CompF4
Findings and Recommendations

Wahid Bhimji (LBL), Meifeng Lin (BNL), Frank Wuerthwein (UCSD)

Summary

- CompF4 focuses on the **R&D needs** to facilitate future processing and storage resource access for HEP.
    - Capacity requirements are addressed by other relevant CompF topical groups.
- Our report highlights the challenges and research directions driven by the **complexity and diversification** of computational infrastructure, as well as the expected large influx of data in the next-gen experiments and simulations:
    - Diverse heterogeneous architectures, e.g., GPUs, FPGAs, etc.
    - Novel and special-purpose computing hardware, e.g., AI Hardware.
    - Diverse facility environments: HPC, cloud, Grid, etc.
    - Large data volume and data rate will require enhancing the storage and networking infrastructure.
    - Analysis facilities and edge services will also need to adapt to make efficient use of the increasingly heterogeneous and diverse computing resources.
- Significant R&D is required to make efficient use of the diverse resources expected to be available at grid, cloud and HPC facilities.

Recommendation I: Efficiently exploit specialized compute architectures and systems.

To achieve this will require

- the allocation of dedicated facilities to specific processing steps in the HEP workflows, in particular for “analysis facilities”;
- designing effective benchmarks to exploit AI hardware;
- improved network visibility and interaction;
- and enhancements to I/O libraries such as lossy compression and custom delivery of data

Recommendation II: Invest in portable and reproducible software and computing solutions to allow exploitation of diverse facilities

- The need for **portable** software libraries, abstractions and programming models is recognized across all the topics discussed here, and is especially called out in Processing, AI Hardware, and Storage.
- Software frameworks to enable **reproducible** HEP workflows are also greatly needed to ensure a variety of resources can be utilized, recognizing HEP's high standards for precision and reproducibility.

Recommendation III: Embrace disaggregation of systems and facilities

- The HEP community will need to embrace heterogeneous resources on different nodes, systems and facilities and effectively balance these accelerated resources to match workflows.
- To do so will require software abstraction to integrate accelerators, such as those for AI ; orchestration of network resources; exploiting computational storage; as well as exploiting system rack-level disaggregation technology if adopted at computing centers

Recommendation IV: Extend common interfaces to diverse facilities

In order to scalably exploit resources wherever they are available, HEP must continue to

- encourage edge-service platforms on dedicated facilities as well as Cloud and HPC
- develop portable edge-services that are re-usable by other HEP projects and exploit commonality within HEP and other sciences.

These interfaces will also need to extend into all aspects of HEP workflows including data management and optimizing data movement; as well as the deployment of compute resources for analysis facilities.

Questions and Comments?

Draft CompF4 report can be found in the Topical Report section on: https://snowmass21.org/computational/storage

We are seeking broad community feedback during this summer study and afterwards.

Final report is due by the **end of September, 2022**.

We would greatly appreciate your feedback!

Contact us: wbhimji@lbl.gov, mlin@bnl.gov, fkw@ucsd.edu