

General Updates for May 2023 Train

Sakib Rahman
25 May 2023

Conductor's Log Book

- Created account on Open Science Grid and subscribed to EIC project
- Connected to osg login node login05.osgconnect.net
- Cloned https://github.com/eic/job_submission_condor
- Within the job submission folder the following script is executed for each dataset provided in the form of csv files

EBEAM=18 PBEAM=275 DETECTOR VERSION=23.05.1
DETECTOR CONFIG=epic_arches JUG_XL_TAG=23.05.1-stable
./scripts/submit_csv.sh osg_csv hepmmc3 DIS/CC/18x275/minQ2\=100/DIS_CC_18x275_minQ2\=100.csv 2

The diagram illustrates the mapping of command-line arguments to their roles in the job submission process. Arrows point from specific parts of the command line to descriptive labels:

- Format**: Points to `osg_csv`
- Input data type**: Points to `hepmmc3`
- Dataset**: Points to `DIS/CC/18x275/minQ2\=100/DIS_CC_18x275_minQ2\=100.csv`. Below this label is the URL https://github.com/eic/simulation_campaign_datasets.
- Tagged geometry**: Points to `DETECTOR VERSION=23.05.1`
- Tagged container**: Points to `JUG_XL_TAG=23.05.1-stable`
- Target time**: Points to the final argument `2`

- Jobs when submitted goes into idle queue before running. Failed jobs go into hold queue. Use `hold_review_and_release.sh` to monitor reasons for common job failures and resubmit. To release all jobs on hold at once use `condor_release <username>`
- Running both `epic_arches` and `epic_bryce canyon` in May train. Should be the case until we converge to a single config.

Conductor's Log Book

- To view the output from the campaign in the S3 storage:
 - Do this ONCE: `./mc config host add S3 https://eics3.sdcc.bnl.gov:9000 eicS3read eicS3read`
 - Then you can use `./mc ls S3/eictest/EPIC` or `./mc tree S3/eictest/EPIC` to view content
- For the majority of the jobs we only store RECO output. For a small number of datasets, we are storing FULL sim results (roughly 10000 events). So, far we have been submitting duplicate jobs with a modified template that stores full sim output. Deprioritize using `condor_prio` command after first 200 jobs in the array are queued in RUN status. Then kill the jobs when the RUN jobs reach DONE status.

Datasets run so far

```
S3/eicetest/EPIC/FULL/23.05.2
├── epic_arches
│   ├── DIS
│   │   └── NC
│   │       ├── 10x100
│   │       │   └── minQ2=100
│   │       ├── 18x275
│   │       │   └── minQ2=100
│   │       └── 5x41
│   │           └── minQ2=100
│   └── epic_brycecaanyon
│       ├── DIS
│       │   └── NC
│       │       ├── 10x100
│       │       │   └── minQ2=100
│       │       ├── 18x275
│       │       │   └── minQ2=100
│       │       └── 5x41
│       │           └── minQ2=100
```

