BigPanDA project and non-LHC experiments

Pavlo Svirin, Sergey Panitkin

non-LHC experiments overview

- ~9 experiments that were tested with PanDA
    - several more that discussed possible collaboration
- various PanDA Servers used: EC2 instance and private instances
- Payloads tested on Grid (LSST/DESC) and supercomputers (Titan/Summit/NERSC)
- Payloads were run in allocations and backfill modes
- Study performed on third-party tools was required in some cases (checkpoint/restore tools LSST/DESC payloads on Titan in backfill mode)
- Study performed on defining a subset of job control tools and ways to describe jobs, which would be easier than ones for ATLAS

IceCube

- Job submission and execution chain has been set up and tested on Titan
- Many IceCube payloads require 2+ hours
  - will use either their allocation on Summit or backfill (if any?)

nEDM

- Job submission and execution chain has been set up and tested on Titan

- Short payloads had been tested in January, 2018

- In January 2019 Leah Broussard asked for instructions on how to install an instance of PanDA Server

    - instructions written by Ruslan were provided

DUNE

- Job submission and execution chain has been set up and tested on Summit
- Sergey Panitkin managed to configure a virtual environment for DUNE ML payloads and managed to run it with PanDA

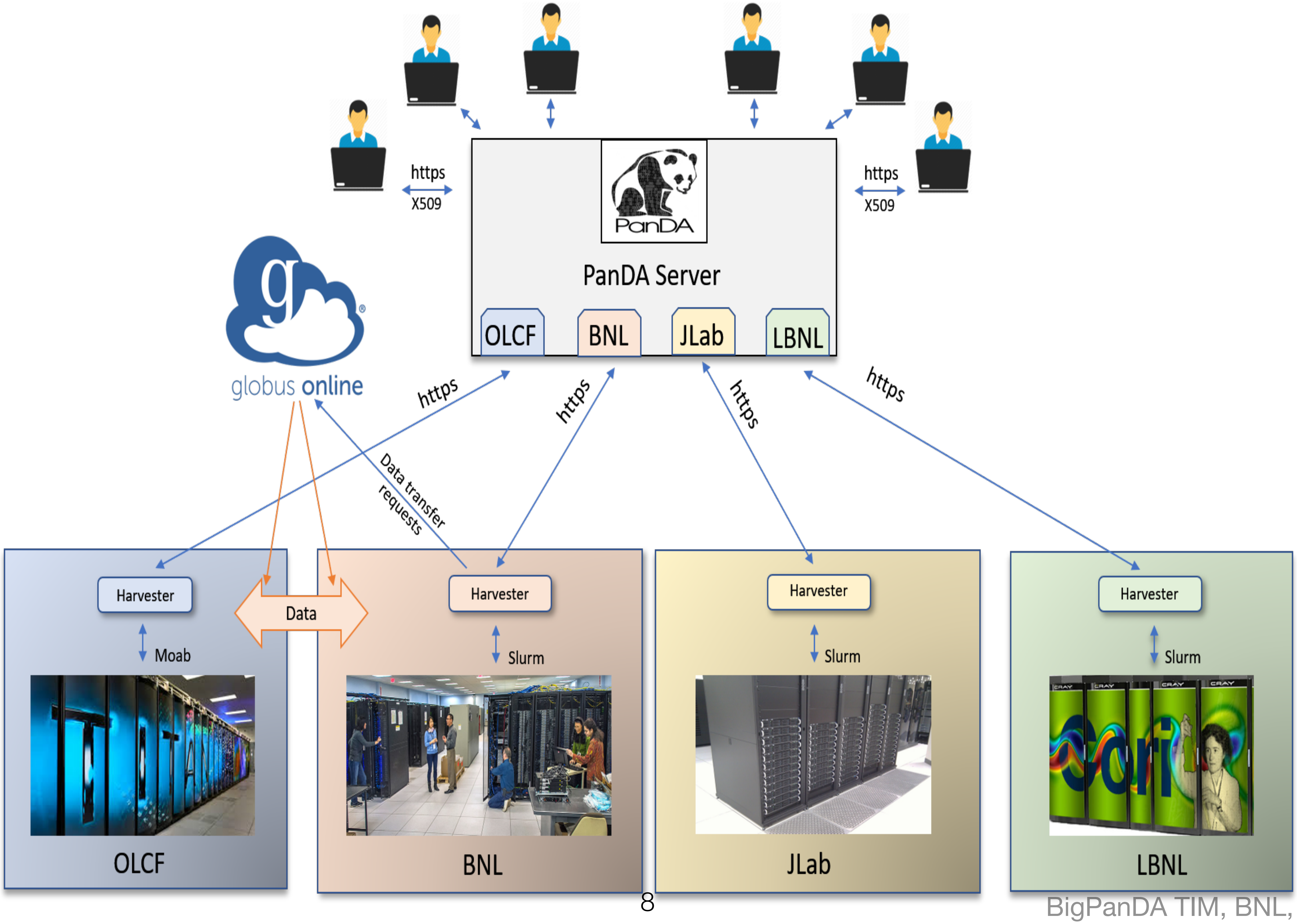
LSST/DESC

- Grid environment with ~30 computing endpoint was created and tested using job submission via PanDA Server
- LSST/DESC DC1 payloads were tested on Titan
  - allocation or checkpoint/restore functionality or software will be needed to run in backfill mode
- unfortunately, we did not receive data to test 2-point analysis workflow
- in November DESC informed about lack of computing resources
  - Bellarmine reacted and offered to use their site

