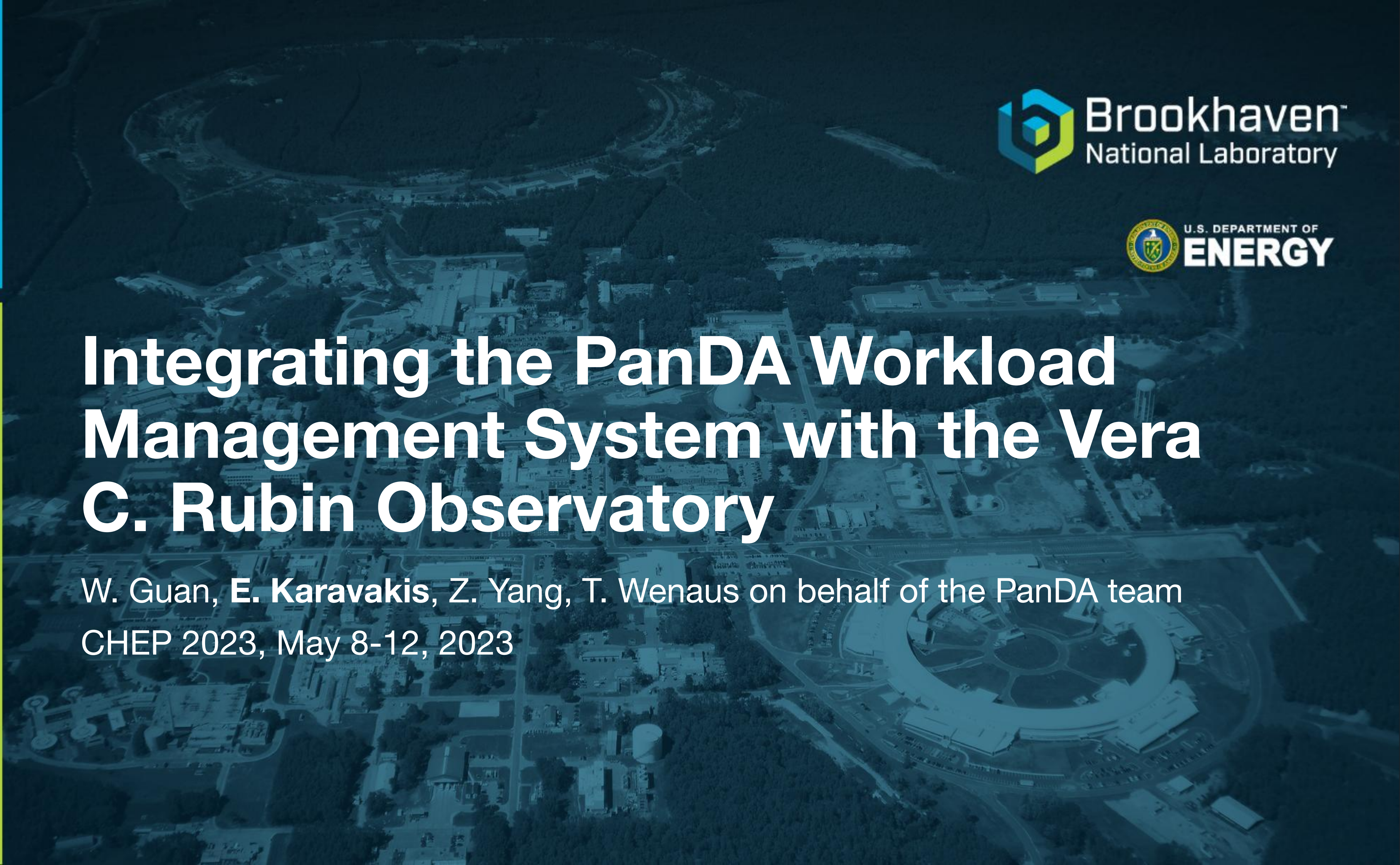


Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Integrating the PanDA Workload Management System with the Vera C. Rubin Observatory

W. Guan, **E. Karavakis**, Z. Yang, T. Wenaus on behalf of the PanDA team

CHEP 2023, May 8-12, 2023

Contents

- Overview of the Vera C. Rubin Observatory
- PanDA developments for Vera C. Rubin
  - Multi-Site Processing
  - Deployment at SLAC on K8s
  - Monitoring Improvements
  - Near Real Time Log Access
  - Prompt Processing
- Summary

VERA C. RUBIN
OBSERVATORY

PanDA

Vera C. Rubin Observatory

- New astronomical facility in Northern Chile starting operations early 2025 - world's largest digital camera!
  - Will conduct a ten-year survey of the Southern Hemisphere sky (referred to as the Legacy Survey of Space and Time - LSST)
  - Every night it will take images of the sky using a 3.2 gigapixel camera
  - Telescope will image entire visible sky every 3-4 nights making it good at detecting objects that have changed in brightness, like supernovae, or in position, like asteroids
  - Its light-collecting power and sensitive camera will discover about 20 billion galaxies and a similar number of stars

- 60 seconds to transfer an image from Chile to California, compare new image to older images to identify changes, and generate alerts based on changes

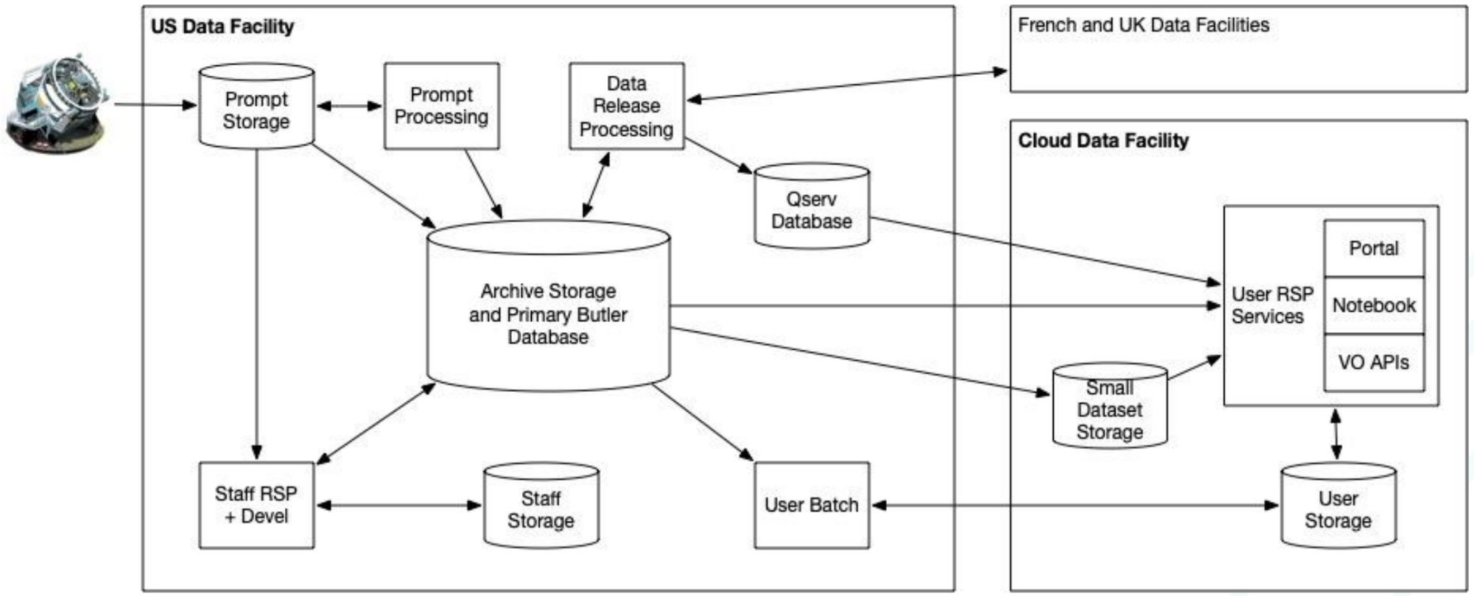
- Will generate ~20 terabytes every 24 hours. At the end of its run, it will have generated ~60 petabytes of raw image data

