Argonne NATIONAL LABORATORY

An Edge Service for Managing HPC Workflows

J. Taylor Childers

with Thomas D. Uram (ANL)

Douglas P. Benjamin (Duke Univ.)

Thomas J. LeCompte (ANL)

Michael E. Papka (ANL)

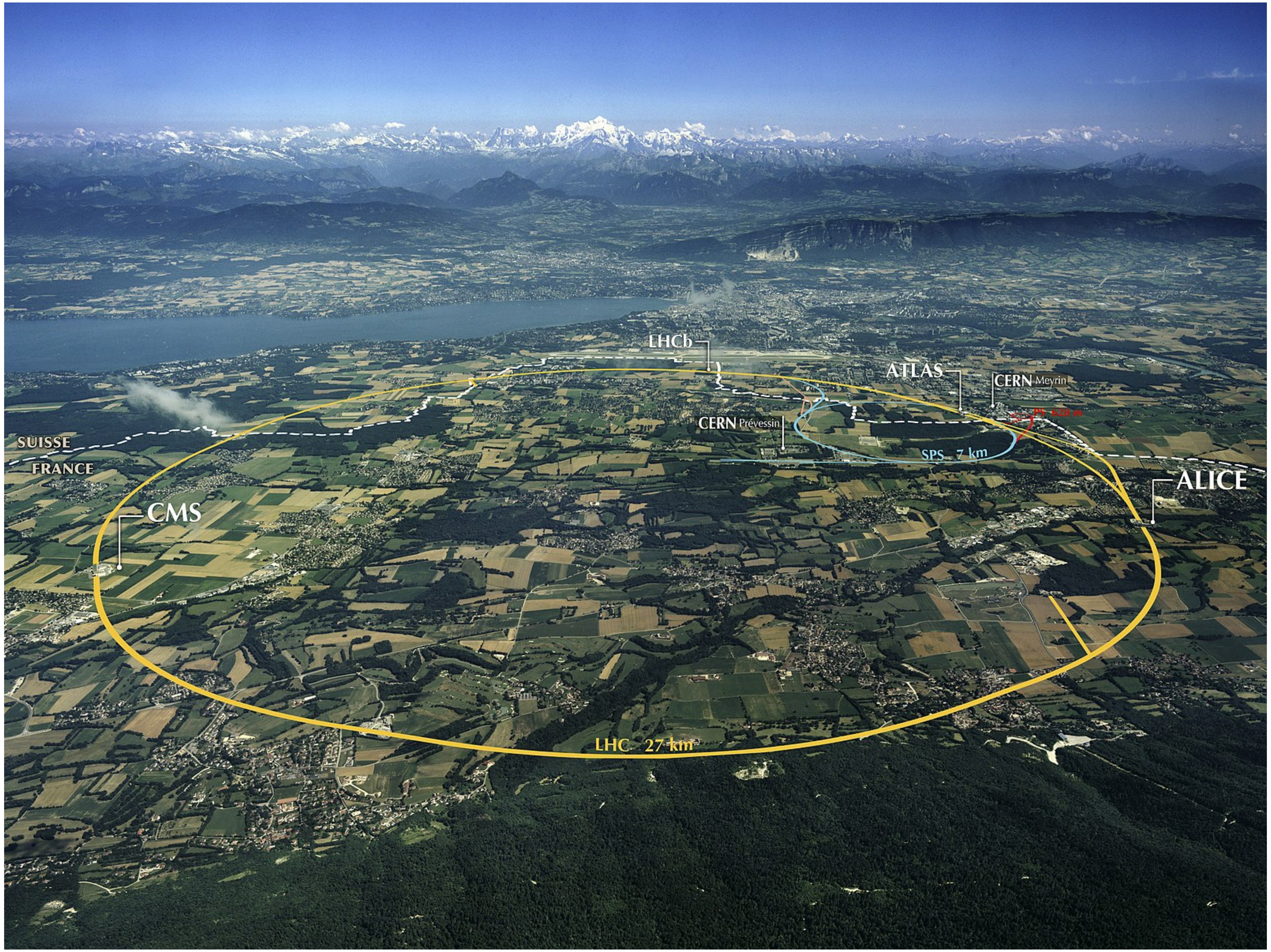
SC17 Denver, CO hpc connects.

HUST 2017

U.S. DEPARTMENT OF ENERGY

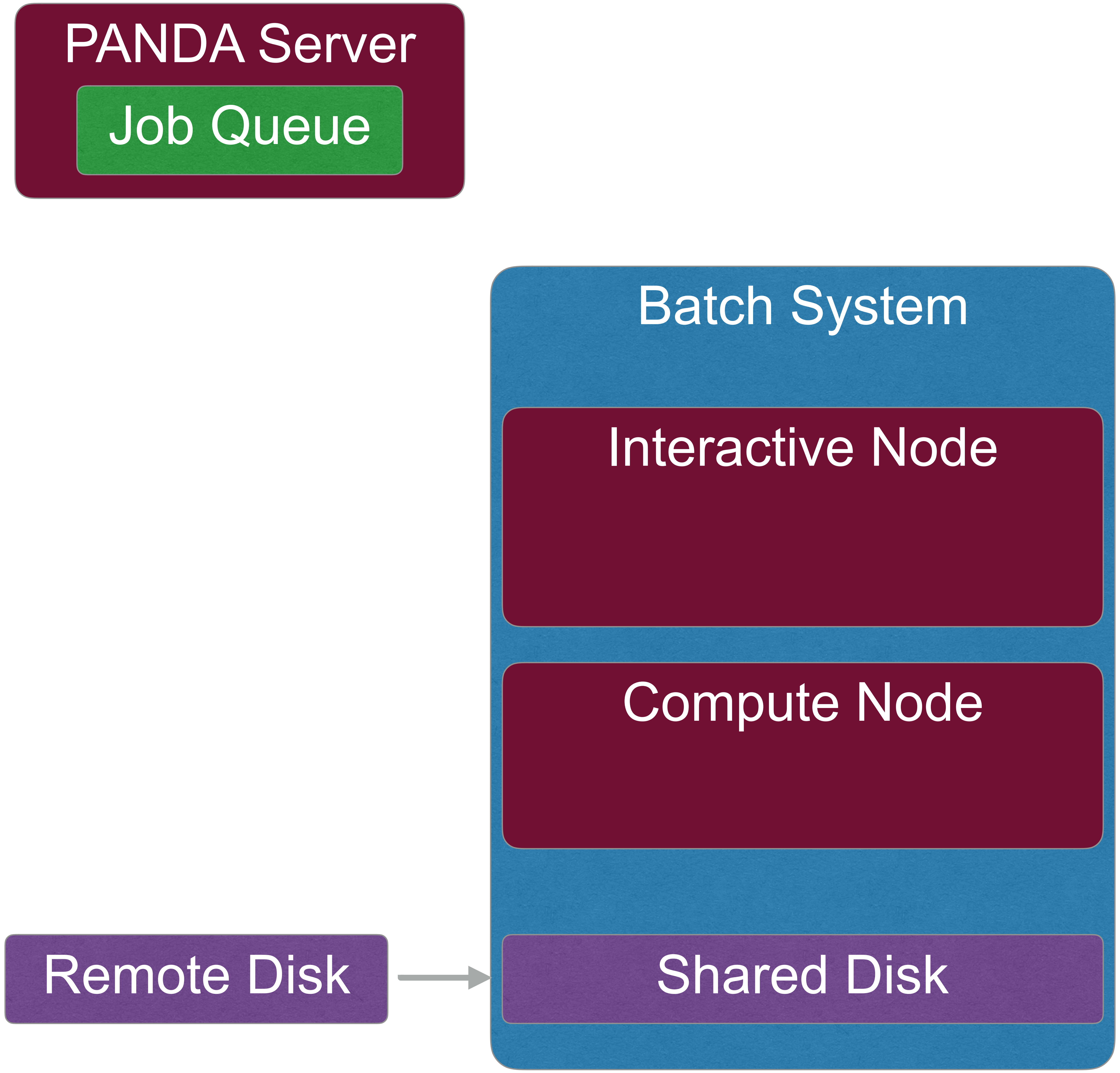
Large Hadron Collider at CERN

- LHC collides protons at the highest energy collider in the world to understand the smallest particles known to us.
- Typical LHC experiments records data at rates of ~1 GB/s.
- This data is farmed out to the Worldwide LHC Grid for processing
- In addition, the Grid is used for simulating much larger datasets to assist in searching for new discoveries.
- This work grew out of wanting to use US DOE HPCs for LHC simulations in an integrated way with experiment's job management systems.



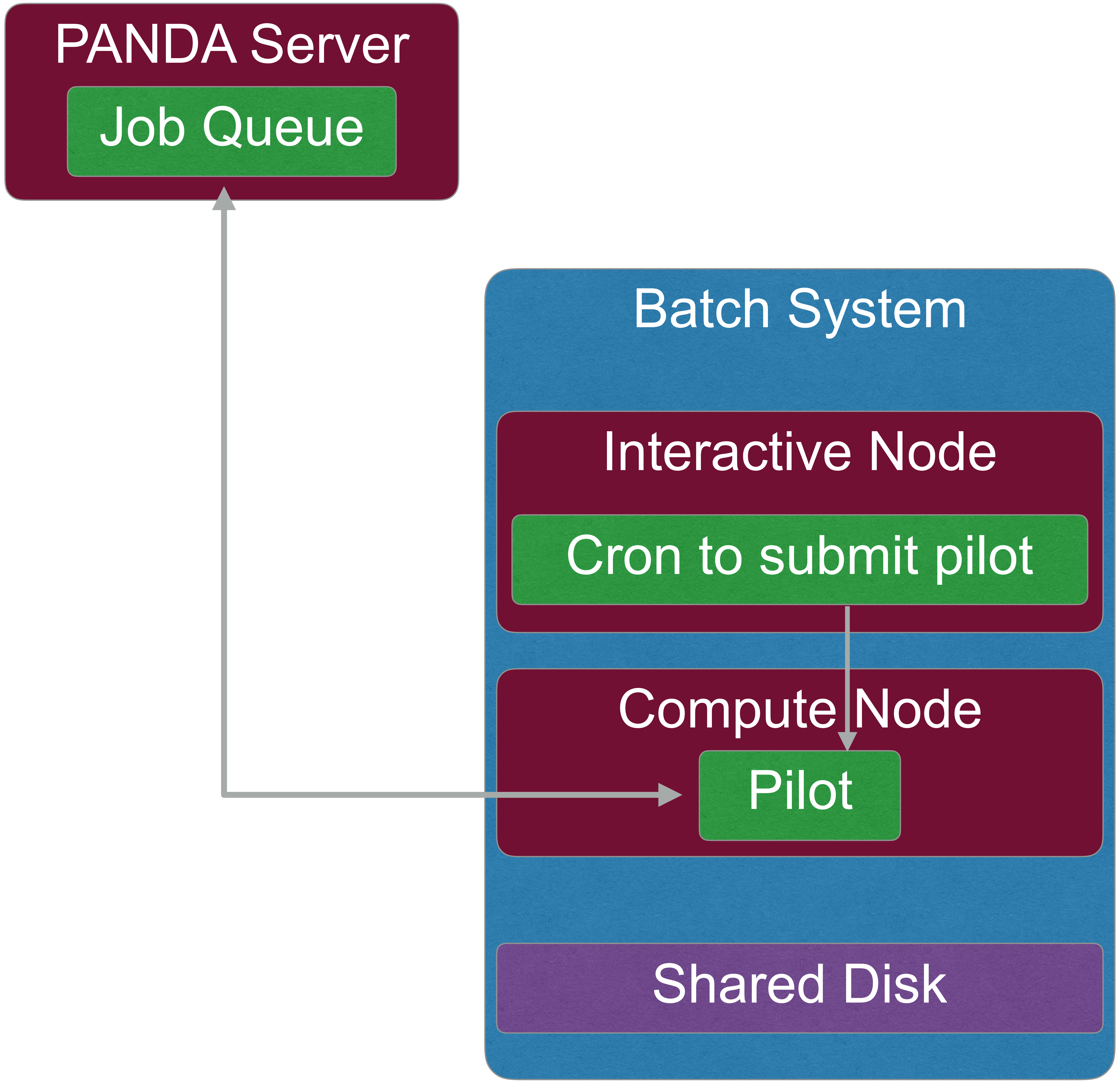
ATLAS Distributed Workflow

- ATLAS uses the PanDA (Production and Distributed Analysis) system (SciDAC project)
- PanDA hosts a job queue
- PanDA stages the needed data to the local shared filesystem via GridFTP or other services



