



Comunidad  
de Madrid

Dirección General  
de Investigación  
e Innovación Tecnológica  
CONSEJERÍA DE EDUCACIÓN,  
CIENCIA Y UNIVERSIDADES

UNIÓN EUROPEA  
Fondos Estructurales  
*Invertimos en su futuro*



# PROGRAMAS DE I+D EN TECNOLOGÍAS 2018

ACRONIMO: CABAHLA-CM

TITULO PROGRAMA: Convergencia Big dAtaHpc: de Los sensores a las  
Aplicaciones

PRESUPUESTO CONCEDIDO: 869.400,00 Euros

Madrid, 17 y 18 de abril de 2024



# CABAHLA-CM - ¿Quiénes participamos?

- **Arquitectura y Tecnología de Sistemas de Computación (ARTECS)**  
Universidad Complutense de Madrid / Facultad de Informática  
(coordinador)
- **Arquitectura, Comunicaciones y Sistemas de Computación (ARCOS)**  
Universidad Carlos III de Madrid / Escuela Politécnica Superior
- **Ontology Engineering Group (OEG)** Universidad Politécnica de Madrid /  
E.T.S.I. Sistemas Informáticos
- **Scientific IT Research Activities and Knowledge (Sci-Track)** Centro de  
Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT)





# CABAHLA-CM - ¿Qué objetivos planteamos?

- **Objetivo global:** Mejorar la integración de los paradigmas de computación de altas prestaciones (HPC) y Big Data.
  - Diseñar e implementar el framework de CABAHLA, capaz de explotar el paralelismo a nivel de nodo y aceleradores y, a partir de él, ejecutar flujos de trabajos intensivos en datos sobre infraestructuras heterogéneas de gran escala minimizando los cuellos de botella de la localidad de los datos.
  - Por un lado, mejorar la eficiencia energética y computacional de los supercomputadores, disminuyendo así su consumo y la emisión de CO<sub>2</sub>, y, por el otro, aumentar el rendimiento en los casos de uso seleccionados:
    - Captación y modelado de datos de sensores para la predicción de radiación solar con alta resolución espacio-temporal.
    - Procesamiento de datos masivos en imagen médica del cerebro para detectar patrones anómalos.





# CABAHLA-CM - ¿Qué resultados hemos obtenido?

- **Objetivo 1:** Se ha proporcionado un marco integrado para el análisis eficiente de datos en infraestructuras de HPC.
- **Objetivo 2:** Se han desarrollado distintas propuestas para la mejora del rendimiento de los nodos aprovechando heterogeneidad y gestionando los recursos de manera eficiente.
- **Objetivo 3:** Se han diseñado mecanismos de gestión de datos para mejorar el rendimiento de sistemas de gran escala, balancear la carga y optimizar el movimiento y almacenamiento de datos y para soportar almacenamiento en memoria para análisis de datos en tiempo real.
- **Objetivo 4:** Se han desarrollado sistemas de monitorización de aplicaciones en sistemas de gran escala y se han utilizado para la planificación y el aprovisionamiento dinámicos de las mismas.





## CABAHLA-CM - ¿Qué resultados hemos obtenido?

- **Objetivo 5:** Se ha diseñado, fabricado y calibrado un conjunto de nodos sensores para medición de radiación solar y se han desarrollado modelos predictivos de la irradiación solar basados en aprendizaje automático profundo.
  - No se ha podido realizar el despliegue de la red de nodos sensores debido a los problemas causados por la falta de suministro y la pandemia del Covid19.
- **Objetivo 6:** Se han utilizado distintas propuestas de los objetivos anteriores para mejorar el rendimiento de sistemas de imagen médica como TAC.





# CABAHLA-CM - ¿Qué resultados hemos obtenido?

- Publicaciones en revistas: casi 70 indexadas en JCR (+30 Q1, +20 Q2)
- Publicaciones en congresos internacionales: +70
- Trabajos Fin de Máster: +40
- Tesis doctorales: 15 + 6 en 2024
  - David del Río (2019), Silvina Caino (2020), Saúl Alonso (2022) y Luis Costero (2022) han obtenido el Premio Nacional de Informática SCIE-BBVA en su modalidad Jóvenes Investigadores.
  - Saúl Alonso (2022) ha obtenido el premio Enrique Fuentes Quintana - Funcas a la mejor tesis a nivel nacional en la categoría de Ingeniería, Matemáticas, Física y Arquitectura.
  - Pierre Matri ha obtenido el premio extraordinario de doctorado de la UPM.
- Workshops organizados: 8





# CABAHLA-CM - ¿Qué resultados hemos obtenido?

- **Colaboraciones internacionales:**
  - CINVESTAV, México.
  - ARGONNE Labs, Chicago, EEUU.
  - Northeastern Univ., Boston, EEUU.
  - Harvard Univ., Boston, EEUU.
  - INRIA, Bordeaux, Francia.
  - TU Darmstadt, Alemania.
  - EPFL, Lausana, Suiza
  - IMEC, Lovaina, Países Bajos
- **Proyectos internacionales** establecidos relacionados con CABAHLA:
  - ADMIRE - Adaptive multitier intelligent data manager for Exascale
  - EOSC-SYNERGY - European Open Science Cloud - Expanding Capacities by building Capabilities
  - DARE - Digital Autonomy for RISC-V in Europe
  - ENTENTE - European Database for Multiscale Modelling of Radiation Damage
  - EoCoE-III - Fostering the European Energy Transition with Exascale





# CABAHLA-CM – ¿Cómo hemos continuado?

- Se ha continuado trabajando en las líneas del proyecto:
  - Advanced simulation-based predictive modelling for solar irradiance sensor farms
  - Xel: A cloud-agnostic data platform for the design-driven building of high-availability data science services
  - Hercules: Scalable and Network Portable In-Memory Ad-Hoc File System for Data-Centric and High-Performance Applications
  - Divide&Content: A Fair OS-Level Resource Manager for Contention Balancing on NUMA Multicores
  - Big-PERCIVAL: Exploring the Native Use of 64-Bit Posit Arithmetic in Scientific Computing
  - SYCL in the edge: performance and energy evaluation for heterogeneous acceleration
- Se han aplicado los resultados del proyecto a nuevos campos:
  - A multivariable sensor-agnostic framework for spatio-temporal air quality forecasting based on Deep Learning
  - A novel approach for large-scale environmental data partitioning on cloud and on-premises storage for compute continuum applications
  - ACORDE: A new application for estimating the dose absorbed by passengers and crews in commercial flights
  - Assessing opportunities of SYCL for biological sequence alignment on GPU-based systems
- Estamos elaborando una nueva propuesta.