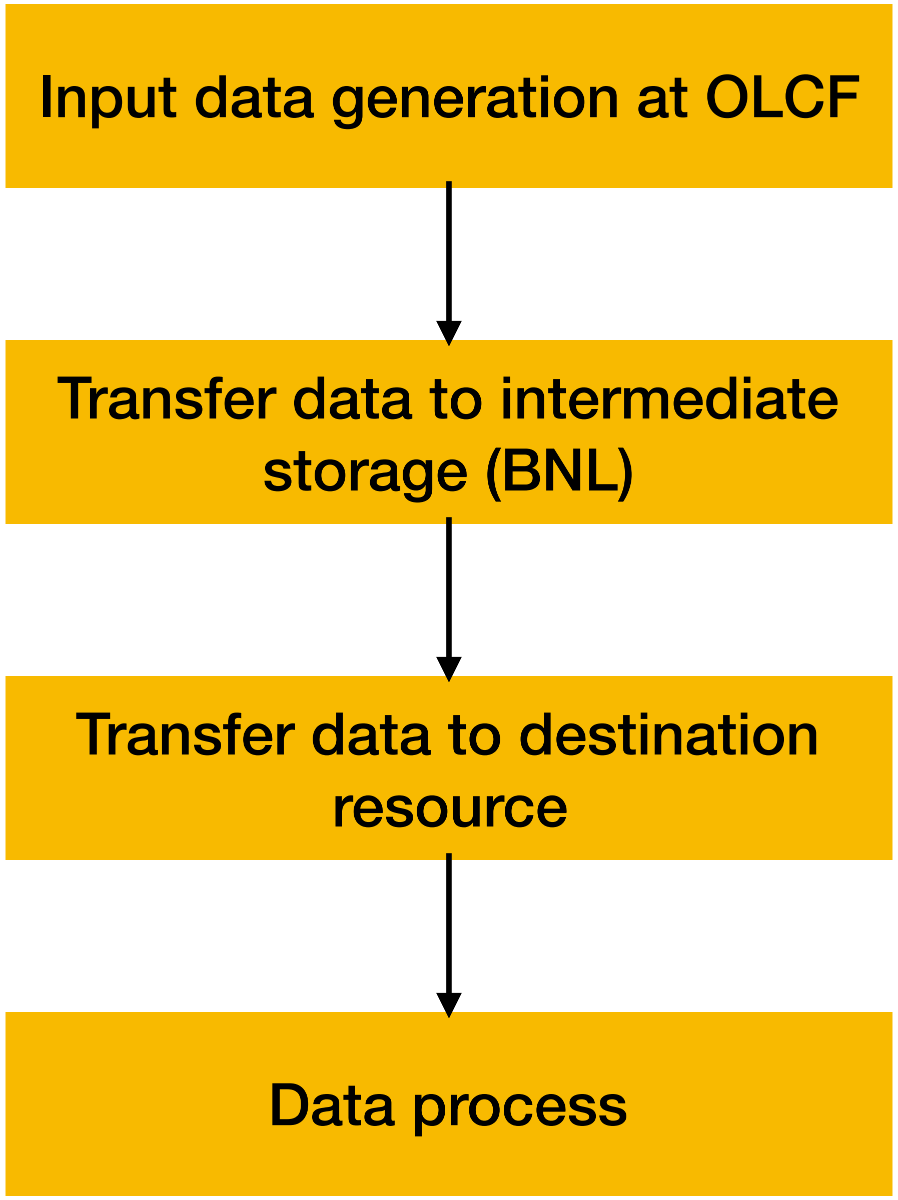
LQCD

- BNL group:
  - successful tests on Titan in November 2017
    - GPU-based, multi-node, occupying ~8000 nodes per job, ~20 hours per each job, independent jobs
    - this was a starting point for a review and redesign of job description and job control tools
  - Production in April/June 2018 on BNL IC
    - GPU-based, single-node, ~12 hours each, simple workflows
    - 13 TB of input data processed, 176 GB of output data, 15,000 GPU hours consumed
- JLab: first third-party testers for Harvester installation tools, job submission/control tools, YAML job description
  - extended YAML description with their parameters and created their own custom submitter module

Future computing environment for LQCD



LQCD: data transfer from OLCF


Input data generation at OLCF

Transfer data to intermediate storage (BNL)

Transfer data to destination resource

Data process

Transfer jobs will have no actual payload (`date` ?)

Pegasus

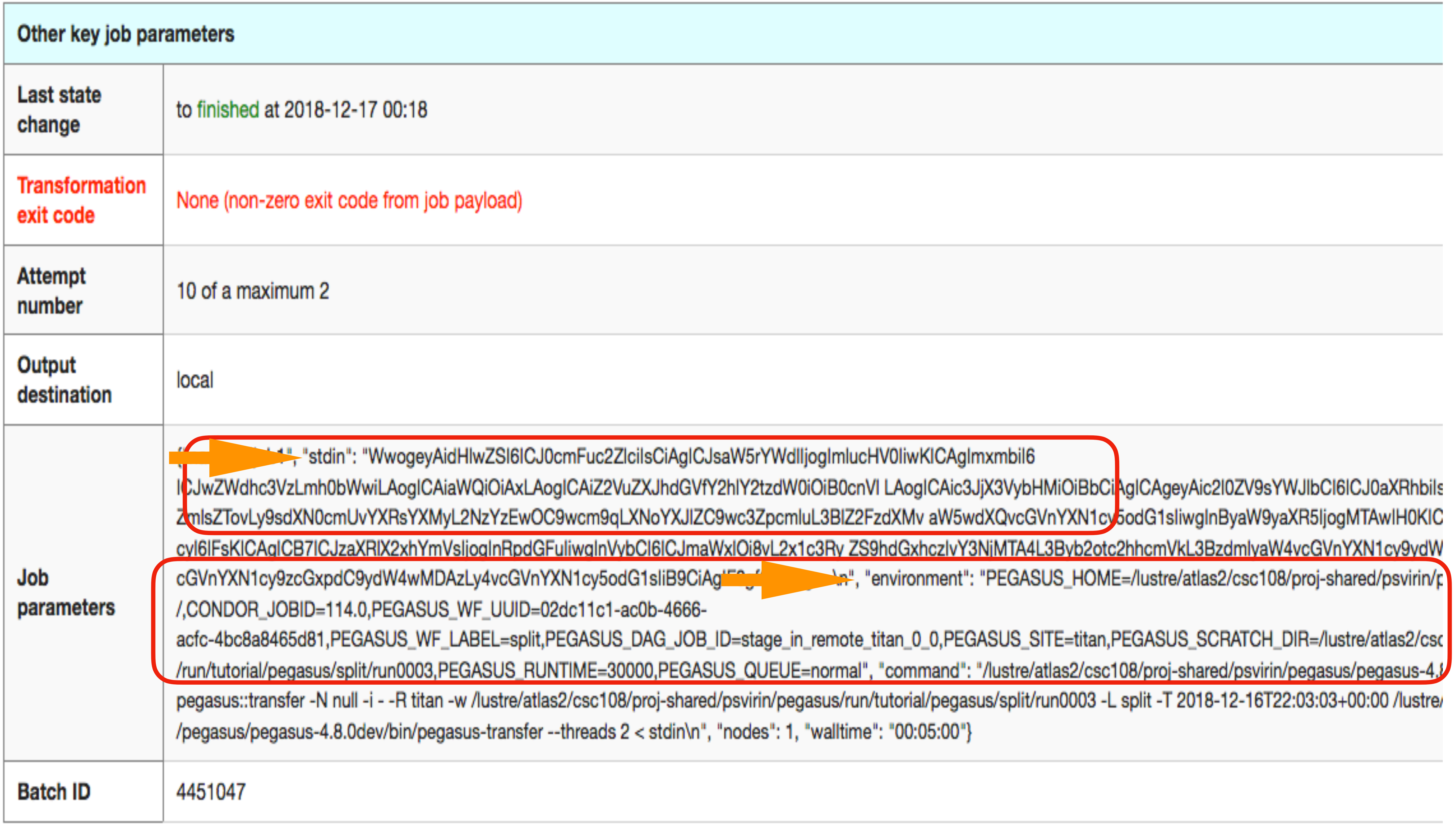
- a collaboration between Pegasus and PanDA teams started in September 2018

