



## eXact-lab

### **Offerta: lavoro e stage**

The following is a list of projects for internship and thesis (bachelor/masters) to be performed at eXact-lab srl. Each project is part of services and products we are currently developing within our company; it is briefly presented with a list of tasks to be performed. We also indicate the required competences and, if available, some links/references.

A few mandatory requirements are common to all of them:

1. great interest and enthusiasm in Information technology applied to scientific and technical problems.
2. A good knowledge of Linux Operating System
3. Working knowledge of English language

Length of each project is not indicated but a commitment of at least 3 months is required. Work can be partially performed remotely but students are supposed to come periodically to Trieste and spend some time to work at our headquarter for some short periods (from 2 days up to 1 week).

Projects can of course last longer upon mutual satisfaction.

- [1. Software reengineering and parallelization/optimization of GEOtop scientific application \(Stefano Cozzini\)](#)
- [2. Comparing HPC vs Cloud; TCO of a weather forecast workflow on AWS and HPC cluster. \(Francesco De Giorgi\)](#)
- [3. Software reengineering of the python MIRTO package \(Stefano Piani\)](#)
- [4. PIGMATO \(Stefano Piani\)](#)
- [5. Benchmark analysis of the REGCM open source package for climate simulation \(Stefano Cozzini\)](#)
- [6. Development of a GUI for scientific data management \(Rossella Aversa/Stefano Cozzini\)](#)
- [7. Massive parallel processing for image classification with Spark \(Rossella Aversa/Stefano Cozzini\)](#)
- [8. OpenCL programming on FPGA accelerator \(Giuseppe Piero Brandino\)](#)
- [9. Deployment of a metadata search engine for scientific images \(Rossella Aversa/Stefano Cozzini\)](#)
- [10. energy efficiency profiling of HPC applications, per-job energy consumption reporting and energy-based scheduling \(Moreno Baricevic\)](#)
- [11. Implementation of an OpenStack security policy \(Moreno Baricevic, Piero Brandino\)](#)
- [12. OpenStack distributed deployment \(Moreno Baricevic, Piero Brandino\)](#)

# 1. Software re-engineering and parallelization/optimization of GEOTop scientific application (Stefano Cozzini)

Goal of the project is to complete the software reengineering eXact-lab is doing on the open source GEOTop package. eXact lab team is recently very active in restructuring the code and is ready to implement a parallel version.

Tasks include

- Data structure enhancements
- Profiling study
- Optimization study
- Parallel implementation of the computational core of the application.

Work done in collaboration with the GEOTop foundation.

Competences/Skills Needed:

- Deep Knowledge of C++ language
- Knowledge of GIT
- Knowledge of parallel programming (MPI/openMP)

References:

<https://github.com/geotopmodel>

## 2. Comparing HPC vs Cloud; TCO of a weather forecast workflow on AWS and HPC cluster. (Francesco De Giorgi)

Goal of the thesis is to port on the Amazon Cloud the daily weather forecast system we run for FVG region in our HPC infrastructure, comparing performance and total cost of ownership.

Task include:

- Setup the computational infrastructure on Amazon
- Plan and deploy the Weather Forecast system on Amazon
- Automate the workflow
- Analyze performance and cost
- Compare against the in-house solution

Competences/Skills Needed:

- Deep Knowledge of Linux shell and scripting (python/bash)
- Some familiarity with HPC systems
- Some familiarity with public cloud infrastructures and AWS platform

References:

- [https://depts.washington.edu/learnit/techconnect/cloudday/wordpress/wp-content/uploads/Kevin-Jorissen\\_Amazon\\_HPC-on-AWS-cfnCluster-and-WRF.pdf](https://depts.washington.edu/learnit/techconnect/cloudday/wordpress/wp-content/uploads/Kevin-Jorissen_Amazon_HPC-on-AWS-cfnCluster-and-WRF.pdf)
- <http://www.wrf-model.org/index.php>

### 3. Software reengineering of the python MIRTO package (Stefano Piani)

Goal is to create a robust open source package applying modern software engineering techniques to MIRTO, a python package to elaborate data coming from meteorological satellites. The package is at the moment under development in a joint collaboration among eXact lab and university of Wisconsin.

