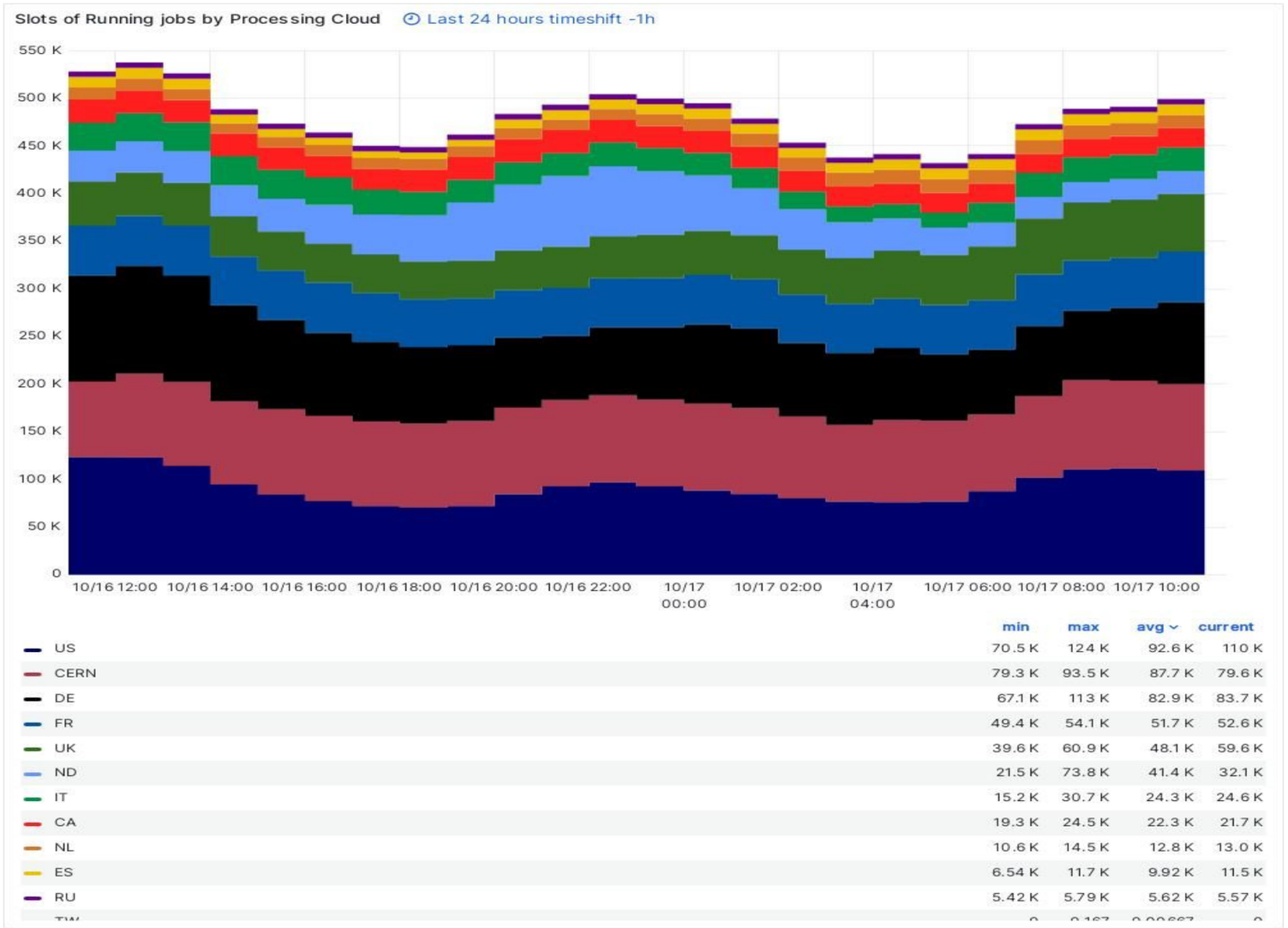
Globus Compute integration with Harvester for PanDA jobs

Wen Guan (BNL), Christian Weber (BNL), Rui Zhang (WISC)
Tadashi Maeno (BNL) and Torre Wenaus (BNL)
on behalf of the iDDS team

