Contribution ID: **345** Type: **Oral**

Physics Data Production on HPC: Experience to be efficiently running at scale

Thursday, November 7, 2019 11:45 AM (15 minutes)

The Solenoidal Tracker at RHIC (STAR) is a multi-national supported experiment located at Brookhaven National Lab and is currently the only remaining running experiment at RHIC. The raw physics data captured from the detector is on the order of tens of PBytes per data acquisition campaign, which makes STAR fit well within the definition of a big data science experiment. The production of the data has typically run using an High Throughput Computing (HTC) approach either done on a local farm or via Grid computing resources. Especially, all embedding simulations (complex workflow mixing real and simulated events) have been run on standard Linux resources at NERSC/PDSF. However, as per April 2019 PDSF has been retired and High Performance Computing (HPC) resources such as the Cray XC-40 Supercomputer known as "Cori" have become available for STAR's data production as well as embedding. STAR has been the very first experiment to show feasibility of running a sustainable data production campaign on this computing resource. In this contribution, we hope to share with the community the best practices for using such resource efficiently.

The use of Docker containers with Shifter is the standard approach to run on HPC at NERSC – this approach encapsulates the environment in which a standard STAR workflow runs. From the deployment of a tailored Scientific Linux environment (with the set of libraries and special configurations required for STAR to run) to the deployment of third-party software and the STAR specific software stack, we've learned it has become impractical to rely on a set of containers containing each specific software release. To this extent, a solution based on CVMFS for the deployment of software and services has been deployed but it doesn't stop there. One needs to make careful scalability considerations when using a resource like Cori, such as avoiding metadata lookups, scalability of distributed filesystems, and real limitations of containerized environments on HPC. Additionally, CVMFS clients are not compatible on Cori nodes and one needs to rely on an indirect NFS mount scheme using special DVS servers designed to forward data to worker nodes. In our contribution, we will discuss our strategies from the past and our current solution based on CVMFS. The second focus of our presentation will be to discuss strategies to find the most efficient use of database Shifter containers serving our data production (a near "database as a service" approach) and the best methods to test and scale your workflow efficiently.

Consider for promotion

No

Primary authors: LAURET, Jerome (Brookhaven National Laboratory); POAT, Michael (Brookhaven National Laboratory); PORTER, Jeff (Lawrence Berkeley National Lab. (US)); BALEWSKI, Jan (Lawrence Berkeley National Lab. (US))

Presenter: POAT, Michael (Brookhaven National Laboratory)

Session Classification: Track 9 – Exascale Science

Track Classification: Track 9 – Exascale Science