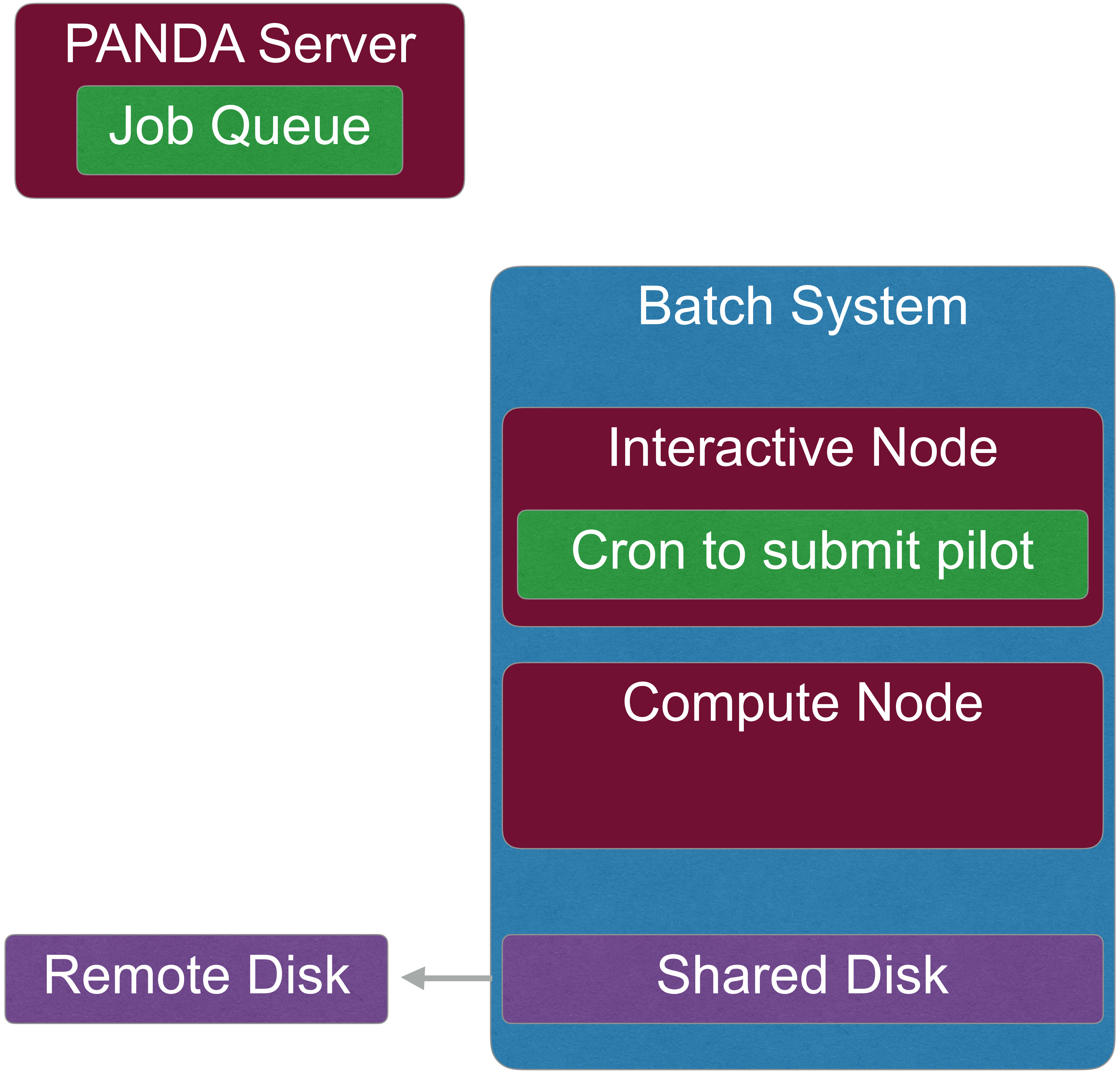
ATLAS Distributed Workflow

- ATLAS uses the PanDA (Production and Distributed Analysis) system (SciDAC project)
- PanDA hosts a job queue
- PanDA stages the needed data to the local shared filesystem via GridFTP or other services
- Pilots are submitted to the batch queues
- At startup, the pilot it reaches out (curl) to the PanDA server and retrieves an analysis job



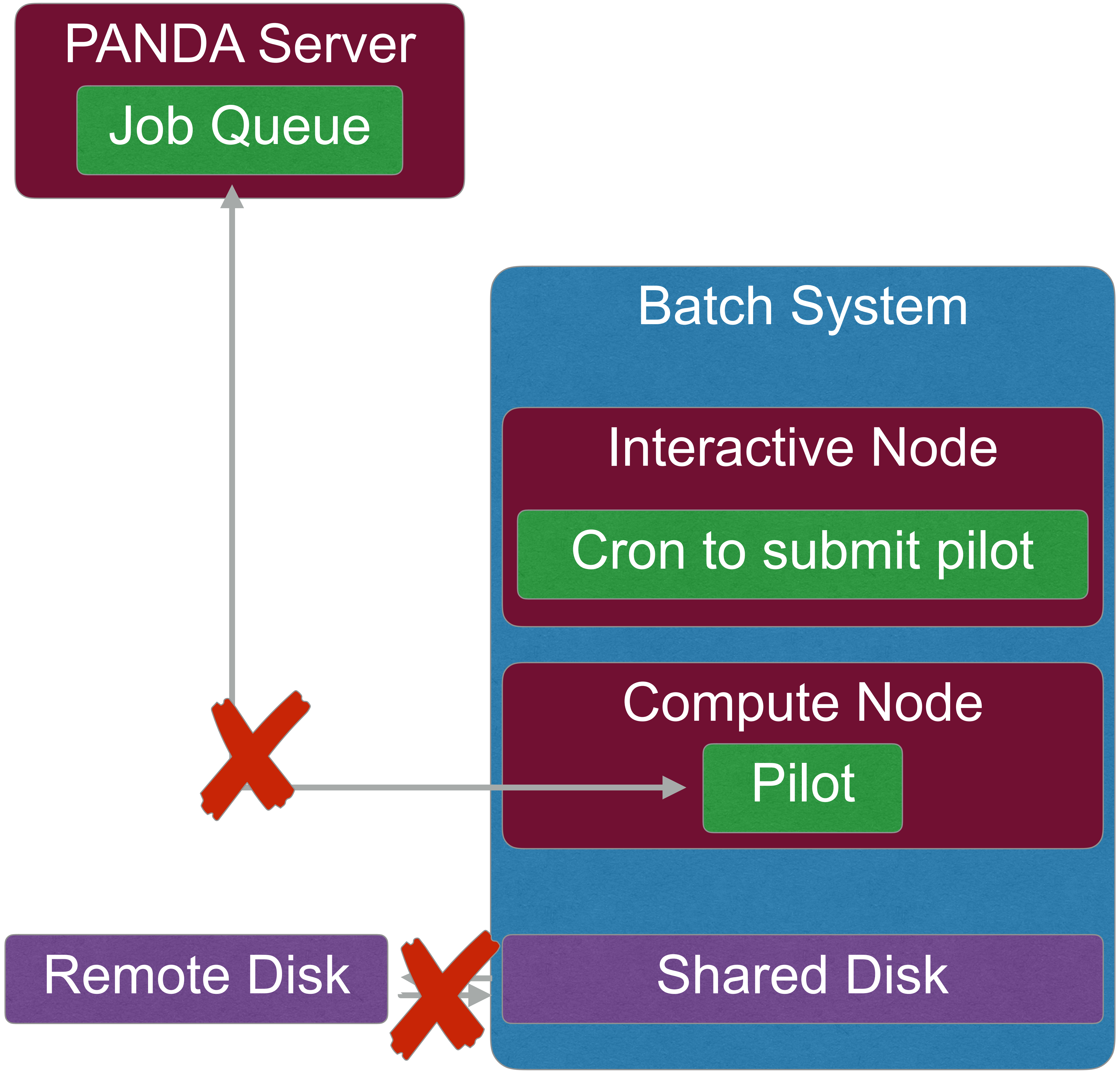
ATLAS Distributed Workflow

- ATLAS uses the PanDA (Production and Distributed Analysis) system (SciDAC project)
- PanDA hosts a job queue
- PanDA stages the needed data to the local shared filesystem via GridFTP or other services
- Pilots are submitted to the batch queues
- At startup, the pilot it reaches out (curl) to the PanDA server and retrieves an analysis job
- Once completed, output data can be retrieved by PanDA



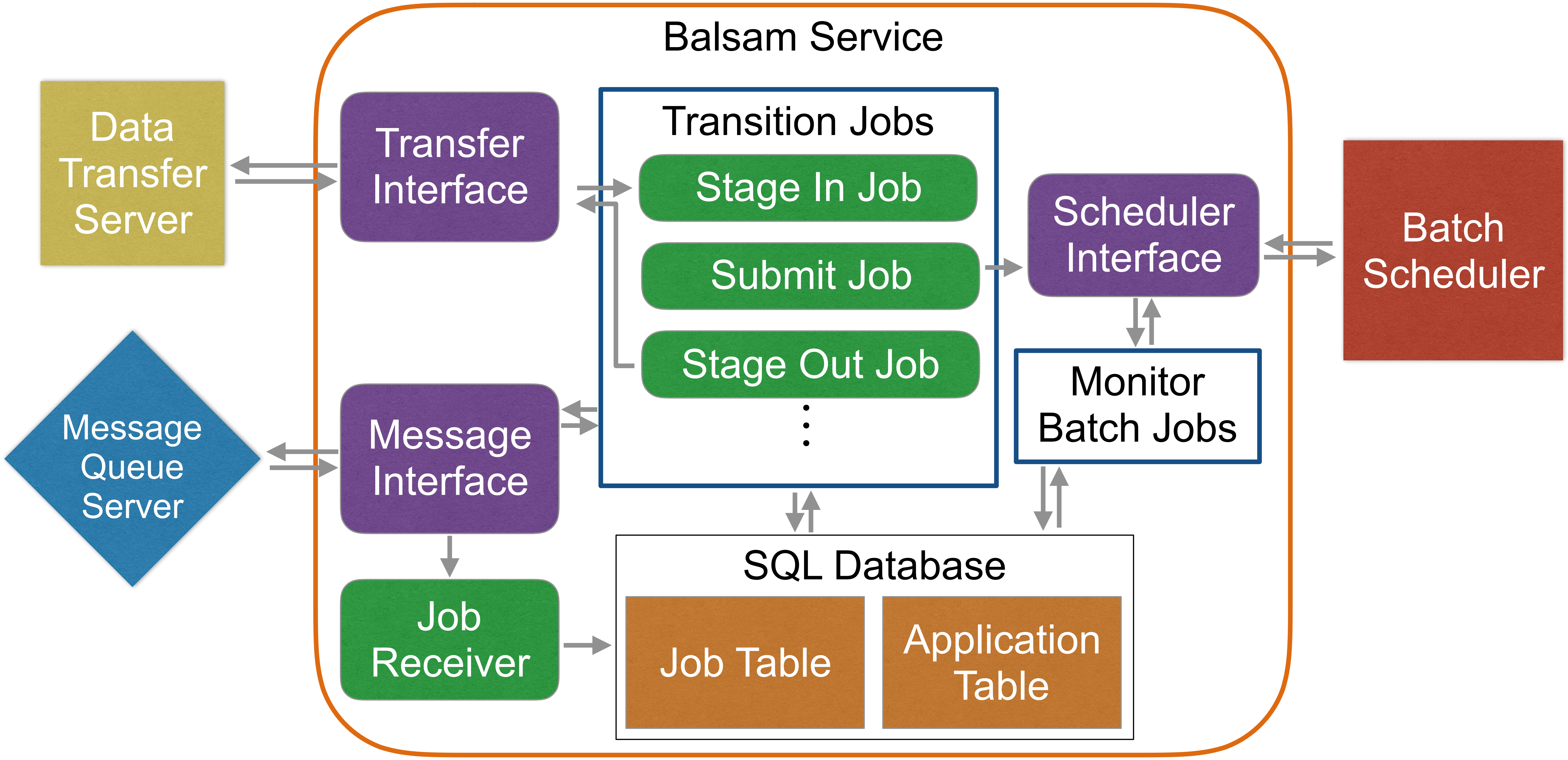
Challenges with DOE HPCs

- Two-factor Authentication
  - PanDA can not push data into site without a site certificate, but PanDA uses LHC Grid certificates, not ALCF certificates.
- Network on Worker Nodes
  - Leadership machines typically restrict network access on worker nodes, therefore, pilots cannot reach out to PanDA for jobs.
- Job Shaping
  - LHC jobs are typically malleable which allows them to be shaped to fit into gaps in schedulers of draining HPCs.



