

Servicio de Administración y Mantenimiento de Sistemas
Coordinación de Telemática
Dirección de Producción e Infraestructuras

NUEVO SISTEMA DE SUPECOMPUTACIÓN PARA AEMET

Tras el concurso para el suministro mediante arrendamiento sin opción a compra del sistema de cálculo de altas prestaciones, la empresa CAIXABANK EQUIPMENT FINANCE con el respaldo tecnológico de ATOS IT SOLUTIONS AND SERVICES resultó adjudicataria del contrato.

El proyecto consiste en un contrato de arrendamiento financiero sin opción a compra junto con servicios de soporte y mantenimiento hardware y software, administración de sistemas y horas de consultoría sobre optimización de aplicaciones. Según el Pliego de Prescripciones Técnicas (PPT), tiene 2 fases bien diferenciadas y en este documento se describen los componentes de los 2 clústeres incluidos en la primera de ellas más la ampliación esperada según el roadmap del fabricante del procesador.

Según los tiempos requeridos en el PPT, el primero de los 2 clústeres (Clúster Inicial 1) se instalará en no más de 3 meses a contar desde la firma del contrato y habrá 2 meses más para la migración de las aplicaciones. El segundo de los clústeres (Clúster Inicial 2) estará instalado en 5 meses a partir de la firma del contrato, según mejora de un mes ofertada por CAIXABANK, siempre que se haya podido retirar el sistema actual ya que necesita de su infraestructura eléctrica.

Como resumen comparativo con el clúster actual, los números de los 2 nuevos sistemas, su ampliación en 2023 y del actual son los siguientes:

Tipo de recurso	Sistema BULL actual (2014-2020)	Sistema ATOS (2020-mitad 2023) Clústeres iniciales	Sistema ATOS (mitad 2023-2025) Clústeres finales
Procesador	Intel Xeon E5 2600V2 (12 cores)	AMD EPYC™7742 (64 cores)	AMD EPYC™7742 (64 cores) + AMD EPYC™ Genoa (64 cores)
Nodos	324	280	376
Nº de procesadores	648	560	752
Nº de cores	7776	35.840	48.128
Memoria (TB)	31,5	70	106
Almacenamiento usable (TB)	360	500 (API) + 500 (AP2) + 2.500 (AD)	650 (API) + 650 (AP2) + 3.000 (AD)
Potencia eléctrica total (IT + refrigeración) media (KW)	78 (medida)	320 (estimada 160 por clúster)	320 (estimada 160 por clúster)
Potencia teórica de cálculo (TFlops)	168	1350	1680
Consumo energético máximo por operaciones (Watt/TFlop)	464,3	237,0	190,5



AEMet

La refrigeración vendrá proporcionada por puertas frías colocadas a la espalda de cada rack y con condensadoras de expansión directa mediante gas refrigerante R449a, todos los equipos de la marca Focus Cooling.

Cada uno de los clústeres iniciales está formado por:

