GB/s)	20	
Espacio neto en datos (PB)	3	
Espacio neto en metadatos (TB)	100	

1.3.- Servidores LifeStar HSM

Ref	Descripción
R1	Se deberán de proveer el número de servidores requerido para realizar las funciones de HSM para "StarLife", siendo estos un mínimo de 2 para poder funcionar en alta disponibilidad activo-pasivo o activo-activo. Un servidor HSM deberá ser capaz de gestionar hasta 4800 millones de ficheros y un servidor deberá ser capaz de maximizar el rendimiento ofrecido por los drives ofertados.
R2	Los servidores HSM deberán montar de forma nativa el filesystem paralelo que se establece como primer nivel de almacenamiento de HSM, y tendrán la función principal de copiar/mover datos entre éste y las cintas. A la vez se encargarán de gestionar la información administrativa necesaria para las funciones de HSM (inventario de datos en cinta, petición montaje cintas, etc.) Los servidores de HSM no podrán usar ningún otro almacenamiento intermedio para la función de HSM, deberán leer/escribir directamente desde/hacía el primer nivel de almacenamiento del filesystem paralelo.
R3	Cada servidor de HSM deberá tener las siguientes conexiones: - Tantos puertos FC conectados como drives tenga la oferta, funcionando a FC16 - 2 enlaces 10 Gbit Ethernet - 1 enlace Gbit gestión SSH - 1 puerto gestión out-of-line (shared o dedicated)
R4	Los servidores deberán estar configurados a nivel de características técnicas (cpu, memoria, etc.) para realizar las funciones que se les asignan de HSM, y deberán ser capaces de leer/escribir entre los dos almacenamientos usando el máximo rendimiento esperado por ellos sin ser ningún factor limitante.
R6	Se deberán de proveer tantos servidores de administración como sean necesarios para gestionar el resto de servidores que conforman la solución HSM descrita en el Lote2. Como mínimo deberá ser un servidor, el cual será responsable de la instalación inicial de todos los servidores y hacerse cargo de servicios como DNS, NTP, repositorio OS, gestión del resto de servidores mediante IPMI, etc. Los servidores deberán disponer de tantas interficies de red como sean necesarias para administrar el resto de elementos con una extra para la conexión a la LAN del BSC.
R7	Se deberán de proveer un mínimo de 2 servidores con las funciones de "StarLife" datamovers y export, que deberán funcionar en modo de alta disponibilidad activo-activo. Dichos servidores deberán montar nativamente los filesystems paralelos del HSM y el espacio compartido "StarLife" del Lote 1, este segundo como cluster remoto. Estos servidores cumplirán 2 funciones: - Exportar mediante NFS en modo activo-activo con IP's virtuales (Cluster NFS)

Ref	Descripción
	a nodos específicos del cluster Lote1, para su acceso interactivo. - Ejecutar copias de datos entre almacenamiento del Lote1 y el primer nivel de HSM mediante un sistema batch.
R8	Los servidores datamovers deberán tener las siguientes inferficies: - 4 enlaces 10 Gbit Ethernet (2 para HSM, 2 para conexión con Lote1 "StarLife") - 1 enlace Gbit gestión SSH - 1 puerto gestión out-of-line (shared o dedicated)
R9	Los servidores proporcionados deberán disponer de una metodología de administración out-of-line: consola remota, power on y off, monitorización de alarmas y temperaturas, etc. que deberá conectarse a la red de gestión privada del cluster de cómputo.
R11	Se deberá de proveer de un KVM (Keyboard Video Mouse) de 1U integrado en el cluster que permita acceder a la consola de todos los servidores de la solución HSM.