An Edge Service Named Balsam

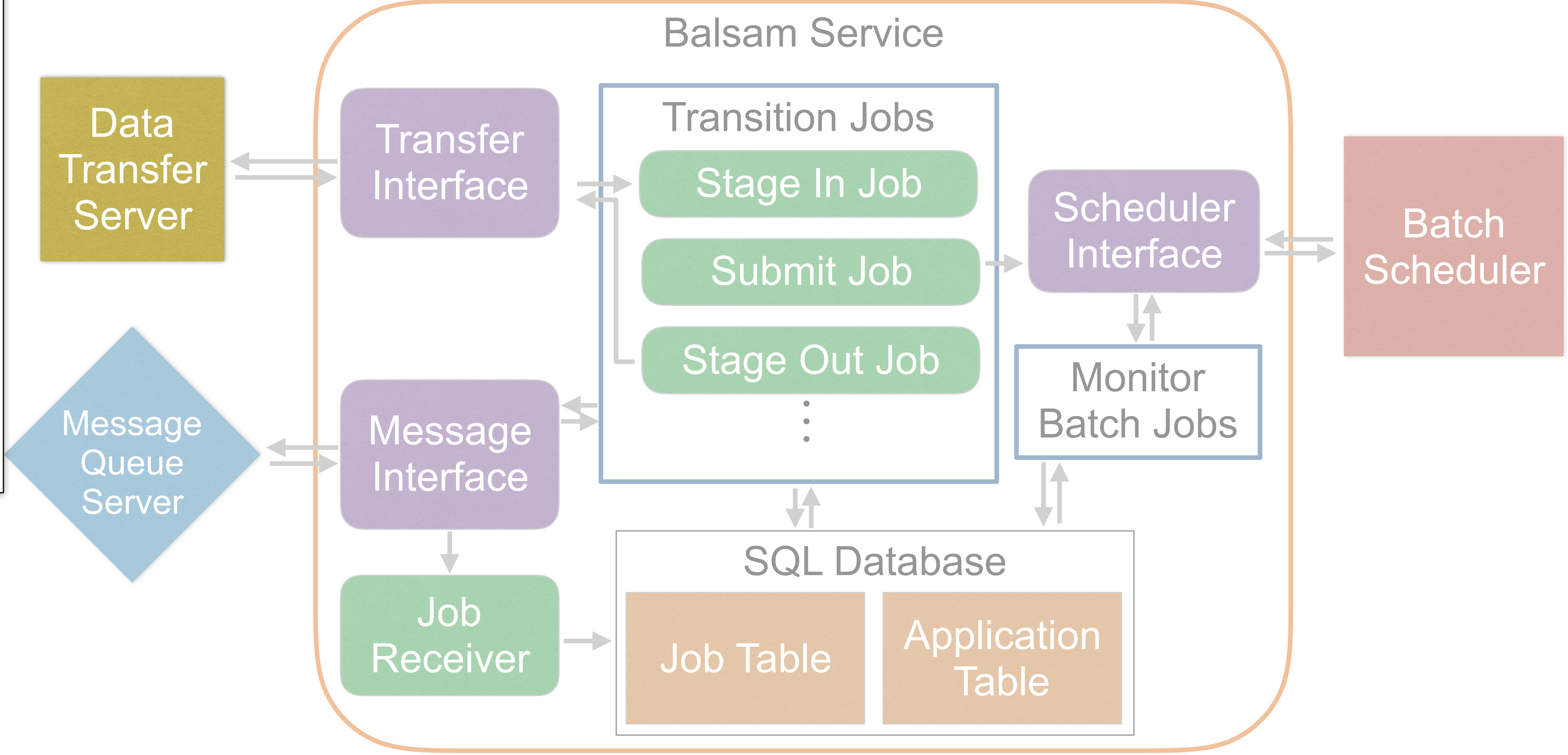
- Created an edge service to pull work into an HPC from the outside.
- Built this service on Django for its database interface and the possibility to author user interface for job submission and monitoring
- Used message queues as the connection to the outside world to avoid hand-made sockets or other custom solutions.



An Edge Service Named Balsam

Data Transfer Server

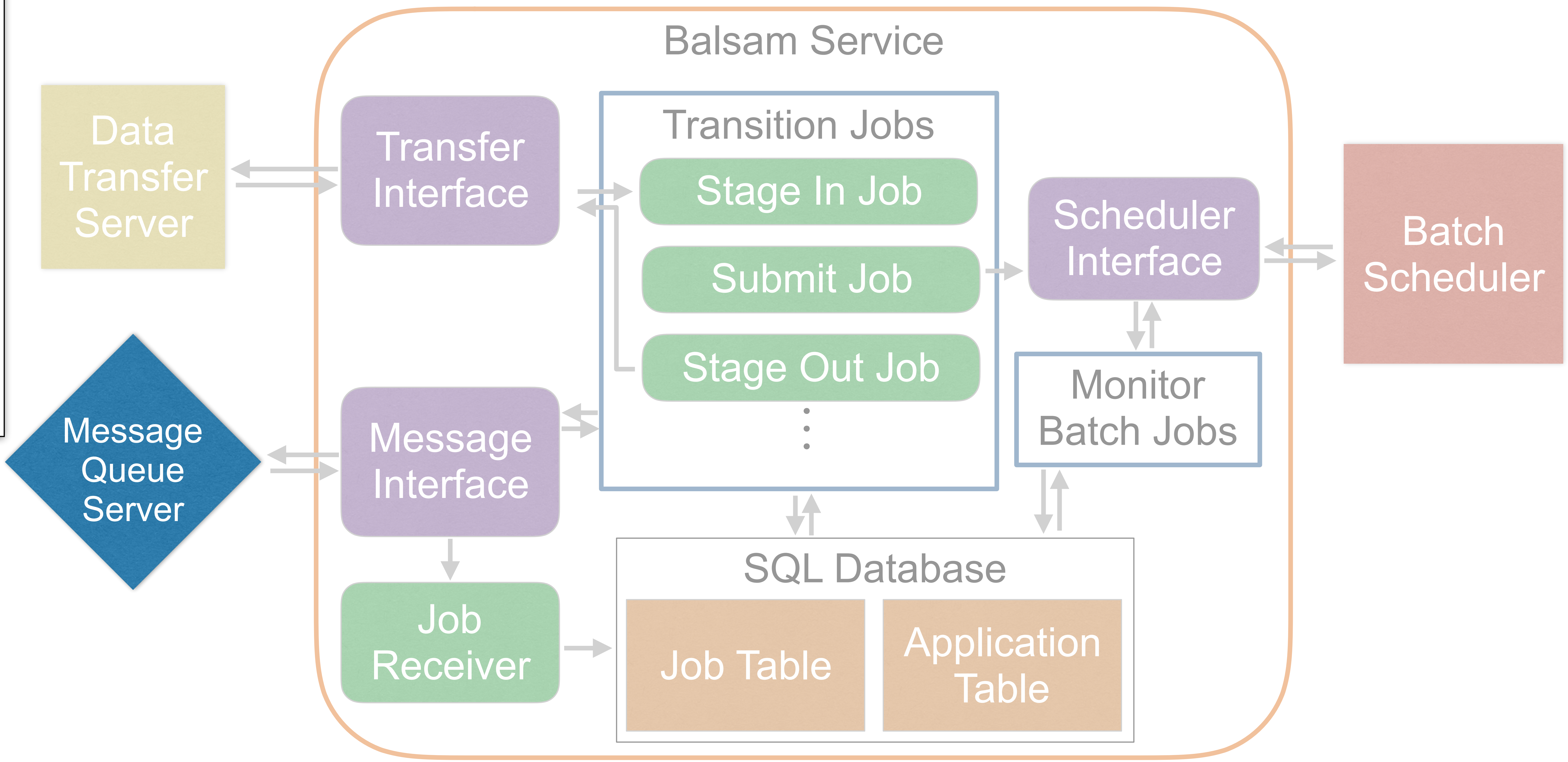
- Sits outside two-factor authentication network.
- Users specify data transfer protocol, server address, and file paths in job definition.
- Balsam transfers data in from specified server.



An Edge Service Named Balsam

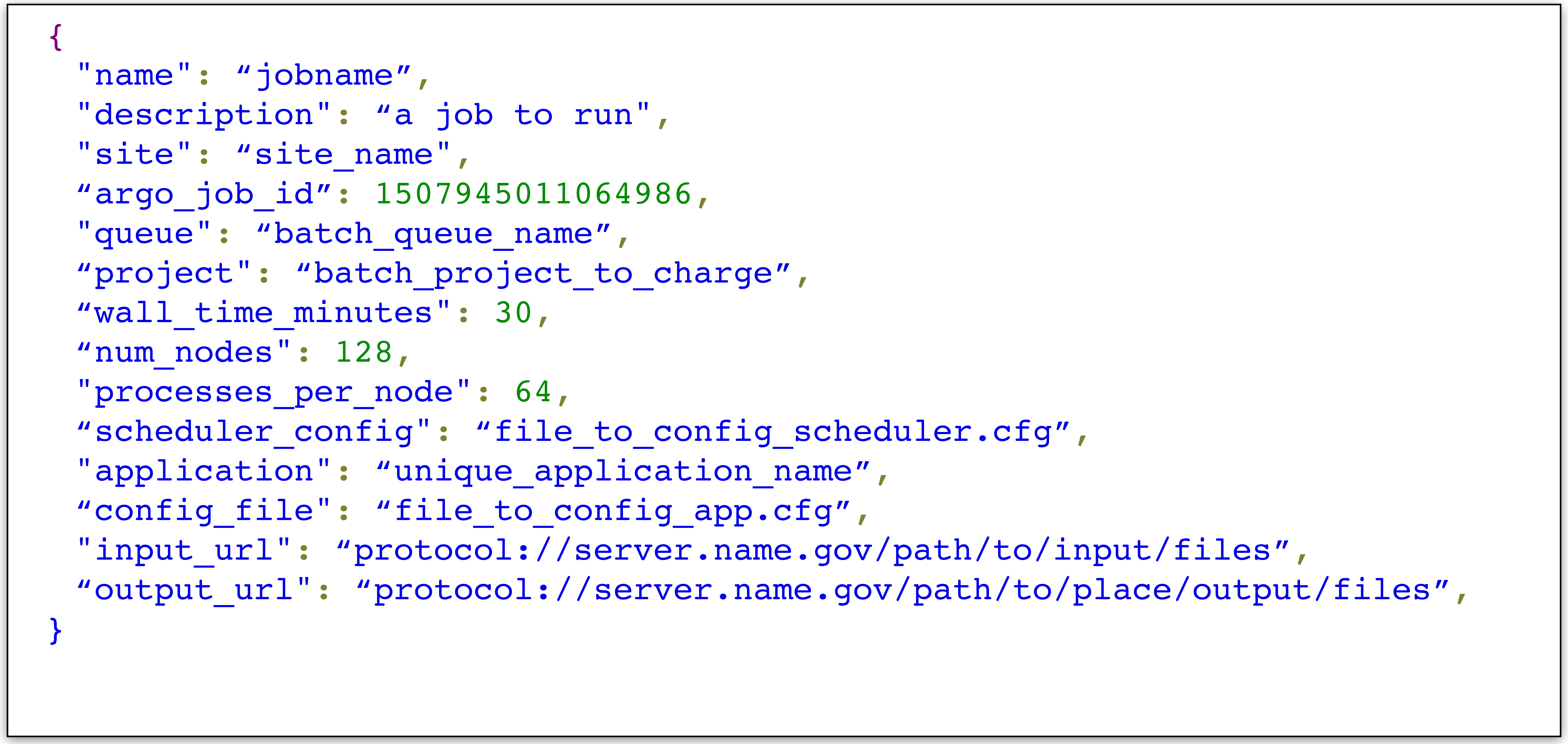
Message Queue Server

- Sits outside two-factor authentication network.
- Users or job management system submits JSON formatted job definitions to the site queue.



Balsam Job Definitions

- Incoming job definitions are JSON formatted Python dictionaries
- The definition defines:
  - Application name
  - input file path (can be None)
  - where to stage output files (can be None)
  - configuration inputs if needed
  - number of nodes and processes per nodes to run for this application as well as scheduler configuration information
- For Security reasons, Balsam contains a separate SQL table that defines the applications it is able to run.
  - User must specify unique app name in job definition
  - User provides a text file with configuration information for the application
  - This file is used to parse command line options, etc, to avoid command injection
  - Text from the user is never directly executed on the command line by Balsam

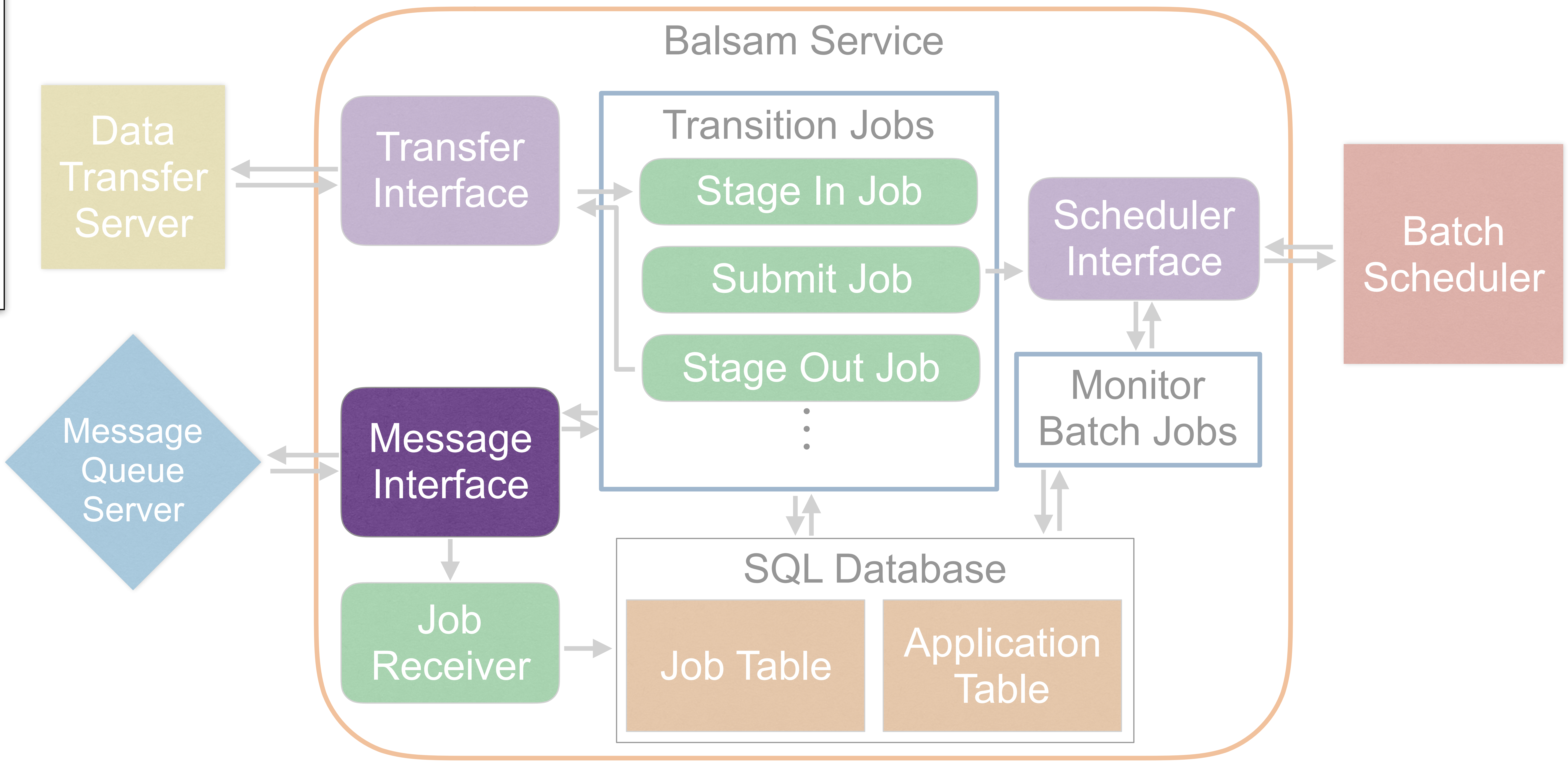


```
{
  "name": “jobname”,
  "description": “a job to run",
  "site": “site_name",
  “argo_job_id": 1507945011064986,
  "queue": “batch_queue_name”,
  “project": “batch_project_to_charge”,
  “wall_time_minutes": 30,
  “num_nodes": 128,
  "processes_per_node": 64,
  “scheduler_config": “file_to_config_scheduler.cfg”,
  "application": “unique_application_name”,
  “config_file": “file_to_config_app.cfg”,
  "input_url": “protocol://server.name.gov/path/to/input/files”,
  “output_url": “protocol://server.name.gov/path/to/place/output/files”,
}
```

