cta cherenkov telescope array

CTA Batch Use Cases for ESCAPE

G Hughes

Outline

- CTA has 3 use case that envisage the use of batch processing with ESAP
  - 3. Generation of Instrument Response Function (IRF)
  - 13. CONCORDIA
  - 12. Analysis of (simulated) CTA science data by a Principal Investigator (PI)
- There is a linked use case that was developed for DAC21 that could also be linked
  - 2. Data Reprocessing

https://docs.google.com/document/d/1HRd_jzw2lro_CB2_GyfytVYgnErsYd6J8RqvVGG50X8

Outline

- CTA has 3 use case that envisage the use of batch processing with ESAP
  - 3. Generation of Instrument Response Function (IRF)
  - 13. CONCORDIA
  - 12. Analysis of (simulated) CTA science data by a Principal Investigator
- There is a linked use case that was developed for DAC21 that could also be linked
  - 2. Data Reprocessing

CTA Workflow: 2 – 3 - 12

3. Generation of Instrument Response Function

1. User navigates to ESAP and logs in
2. **User input** of parameters.
   1. (Time period, model, systematic uncertainty)
   2. Validation of parameters
(collection of further information from DB)
   3. Confirm valid simulation does not already exist
   4. Calculate number of required events
3. Search for **compute** resources
4. Job **submission** via WLMS (e.g. DIRAC)
5. Job **monitoring**
6. Validation of simulated IRF
7. Ingest IRF in the data lake with appropriate metadata

13. CONCORDIA

1. User navigates to the ESAP
2. logs in and **recognised** as an ESCAPE & CTA/KM3NeT member (EGI DIRAC)
3. Navigate to CONCORDIA page / batch page on ESAP
4. They are able to configure a container
    1. if the container exists it is selected
    2. if not it is created
5. The job is then **submitted** to EGI using DIRAC
6. The job can be **monitored**
7. Results and logs can be accessed/downloaded

12 Analysis of (simulated) CTA science data by a Principal Investigator

1. User logs in to the ESAP and is **recognised** as a CTA project PI
2. Search for Data in the data lake
   Search for (simulated) CTA DL3 level data by project ID
   Select data from search results or select all
   Add to shopping basket.
3. Search for corresponding IRF (instrument response function) for the data selected
   Add to shopping basket.
4. [Optional] Search for corresponding metadata, log files etc ….
5. Batch submission
   1. Search for **compute resources**
   2. Data is transferred to resource cache
   3. **Submit** job(s)
   4. **Monitor** job(s)
6. Save data to data lake or download final results

2. Data Reprocessing (RUCIO-DIRAC DAC21)

1. Raw (DL0) data is identified on tape (obsid or time range) e.g. via metadata e.g. using getMetaData method https://tinyurl.com/rnyyvab5
2. Data volume is calculated.
3. Data is staged from tape (COLD) storage to temporary disk (HOT).
4. Data is reprocessed using CTA pipeline software via the workload management system (WMS, based on DIRAC) using a cache area for on-the fly, transient data products
5. Final data products (DL3) are verified.
6. Cache and temporary data is cleared.
7. Ingest the resulting new DL3 data into the datalake.
8. Update the corresponding metadata.

Common themes

- Navigating and logging into ESAP via IAM
- IAM give experiment and credentials for resources and access
- Job preparation:
  - user input or DB input
  - data and IRF from data lake or container creation
- Finding compute resources
- Job submission & monitoring
- Retrieving the data

Q&A

- How is the software to be executed in batch packaged?
  - Available in containers
    - from repositories?
  - Already installed on the processing system, script on the backend
- Is support for specific execution systems required?
  - DIRAC (CTADIRAC and EGI)
  - Interested in looking at other systems e.g. Rosetta
- How is the workflow to be executed described? Is user interaction required, to e.g. insert job parameters or configuration?
  - UC2 & 13 require user input
  - UC12 depends on the data selected
- Where is the input data for the workflow drawn from? How is it exposed to the workflow system?
  - UC12 Data lake
- How will the user retrieve their results?
  - Open question … reverse of getting data from DL.

Final thoughts

- Caching the data: might depend on compute resource or WLM
  - WF depends on where you submit it?

- DIRAC helloworld example

- RCAuth discussion notes
  - DESY-H dCache demo