

# *Open calls Accademiche ICSC/Spoke2*

## **1. Development of data acquisition and trigger strategies in the context high background, rare event experiment, using acceleration on hybrid CPU/FPGA architectures, in the context of the activities of the Spoke2 in ICSC**

- a. Objectives: The applicant is requested to participate in the activities of Spoke 2 / WP2, by proposing projects for the development of trigger algorithms that can be deployed on FPGAs, possibly taking advantage of the presence of fast communication bus with a CPUs, e.g. using Zynq or similar boards hosting both an FPGA and a CPU, in the context of a DAQ + Trigger development. The physics context should be related to High Energy Physics experiments researching ultra-rare events in high background conditions, which is currently under-represented in the institutions that are part of Spoke 2 / WP2 of the ICSC. The applicant will gain access to an FPGA cluster (soon to become available at the ICSC) consisting of Xilinx and Altera boards with full-mesh capabilities. The cluster will allow for both the simulation of the incoming data stream with the required throughput and should serve as a testbed for algorithmic development.

## **2. Development of a multipurpose/multi experiment high level cloud infrastructure for small/medium astroparticle experiments, on the infrastructure of ICSC and for the benefit of Spoke 2 use cases**

- a. Objectives: The applicant is requested to participate in the activities of Spoke 2 / WP3 / Task XXX, by taking part in the realization of a SaaS infrastructure needed for data elaboration, analysis and storage for small/medium astroparticle experiments.

The infrastructure should:

- i. elaborate online generic tools to monitor data quality and experiments features;
- ii. develop access to CPU/GPU resources for data analysis and simulation (queue on demand);
- iii. develop software accessibility and maintenance system for data analysis and simulation;
- iv. develop data and metadata systems from DAQ to tape (storage/backup).
- v. elaborate an orchestrator able to provide resources on demand, minimizing resource requirements and optimizing the thor exploitation.

In particular it is expected to realize generic containers on the cloud infrastructure able to provide all the above services, easily configurable and exploitable by the experiments. Moreover the developed services must be able to exploit the common pledges and storage by means of elaborate/integrate authentication and accounting systems able to guarantee the secure and correct explanation of resources.

The software must adhere to high quality standards and include:

- vi. full documentation;
- vii. tests, benchmarks and performance expected;
- viii. comprehensive examples.

**3. Development of a GPU library for massively parallelized simulations of QCD and QCD+QED on the lattice in the context of the activities of the Spoke2 in ICSC**

- a. Objectives: The applicant is requested to participate in the activities of Spoke 2 / WP1 / Task XXX, by realizing a library to perform massively parallelized lattice simulations of QCD and QCD-like theories on GPU architecture.

The package should:

- i. explicitly target Wilson-like and staggered regularizations;
- ii. implement the correlation functions needed to study QCD and QCD+QED theory;
- iii. support, but not be tied to, the GPU architecture in use at the BolognaTecnopolo/Leonardo (NVIDIA).

In particular, the support for the NVIDIA GPU architecture must be implemented in a way to allow the execution on pure CPU systems for code validation. A strategy allowing future extensions to different vendors/architectures must be adopted.

The package must be able to scale up to O(200) GPUs. High performances are required, and a list of optimizations implemented must be presented, allowing proper measurement of the improvement achieved on a set of tasks central to the calculation.

- b.  
c.

**4. Development and performance optimization of typical simulation code in HEP and Astro, for the use cases of Spoke 2**

- a. Objectives: Simulating the production of rare "beyond standard model" physics events in experiments with particle beams impinging on a target often requires dealing simultaneously with several aspects occurring at distinct spatial and energy scales. On the one hand, on a "macroscopic" scale, it is necessary to simulate the propagation of the primary particle in the target material, and on the other, it is necessary to consider the microscopic mechanisms associated with the production of the events of interest, induced by the propagating particles in the target itself. Existing simulation codes are typically optimized to address only one of these aspects. The purpose of this activity is to study and implement an effective strategy to "connect" the different existing computational systems for the "microscopic" part with those suitable for the "macroscopic" part, in order to create a single solution that allows simulation of rare events in a beam-dump experiment.