ParslFest 2023

Oct 20, 2023

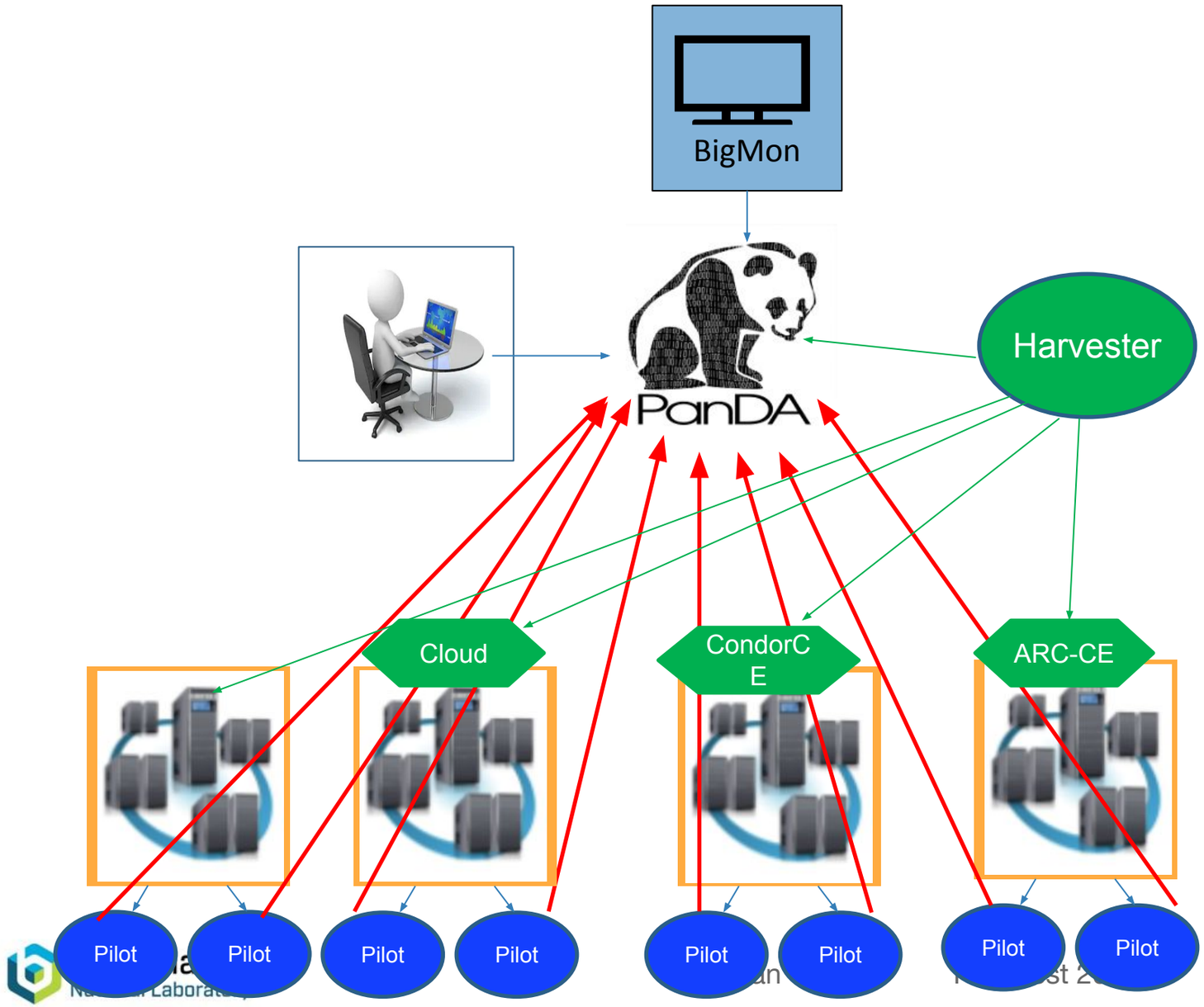
PanDA (Production and Distributed Analysis) System



	min	max	avg	current
US	70.5 K	124 K	92.6 K	110 K
CERN	79.3 K	93.5 K	87.7 K	79.6 K
DE	67.1 K	113 K	82.9 K	83.7 K
FR	49.4 K	54.1 K	51.7 K	52.6 K
UK	39.6 K	60.9 K	48.1 K	59.6 K
ND	21.5 K	73.8 K	41.4 K	32.1 K
IT	15.2 K	30.7 K	24.3 K	24.6 K
CA	19.3 K	24.5 K	22.3 K	21.7 K
NL	10.6 K	14.5 K	12.8 K	13.0 K
ES	6.54 K	11.7 K	9.92 K	11.5 K
RU	5.42 K	5.79 K	5.62 K	5.57 K

