



FENIX

RESEARCH INFRASTRUCTURE

Especificaciones técnicas para la implementación de FENIX Research e-Infraestructure en el BSC

The ICEI project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 800858.

© 2018 ICEI Consortium Partners. All rights reserved.



Índice

1. Contexto del proyecto	3
1.1 Proyecto global	3
1.1.1 The Human Brain Project (HBP)	3
1.1.2 El Proyecto ICEI y Fenix Research Infrastructure	3
1.1.3 Implementación	4
2. Definiciones	5
2.1 Definiciones de procedimiento	5
2.1.1 Definiciones	5
2.1.2 Categoría de requerimientos	5
2.2 Glosario	5
2.3 Unidades	6
3. Contexto	7
3.1 BSC	7
3.2 Objeto del pliego	7
3.3 BSC Data Center	9
3.4 Lote 1 Interactive Computing cluster	9
3.5 Lote 2 Archival Data Repository	9
4. Especificaciones técnicas	11
4.1 Lote 1 Interactive Computing cluster	11
4.1.1 Hardware	11
4.1.2 Software	12
4.2 Lote 2 Archival Data Repository	12
4.2.1 Hardware	12
4.2.2 Software	15
5. Benchmarks	16
5.1 Lote 1 Interactive Computing cluster	16
5.2 Lote 2 Archival Data Repository	17
6. Mantenimiento	17
6.1 Lote 1 Interactive Compute cluster	18
6.2 Lot 2 Archival Data Repository	18
7. Transferencia de conocimientos	18
7.1 Lote 1 Interactive Compute cluster	19
7.2 Lote 2 Archival Data Repository	19
8. Instalación y aceptación	19

1. Contexto del proyecto

1.1 Proyecto global

1.1.1 The Human Brain Project (HBP)

El **Human Brain Project** (HBP) es un Proyecto estratégico H2020 (H2020 FET Flagship) financiado por la comisión Europea. Su objetivo es acelerar los campos de la neurociencia, computación y medicina relacionada con el cerebro. Éste aglutina una federación de programas de investigación en neurociencia fundamental, simulación avanzada y modelado a multi-escala con la construcción de una infraestructura de investigación.



Human Brain Project

1.1.2 El Proyecto ICEI y Fenix Research Infrastructure

1.1.2.1 Presentación

El Proyecto ICEI (Interactive Computing e-Infraestructure) está financiado por la comisión europea y está formado por los centros de supercomputación punteros en Europa como BSC (España), CEA (Francia), CINECA (Italia), CSCS (Suiza), and FZJ/JSC (Alemania).

El plan de este Proyecto es crear un conjunto de e-infraestructuras que será federadas para formar la **Fenix Research Infrastructure**. El Human Brain Project será el primer Usuario de esta plataforma.

1.1.2.2 Servicios de la Fenix Infrastructure

La figura siguiente muestra como los usuarios de las diversas comunidades interactuaran con la Fenix infraestructure.

En esta infraestructura federada, cada centro deberá proveer los siguientes servicios y

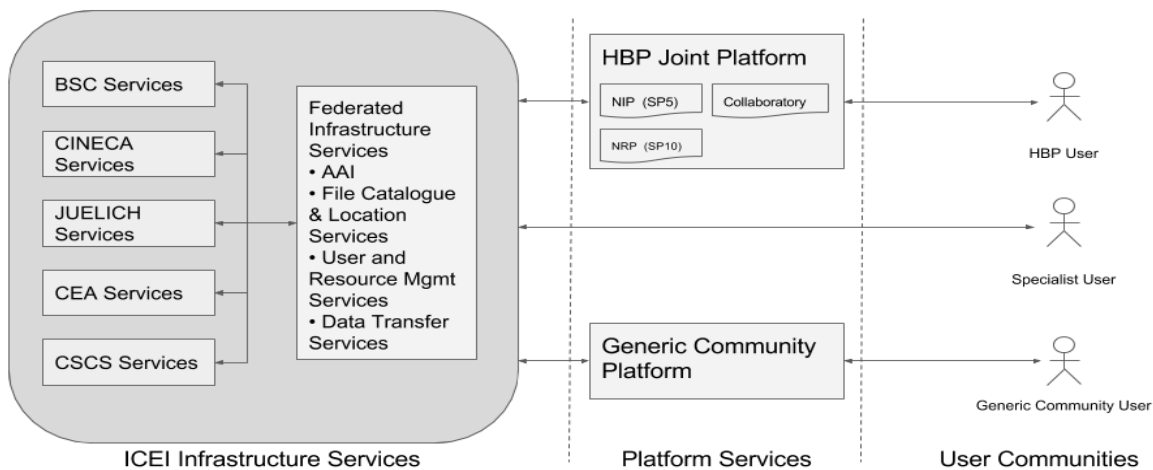


Figure 1 – Federación de la FENIX infraestructure

recursos:

- interactive computing services (IAC);
- scalable computing resources (SCC);
- virtual machines services (VM);
- active data repository (ACD);
- archive data repository (ARD).