Brookhaven National Laboratory

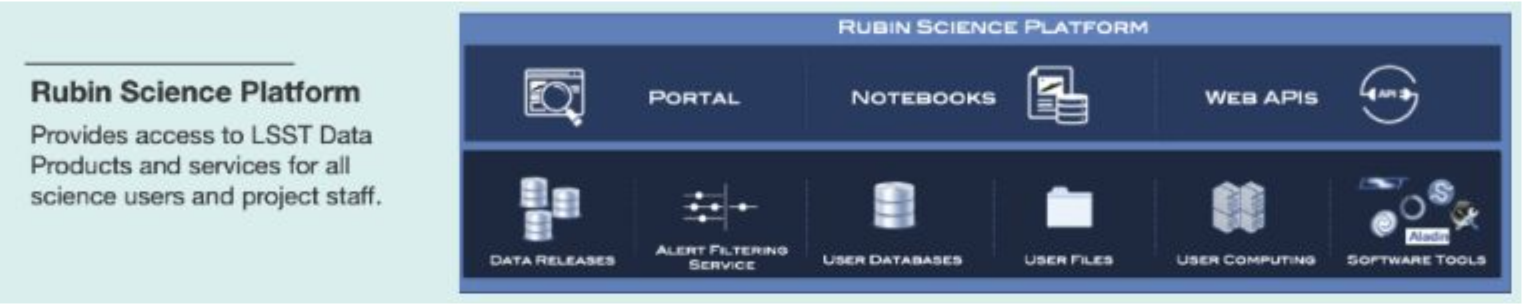
https://rubinobservatory.org

Rubin’s Data Facilities

- There are 3 main data facilities (USDF, FrDF, UKDF) and 1 cloud-based IDF (Google)
  - USDF (S3DF at SLAC National Accelerator Laboratory, CA, USA): All prompt processing, **35%** of data release processing, and data access services to the US and international community
  - UKDF (IRIS and GridPP, UK): **25%** of data release processing
  - FrDF (CC-IN2P3, Lyon, FR): **40%** of data release processing, back up of raw data and published products
  - IDF: Cloud-based Interim Data Facility, used for pre-ops activities



For more details on Rubin see: [Image processing infrastructure to produce the Legacy Survey of Space and Time (LSST)](), track 1, today 11:45.



Rubin Science Platform

Provides access to LSST Data Products and services for all science users and project staff.

PanDA/iDDS

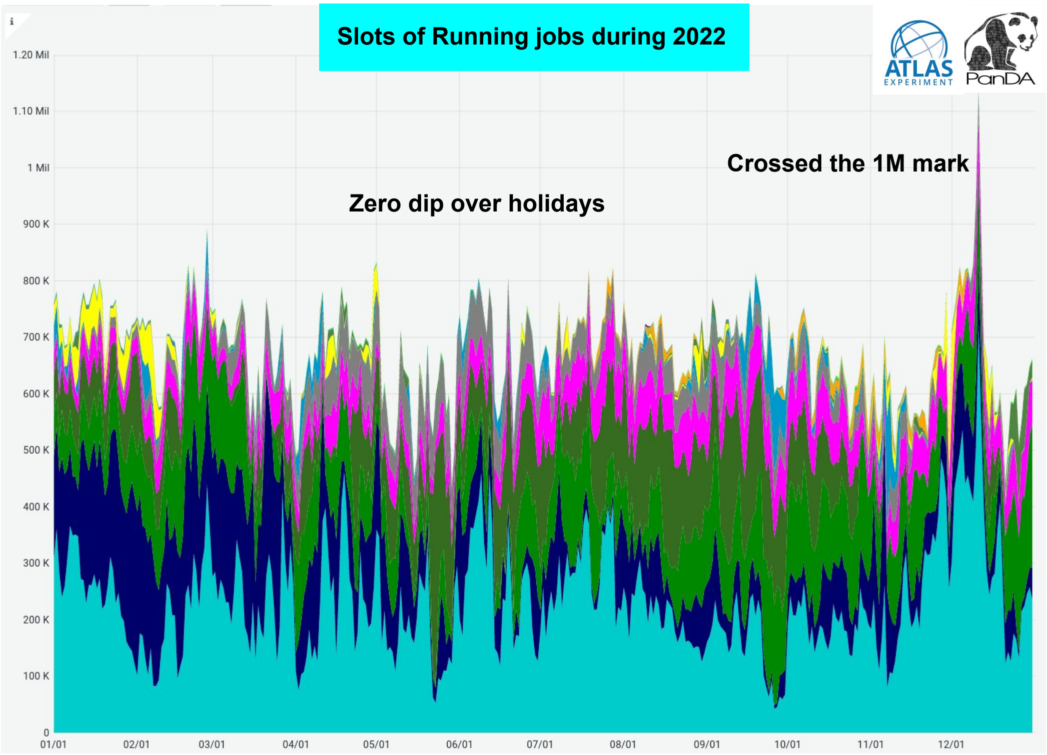
For more details on PanDA see: [Utilizing Distributed Heterogeneous Computing with PanDA in ATLAS](), track 4, 11th May, 12:00.

