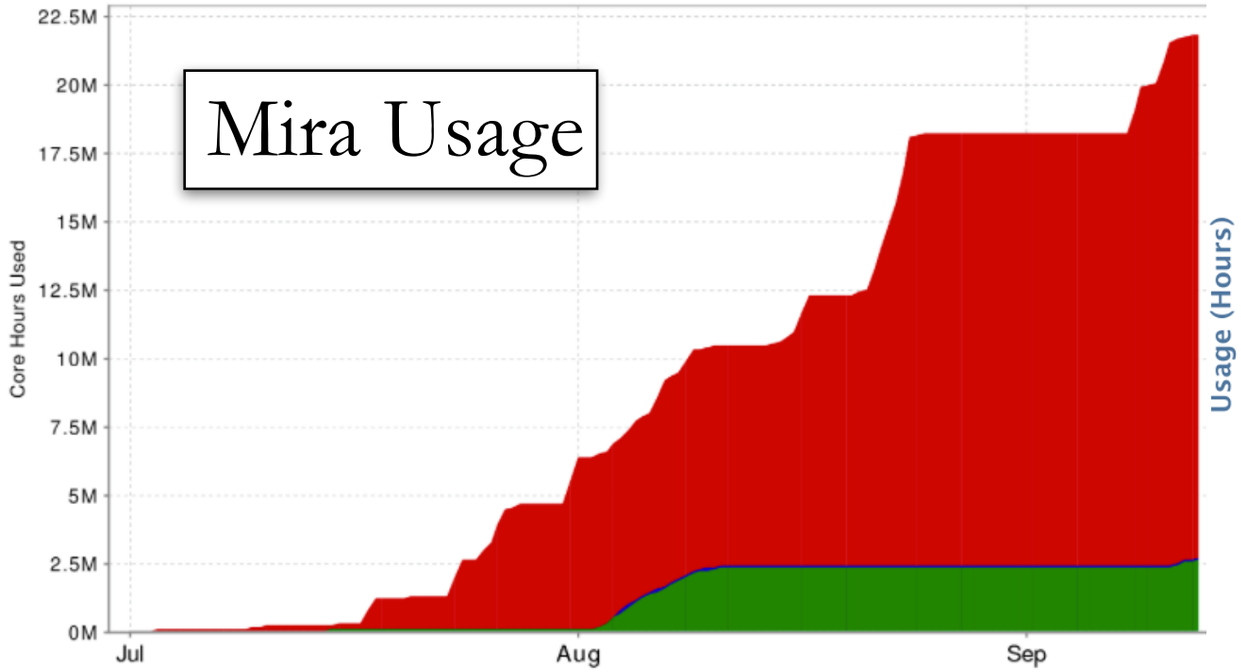
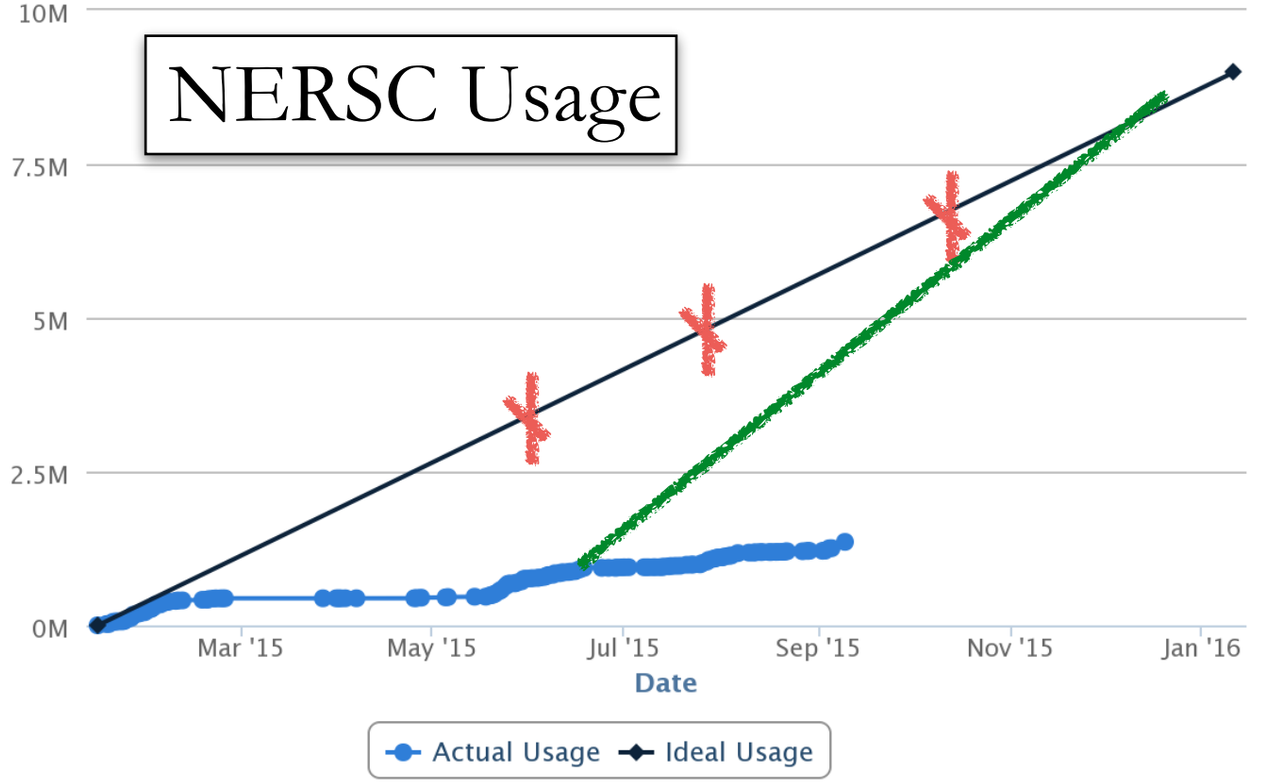
ALCF/NERSC/ALCC Update

