Charge for experiments for tracking the evolution of the Analysis Infrastructure

Context:

A focus session on Analysis Facilities was organized during the 154th LHCC week, where the R&D work happening within the community and discussed in the HSF Analysis Facilities Forum was presented. The LHCC recommended that experiments engage in the process of developing and defining the structure of the future Analysis Facilities and requested they produce a document which defines the use cases in order to establish realistic benchmarks. This process should be coordinated with the HL-LHC Computing and SW review panel. The document is expected to be regularly updated in the process towards HL-LHC.

The WLCG community agreed during the WLCG/HSF workshop in May 2024 on a set of questions for experiments to better define the requirements for the Analysis Facilities and use them in a process to help sites better understand what running this infrastructure entails. These were presented during the 158th LHCC week and form the basis of this charge.

Charge:

Experiments are requested to answer the following questions in a focus session at the upcoming LHCC in March 2025:

1. Description of the current Run-3 analysis model
    a. Main analysis workflows and data reduction steps, including how closely chained they need to be.
    b. Data formats used for analysis, including their size and level of adoption (current and Run-3 final goal).
    c. How much compute, storage and network resources are used for Run-3 analysis. Which fraction are pledged and which fraction are used in interactive mode (as opposed to batch).
    d. Comment on what is working well and what is not, both from the point of view of users as well as providers (experiment S&C teams and sites).

2. Future analysis model in Run-4 and Run-5
    a. Comment on which aspects of the current Run-3 analysis model will not scale for Run-4.
    b. Describe the relevant changes in the model and their impact in resources: policies for number of versions and replicas, fraction of data which is managed vs. unmanaged (e.g. caches), remote vs. local data access, batch vs. interactive cpu/gpu access, need of access to external DBs, or any other.
    c. Annual volume expected for the different data formats, both data and MC.

3. Managing the evolution of the Analysis Infrastructure
    a. Describe the user requirements for analysis in HL-LHC and the processes that will be used to track their evolution in the next few years.
    b. Comment on which new technologies or emerging paradigms you expect to be needed or have a relevant impact on the future Analysis Infrastructure and which mechanisms can be set up to manage this evolution as new technology will appear (e.g. ML, GPUs/FPGAs, etc).
    c. Describe the plans to develop specific use cases that can be used to benchmark different building blocks of the Analysis Infrastructure so that a comparison can be made between different implementations.
    d. Comment if you think that support for analysis workflows in Run-4 will need specialized infrastructure different from the Grid. If so, please describe what features that Analysis Infrastructure will need to provide to expand the one in the Grid.
    e. Describe the current status and the R&D work that is underway towards implementing relevant Analysis Infrastructure functionality.