

NEARDATA: Extreme Near-Data Processing Platform

Description

The main goal is to design an Extreme near-data platform to enable consumption, mining and processing of distributed and federated data without needing to master the logistics of data access across heterogeneous datalocations and pools. We go beyond traditional passive or bulk data ingested from storage systems towards next generation near-data processing platforms both in the Cloud and in the Edge.

In our platform, Extreme Data includes both metadata and trustworthy data connectors enabling advanced data management operations like data discovery, mining, and filtering from heterogeneous data sources. The three core objectives are to:

- Provide high-performance near-data processing for Extreme Data Types, including the creation of a novel intermediary data service (XtremeDataHub) providing serverless data connectors that optimize data management operations(partitioning, filtering, transformation, aggregation) and interactive queries (search, discovery, matching,multi-object queries) to efficiently present data to analytics platforms;
- Support real-time video streams but also event streams that must be ingested and processed very fast to seamlessly combine streaming and batch data processing for analytics;
- Create a Data Broker service enabling trustworthy data sharing and confidential orchestration of data pipelines across the Compute Continuum.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 11 ago 2024 - 13:11): <https://www.bsc.es/ca/research-and-development/projects/neardata-extreme-near-data-processing-platform>