NEARDATA

Extreme Near-Data Processing Platform

ABOUT THE PROJECT

Project title | NEARDATA: Extreme Near-Data Processing Platform
Programme | Horizon Europe
Topic | HORIZON-CL4-2022-DATA-01-05
Duration | January 2023 - December 2025 (36 months)
Project Coordinator | Universitat Rovira i Virgili
Funding from the EC | €3,913,585.00

PARTNERS

Universitat Rovira i Virgili (Spain)
Barcelona Supercomputing Center (Spain)
Technische Universität Dresden (Germany)
Deutsches Krebsforschungszentrum Heidelberg (Germany)
European Molecular Biology Laboratory (Germany)
EMC Information Systems International Unlimited Company (Ireland)
KIO Networks España SA (Spain)
Sano - Centre for Computational Medicine (Poland)
Scontain GMBH (Germany)
UK Health Security Agency (United Kingdom)

neardata.eu
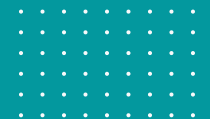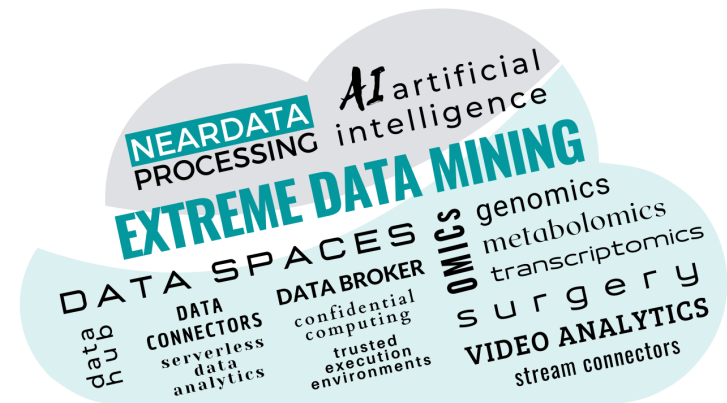
OBJECTIVES

The goal of NEARDATA is **to create an extreme data infrastructure mediating data flows between Object Storage and Data Analytics platforms across the Compute Continuum.**

Our novel **XtremeDataHub platform is an intermediary data service that intercepts and optimises data flows** (S3 API, stream APIs) with high performance near-data connectors (Cloud/Edge).

Finally, our unique **Data Broker service will provide confidential data access and orchestration of dispersed data sources** thanks to TEEs and federated learning architectures. Our NEARDATA platform is a novel technology for data mining of large and dispersed unstructured data sets that can be deployed in the Cloud and in the Edge (HPC, IoT Devices), that leverages advanced AI technologies and offers a novel confidential layer for trusted data computation.

NEARDATA

● ● ●   Provide high-performance near-data processing for Extreme Data Types

● ● ●   Support real-time video streams but also event streams that must be ingested and processed very fast to Object Storage

● ● ●   Create a Data Broker service enabling trustworthy data sharing and confidential orchestration of data pipelines across the Compute Continuum

EDGE

CLOUD

HPC   HPC

⊕ ACTIVEHUB   ⚙ DATA CONNECTOR   ⛁ OBJECT STORAGE   ▦ DATA BROKER
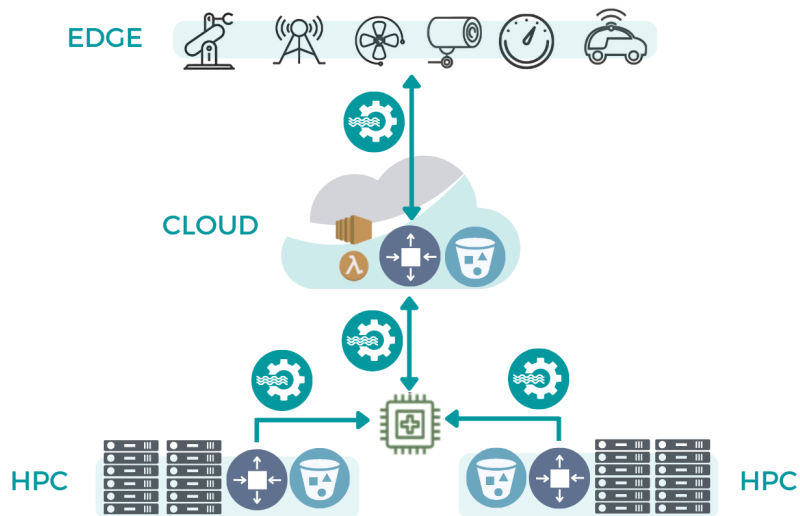
USE CASES

Genomics

Creation of methods, fast storage, and communications infrastructures to communicate distributed computing power with scalable storage systems, allowing efficient distribution of datasets across the system

Metabolomics

Expand the analysis of metabolomics raw data and boost external access and efficient re-use of open data. Creation of federated and Hybrid distributed architecture and ensuring data privacy but also shared global computations

Surgery

Create generalised machine-learning models that can aid surgeons during surgery and allow video data to be analysed in real-time and with low latency