J. Taylor Childers
with Tom LeCompte (ANL),
Doug Benjamin (Duke),
Tom Uram (ALCF),
and others…

2015-2016 ALCC Status

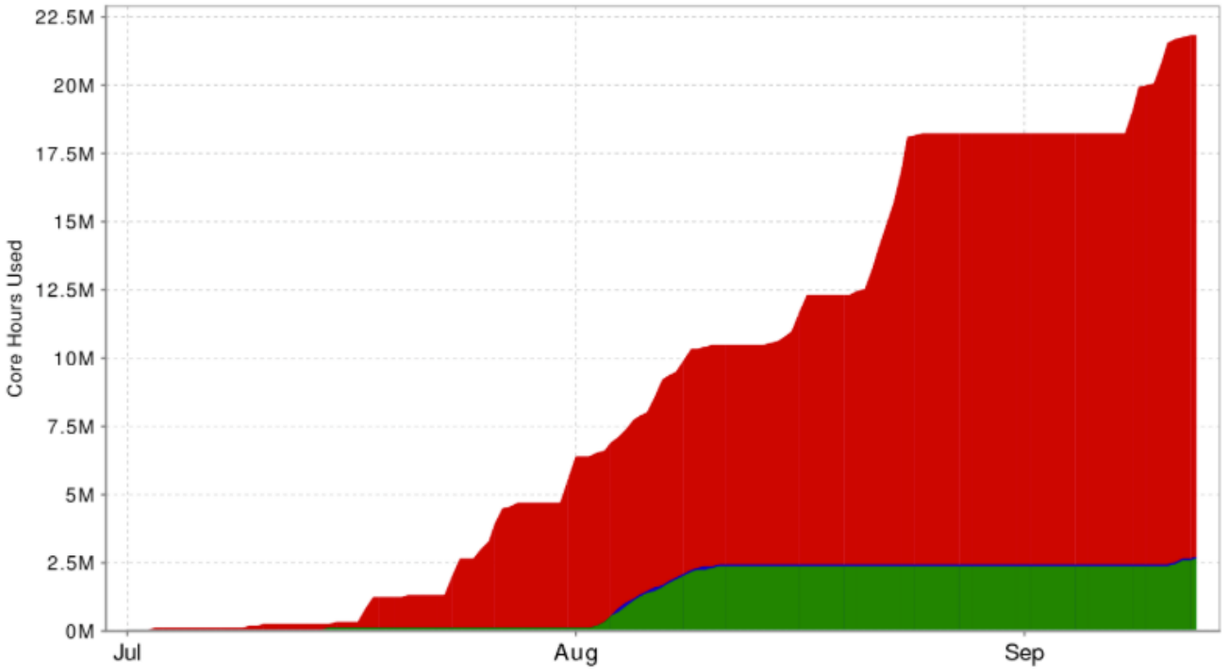
- July 2015 - June 2016 Award
- 62M core-hours on Mira
  - Already used 22M core-hours
- 16M core-hours at NERSC
  - awarded in pieces from Jul-Dec and Jan-Jun





More Alpgen…

- ATLAS Modified Alpgen inputs at the last minute
- Resulted in x2 reduction in Pythia Shower efficiency in V+5jets
- This is the largest sample we produce
- Requires 30M more core-hours than previously reported
- Good thing we have an allocation for this