1.4.- Switches y redes Ethernet

Todos los elementos que forman parte del sistema HSM estarán conectados a través de 2 redes disjuntas físicamente, con lo que se deberá proveer de todos los elementos para su formación:

- *Red de administración*: Conectando todos los elementos de gestión del cluster (switches FC, IPMI servidores, interficies SSH gestión) y sólo accesible por los administradores de la infraestructura
- *Red de datos filesystem paralelo HSM*: Red basada en tecnología 10/40 Gbit Ethernet visible entre los nodos del HSM que permitirá el montaje del filesystem paralelo en todos los servidores.

A continuación, se detallan los requisitos comunes para todas las redes del sistema HSM "StarLife" y en las consiguientes tablas los requisitos específicos para cada una de las redes.

Ref	Descripción
R1	Se deberán de proveer de esquemas de conexionado físico de cada una de las redes que conforman el sistema HSM de la infraestructura "StarLife". También se debe describir el ancho de banda disponible a cada nivel de las redes y la latencia introducida por cada elemento hardware. Cada una de las redes descritas debe ser completamente disjunta a nivel físico.

R2	Todos los switches de cualquier red deberán tener doble fuente de alimentación, y redundancia a nivel de ventiladores. Todos estos componentes deberán ser modulares y poderse cambiar en caliente, sin la parada del switch en cuestión. Si dentro de una misma red se proveen switches de fabricantes diferentes y se detecta cualquier incompatibilidad a la hora de conectarlos entre ellos (GBIC, fibra, limitación funcionalidades, rendimiento), el licitante deberá sustituir los switches necesarios para que todos sean del mismo fabricante para eliminar la incompatibilidad.
R3	Para cada una de las redes y una vez todos los componentes estén conectados deberá existir un 10% de puertos libres físicos (sin break-outs) en cualquiera de los switches.
R4	Se requiere que se rellene la tabla (Tabla 6- Descripción Sitches y redes HSM "StarLife"), en ella se especifican los valores mínimos a cumplir. En el caso de proporcionar más de un tipo de switch por red, se deberán rellenar los datos de la tabla 6 por cada tipo de switch proporcionado.
R6	En todas las redes que hayan conexiones de velocidades diferentes, los switches deberán incorporar los buffers necesarios para ofrecer los rendimientos line-rate entre las diferentes velocidades.

Ref	Descripción
Red de administración	
R7	Se deberá proveer del hardware necesario (switches, cables, etc.) para poder establecer la red interna del cluster con tecnología 1 Gigabit Ethernet. Todos los cables y fibras de esta red física que vayan a la misma velocidad deberán ser del mismo color y de un color diferente a las otras redes de la máquina, de tal manera que puedan distinguirse visualmente.
R8	Todos los puertos de cada tipo de un mismo switch deberán ser line-rate entre ellos sin ningún tipo de sobre-suscripción.
R9	Requerimientos de funcionalidades de los switches de esta red: - Soporte Jumbo Frames (MTU > 9000) - Line-rate Nivel 2 switching - Definición de Access-list a nivel 2 y nivel 3 - Spaning-tree (MSTP y RSTP) - Capacidad para filtrar los paquetes BPDU a nivel de puerto físico del equipo - Port mirroring - Broadcast storm control - QoS - Snmp - SSH - Minimum 256 VLANs - LACP (Soporte hash LACP L3 + L4) - Flowcontrol

