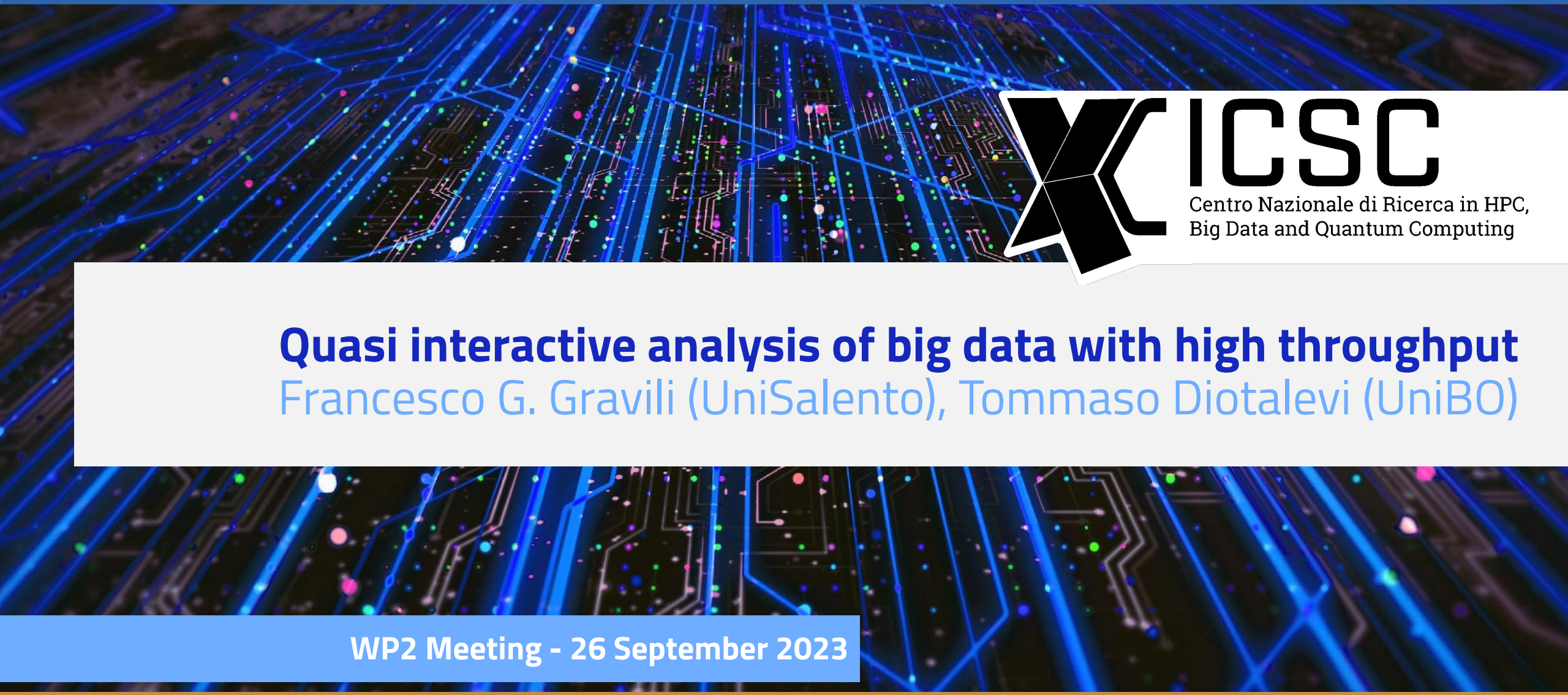
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing



ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Quasi interactive analysis of big data with high throughput