Dentro de cada centro, estos servicios interactuarán según el siguiente gráfico:

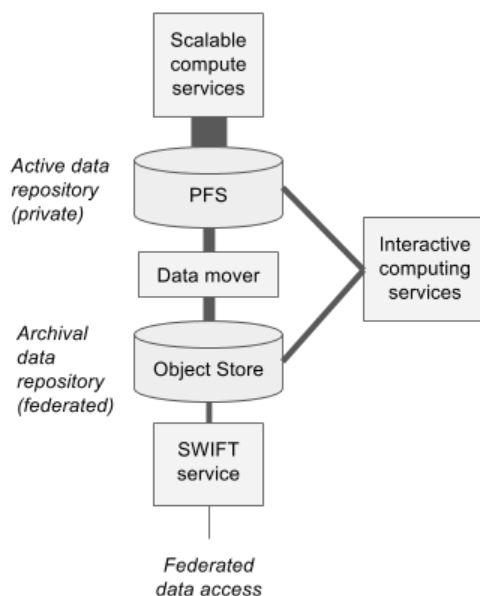


Figure 2 - Interacción entre componentes

1.1.3 Implementación

La Fenix e-infrastructure será adquirida mediante unos concursos coordinados. El resultado del global de la infraestructura se debe a un diseño conjunto entre los diversos miembros del Proyecto ICEI.

Cada centro es responsable para liderar la adquisición del su hardware y los servicios que van a proporcionar.

2. Definiciones

2.1 Definiciones de procedimiento

2.1.1 Definiciones

Term	Description
Candidato	Empresa participante en este concurso
Proveedor	El candidato que ha sido galardonado con el contrato como parte de esta licitación.

2.1.2 Categoría de requerimientos

Los requerimientos y características en este documento se categorizan de la siguiente forma:

Categoría	Descripción
MRQ	Mandatory Requirements son considerados esenciales para el Sistema que se está adquiriendo y deben cumplirse por las propuesta final. Mandatory Requirements se verificarán por cada propuesta. Las propuestas finales que no cumplan con todos los Mandatory Requirements serán descalificadas.
TC-1 TC-2	Target Capabilities son características deseables o niveles de rendimiento del Sistema a adquirir. A diferencia con los Mandatory Requirements, el no cumplimiento de Target Capabilities no implicará la descalificación de la propuesta. Propuestas que provean los Target Capabilities o los mejoren recibirán mejor puntuación. Target Capabilities están priorizadas. Nivel uno Target Capabilities (TC-1) se consideran de más importancia de nivel 2 Target Capabilities (TC-2).

2.2 Glosario

Term	Description
Backbone	Red Ethernet (10Gb, 25Gb, 40Gb or 100Gb).
BSC-CNS	Barcelona Supercomputing Center – Centro Nacional de Supercomputación
ESS	Elastic Storage Server
GPFS	IBM Parallel Filesystem
GPGPU	General-Purpose computing on Graphics Processing Units.
HA	High-Availability. Mecanismo para asegurar la disponibilidad de un servicio en caso de un fallo de un componente.
HBP	Human Brain Project. H2020 FET Flagship financiado por la Comisión Europea
HPC	High-Performance Computing.
HSM	Hierarchical Storage Management. Almacenamiento que puede usar varios niveles de tecnología (discos, cintas,...).
ICEI	Interactive Computing e-Infrastructure.
IOPS	Input Output oPerations per Second
LTFS	Linear Tape Filesystem
OpenStack	Plataforma de código abierto de cloud computing
PDU	Power Distribution Unit.
PRACE	Partnership for Advanced Computing in Europe.

RAID	Redundant Array Of Inexpensive Disks. Mecanismo para prevenir de fallo de un disco, almacenando información redundante en discos adicionales (mirror, paridad, ...)
Spectrum Archive	Hierarchical storage magement solución desarrollada por IBM
SPOF	Single Point Of Failure. Parte de un Sistema que en caso de fallo, impide el funcionamiento normal del sistema a nivel global.
SPOM	Single Point Of Management. Servidor(es) que proven de forma centralizada monitorización y servicios de administración.
Swift	Object storage de la plataforma OpenStack.
UPS	Uninterruptible Power Supply.
VM	Virtual machine / Máquina Virtual

2.3 Unidades

En este documento, la siguiente convención se ha usado para medir las capacidades de almacenamiento:

- 1 kilo-byte, escrito 1KB son 1,000 bytes;
- 1 mega-byte, escrito 1MB son 1,000 KB;
- 1 giga-byte, escrito 1GB son 1,000 MB;
- 1 tera-byte, escrito 1TB, son 1,000 GB;
- 1 peta-byte, escrito 1PB son 1,000 TB.

Estas unidades también se usan para medir el rendimiento, por ejemplo, 20GB/s corresponde a 20 giga-bytes (20×10^9 bytes) por segundo.

