

Integrating XNAT platform with the HPC infrastructure @INFN-Pisa

Antonino Formuso¹, Camilla Scapicchio¹, Silvia Arezzini¹, Alessio Fiori¹, Francesco Laruina¹, Francesca Lizzi¹, Simone Lossano¹, Enrico Mazzoni¹,
Alessandra Retico¹

¹ Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Pisa, Italia

The rising applications of data processing to medical imaging have increased the demand for IT platforms that are capable of storing and managing large amounts of structured and unstructured data. In a typical scientific research workflow, the core step is the analysis of such data, which often requires considerable computing resources.

We have deployed an instance of the XNAT (Extensible Neuroimaging Archive Toolkit, <https://www.xnat.org/>) platform at INFN-Pisa data center. In addition to managing the data lifecycle, XNAT features the definition of customized analysis pipelines, thereby promoting the reuse of semi-processed data and analysis code. This optimization is of crucial importance as it enables the creation of checkpoints and procedures (Workflow) which are tested and validated by the entire project team. In a first implementation of XNAT, the analysis workflows were processed within the same application server, with obvious limitations in terms of resource scalability. We have therefore investigated the possibility of interfacing XNAT with the HPC resources available at the INFN-Pisa data center. In this process, we have remarked the importance of decoupling the development of pipelines from the HPC environment (Scheduler), thus rendering the XNAT interface independent of the environment hosting the job execution.

To achieve our goal, we have resorted to technologies which are consolidated in the HPC world, such as container engines (e.g. Docker and Singularity), and job schedulers (e.g. Slurm). We have worked toward an approach that is easily replicable and fully transparent to the XNAT user. In the course of this work, we have explored several scenarios and we have experimented with the submission of pipelines from the XNAT web interface. With this configuration, pipelines will be Slurm jobs submitted to an HPC cluster partition, and eventually, results will be automatically uploaded to the XNAT project's data store.

Primary author: FORMUSO, Antonino (Istituto Nazionale di Fisica Nucleare)

Co-authors: RETICO, Alessandra (Istituto Nazionale di Fisica Nucleare); FIORI, Alessio (Istituto Nazionale di Fisica Nucleare); Dr SCAPICCHIO, Camilla (Istituto Nazionale di Fisica Nucleare); MAZZONI, Enrico (Istituto Nazionale di Fisica Nucleare); LIZZI, Francesca (Istituto Nazionale di Fisica Nucleare); LARUINA, Francesco (Istituto Nazionale di Fisica Nucleare); AREZZINI, Silvia (Istituto Nazionale di Fisica Nucleare); LOSSANO, Simone (Istituto Nazionale di Fisica Nucleare)

Presenter: FORMUSO, Antonino (Istituto Nazionale di Fisica Nucleare)

Session Classification: Poster

Track Classification: Calcolo distribuito