

[Inici](#) > EXTRACT: A distributed data-mining software platform for extreme data across the compute continuum

---

## **EXTRACT: A distributed data-mining software platform for extreme data across the compute continuum**

### **Description**

Data has become one of the most valuable assets, driving the digital transformation across many sectors. Current data mining solutions are optimized to deal with specific data requirements, but fail to cope as the data characteristics become extreme. There is therefore an urgent need for novel and holistic approaches to enable the development, deployment and efficient execution of datamining workflows across a heterogeneous, secure and energy-efficient compute continuum, while fulfilling the diverse extreme data characteristics.

To fill this technological gap, EXTRACT will deliver a data-driven open-source software platform integrating the most relevant technologies, to facilitate the development of trustworthy, accurate, fair and green data mining workflows able to generate high-quality actionable knowledge. The EXTRACT platform will improve the complete lifecycle of extreme data mining workflows, significantly enhancing performance, energy-efficiency, scalability and security, while fulfilling the extreme data characteristics in a holistic way. Moreover, multiple computing technologies, from edge to cloud to HPC, will be integrated into a unified and secure compute continuum. Specifically, the platform will feature enhanced data infrastructures and AI & big-data frameworks, novel data-driven orchestration and distributed monitoring mechanisms, a unified continuum abstraction and cybersecurity and digital privacy across all software layers.

The EXTRACT platform will be validated in two real-world use-cases with different extreme data requirements:

- Personalized Evacuation Route service, integrating data from the European data sources, Copernicus and Galileo, with 5G localization signals and smart city IoT sensors for civilian-centric crisis management; and
- Transient Astrophysics with a SKA pathfinder, processing extreme data from 2000 radio-telescopes for the real-time assessment of solar activity, generating knowledge for further scientific exploitation.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

**Source URL (retrieved on 29 set 2023 - 13:20):** <https://www.bsc.es/ca/research-and-development/projects/extract-distributed-data-mining-software-platform-extreme-data>