

Reference ID: 11226437940_Gutsche

Reference ID: 11226437940_Gutsche

Submission Date and Time: 12/16/2019 12:57:56 PM

This contribution was submitted to the National Science Foundation in response to a Request for Information, <https://www.nsf.gov/pubs/2020/nsf20015/nsf20015.jsp>. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Consent Statement: "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publicly available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

Consent answer: I consent to NSF's use and display of the submitted information.

Author Names & Affiliations

Submitting author: Oliver Gutsche - Fermilab

Additional authors: For the U.S. CMS S&C Operations program with special acknowledgement to; James Letts, UCSD; Frank Wuerthwein, UCSD; Brian Bockelman, Morgridge Institute; Mike Hildreth, Notre Dame University; Ken Bloom, University of Nebraska–Lincoln

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

High Energy Physics;LHC;HL-LHC

Title of Response

Analysis facilities for HL-LHC: dynamically cached storage hierarchies and new analysis paradigms

Abstract

A transformative change of the analysis facilities is needed to support HL-LHC analysis because of the significant increase to hundreds of billions of collision events to analyze in HL-LHC. Analysis input data of

different sizes and event content need to be cached on a hierarchical storage architecture combining high latency and low latency storage technologies. Quasi-interactive access to data is needed for the smallest data sizes, while transparent access to larger data sizes still needs to be possible. The interaction with the data needs to be facilitated using new analysis paradigms based on industry technologies, accessing data in the most efficient way with the smallest possible latency, providing batch-like quasi-interactive access to events hiding the technical complexity from the user community.

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

The HL-LHC is expected to record an unprecedented number of proton-proton collisions during its planned run from 2027-2037. Compared to LHC (starting its 3rd running period in 2021 and ending in 2024), the number of events recorded by the CMS detector will increase by almost an order of magnitude, from 7 billion collisions to around 50 billion collisions per data taking year, with similar increases for simulated collisions needed for physics analysis from 15 billion to around 100 billion events. The LHC uses a distributed infrastructure of computing resources interconnected through the R&E networks in Europe, Asia and the U.S. Currently, the CMS infrastructure is comprised of over 250k x86-compatible compute cores, 150 PB of disk, and 300 PB of tape to support central activities to record and simulate proton-proton collisions and perform their subsequent reconstruction to extract physics quantities from the measurements, as well as subsequent analysis activities driven by the physicists of the 2,500 large CMS collaboration. LHC analysis is based on individuals or groups of physicists processing centrally provided reconstructed events with batch techniques to produce individualized ntuple style data views that are further reduced to allow for quasi-interactive analysis of the data. To that effect, CMS uses a combination of placement and caching of the centrally produced events to optimize access for the collaboration. Because of the significant increase of the number of collision events to analyze in HL-LHC, a transformative change of the analysis facilities is needed to support HL-LHC analysis.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

The HL-LHC is expected to record an unprecedented number of proton-proton collisions during its run planned from 2027-2037. Compared to the current LHC (starting its 3rd running period in 2021 and ending in 2024), the number of collisions recorded by the CMS detector will increase by almost an order of magnitude from 7 billion to around 50 billion per data taking year, with similar increases for simulated collisions from 15 billion to around 100 billion. We distinguish “central production” from “community data analysis” in that a small production team produces and curates data for the global CMS community of thousands of scientists. This response to the RFI covers only the community data analysis aspects using the centrally produced data as input. Many different analysis strategies are pursued at the LHC to sift through the trillions of proton-proton collisions searching for hints of a multitude of theories of new physics. The complexities of these analyses, whether performed by an individual or a coordinated group, will only increase in the HL-LHC era. Analysis of recorded and simulated collisions starts after central processing has provided centrally reconstructed data samples. CMS shares common event formats for analyses at different levels of complexity among all researchers. The formats range from kB to MB in size per collision. CMS expects to use up to three complexity levels of analysis event formats. Because of the increase in the number of events in HL-LHC, CMS expects to use a hierarchical storage concept where only the small formats reside on direct low latency access media (e.g. HDD or SSD) while the large format resides on cheap high latency media (e.g. tape) To support quasi-interactive analysis for a community of researchers, a transformational change in how analysis facilities for HL-LHC are provided is needed. As placement of complete working sets of centrally produced samples across all three analysis formats will be impossible, new facilities will have to be used that dynamically cache data across the storage hierarchy. Providing access to the smallest formats with the shortest latencies while allowing access to larger formats with higher latencies but still sufficient to allow for reasonable analysis turn-around of days to weeks. This is combined with end-analysis facilities based on industry technologies that use the dynamically cached storage hierarchy to access data in the most efficient way with the smallest possible latency, providing batch-like quasi-interactive access to events hiding the technical complexity from the user community. The facility concept needs to support flexible deployment on centralized resources at national labs as well as easy deployment at campus resources of universities that host physics groups for CMS, using a combination of local support as well as centralized deployment orchestration. The latter may possibly be shared with other infrastructures that the universities deploy for other disciplines. We expect that overlap between disciplines will be high if open-source and industry-standard solution components can be used.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

The evolution of the community analysis facilities will require education of both the analysis community and the site administration teams at the sites in the utilization of new technologies. The sites and central support teams will have to be educated in the centralized deployment of services and software following industry standard tools. These new skills are to the benefit of the whole infrastructure, national labs and campus organizations alike, as these techniques and tools can be applied to most of

the services run, not only the services needed for the analysis facilities. The analysis community will have to be educated in new analysis paradigms that are closer to industry practices than to the community-own solutions that are mostly specific to the HEP community. These new skills will increase the ability of the scientific community to compete on the global job market.

-- End Submission --