Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

MCP in PanDA

”Ask PanDA” Update

Paul Nilsson (BNL)

September 10, 2025

NPPS meeting

@BrookhavenLab

News

***Recap**: "Ask PanDA" is an MCP based AI application that enables interactive analyses of PanDA jobs, tasks and metadata, and comes with chat functionality using RAG technology*

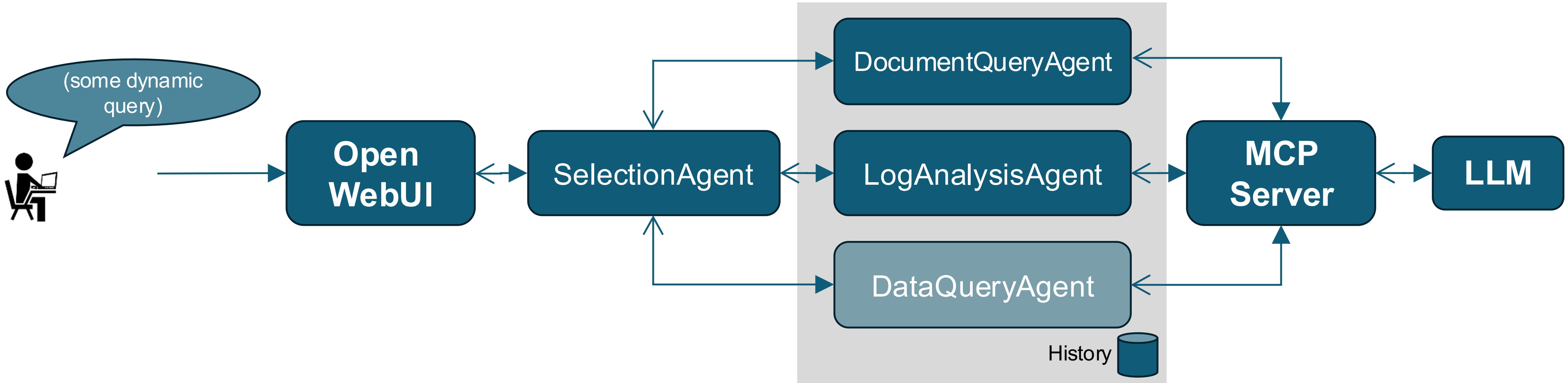
- Support for Gemini^{free}, Anthropic, Open AI, Llama^{local} and Mistral^{*new*} via APIs
- Chat history added to all agents
  - Enables follow-up questions (e.g. after a log analysis of a failed job)
- Log analysis can now handle user job failures
  - Previously only the pilot log was analyzed
    - I.e. error explanations was previously only provided for site/non-user related issues
  - Agent is now downloading the payload output when necessary and performing analysis on that log file instead
- Log agent can also now handle jobs in any state
  - Previously only for failed jobs
  - A job (in any state) will be analyzed using its current metadata
- Integration with Open WebUI

Open WebUI

- Open WebUI is an open-source, extensible web interface for interacting with LLMs locally or remotely
  - https://docs.openwebui.com
- Offer web interface to MCP agents via “pipe function”
  - Pipes allow to add new "model" with custom logic and processing to Open WebUI
  - Interface written for Ask PanDA which appears as a selectable model in the UI
- Challenges
  - Requires older version of chromadb
    - Solved by downgrading chromadb in Ask PanDA – not ideal solution, e.g. db corruption occurred earlier this week and this didn’t happen before with a newer version of chromadb
  - Not obvious how to get access to chat id in pipe function (sqlite key needed for chat history)
    - The chat id was accidentally removed in a PR last year. When it was fixed, it was not documented how to get it back due to heavy refactoring of Open WebUI. At least ChatGPT hallucinates on the matter but eventually suggested that it could be brought back by adding metadata object to function header
  - Later versions include support for MCP – which complicates use of existing stand-alone MCP agents
    - Broke the Ask PanDA interface since it lead to circular imports of some python modules
    - Solved by refactoring and insulation of problematic modules