3. Contexto

3.1 BSC

El Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), establecido en 2005, sirve como el Centro Nacional de supercomputación en España. El centro alberga MareNostrum, uno de los superordenadores más potentes de Europa.

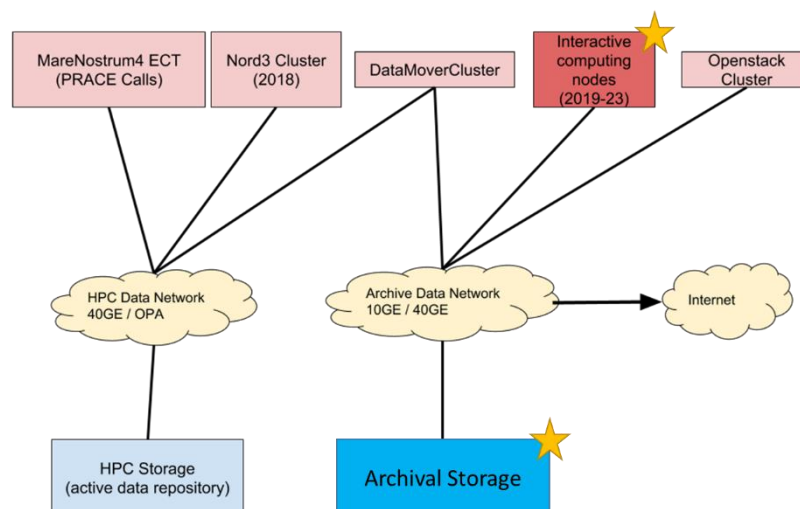
La misión del BSC-CNS es investigar, desarrollar y gestionar tecnologías de la información con un objetivo mayor de proporcionar innovación científica. Trabaja para conseguir estos objetivos en campos de la ciencia de la computación, ciencias de la vida y de la tierra.

BSC es un activo y reconocido participante en diversas iniciativas para la integración y consolidación de la supercomputación y gestión de datos en Europa y España. El BSC gestiona toda la información generado por las simulaciones de altas prestaciones ejecutados en sus clusters. En orden de poder cumplir con los requisitos de datos de los científicos, BSC dispone de más de 20 PB de disco disponible y una librería de 6 PB. A nivel internacional, BSC está trabajando en más de 42 proyectos europeos. A nivel de infraestructura, los principales proyectos de e-infraestructura que el BSC está participando son: EOSC-Hub, PRACE (donde el BSC es uno de los centros Tier-0) y HBP, en todos estos proyectos el BSC está participando como centro de datos como centro de supercomputación.

3.2 Objeto del pliego

El objetivo de este pliego es la adquisición de unos nuevos componentes de una infraestructura que va a ser usada en conjunto que infraestructura actual del BSC, para poder proporcionar los recursos suficientes para poder dar soporte a los requerimientos científicos de diversas comunidades como el Human Brain Project y PRACE.

Siguiendo la referencia de arquitectura técnica que se ha definido dentro del proyecto ICEI, la infraestructura local dentro del BSC tendrá los siguientes componentes (se marcan con una estrella aquellos que se adquirirán a través de este concurso):



Los siguientes componentes se proveerán con infraestructura ya existente en dependencias del BSC:

- **Scalable compute services** se proveerá mediante el cluster Nord3, un cluster basado en Intel SandyBridge y red Infiniband
- **Active Data repository** se proveerá con el almacenamiento actual de HPC que ya se provee a los diversos clusters de supercomputación
- Actuales BSC **Data Mover service** proveerán el servicio que el mismo nombre tiene para ICEI
- **Openstack cluster** será proporcionado parcialmente por el cluster Nord3 y algún servidor extra

Nuevos componentes deben ser adquiridos para completar la arquitectura referencia y así poder proveer de los servicios ICEI, concretamente: Interactive computing node cluster y el Archival Storage.

Todos los requerimientos de este pliego han sido definidos teniendo en cuenta el uso futuro de esta infraestructura dentro de Fenix y las comunidades de usuarios y sus necesidades dentro del ámbito del Human Brain Project, PRACE como cualquier otra comunidad científica Europea con altas necesidades de computación y uso de datos.

Este concurso está dividido en 2 lotes, uno por cada uno de los componentes a adquirir. Todo el texto que no está específico en un apartado de un lote, aplicará a los dos lotes por igual.

Lote 1 – Interactive Computing Cluster

Cluster de computo que tienen de ser capaces de soportar ejecución de jobs HPC y al mismo tiempo ser capaz de soportar sesiones interactivas, por ejemplo de visualización o de análisis de ejecuciones que los usuarios puedan interactuar durante la ejecución de su simulación.

