

Porting the IRCCS Sant'Orsola Computational Genomic platform on INFN Cloud: a first proof of concept (Remote Presentation)

Wednesday, March 27, 2024 4:00 PM (30 minutes)

Modern technologies for DNA and RNA sequencing allow for fast, parallel reading of multiple DNA lines. While sequencing the first genome took 32 years, today with Next Generation Sequencing technologies we are able to sequence 40 genomes in about 2 days, producing 4 TB of text data (a file of about 100 GB per genome). This ability poses a challenge to computing infrastructures, which need to be able to ingest this amount of data and to process it through efficient genomic pipelines, exploiting heterogeneous resources such as CPUs, GPUs, HPC clusters and storage exposing different Quality of Service (QoS) to perform the analysis with the optimal cost-performance balance. At the same time, the computing platform needs to have user-friendly interfaces to be exploited by a plethora of different scientists like biologists, physicists, engineers, medical doctors and others. Finally, the API and data formats need to meet international de-facto standard and to be interoperable, to maximize their portability and to be runnable on cloud federations.

In this talk we describe the status of the Computational Genomic platform under development in the context of the collaboration between INFN and IRCCS AOU Sant'Orsola (the main research hospital in Bologna, Italy). The platform is deployed as a series of Openstack projects on EPIC (Enhanced Privacy and Compliance) Cloud: the high security partition of INFN Cloud certified ISO 27001 27017 27018. Presently it consists of about 1000 CPU cores with 5,8 TB RAM, 2 NVIDIA A100 GPU and 320 TB of storage (HDD, SSD, tape). We'll provide information about the performance reached on some sample genomic pipelines and on security measures adopted to guarantee GDPR compliance. Finally, we'll discuss possible synergies and interactions with other similar and broader initiatives at both national and international level.

Primary authors: COSTANTINI, Alessandro (INFN CNAF); MARTELLI, Barbara (INFN - CNAF); MICHELOTTO, Diego; SINISI, Francesco (INFN); SERGI, Giusy (INFN CNAF); GASPARETTO, Jacopo (INFN CNAF); MAGENTA, Letizia (INFN CNAF); GIANGREGORIO, Tania (IRCCS AOUBO); PIPPUCCI, Tommaso (IRCCS AOUBO)

Presenter: GASPARETTO, Jacopo (INFN CNAF)

Session Classification: Health & Life Science Applications

Track Classification: Track 2: Health & Life Sciences (including Pandemic Preparedness Applications)