An Edge Service Named Balsam

Message Interface

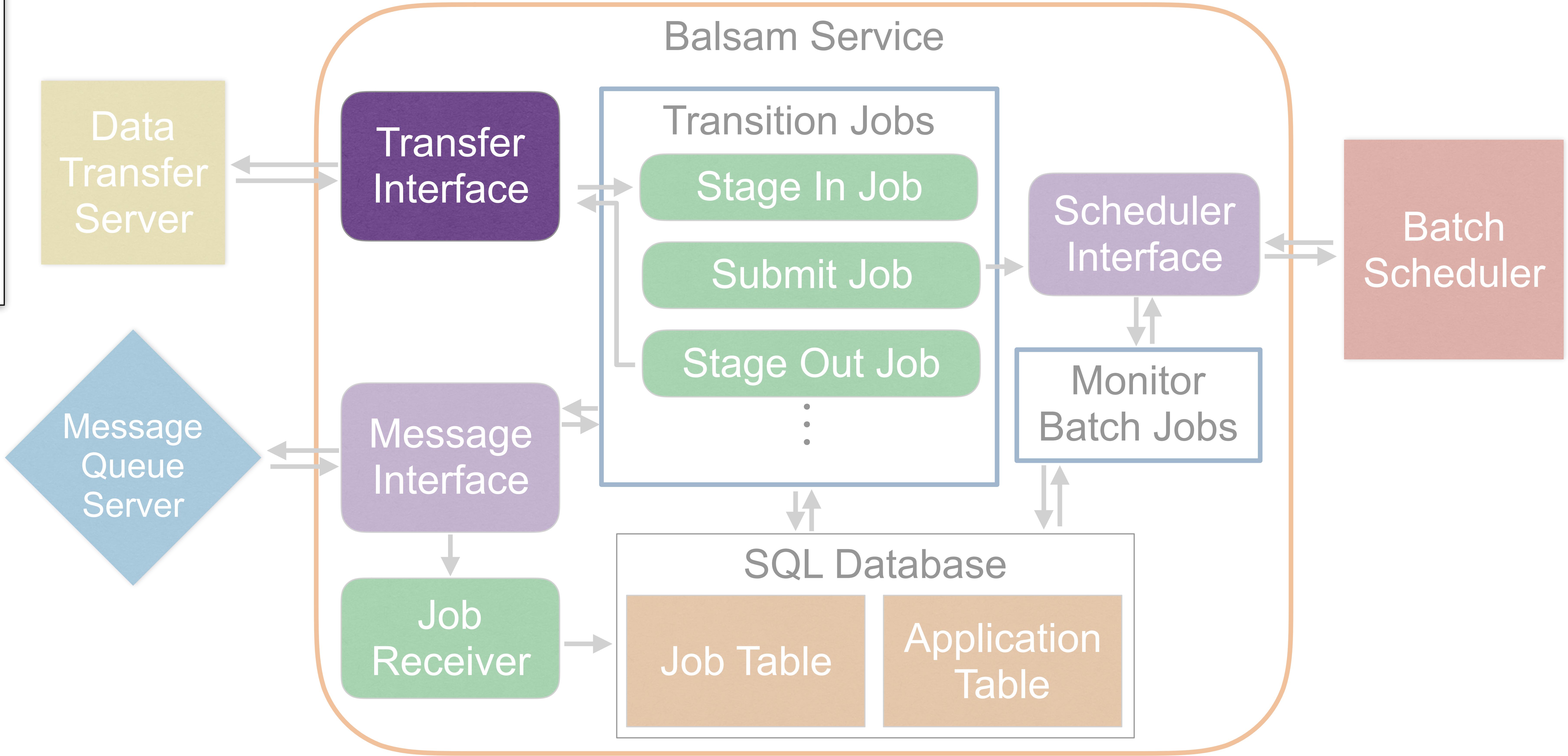
- Abstraction of the message interface. In our case we used RabbitMQ, but others can be substituted



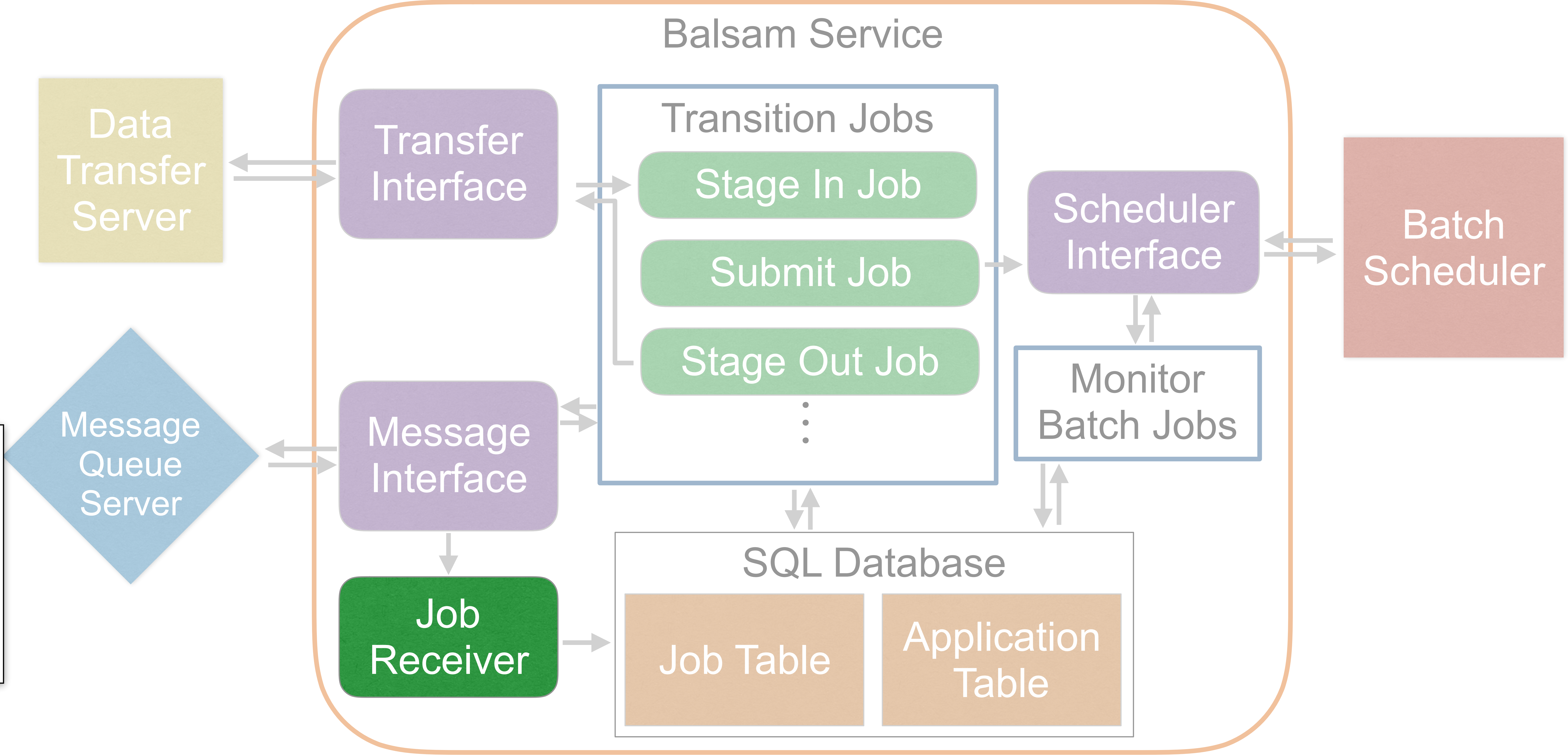
An Edge Service Named Balsam

Transfer Interface

- Abstraction of the transfer mechanism. In our case we implemented GridFTP, SCP, and CP, but others can be added.