Después del estudio de los casos de uso de las diversas comunidades científicas que harán uso de este cluster, los nodos se espera que tengan:

- GPUs : Para ayudar en la ejecución de diversas aplicaciones
- Alta capacidad de memoria (diversos TB): Implementado con diferentes tecnologías (NVDIMM/NVMe, ...)
- Alta densidad de almacenamiento local, para complementar la alta densidad de memoria principal
- Procesadores capaces de ejecutar de forma efectiva nuevos paradigmas de programación como containers, virtualizaciones, códigos de machine learning.

Lote 2 – Archival Data Repository

Para implementar el archival data repository para la Fenix Research Infrastructure dentro del proyecto ICEI, BSC extenderá su infraestructura de archivado tanto en capacidad como en rendimiento.

BSC dispone de una solución de almacenamiento por capas basado en las tecnologías:

- Robot TS4500 con cintas LTO8
- ESS GL6 actuando como primer nivel de almacenamiento de HSM
- Modulo metadatos ESS GS2S para almacenar los metadatos del filesystem
- Spectrum Archive actuando como la solución software que implementa HSM

Este lote del pliego debe proporcionar una extensión en capacidad y rendimiento para todos los niveles de almacenamiento de la solución HSM:

- Almacenamiento Metadatos
- 1^{er} nivel almacenamiento basado en discos duros y ESS Spectrum scale
- 2^o nivel de almacenamiento robot en cintas TS4500

La extensión ha de ser lo suficientemente grande para poder implementar los casos de uso de las comunidades de la futura Fenix Research Infrastructure como el resto de comunidades europeas científicas que usaran esta infraestructura.

Además, se deberá incluir los servicios necesarios para añadir la capacidad de federar este almacenamiento HSM con el resto de partners ICEI, usando la tecnología Swift.

3.3 BSC Data Center

Todo el archive storage y Interactive Compute cluster se instalará en el nuevo datacenter del BSC localizado en el sótano del edificio TG. En el siguiente esquema se presenta la sala y en verde están marcado los racks que pueden ser destinados a esta instalación.



3.4 Lote 1 Interactive Computing cluster

El licitador del lote1 deberá enrackear sus equipos en uno de los racks de archivo activo existentes o en alguno de los que proporcione la empresa que sea adjudicataria del lote2.

3.5 Lote 2 Archival Data Repository

El licitador del lote2 deberá proporcionar nuevos racks para los componentes de almacenamiento tanto para el robot como racks standard para proporcionar la ampliación de almacenamiento del primer nivel de HSM.

Actualmente los espacios de rack standard (RDHX) marcados en verde están ocupados, el lote 2 deberá incluir el descableado y desmontaje de todos esos racks. Los racks salientes se deberán preparar y trasladar a un almacén que el BSC indicará en el momento de implementación del proyecto.

Todo rack standard y cableado asociado que forme parte de la arquitectura de archivo activo 2 (PPI1, PPI2, StarLife-HSM) deberá trasladarse en conjunto con lo ofertado en este pliego a las posiciones indicadas de color verde, que habrán quedado liberadas por el transporte antes indicado.

Se deberá incluir cualquier tarea de cableado o infraestructura para implementar los movimientos antes descritos.

El espacio total destinado para toda la infraestructura de Archivo Activo2 y para esta licitación está formada por un máximo de:

- Espacio para 9 racks standard de 42U 19" (600mm ancho)
- Espacio para 11 racks robot TS4500

Todos los racks standard deben venir configurados con refrigeración por puerta trasera de agua que elimine el 90% del calor generado. El consumo eléctrico máximo por rack es de 28kW y tendrán 4 32A conectores monofásicos por rack.

El suelo soporta hasta 2000kg/m², todo cableado de alimentación y datos debe ir por encima de los racks por las bandejas existentes.

Durante la fase de instalación, cualquier acción o modificación necesaria para adaptar los componentes adquiridos al datacenter deberá estar incluido.

4. Especificaciones técnicas

En este capítulo, vamos a describir la especificación técnica para los 2 lotes de este concurso. Para ambos lotes se debe de proveer todo componente debidamente integrado e instalado con el resto de la infraestructura del BSC para su uso en producción.

4.1 Lote 1 Interactive Computing cluster

4.1.1 Hardware

	Description	Category
Interactive Compute nodes	Mínimo de 4 nodos de cómputo para interactive computing que deben ser capaz de ejecutar aplicaciones de análisis de forma eficiente. <ul style="list-style-type: none"> Mínimo de 512 GB de memoria principal volatil, configurada en un modo balanceado entre todos los canales disponibles y máximo rendimiento Non-volatile Memory tecnología para complementar la memoria anterior con una capacidad mínima de 4 TB Mínimo 4 GPUs por nodo para cómputo y visualización de la tecnología más avanzada Conectividad para montar el filesystem de HSM y conectividad hacia internet de mínimo 2 links de 25/40Gbit Ethernet por nodo 	MRQ
	El número de nodos de computo ofrecidos por encima del mínimo será evaluado	TC-1
	Características hardware para los nodos de Interactive computing nodes ofrecidos serán comparados y evaluados a nivel de: <ul style="list-style-type: none"> Capacidad de computación; Memoria ofrecida por nodo; NVMRAM ofrecida; Almacenamiento local ofrecido. 	TC-2
	Los nodos deben soportar cualquier tipo de virtualización y sus extensiones, como poder configurar las GPUs en modo pass-through o poder ofrecerlas a las máquinas virtuales.	MRQ
	Los servidores deberán poder ser administrados via out-of-line via una interfaz ethernet (ipmi, openbmc, o similar)	MRQ
	Estos nodos deberán integrarse en la red de administración del archive storage system, su instalación de sistema operative será gestionada por dicho cluster. Cualquier modificación de estos nodos puedan necesitar para adaptarse a dicha infraestructura deberá estar incluido (hardware/software).	MRQ
	Network	Es necesario proveer todos los componentes de red necesarios para poder conectar estos componentes a las diversas redes del sistema de HSM. Esas conexiones se