Ref	Descripción
	- Soporte de más de 10000 MACs en la tabla de forwarding - 802.1Q - Fuentes redundantes y hot-swap - Ventiladores redundantes y hot-swap - MC-LAG (Multi-Chassis Link Aggregation Group) ó VLT (Virtual Link Trunking) en todos los switches de esta red
R10	En esta red física se conectará: - Puertos gestión out-of-line de todos los servidores del lote2 - Puertos gestión mediante SSH - Cualquier interfaz de gestión de cualquiera de los componentes del cluster (racks, IPMI servers, puertas frías, PDU, switches FC, etc.) - Interfaces controladoras de disco/flash - Interfaz 1 Gbit Ethernet de cada servidor de administración central para acceder a esta red
Red datos filesystem paralelo HSM	
R11	Se deberá proveer del hardware necesario (switches, cables, etc.) para poder establecer la red de datos del filesystem paralelo HSM basado en tecnología 10/40 Gigabit Ethernet y que permita conectar los servidores a la LAN del BSC. Todos los cables y fibras de esta red física que vayan a la misma velocidad deberán ser del mismo color y de un color diferente a las otras redes de la máquina, de tal manera que puedan distinguirse visualmente. Para dicha red se propone equipos con las siguientes características: - 1 Switch con un mínimo de 32 puertos 40 Gbit Line-rate - 1 Switch con un mínimo de 48 puertos 10 Gbit con 6 puertos 40 Gbit todos ellos line-rate - Estos dos switches deberán conectarse entre ellos con un mínimo de 4 uplinks de 40 Gbits
R12	Se requerirán los GBICs necesarios para conectar todos elementos de la infraestructura a esta red. Adicionalmente se deberán proveer de los siguientes GBICs y puertos libres en los switches de esta red para las redes externas del cluster: - 8 GBIC Long Range 10 Gbit Ethernet para conectar el sistema HSM al resto de LANs del BSC (conexión con edificio Torre Girona) - 8 GBIC Long Range 10 Gbit Ethernet (Dell GP-10GSFP-1L o similar); para poderse instalar en el equipo que el BSC dispone en el edificio Torre Girona (Dell S4048) para habilitar la conexión en el otro extremo - 4 GBIC Short Range 10 Gbit de spare para conexiones futuras - 2 GBIC Short Range QSFP 40 Gbit de spare para conexiones futuras
R14	Todos los puertos de cada tipo de un mismo switch deberán ser line-rate entre ellos sin ningún tipo de sobre-suscripción.
R15	Requerimientos de funcionalidades de los switches de esta red: - Soporte Jumbo Frames (MTU > 9000) - Line-rate Nivel 2 switching - Line-rate Nivel 3 routing - Definición de Access-list - Routing (dinámico y estático)

Ref	Descripción
	- Spaning-tree (MSTP y RSTP) - Capacidad para filtrar los paquetes BPDU a nivel de puerto físico del equipo - Port mirroring - Broadcast storm control - QoS - Snmp - SSH - Minimum 256 VLANs - LACP (Soporte hash LACP L3 + L4) - Flowcontrol - Soporte de más de 10000 MACs en la tabla de forwarding - 802.1Q - Fuentes redundantes y hot-swap - Ventiladores redundantes y hot-swap - MC-LAG (Multi-Chassis Link Aggregation Group) ó VLT (Virtual Link Trunking) (requerimiento para switches de segundo nivel)
R16	En esta red física se conectará: - Servidores de datos del almacenamiento primer nivel HSM - Servidores de metadatos del almacenamiento primer nivel HSM - Servidores HSM - Servidores “StarLife” datamovers y exports - Los 3 servidores actuales TSM del BSC encargados del backup mediante links 10 Gbit Ethernet - La conexión hacia otras redes del BSC y los enlaces hacia la red de datos del almacenamiento “StarLife” del lote1 - Los servidores de administración

Tabla 6 – Descripción hardware switches y redes HSM “StarLife”

Concepto	Valor mínimo	Valor ofertado
Red administración		
Número de switches proporcionados		
Marca switch		
Modelo switch		
Número de puertos 1GE por switch		
Número de puertos 10GE por switch		
Número de puertos libres		
Latencia introducida por el switch		
Red datos filesystem paralelo HSM		
Número de switches proporcionados		
Marca switch		
Modelo switch		
Número de puertos 1GE por switch		
Número de puertos 10GE por switch		
Número de puertos 40GE por switch		
Número de puertos libres		
Latencia introducida por el switch		

2.- Operacional

En este apartado se describen los requerimientos operacionales relacionados con los componentes del HSM del proyecto “StarLife”. Los componentes del Lote 2 de “StarLife” se instalarán en el CPD del sótano 1 del edificio D6 del campus Nord de la UPC, en la esquina entre la calle Dolcet y Sor Eulalia de Azinzu.