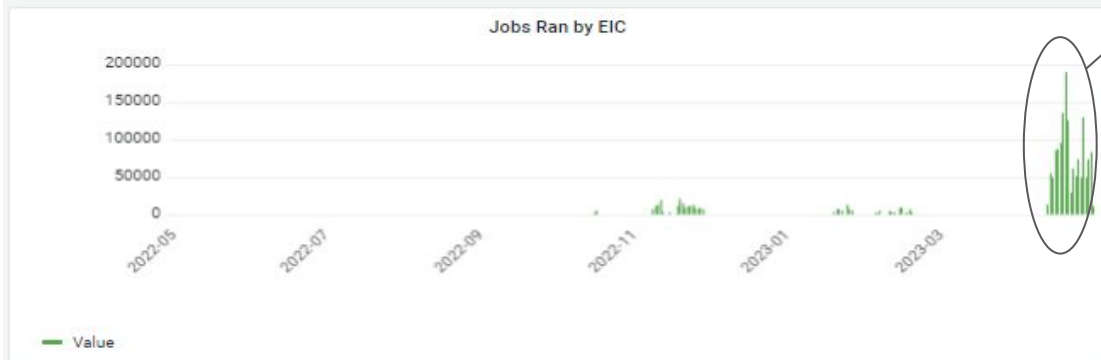
```
S3/eicetest/EPIC/RECO/23.05.2
├── epic_arches
│   ├── DIS
│   │   └── NC
│   │       ├── 10x100
│   │       │   └── minQ2=100
│   │       ├── 18x275
│   │       │   ├── minQ2=1
│   │       │   ├── minQ2=10
│   │       │   ├── minQ2=100
│   │       │   └── minQ2=1000
│   │       └── 5x41
│   │           └── minQ2=100
│   └── epic_brycecaanyon
│       ├── DIS
│       │   ├── CC
│       │   │   ├── 10x100
│       │   │   │   ├── minQ2=100
│       │   │   │   └── minQ2=1000
│       │   │   ├── 18x275
│       │   │   │   ├── minQ2=100
│       │   │   │   └── minQ2=1000
│       │   │   └── 5x41
│       │   │       └── minQ2=100
│       │   └── NC
│       │       ├── 10x100
│       │       │   ├── minQ2=1
│       │       │   ├── minQ2=10
│       │       │   ├── minQ2=100
│       │       │   └── minQ2=1000
│       │       ├── 18x275
│       │       │   ├── minQ2=1
│       │       │   ├── minQ2=10
│       │       │   ├── minQ2=100
│       │       │   └── minQ2=1000
│       │       ├── 5x100
│       │       │   ├── minQ2=1
│       │       │   ├── minQ2=10
│       │       │   ├── minQ2=100
│       │       │   └── minQ2=1000
│       │       └── 5x41
│       │           ├── minQ2=1
│       │           ├── minQ2=10
│       │           └── minQ2=100
```

Roughly 1/4th through the jobs producing RECO outputs for brycecaanyon. Jobs for arches just started getting queued to run.

Roughly 10-20K events available for the FULL outputs listed here

Comparison with Past Campaign ([GRACC Dashboard](#))

How many jobs has EIC ran on the OSPool?



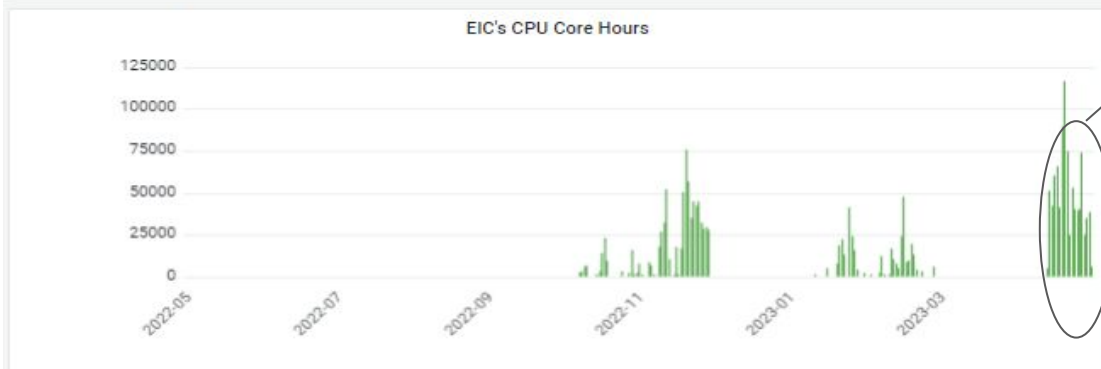
May 2023
Campaigns

Two tagged trains

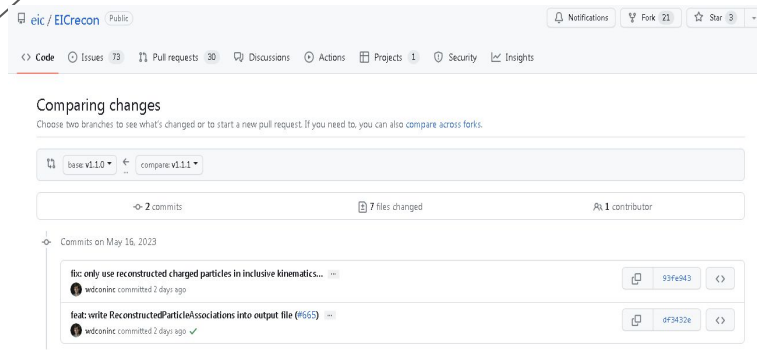
23.05.1: Released on
May 7th.
Deprioritised.

23.05.2:
Released on May
18th.

How many compute resources has EIC been using?



May 2023
Campaigns



Resource Locations

Where has EIC been running their jobs?