Dynamic Questions in Chatbot

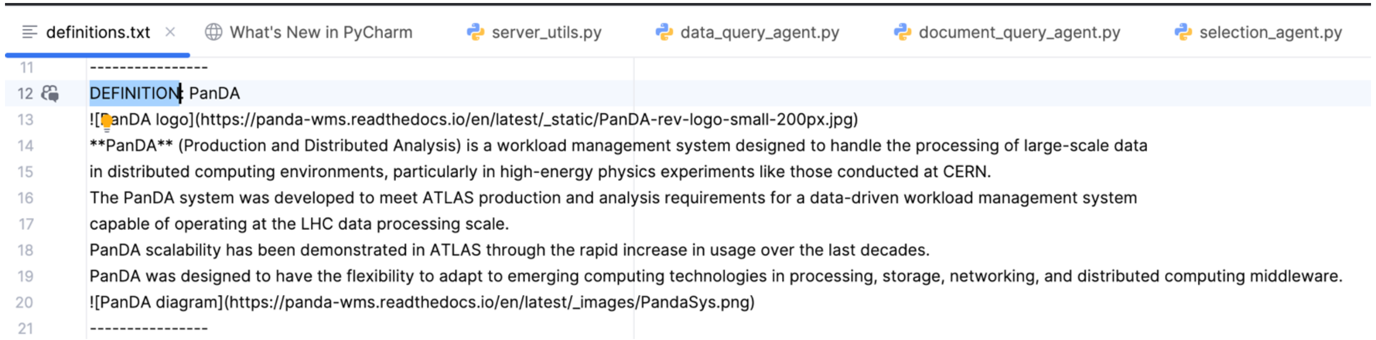
- In case the user asks dynamic questions, the pipe function must select the correct agent
  - “Why did my job NNN fail?”, “How many of my jobs in task MMM failed?”
- Idea is to use the LLM for help
  - Describe to it which agents are used for what, then ask it to select the proper agent for a given prompt
  - Handled by the ”Selection Agent”, called by the pipe function



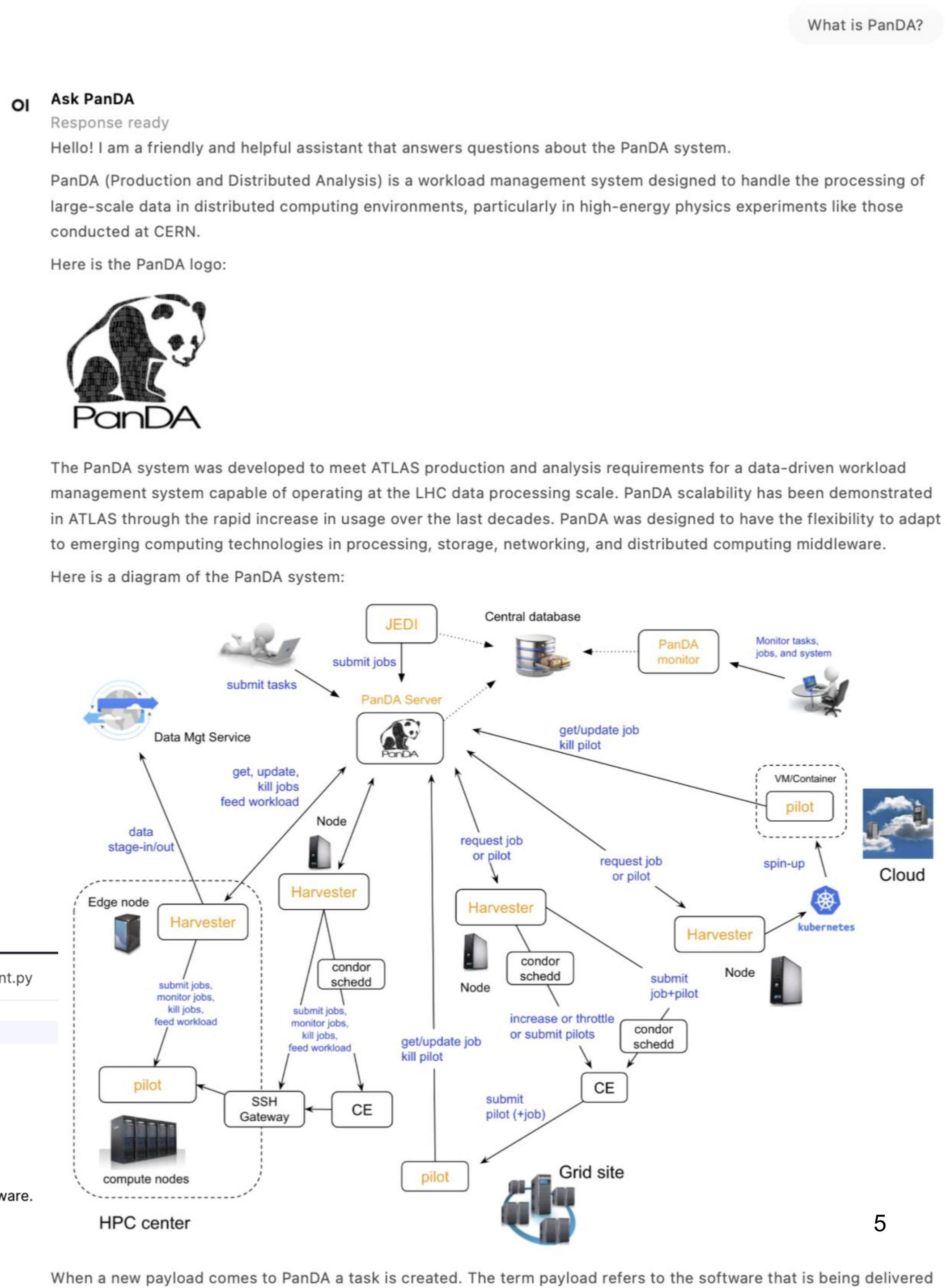
RAG example

- **Prompt**: What is PanDA?