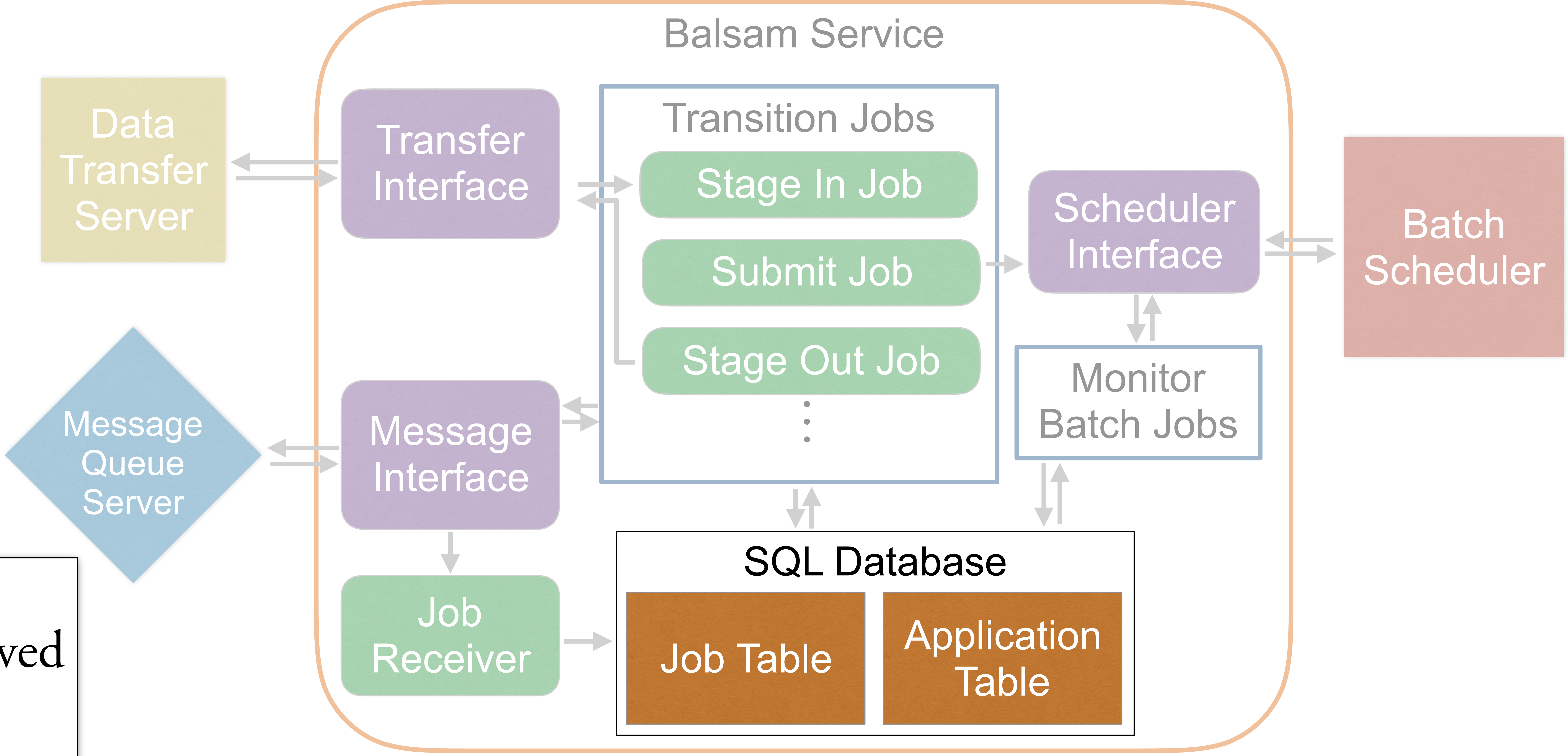
An Edge Service Named Balsam



Job Receiver

- Processes incoming jobs into database entries via Django model objects

An Edge Service Named Balsam



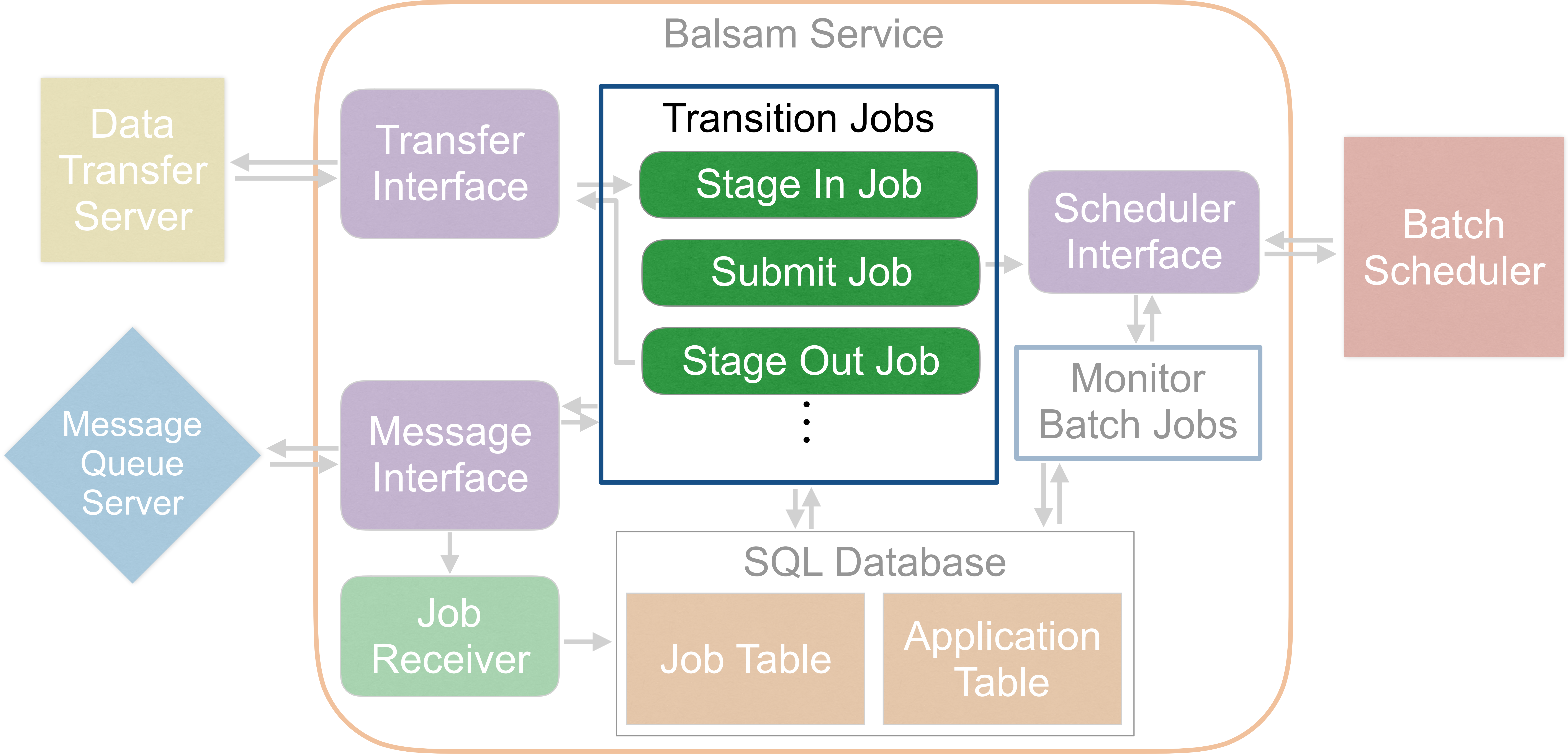
Database

- Table of allowed applications
- Table of jobs

An Edge Service Named Balsam

Transition Jobs

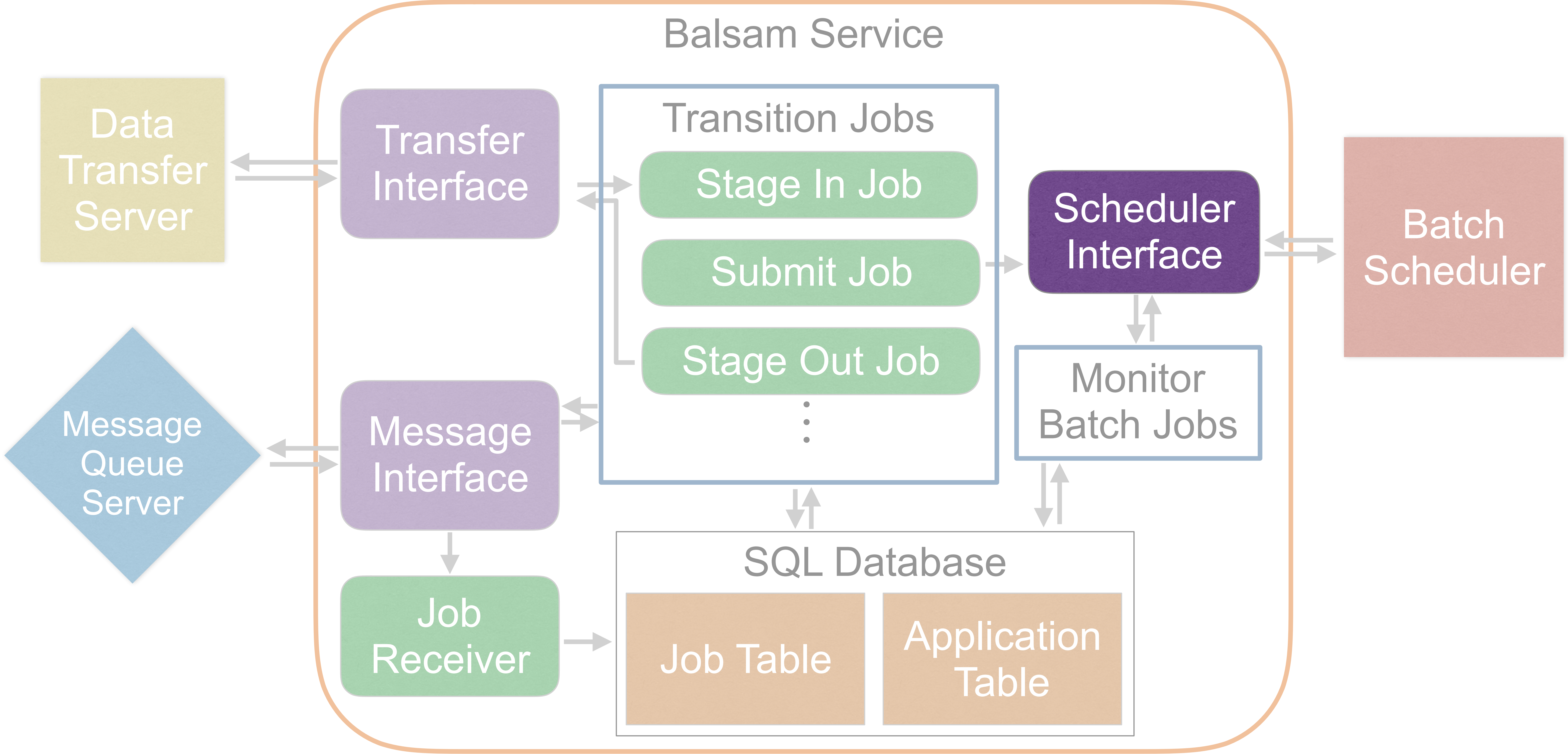
- Checks DB for jobs waiting for next state transition
- Triggers job transitions in sub-threads, such as data staging, preprocessing and post processing, and batch queue submission
- Job state changes are reported on a message queue available to the outside for monitoring



An Edge Service Named Balsam

Scheduler Interface

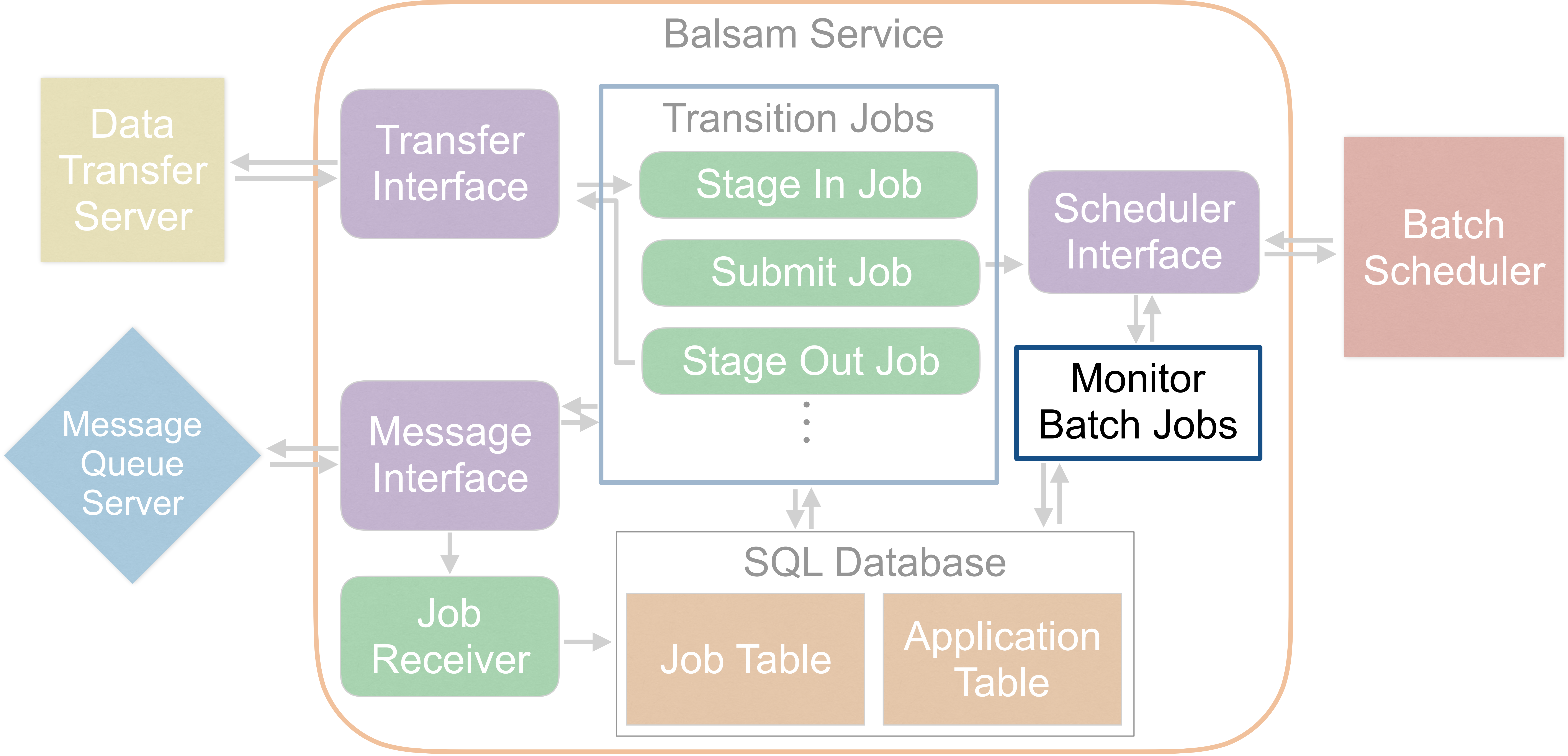
- Abstraction of the batch scheduler interface.
- Currently have plugins for Condor, Cobalt, Slurm, and Torque.
- Others can be added.



