Batch Exercise

Introduction

Christoph Wissing (DESY)

FH Sustainable Computing Workshop
October 2024

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

Batch Exercise

Typical use-case in an analysis flow

- Large data sample in a “heavy” format
  - Here: CMS AOD (Analysis Object Data), ~ 150kb/Event from Run1 - OpenData
  - Total dataset size order a few Terabytes
  - Not very handy for regular analysis cycle
- Data “slimming”
  - Save only relevant variables (or objects)
  - Typical size reduction factor ~100
  - Requires a full run over input AOD dataset to write out the small (custom) format

Note: Our example violates best practices (s. next slide)

Batch Exercise

Batch farm allows access to a lot of CPU resources

- Mistakes can lead to huge waste of resources
- There is a lot of bad-put each and every day on the NAF and in the Grid!

A few considerations

- Plan of processing
  - Has a similar processing already happened?
  - Can I use a smaller or pre-processed existing sample?
  - Consult with colleagues and team up for similar processing campaigns
- Prepare the processing
  - Test the code and validate the output on small samples
  - Know the requirements of your batch jobs
  - Submit a small number of jobs and double check the expected outcome

Batch Exercise

Batch farm allows access to a lot of CPU resources

- Mistakes can lead to huge waste of resources
- There is a lot of bad-put each and every day on the NAF and in the Grid!

A few considerations

- Plan of processing
  - Has a similar processing already happened?
  - Can I use a smaller or pre-processed existing sample?
  - Consult with colleagues and team up for similar processing campaigns
- Prepare the processing
  - Test the code and validate the output on small samples
  - Know the requirements of your batch jobs
  - Submit a small number of jobs and double check the expected outcome

For our example AOD dataset there is already a nanoADO available (2kB/evt)!!!

Let’s get started

Find the tutorial on the DESY Gitlab:

- DESY-Gitlab:sustainable-coding-tutorial/batch-exercises
- Exercise needs to run on NAF workgroup-servers

Acknowledgements

- Tutorial is based on a CMS Opendata example
  - Gitlhub:cms-opendata-analyses/DimuonSpectrum2011
- Significant help to adopt it for the NAF and document it for our FH sustainable computing workshops came from Ben Brüers (PhD student in the DESY ATLAS group)