	deberán realizar de manera que se garantice la redundancia y balanceo de carga, siempre que la red no sea un factor limitante en el cluster de interactive computing.	
	Los componentes de red ofertados serán evaluados y comparados entre las diversas ofertas recibidas, tanto a nivel de redundancia, balanceo y rendimiento.	TC-1

4.1.2 Software

	Description	Category
General	La solución debe proveer con todo el software necesario para que funcione el interactive compute cluster como cada uno de los componentes por los que están formados.	MRQ
OS software	El sistema operativo ha de ser Linux y debe estar soportado por el resto de componentes hardware o software.	MRQ

4.2 Lote 2 Archival Data Repository

Para implementar el Archival Data Repository para ICEI, este lote propone extender la solución actual del BSC de archivado HSM basado en las siguientes tecnologías:

- Robot TS4500 con cintas LTO8
- ESS GL6s como primer nivel de almacenamiento de HSM
- ESS GL2S como módulo de metadatos para el filesystem
- Spectrum Archive como solución software para implementar HSM

Una extensión con similar o componentes compatibles deberá ser provista, para poder incrementar el rendimiento y la capacidad del almacenamiento de archivado del BSC, siguiéndose ver como una única infraestructura de datos.

La solución propuesta ha de ser modular y cada módulo has de ser capaz de incrementar la capacidad y rendimiento de cada uno de los componentes de almacenamiento que forma el sistema de archivado del BSC, concretamente:

- Almacenamiento de metadatos basado en Flash o SSD
- Filesystem paralelo basado en tecnología ESS (1er nivel de HSM) tecnología disco duro
- Tecnología de cinta en un robot TS4500 (2o nivel de HSM)

Todos los componentes ofrecidos deben integrarse como una única infraestructura administrativa; la cual también pueda configurarse virtualmente en varias particiones para asignar recursos a las diversas comunidades científicas.

4.2.1 Hardware

	Description	Category
General	Toda la infraestructura no debe presentar SPOF (single point of failure). Se deberá de proveer cualquier componente extra para poder cumplir con este requerimiento.	MRQ
	Toda la infraestructura deberá estar completamente balanceada, y capaz de distribuir la carga entre los diversos componentes hardware existentes.	MRQ
	En el evento de un fallo del almacenamiento la solución debe de proveer de un Sistema de recuperación automática para recuperar la situación. La duración del proceso de recuperación no debe exceder las 4 horas y el impacto de la reconstrucción en rendimiento debe ser menor al 10%. El proceso no debe afectar la integridad de los datos o su normal acceso.	MRQ
	La solución de almacenamiento debe proporcionar herramientas para reportar y proveer de una idea global del uso, tipo de datos y accesos que el almacenamiento está teniendo. Estas herramientas han de recolectar esta información de forma eficiente sin molestar la normal operativa del Sistema y los resultados deben ser presentados en casi tiempo real.	MRQ
	Todo servidor ofrecido deberá disponer de una interfaz de gestión out-of-line Ethernet interface (ipmi, openbmc, o similar)	MRQ
Metadata	Una capacidad neta del 3% sobre la capacidad total disponible en disco duro deberá ser proporcionada en Flash/SSD protegido por RAID o similar. Que sirva como referencia el siguiente cálculo: 10.5PB (ampliación HDD mínima de esta licitación) + 5.4PB (capacidad HDD actual) = 15.9PB (3% de este valor implica mínima capacidad total SSD de la solución total 477TB netos)	MRQ
	El almacenamiento de metadatos debe esta optimizado para IOPS y operaciones de metadatos	MRQ
1er nivel de HSM	Proporción de uno o diversos módulos de datos, con la finalidad de expandir el actual primer nivel de almacenamiento HSM del BSC con un mínimo de 10.5 PB de capacidad neta extra. El almacenamiento proporcionado se deberá verse como un único filesystem en conjunto con el actual ya existente. El almacenamiento ha de ser accesible y montado simultáneamente en miles de clientes con una interfaz POSIX.	MRQ
	El total de capacidad neta ofrecida será comparada y evaluada.	TC-1
	Un rendimiento mínimo extra de 50 GB/s es necesario proporcionar al actual primer nivel de almacenamiento de HSM.	MRQ
	El valor de rendimiento ofrecido será comparado y evaluado.	TC-1
	El diseño de la infraestructura de datos propuesta será evaluada.	TC-1