- a kick-off meeting was held in October 2018 at USC

- regular meetings on BlueJeans were established for discussions of the ways to integrate these systems

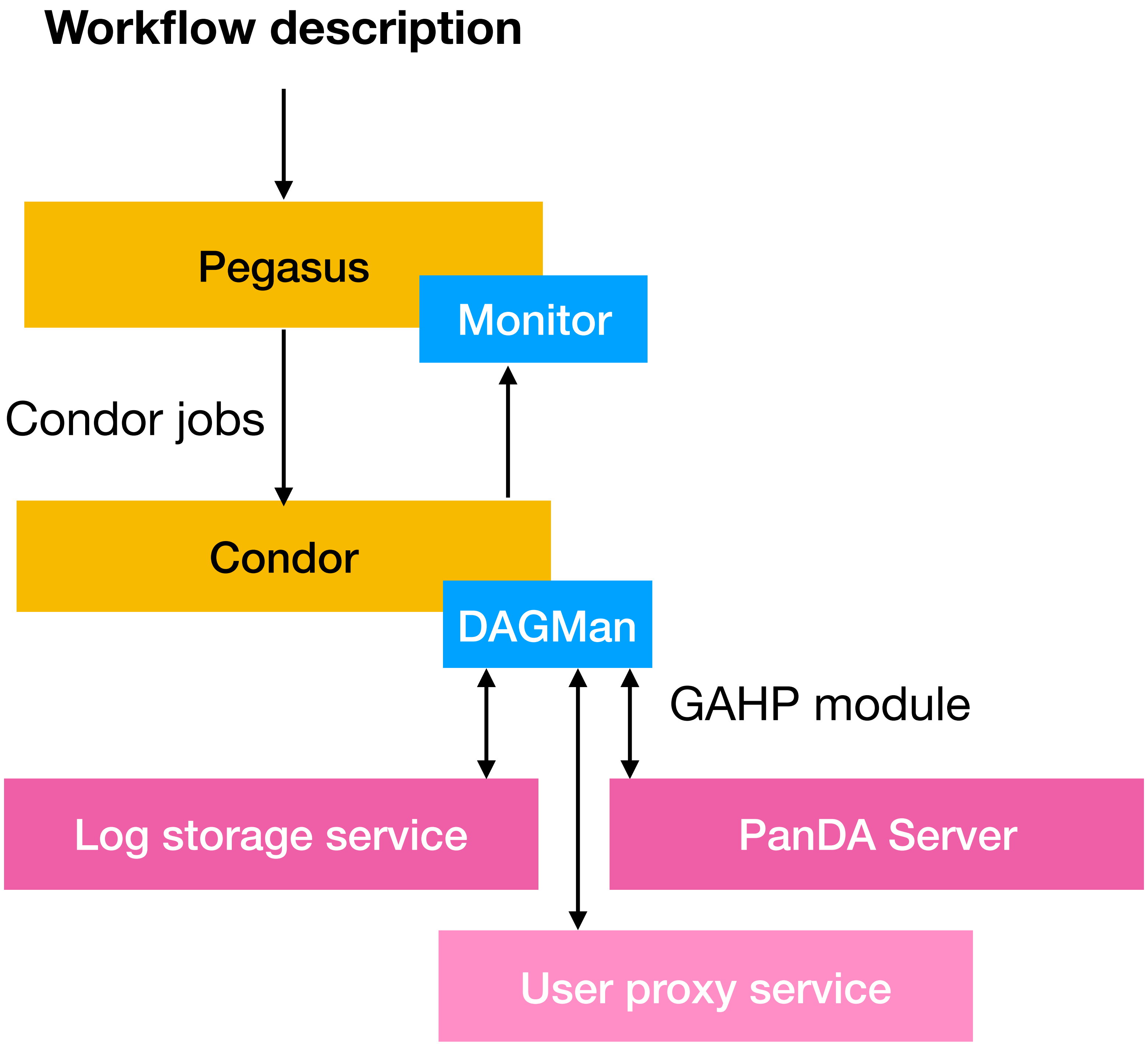
- The first prototype was announced working in December 2018

