EuCAIF WG: 4 White paper JENA WP 4"

Andreas Ipp, Sascha Caron

May 2, 2024

Why this working group?

- The **Joint ECFA-NuPECC-APPEC Computing Initiative** had a workshop on 12-14 June 2023 in Bologna: https://agenda.infn.it/event/34738/
- Joint ECFA-NuPECC-APPEC (JENA) Seminar in May 2022 in Madrid (https://indico.cern.ch/event/1040535/) revealed that there is an increased need for discussions on the strategy and implementation of **European federated computing** at future large-scale research facilities. In particular, synergies between the three fields should be identified, as well as computing requirements in the next decade, where all three communities (particle physics, nuclear physics and astroparticle physics) as well as neighbouring research fields such as astrophysics or cosmology can benefit.
- Main conclusion at the Bologna-Workshop was the **creation of five working groups** in order to coordinate a white paper as input for the next JENA Symposium:
  1. HTC, WLCG and HPC (HPC)
  2. Software and Heterogeneous Architectures (Software)
  3. Federate Data Management, Virtual Research Environments and FAIR/Open Data (Data)
  4. **Machine Learning and Artificial Intelligence (AI)**
  5. Training, Dissemination, Education (TDE)

Mandate ML and AI group WP4

Machine Learning and Artificial Intelligence (AI):

These data analysis methodologies have seen a rapid expansion in the last years in most fields of science, including the ENA domains.
A Working Group will be set up to

- ***follow the technologies in this fast evolving field***
- ***and analyse the potential impact on the ENA computing infrastructure needs.***
  The focus will be to quantify the resource needs and to define the interfaces and services that are needed by physicists to run ML workloads (looking at both training and inference)

  → https://indico.scc.kit.edu/event/3813/

Timescale ? Vision ? Mandate

- Next funding rounds , i.e. upcoming 5 years ?
- Timescale: Paper ready for review by end of this year (to ENA chairs by **30 November 2024**)
- Size : 10 pages + 1 page summary intended for funding agencies

How ? Paper structure (about 8 pages)

- **Identify + Describe (new) technologies:** *machine learning/AI algorithms, hardware accelerators, distributed computing frameworks, data management techniques, protocols, interfaces etc.*
  ⇒ *see google docs*
- **Analyse Impact:** *How do 1. influence the computing infrastructure needs for ENA*
- **Consult experts / Survey** *: ??*
- **Describe outcomes + implementation plan / needed resources:**

Participation in creation process

Please enter your ML field and topic into the Google Doc:

https://docs.google.com/document/d/1szaLjc5edw90jGnUS_nHDBEsGYSv5qWs-3EFM-asb40/edit?usp=sharing

Or use the google of the conference page:
https://bit.ly/eucaifcon24-wg4

Please add where you can contribute.

Survey

- Here is the current state of the survey (draft only - please do not distribute this yet) https://docs.google.com/forms/d/e/1FAIpQLSet3CbIDD8zzjnutMDRGpB86IxDAuAc3O0GqCkJ7NnAvMhXKw/viewform?usp=sf_link
- If you have any suggestions for further questions, please feel free to add them to the following document (which should contain exactly the same questions) https://docs.google.com/document/d/1LJVcbawqMAm0xqwsmecSrJdcuYznMQt1YCwd2X2zUBg/edit?usp=sharing
-