

**SUMINISTRO DE UNA AMPLIACIÓN DE LA PLATAFORMA DE PROCESAMIENTO HADOOP PARA LA
GESTIÓN INTEGRAL DE DATOS CIENTÍFICOS MASIVOS DEL PIC**

**PLIEGO DE PRESCRIPCIONES TÉCNICAS
EXPEDIENTE: IFAE-2024-06**

SUMINISTRO DE UNA PLATAFORMA DE PROCESAMIENTO HADOOP PARA LA GESTIÓN INTEGRAL DE DATOS CIENTÍFICOS MASIVOS DEL PIC

EXPEDIENTE: IFAE-2024-06

El PIC es un centro científico-tecnológico orientado al procesamiento y gestión de grandes cantidades de datos científicos. Ubicado en el edificio de los servicios informáticos del campus de la Universidad Autónoma de Barcelona, está mantenido mediante un acuerdo de colaboración entre el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) y el Institut de Física d'Altes Energies (IFAE), respaldado por las administraciones central y autonómica.

El PIC se creó en 2003 con dos objetivos fundamentales: por un lado contribuir al procesamiento de datos de los experimentos del acelerador de partículas Large Hadron Collider (LHC) del Laboratorio Europeo de Física de Partículas, CERN, y por otro aplicar el conocimiento, metodologías y tecnologías derivadas del servicio del LHC a otras disciplinas científicas en las que también se necesitan gestionar grandes cantidades de datos experimentales.

El PIC es un centro Tier-1 dentro del proyecto para la Red de Computación Global para el LHC (WLCG) junto a otros 13 centros alrededor del mundo que custodian los datos experimentales producidos por los detectores de partículas ATLAS, CMS y LHCb que están ubicados en el LHC. Además, el PIC es el centro de datos de referencia para la colaboración MAGIC (Major Atmospheric Gamma-ray Imaging Cherenkov Telescope) y el survey de galaxias PAU (Physics of the Accelerated Universe). Es el centro de datos científico español de la misión Euclid, que es un proyecto de la Agencia Espacial Europea (ESA), y da soporte a los proyectos Dark Energy Survey (DES), Dark Energy Spectroscopic Instrument (DESI), LIGO/Virgo y Large Size Telescope - Cherenkov Telescope Array (CTA-LST1).

El PIC utiliza tecnologías de computación distribuida de alto rendimiento, incluyendo clústers, Grid, Cloud y Big Data, para desarrollar servicios de análisis de grandes cantidades de datos científicos. Es uno de los mayores centros de procesamiento de datos científicos de España, con 60 Petabytes de almacenamiento, 10000 núcleos de procesamiento, y conexión de red de área amplia de alto rendimiento a 200 Gbps.

El PIC ofrece servicios especializados para la comunidad científica orientados al soporte al análisis de datos. Los principales son el servicio de Almacenamiento Masivo de Datos, que actualmente cuenta con un espacio total de 60 PB, y el servicio de Computación,

Plan de Recuperación, Transformación y resiliencia - Financiado por la Unión Europea - Next GenerationEU

Procesamiento y Análisis de Datos, que actualmente dispone de una potencia de cómputo de en torno a 10000 núcleos de CPU. Ambos servicios se complementan para proporcionar la funcionalidad requerida para gestionar el ciclo de vida completo de los datos en una investigación científica. En la página web del centro bajo <https://www.pic.es/services> se describen brevemente todos los servicios ofrecidos por el PIC.

Objeto del Contrato

El objetivo de esta actuación es la ampliación de la plataforma de procesamiento (clúster) Hadoop en el Port d'Informació Científica (PIC) con el que los investigadores pueden generar, almacenar y analizar grandes conjuntos de datos, además de compartirlos y distribuirlos con toda la comunidad científica.

El clúster, que se ubica en el servicio común del PIC denominado Big Data, necesita aumentar la capacidad de la plataforma ya existente y posibilitará gestionar integralmente un flujo de trabajo de datos dentro del mismo servicio. Basado en el sistema de ficheros distribuido HDFS y todo su ecosistema tecnológico, va a facilitar la interacción y el manejo eficiente de grandes volúmenes de datos

Una parte fundamental del éxito de cualquier proyecto científico se mide por el impacto de sus resultados en la comunidad científica. Para ello, este clúster está conectado a 200 Gbps con redes de datos externas y facilitará, a través de diferentes herramientas, la creación, análisis, exploración, visualización y distribución de los datos, para impulsar su uso, teniendo en cuenta los principios de la ciencia abierta.

