

Automation and future plans in WFMS

Automation

- Recent improvements in development and infrastructure management
 - JIRA rules to keep our tracker tidy
 - Pre-commit hooks
 - Automated alerts and internal chat-ops: one-stop channel for ATLAS-PanDA operations (e.g. service alerts, server updates)
 - Ongoing: staging environment and CI/CD
- Workload management automation
 - Day to day improvements as proposed by DPA (e.g. shortening jobs for task final 10%)
 - Scouting prediction
 - Error classification
 - Possible future project: anomaly detection and automatic retry rule creation


Future plans

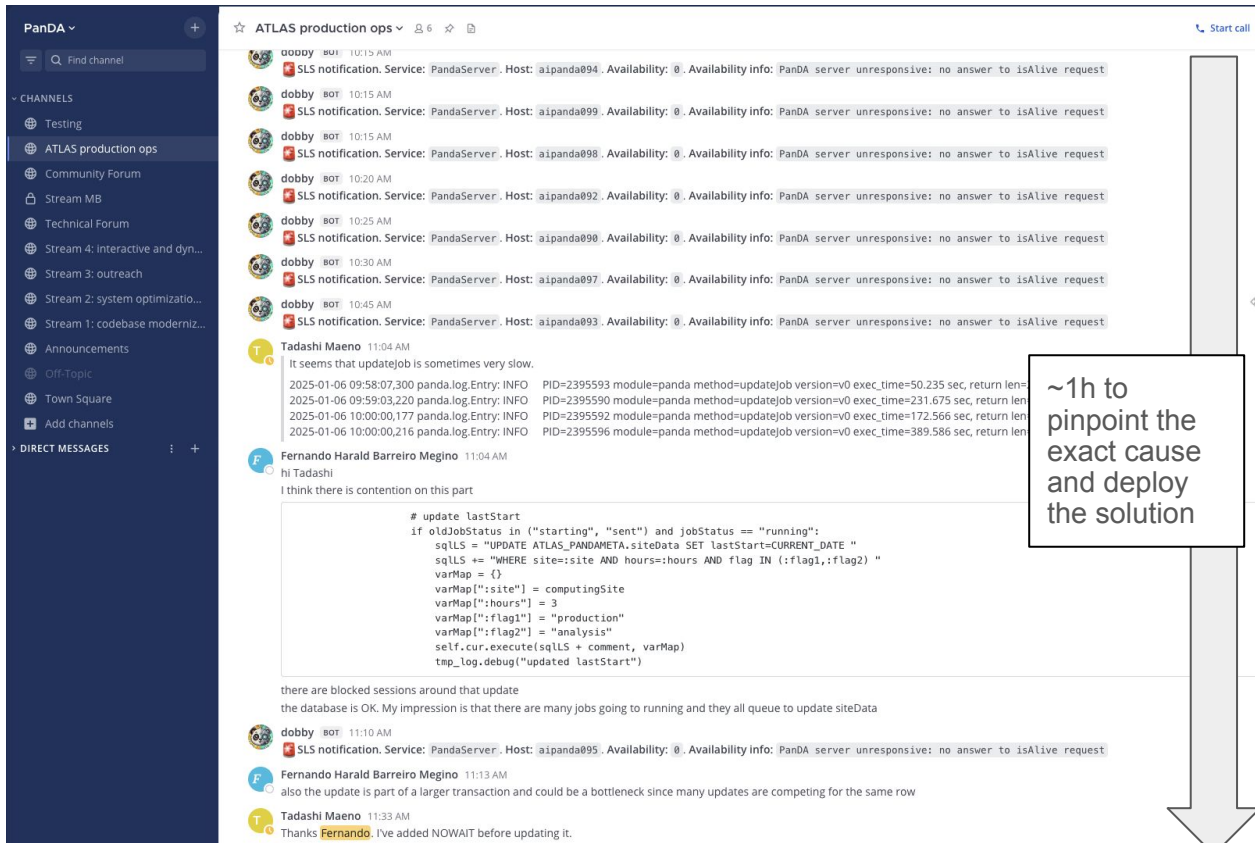
- Stream structure aligned with important activities
- Stream 1: codebase & infrastructure modernization
 - Improving and modernizing codebase. Obsoleting unnecessary code
 - Better use of coding tools and standards
 - Reimplementing API: better code, better structure, documentation, obsolete old technologies
 - Staging environment
- Stream 2: development
 - Catch-all for operations-driven and planned development
 - Current topics: token migration, worker node map, alternative stage-out, HIMEM, metrics, boosting tasks (job cloning, adapting job duration, etc.), better data handling
- Stream 3: community and outreach
 - Improving feedback channels
 - Building multi-experiment community and adoption of new experiments
- Stream 4: workflow and interface evolution
 - Data carousel for analysis and eventual analysis-production consolidation
 - Expansion towards workflow-aware core
 - BigPanDA 2.0 expanding from monitoring to also include interfaces to interact with the system (as operator and/or end user)