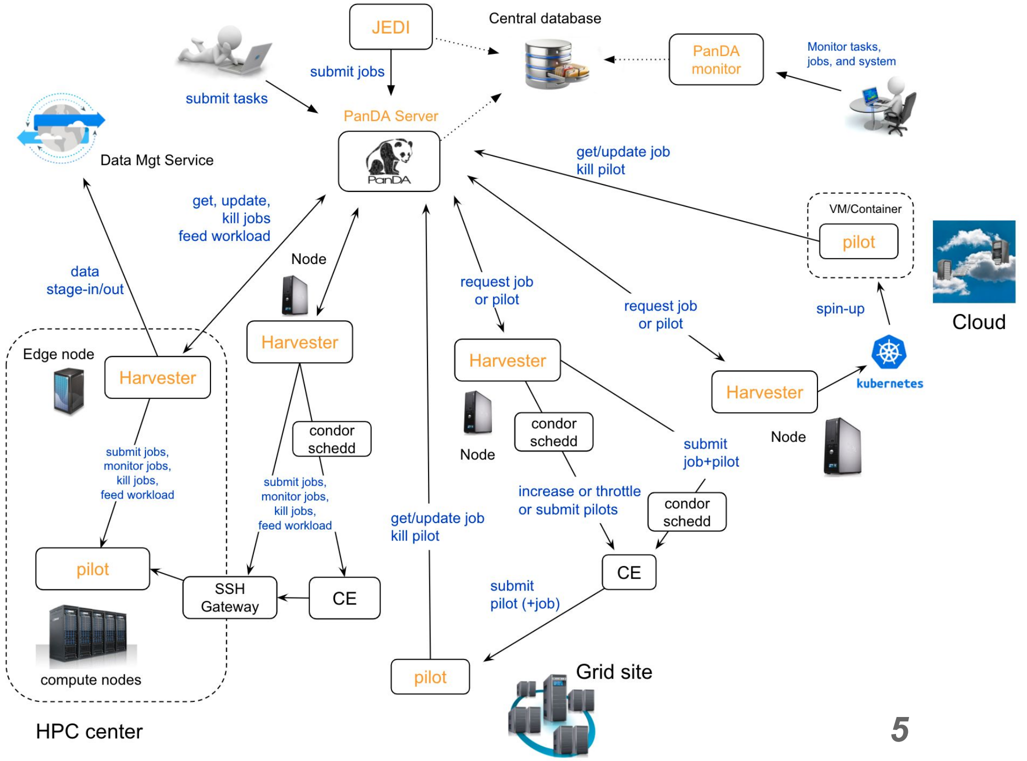
- PanDA: the workload manager
  - Manages 24x365 processing on ~800k concurrent cores globally for ATLAS, all workflows, all resource types, ~1500 users, ~300M jobs/yr, in tandem with Rucio for DM
  - Smooth horizontal scaling (K8s support improves it further)
  - Easy to use submit client, python script or Jupyter notebook
  - **Rubin expects ~200k concurrent jobs (tested)**
  - Particularly relevant for Rubin: improved PostgreSQL support, K8s based services, BNL-led ATLAS-Google project, Jupyter support, OAuth2 authentication, the rise of iDDS

- iDDS: intelligent Data Delivery System
  - Supports arbitrarily complex fine-grained workflows defined via Directed Acyclic Graphs (DAGs) or workflow description languages
  - Used in ATLAS for data carousel (orchestrated disk-efficient tape staging) and a growing list of ML and analysis workflows
  - The basis for supporting Rubin's workflows

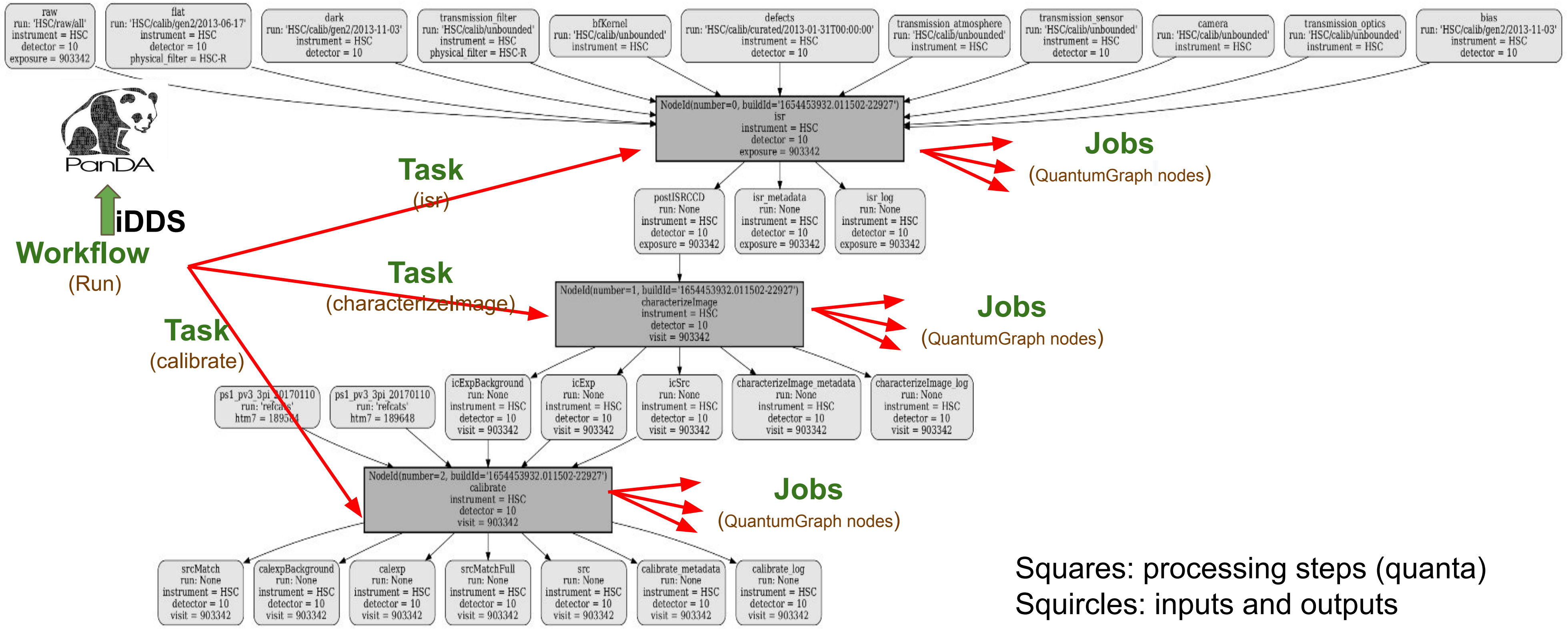
For more details on iDDS see: [Distributed Machine Learning with PanDA and iDDS in LHC ATLAS](), track 4, 8th May, 11:00.

