EXTRACT

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

Project Overview

DataWeek Jun14, 2023

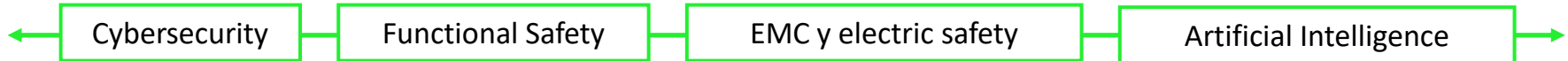Andoni Amurrio González, PhD

Researcher at Ikerlan, ICT

From **SENSORS** to the **CLOUD**

- **Embedded Systems**.
  - Hardware.
  - Software.
  - Communications.

| Cybersecurity | Functional Safety | EMC y electric safety | Artificial Intelligence |

IKERLAN – Digitization

From connected devices to the analytics in the cloud

CONNECTIVITY
IoT & EDGE COMPUTING

→ Cybersecure connection to cloud: **5G**

→ *Distributed* Intelligence: **Edge-to-Cloud Continuum**

→ Remote updating, management and acting

→ Gateways & Industrial protocols

Cloud PLATFORMS

→ Public (AWS, Azure, GCP), private and hybrid *Clouds*

→ *Datalake* creation

→ *Scalable* platforms:

From 10 to 100.000 data sources with the same architecture

→ Productivization *of Digital twins*

→ Device *fleets* management

Data Analytics & AI

→ Creation of AI & ML *models*

→ AI integrated, productivized and *optimized* in the cloud

→ Model *deployment* in Edge devices.

→ Ex: Predictive manteinance, image/video recognition, anomaly detection, etc.

EXTRACT Context

- Data has become one of the most valuable assets driving the digital transformation across a variety of sectors

- Current data mining solutions can deal with specific data requirements but fail to cope as the data characteristics become extreme

There is an **urgent need for novel and holistic approaches** to enable the development, deployment and efficient execution of data mining workflows **across** a heterogeneous, secure and energy-efficient **compute continuum**, while fulfilling the diverse extreme data characteristics

EXTRACT vision

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

- **Compute Continuum**
  - *HPC* to support massive parallel processing capabilities and advanced acceleration features
  - *Edge computing* to reduce data transmission latencies, minimize data privacy and security risks, and reduce energy consumption
  - *Cloud computing* to provide highly scalable storage systems and on-demand analytics technologies

- **Software Platform**
  - *Data infrastructures* and *AI & Big-data frameworks* to facilitate the development of complex workflows, including data processes and analytics methods
  - *Data-driven orchestration* to select the most appropriate computing resources to address extreme data
  - *Interoperability* of programming paradigms and execution models used across the compute continuum

> Addressing the above challenges in a holistic manner is fundamental for taking full benefit of extreme data

5

Objectives

- Enable the development of complex and secure data mining workflows
- Develop novel data-driven orchestration mechanisms to efficiently deploy and execute data mining workflows
- Fully exploit the performance capabilities of the compute continuum to address extreme data characteristics
- Deliver the EXTRACT software platform and demonstrate its benefits in 2 use cases
- Foster the adoption of EXTRACT technology by industrial and academic communities