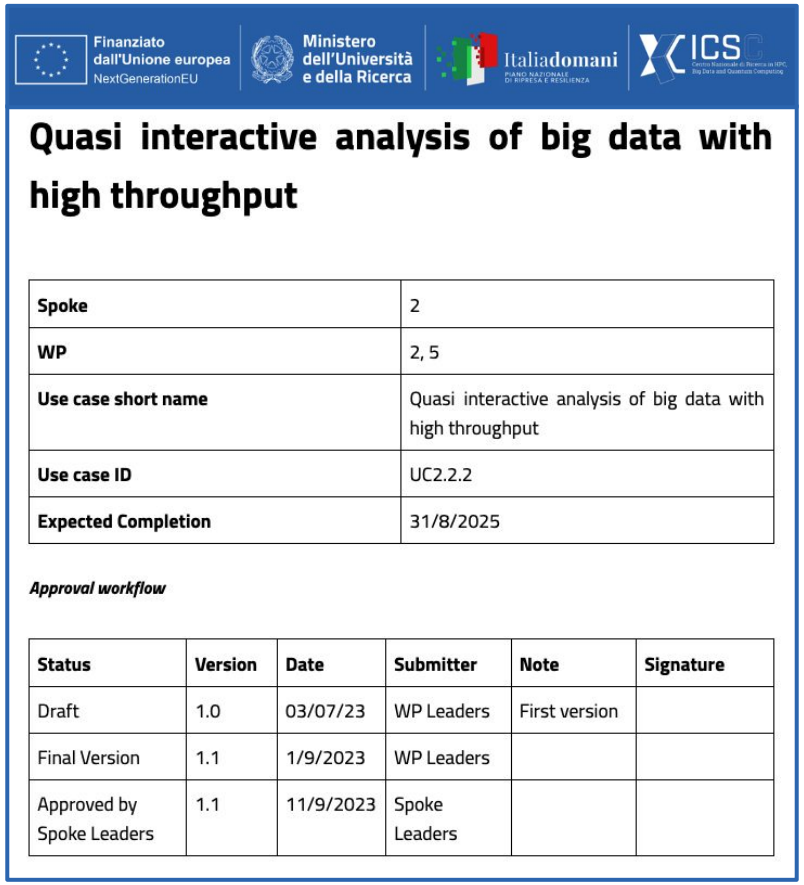
Francesco G. Gravili (UniSalento), Tommaso Diotalevi (UniBO)

WP2 Meeting - 26 September 2023

“Definition and workflow” document

Introduction

The "**Definition and workflow**" document, for the "*Quasi Interactive analysis of big data with high throughput*" use case, is now complete and approved.



Quasi interactive analysis of big data with high throughput

Spoke	2
WP	2, 5
Use case short name	Quasi interactive analysis of big data with high throughput
Use case ID	UC2.2.2
Expected Completion	31/8/2025

Approval workflow

Status	Version	Date	Submitter	Note	Signature
Draft	1.0	03/07/23	WP Leaders	First version	
Final Version	1.1	1/9/2023	WP Leaders		
Approved by Spoke Leaders	1.1	11/9/2023	Spoke Leaders		

Link to document

Use case definition

The need to interject, process and analyze large datasets in an as-short-as-possible amount of time is typical of BigData use cases. Data can be of the most generic type (medical data, data from industrial processes, data from scientific instruments), and processed via different mechanisms. One of these is processing on distributed computing systems, like those which will be deployed by the ICSC DataLake.

The data analysis in High Energy Physics at CERN in particular will require, ahead of the next phase of high-luminosity at *LHC* [1], access to big amounts of data (order of 100PB/year). However, thanks to continuous developments on resource handling (i.e. *WLCG* [2] and Cloud Computing), it is possible to offer users a more flexible and dynamic data access. Furthermore, thanks to software improvements, libraries and data analysis frameworks, together with heterogeneous and extensive computing infrastructures, a new paradigm of data analysis is beginning: from a batch-based approach to an interactive *Analysis Facilities* (AF), based on a parallel and geographically distributed back-end.

Currently, among the several LHC experiments collaborations, discussions about possible strategies to build such facilities are ongoing and span across different solutions carried on by various institutions (e.g. UNL [3], CERN [4], DESY [5], Fermilab [6], ...).

On a national-wide level, the INFN is also pursuing a similar effort in the development of AFs: the current state of the art is a testbed infrastructure developed by members of the CMS collaboration [7]. This prototype offers a user-friendly environment, adopting open-source industry standards such as *Dask* [8], *Jupyter Notebooks* [9] and *HTCondor* [10]. This environment enables streamlined and rapid data processing, while integrating dispersed and heterogeneous computing resources. Transparent resource integration is facilitated by HTCondor, which also ensures fair-share-based resource allocation. The technical architecture provides a solid foundation for testing and benchmarking different approaches, validating new frameworks (e.g. *ROOT RDataFrame* [11] with multi-threading and other low-level optimisations, *Coffea* [12]), and exploring innovative solutions for resource provisioning. Users will exploit all the resources available on their machines in a completely transparent way. In the initial stages, the testbed is using a fraction of the existing resources of the CMS WLCG Tier-2 computing center of Legnaro (Italy) - details on the "*Resources already secured*" section of this document.

