

Federated engine for information exchange (Fenix)

Thursday, 30 November 2017 12:00 (15 minutes)

In contrast to experimental high-energy physics community and others, that already operate federated data infrastructures, neuroscience has to cope with a diverse set of data sources with their specific formats, modalities, spatial and temporal scales, coverage, and more (from high-resolution microscopes and magnetic-resonance data to electro-physiological data, from multi-electrode array measurements to brain simulations on HPC or neuromorphic systems) and with no fixed relationship between them. Thus, the scientific approaches and workflows of this community are a much faster moving target compared to, e.g., high-energy physics. Furthermore, there is the need for using HPC resources for processing these data. However, robust solutions do not exist currently for the federation of data services that can be readily adopted to fulfill the requirements of the neuroscience community.

Fenix is based on a consortium of five European supercomputing and data centres (BSC, CEA, CINECA, CSCS, and JSC), which agreed to deploy a set of infrastructure services (IaaS) and integrated platform services (iPaaS) to allow the creation of a federated infrastructure and to facilitate access to scalable compute resources, data services, and interactive compute services.

The setup of this federated data infrastructure is guided by the following considerations:

- Data are brought in close proximity to the data processing resources at different compute and data infrastructure service providers in order to take advantage of high bandwidth active data repositories as well as data archival services.
- Federating multiple data resources enables easy replication of data at multiple sites. This capability can be exploited to improve data resilience, data availability as well as data access performance.
- Services are implemented in a cloud-like manner that is compatible with the work cultures in scientific computing and data science. Specifically, this entails developing interactive supercomputing capabilities on the extreme computing and data platforms of the participating data centres.
- The level of integration is kept as low as possible in order to reduce operational dependencies between the sites (to avoid, e.g., the need for coordinated maintenance and upgrades) and to allow for the site local infrastructures to evolve following different technology roadmaps.

The Fenix federated infrastructure includes as main components:

- Scalable Compute Services;
- Interactive Compute Services;
- Active Data Repositories based on fast memory and active storage tiers;
- Archival Data Repositories; and
- Information/catalogue services.

The major advantages of the proposed architecture are: the use case driven design (it is being co-designed with continuous analysis and consideration for scientific neuroscience use cases), the scalability of the services, the easy extensibility which allows in the future to move to new state of the art solutions or to enable workflows also for other scientific communities.

The Fenix infrastructure will primarily offer resources to the neuroscience community as part of the Human Brain Project but it is meant to grow into a more generic provider.

The goal of this abstract is to present the status of the infrastructure, the technological choices made so far and the future plans.

Topic Area

The EOSC & EDI building blocks

Type of abstract

Presentation (15 minutes)

Primary author: FIAMENI, Giuseppe (CINECA - Consorzio Interuniversitario)

Co-authors: Mrs LUHRS, Anna (Juelich); Mr MCMURTRIE, Colin (CSCS); Mr MEZZANOTTE, Cristian (CSCS); TESTI, Debora (CINECA); Prof. PLEITER, Dirk (Juelich); Mr MUCCI, Roberto (CINECA)

Presenter: FIAMENI, Giuseppe (CINECA - Consorzio Interuniversitario)

Session Classification: EOSC building blocks presentations