Production and Distributed Analysis System





Mapping Rubin DAG to PanDA workload



iDDS

Workflow
(Run)

Task
(isr)

Task
(characterizeImage)

Task
(calibrate)

Jobs
(QuantumGraph nodes)

Jobs
(QuantumGraph nodes)

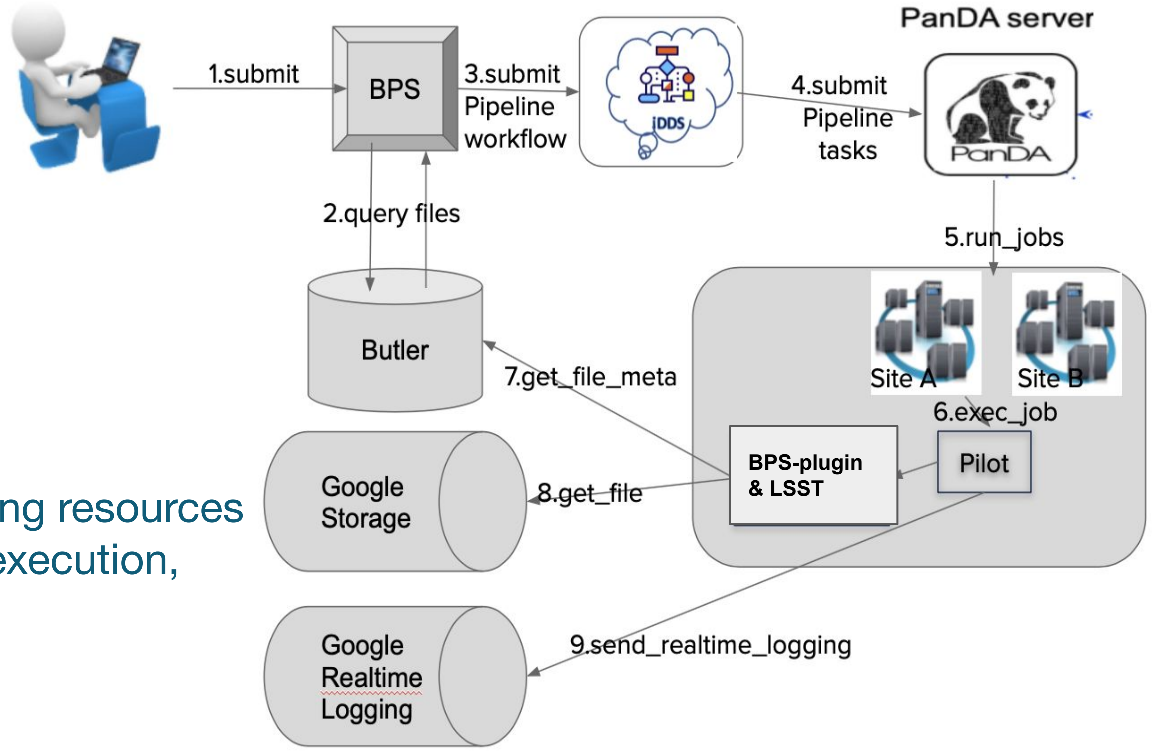
Jobs
(QuantumGraph nodes)

Squares: processing steps (quanta)
Squircles: inputs and outputs

- Processing in Rubin is described with quantum graphs
  - DAG defining inputs and outputs for every node (quantum)
- DAG support in iDDS was developed originally for Rubin and backported to ATLAS for complex analysis workflows

PanDA & Rubin Integration

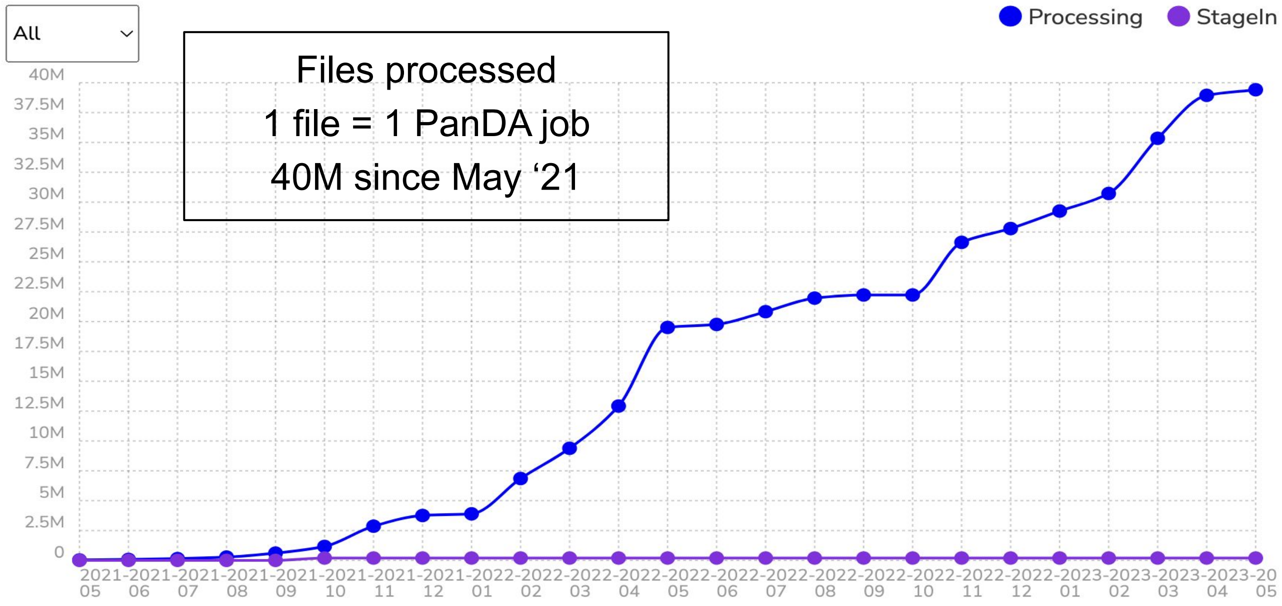
- LSST Science Pipelines (stack)
  - Butler + pipeline framework
- Butler: Data access
  - Interface between data and pipeline tasks
- BPS: Batch Production Service
  - Interface between Butler and PanDA
  - Integrate Rubin with PanDA/iDDS client
- PanDA: Workload management system
  - Manage and schedule Rubin workload to distributed computing resources
  - PanDA pilot integrates Rubin Butler access, Rubin workload execution, Google storage access and real-time logging
- Google Cloud
  - Pilot logs storage and real-time logging
  - GKE clusters (for the Interim Data Facility)



Brookhaven National Laboratory

DRP: Data Release Production

- 2022 production campaigns used PanDA
  - Data Preview 0.2 (DP0.2): **16M** jobs@IDF
  - **Hyper Suprime Cam (HSC)** reprocessing: **8M** jobs@USDF



Files processed

1 file = 1 PanDA job

40M since May '21

- DP0.2: exercise - 3M files, 50 TBs simulated Rubin images generated by the Dark Energy Science Collaboration
  - Reprocessing using latest pipelines
  - Integration test of processing pipelines, data management system, and infrastructure
  - Introduction of workflow automation