An Edge Service Named Balsam

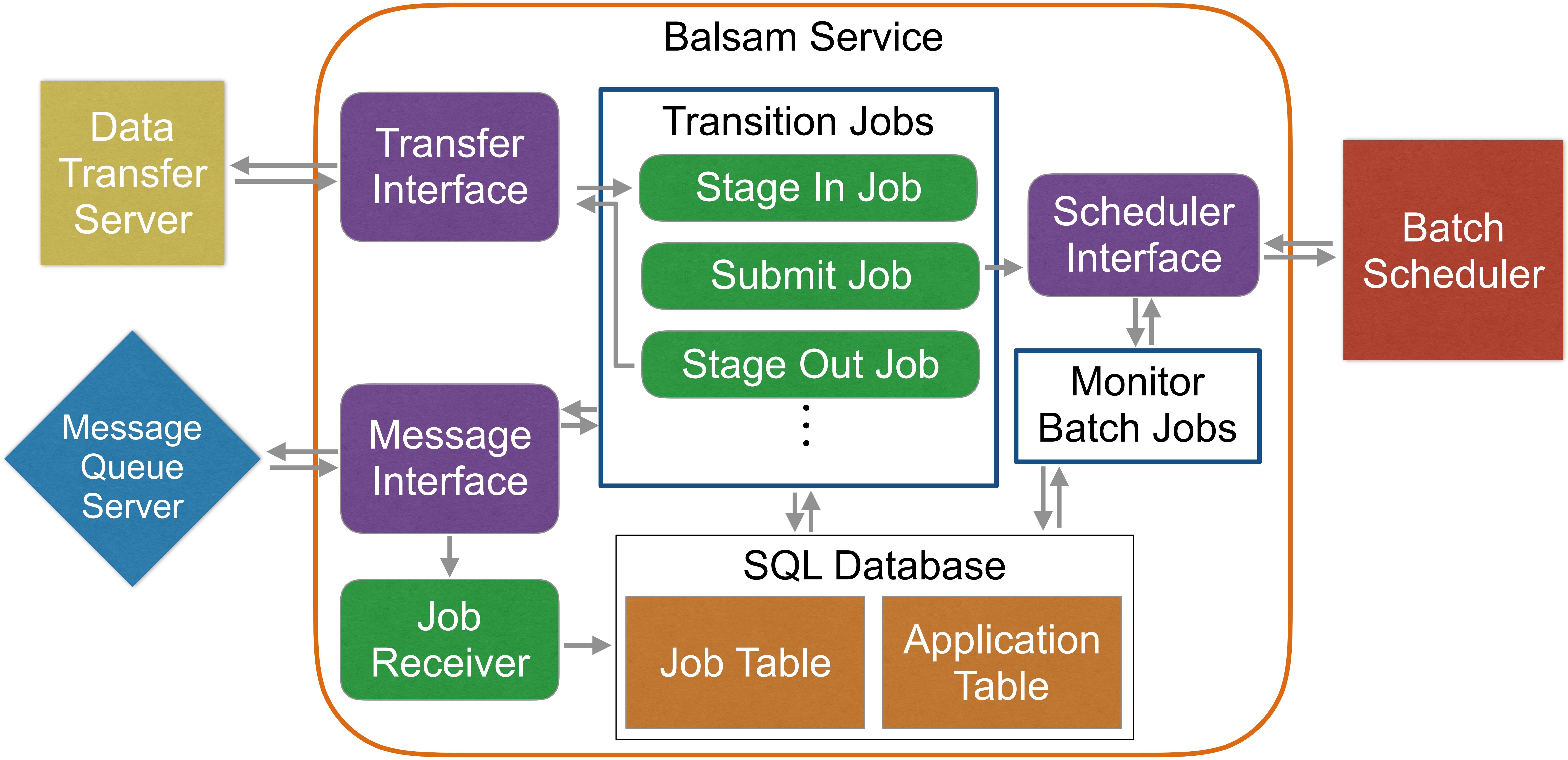
Monitor Batch Jobs

- Monitors jobs in the DB that are running on the batch system.
- Upon job exit, changes job state to trigger Transition Job thread



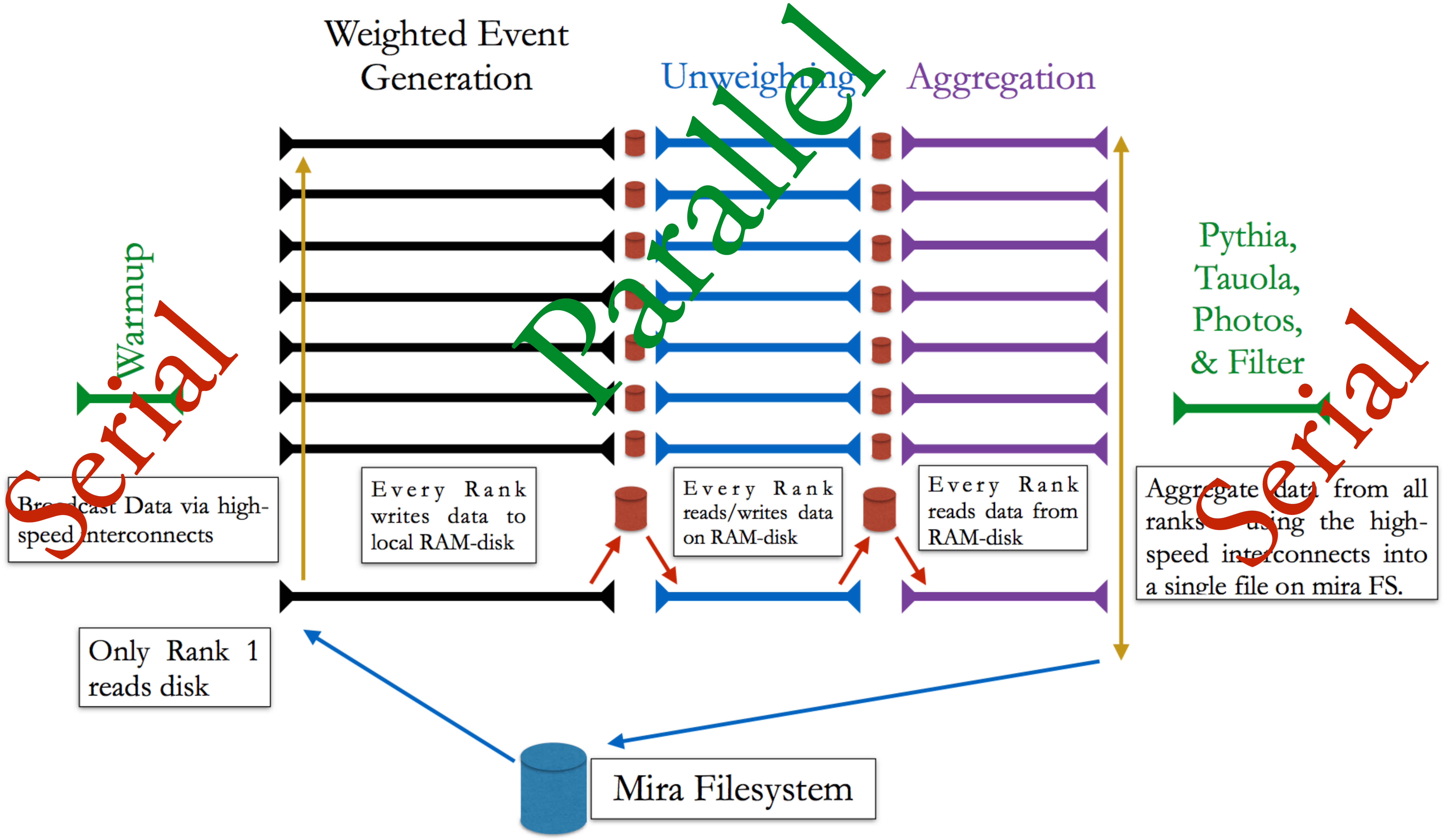
An Edge Service Named Balsam

- Now we were ready to submit lots of jobs to Mira for LHC simulation workflows, right?
- Well, not quite.



Mixed Scale Workflows

- The applications we were trying to scale on Mira contains a 4-5 hour serial pre-processing step and a quick serial post-processing step.
- We did not want to run these on the login nodes and wanted this to be seamless in the PanDA system.



Weighted Event Generation

Unweighting

Aggregation

Parallel

Warmup

Serial

Serial

Pythia, Tauola, Photos, & Filter

Broadcast Data via high-speed interconnects

Every Rank writes data to local RAM-disk

Every Rank reads/writes data on RAM-disk

Every Rank reads data from RAM-disk

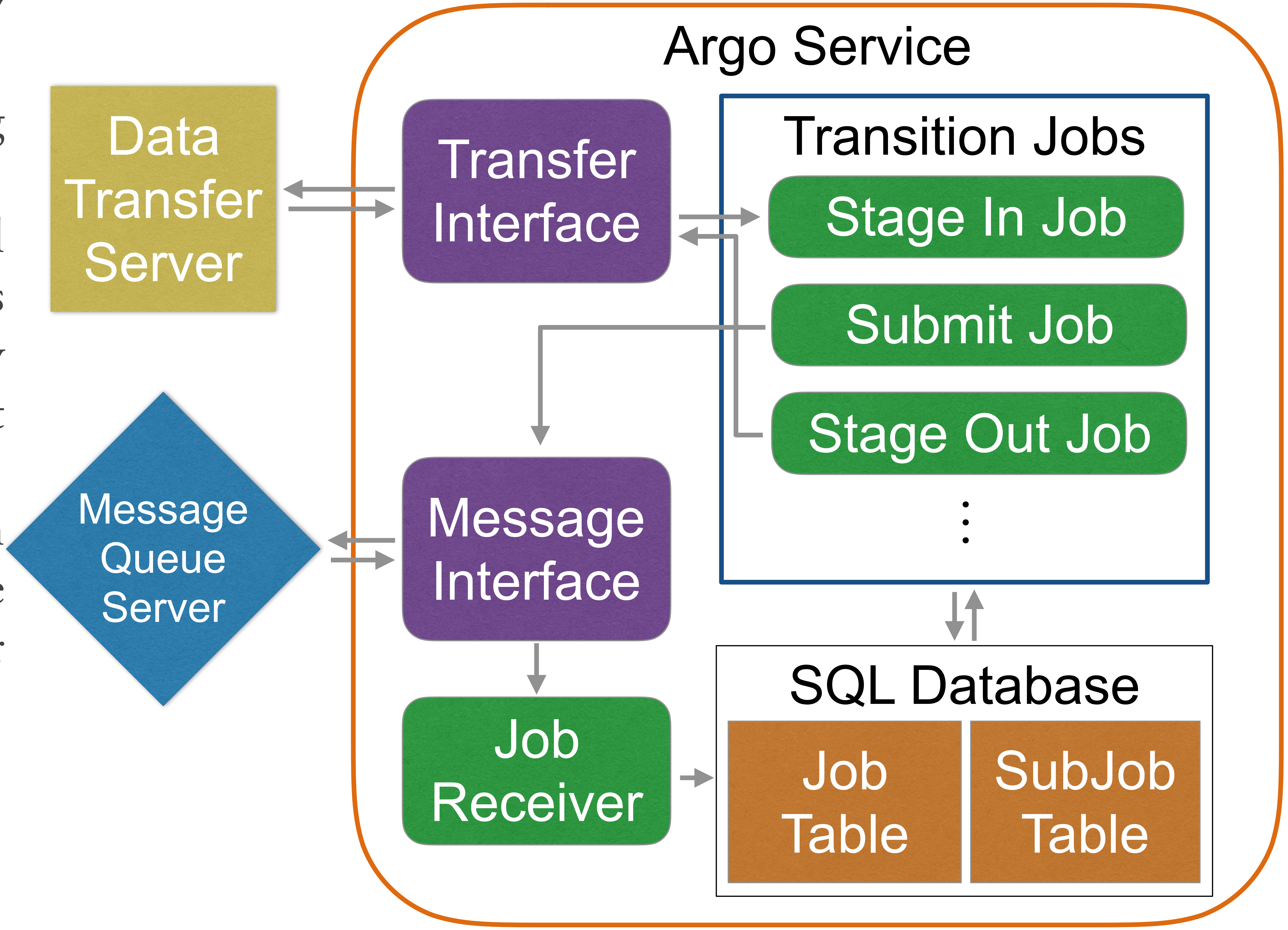
Aggregate data from all ranks using the high-speed interconnects into a single file on mira FS.

Only Rank 1 reads disk

Mira Filesystem

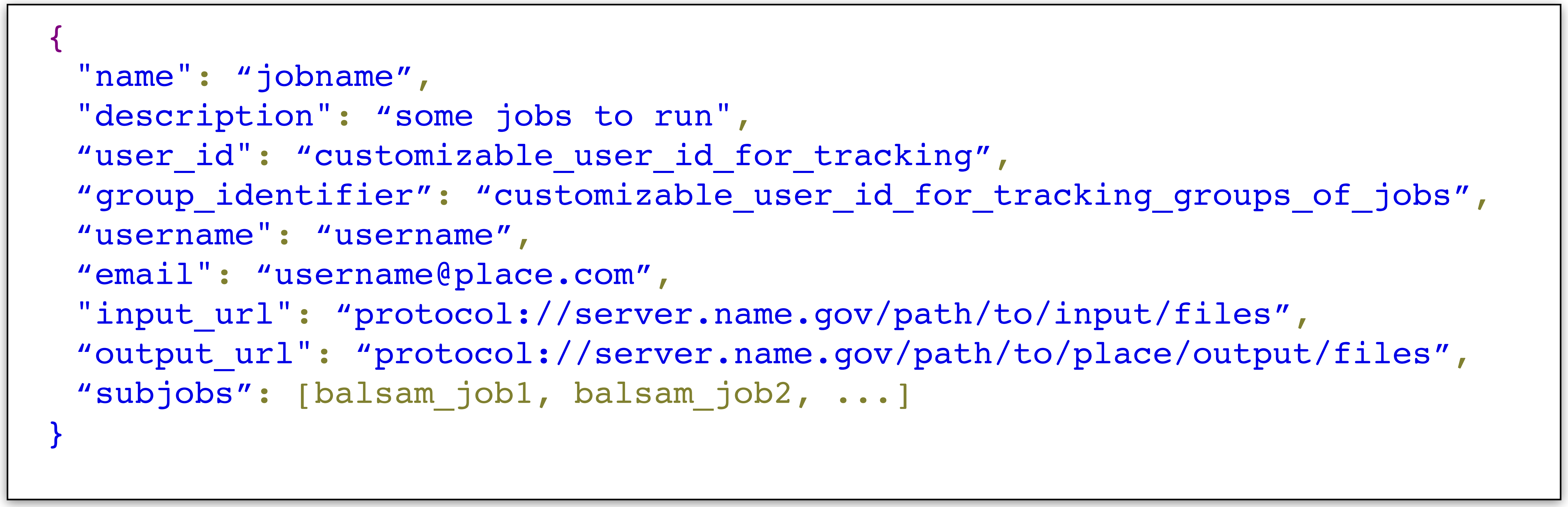
An Edge Service Named Argo

- Argo is a service run independently from Balsam.
- Argo is very Similar to Balsam re-using many parts.
- Argo does not submit jobs to a local batch queue, instead executing jobs made of sub-jobs that are run by Balsam instances running on different sites.
- Communication/Data-flow between Argo and Balsam is achieved using the Message Queue and Data Transfer Servers



