ICSC

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

WP5

Meeting WPLeaders Review Use cases e Deliverables
25.07.2023. Bologna

Elvira Rossi (UniNA)

Daniele Spiga (INFN - PG)

• *Kick off meeting* 25/26 novembre 2022, Bologna

Outline

- WP5 quick recap

- Landscape document
  - status and next steps

- WP5 and Flagship use cases

A quick recap on the WP5 assets

Generally speaking the mission is to help "scientific and industrial" use case to scale out, exploiting the ICSC resources
- should be in strict contact with WP1-2-3 and 6… and try to technically/technologically liaise with Spoke0 [for the service not the HW]

Since the kickoff we proposed a bottom-up approach suggesting two main assets:
- **High Rate Data Analysis**
  - a solution to use possibly a heterogeneous set of resources (seen from the user perspective: to transparently exploit different type of providers…. HPC, Cloud HTC)
  - a solution to access large amount of resources ALSO for (quasi)-interactive processing (to process a huge   amount   of   data   interactively)

- **Data Management in the Data Lake**
  - a solution to abstract the physical layer of the storage (user don't need to care about different backend)
  - track data location
  - automatically replicate data between different locations in a transparent manner (third party copy of data)

How we proposed to work… the Testbeds

At the time of the kick-off we proposed the testbed idea…
- life is not ideal… we are keep pushing for these but we're a bit behind our schedule. This is also due to the lack of ICSC resources

Few tests on high rate analysis have been done using Experiment specific/dedicated resources and setup… but this is not really what we want
- recently we've got some INFN resources for LHC R&D and we decided to start using them for our early exercises

Meanwhile we've produced and shared recipes for the technical deployment of an Analysis Facility and members of WP5 are contributing to those in view of the ICSC resources
- this could allow us to anticipate a bit the program of work

Best Practices Document

We are working on:

"**Best practices for data lake solutions and technologies for high rate analysis activities**"

in short this includes three topics:
- data management
- data access for high rate analysis
- heterogeneous resource exploitation

(Not surprisingly) This is tightly coupled to the assets that we already identified and discussed. Based on the flagship use cases these seems to be two highly sinergic topics

- and thus this is how we plan to support the scientific use cases
- (not only the flagship)

Status of the document

The document is available here

Why we need changes to the "current" infrastructure

How things are evolving. Current trends in HEP and beyond (in EU)

Current technological trends for managing data in a Scientific Data Lake

Current technological trends for managing high rate data analysis

What we will do and link to the flagship that we support

Best practices for data lake solutions and technologies for high rate analysis activities

Introduction
Data Management and Data Access: needs and projections
Data Lakes, Data Access and Processing
Data Management and storage federation: Technologies
High Rate Analysis: Technologies
Summary and relationship with flagship use cases

Still To Do

Best Practices doc: next steps

To get a **feedback and comments**

- is the content ok?
- anything, content wise, still missing or worth to add/remove?
    - i.e.
    - too LHC/WLCG oriented/biased?
    - more quantitative?

To finish the **missing chapters**

- priority on summary and connections with flagships
    - see next slides

Anything else?

- i.e do we need to add something about the resource access methods?
    - a new small paragraph ? to extend what is already there?

WP5 and connections with flagship (explicitly discussed)

WP5 does not own dedicated flagship. It support, technologically wise others.

High Rate Data Analysis. We support two flagship

- **WP2:**
   Quasi interactive analysis of big data with high throughput

   - **PIs:** Tommaso Diotalevi (BO), Francesco G. Gravili (LE)
   - Synergies on testing the framework, together with the AF developers, and improving the technical documentation

   Advanced Machine Learning: Flash simulation and other bleeding edge applications

   - **PI:** Lucio Anderlini (FI)
   - Synergies on data management and cloud-based resource provisioning

- **WP6:**
   - **Add here Titolo**
   - PI: Alessandro Bombini (INFN-PI)

   Reuse the same technology. Added value to demonstrate how the solution is actually generic enough to accommodate very diverse use cases (and data)

Still WP5 and Flagship use cases… to verify/discuss

There are more "on-going discussions"

- **WP2:**

  Use case porting GPU
  - **PI** A.Di Florio.
  - idea: To support the scouting about solutions to access dispersed and specialized resources (GPUs) possibly in a transparent manner (i.e. transparently access Leonardo's GPU)
  - NOTE: not necessarily relates to huge amount of resources

  Flagship "Development of ultra-fast algorithms running on FPGAs"
  - **PI:** Bernardino Spisso (NA), Simone Gennai (MiB)
  - Synergies on exploitation of GPU resources (single node or multi node) ie. for ML trainings

- **WP3:**
  - soluzioni di data management per i dati dal DAQ al Tape nel data lake
  - PI? (G. Mazzitelli ?)
  - this would be fully focused on Data management.. never discussed in details.. this could be a good opportunity..

School on Open Science Cloud (SOSC 2023)

23–27 Oct 2023 - INFN Perugia

The 5th edition of the international School on Open Science Cloud (SOSC 2023) will be held in **Perugia, from 23 to 27 October 2023**. The school is organized by INFN, Department of Physics and Astronomy "Augusto Righi" of the University of Bologna and the Departments of Physics and Geology of the University of Perugia.

The School is multi-disciplinary and targeted at postgraduate researchers including bachelor degree or equivalent in fields such as physics, statistics, computer science, computer vision, biology, medicine, bioinformatics, engineering, working at any research institute, with some experience and interest in data analysis, in computing or in related fields. Applications by university students (undergraduate) will be considered depending on availability and must be accompanied by a letter of reference from a university professor. **We welcome applications from all nationalities, and encourage all qualified persons to apply.**

https://web.infn.it/SOSC23

Backup

WP2 slides:

https://docs.google.com/presentation/d/1_fUH_EXN5eLt7KOABst v6JMN4kRcSQfwICSssY1OLUA/edit#slide=id.g25ab104a940_0_ 11