2o nivel de HSM	Mínimo de 30 PB para expandir el segundo nivel de almacenamiento del Sistema HSM del BSC (basado en tecnología de Cintas LTO8) debe ser proporcionado. Mínima extensión de 48 drives para leer y escribir en esas Cintas debe ser proporcionado. Se deben de proveer las extensiones de rack necesarias al robot del BSC TS4500 para poder instalar todo el material que la oferta proponga, en caso que no se disponga de espacio en el robot del BSC.	MRQ
	Se valorará la mejora en capacidad en cintas como en el número de drives ofertados.	TC-1
VM & Cloud Service	Mínimo de 4 servidores que deberán de proveer y expandir el actual servicio de VM y cloud que la infraestructura de archive dispone. Cada servidor deberá ofrecer un mínimo de 256GB de memoria RAM, soporte virtualización KVM y pci-passthrough para la optimización y uso de los dispositivos del servidor en las VMs. Cada servidor debe proveer conectividad con un mínimo de 50 Gbits por host.	MRQ
Administrative services	Se debe de proveer cualquier hardware necesario para poder realizar las tareas de administración de la infraestructura antes descrita y los servicios a proporcionar.	MRQ
	La infraestructura administrativa actual deberá ser expandida con nuevo hardware para mirar de cumplir con los requerimientos de rendimiento y capacidades aquí descritos.	MRQ
Export servers	Mínimo de 4 servidores conectados con al menos 2 links de 25/40 Gbit Ethernet cada uno, para ofrecer el almacenamiento HSM vía protocolos como: Swift object store protocol, NFS y CIFS. La configuración hardware ofertada debe ser la necesaria para poder saturar el ancho de banda que el servidor ofrezca.	MRQ
	Las características hardware ofertadas serán evaluadas y comparadas	TC-2
Network	Es necesario de proveer todo aquel hardware de comunicaciones necesario para conectar los nuevos componentes, garantizando la redundancia y balanceo de carga. Por otro lado, para que la red no sea un factor limitante en términos de rendimiento entre los diversos servidores y servicios a ofrecer. Mínimo del 10% de puertos ofertados deberán quedar libres para futuras ampliaciones en los switches Ethernet. Los switches fiberchannel deberán disponer de un mínimo de 4 puertos libres.	MRQ
	Los diversos componentes de red como su diseño serán evaluados y comparados con las diversas ofertas recibidas, a nivel de redundancia, balanceo de carga, y rendimiento.	TC-2
	La red ofertada se deberá tener en cuenta que se deberán de conectar los servidores que se ofrezcan en el lote 1.	MRQ

4.2.2 Software

	Description	Category
General	La solución propuesta debe venir con todos los componentes software necesarios para el funcionamiento óptimo de toda la solución de almacenamiento como una única, como cada uno de los componentes que la forman. El software ofertado debe ser completamente compatible con el software stack actual de la solución de archivado que dispone el BSC.	MRQ
	La infraestructura debe proveer de un Sistema de monitorización de todos los componentes y generar alertas de cualquier situación errónea.	MRQ
	La solución de monitorización ofertada será evaluada	TC-2
Data software	Todo el espacio de archivado, debe verse como un único espacio de filesystem.	MRQ
	<p>El filesystem propuesto ha de quedar completamente integrado con el actual existente en el BSC y proveer las siguientes características:</p> <ul style="list-style-type: none"> • Definir cuotas por inodos/ficheros y uso de datos (en total y por capa) • Definir políticas basadas en un lenguaje “SQL-like” que esté completamente integrado en el código del sistema de ficheros • Procesado y ejecución paralela y distribuida de la ejecución de las políticas • Las políticas de data management deben incluir parámetros como: tamaño, tiempo acceso/modificación, límite de capacidad por capa • Ejecución de migración distribuida y paralela de datos entre capas • Posibilidad de definir el número de copias de un mismo dato en las diversas capas • Definición de fileset para poder dividir el espacio del Sistema de ficheros por proyectos o servicios. Pudiendo definir cuotas o reglas/ políticas diferentes para cada uno. • Poder definir storage pools con diferentes blocksizes dentro de un mismo filesystem • Proveer un Sistema de metadatos distribuido y balanceado con “data in inode” • Debe incluir módulos para integrar Hadoop y object store de forma nativa • Posibilidad de conectar este Sistema de ficheros con varios clusters (multi-cluster) 	MRQ
	Los datos has de ser migrados automáticamente entre las diversas capas de almacenamiento, esto debe incluir:	MRQ
	<ul style="list-style-type: none"> • Definición de número de copias por capa • Configuración de copia (síncrona o asíncrona) 	