Ref	Descripción
Requerimientos operacionales	
R1	Para la instalación de toda esta infraestructura se deberán usar 3 racks Enterprise de machine type 7014 de la marca IBM, los cuales son racks standard de 42U propiedad del BSC. Estos racks están situados en un almacen del campus nord, en el proyecto se deberá incluir el transporte al edificio D6 como cualquier tarea asociada para poder ser usados, entre otras cosas se deberá: - Vaciar los racks de cualquier material que tengan enrackado - Proveer e instalar de tal forma que cada rack tenga 4 PDUs redundantes N+N (faltan 6 PDUs) - Proveer cables de alimentación monofásicos de 32A (faltan 10 cables) - Instalación de los laterales metálicos en el caso de que no los tuvieran puestos. De la misma manera, se deberá de incluir las tareas de desenracke y envío a almacen del campus nord de los racks que actualmente están instalados en la sala D6, que se necesitan retirar para sustituirlos por los 3 racks del tipo 7014.
R2	Como máximo todos los componentes del HSM y los tres servidores Tivoli actuales del BSC (3 servidores de 2U) deberán ocupar como máximo los 3 racks standard 42U antes mencionados.
R3	Para la instalación eléctrica para cada rack, existen 4 alimentaciones monofásicas II/16A Vigi 40A 30mA SI en cuadro eléctrico, con 4 bases Cetac aérea II/16A que se deberán sustituir por 4 bases Cetac aérea II/32A por rack. Si el consumo fuera superior al establecido con las protecciones en el cuadro eléctrico, se deberán incluir las modificaciones en la sala necesaria para poder alimentar la infraestructura actualmente en la sala existe (Cambio de protecciones eléctricas, tirada nuevos cables eléctricos, etc.), incluyendo la legalización de la instalación eléctrica “Memoria técnica de diseño” en caso de cualquier modificación.
R4	El peso de cada rack no deberá nunca superar más de 1100 Kg. x m2.
R5	Se deberá presentar en la documentación un esquema frontal con la ocupación de los racks de los diversos equipos presentados en la solución. En ella se deberá claramente especificar el hardware ofertado, como las U’s que ocupa cada componente de la solución. También se deberán de especificar por cada tipo de rack el número de cables/fibras que sale de cada rack para cada una de las redes definidas. Todos los elementos deberán enrackarse en los 3 racks indicados a excepción de los switches Ethernet que deberán enrackarse en el rack existente de comunicaciones que estará colindante a los 3 racks a instalar. El BSC se reserva el derecho de verificar y modificar la proposición de enrackado presentado por las empresas licitadoras, sin que dicho cambio le represente ningún coste adicional.
R6	Todos los servidores, switches y componentes de la solución HSM deben disponer de