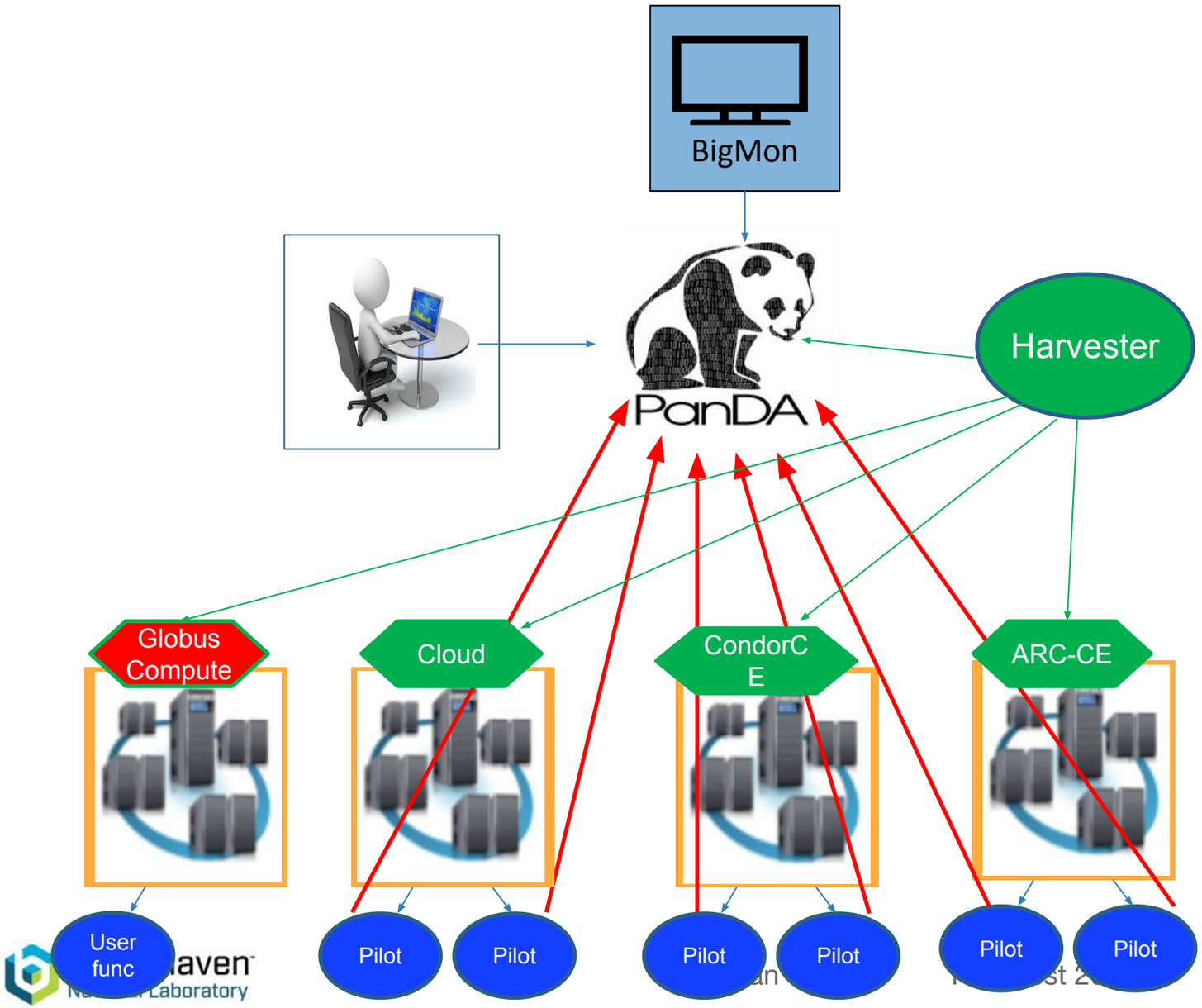
- PanDA is a global workload management system
  - Employed in ATLAS (LHC at CERN), Rubin, sPHENIX experiments
  - For ATLAS, it has more than 100 sites across the world, 450k ~ 500k concurrent core slots.

