Other identified i-Spaces

General Information

i-Space Name	IOP4HPDA – Italian Open Platform for High Performance Data Analysis by Cineca

Cineca is a non-profit organization, founded in 1969 as a Consortium of Universities. With its High Performance Computing (HPC) facility, and with excellent scientific skills, Cineca supports the world of the public and private research: it is the most powerful supercomputing center in Italy devoted to scientific and industrial research, and one of the most important worldwide. Moreover, Cineca develops IT systems for universities administration offices, for the MIUR, and for companies, health care Institutions, and public administration.

In particular the mission of the HPC department of Cineca is to accelerate the scientific discovery by providing HPC resources, data management and storage systems and tools. It also provides expertise on numerical simulation and data science in an Open Innovation paradigm. Member of BDVA, ETP4HPC, PRACE, core partner in EUDAT and Elixir, partner in Human Brain, Fortissimo2 and I4MS, in 2016 the HPC department of Cineca supported 1140 research projects and directly participated in 31 EU research projects, 40 research agreements and 12 industry projects.

The IOP4HPDA is made available by the HPC department of Cineca, as an environment for Big Data for research and innovation.

Address	Cineca, via Magnanelli 6/3, 40033, Casalecchio di Reno (BO), Italy
Contact Phone	+39 051 6171411
Contact Mail	r.turra@cineca.it
Partner Organisations	70 Italian Universities, 8 Research Centers, the Italian Ministry of Education, Universities and Research (MIUR)
Web site	https://www.cineca.it/

Other identified i-Spaces

BDV BIG DATA VALUE ASSOCIATION

Platform and Services Information

Platform(s) & Service(s)

Resource	Value
Hardware	20 Pflops performance; 360.000 cores (75.000 servers with 48 cores per server); 80 Nvidia Tesla K80; 200 Nvidia Tesla P100; 4 FPGA Intel DLIA; storage: 30 PB online and 30 PB offline; total RAM 1 PB; network: internal 100 Gb/s OPA 25 Gb/s ETH, external 10 Gb/s.
Software	Big Data Apache suite AAS; Deep Learning: Caffe, Theano, TensorFlow optimized for platform hardware characteristics in collaboration with Intel and Nvidia; Data Analytics: R, H2O, Octave, math libraries, I/O libraries; organized data repository for data deposit, retrieval and preservation; high performance data base PGSQL, MySQL, NEO4J; specific tools for bioinformatics (NGS pipelines) and for visualization.

Provided Services

The IOP4HPDA enables, supports and promotes the exploitation and development of Big Data applications by providing the following services:

- Exploitation of the Infrastructure: open access to HPC/HPDA storage and computing resources, Cloud Computing, Computing in batch, interactive and streaming modes.
- Advanced middleware and software tools.
- Data management: collection, preparation, annotation, curation, linking, security, access control, long-term preservation, post-processing.
- Data analytics: Predictive modelling, Supervised and unsupervised learning, Association rules, Sequential patterns, Link analysis, Recommenders, Natural Language Processing, Named Entities Recognition, Information Extraction, Automatic classification, Sentiment Analysis, Semantic metadata generation, Automatic annotation, Speaker segmentation, Automatic Speech Recognition, Video segmentation, Keyframes extraction, Semantic metadata generation from video items, Image recognition.
- Visualization: Remote visualization, Computer vision and visual computing, Computer Graphics, 3d modelling and rendering, Immersive device programming, Render farm service, Virtual Reality, Augmented Reality, Virtual museum and exhibition design.
- User support and Specialist support covering different scientific fields, technologies, programming languages, and techniques.
- Training and Education: Specialized training (workshops on massive data analysis and international summer schools on parallel computing, data analytics and computer graphics), Cooperation with universities (lab activity for master programs, post-doc programs), Knowledge transfer during the projects life cycle.

Other identified i-Spaces

BDV BIG DATA VALUE ASSOCIATION

- <u>Technology transfer and consulting</u>: Development of proof of concept and innovation projects for businesses to demonstrate the added value and ROIs.

Selected Projects and/or Success Stories

Risk Management Code Optimization for a Large Insurance Company

The risk assessment in the life insurance field may require considerable computing power. The algorithm that the Large Insurance Company was using took many hours and would not allow them to calculate the risk measurement with a nested Monte Carlo approach. In fact, nested Monte Carlo, involve two stages, scenario generation (outer stage) and portfolio re-valuation (inner stage), that produce millions of Monte Carlo trajectories to be executed for each of the millions of life policies. The simulation becomes very quickly a computational challenge. The Large Insurance Company contacted the HPC department of Cineca – IOP4HPDA for a PoC to demostrate the improved efficiency that could be obtained with efficient code parallelization and optimization. Nested Monte Carlo with parameters 100000x100 for all the 12M of policies could be achieved. The Large Insurance Company then decided to establish a commercial contract with Cineca for the provision of the service.

Internet as a data source for ISTAT, the Italian National Statistical Office

The HPC department of Cineca – IOP4HPDA started a collaboration with ISTAT to investigate the use of web scraping techniques, associated with text mining algorithms, in order to replace traditional tools of data collection and estimation, and/or to combine them into an integrated strategy.

In the first experiment, a sample of 8,600 enterprises' websites were "scraped" and the acquired texts (more than 200 M textual records) were processed in order to extract the same information that is provided by the standard questionnaire "Survey on ICT Usage and e-Commerce in Enterprises". The results were encouraging, with a satisfactory predictive capability of the fitted models. Within this experiment, the HPC department of Cineca – IOP4HPDA developed novel methods for extracting information from unstructured data and the usage of supercomputers resulted in significantly reducing the needed computational time. As a result a more extensive application of the approach in a "Census like manner", by considering all the Italian enterprises, has been performed and ISTAT is now evaluating the production phase.

Tax Fraud Detection for SOGEI, the Italian Revenue Agency Computing Centre

Cineca, with its IOP4HPDA data scientists, developed predictive models of the fraudulent behavior of companies in the entailment of tax credit and provided methodological solutions for impact and compliance assessment, in particular relating to training sample bias and model estimation and evaluation. The fraudulent behavior model enabled to increase the auditing success rate from 39% to 65% (precision).

Managing scientific data for various scientific communities

Among the scientific research projects that the HPC department of Cineca supports many can be reported as being both very successful and data intensive projects, eg **EMODnet** (European Marine Observation and Data Network) http://www.emodnet-chemistry.eu/ and **SPHINX** (Data Storage and Preservation of High resolution climate experiments) http://sansone.to.isac.cnr.it/sphinx/ .