- b. Objectives: The CORSIKA software is used in the field of very high energy physics to simulate the so-called extended air showers originating from the interaction of cosmic rays in the atmosphere. The version currently in use is version 7.7420 (May 20, 2022), written mainly in the FORTRAN 77 language with some parts in C, which runs on a single core.

The development of a new version (CORSIKA 8) in C++ is currently underway by the Air Showers group of the Karlsruhe Institute of Technology, made public on their gitlab instance: it is an excellent opportunity to obtain concrete results,

given the difficulty of the task, unthinkable to deal with just one person even if 100% dedicated.

A version 7 profiling is already available, which identifies the most critical functions, to be addressed first from an optimization point of view.

An incremental approach is expected, such as to use the tools made available by OpenACC to take advantage of the available hardware accelerators: in fact, with these tools, it is possible to run the same code in parallel way on both CPU and GPU, and then rewrite the code where the maximum optimization is achievable.

A further activity of interest is to make the software as modular as possible, and consequently make it usable as a service in a client/server logic: the final idea is to use CORSIKA as a module of an HPC system managed via queues, and trigger executions through https requests that specify only the configurations necessary for the run.

c. Objective: performance optimization of specific Geant4 modules

**5. Advanced algorithms for GW experiments (Virgo and ET) in the context of the activities of the Spoke2 in ICSC**

a. Objectives: The applicant is requested to participate in the activities of Spoke 2 / WP3 / Task 1, by enhancing and developing code to search for gravitational waves transients and GW interferometric detectors characterization. In particular, it will be requested to concentrate on the porting of Multi Band Template Analysis (MBTA) code. The applicant will evaluate high performance computing GPUs architecture and will run the code on the infrastructures of the ICSC.

The code will need to operate under the following conditions and show the following characteristics:

- i. Capability to execute MBTA code on the ICSC infrastructure.
- ii. Performance improvement with respect the existing code implementation.
- iii. Possibility of new implementations through the use of Machine Learning.

**6. Solving the quantum Boltzmann equation following neutrino flavour conversion in a dense environment in the context of the activities of the Spoke2 in ICSC**

a. Objectives: The applicant is requested to participate in the activities of Spoke 2 / WP2 / Task 2, by realizing an algorithm to solve the quantum Boltzmann equation following neutrino flavor conversion in a dense environment. The code will run on the GPUs provided by the Infrastructures of the ICSC.

The algorithm will need to operate under the following conditions and show the following characteristics:

- i. Capability to execute on the ICSC infrastructure.
- ii. Implementation of multi momentum and multi flavour computation of the neutrino distribution function;
- iii. 1% precision in the computation of the neutrino density.

**7. Development of algorithms for boosted topologies at LHC/FCC, in the context of the activities of the Spoke2 in ICSC**

a. Objectives: the applicant is requested to participate in the activities of Spoke 2 / WP2, by proposing projects for the development of Machine Learning techniques focused on boosted topologies as a powerful tool for signal characterization and possible probe for new physics.

In processes where sensitivity to new physics is enhanced, such as Vector Boson Scattering, boosted topologies can be heavily investigated using the LHC Run3 data, and optimized for HL-LHC. Furthermore, in the perspective of a future collider, boosted topologies can be fine tuned and the jet substructures analysis fully exploited, in particular in the scenario of FCC-ee, where the background from pile up is naturally removed.

In this context, the proposed work is expected to provide a benchmark for new analysis model studies that have a natural allocation within the Spoke 2 / WP2.