Descripción del Suministro

Para cumplir con el objetivo antes mencionado, se requiere desplegar un conjunto de Servidores Nodos de Cómputo de Plataforma Hadoop con las siguientes especificaciones:

1. Características de los Servidores Nodos de Cómputo de Plataforma Hadoop

Características para cada servidor:

- Chasis 2U, con 12 bahías frontales 3.5" hot-swap
- 24 cores x64 a 2.8GHz o superior (ej: AMD EPYC 7413)
- 512 GB de memoria RAM (DDR4 3200MHz o superior)
- 12 discos SATA 7200rpm 6 Gbit/s (perfil NAS) de 12 TB
- Dos (2) Discos NVMe iguales (DWPD >= 3) de 750GB a 1000GB cada uno
- Tarjeta de red de fibra dual 25G SFP28 con soporte PXE
- Consola de gestión remota dedicada IPMI/BMC
- No se requiere PSU redundada
- Controladora de discos modo JBOD/IT (no se requiere soporte RAID)

2. Requisitos eléctricos y refrigeración

- Los equipos serán instalados en una sala CPD con una distribución de pasillo caliente cerrado. La temperatura del aire de impulsión está entre los 22-25°C y en los pasillos calientes entre 37-42°C. Los equipos deben ser compatibles con estas condiciones de trabajo.
- La circulación de aire en los equipos deberá ser de delante a atrás (front to rear).
- Las fuentes de alimentación deben tener una certificación de eficiencia mínima 80PLUS Gold.
- La conexión de los nodos servidores de disco a la red eléctrica deberá ser descrita y especificada en detalle.

Los servidores se instalarán en racks de 19', 600 mm de anchura y 900/1000 mm de profundidad.

3. Requisitos de Software y documentación

3.1. Sistema Operativo

- Los equipos suministrados y sus componentes deberán ser plenamente compatibles con sistemas operativos GNU/Linux tipo RHEL 9 o derivados (Rocky Linux, Alma Linux, CentOS Stream).
- Los componentes o dispositivos que requieran drivers o software específico deben ser compatibles con el sistema operativo, y ser proporcionados.

3.2. Software de gestión

- Si aplica, se debe proveer del software de gestión de la(s) controladora(s) RAID(s) para cada nodo de la solución. Este punto es opcional si y sólo si dicho software está disponible en los repositorios del sistema operativo escogido en la solución.
- El software de gestión debe ser compatible con el sistema operativo Linux escogido en la solución.

3.3. Configuración

Todos los servidores de discos deberán ser entregados con la misma configuración. En particular: configuraciones de discos, versiones de firmware (i.e. BIOS, BMC, controladora(s) RAID, discos, configuraciones de BIOS). Previamente a la entrega, el proveedor deberá contactar con Port d'Informació Científica para detallar dichas configuraciones.

4. Servicios

Se requiere incluir la instalación del hardware en los racks del PIC en el precio de la oferta, así como el desalojo y/o reciclaje de su embalaje.

Plan de Recuperación, Transformación y resiliencia - Financiado por la Unión Europea - Next GenerationEU

Los herrajes necesarios para montar los servidores de cómputo y de gestión se deben proveer, así como sus cables de corriente.

Se deben proporcionar los rieles para el desplazamiento horizontal con las características adecuadas para soportar con seguridad el peso y proveer estabilidad al equipo.

El adjudicatario deberá disponer de capacidad de servicio y soporte técnico in situ en el área de Barcelona, así como de los medios personales y materiales para cumplir las obligaciones contractuales establecidas en el presente PPT y, en especial, las relativas al mantenimiento y garantía de los productos objeto de suministro.