- Document agent selected by the pipe function and Selection agent
- Corresponding context is chosen and enhanced prompt is sent to the LLM
  - Project file “definitions.txt” contain explanations and image references, stored in chromadb



```
----------------
DEFINITION: PanDA
![PanDA logo](https://panda-wms.readthedocs.io/en/latest/_static/PanDA-rev-logo-small-200px.jpg)
**PanDA** (Production and Distributed Analysis) is a workload management system designed to handle the processing of large-scale data
in distributed computing environments, particularly in high-energy physics experiments like those conducted at CERN.
The PanDA system was developed to meet ATLAS production and analysis requirements for a data-driven workload management system
capable of operating at the LHC data processing scale.
PanDA scalability has been demonstrated in ATLAS through the rapid increase in usage over the last decades.
PanDA was designed to have the flexibility to adapt to emerging computing technologies in processing, storage, networking, and distributed computing middleware.
![PanDA diagram](https://panda-wms.readthedocs.io/en/latest/_images/PandaSys.png)
----------------
```



Ask PanDA

Response ready

Hello! I am a friendly and helpful assistant that answers questions about the PanDA system.

PanDA (Production and Distributed Analysis) is a workload management system designed to handle the processing of large-scale data in distributed computing environments, particularly in high-energy physics experiments like those conducted at CERN.

Here is the PanDA logo:

The PanDA system was developed to meet ATLAS production and analysis requirements for a data-driven workload management system capable of operating at the LHC data processing scale. PanDA scalability has been demonstrated in ATLAS through the rapid increase in usage over the last decades. PanDA was designed to have the flexibility to adapt to emerging computing technologies in processing, storage, networking, and distributed computing middleware.

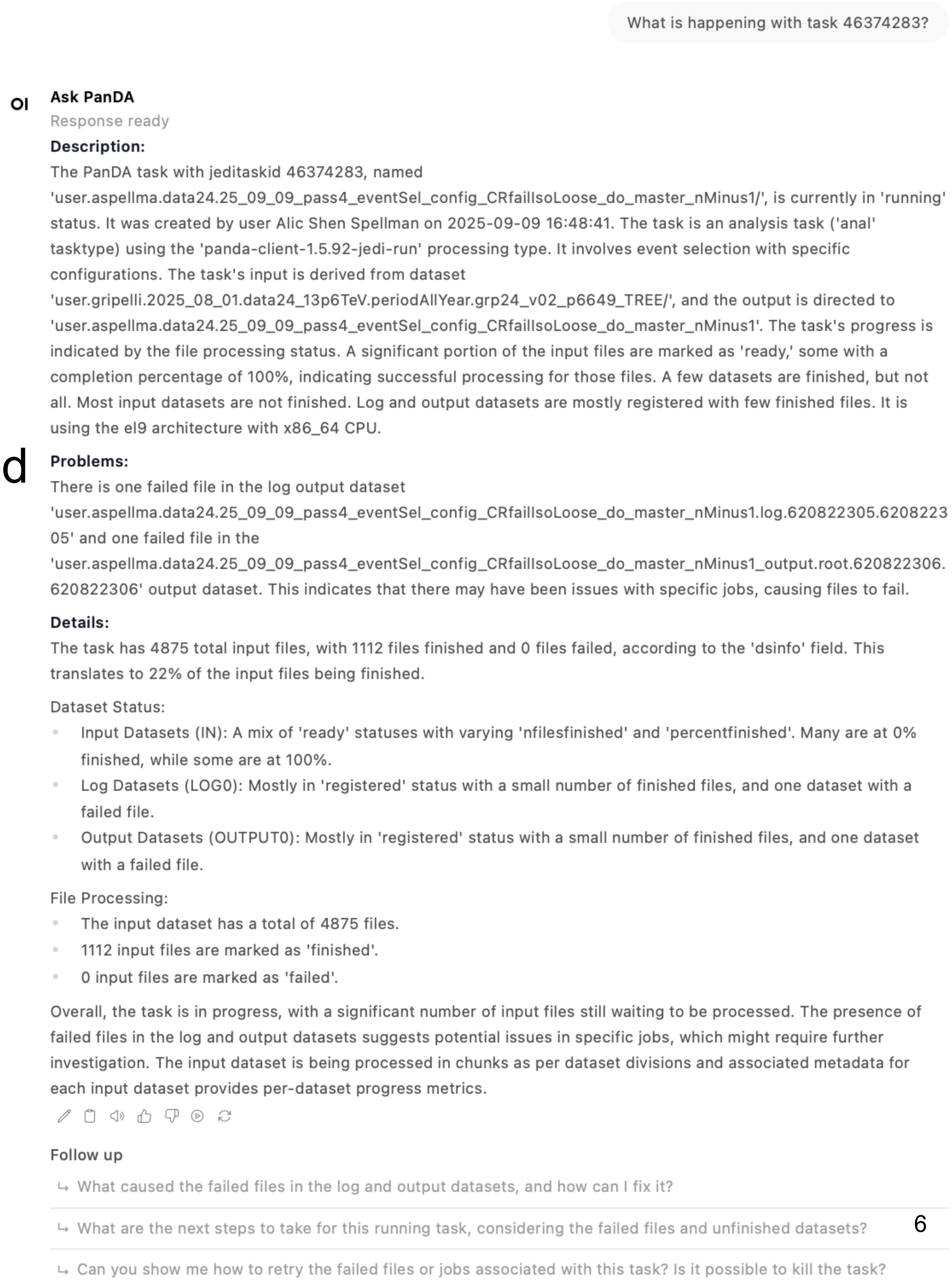
Here is a diagram of the PanDA system:

When a new payload comes to PanDA a task is created. The term payload refers to the software that is being delivered

Task metadata info

- **Prompt**: What is happening with task 46374283?

- Data query agent selected by the pipe function and Selection agent
- Data query agent downloads metadata for the given task
- Prompt is enhanced with special instructions and everything is sent to the LLM



What is happening with task 46374283?

Ask PanDA

Response ready

Description:

The PanDA task with jeditaskid 46374283, named 'user.aspellma.data24.25_09_09_pass4_eventSel_config_CRfailIsoLoose_do_master_nMinus1/', is currently in 'running' status. It was created by user Alic Shen Spellman on 2025-09-09 16:48:41. The task is an analysis task ('anal' tasktype) using the 'panda-client-1.5.92-jedi-run' processing type. It involves event selection with specific configurations. The task's input is derived from dataset 'user.gripelli.2025_08_01.data24_13p6TeV.periodAllYear.grp24_v02_p6649_TREE/', and the output is directed to 'user.aspellma.data24.25_09_09_pass4_eventSel_config_CRfailIsoLoose_do_master_nMinus1'. The task's progress is indicated by the file processing status. A significant portion of the input files are marked as 'ready,' some with a completion percentage of 100%, indicating successful processing for those files. A few datasets are finished, but not all. Most input datasets are not finished. Log and output datasets are mostly registered with few finished files. It is using the el9 architecture with x86_64 CPU.

Problems:

There is one failed file in the log output dataset 'user.aspellma.data24.25_09_09_pass4_eventSel_config_CRfailIsoLoose_do_master_nMinus1.log.620822305.620822305' and one failed file in the 'user.aspellma.data24.25_09_09_pass4_eventSel_config_CRfailIsoLoose_do_master_nMinus1_output.root.620822306.620822306' output dataset. This indicates that there may have been issues with specific jobs, causing files to fail.

Details:

The task has 4875 total input files, with 1112 files finished and 0 files failed, according to the 'dsinfo' field. This translates to 22% of the input files being finished.

Dataset Status:

- Input Datasets (IN): A mix of 'ready' statuses with varying 'nfilesfinished' and 'percentfinished'. Many are at 0% finished, while some are at 100%.
- Log Datasets (LOG0): Mostly in 'registered' status with a small number of finished files, and one dataset with a failed file.
- Output Datasets (OUTPUT0): Mostly in 'registered' status with a small number of finished files, and one dataset with a failed file.

File Processing:

- The input dataset has a total of 4875 files.
- 1112 input files are marked as 'finished'.
- 0 input files are marked as 'failed'.

Overall, the task is in progress, with a significant number of input files still waiting to be processed. The presence of failed files in the log and output datasets suggests potential issues in specific jobs, which might require further investigation. The input dataset is being processed in chunks as per dataset divisions and associated metadata for each input dataset provides per-dataset progress metrics.

Follow up

↳ What caused the failed files in the log and output datasets, and how can I fix it?