Ref	Descripción
	fuentes de alimentación redundadas N+N. Que se conectaran a las PDU correspondientes redundadas.
R7	Se deberá presentar esquema de conexionado eléctrico interno de cada tipo de rack de la solución. Mostrando la redundancia de todos los elementos a nivel de alimentación a través de diferentes grupos de PDU.
R9	La sala dispone de un falso suelo de 25 cm para el paso de cableado de datos, como alimentación eléctrica. Se podrá establecer una visita al CPD confirmado día y hora por parte del BSC.
R10	Se deberá incluir todas las tareas de cableado de los componentes según el diseño acordado entre el BSC y la empresa licitadora.
R11	Los servidores deberán disponer de un sistema para poder consultar su temperatura de ambiente de trabajo, para poder definir alertas y avisos. De la misma manera, debe soportar la parada automática en caso de detectar temperatura muy alta.
R12	Se deberá incluir todos los elementos y componentes necesarios para el transporte y ubicación de los rack totalmente instalados en dicha sala CPD. De la misma manera se deberán de transportar los discos spare del BSC desde el almacén del campus nord al D6, en el caso de usarlos. Dado que la sala no dispone de almacén en el edificio D6 se deberá de realizar limpieza diaria del material sobrante a medida que se vaya instalando, la empresa licitadora se deberá hacer cargo de la limpieza del material sobrante. En la retirada de los racks actuales se deberá sanear el falso suelo del cableado de datos y eléctrico actual, como la retirada a un almacén de diverso material informático que dispone el BSC en ese CPD. Para cualquier movimiento de racks, hacer notar que las puertas de acceso tanto en el almacén como en el D6 son de 2 metros dificultando el paso de racks verticales.
R13	Se deberá establecer un plan que minimice los cortes de producción de los servicios que actualmente está funcionando en el CPD de instalación del HSM.
R14	Se deberán conectar todos los rack´s a la red de equipotencial existente. O la modificación de dicha red si hiciera falta.
R15	Se exigirá en la instalación cableado (fibra, eléctrico, cobre Ethernet, etc.) ordenado dado que el falso suelo es muy bajo, y todo el paso de cables se debe hacer por él.
R16	Todo cable o fibra que forme parte de la misma red y tecnología deberá ser del mismo color en toda la máquina y entre cualquiera de los diferentes componentes hardware que formen esa red. Cada red física deberá usar un color diferente entre ellas.
R17	Todo componente de la solución (rack, server, switch, cable, fibra, …) deberá ir debidamente etiquetado, para ser identificado físicamente de forma única según nomenclatura que se establezca entre el BSC y la empresa instaladora.

Ref	Descripción
	En los cables y fibras se deberá indicar origen y destino de la conexión usando la misma nomenclatura.
R18	La solución deberá incluir el montaje en racks de toda la solución, además de la recogida de todos los materiales sobrantes de la instalación.
R19	Se deberá presentar un diagrama de Gantt con la indicación de las diversas fases de la instalación del HSM desde el inicio del contrato hasta la finalización de la instalación del mismo.
R21	Se deberán retirar los rack existentes en el CPD del Edificio D6, que no están operativos, incluyendo los componentes que estén en su interior a un almacén del BSC situado el campus Nord a un vertedero oficial, se deberá realizar una limpieza diaria del material sobrante a medida que se vaya instalando, la empresa licitadora se deberá hacer cargo de la limpieza del material sobrante y de la limpieza final de obra, la recogida de residuos se deberá realizar a un vertedero oficial para su correcto reciclaje.
R22	Con la finalización del proyecto se deberá presentar la siguiente documentación en formato Autocad 2000 o superior y PDF: - Planos Asbuilt con todos los elementos instalados. - Fichas técnicas de todos los elementos instalados.

3.- Software

En este apartado se describe el software a proporcionar para el sistema HSM para el proyecto "StarLife".