PanDA Architecture



- PanDA centrally manages jobs
- Harvester submits jobs to different resources and monitor the job status
- It has integrated a lot of different resources
  - Grid
  - Cloud
  - HPC
  - K8S
  - Volunteer Computing: BOINC
  - etc.
- Here the idea is to integrate Globus Compute as a resource provider of PanDA.

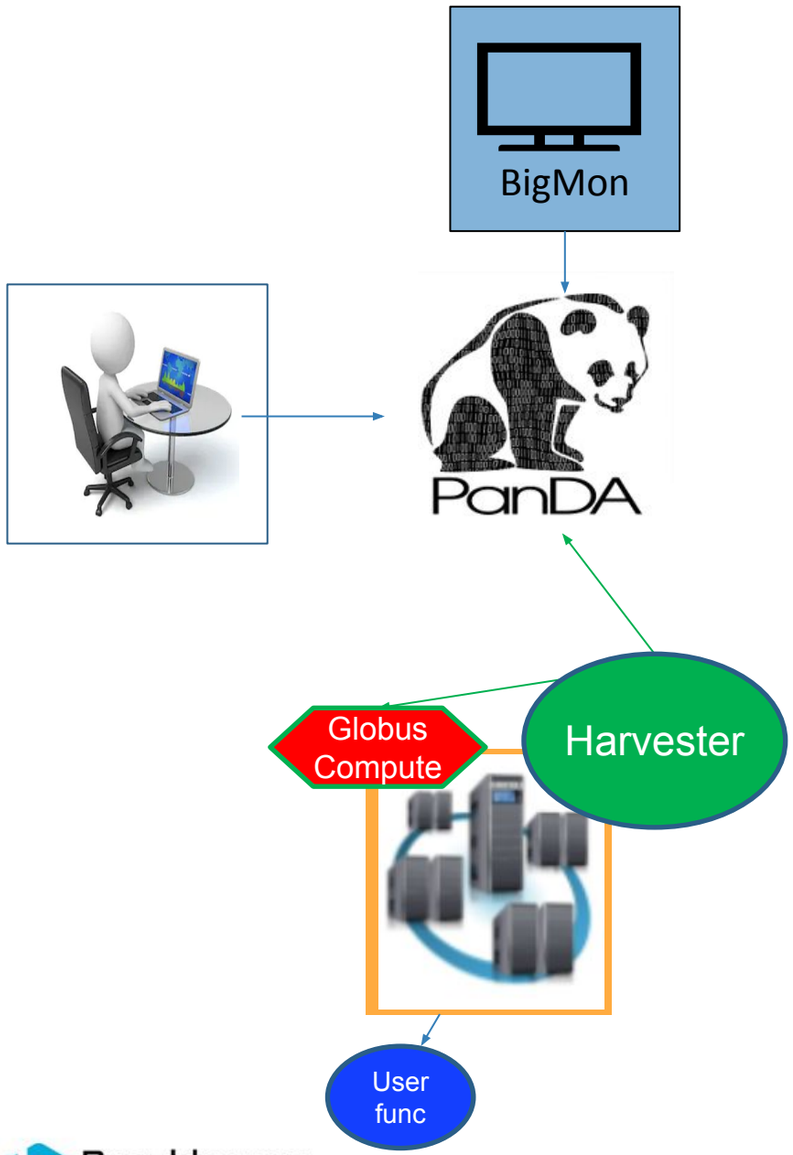
Integrate Globus Compute with PanDA system



- Globus Compute
  - a resource provider
- Harvester
  - downloads PanDA jobs
  - submits PanDA jobs to Globus Compute
  - Monitor the job status
  - Get the outputs

Here is the structure we expected. It works for simple tests (single function, single parsl function, …). However, for complex functions in which user codes call other user functions, it cannot upload thecodes for the other functions