5. Garantía

- El período de garantía y mantenimiento y licencias (si son necesarias) será de cinco (5) años a partir de la fecha de entrega de los equipos.
- La garantía cubrirá cualquier tipo de fallo hardware de los servidores bajo condiciones de operación continua 24x7 a máxima carga en todos sus parámetros y en el entorno habitual del PIC.
- La subsanación de fallos para los Servidores Nodos de Cómputo se efectuará en un máximo de 5 días laborables a partir de la recepción del reporte de avería. Los días laborables serán definidos por el calendario usado por las instituciones públicas en la ciudad de Barcelona.
- La empresa adjudicataria deberá proveer para los Servidores Nodos de Cómputo un mínimo de 2 discos SATA, 1 disco NVME y 1 PSU de stock con las mismas características de los contenidos en los Servidores Nodos de Cómputo para el reemplazo inmediato en caso de fallo ("stock on-site").
- El coste de transporte de equipos o piezas de repuesto desde y hacia las dependencias del PIC para efectuar reparaciones o reemplazos será incluido como parte de la garantía. Asimismo serán incluidos aquellos costes asociados a los traslados del personal de la empresa adjudicataria.
- Cualquier reemplazo se realizará con componentes que sean exactamente iguales a los originales en todas sus características. Excepcionalmente, y bajo acuerdo explícito del PIC, el reemplazo podrá ser compatible con el original pero de capacidad superior.
- En caso de haber fallos que afecten a más del 25% de alguno de los componentes del Sistema durante el primer mes de operación, el proveedor deberá sustituir el 100% de dichos componentes por otros con iguales prestaciones y cuyas especificaciones sean acordadas con el IF AE. En caso de no hacer dicha sustitución, el IF AE se reserva el derecho a cancelar la orden de compra y devolver los equipos.
- En caso de haber tres o más fallos de hardware en un mismo componente del Sistema durante cualquier período de seis meses dentro del periodo de garantía, no se

Plan de Recuperación, Transformación y resiliencia - Financiado por la Unión Europea - Next GenerationEU

aceptará ninguna reparación adicional de dicho componente y el proveedor deberá proveer un componente nuevo de reemplazo.

- El mantenimiento cubrirá todo lo relacionado con el soporte técnico de los equipos, para lo cual la empresa adjudicataria debe disponer de un centro de soporte técnico que atienda al cliente en al menos uno de los siguientes idiomas: catalán, castellano o inglés. La disponibilidad del soporte técnico in situ debe ser como mínimo 8x5 (estándar de horas laborables durante días laborables).
- Como consecuencia de las condiciones especificadas anteriormente, no se necesitan contratos de actualización, mantenimiento y soporte para el hardware y el software del sistema comprado durante el período de garantía especificado.
- La empresa adjudicataria deberá demostrar su capacidad de servicio y soporte técnico in situ en el área de Barcelona con antigüedad a 10 años, así como también documentar las instalaciones en España de los equipos de la marca ofertada.
- Todos los costes de proveer la garantía y mantenimiento especificado serán incluidos en el coste total de la oferta.

6. Entrega

Los servidores se deberán entregar en las dependencias del PIC en un máximo de 90 días naturales a partir de la firma del contrato correspondiente.

7. Documentación de la solución por parte de las empresas licitadoras

Será necesario incorporar como mínimo los siguientes puntos en la documentación de la propuesta:

- Descripción de la solución donde se indiquen los datos que permitan validar que la propuesta cumple con las especificaciones técnicas requeridas.
- Enlaces a las especificaciones técnicas de la solución.
- Cumplimentar Anexo I de este Pliego, Tabla de Especificaciones Técnicas del equipamiento
- Certificado del fabricante de la solución ofertada, que acredite que la empresa licitadora es un proveedor homologado del fabricante.
- Declaración Responsable de la empresa licitadora sobre su capacidad de servicio y soporte técnico in situ en el área de Barcelona, así como de la disposición de los medios personales y materiales; para cumplir lo detallado en el punto “5. Garantía”

Todas las obligaciones del contratista establecidas en el presente PPT se consideran obligaciones contractuales esenciales y su incumplimiento dará lugar a la resolución del contrato.

8. Nº Mínimo servidores

El número mínimo de servidores a ofertar son 7.

Bellaterra, a 17 de julio de 2024

Eugenio Coccia

Director

TABLA DE ESPECIFICACIONES TÉCNICAS DEL EQUIPAMIENTO

Servidores Nodos de Cómputo de Plataforma Hadoop

Número de Servidores	
Fabricante	
Modelo	
Número de U's Chasis	
Número de CPU	
Modelo CPU/Numero de Cores	
Memoria RAM Total por Servidor	
Número de DIMM y Tamaño	
Tipo de Memoria RAM	
Modelo Controladora	
Tipo Disco Interno	
Tamaño Disco Interno	
Número de Discos Internos	
Tipo disco	
Capacidad disco	
Número de discos	
Modelo Tarjeta Gestión Remota	
Modelo tarjeta de Red / Número de Puertos	
Modelo fuente	
Peso equipo	
Consumo eléctrico	
Stock on-site: 2 discos SATA, 1 disco NVME y 1 PSU	
Soporte	