	<ul style="list-style-type: none"> El usuario debe poder consultar en que capa tiene un fichero almacenado 	
	La solución de datos presentada debe ser compatible con el actual sistema de archivado del BSC basado en (Spectrum Archive EE) y el formato LTFS	MRQ
	La solución ha de incluir todo el software necesario para que la infraestructura de datos pueda ser accedida por otros protocolos no sólo POSIX, concretamente: Swift (Object Store), NFS y CIFS	MRQ
OS software	El Sistema operativo ofrecido deberá ser Linux y deberá ser soportado por todos los componentes hardware y software, ya existentes en la infraestructura de archivado del BSC, como los nuevos que este concurso proporcionará.	MRQ

5. Benchmarks

Una lista de benchmarks han sido seleccionados para ayudar a evaluar el rendimiento técnico de las soluciones presentadas.

Los benchmarks seleccionados se pueden organizar en 2 grupos:

- Benchmarks genéricos o sintéticos que miden el rendimiento general
- Benchmarks de aplicación derivados del grupo denominado: ICEI application Benchmark suite. Este grupo se usará para medir el rendimiento de aquellas aplicaciones y casos de uso que esta infraestructura proveerá servicio en un futuro a través de la Fenix-RI.

El input, software e instrucciones de ejecución para los benchmarks se derivan del ICEI application benchmark suite, que será proporcionado durante el procedimiento de este concurso, aunque se describe en la siguiente página web:

<https://indico-jsc.fz-juelich.de/event/87/material/slides/0.pdf>

Para cada uno de los dos lotes un diferente subconjunto de benchmarks has sido seleccionados, debido a las diferentes características que cada lote tiene. (Lote 1 infraestructura de computación y Lote 2 infraestructura de datos)

Cada proveedor debe proveer resultados reales o una estimación de rendimiento que producirá su solución en los benchmarks seleccionados. Esos valores serán confirmados luego con la infraestructura durante las pruebas de aceptación.

5.1 Lote 1 Interactive Computing cluster

Para el cluster interactive computing un grupo de benchmark sintéticos ha sido seleccionado, se deben de proveer valores con ejecuciones reales en nodos como los ofertados o una estimación de los valores que se espera de ellos en los nodos ofertados.

En caso de presentar una estimación, se deberá razonar los valores presentados, dichos valores se deberán luego validar en la fase de validación de la solución presentada.

- HPL peak performance
- HPL real performance
- HPCG
- Stream
- IOR (medir rendimiento de la memoria no-volátil)

Un Segundo grupo de benchmark de aplicaciones de usuario ha sido seleccionado; estos benchmarks representa las aplicaciones que se espera que sean ejecutados en estos nodos.

- Neuron
- Nest
- Neuroimaging Deep Learning
- Elephant ASSET

5.2 Lote 2 Archival Data Repository

Para el lote del Archival Data Repository, diferentes benchmarks de entrada/salida han sido seleccionados para evaluar el rendimiento de las soluciones ofertadas.

Como benchmarks sintéticos, los siguientes valores deben proporcionarse por el proveedor de la solución:

- IOPS esperados del almacenamiento de metadatos
- Resultado de IOR del primer nivel de HSM

Para los benchmarks de aplicaciones de usuario, se ha seleccionado un benchmark basado en swift object store:

- Cosbench

6. Mantenimiento

Mantenimiento debe ser proporcionado (correctivo y preventivo) para cualquiera de los componentes de la solución que se provee (hardware o software), como de la solución como una unidad.

En el evento de un fallo, una respuesta deberá proporcionarse dentro de las siguientes 4 horas dentro del horario laboral del BSC (08:00 – 17:00) y se deberá de proveer de un servicio de mantenimiento de 'next business day'.

En el evento de un fallo crítico que implique un fallo global de toda la infraestructura se deberá realizar un seguimiento 24x7 hasta que el incidente quede resuelto.

Durante el periodo de mantenimiento, el proveedor tendrá la responsabilidad de todas las tareas relacionadas con la sustitución de cualquier componente hardware o software que falle.

El proveedor proveerá al BSC con el acceso a todas las actualizaciones de los diversos componentes software que la solución está compuesta, como pueden ser: paquetes de sistema operativo, software de datos, sistema de ficheros paralelo, firmware, etc.

Deberá existir un único punto de contacto para que el BSC pueda reportar problemas de cualquier de los componentes que forma la solución ofertada, se debe proveer de un teléfono y una web para poder abrir dichos casos.

Mantenimiento pro-activo es obligatorio, donde el proveedor proporcionará al BSC recomendaciones sobre actualizaciones firmware/software, indicando también que mejoras o fallos solventan dichas actualizaciones.

El candidato debe de proveer una lista de riesgos que pueden afectar negativamente a la instalación del sistema. Para cada riesgo, el candidato deberá dar una indicación de probabilidad que pase y proveer de una descripción del impacto esperado como las acciones de mitigación que se implementarían, todo esto como parte del contrato.