Thanks to the activities of this UC, a complementary approach was identified by some members of the ATLAS Collaboration and a local testbed infrastructure has been developed in Naples (Italy). In this approach, the local deployment is based on the *Open-Stack IaaS* [13] and it is using the *Rancher Kubernetes Engine* (RKE) [14] with a *Docker* [15] backend, to simplify the setup and update workflows. This strategy allows for a coherent setup to compare benchmark results and performance. Other ATLAS and CMS local AFs are under R&D, e.g. at the Bari ReCaS center, aiming for a global integration in the future ICSC infrastructure. These AFs are the first steps towards the creation of a much larger implementation, ranging from a HPC/HTC infrastructure to a decentralized DataLake model accessible by the entire scientific community, serving multiple experiments and diverse data analysis use cases: they can range from physics activities to deep learning pipelines for any kind of data. On top of these use cases, already established, we expect to raise interest from now-unforeseen activities, including those coming from the productive system as Open Calls to be issued in late 2023.

Participating Institutions

- **PIs:** Tommaso Diotalevi (Università di Bologna), Francesco G. Gravili (Università del Salento)
- **INFN** (including all sections)
- **Università del Salento**:
  - Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum, using LHC Run2 data. Porting of the analysis in the AF framework.
- **Università di Bologna:**
  - Search of a heavy neutral lepton (bump hunt) in the D_S meson decays with Run2 data. Porting of the analysis in the AF framework;
  - CMS Muon detector performance analysis: target quasi-interactive performance studies of phase space corners using large datasets; driven by the need to accurately assess analysis systematics (e.g. high-energy muons).
- **Politecnico di Bari:**
  - Testbed implementation of the AF in the ReCaS computing center
- **Università degli Studi di Bari:**
  - Charm and multi-Charm baryon measurements.
- **Università degli Studi di Napoli:**
  - Top quark + MET search. Full analysis implementation with Run3 data.
  - Analysis Facility deployment on local cloud infrastructure.
- **Università degli Studi di Firenze:**
  - Developing common analysis tools based on analysis facilities for the CMS experiment.
- **Università degli Studi Ferrara:**
  - Adopting tools for quasi interactive analysis for the IDEA experiment at the FCC-ee.
- **Università della Calabria.**

Please, check the correctness of the contact names in the document!

Use case expected activities

- **0-6 Months**: Identifying target analyses (from different experimental collaborations) and porting a first complete analysis on a first operational AF testbed;
- **6-13 Months**: Adding more experimental applications to the existing infrastructure testbed as well as performance comparisons for speed-up, with respect to traditional approaches;
- **13-22 Months:** Harmonization of the different efforts into a single infrastructure, also consolidating the relative documentation, aiming for a multitenant service. Concurrently, organizing a first hands-on workshop on AF usage for users;
- **Month 22 - MILESTONE 8**: Prototypal central infrastructure on available resources;
- **22-36 Months**: Scaling the local testbed facilities to a national-wide DataLake infrastructure. Scientific reports at conferences and/or workshops;
- **Month 35 - MILESTONE 10**: Final central infrastructure on available resources, with long-time stability and downtime troubleshooting.

KPI - Key Performance Indicator

These parameters, which are measurable objects, will reflect the final success of the entire use case.
Eventually, it is possible to nominate, for each KPI, a contact person. This will be evaluated in the next weeks.

KPI ID	Description	Acceptance threshold
KPI2.2.2.1	Implementation of *N* data analyses in the AF	$N \geq 2$
KPI2.2.2.2	Reference documentation of the AF	≥ 1 dedicated web site
KPI2.2.2.3	Hands-on workshops for AF users	≥ 1 workshops
KPI2.2.2.4	Scaling up the testbed AF infrastructure, serving *k* tenants, for a total of *N* data analyses	≥ (200·*N*) cores
KPI2.2.2.5	Talks at conferences/workshops about AF activities	≥ 1 talk

Computing resources

WP5 synergy

Already secured (not ICSC resources):

- CMS AF current testbed: CMS Tier-2 of Legnaro: 3 nodes, each with 32 logic CPUs - 128GB of RAM - 1Gb/s network;
- ATLAS testbed: 2 virtual machines, one controller and one worker, each one with 1-4 logic CPUs and 8GB of RAM. Rocky Linux 8.6 OS.

Projected needs (purely an estimation - discussed also within WP5):

- Requirements for setup and development (per single analysis):
  - 200 logical CPU cores;
  - 800GB of RAM;
- Requirements for scale up (per single analysis):
  - 1000 logical CPU cores;
  - 4TB of RAM;

Storage (assuming that input data are not read by any external existing storage area, i.e. the Worldwide LHC Computing Grid (WLCG) for the LHC experiments):

- ~400TB of disk (considering input data and simulations) with "read-only" permissions, per tenant (i.e. experimental collaboration);
- ~100TB of disk per single analysis, for stage-out and including a fixed scratch area quota per single core, with "read-write" permissions.

