STAR Data Production at NERSC/Cori, an adaptable Docker container approach for HPC

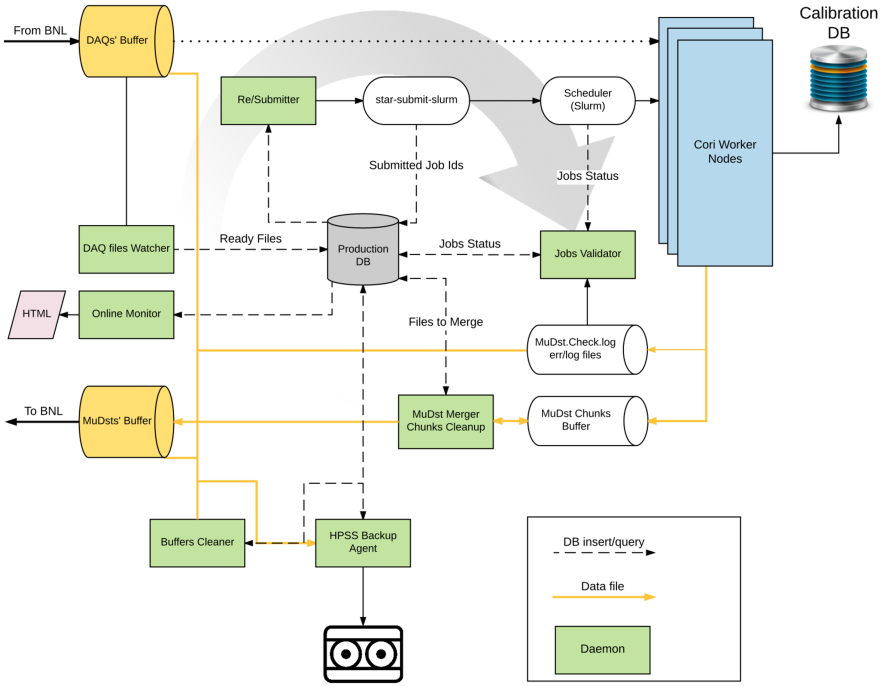
Goal: to carry large scale HEP/NP data production on HPC systems utilizing Docker/Shifter containers for customized software stack.

Project Status:

- STAR software Docker image built
- Fault tolerant (finite state) pipeline built and tested on +100k CPU hours
  - ~99% walltime efficiency demonstrated
  - +96% overall pipeline success rate achieved
- End-to-end optimization:
  - Optimized NY-CA data transfer over Esnet
  - Scalable DB accessibility

Plan:

Requested 25M CPU hours to process 1.5PB of Au+Au collisions data.



Highlight: Linux containers enable use of cloud-type and HPC systems to provide vast elastic resources suitable for HEP/NP real data and simulation productions. This has the potential to address current and future challenging data processing needs of experiments.

STAR experiment demonstrated equally efficient data production workflows on Cori/HPC, comparable to standard Linux clusters.

BERKELEY LAB
Lawrence Berkeley National Laboratory

Poster-420: M. Mustafa, J. Porter, J. Balewski, J. Lauret