Ref	Descripción
R1	El sistema operativo deberá ser UNIX like y compatible con el X/Open Standard POSIX 1003 (IS/IEC 9945). El sistema operativo deberá ser Linux, todos los componentes deberán llevar la misma versión de sistema operativo. Dicho sistema operativo deberá proporcionar soporte Enterprise y estar soportado por cualquiera del resto de componentes del software stack: Sistema de clustering, sistema de colas, sistema de ficheros, sistema de HSM, drivers tapes, etc. El sistema operativo ofertado deberá estar soportado por la última versión del filesystem paralelo ofertado y soportar la instalación de virtualización y containers. El sistema operativo ofertado deberá ser la última versión disponible cumpliendo los requerimientos anteriores.
R2	Se deberá aportar también todo el software necesario para la gestión de todos los componentes que formen la solución: Switches, Controladoras disco, etc.
R3	Cualquier elemento hardware (servidores de administración, servidores HSM, servidores datamovers) deberá soportar el standard IPMIv2, y deberá soportar estar gestionado por el sistema de clustering xCAT.
R4	Se deberán incluir el software y las licencias del sistema de ficheros paralelo (como por ejemplo GPFS) de tipo servidor y cliente para todos los nodos del sistema HSM ofertados para la infraestructura del lote 2 de "StarLife". De la misma manera las licencias adecuadas para el sistema de HSM.
R5	Se deberá de proveer de una solución software HSM usando nativamente el filesystem paralelo propuesto como primer nivel y el robot del BSC con su ampliación como segundo nivel. El sistema de HSM no podrá usar ningún almacenamiento intermedio y deberá copiar/mover datos directamente entre las cintas y el filesystem de primer nivel. El sistema HSM deberá ser High availability entre los dos servidores (activo-pasivo o activo-activo). Dicho HSM deberá cumplir los siguientes requerimientos/funcionalidades, los cuales se deberán poder definir por filesets de un mismo filesystem: - Posibilidad de definir cuotas por número de inodos y espacio ocupado online (en disco). Posibilidad de limitar el espacio máximo a ocupar en el espacio offline (cintas) - Definición de políticas de migración, mediante lenguaje tipo "SQL" integrado con el filesystem paralelo (ILM) - Ejecución de políticas de migración de forma paralela, distribuida entre diversos nodos - Políticas de migración por tamaño fichero - Políticas de migración por access/modification time - Políticas de threshold/watermark de ocupación - Posibilidad de definir número de copias en cinta de un dato - La copia en cinta se debería poder configurar para hacerse de forma síncrona o asíncrona - Cada usuario deberá poder consultar el estado de sus ficheros (online, premigrado, migrado) desde un cliente del filesystem paralelo

Ref	Descripción
R6	Sistema de clustering basado en open-source como puede ser xCAT que permita la instalación de los diversos servidores de la infraestructura HSM.
R7	La solución HSM proporcionada deberá disponer de un sistema de monitorización que permita: - Generación de alertas en el caso de fallo de cualquier componente - Consultar el estado de los diversos componentes de la infraestructura - Reporting de ocupación de los diversos filesets tanto en online como en offline
R8	Se deberá incluir en el proyecto la creación de las reglas de migración según los requerimientos recibidos por los diversos proyectos que van a ejecutar en la plataforma “StarLife”
R9	El licenciamiento del software de HSM deberá permitir de crecer en capacidad de uso tanto de primer nivel de almacenamiento como del segundo de forma infinita

4.- Mantenimiento y soporte

Ref	Descripción
R1	Se deberá proporcionar un mantenimiento de la solución HSM (hardware y software) propuesta con una duración mínima de 3 años desde el momento de aceptación del lote 2. Delante de fallos hardware se deberán reparar con una respuesta en 4 horas dentro de las horas de oficina (08:00 – 17:00) y con un servicio de soporte de Next Business Day. En caso de incidencias muy críticas que impliquen una afectación global de la producción de los clusters, se deberá proveer un seguimiento continuo 24x7 hasta la resolución de la incidencia.
R2	La empresa licitadora se hará cargo de la reparación y sustitución durante el periodo del proyecto “StarLife” de cualquier componente hardware adquirido para el sistema HSM, incluyendo los discos duros en spare del BSC (en caso de usarlos).
R4	El proyecto de instalación incluirá la instalación “llave en mano” del sistema HSM y de la integración del sistema actual de backup del BSC a la ampliación a realizar. El proyecto incluirá el diseño de las reglas de migración/recall y la optimización del sistema HSM para su producción.
R5	Se exigirá un trabajo en equipo con el departamento de operaciones del BSC, para la coordinación del diseño hardware/software de la solución. Cualquier plan o toma de decisión se deberá verificar con el departamento de operaciones del BSC antes de llevarla a cabo.
R6	Se proporcionará (mientras dure el mantenimiento): - Acceso a todo el software upgrade (incluyendo sistemas operativos, clientes filesystem paralelo y firmware) de todos los componentes de la solución - Punto único de suporte para el aviso de problemas e incidencias de cualquier componente que componga la solución
R7	Se exigirá soporte pro-activo, notificando y recomendado subidas de versión tanto de software como de firmware de cualquier componente de la solución.
R8	Se deberá entregar al final de la instalación una documentación digital en formato office en la que se describa: - Descripción general de los componentes de la solución - Esquemas físicos de los conexionados finales realizados - Versiones de firmware instalados y recomendaciones de configuración BIOS/firmware - Explicación procedimientos hardware para: Puesta en marcha, parada y diagnostico delante de problemas hardware - Proyecto de instalación y configuración del sistema HSM
R9	Toda la instalación y desarrollo del proyecto se deberá hacer on-site en las