- With successful processing of DP0.2, PanDA was endorsed for DRP processing

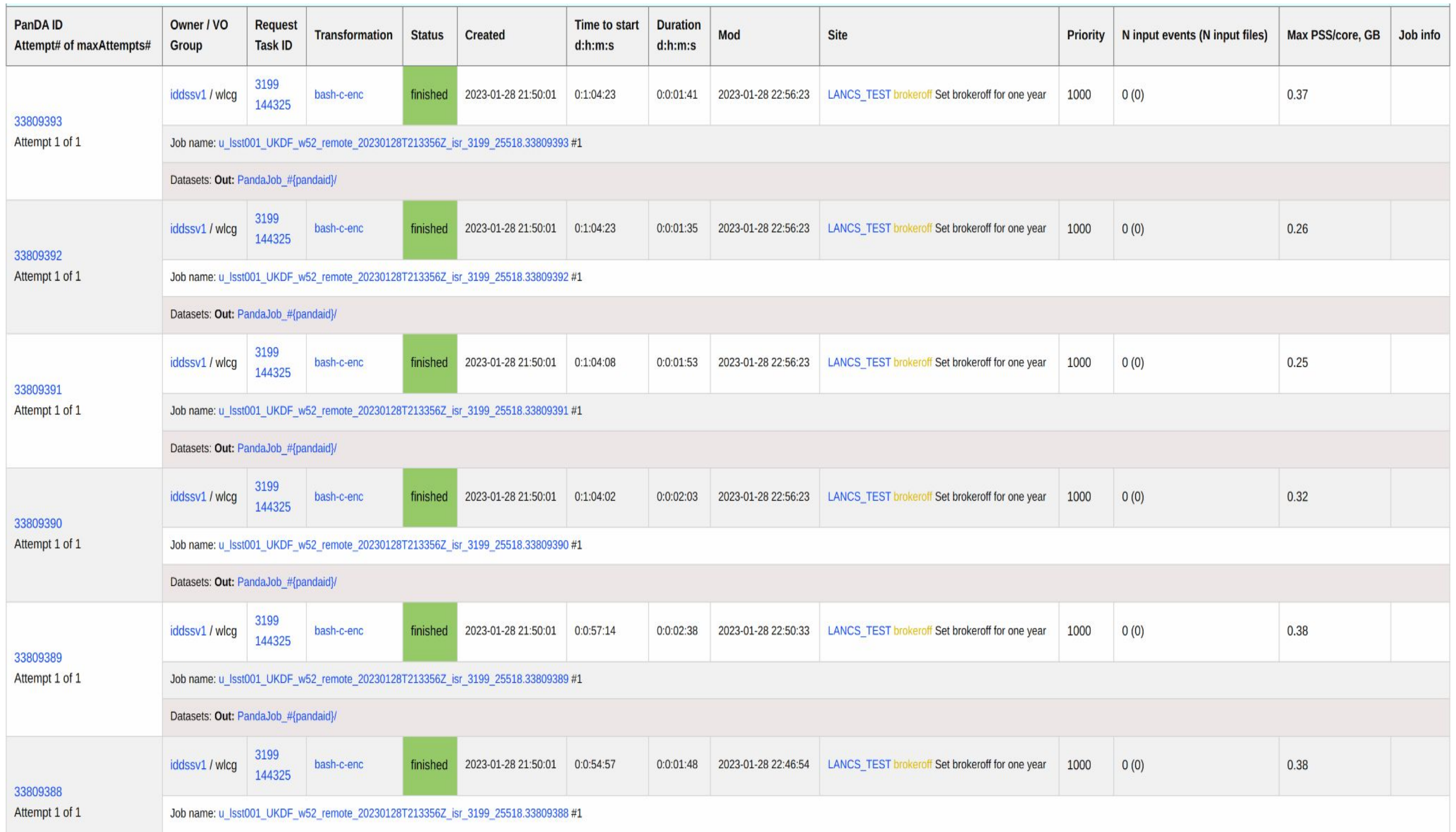
- 2023 DRP estimated to have **~36M** jobs for Public Data Release 2 (HSC) (PDR2), **~8M** for HSC reprocessing

Multi-Site Processing

- Constraints from Butler in order to be able to process Rubin workflows in multi-DF
  - Quantum graph and execution Butler created at one DF not portable to another DF
  - After the processing of all pipeline tasks, need to merge outputs and metadata back to main Butler registry. Current Butler doesn't support this remotely

- The support for multi-DF processing needed development both in Rubin's DM middleware and in PanDA/iDDS

Rubin pipeline jobs submitted remotely



PanDA ID Attempt# of maxAttempts#	Owner / VO Group	Request Task ID	Transformation	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Mod	Site	Priority	N input events (N input files)	Max PSS/core, GB	Job info
33809393 Attempt 1 of 1	iddssv1 / wlcg	3199 144325	bash-c-enc	finished	2023-01-28 21:50:01	0:1:04:23	0:0:01:41	2023-01-28 22:56:23	LANCS_TEST brokeroff Set brokeroff for one year	1000	0 (0)	0.37	
	Job name: u_lsst001_UKDF_w52_remote_20230128T213356Z_isr_3199_25518.33809393 #1												
	Datasets: **Out:** PandaJob_#{pandaid}/												
33809392 Attempt 1 of 1	iddssv1 / wlcg	3199 144325	bash-c-enc	finished	2023-01-28 21:50:01	0:1:04:23	0:0:01:35	2023-01-28 22:56:23	LANCS_TEST brokeroff Set brokeroff for one year	1000	0 (0)	0.26	
	Job name: u_lsst001_UKDF_w52_remote_20230128T213356Z_isr_3199_25518.33809392 #1												
	Datasets: **Out:** PandaJob_#{pandaid}/												
33809391 Attempt 1 of 1	iddssv1 / wlcg	3199 144325	bash-c-enc	finished	2023-01-28 21:50:01	0:1:04:08	0:0:01:53	2023-01-28 22:56:23	LANCS_TEST brokeroff Set brokeroff for one year	1000	0 (0)	0.25	
	Job name: u_lsst001_UKDF_w52_remote_20230128T213356Z_isr_3199_25518.33809391 #1												
	Datasets: **Out:** PandaJob_#{pandaid}/												
33809390 Attempt 1 of 1	iddssv1 / wlcg	3199 144325	bash-c-enc	finished	2023-01-28 21:50:01	0:1:04:02	0:0:02:03	2023-01-28 22:56:23	LANCS_TEST brokeroff Set brokeroff for one year	1000	0 (0)	0.32	
	Job name: u_lsst001_UKDF_w52_remote_20230128T213356Z_isr_3199_25518.33809390 #1												
	Datasets: **Out:** PandaJob_#{pandaid}/												
33809389 Attempt 1 of 1	iddssv1 / wlcg	3199 144325	bash-c-enc	finished	2023-01-28 21:50:01	0:0:57:14	0:0:02:38	2023-01-28 22:50:33	LANCS_TEST brokeroff Set brokeroff for one year	1000	0 (0)	0.38	
	Job name: u_lsst001_UKDF_w52_remote_20230128T213356Z_isr_3199_25518.33809389 #1												
	Datasets: **Out:** PandaJob_#{pandaid}/												
33809388 Attempt 1 of 1	iddssv1 / wlcg	3199 144325	bash-c-enc	finished	2023-01-28 21:50:01	0:0:54:57	0:0:01:48	2023-01-28 22:46:54	LANCS_TEST brokeroff Set brokeroff for one year	1000	0 (0)	0.38	
	Job name: u_lsst001_UKDF_w52_remote_20230128T213356Z_isr_3199_25518.33809388 #1												

PanDA Deployment at SLAC K8s

- Main components are all deployed:
  - PanDA Server and JEDI, Indigo IAM authentication, Harvester, iDDS, PanDA monitor, ActiveMQ