Argo Job Definitions

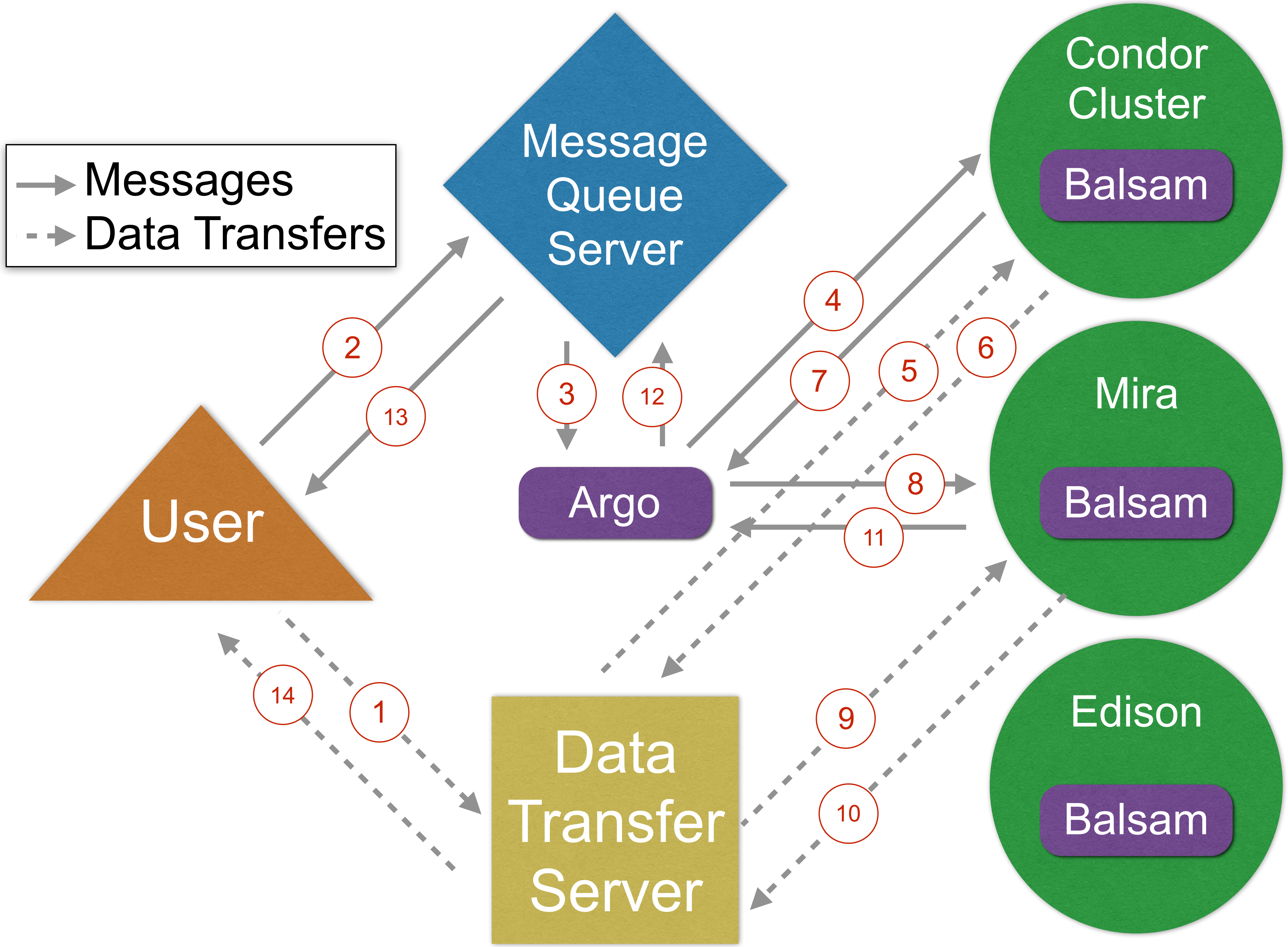
- Incoming job definitions are JSON formatted Python dictionaries
- The main difference from Balsam is that the definition does not specify one job, but a JSON list of jobs.
- Each job in the ‘subjobs’ list follow the Balsam job definition described previously.



```
{
  "name": “jobname”,
  "description": “some jobs to run",
  “user_id": “customizable_user_id_for_tracking”,
  “group_identifier”: “customizable_user_id_for_tracking_groups_of_jobs”,
  “username": “username”,
  “email": “username@place.com”,
  "input_url": “protocol://server.name.gov/path/to/input/files”,
  “output_url": “protocol://server.name.gov/path/to/place/output/files”,
  “subjobs”: [balsam_job1, balsam_job2, ...]
}
```

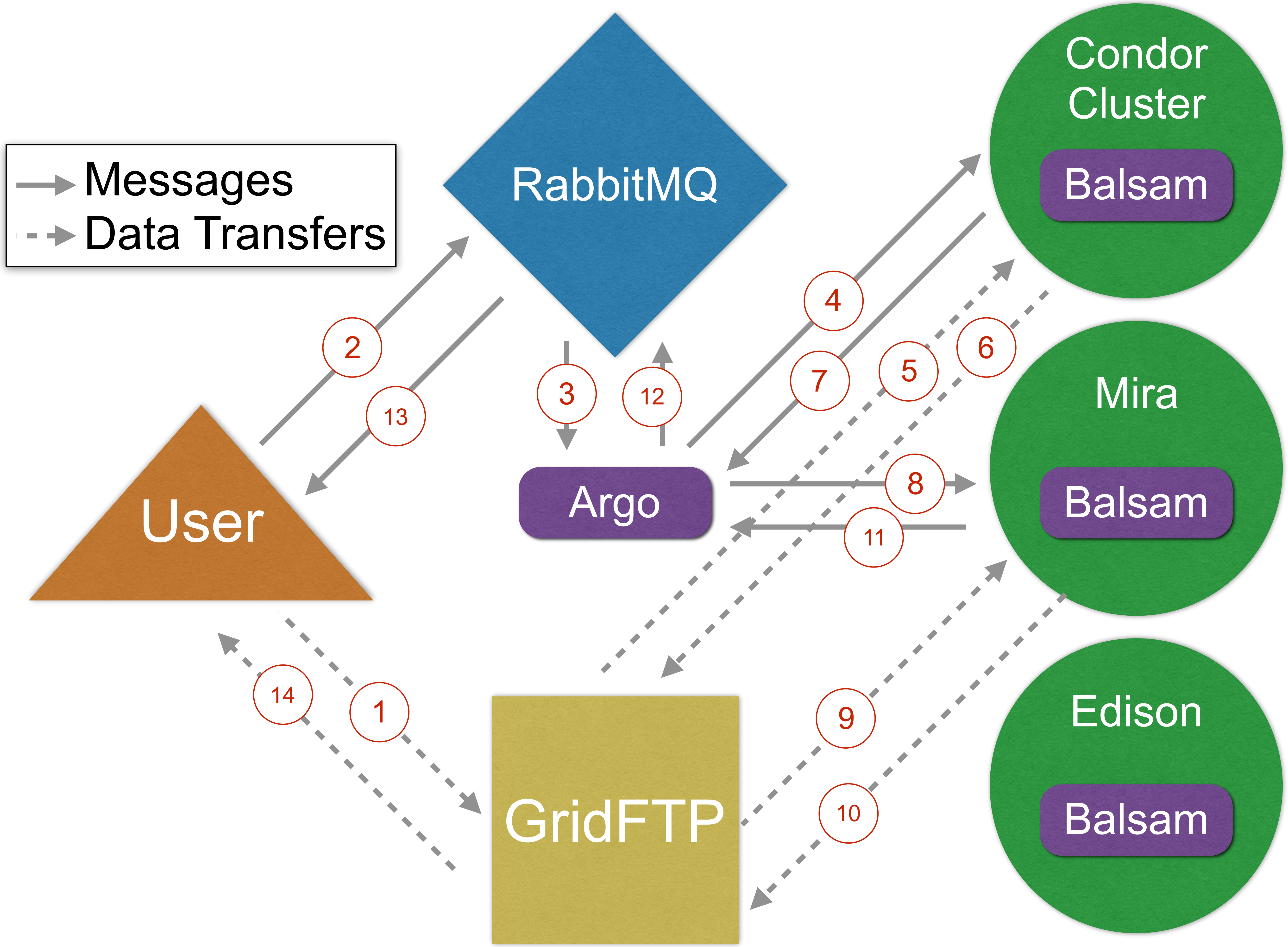
An Edge Service

- Together, Argo & Balsam enable the workflow for the LHC simulation.
- The ATLAS experiment's job manager could submit jobs via the message queue to Argo.
- This Argo job would be composed of three Balsam subjobs.
- The first and third would run serially on a Condor Cluster.
- The second job would run the highly parallel simulation on the Mira or Edison supercomputers.



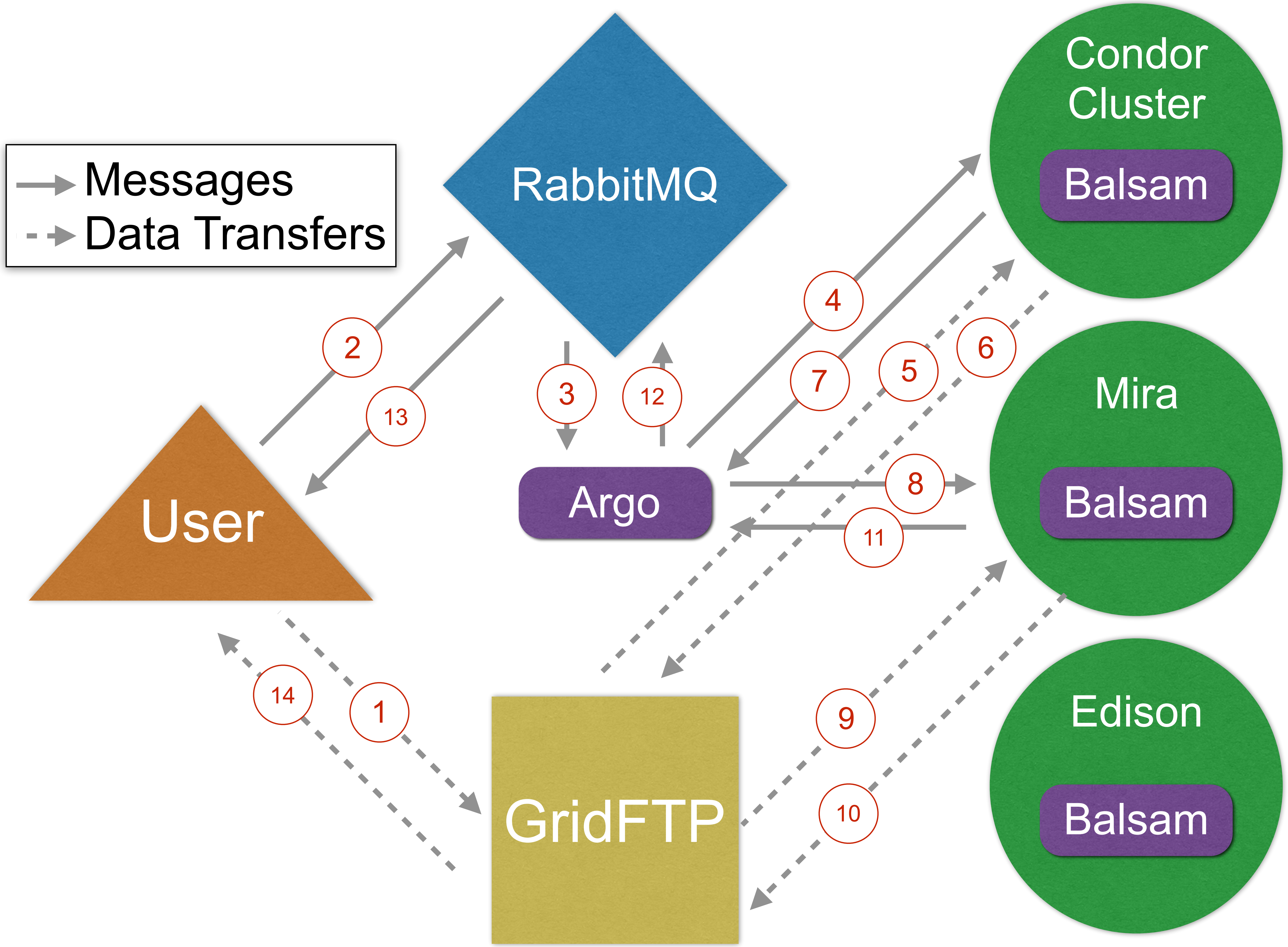
An Edge Service in Use

- In the use case for ATLAS, we used:
  - GridFTP servers for data transfers and
  - RabbitMQ for the message queue server.
- We did preliminary work to create the PanDA side code to submit jobs from ATLAS via the message queue, but this never went into production.
- Primarily ATLAS ran targeted large scale simulations that benefit from the Mira supercomputer.
- These jobs were submitted to RabbitMQ by command line.
- Once completed, on the order of 10TB of data were uploaded to the LHC Grid for downstream analyses.