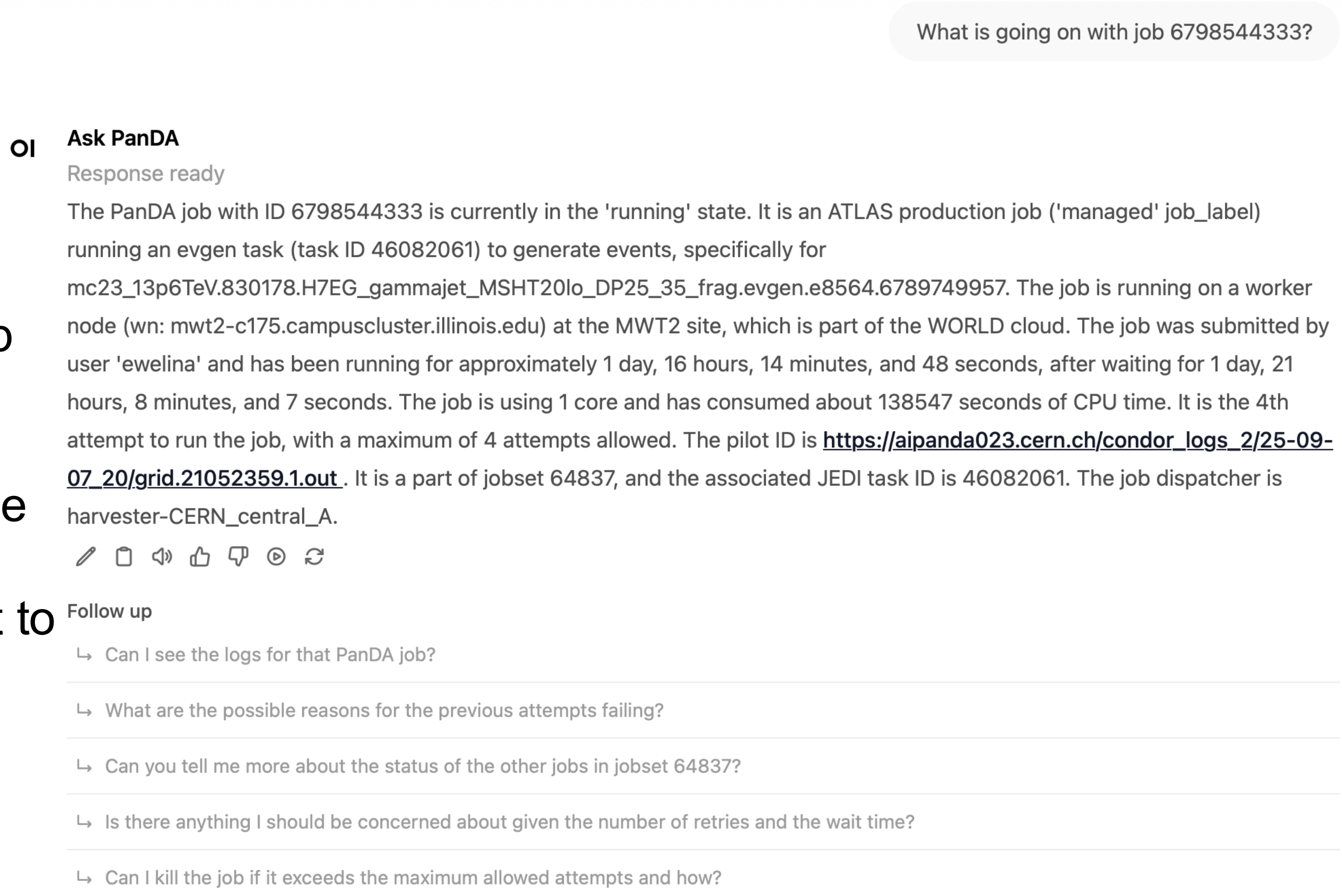
↳ What are the next steps to take for this running task, considering the failed files and unfinished datasets?

↳ Can you show me how to retry the failed files or jobs associated with this task? Is it possible to kill the task?

Running job

- **Prompt**: What is going on with job 6798544333?

- Log Analysis Agent chosen by pipe function and Selection Agent
- Metadata is downloaded and sent to LLM for analysis



What is going on with job 6798544333?

Ask PanDA

Response ready

The PanDA job with ID 6798544333 is currently in the 'running' state. It is an ATLAS production job ('managed' job_label) running an evgen task (task ID 46082061) to generate events, specifically for mc23_13p6TeV.830178.H7EG_gammajet_MSHT20lo_DP25_35_frag.evgen.e8564.6789749957. The job is running on a worker node (wn: mwt2-c175.campuscluster.illinois.edu) at the MWT2 site, which is part of the WORLD cloud. The job was submitted by user 'ewelina' and has been running for approximately 1 day, 16 hours, 14 minutes, and 48 seconds, after waiting for 1 day, 21 hours, 8 minutes, and 7 seconds. The job is using 1 core and has consumed about 138547 seconds of CPU time. It is the 4th attempt to run the job, with a maximum of 4 attempts allowed. The pilot ID is https://aipanda023.cern.ch/condor_logs_2/25-09-07_20/grid.21052359.1.out . It is a part of jobset 64837, and the associated JEDI task ID is 46082061. The job dispatcher is harvester-CERN_central_A.