PanDA job submitted via Pegasus

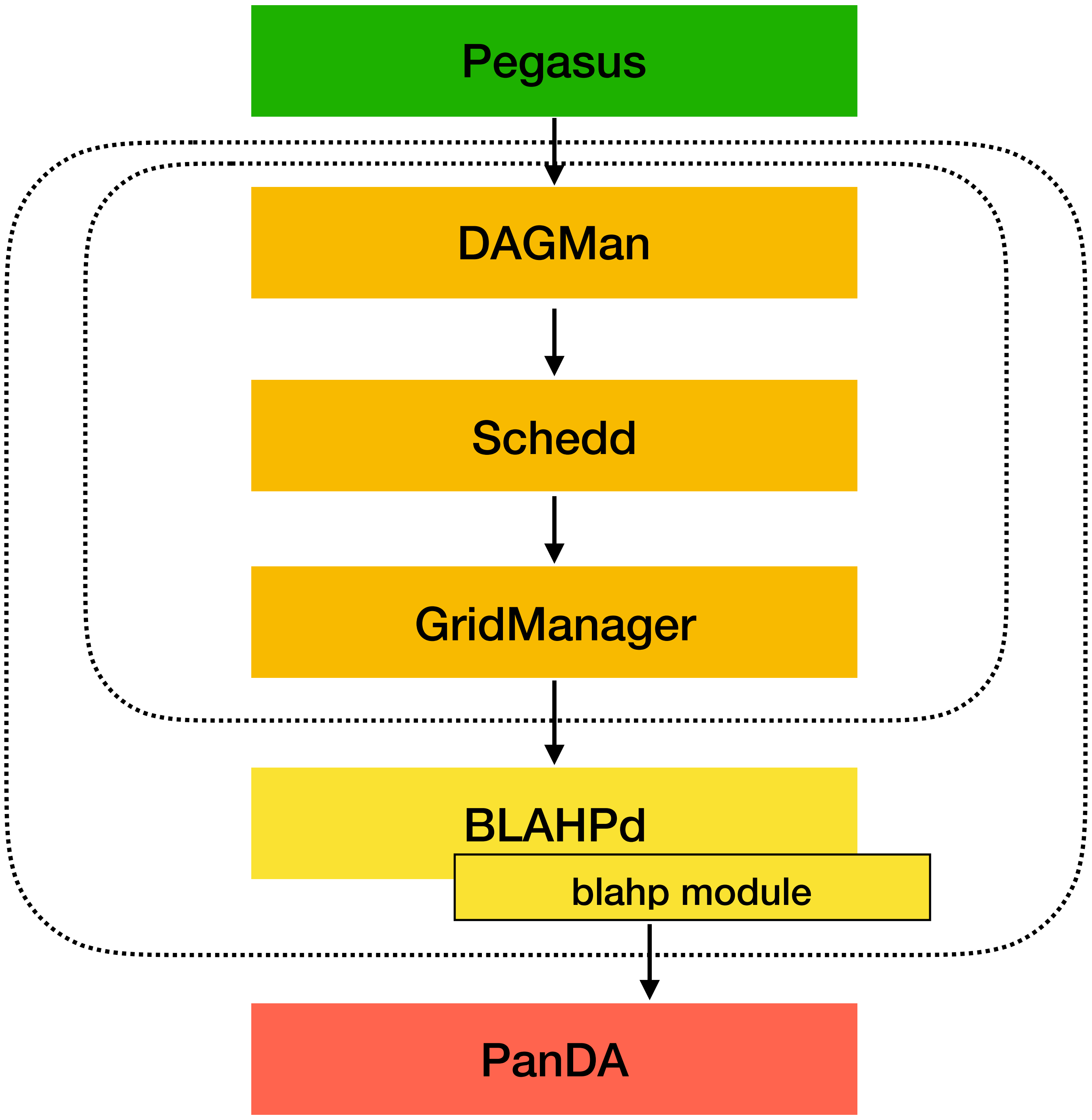


Other key job parameters	
Last state change	to finished at 2018-12-17 00:18
Transformation exit code	None (non-zero exit code from job payload)
Attempt number	10 of a maximum 2
Output destination	local
Job parameters	{[illegible]1", "stdin": "WwogeyAidHlwZSI6ICJ0cmFuc2ZlcilsCiAgICJsaW5rYWdlljogImlucHV0liwKICAgImxmbil6 ICJwZWdhc3VzLmh0bWwiLAogICAiaWQiOiAxLAogICAiZ2VuZXJhdGVfY2hlY2tzdW0iOiB0cnVl LAogICAic3JjX3VybHMiOiBbCiAgICAgeyAic2l0ZV9sYWJlbCI6ICJ0aXRhbils ZmlsZTovLy9sdXN0cmUvYXRsYXMyL2NzYzEwOC9wcm9qLXNoYXJlZC9wc3ZpcmluL3BlZ2FzdXMv aW5wdXQvcGVnYXN1cy5odG1sliwgInByaW9yaXR5IjogMTAwIH0KIC cyl6IFsKICAgICB7ICJzaXRlX2xhYmVsIjogInRpdGFuIiwgInVybCI6ICJmaWxlOi8vL2x1c3Ry ZS9hdGxhczlvY3NjMTA4L3Byb2otc2hhcmVkL3BzdmlyaW4vcGVnYXN1cy9ydW cGVnYXN1cy9zcGxpdC9ydW4wMDAzLy4vcGVnYXN1cy5odG1sliB9CiAg[illegible]\n", "environment": "PEGASUS_HOME=/lustre/atlas2/csc108/proj-shared/psvirin/p /,CONDOR_JOBID=114.0,PEGASUS_WF_UUID=02dc11c1-ac0b-4666-acfc-4bc8a8465d81,PEGASUS_WF_LABEL=split,PEGASUS_DAG_JOB_ID=stage_in_remote_titan_0_0,PEGASUS_SITE=titan,PEGASUS_SCRATCH_DIR=/lustre/atlas2/csc /run/tutorial/pegasus/split/run0003,PEGASUS_RUNTIME=30000,PEGASUS_QUEUE=normal", "command": "/lustre/atlas2/csc108/proj-shared/psvirin/pegasus/pegasus-4.8 pegasus::transfer -N null -i - -R titan -w /lustre/atlas2/csc108/proj-shared/psvirin/pegasus/run/tutorial/pegasus/split/run0003 -L split -T 2018-12-16T22:03:03+00:00 /lustre/ /pegasus/pegasus-4.8.0dev/bin/pegasus-transfer --threads 2 < stdin\n", "nodes": 1, "walltime": "00:05:00"}
Batch ID	4451047

Services


Workflow description
Pegasus
Monitor
Condor jobs
Condor
DAGMan
GAHP module
Log storage service
PanDA Server
User proxy service