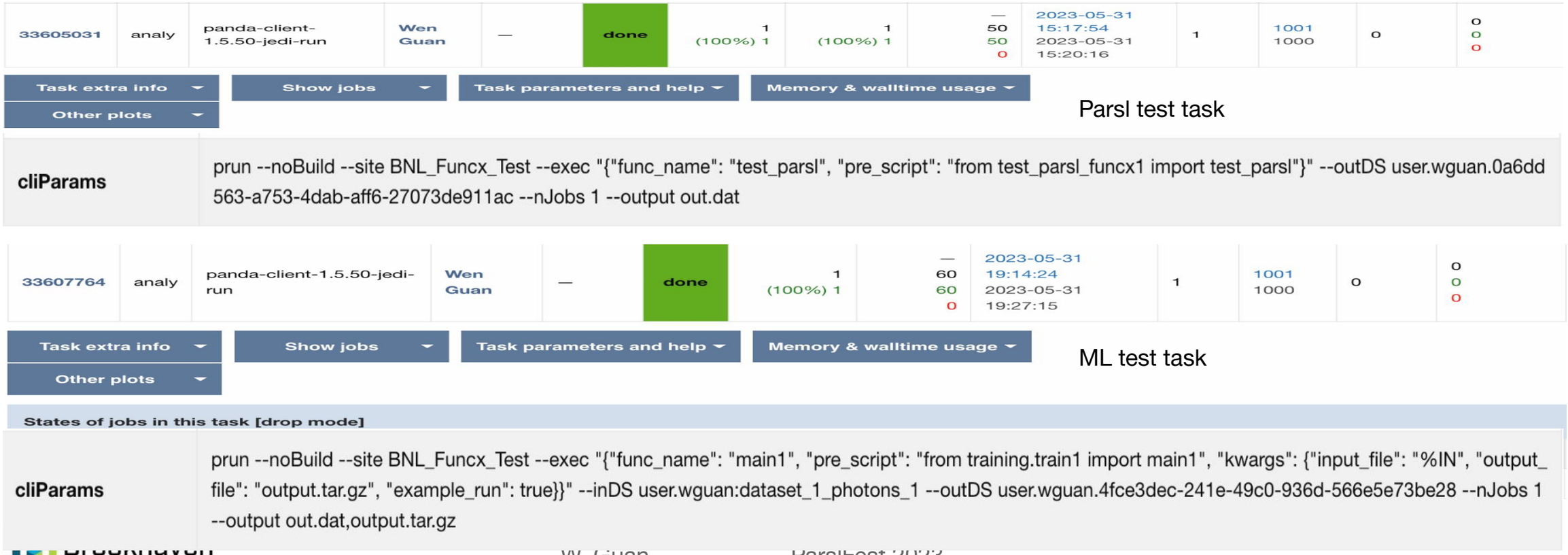
Integrate Globus Compute with PanDA system



- To ship user codes to globus compute
  - Register other user functions to globus compute
    - Too many functions if user codes call other user funcs
    - How to register a class with attributes shared with different functions
  - Globus compute to support container to ship user codes (not ready during this integration work)
  - Ship all user codes to globus compute within a tar file
    - PanDA client collects user codes into a tar file
    - Ship the tar file to PanDA http cache
    - Install harvester at the Globus Compute Endpoint node
    - Harvester downloads the user tar file from PanDA cache and install them on the globus compute endpoint.
    - Have Implemented this model.
    - Limitations:
      - Harvester is required for every GC endpoint