#### **8. Accelerated analysis of Astrophysical data in the Photon Counting Big Data era, in the context of the activities of the Spoke2 in ICSC**

- a. Objectives: We aim at the creation of an innovative approach for the reduction and scientific analysis of the data produced by the INAF ASTRI Mini-Array in "photon counting intensity interferometry" mode and the optimization of cross-correlation algorithms within this framework. The first step in this approach is the optimization of the pipelines for the treatment of the time series acquired at extremely high count rates with the entire array (in the regime of BigData). The second activity consists in the development of efficient and innovative algorithms for the cross-correlation of the arrival times in large time series, exploiting what up to now implemented by our team in dedicated pilot projects. This activity, in turn, includes two distinct threads: on the one hand the parallelization of some components of the algorithm and on the other the computational acceleration of some segments on specific platforms (CUDA) and architectures (FPGA). Within this activity we will proceed with the definition of the modeling algorithms for astrophysical instrumentation to be codified in CUDA and to be made scalable in the context of large ground-based experiments. The third planned activity consists in the development of a preliminary version of a dedicated pipeline for the synthesis of images starting from the interferometric data and the data related to the instrumental simulations.

## *Open calls Industriali*

### **1. Data Management tests using Spoke 2 solutions in industrial environments**

- a. Objectives: The Spoke will develop, deploy and operate data management services for the integration of datasets and data sources with the CN Datalake. We are searching for 1-3 (depending on the evaluation) proposals for testing them in non-academic environments.

The proposals should:

- i. contain a real/realistic use case of interest of the proponents, with details about the improvement expected from a datalake solution
- ii. a description of the data to be used (with their possible IP concerns)
- iii. contain patterns for data access / data movement as desired in the test
- iv. contain a description of the protocols needed / possible

- v. contain a list of needed services / resources (storage and computing, with their locations) which will need to undergo the scrutiny of the CN resource allocation mechanisms
- vi. the length of the proposed test

The Spoke2 team will offer support and technical explanations; still, the operations and the realization of the use case will be realised on manpower from the proponent.

**2. Support for R&D activities on the ARM architecture (support + access to remote resources not available in the CN/Spoke0 + help in the sw porting)**

- a. Objectives: The applicant should provide support for the experimentation of the ARM computing architecture for the use cases of Spoke2 in the CN.

Specifically, it is required to:

- i. provide, deploy and support during operations an ARM cluster of at least 1024 physical cores, with architecture ARMv8 or ARMv9. The cluster will need to be installed at the BolognaTecnopolo in order to profit from the high bandwidth and the local storage.
- ii. The support offered for system administration must be in working hours (9x5) and should be realized with local personnel.
- iii. The nodes of the cluster must be configured with:
  - 1. at least 4 GB memory per physical core
  - 2. at least 40 GB of fast local disk (NVMe, SSD) per physical core
  - 3. at least 40 Mbit/s networking per physical core (with the specifications needed to connect to the CNAF Tier1 networking core)
  - 4. a recent linux 64 bit distribution at least at the level of Ubuntu 22.04 or RHEL 8, or Almalinux 8 (similar configurations to be proposed and agreed)
- iv. On top of the support for the hardware, the applicant must support the software environment as needed by the test use cases (typically using WLCG software stack or Python ecosystems), via a ticketing / helpdesk system.

The cluster will be used to validate use cases as needed by Spoke 2. We require a 9x5 support (working hours), and an average uptime of > 80% along the 24 months.

**3. Porting and optimization (on GPU, on FPGA, on CPU) of algorithms of interest of AstroParticle experiments**

- a. The spoke 2 will develop a set of optimized algorithms and analysis pipelines applied on dataset of astroparticle experiments. We are searching a number of proposals aiming at:
  - i. Port selected algorithms to GPU, FPGA, and CPU architectures, ensuring compatibility and efficiency.
  - ii. Optimize these algorithms for each target platform, taking full advantage of the hardware capabilities and parallel processing.
  - iii. Establish clear performance metrics to evaluate the effectiveness of porting and optimization efforts.

- iv. Measure factors like execution time, throughput, and power consumption for each algorithm on different platforms.
- v. a. Ensure seamless integration of the ported and optimized algorithms with existing AstroParticle experiment software and hardware systems.
- vi. b. Address compatibility issues and provide support for a smooth transition.
- vii. Create comprehensive documentation for the ported and optimized algorithms, including installation, configuration, and usage guides.
- viii. Facilitate knowledge transfer among team members and industry collaborators through workshops, seminars, and training sessions.