Components

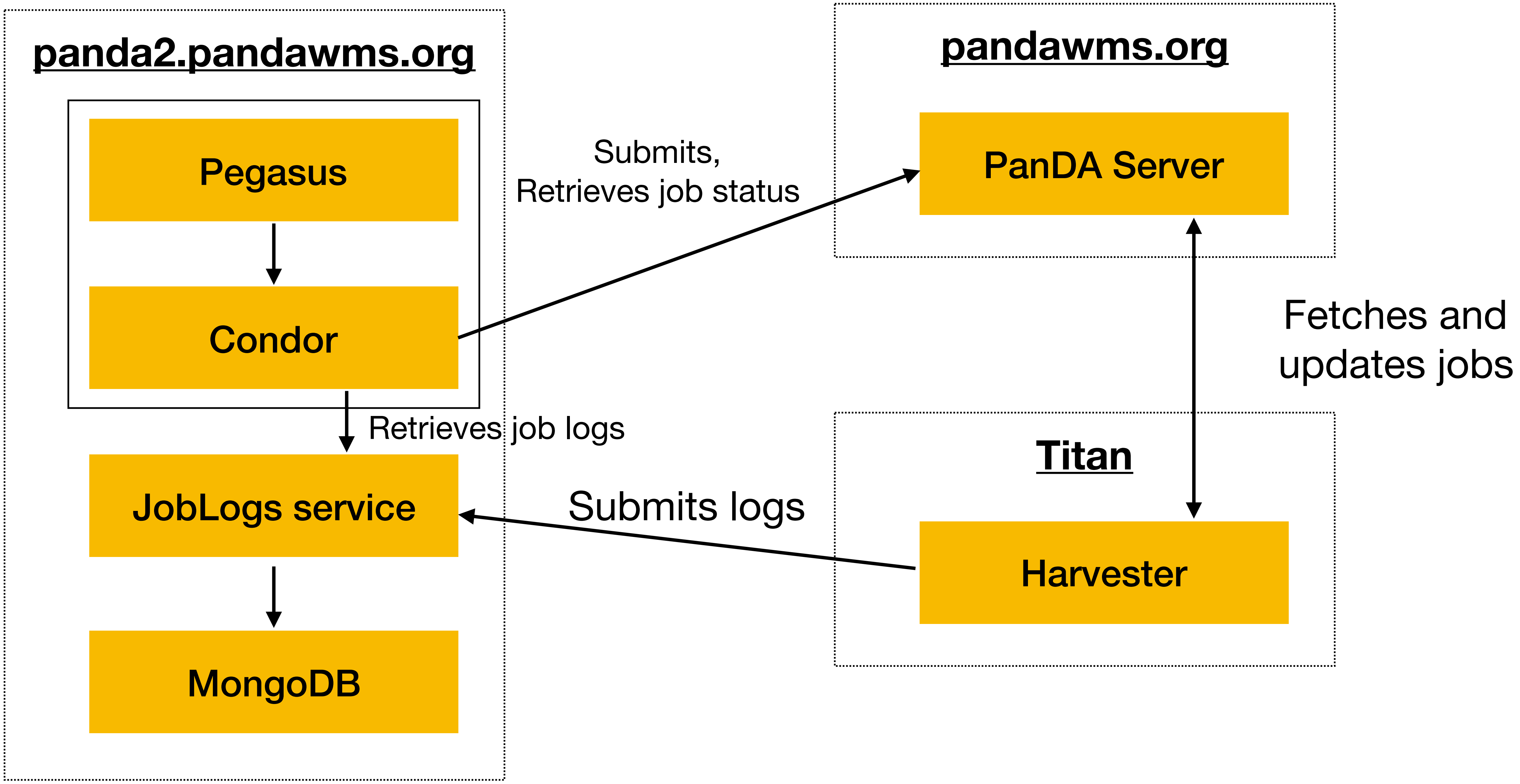


The object of the Grid ASCII Helper Protocol (GAHP) is to allow the use of the client library or package of a grid or cloud service via a simple ASCII-based protocol. A process which implements GAHP is referred to as a GAHP server. GAHP is designed to handle both synchronous (blocking) and asynchronous (non-blocking) calls.

There are several GAHP protocols available, each designed to communicate with a specific remote job submission system. At the time of this writing they are:

- **Globus 2 GAHP** (the original GAHP).
- **Globus 3.2 GAHP**
- **Globus 4 GAHP**. *Spec coming soon*
- **C-GAHP** (HTCondor GAHP).
- **BLAHP** (from INFN)
- **EC2 GAHP**
- **GCE GAHP**
- **Azure GAHP**

Deployment

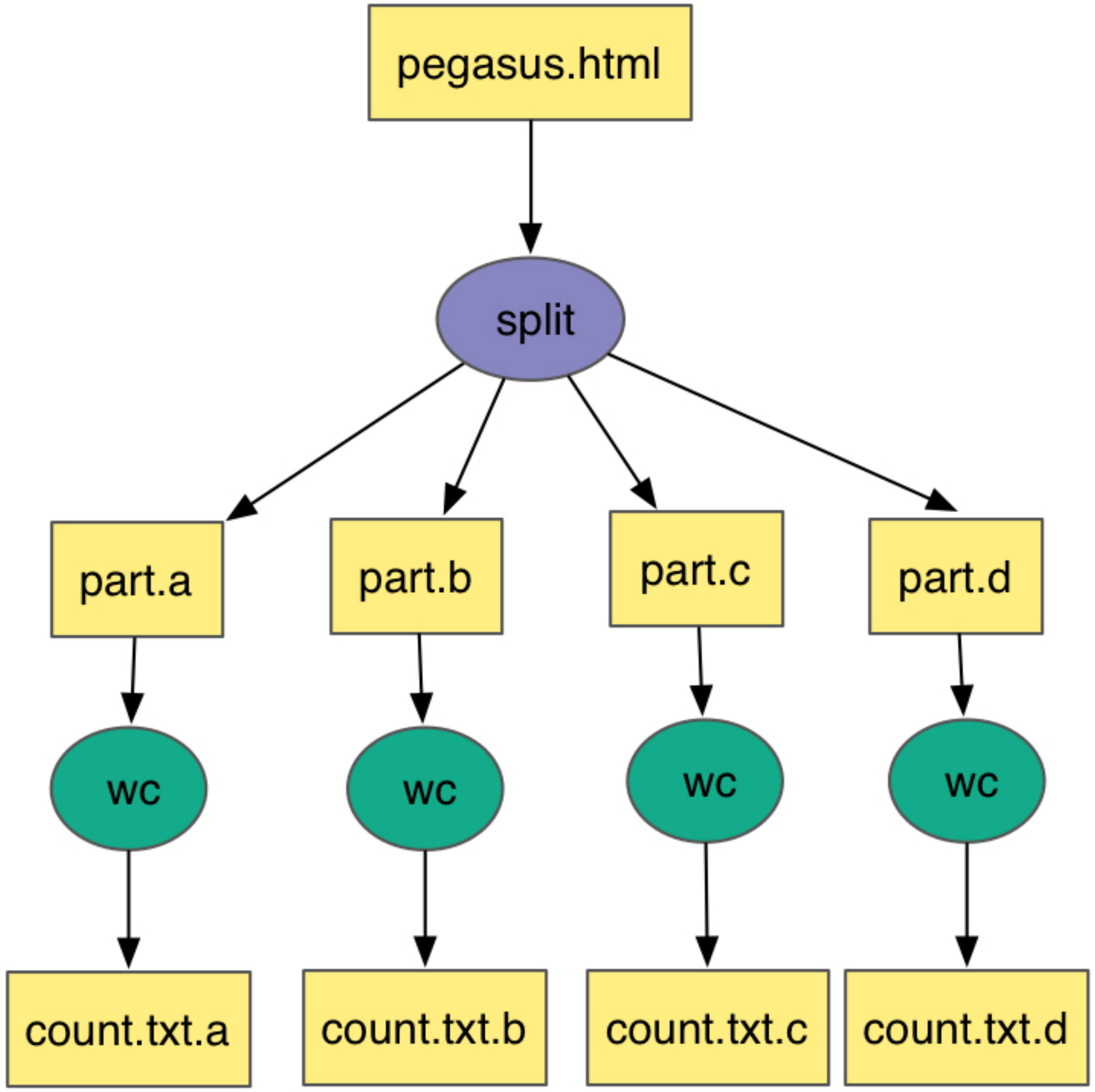

panda2.pandawms.org
Pegasus
Condor
JobLogs service
MongoDB
Submits,
Retrieves job status
Retrieves job logs
pandawms.org
PanDA Server
Fetches and
updates jobs
Titan
Submits logs
Harvester