	instalaciones del BSC bajo la supervisión del grupo de sistemas del BSC. En ningún caso se permitirá el acceso externo o remoto para la configuración o instalación de la solución presentada. Durante el mantenimiento no se permite el acceso remoto y toda modificación se debe hacer on-site.
R10	En la implantación de la solución presentada se exigirá la participación activa y presencial de los expertos de cada uno de los componentes que forman la solución: - Responsables de hardware/ desarrolladores de robótica, filesytem paralelo, sistema HSM - Desarrolladores o responsables técnicos de redes o switches ethernet Teniendo la posibilidad el personal del BSC poder intercambiar emails de forma directa con dichas personas con el fin de solucionar cualquier problema que surja durante el desarrollo e instalación de la máquina.
R11	El equipo técnico encargado de la instalación hardware y software deberá disponer de la formación y capacidades técnicas para la realización de este tipo de instalaciones, ya que es imprescindible para la correcta ejecución del contrato. Las personas asignadas deberán ser especialistas en sistemas de almacenamiento HSM ofertado, habiendo realizado con anterioridad instalaciones de sistemas HSM del mismo tipo. (Instalaciones HSM de más de 15 PB, con un primer nivel basado en filesystem paralelo como la solución aquí propuesta) Si la persona técnica designada en el desarrollo del proyecto por parte de la empresa licitadora no dispone de la experiencia adecuada, se exigirá la presencia de otro experto on-site para la realización del proyecto.
R12	Al final del proyecto se deberá realizar una formación de un mínimo de 1 jornada laboral sobre la solución HSM instalada al equipo de sistemas del BSC.
R13	Para la aceptación de la solución HSM se deberá de verificar todos los requisitos funcionales y rendimientos que se especifican en este lote 2. A continuación se listan los más destacados (sin que esta lista sea limitante): - Rendimiento de acceso a primer nivel de disco - Pruebas de redundancia de los diversos servidores que conforman la solución (servidores datos primer nivel de almacenamiento, servidores metadatos, Servidores HSM, servidores datamovers “LifeStar”) - Rendimiento de migración y recall siendo capaz de alcanzar el ancho de banda esperado de todos los drives ofertados - Escaneo del filesystem paralelo y generación de lista candidatos a migrar en línea con las IOPS que establece la solución de metadatos - Correcto funcionamiento de las reglas (migración y copias) definidas por el BSC - Comprobación del buen funcionamiento de la configuración de diversas copias síncronas o asíncronas - Ejecución y buen funcionamiento de los procesos administrativos (backup

	base de datos, etc.)
R14	Una vez acabada la instalación se deberá de comprobar que todos los requerimientos de la operativa (apartado 2) establecidos en este lote se cumplen. Como, por ejemplo, sin estar limitados a: - Redundancia eléctrica allí donde se requiera - Adecuación y optimización de la infraestructura - Repartición de las fases eléctricas - Cableado óptimo para el flujo de aire y refrigeración de todos los componentes - Etc.