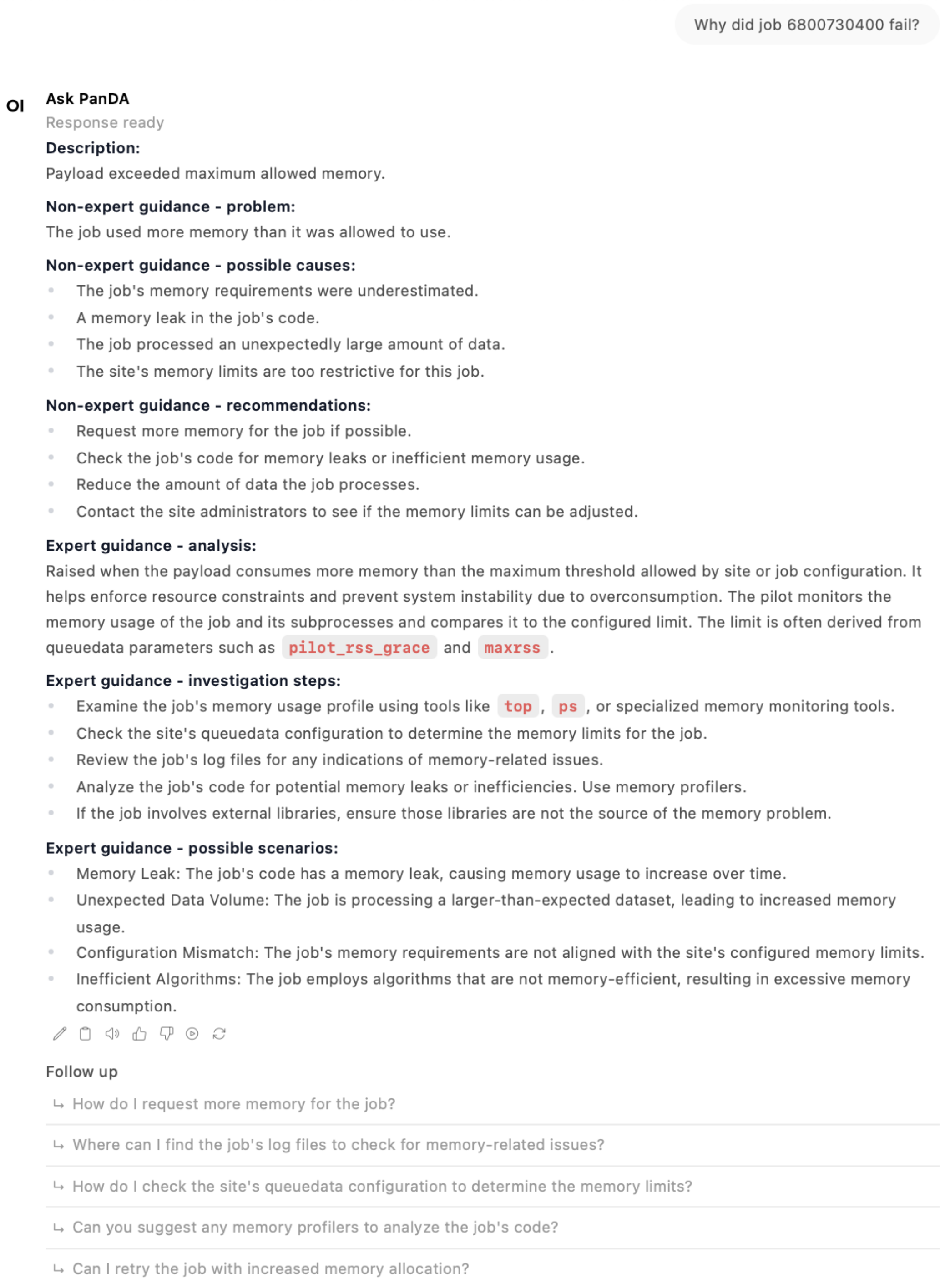
Follow up

- ↳ Can I see the logs for that PanDA job?
- ↳ What are the possible reasons for the previous attempts failing?
- ↳ Can you tell me more about the status of the other jobs in jobset 64837?
- ↳ Is there anything I should be concerned about given the number of retries and the wait time?
- ↳ Can I kill the job if it exceeds the maximum allowed attempts and how?

Failed job

- **Prompt**: Why did job 6800730400 fail?

- Log Analysis Agent selected
- Metadata and job log (if available) is downloaded
- Context window is extracted from job log and sent to the LLM (or only the metadata is sent if the log is not available)



Why did job 6800730400 fail?

Ask PanDA

Response ready

Description:

Payload exceeded maximum allowed memory.

Non-expert guidance - problem:

The job used more memory than it was allowed to use.

Non-expert guidance - possible causes:

- The job's memory requirements were underestimated.
- A memory leak in the job's code.
- The job processed an unexpectedly large amount of data.
- The site's memory limits are too restrictive for this job.

Non-expert guidance - recommendations:

- Request more memory for the job if possible.
- Check the job's code for memory leaks or inefficient memory usage.
- Reduce the amount of data the job processes.
- Contact the site administrators to see if the memory limits can be adjusted.

Expert guidance - analysis:

Raised when the payload consumes more memory than the maximum threshold allowed by site or job configuration. It helps enforce resource constraints and prevent system instability due to overconsumption. The pilot monitors the memory usage of the job and its subprocesses and compares it to the configured limit. The limit is often derived from queuedata parameters such as `pilot_rss_grace` and `maxrss`.

Expert guidance - investigation steps:

- Examine the job's memory usage profile using tools like `top`, `ps`, or specialized memory monitoring tools.
- Check the site's queuedata configuration to determine the memory limits for the job.
- Review the job's log files for any indications of memory-related issues.
- Analyze the job's code for potential memory leaks or inefficiencies. Use memory profilers.
- If the job involves external libraries, ensure those libraries are not the source of the memory problem.

