



# **REANA / PanDA integration for Active Learning**

Wen Guan, Tadashi Maeno, Christian Weber, Torre Wenaus, <u>Rui Zhang</u> University of Wisconsin-Madison, Wisconsin ADC WFMS weekly Mar 03, 2022

#### Introduction

- The PanDA workload management system is widely used for long, resource intensive jobs in ATLAS
- Active learning is one of the ideas that require complex logics between jobs in a single task
  - Loop
  - Conditional
  - nested dependence
- REANA (Reusable Analyses) is a platform for an easy access to the analysis environment and workflow at any time, for the purpose of analysis preservation
- As an end user, they might need to deal with all aspects at some point

#### **Running workflow in PanDA and iDDS**

- Running workflow is possible with PanDA, thanks to iDDS
  - Workflow is a series of tasks with loops, conditional branching, nested dependence, etc.
  - Each job can be a PanDA job (prun, pathena, ...) or a sub-task (another workflow, HPO, ...)
  - Dependence of jobs are defined in a CWL file
  - More info in the doc



## **REANA** tasks in PanDA

- In a view of a bigger picture, REANA tasks might be one step of the whole physics analysis from Monte-Carlo generation to New Physics discovery
- Take advantage of both by combining



Param space

# Integration of PanDA/REANA

- Need to setup communications between the two independent systems
  - PanDA → REANA: PanDA submit REANA task on behalf of user authenticated using PanDA secrets
  - PanDA ← REANA: periodically check if the task is finished (right now) or extend callback receiver in PanDA (in plan)



R. Zhang

ADC WFMS weekly

### Preparation of secrets and payload

#### REANA\_ACCESS\_TOKEN

- REANA requirement for task submission
- PanDA needs to know its value (a string) in order to submit task on your behalf
- Tell PanDA using set\_secret() function in pbook
- Grid PEM files
  - To allow REANA accessing Rucio, you could provide your PEM files
- keytab
  - To allow REANA accessing EOS, you could provide a keytab file
  - Check REANA doc on how to access EOS
- At the moment, using a simple payload for testing
  - Access an input from EOS and produce a plot repo
  - Access an input from Rucio and produce a plot <u>repo</u> (need to containerise python3, voms-clients-java, ruico and gfal2)

## Submit `pchain` tasks

- `pchain` is a new tool in PanDA to submit a workflow task (doc)
  - A CWL script to define the workflow
  - A Yaml file to define input
- Inside the CWL, user can execute any bash commands
  - Can put all commands in the shell script
  - In the script, define how you would submit REANA tasks from a Grid node the same way you would do from on your local machine
    - An example of such a script is Link
- Verify syntax and submit
  - pchain --cwl <cwl\_file> --yaml <inputs.yaml> --outDS user.\$USER.\$taskname --check
  - Drop `--check` for a real submission

#### Examples

- Several working examples are available
  - A simple top-bottom task chain repo
  - A task with complex dependence repo
  - A loop task <u>repo</u>
- All examples mix both the EOS-version and Rucio-version payload in the workflow and are tested successfully at the "CERN" site



#### Conclusion

- Integrating PanDA and REANA will lead to a more powerful system
- We successfully demonstrate complex workflow that involves REANA tasks in PanDA
- The workflow is ready to use for real physics cases
- We don't have a real physics example at hand. Should identify and adapt one for demonstration using the final physics results with the Active Learning community.
  - Christian will join the effort as well
- In the meantime, we will extend the PanDA callback receiver to accept REANA messaging