El candidato debe describir los roles de responsabilidad de todos los elementos implicados durante la instalación y pre-producción del sistema en la forma de RACI- (Responsible, Accountable, Consulted, Informed)-model.

6.1 Lote 1 Interactive Compute cluster

Mantenimiento del Lote 1 ha de ser de 4 años desde el momento que el lote ha sido aceptado y se encuentra en producción.

6.2 Lot 2 Archival Data Repository

Mantenimiento del lote 2 ha de ser de 4 años desde el momento en que el lote es aceptado y está en producción.

El licitador debe de incluir una extensión de mantenimiento de hasta 4 años (desde la fecha de que este lote ha sido aceptado), de todos los componentes actuales del sistema de archivado del BSC.

Se valorará la ampliación de mantenimiento por encima de 4 años.

7. Transferencia de conocimientos

Como parte de la instalación del Proyecto, el proveedor debe proporcionar la documentación describiendo el diseño y el log de instalación de todo el proyecto, explicando las diversas decisiones de diseño durante la fase de instalación.

El proveedor debe realizar reuniones regulares durante la instalación y trabajar en conjunto con el equipo de operaciones del BSC en las tareas de instalación.

El proveedor deber proporcionar sesiones de formación durante la fase de instalación y trabajará de forma conjunta en la instalación con el grupo de operaciones del BSC.

Se debe de proveer un mapa físico de los racks a proveer, junto con el cableado que va a ser conectado a cada rack y a cada switch.

Mapas físicos de todas las redes se deberán de proveer, indicando claramente que está conectado en cada puerto de toda la solución.

La documentación también debe incluir todos los procedimientos operacionales que se necesitan para mantener la infraestructura en funcionamiento óptimo.

Los documentos proporcionados han de ser en formato editable (Office).

7.1 Lote 1 Interactive Compute cluster

No se requiere nada extra para el Lote 1.

7.2 Lote 2 Archival Data Repository

Al final de la instalación, una sesión de training (mínimo 1 día entero) se deberá de proporcionar para explicar el uso, configuración y administración del object store Swift y su funcionamiento.

Todo el material de formación deberá ser entregado al BSC en formato editable.

Se valorará la inclusión de una formación de instalación y administración de una infraestructura openstack.

8. Instalación y aceptación

Instalación complete se espera que se complete en 4 meses después de la notificación del contrato, la calificación de pre-producción tendrá una duración mínima de 1 mes.

La instalación de ambos lotes se divide en las siguientes fases:

1. Instalación Hardware y software;
2. Aceptación provisional;
3. Cualificación de pre-producción;
4. Aceptación final

Instalación Hardware y software

La instalación software y hardware se completa una vez el proveedor ha entregado e instalado todos los elementos que conforman el sistema entero de acuerdo con la oferta presentada en este concurso.

Aceptación provisional

Después de la finalización de la instalación de hardware y software, y siguiendo a la declaración de buena disposición por el proveedor, el sistema se validará en esta fase.

El proveedor reproducirá los valores de rendimiento comprometidos en su oferta. También deberá demostrar los diversos Target Capabilities (TC-1 y TC-2) propuestos en su oferta.

Las siguientes pruebas serán realizadas:

- Comprobar concordancia con el contrato a nivel de hardware y software entregado;
- Pruebas realizadas por los representantes del proveedor y personas técnicas del BSC para poder validar el funcionamiento adecuado del sistema y su entorno;
- Resultado de los benchmarks.

Aceptación provisional será proporcionada después de esta fase.

Cualificación de pre-producción

El objetivo de la cualificación de pre-producción es comprobar que el sistema se comporta como es esperado durante una fase inicial de operativa. Elementos clave son la estabilidad, fiabilidad y rendimiento del sistema. Durante este periodo, el funcionamiento adecuado de los mecanismos necesarios del sistema en producción será validados. Durante este periodo, el proveedor deberá realizar los ajustes necesarios para hacer que la disponibilidad del Sistema sea la esperada según el pliego técnico.

La estabilidad de la máquina se verificará usando la metodología proporcionada en este lote técnico. Actividades como la transferencia de conocimiento se podrán realizar durante esta fase también.

Aceptación final

La aceptación final validará el funcionamiento adecuado del sistema a nivel global durante el periodo de cualificación de pre-producción.

Independientemente de estas fases, los siguientes requerimientos se deben de tener en cuenta:

- La infraestructura se debe entregar completamente instalada, y lista para ser usada (llaves en mano), de acuerdo con el diseño proporcionado por el BSC;
- Toda instalación y configuración debe realizarse en las dependencias del BSC; no se permitirá el acceso remoto para trabajar en las tareas de instalación. La instalación se deberá realizar de forma conjunta con el equipo de operaciones del BSC para facilitar la transferencia de conocimientos durante la instalación;
- Cualquier decisión que se deba realizar o plan a implementar durante la instalación deberá ser aprobada por el grupo de operaciones del BSC antes de implementarse.