Tasks include:

- Refactoring of the Mirto code
- Adding external plug-ins to expand the features of Mirto to different setups

Competences/Skills Needed:

- Deep Knowledge of python language
- Some familiarity with HPC systems
- Experience with parallel computing
- Knowledge of OOP paradigm
- Experience using python packages for scientific computation (numpy, scipy, ...)

References:

<http://mkwc.ifa.hawaii.edu/satellite/polar/>

## 4. PIGMATO (Stefano Piani)

PIGMATO is an open source library that virtualize a Slurm cluster using some docker containers. It is particularly suitable for embarrassing parallel problems and for the situation where a fast deployment is strongly desirable. Currently, the library still has to be released and needs an accurate phase of testing and developing.

Tasks include:

- Add a REST interface for the library
- Test the library in different setups
- Add other functionalities
- develop a web frontend for the REST interface

Competences/Skills Needed:

- Good Knowledge of python language
- Familiarity with Docker working in HPC systems
- Familiar with basic networking concepts and, in particular, with VPNs

References:

[www.docker.com](http://www.docker.com)

## 5. Benchmark analysis of the REGCM open source package for climate simulation (Stefano Cozzini)

Goal is to provide some benchmarking reference number for the last releases of the RegCM scientific package for climate simulation.

Tasks include:

- Set up regression and benchmarking test suites.
- Implement continuous integration approach
- Run benchmarks and compare different versions of the code
- Accurate profiling of the package

Work done in collaboration with Graziano Giuliani from ICTP.

Competences/Skills Needed:

- Good Knowledge of Linux shell and scripting
- Familiarity with HPC systems
- Knowledge of F90 programming language

References:

<http://gforge.ictp.it/gf/project/regcm/>

## 6. Development of a GUI for scientific data management (Rossella Aversa/Stefano Cozzini)

The goal of the project is to develop and test a Graphical User Interface to the REST interface of the KIT Data Manager used within the NFFA-Europe project. At the moment, we developed a CLI for Linux based on Python, that should be wrapped by the GUI.

Tasks include:

- explore available graphical tools (PyGTK, Qt, etc..)
- develop the graphical interface
- add extra functionalities with the respect to the CLI (e.g., drag and drop)
- port and test on other operative systems (Windows and MacOS)

Work done in collaboration with NFFA-Europe project at CNR-IOM Institute and Karlsruhe Institute of Technology

Competences/Skills Needed:

- Knowledge of Python language
- Some familiarity with data management systems
- Experience in GUI development

References:

- <http://datamanager.kit.edu/index.php/kat-data-manager>
- <http://nffa.eu/about/>



## 7. Massive parallel processing for image classification with Spark (Rossella Aversa/Stefano Cozzini)

The goal of the project is to set up a semiautomatic procedure to spawn a Spark cluster on a cloud/HPC infrastructure to process a massive amount of scientific images. The processing mainly includes the image recognition task, and the code has been already written. A prototype implementation in Python has been already deployed on HPC facility using Docker containers.

Tasks include:

- Familiarize with image classification code and Tensorflow library
- Integrate Docker within HPC environment
- Develop a Spark cluster within a Docker compose or a similar tool
- Run and benchmark the massive processing
- Explore the impact in performance of different parallel file systems

Work done in collaboration with NFFA-Europe project at CNR-IOM Institute.

Competences/Skills Needed:

- Good knowledge of Python language
- Some familiarity with Spark and Docker containers
- Knowledge of parallel programming concepts
- Basic knowledge of parallel file systems

References:

- <http://nffa.eu/about/>
- <http://spark.apache.org/>
- <https://www.docker.com/>
- <https://www.tensorflow.org/>

## 8. OpenCL programming on FPGA accelerator (Giuseppe Piero Brandino)

The goal of the project is to port some openCL kernel on FPGA accelerators and measure the performance obtained. We target a minimal Molecular Dynamics code (MiniMD) composed by several openCL kernels.

Tasks include:

- familiarize with MiniMD application
- measure openCL performance on standard accelerator (GPUs)
- familiarize with FPGA platform and associated openCL SDK
- identify openCL computational kernel to be executed on FPGA
- port and analyze performance

Competences/Skills needed:

- Deep knowledge of C/C++
- Some Knowledge of openCL language
- Knowledge of parallel programming concepts

Work done within the EXanest European project.