ICSC Resources available: n.a. / Waiting for news about RAC, recently discussed also by CN governance in the annual ICSC meeting.

Risk Analysis

Identifier	Description	Mitigation
R1	The CN is unable to provide the needed resources	Set a minimum amount of resources (defined in *"projected needs"*, under *"Requirements for setup and development per single analysis"*), at least to run a single analysis within the AF. This will allow only minimum functionalities, with no upscaling, for a single analysis.
R2	The underlying AF software, as well as the analysis-specific libraries, may be unstable	By design of the AF framework, different technological solutions can be proactively developed simultaneously, allowing backup solutions in case of any software instabilities. This can also be applied to analysis-specific software (e.g. *RDataFrame* [11] *vs Coffea* [12]).
R3	Lack of person power for long-term maintenance of the infrastructure	Define troubleshooting checklists to minimize downtimes and/or outages. Provide a recovery system for continuous operations in the final stages of the project (and beyond).

Current status

Activities ongoing

CMS:

- Vector Boson Scattering ssWW analysis in hadronic tau and light lepton (INFN e Univ. Perugia);
- Heavy Neutral Lepton search on heavy neutrinos in the D_S decays (INFN e Univ. Bologna);
  - Studies for performance comparison speed-up, using in-site metrics and RDF monitoring, in progress
- top quark+MET analysis (INFN e Univ. Napoli). Still in an explorative phase, talk presentato a WP5;
- [WP5 synergy] Testbed implementation of the INFN-CMS Analysis Facility in the ReCaS (Bari) computing center.

} Joint WP2 meeting

Future talk (WP5 synergy)

ATLAS:

- [WP5 synergy] Deployment of an Analysis Facility (INFN e Univ. Napoli);
  - Recent progresses made by the involved group. Started testing first analysis applications.
- Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum (UniSalento e INFN Lecce). Interest towards an interactive analysis, porting strictly connected to the previous bullets for the AFs.

Future talk (WP5 synergy)

\+ Other activities, listed in the *"Participating Institution"*, to be reported. Reach out to us, to report an update or ask for support! (*see next slide*)

Reporting news for upcoming talks

During the next weeks, in the bi-weekly WP2 meetings, the different flagship use cases will report progresses on the ongoing activities.

- We have established an internal mailing list, that will work in a bi-directional way: for us to report news related to our use case, and for the involved institutions and members to share updates, achievements, and developments.

cn1-spoke2-wp2-analysisfacility@lists.infn.it (click here to subscribe or get in touch in case of issues)

We strongly encourage every WP2 member involved in this use case to start providing us details (and progresses) of their work.

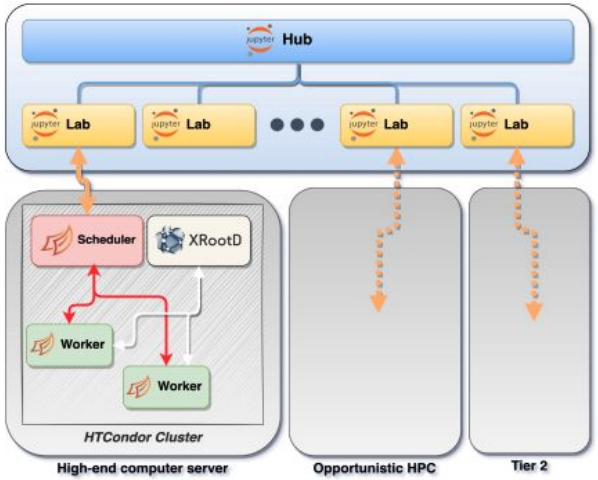
To schedule a future update talk, please consider the planned roadmap of the next Analysis Facility-specific meetings in WP2:

- 07/11/2023
- 05/12/2023
- 30/01/2024
- 27/02/2024
- 09/04/2024

BACKUP

Analysis Facility testbeds

State of the art:



INFN-CMS *Analysis Facility*:

- Access to a single HUB and authentication via token (INDIGO-IAM)
- Based on standard industry technologies
- Customisable python kernel
- Workarea fully containarisable
- Overlay based on HTCondor (also available standalone)
- DASK library (python) for distributed computing
  - Scale the execution from 1 to N cores
- Possible implementation on heterogeneous resources (HTC/HPC/Cloud)
- Data access configurable with WLCG (xrootd, WebDAV, ...)

Similar approach for the INFN-ATLAS Analysis Facility:

- Jupyter Hub/Lab for interactive environment, supporting multiple accesses
- DASK library for scalability
- Kubernetes for container management and connection
  - Ongoing tests on this container technology to achieve a higher scalability