Contribution ID: **469** Type: **Oral**

STAR Data Production Workflow on HPC: Lessons Learned & Best Practices

Thursday 14 March 2019 19:00 (20 minutes)

The Solenoidal Tracker at RHIC (STAR) is a multi-national supported experiment located at Brookhaven National Lab. 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 on standard nodes or on standard Grid computing environments. All embedding simulations (complex workflow mixing real and simulated events) have been run a standard Linux resources at NERSC aka PDSF. However, HPC resources such as Cori have become available for STAR's data production as well as embedding, and STAR has been the very first experiment to show feasibility of running a sustainable data production campaign on this computing resource.

The use of Docker containers with Shifter is required to run on HPC @ NERSC –this approach encapsulates the environment in which a standard STAR workflow runs. From the deployment of a tailored Scientific Linux environment (requiring many of its own libraries and special configurations required to run) to the deployment of third-party software and the STAR specific software stack, it has become impractical to rely on a set of containers containing each specific software release. To this extent, solutions based on CVMFS for the deployment of software and services have been employed in HENP, but one needs to make careful scalability considerations when using a resource like Cori, such as not allowing all software to be deployed in containers or bare node. Additionally, CVMFS clients are not compatible on Cori nodes and one needs to rely on an indirect NFS mount scheme. In our contribution, we will discuss our strategies from the past and our current solution based on CVMFS. Furthermore, running on HPC is not a simple task as each aspect of the workflow must be enabled to scale, run efficiently, and the workflow needs to fit within the boundaries of the provided queue system (SLURM in this case). Lastly, we will also discuss what we have learned to be the best method for grouping jobs to maximize a single 48 core HPC node within a specific time frame and maximize our workflow efficiency.
We hope both aspects will serve the community well as well as those following the same path.

Authors: POAT, Michael (Brookhaven National Laboratory); LAURET, Jerome (Brookhaven National Laboratory)

Co-authors: PORTER, Jefferson; BALEWSKI, Jan

Presenters: POAT, Michael (Brookhaven National Laboratory); PORTER, Jefferson; BALEWSKI, Jan

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research