Integration results

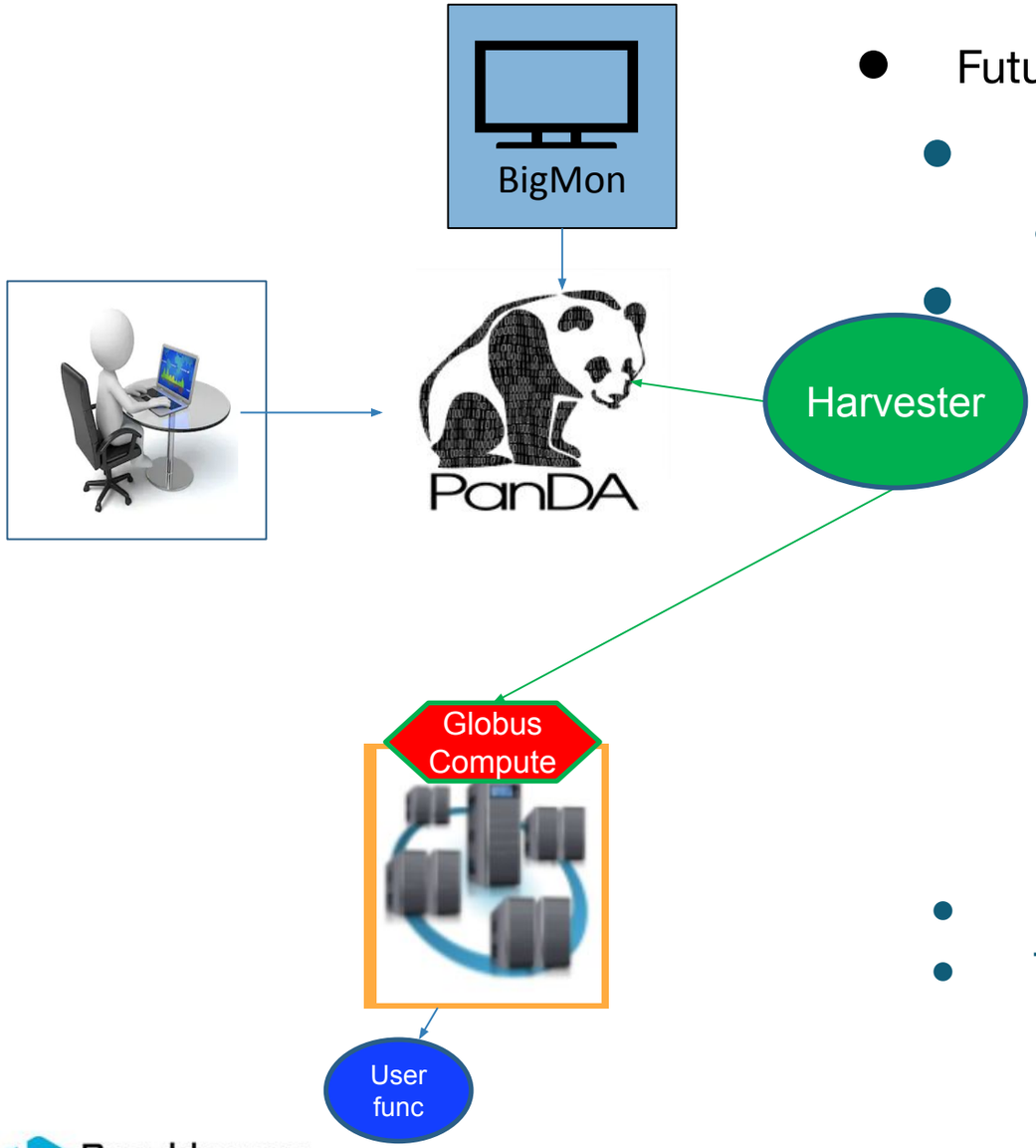
- With harvester installed on the GC endpoint, we are able to
  - Run simple functions
    - with/without parsl
  - Run complex ML (https://github.com/chnzhangrui/funcXtraining)
  - All jobs are submitted from CERN lxplus to ATLAS PanDA instances.



Integration results

- Limitations:
  - Lack of methods in GC to upload complex user codes (currently installed harvester to make it work)
    - Container can be a solution if fully supported
  - Lack of methods in GC to download logs of user codes
    - Without logs, it's difficult to debug
    - (Finally found the logs in some worker directory in GC endpoint. However, no methods to ship the logs to the user side)
  - Some other limitations:
    - For batch run, if one job fails, even all other jobs finish successfully, get_batch_result will raise exceptions. It's good to be able to get the successful results.
    - No kill/abort functions to kill jobs

Future ideas



- Future ideas
  - In the future, we still want harvester to run remotely to use GC as a resource provider
  - PanDA and Harvester can start two jobs for a task
    - The first build job
      - Run a function on GC endpoint, inside this function:
        - Download user codes from PanDA
        - Install the user codes locally
    - The second execution job
      - Call the user function, which can call other user codes.
  - It requires GC to provide some areas to install user codes.
  - This model will also be useful even GC can support container
    - Users can the use container as a base
    - The tar file can include specific codes (not convenient to build the container for every line update).