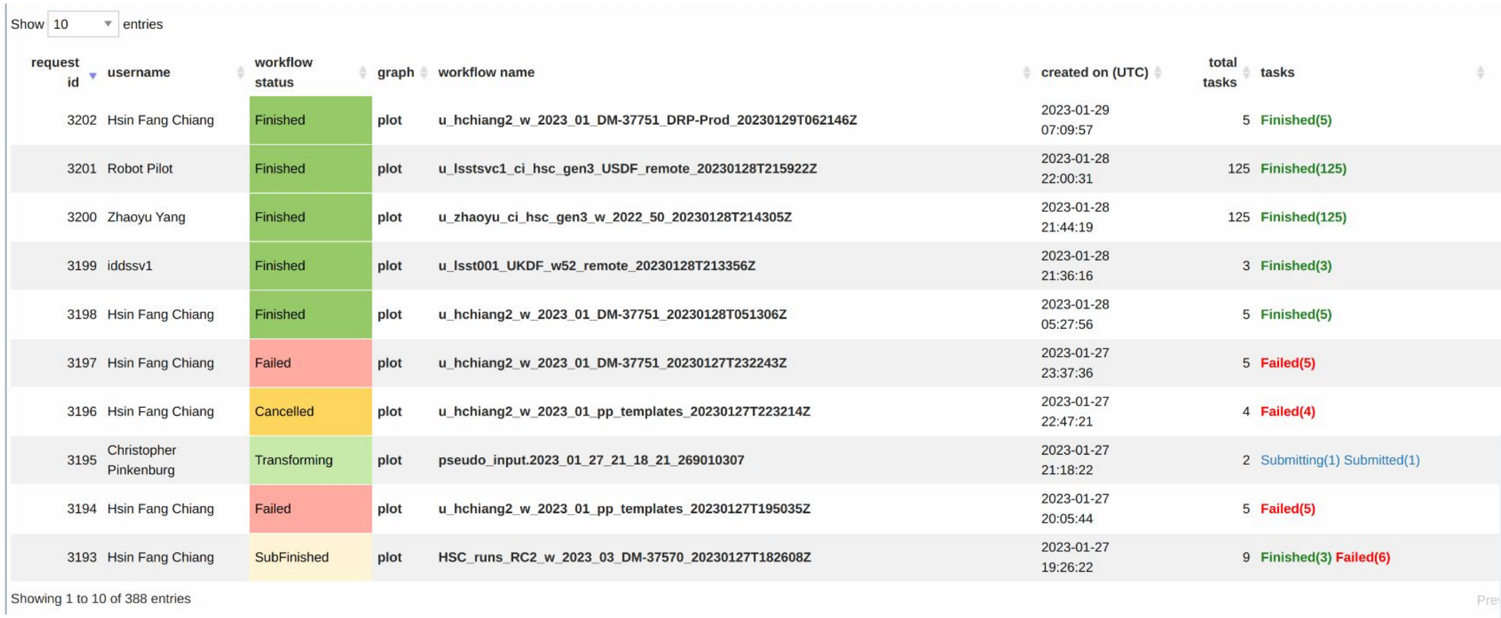
- PostgreSQL
  - DB deployed with CNPG (CloudNativePostgreSQL) for a highly available PostgreSQL DB cluster with a primary/standby architecture
  - Relies on Kubernetes API server to maintain the state of PostgreSQL cluster
  - Provides cloud native capabilities like self-healing, high availability, rolling updates, scale up/down of read-only replicas, resource management, ..

- Restricted network in/out access at SLAC
  - No outbound access to FrDF and UKDF
  - Using squid proxy and NAT

- PanDA monitor available
  - https://rubin-panda-bigmon-dev.slac.stanford.edu:8443 with IAM to support login with institute's credentials

PanDA Monitor Development

- The DOMA instance of the PanDA monitor was developed for Rubin job monitoring
  - DOMA is a CERN/LHC R&D project that offers a playground for non-ATLAS experiments to try PanDA, iDDS,..
- Many features have been added for the Rubin workflow monitoring
  - Hierarchical navigation at different levels: workflow->tasks->jobs->logs
  - Dedicated workflow progress monitoring
  - Memory usage monitoring using prmon (open source tool originally from ATLAS - now HSF)
  - Direct access to the logs from the monitor
- The same monitor is used by all non-ATLAS experiments, e.g. sPHENIX

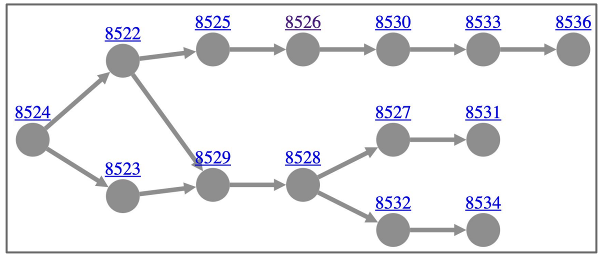


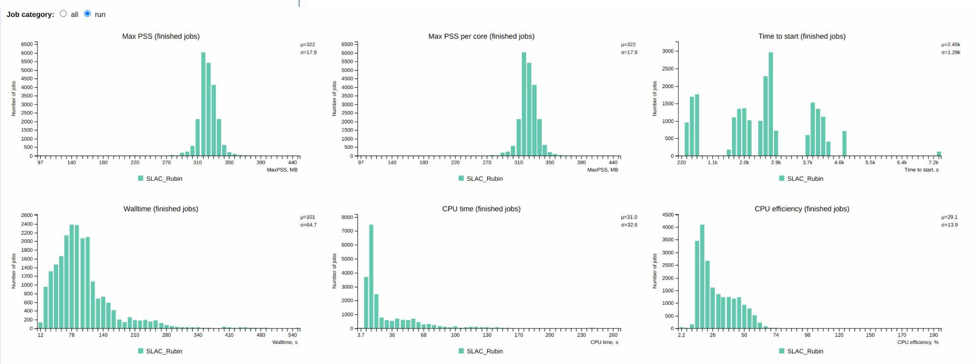
Show 10 entries

Search:

request id	username	workflow status	graph	workflow name	created on (UTC)	total tasks	tasks	remaining files	released files	total files
3202	Hsin Fang Chiang	Finished	plot	u_hchiang2_w_2023_01_DM-37751_DRP-Prod_20230129T062146Z	2023-01-29 07:09:57	5	Finished(5)	0	1298	1298
3201	Robot Pilot	Finished	plot	u_lsstsvc1_ci_hsc_gen3_USDF_remote_20230128T215922Z	2023-01-28 22:00:31	125	Finished(125)	0	1147	1147
3200	Zhaoyu Yang	Finished	plot	u_zhaoyu_ci_hsc_gen3_w_2022_50_20230128T214305Z	2023-01-28 21:44:19	125	Finished(125)	0	1147	1147
3199	iddssv1	Finished	plot	u_lsst001_UKDF_w52_remote_20230128T213356Z	2023-01-28 21:36:16	3	Finished(3)	0	35	35
3198	Hsin Fang Chiang	Finished	plot	u_hchiang2_w_2023_01_DM-37751_20230128T051306Z	2023-01-28 05:27:56	5	Finished(5)	0	21901	21901
3197	Hsin Fang Chiang	Failed	plot	u_hchiang2_w_2023_01_DM-37751_20230127T232243Z	2023-01-27 23:37:36	5	Failed(5)	21899	2	21901
3196	Hsin Fang Chiang	Cancelled	plot	u_hchiang2_w_2023_01_pp_templates_20230127T223214Z	2023-01-27 22:47:21	4	Failed(4)	21899	1	21900
3195	Christopher Pinkenburg	Transforming	plot	pseudo_input.2023_01_27_21_18_21_269010307	2023-01-27 21:18:22	2	Submitting(1) Submitted(1)			
3194	Hsin Fang Chiang	Failed	plot	u_hchiang2_w_2023_01_pp_templates_20230127T195035Z	2023-01-27 20:05:44	5	Failed(5)			
3193	Hsin Fang Chiang	SubFinished	plot	HSC_runs_RC2_w_2023_03_DM-37570_20230127T182608Z	2023-01-27 19:26:22	9	Finished(3) Failed(6)			

Showing 1 to 10 of 388 entries



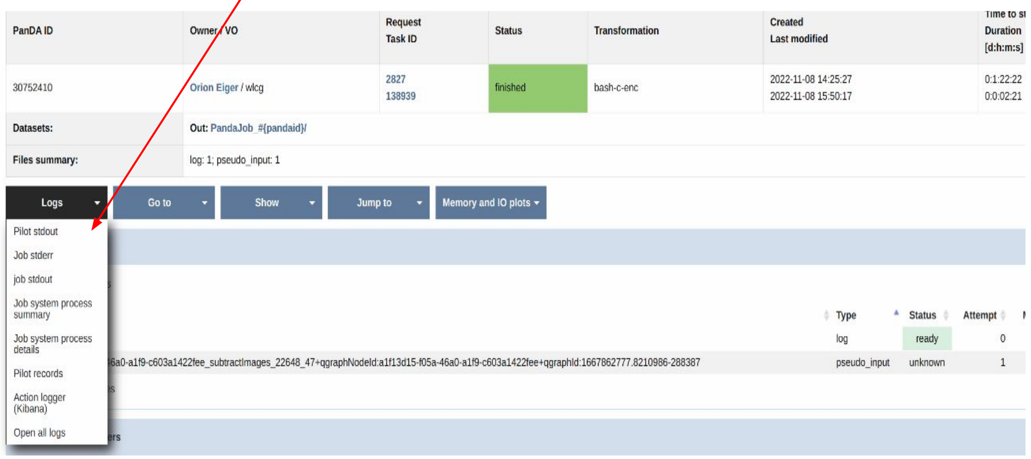


For more details on PanDA Monitor see:
BigPanDA monitoring system evolution in the ATLAS experiment, track 4, 9th May, 14:15.

Near Real-Time Logging

- Conventional log access
  - At end of job execution, pilot uploads the logs including full pilot log, payload stdout and payload stderr dump to the Google cloud (GCS) bucket
- New near real-time log access
  - In Rubin, pilot captures the payload log and sends it as json to **Google Cloud Logging**
  - In addition to the payload logs, pilot sends its own logs to Google Cloud Logging
- Real-time logs provide complementary information for monitoring and debugging
- **Experiment agnostic.** Strong interest from ATLAS. Will be enabled for sPHENIX @ BNL as well

pilot logs uploaded to GCS
(N/A if job is killed)



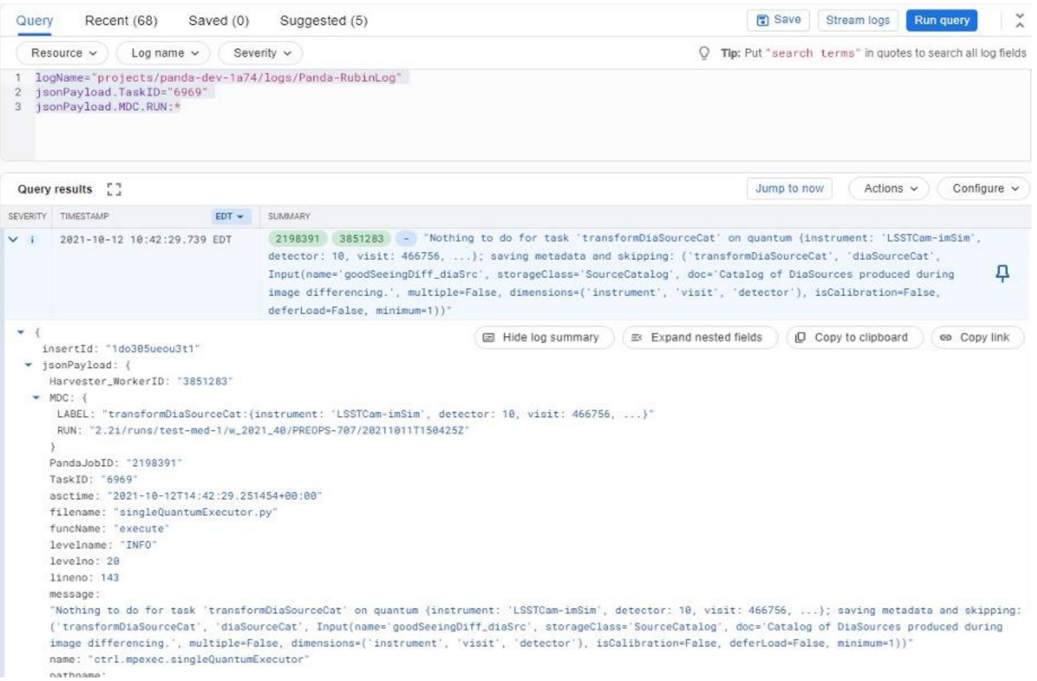
PanDA ID	Owner / VO	Request Task ID	Status	Transformation	Created Last modified	Time to st Duration [d:h:m:s]
30752410	Orion Eiger / wlcg	2827 138939	finished	bash-c-enc	2022-11-08 14:25:27 2022-11-08 15:50:17	0:1:22:22 0:0:02:21

Datasets: Out: PandaJob_#{pandaid}/

Files summary: log: 1; pseudo_input: 1

	Type	Status	Attempt
	log	ready	0
6a0-a1f9-c603a1422fee_subtractImages_22648_47+qgraphNodeId:a1f13d15-f05a-46a0-a1f9-c603a1422fee+qgraphId:1667862777.8210986-288387	pseudo_input	unknown	1