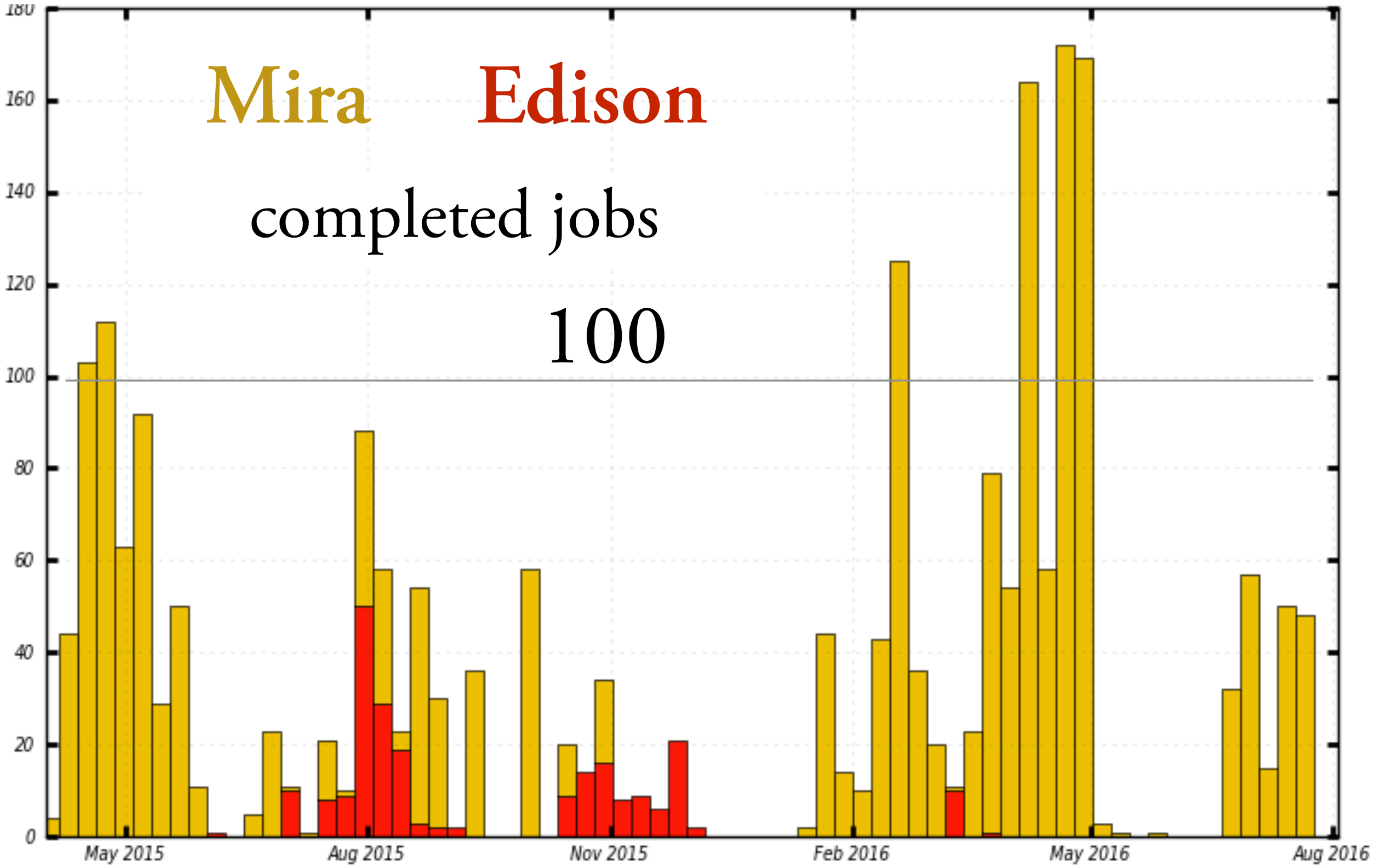


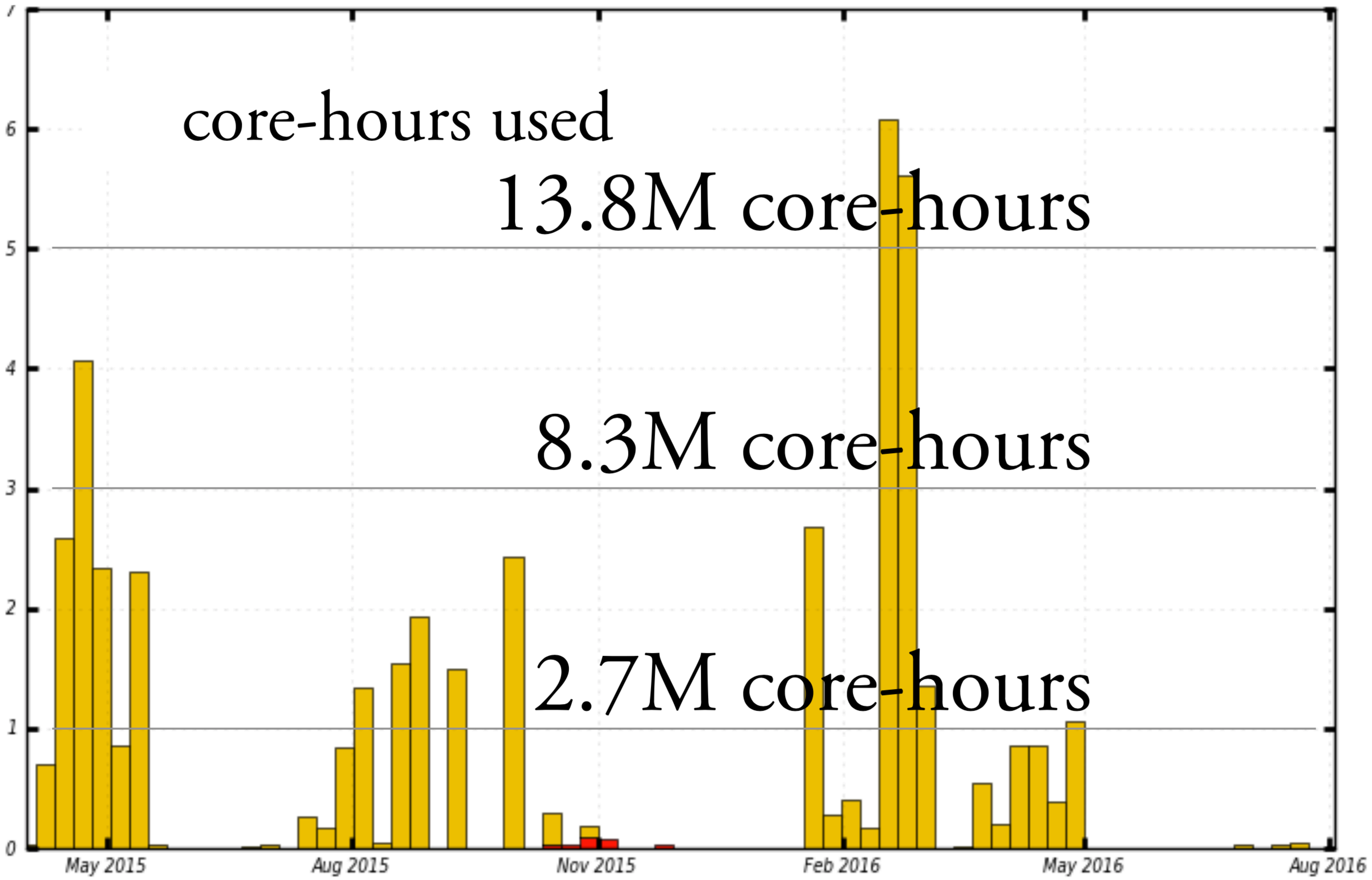
A Note on Security

- Balsam is operated at sites with two-factor authentication and design choices were made based on respecting this security.
- RabbitMQ is configured to use SSL certificates to authenticate users who post and view messages.
  - This allows for using local MyProxy certificates to authenticate to the service, which are generated using the local two-factor authentication methods. The certificates have 1 week lifetimes and therefore must be renewed periodically.
- Balsam restricts what applications can be run on the batch system, and no code or setting from a user is ever run on a command line.
- Services only contact known specified servers and do not listen for outside connections.



An Edge Service in Use

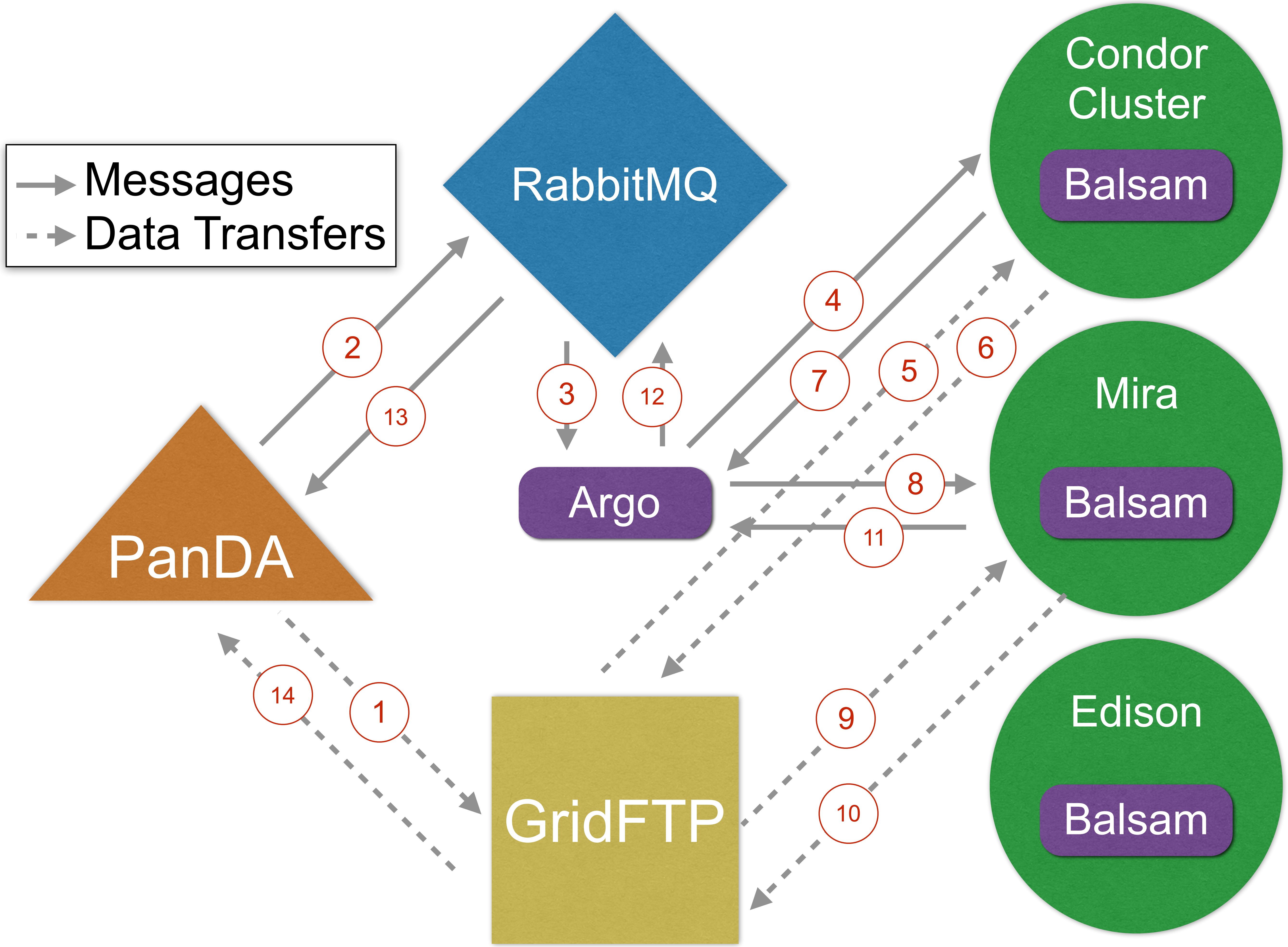




- About 122M core-hours were used during this production campaign
- Resulted in data used in 20 papers so far and still in use.

Summary

- We designed an HPC edge service that can submit jobs across multiple HPC sites, particularly designed for those with strict access control.
- These jobs can have both serial and parallel steps to be run at different sites.
- Maintaining the security of a site was an important factor in design choices.
- This work has led to over 20 scientific publications in the ATLAS experiment.



Messages
Data Transfers
RabbitMQ
PanDA
Argo
GridFTP
Condor Cluster
Balsam
Mira
Balsam
Edison
Balsam
1
2
3
4
5
6
7
8
9
10
11
12
13
14