References: [https://www.altera.com/content/dam/altera-www/global/en\\_US/pdfs/literature/po/ps-opencl.pdf](https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/po/ps-opencl.pdf)

## 9. Deployment of a metadata search engine for scientific images (Rossella Aversa/Stefano Cozzini)

The goal is the deployment within the KIT Data Manager of a metadata search engine, based on ElasticSearch, for scientific images. This will include the implementation of a web service connecting the search engine and the client. The database interrogation should be structured to allow the scientific metadata to be searchable.

Tasks includes:

- Implementation of a customized metadata schema
- Extraction of the metadata from scientific images, preferably into an XML representation, and their publication to a search index
- Implementation of a web service

Work done in collaboration with NFFA-Europe project at CNR-IOM Institute and Karlsruhe Institute of Technology

Competences/Skills needed:

- Knowledge of XML documents
- General knowledge of ElasticSearch (or other NoSQL database)
- Experience in web service implementation

References:

- <http://nffa.eu/about/>
- <http://datamanager.kit.edu/index.php/kit-data-manager>
- <https://www.elastic.co/products/elasticsearch>

## 10. energy efficiency profiling of HPC applications, per-job energy consumption reporting and energy-based scheduling (Moreno Baricevic)

Goal of this project is to further develop and expand the work already done on energy efficiency profiling of HPC applications, introducing energy-based metric reporting and energy-based scheduling on widely used batch/queue systems and schedulers (e.g. torque+maui, slurm, ...).

Task include:

- testing and improving power consumption acquisition programs
- detailed energy efficiency profiling of RegCM and other HPC applications of interest
- integration of acquisition programs into batch system (namely torque/maui and slurm)
- introduction and reporting of new energy-based accounting metrics
- definition of a basic energy-based scheduling algorithm and power-capping policies

Competences/Skills needed:

- C programming skills
- Good knowledge of Linux shell and bash scripting (bash, perl, python)
- Basic knowledge of microprocessor architecture
- Familiarity with HPC systems, queue systems and basic scheduling algorithms

Work done in collaboration with CNR/IOM research institute.

References:

- <http://urania.sissa.it/xmlui/handle/1963/35155>

## 11. Implementation of an OpenStack security policy (Moreno Baricevic, Piero Brandino)

One of the challenges that we are facing comes from the fact that we need to provide a cloud-based internet-exposed interface to internal HPC resources (e.g. remote visualization and post-processing of HPC-originated data, submission of job to the HPC queue system from the VMs, ...). Leaving full control to the cloud users isn't an option, limiting too much the flexibility of OpenStack isn't an option either. The goal of this project is to study and implement a security policy suitable for such scenario.

Tasks include:

- implement a security policy using Role-Based Access List (RBAC) to control the access to OpenStack objects and resources
- investigating Open vSwitch flows for monitoring purposes, statistics collection, and low level control, QOS and filtering
- investigating alternative strategies and methods to improve the security for various use-cases, like remote job submission, remote visualization, distributed access to shared filesystems, single-sign-on

Competences/Skills needed:

- Good knowledge of Linux and shell scripting
- Good knowledge of Networking
- Strong interest and enthusiasm concerning security aspects of ICT

References:

- <https://docs.openstack.org/draft/networking-guide/config-rbac.html>
- <http://openvswitch.org/support/dist-docs-2.5/tutorial/Tutorial.md.html>
- <http://docs.openvswitch.org/en/latest/faq/openflow/>
- <http://en.community.dell.com/techcenter/networking/w/wiki/3820.openvswitch-openflow-lets-get-started>

## 12. OpenStack distributed deployment (Moreno Baricevic, Piero Brandino)

The goal of this project is to explore TripleO OpenStack deployment and OpenStack HA services, as well as deployments methods alternative to standard virtual machines, like bare metal provisioning and containers.

Tasks include:

- Setup and testing of TripleO
- Setup of OpenStack services in High Availability (HA)
- Bare metal provisioning (Ironic)
- Deploying containers

Competences/Skills needed:

- Good knowledge of Linux and shell scripting
- Good knowledge of Networking

References:

- <https://wiki.openstack.org/wiki/TripleO>
- <https://wiki.openstack.org/wiki/Packstack>
- <https://docs.openstack.org/ha-guide/>
- <https://wiki.openstack.org/wiki/Ironic>
- <https://www.openstack.org/containers/>