Backup



Automation: chat-ops



Updates on who is updating production infrastructure



-  doobby BOT 10:01 AM   panda-jedi@master upgrade on: aipanda084.cern.ch by tmaeno.
-  doobby BOT 10:01 AM   panda-jedi@master upgrade on: aipanda086.cern.ch by tmaeno.
-  doobby BOT 10:01 AM   panda-jedi@master upgrade on: aipanda082.cern.ch by tmaeno.







☆ ATLAS production ops 6 1 Start call



 doobby BOT 10:15 AM  SLS notification. Service: PandaServer. Host: aipanda094. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request



 doobby BOT 10:15 AM  SLS notification. Service: PandaServer. Host: aipanda099. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request


 doobby BOT 10:15 AM  SLS notification. Service: PandaServer. Host: aipanda098. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request

 doobby BOT 10:20 AM  SLS notification. Service: PandaServer. Host: aipanda092. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request


 doobby BOT 10:25 AM  SLS notification. Service: PandaServer. Host: aipanda090. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request

 doobby BOT 10:30 AM  SLS notification. Service: PandaServer. Host: aipanda097. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request

 doobby BOT 10:45 AM  SLS notification. Service: PandaServer. Host: aipanda093. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request



 Tadashi Maeno 11:04 AM
It seems that updatejob is sometimes very slow.


```
2025-01-06 09:58:07,300 panda.log.Entry: INFO PID=2395593 module=panda method=updatejob version=v0 exec_time=50.235 sec, return len=
2025-01-06 09:59:03,220 panda.log.Entry: INFO PID=2395590 module=panda method=updatejob version=v0 exec_time=231.675 sec, return len=
2025-01-06 10:00:00,177 panda.log.Entry: INFO PID=2395592 module=panda method=updatejob version=v0 exec_time=172.566 sec, return len=
2025-01-06 10:00:00,216 panda.log.Entry: INFO PID=2395596 module=panda method=updatejob version=v0 exec_time=389.586 sec, return len=
```


 Fernando Harald Barreiro Megino 11:04 AM
hi Tadashi
I think there is contention on this part

```
# update lastStart
if oldJobStatus in ("starting", "sent") and jobStatus == "running":
    sqlLS = "UPDATE ATLAS_PANDAMETA.siteData SET lastStart=CURRENT_DATE "
    sqlLS += "WHERE site=:site AND hours=:hours AND flag IN (:flag1,:flag2) "
    varMap = {}
    varMap["site"] = computingSite
    varMap["hours"] = 3
    varMap["flag1"] = "production"
    varMap["flag2"] = "analysis"
    self.cur.execute(sqlLS + comment, varMap)
    tmp_log.debug("updated lastStart")
```

there are blocked sessions around that update
the database is OK. My impression is that there are many jobs going to running and they all queue to update siteData

 doobby BOT 11:10 AM  SLS notification. Service: PandaServer. Host: aipanda095. Availability: 0. Availability info: PanDA server unresponsive: no answer to isAlive request

 Fernando Harald Barreiro Megino 11:13 AM
also the update is part of a larger transaction and could be a bottleneck since many updates are competing for the same row

 Tadashi Maeno 11:33 AM
Thanks **Fernando**. I've added NOWAIT before updating it.

