EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

TASKA Use Case

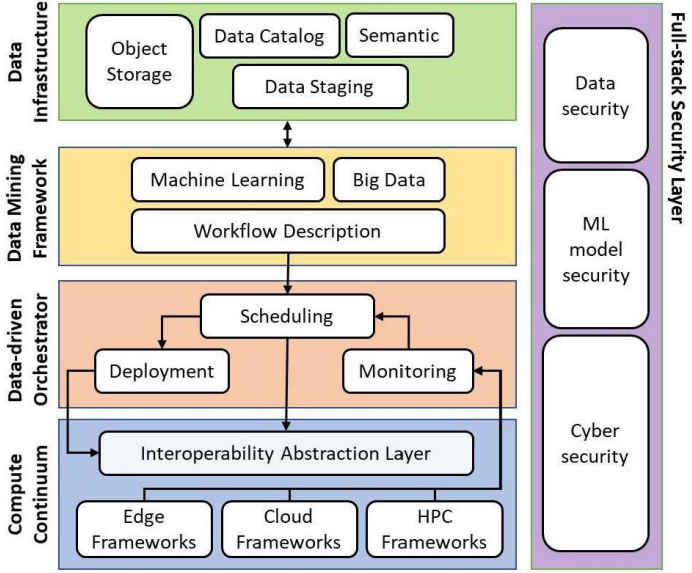
Transient Astrophysics with an SKA pathfinder

B. Cecconi and the EXTRACT & TASKA teams

The EXTRACT Project has received funding from the European Union's Horizon Europe programme under grant agreement number 101093110

Project Goal

- Delivering a data-driven **open-source platform** integrating cloud, edge and HPC technologies for trustworthy, accurate, fair and green **data mining workflows** for high-quality actionable knowledge



Example software platforms to be assessed

- **Lithops** — https://lithops-cloud.github.io/
  Lithops is a Python multi-cloud serverless data processing framework able to run massively parallel functions on data from Object Storage.

- **Nuvla** — https://nuvla.io/
  Edge-to-cloud management platform software, with support for marketplaces of business applications.

- **COMPS** — http://www.bsc.es/compss
  SW development framework for the transparent distribution of workflows in a distributed computing infrastructure (data-centers, clusters, clouds).

- **Ray** — https://www.ray.io/
  Ray is a high-performance distributed execution framework targeted at large-scale machine learning and reinforcement learning applications.

- **DataClay** — https://dataclay.bsc.es/
  Distributed storage system used in HPC, edge/cloud environments to maintain user-defined data consistency, by defining visibility scopes. It is integrated in COMPSs.

Partners

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

sixsq.

LOGOS RI
RESEARCH&INNOVATION

UNIVERSITAT
ROVIRA i VIRGILI

binaré

IBM

MATHEMA

ikerlan
MEMBER OF BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

l'Observatoire
de Paris
PSL

TASKA Use Case

Transient Astrophysics with an SKA Pathfinder

- **NenuFAR**: an SKA Pathfinder, located in Nançay (France)
- **Edge** data processing, in Nançay, real time data analysis:
  « Beam forming » mode, goal = Analog-to-information
  Detect (AI) structures => decision on resolution of data output
- **Cloud** data processing in Datalake (NenuFAR data centre):
  Post-processing of « Imaging » data:
  - orchestration of staging, computing, optimisation of workflow
  - generic processing (calibration, source removal…)
  - specific heavy processing (e.g., dynamic spectrum extraction from visibilities)

  Dynamic imaging of transient / variable sources in visibility space:
  - decomposition of components/calibration in visibility space

EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

Follow us on social media:

www.extract-project.eu

The EXTRACT Project has received funding from the European Union's Horizon Europe programmeunder grant agreement number 101093110