**4. Realization of tools and pilots for the Space Economy domain, using the infrastructure of the ICSC**

- a. Objectives: we accept and evaluate applications from SME in the Space Economy sector, willing to produce a tool / a portal / a pilot application for Space Economy. The solution will need to be integrated with the ICSC Datalake; on the other hand, IT resources from the ICSC will be available for its development / Proof of Concept phases.

Solutions can include (but not limited to):

- i. Algorithms to analyze and predict specific features from images (satellites, aerial, ...). The images should be in principle be procured by the applicant; in some cases the ICSC should be able to provide specific datasets
- ii. Services for the SE domain, like information brokers, data formatting and data fusion, catalogs of applications, market applications ...
- iii. Infrastructure tools to execute specific applications, for example tools to execute demanding applications on the ICSC infrastructure
- iv. Any other solution which can increase the usability, the accessibility and the capacity to extract information from Space Economy datasets.

**5. Realization of a PoC market exchange for Space Economy, on the infrastructure of the ICSC**

- a. Objectives: The context for a market exchange on space data refers to the buying and selling of geospatial data, satellite data, and data from telescopes that are collected from space. In this market exchange, individuals or organizations can upload their data onto a platform where it can be accessed by potential buyers. The platform may provide tools for basic data elaborations and validations to ensure the data's quality and accuracy. To ensure the data's property and provenance, a private blockchain can be utilized. The private blockchain will be set up to record ownership and transfer of data, ensuring that the data's intellectual property rights are protected and that buyers can verify the data's authenticity and source.

The pipelines to validate data quality could be referenced in the blockchain as well, and generate a secondary market of software.

The minimum expected set of services will be:

- i. a software suite to be installed on premise, composed by:

- ii. a backend to handle http requests for CRUD operations on a private blockchain;
- iii. a frontend as a web GUI to issue such actions;
- iv. authentication system on the web interface via oauth 2.0 standard protocol, as well as via private key/certificate (depending upon the blockchain of choice)
- v. a software suite to manage the marketplace outside the blockchain, with payments in FIAT currencies.

**6. Heterogeneous industrial use cases on the Spoke 2 platforms (GPU + FPGA + ARM)**

- a. Objectives: The Spoke 2 will study, develop and document software solutions for executing code on heterogeneous platforms, including GPUs and FPGAs. We are searching for 1-3 (depending on the evaluation) proposals for testing these solutions in non-academic environments.

- i. The proposals should:
- ii. contain a real/realistic use case of interest of the proponents, with details about the improvement expected from heterogeneous solutions
- iii. contain a description of the code/services to be used (with their possible IP concerns)
- iv. contain a list of needed services / resources (storage and computing, with their locations) which will need to undergo the scrutiny of the CN resource allocation mechanisms
- v. the length of the proposed test

The Spoke2 team will offer support and technical explanations; still, the operations and the realization of the use case will be realised on manpower from the proponent.

**7. Support for profiling, code engineering and code quality on Spoke 2 code repositories**

- a. Objective: the collaborators of Spoke 2 will create a number of sw repositories, under some central management systems like GitHub, GitLab or similar. The repositories will contain codes of various types and languages, ranging from C++, Python, CUDA and similar. We search a collaboration for code engineering, code quality and profiling support, in which the applicant will propose regular activities on the Spoke 2 repositories, allowing to spot weak programming techniques, improve code quality by finding typical coding problems, and run regular profiling on a subset of the codes identified by the Spoke 2 Leaders.

**8. Support of Spoke 2 tools and solutions towards industrial cascade call recipients**

- a. Spoke 2 will fund, via the Cascade Calls mechanisms, industrial tests and proofs of concept where the Spoke technologies will be tested on use cases from the productive system. The applicant will serve as a middleman between the Spoke affiliates (for example, the WP leaders) and the industrial systems, providing first line support for the latter. This includes the operations of a ticketing portal, and the creation of knowledge bases and FAQs for the most typically encountered problems.