WORKFLOWS

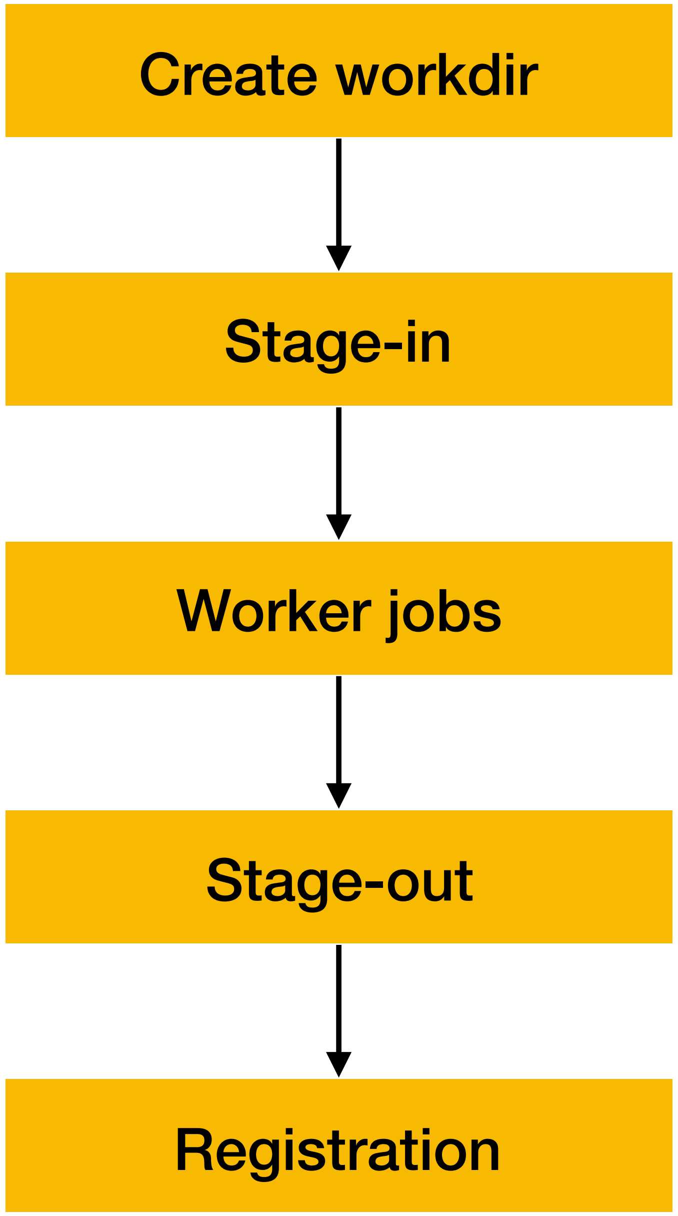
Workflow tested

- Genomics example: did not work out

- Split example was tested with PanDA



Workflow: physical layout and shared disk


Create workdir

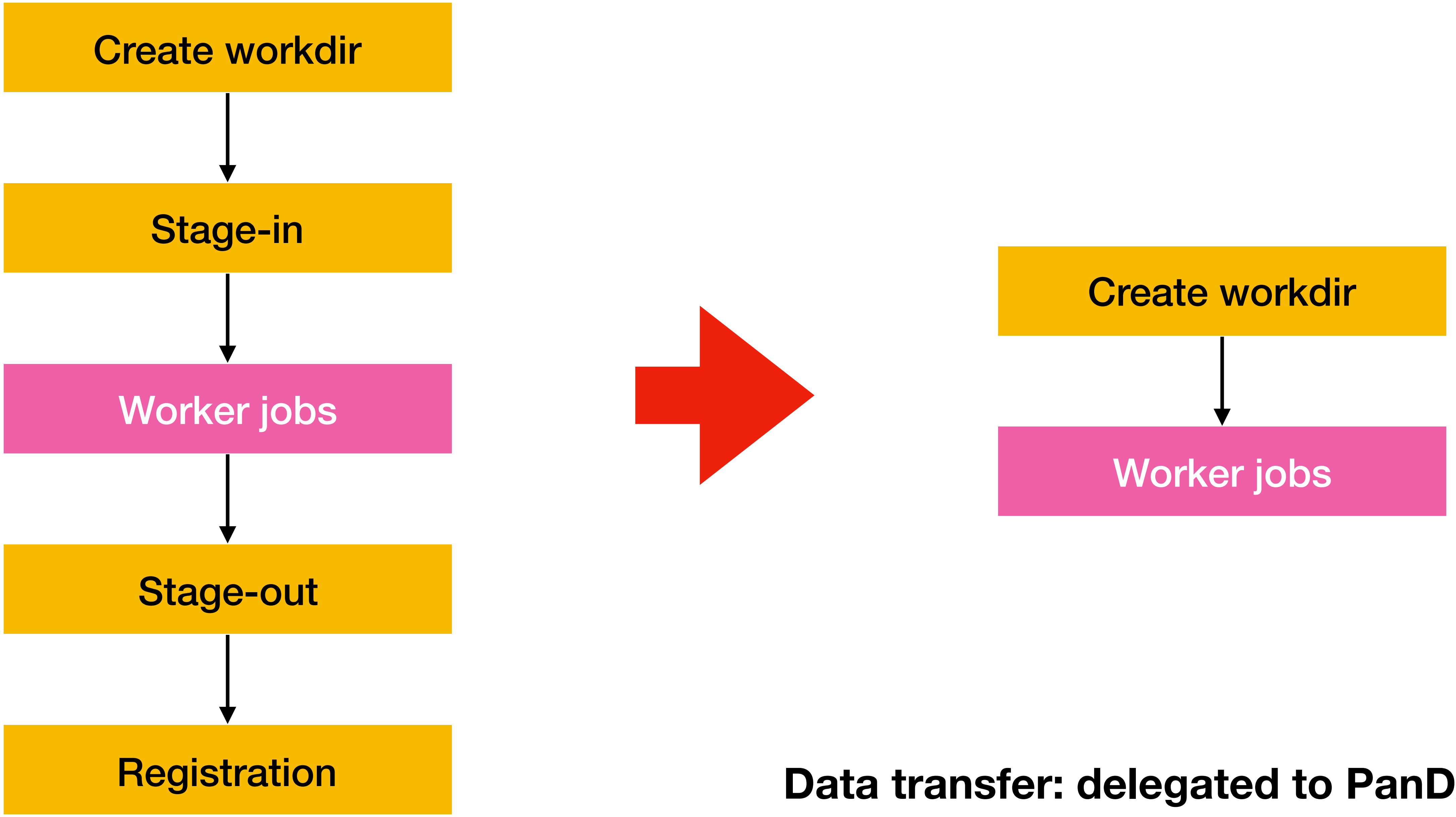
Stage-in

Worker jobs

Stage-out

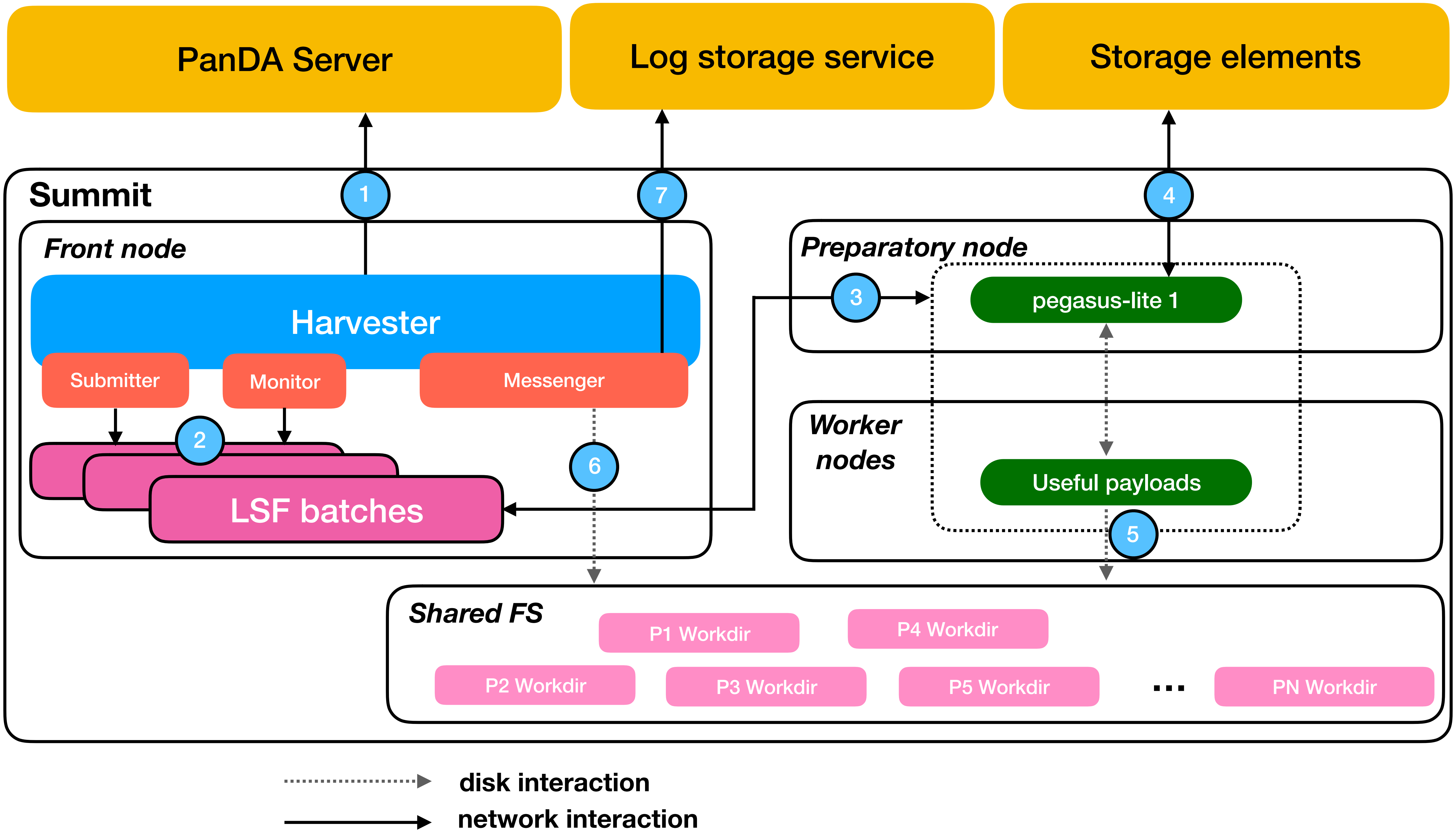
Registration

Workflow: physical layout and shared disk

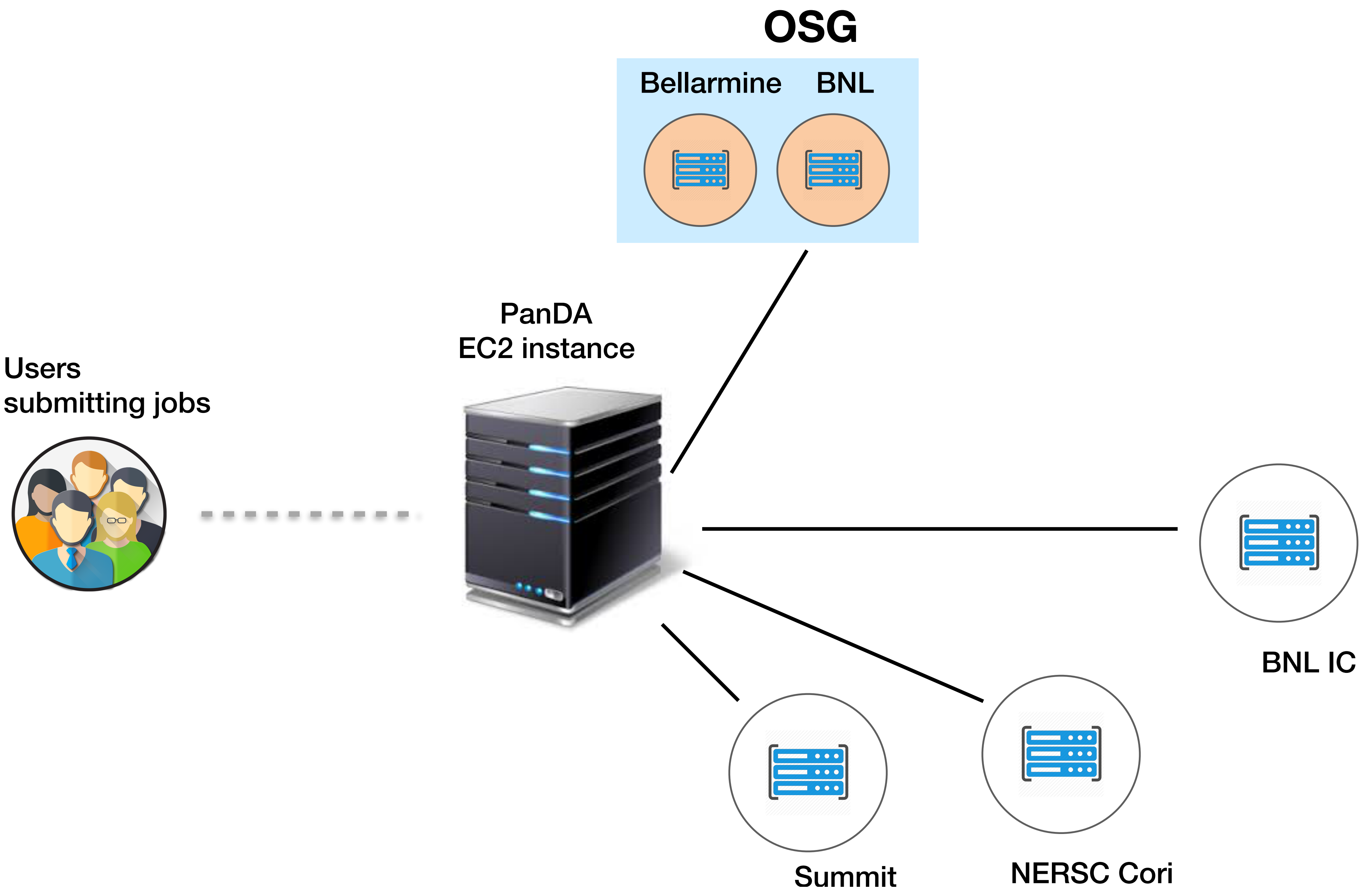


Data transfer: delegated to PanDA

Workflow: physical layout and non-shared disk

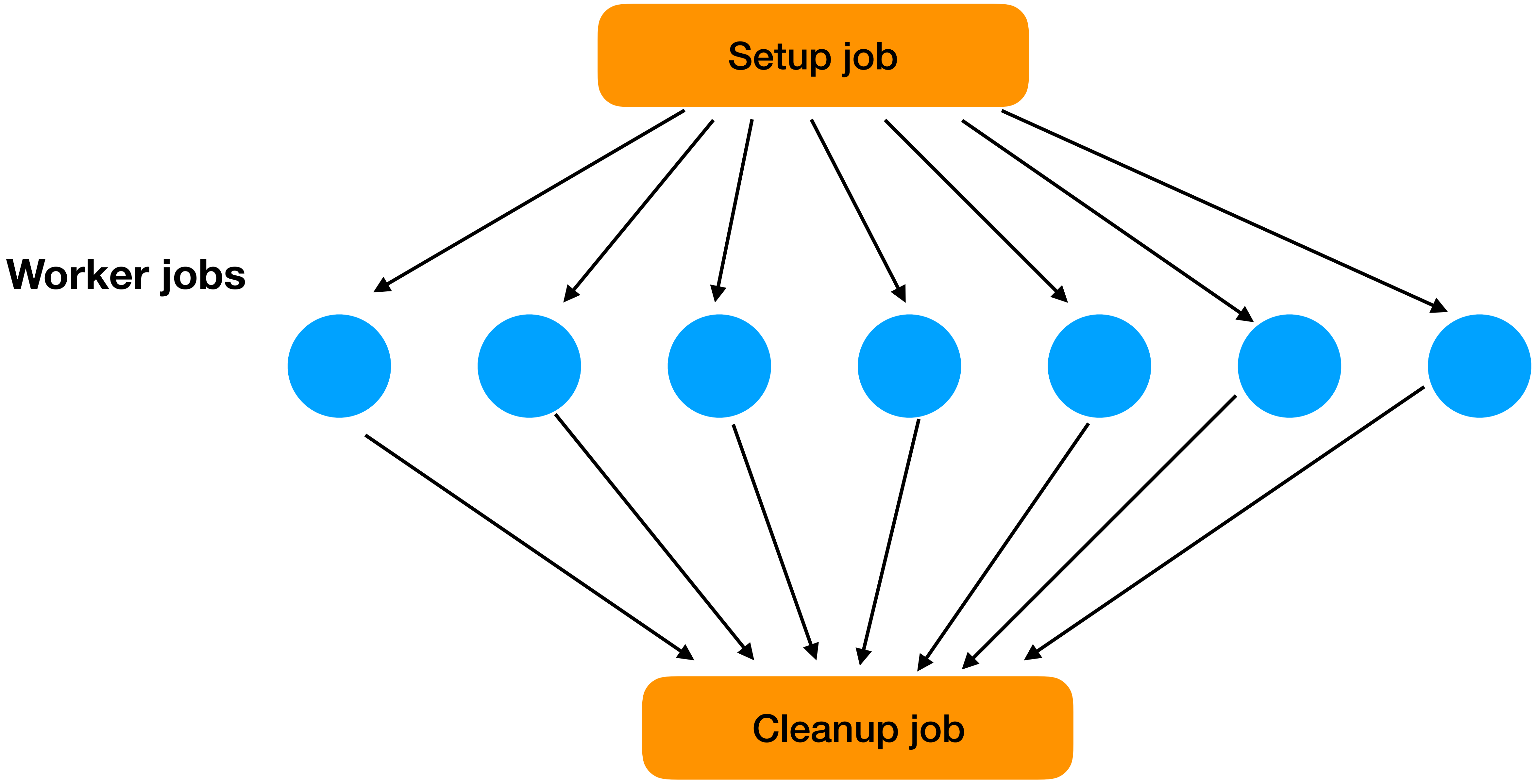


Target computing environment



With data exchange over RUCIO

LQCD Production at BNL: workflow



Setup job

Worker jobs

Cleanup job

Two-level payload management:

- **per workflow**
- **per worker job (resubmit individual worker if it fails, cleanup can be called even if workers fail several times)**

Conclusions

- Achievements:
    - YAML job descriptions & simple job submission/control tools
    - JSON
    - Harvester working with other experiments
        - also got a simple installer
    - PanDA Server upgrade done
        - upgrade to the latest version is needed to work with up-to-date version of Harvester
- Articles/conferences/meetings:
    - achievements with LSST/DESC and LQCD were presented at CHEP2018 in Sofia as a talk and a poster
    - article with these and other experiments were accepted for publication in CHEP2018 Proceedings
    - information about PanDA/Pegasus interaction was passed to Pegasus team for the poster at OLCF Users Meeting

… and future work

- We got instructions on how to install PanDA Server not only on bare metal but in a container (thanks to Ruslan)
  - however, some steps are difficult (e.g.: receiving certificates, CA creation)
- We're missing graphical UIs for administration
  - e.g.: for retry module
- Brokerage module seems to be too ATLAS-specific and hardcoded
- Pegasus and its interaction with PanDA Server/JEDI/NGE/etc.
  - a service which will store user proxies will be needed in order to do stage-in/out for user data
- PostgreSQL: schema ported, data ported, to study whether there are any benefits from this RDBMS