Immediate alerts on infrastructure

~1h to pinpoint the exact cause and deploy the solution

Reimplementing API: documentation

Search...

harvester >

task ▾

POST Retry a given task. Requires a secure connection without a production role to retry own tasks and with a production role to retry others' tasks.

POST Enable job cloning for a given task. Requires secure connection and production role.

POST Disable job cloning for a given task. Requires secure connection and production role.

task

Operations related to task

- ⓘ Retry a given task. Requires a secure connection without a production role to retry own tasks and with a production role to retry others' tasks.

REQUEST BODY SCHEMA: application/json
required

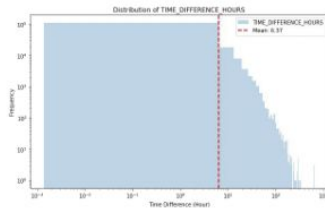
jedi_task_id required	integer JEDI Task ID
new_parameters required	string a json string of new parameters the task uses when rerunning
no_child_retry required	boolean if True, the child tasks are not retried
disable_staging_mode required	boolean if True, the task skips staging state and directly goes to subsequent state
keep_gshare_priority required	boolean if True, the task keeps current gshare and priority

Responses

> 200
Method called correctly
> 403
Forbidden
> 404
Not Found
> 500
INTERNAL SERVER ERROR

Resource Requirement Prediction for Scout Jobs

- **Scout jobs**
 - **Failure rate ~26%**
 - **~6.37 hours scouting time per Task**
- With ML pipeline
 - All 4 resource requirements for each workload upfront, in less than 60 sec
- **The prediction will determine resource requirements for each workload upfront, an crucial building block in dynamically optimizing resource usage and enhancing system resilience**



Work In Progress

- ✓ Automated model prediction pipeline package
- **Deployment in**
 - **Testing environment [Jan 20-24]**
 - **Production [Jan 27-31]**
- **Model Evaluation: (Online Data)**
 - Impact on Total Task completion time
 - Overall and individual model error on a weekly basis
- Paper draft (in progress)

Error Classification

➤ Goals

- Distinguish errors, e.g. temporary vs fatal, user-own vs system-own, etc
- Split catch-all errors to be more fine-grained
- Automate classification based on error patterns, characteristics, etc

➤ Steps

- New error table in the database ✓
- Logic in PanDA ✓
 - E.g. Not to penalize users in case of system-own errors
 - Git repo for data collector [[link](#)]
- Interface for error management ←
- Automation

Step 1: Error table

ID	Source	Code	Message	Description	...	Status
1	Pilot	12	Site error	...		Confirmed
			...			
x	Pilot	135	Site error x	...		Suggested

Jammel+Fernando

Step 3: Interface for error management

The screenshot shows a web interface for error management. At the top right, there is a user profile box for 'Jammel+Tatiana'. Below it is a navigation bar with links for 'Home', 'New', 'List', 'Edit', 'Delete', 'Refresh', and 'Help'. The main content area is divided into two sections: 'Confirmed errors' and 'Suggested errors'. The 'Confirmed errors' section contains a table with columns 'Source', 'Code', 'Message', 'Description', and '...'. One row is visible with 'Pilot' as the source, '12' as the code, and 'Batch crash' as the message. Below this table is a blue button with a right-pointing arrow. The 'Suggested errors' section contains a similar table with a blue header row and several empty rows below it. A blue button with a right-pointing arrow is positioned above this table. At the bottom right, there is a 'Suggest rule' button and a small text box that reads: '(This opens a dialog to choose a PanDA rule it will propose to create a rule from the error)'. A white arrow points from the 'Suggested errors' table up to the 'Confirmed errors' table.

- One of the most crucial ATLAS milestones
 - Joint effort between ATLAS and REDWOOD
- To quantify the system-wide effects resulting from any changes
 - Lack of intrinsic or obvious metrics reflecting system performance due to the system complexity
- Status (details)
 - Defined key metrics to evaluate system performance
 - Averaged composite queuing time
 - Task active time
 - Implemented metrics collector in Dec 2024
- Next steps
 - To identify explanatory variables (metrics potentially affecting the key metrics)
 - To be enhanced to system metrics modeling