Google Cloud Logging



```
logName="projects/panda-dev-1a74/logs/Panda-RubinLog"
jsonPayload.TaskID="6969"
jsonPayload.MDC.RUN:*
```

```
SEVERITY  TIMESTAMP                      EDT   SUMMARY
   i      2021-10-12 10:42:29.739 EDT          2198391  3851283  -  "Nothing to do for task 'transformDiaSourceCat' on quantum {instrument: 'LSSTCam-imSim',
                                               detector: 10, visit: 466756, ...}; saving metadata and skipping: ('transformDiaSourceCat', 'diaSourceCat',
                                               Input(name='goodSeeingDiff_diaSrc', storageClass='SourceCatalog', doc='Catalog of DiaSources produced during
                                               image differencing.', multiple=False, dimensions=('instrument', 'visit', 'detector'), isCalibration=False,
                                               deferLoad=False, minimum=1))"
{
  insertId: "1do305ueou3t1"
  jsonPayload: {
   Harvester_WorkerID: "3851283"
   MDC: {
    LABEL: "transformDiaSourceCat:{instrument: 'LSSTCam-imSim', detector: 10, visit: 466756, ...}"
    RUN: "2.2i/runs/test-med-1/w_2021_40/PREOPS-707/20211011T150425Z"
   }
   PandaJobID: "2198391"
   TaskID: "6969"
   asctime: "2021-10-12T14:42:29.251454+00:00"
   filename: "singleQuantumExecutor.py"
   funcName: "execute"
   levelname: "INFO"
   levelno: 20
   lineno: 143
   message:
   "Nothing to do for task 'transformDiaSourceCat' on quantum {instrument: 'LSSTCam-imSim', detector: 10, visit: 466756, ...}; saving metadata and skipping:
   ('transformDiaSourceCat', 'diaSourceCat', Input(name='goodSeeingDiff_diaSrc', storageClass='SourceCatalog', doc='Catalog of DiaSources produced during
   image differencing.', multiple=False, dimensions=('instrument', 'visit', 'detector'), isCalibration=False, deferLoad=False, minimum=1))"
   name: "ctrl.mpexec.singleQuantumExecutor"
   pathname:
```

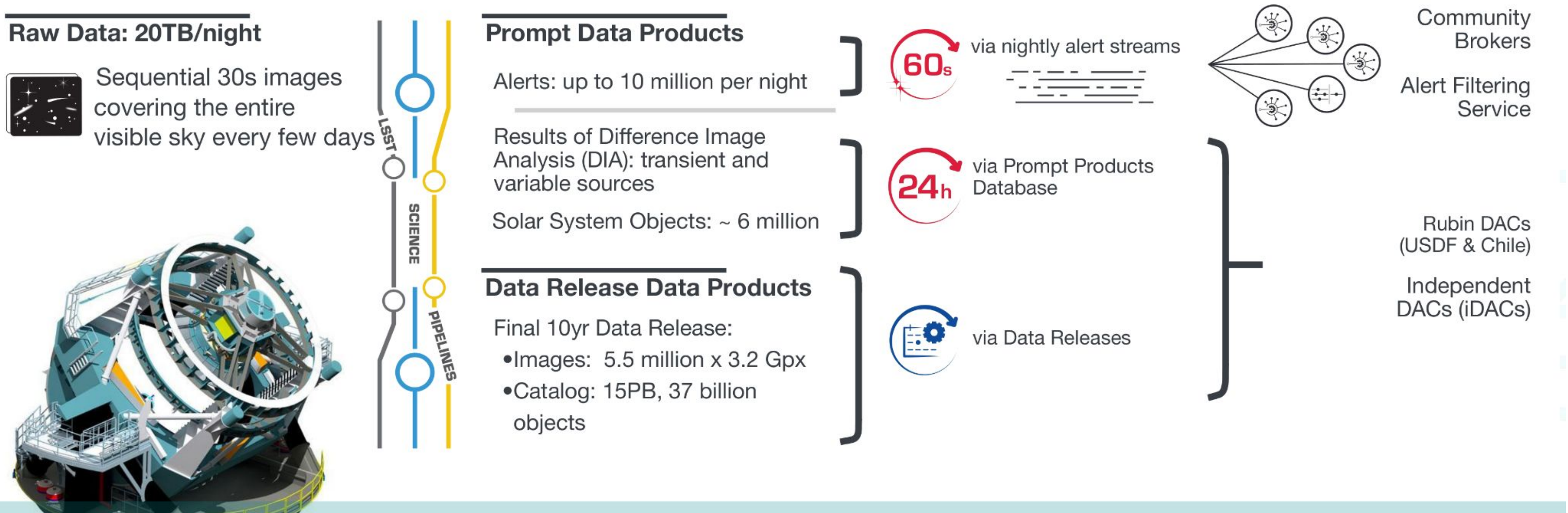
Brookhaven National Laboratory

Prompt Processing

- Prompt processing in Rubin
  - To be able to initiate processing in a few seconds on dedicated resources at SLAC
  - Reuse of WN for each visit to skip downloading calibration data in the processing

- Developments for rapid workload provisioning and processing
  - Semi-persistent pilot up and running on WN
  - Task resurrection via notification to skip overhead before generating jobs
  - Job pushed to the pilot via ActiveMQ
  - Direct communication channel between JEDI and PanDA server

- Mechanism is ready for Rubin to try - also useful to minimize latencies and support pseudo-interactive analysis in ATLAS



Raw Data: 20TB/night
Sequential 30s images covering the entire visible sky every few days
LSST
SCIENCE
PIPELINES
Prompt Data Products
Alerts: up to 10 million per night
Results of Difference Image Analysis (DIA): transient and variable sources
Solar System Objects: ~ 6 million
Data Release Data Products
Final 10yr Data Release:
•Images: 5.5 million x 3.2 Gpx
•Catalog: 15PB, 37 billion objects
60s
via nightly alert streams
24h
via Prompt Products Database
via Data Releases
Community Brokers
Alert Filtering Service
Rubin DACs (USDF & Chile)
Independent DACs (iDACs)

Summary

- PanDA has been endorsed for Rubin Data Release Production processing. The production processing loads will increase steadily

- Current production uses the DOMA PanDA@CERN - deployment of PanDA at SLAC K8s recently completed
  - PanDA@SLAC configuration very similar to PanDA@BNL

- Many new developments for Rubin - applicable to ATLAS and sPHENIX as well:
  - Near real-time logging sends both payload logs and pilot logs to Google Cloud Logging
  - PanDA monitor has been further improved to meet Rubin needs
  - Containerization of PanDA components and helm based deployments
  - Improved PostgreSQL support
  - Prompt processing mechanism for Rubin
  - Support for Multi-DF processing
  - Clustering of pipeline tasks

Resources

- PanDA for Rubin manual: https://panda.lsst.io
- PanDA monitoring for Rubin: https://panda-doma.cern.ch
  https://rubin-panda-bigmon-dev.slac.stanford.edu:8443
- Slack channels
  - Rubin users support
  - Rubin PanDA development

- PanDA docs: https://panda-wms.readthedocs.io/en/latest/
- iDDS docs: https://idds.readthedocs.io/en/latest/
- Harvester docs: https://github.com/HSF/harvester/wiki
- Pilot docs: https://github.com/PanDAWMS/pilot3/wiki

Questions?

PanDA