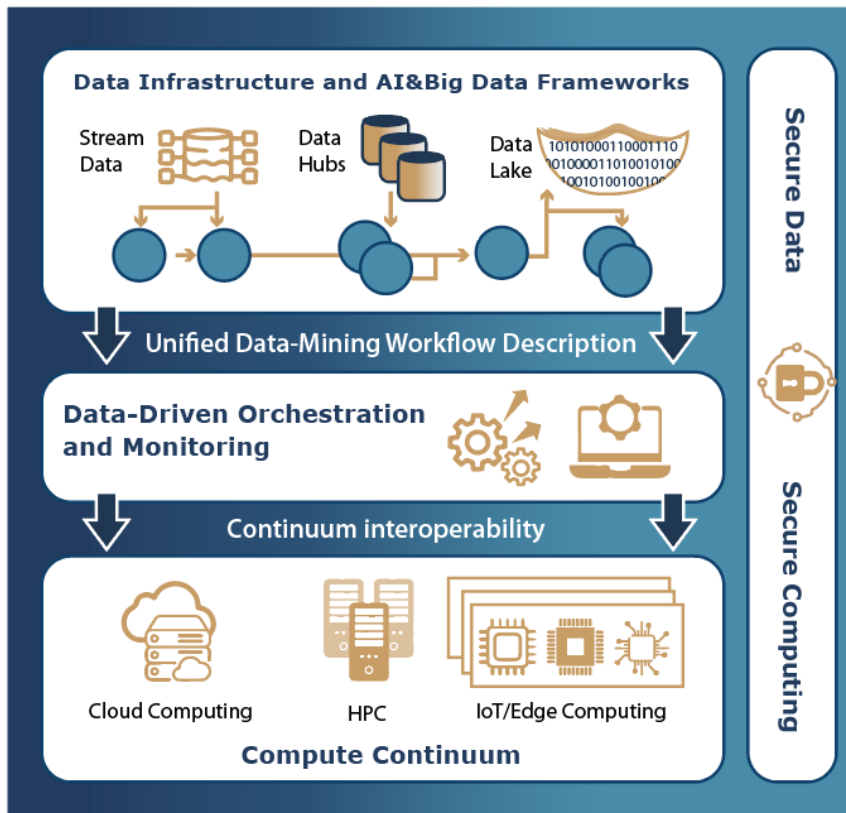# A Secured Compute Continuum Software Platform for Extreme Data

Figure 1. Main EXTRACT platform components . BSC©

EXTRACT Use Cases

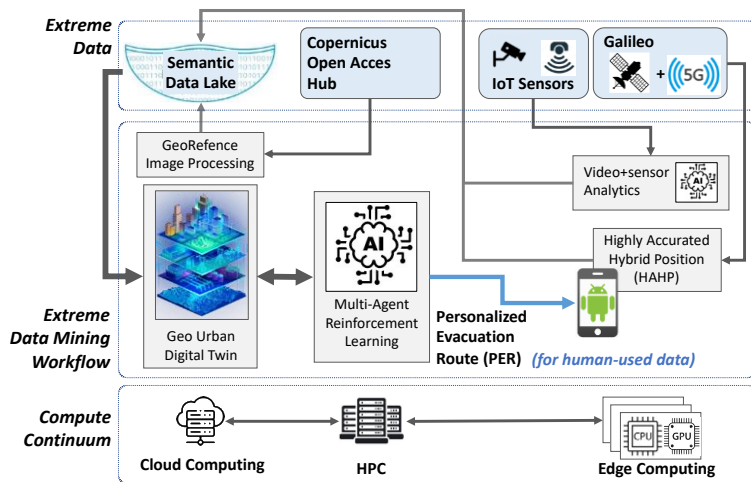**Personalized Evacuation Routing (PER) System**

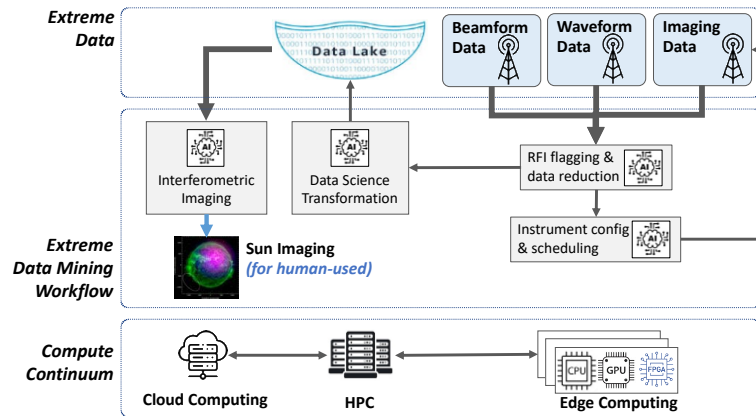**Transient Astrophysics with System a Square Kilometer Array Pathfinder (TASKA)**

Personalized Evacuation Routing (PER) System

Transient Astrophysics with a Square Kilometer Array Pathfinder (TASKA)

ESKERRIK ASKO

Follow us on social media:

www.extract-project.eu