Expert guidance - possible scenarios:

- Memory Leak: The job's code has a memory leak, causing memory usage to increase over time.
- Unexpected Data Volume: The job is processing a larger-than-expected dataset, leading to increased memory usage.
- Configuration Mismatch: The job's memory requirements are not aligned with the site's configured memory limits.
- Inefficient Algorithms: The job employs algorithms that are not memory-efficient, resulting in excessive memory consumption.

Follow up

↳ How do I request more memory for the job?

↳ Where can I find the job's log files to check for memory-related issues?

↳ How do I check the site's queuedata configuration to determine the memory limits?

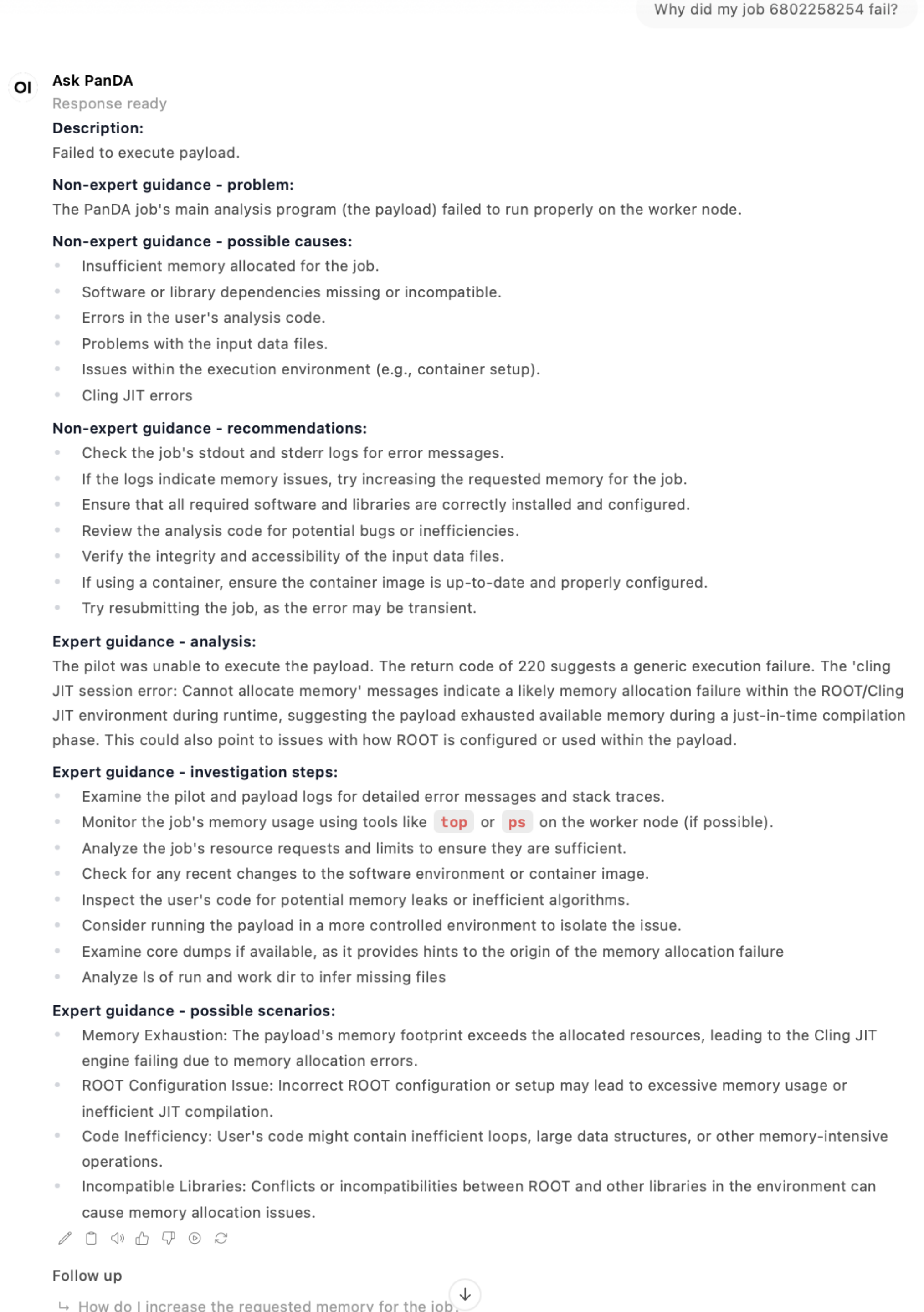
↳ Can you suggest any memory profilers to analyze the job's code?

↳ Can I retry the job with increased memory allocation?

Failed user job

- **Prompt**: Why did my job 6802258254 fail?

- Same workflow as previous example, except that for user payload failures (pilot error code 1305), the log analyzer downloads the user payload stdout instead of the pilot log
- The last ~100 lines of the stdout is sent to the LLM for analysis



Why did my job 6802258254 fail?

Ask PanDA

Response ready

Description:

Failed to execute payload.