- 140 nodos de cómputo montados sobre 35 chasis Bull Sequana X440 A5. Cada nodo tiene:
 - 2 procesadores AMD EPYC™7742 (64 cores)
 - 256 GB DDR4-3200 de memoria
 - 1 disco SSD de 240 GB
- Red de interconexión de altas prestaciones: InfiniBand HDR 100/200 full non blocking 1:1
 - Nivel 1 con 6 switches Mellanox HDR QM8790
 - Nivel 2 con 4 switches Mellanox HDR, 2 QM8700 y 2 QM8790
- Nodos de gestión: 2 Bull Sequana X430 A5 1s:
 - Un procesador AMD EPYC™ 7452 (32 cores)
 - 256 DDR4-3200 de memoria
 - 2 discos 1,9 TB SSD con tarjeta raid por HW
- Nodos de login: 2 Bull Sequana X430 A5 1s:
 - Un procesador AMD EPYC™ 7452 (32 cores)
 - 256 DDR4-3200 de memoria
 - 2 discos 1,9 TB SSD con tarjeta raid por HW
- Nodos de pre- y post-proceso: 4 Bull Sequana X430 A5 1s:
 - Un procesador AMD EPYC™ 7452 (32 cores)
 - 256 DDR4-3200 de memoria
 - 2 discos 1,9 TB SSD con tarjeta raid por HW
- Nodos de E/S: 2 Bull Sequana X430 A5 1s en alta disponibilidad para conexión con las redes de AEMET:
 - Un procesador AMD EPYC™ 7452 (32 cores)
 - 256 DDR4-3200 de memoria
 - 2 discos 1,9 TB SSD con tarjeta raid por HW
- Sistema de ficheros LUSTRE de 500 TB para datos de operación (API y AP2) con las siguientes características para cada uno de ellos:
 - 100 GB/s de ancho de banda acumulado y sostenido
 - 3 Bull Sequana X430 A5 en alta disponibilidad para servidores de datos (OSS)
 - 2 Bull Sequana X430 A5 en alta disponibilidad para servidores de metadatos (MDS)
 - 6 cabinas de almacenamiento Netapp EF600 con 20 discos NVME para datos
 - 2 cabinas de almacenamiento Netapp EF280 con 8 discos SSD para metadatos de Lustre
 - 2 switches Infiniband Mellanox HDR QM8790 de 40 bocas HDR 200 para red interna de datos



AEMet

- Almacenamiento NAS para propósito general (homes de usuario, repositorios SW, ...) de 17 TB netos proporcionado por una cabina de disco de doble controladora NetApp AFF C190 con 24 discos SSD.
- Software:
 - Licencias de Bull Smart Management Center
 - Licencias de sistema operativo Red Hat Enterprise Linux
 - Suite Intel Parallel Studio XE Cluster Edition
 - Suite de depuración Alinea DDT para 64 threads

Además existirá un sistema de almacenamiento LUSTRE para desarrollo (AD), accesible desde ambos clústeres con las siguientes características:

- 2.500 TB (~2,5 PB) netos de almacenamiento
- 90 GB/s de ancho de banda acumulado y sostenido
- 4 Bull Sequana X430 A5 en alta disponibilidad para servidores de datos (OSS)
- 2 Bull Sequana X430 A5 en alta disponibilidad para servidores de metadatos (MDS)
- 16 cabinas de almacenamiento Netapp E5700 con discos NL-SAS para datos
- 2 cabinas de almacenamiento Netapp EF280 con 8 discos SSD para metadatos de Lustre

A los 30 meses de la firma del contrato, en 2023, el clúster inicial 1 serán ampliado con:

- 48 nodos de cómputo en 12 chasis Bull Sequana X440 A6. Cada nodo con:
 - 2 procesadores AMD EPYC™ Genoa de 64 cores
 - 384 GB DDR4-3200 de memoria
- Aumento de 150 TB netos de almacenamiento para el sistema de fichero LUSTRE API
- Aumento a 125 GB/s de ancho de banda acumulado y sostenido para el sistema de ficheros LUSTRE API

A los 33 meses de la firma del contrato, en 2023, el clúster inicial 2 serán ampliado con:

- 48 nodos de cómputo en 12 chasis Bull Sequana X440 A6. Cada nodo con:
 - 2 procesadores AMD EPYC™ Genoa de 64 cores
 - 384 GB DDR4-3200 de memoria
- Aumento de 150 TB netos de almacenamiento para el sistema de fichero LUSTRE AP2
- Aumento a 125 GB/s de ancho de banda acumulado y sostenido para el sistema de ficheros LUSTRE AP2
- Aumento de 500 TB netos de almacenamiento para el sistema de fichero LUSTRE AD
- Aumento a 100 GB/s de ancho de banda acumulado y sostenido para el sistema de ficheros LUSTRE AD