

Joint discussion with the Production WG

September 23, 2024

Extending production status monitoring

Timeliness of benchmark delivery could be improved: the progress of the productions is reported on ~firehose, and this is useful. It would be nice to figure out the next step to report which datasets are complete enough and ready for benchmarking.

- We can monitor firehose or look at disk directly, but we also need to know how many jobs are submitted
- Is there an predefined ordering in which datasets are submitted?
- Do we start at ~80% dataset completion?

Benchmarking on OSG

It seems inevitable that we will need to leverage the computing capacity of OSG for long-running benchmarks. That would require relying on XRootD for storage of intermediate results. Since XRootD throughput is a shared resource we'd like your input on how we could best share it.

- Is there a bandwidth consideration for running on OSG?
 - We may need to download campaign files
 - We will need to upload histogram artifacts
 - Does it even make sense to do campaign benchmarks this way? i.e. this only fits well to fload some eicweb-style benchmarks instead (simulation and benchmarking run together).
- Currently, we upload image artifacts from eicweb to /work/eic2, then copy to the eic3 disk.

Benchmarking during campaign production

It should be possible to combine simulation and benchmarking step.

- How to publish and manage benchmarking payloads?
 - Payload type depends on dataset
 - Payloads need to be available for reuse for offline analysis (development, benchmark redo)
- How can robust can we make campaign job against failure in payload?
- Artifacts would have to be uploaded to XRootD