Non-expert guidance - problem:

The PanDA job's main analysis program (the payload) failed to run properly on the worker node.

Non-expert guidance - possible causes:

- Insufficient memory allocated for the job.
- Software or library dependencies missing or incompatible.
- Errors in the user's analysis code.
- Problems with the input data files.
- Issues within the execution environment (e.g., container setup).
- Cling JIT errors

Non-expert guidance - recommendations:

- Check the job's stdout and stderr logs for error messages.
- If the logs indicate memory issues, try increasing the requested memory for the job.
- Ensure that all required software and libraries are correctly installed and configured.
- Review the analysis code for potential bugs or inefficiencies.
- Verify the integrity and accessibility of the input data files.
- If using a container, ensure the container image is up-to-date and properly configured.
- Try resubmitting the job, as the error may be transient.

Expert guidance - analysis:

The pilot was unable to execute the payload. The return code of 220 suggests a generic execution failure. The 'cling JIT session error: Cannot allocate memory' messages indicate a likely memory allocation failure within the ROOT/Cling JIT environment during runtime, suggesting the payload exhausted available memory during a just-in-time compilation phase. This could also point to issues with how ROOT is configured or used within the payload.

Expert guidance - investigation steps:

- Examine the pilot and payload logs for detailed error messages and stack traces.
- Monitor the job's memory usage using tools like `top` or `ps` on the worker node (if possible).
- Analyze the job's resource requests and limits to ensure they are sufficient.
- Check for any recent changes to the software environment or container image.
- Inspect the user's code for potential memory leaks or inefficient algorithms.
- Consider running the payload in a more controlled environment to isolate the issue.
- Examine core dumps if available, as it provides hints to the origin of the memory allocation failure
- Analyze ls of run and work dir to infer missing files

Expert guidance - possible scenarios:

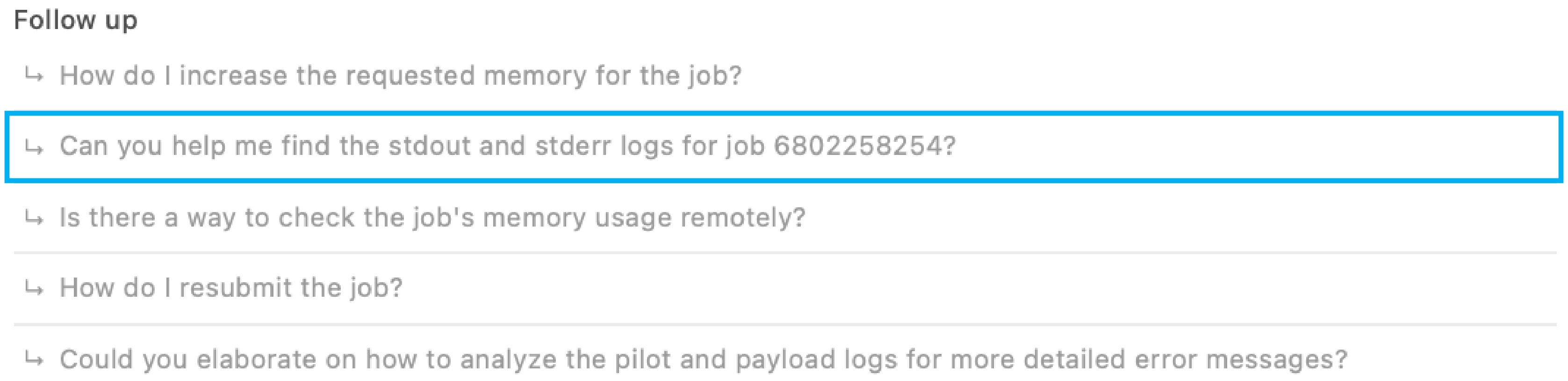
- Memory Exhaustion: The payload's memory footprint exceeds the allocated resources, leading to the Cling JIT engine failing due to memory allocation errors.
- ROOT Configuration Issue: Incorrect ROOT configuration or setup may lead to excessive memory usage or inefficient JIT compilation.
- Code Inefficiency: User's code might contain inefficient loops, large data structures, or other memory-intensive operations.
- Incompatible Libraries: Conflicts or incompatibilities between ROOT and other libraries in the environment can cause memory allocation issues.

Follow up

↳ How do I increase the requested memory for the job.

Next

- Develop custom UI for Ask PanDA based on Open WebUI with new PanDA Monitor
- Sort out stability problems
  - Mistral API currently clogs up localhost, unclear why
  - Chromadb corruption issues
- Add explanations for “follow-up” questions
  - These are suggestions from the LLM and are sometimes (often?) not explained in the documentation
  - Also, some follow-ups require additional coding



Follow up

↳ How do I increase the requested memory for the job?

↳ Can you help me find the stdout and stderr logs for job 6802258254?

↳ Is there a way to check the job's memory usage remotely?

↳ How do I resubmit the job?

↳ Could you elaborate on how to analyze the pilot and payload logs for more detailed error messages?