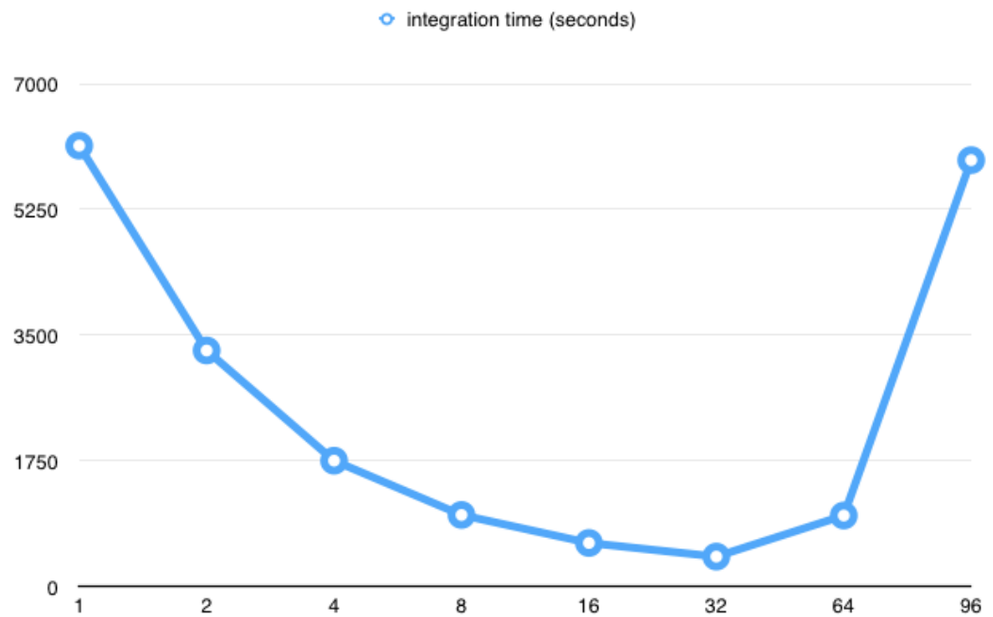


Sherpa Production

- We've produced 128 integration inputs for ATLAS MC15 samples
  - W→eν+jets
  - Z→ee+jets
  - Z→νν+jets
- This amounted to about 200k hours on Edison
- These jobs are typically 3 nodes each with 24 MPI ranks of Sherpa running for 10 hours
- This uses Sherpa 2.1.1 which we independently validated with ATLAS PMG.

Sherpa Development

- Sherpa integration runs optimally at 200-400 ranks
- Event Generation will be online soon, scales well to 1000s of ranks
- Improving these could give major speedups:
  - Very dependent on shared libraries, makes explicit 'dlopen' calls during run time (HPC unfriendly)
  - File-I/O intensive (single ranks make open calls 1000s times)
  - Large Memory Footprint ~1.5GB/rank (limits us to 8 ranks per Mira node)

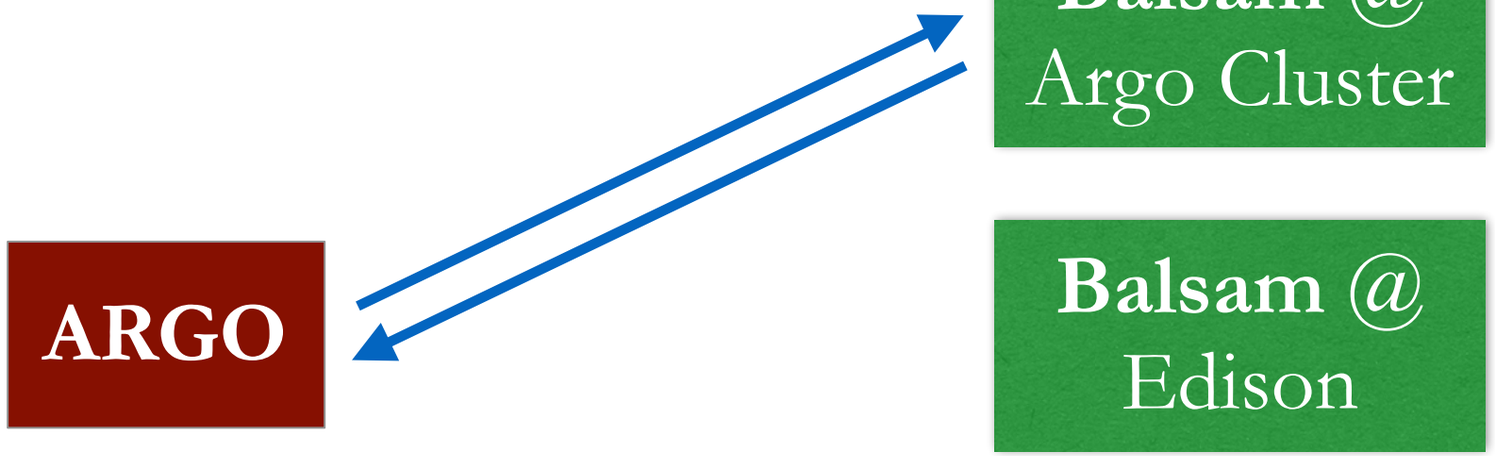


Cooley-nodes (12 ranks/node)

- **Sherpa, unlike Alpgen, has a team of young theorists behind it constantly adding new physics and higher orders.**
- **Our improvements are being deployed in the latest versions.**
- **Did I mention we found a sleep statement in Sherpa?**
- **Sherpa will be with us for a while and is run inside Athena, any improvements we make can be integrated.**

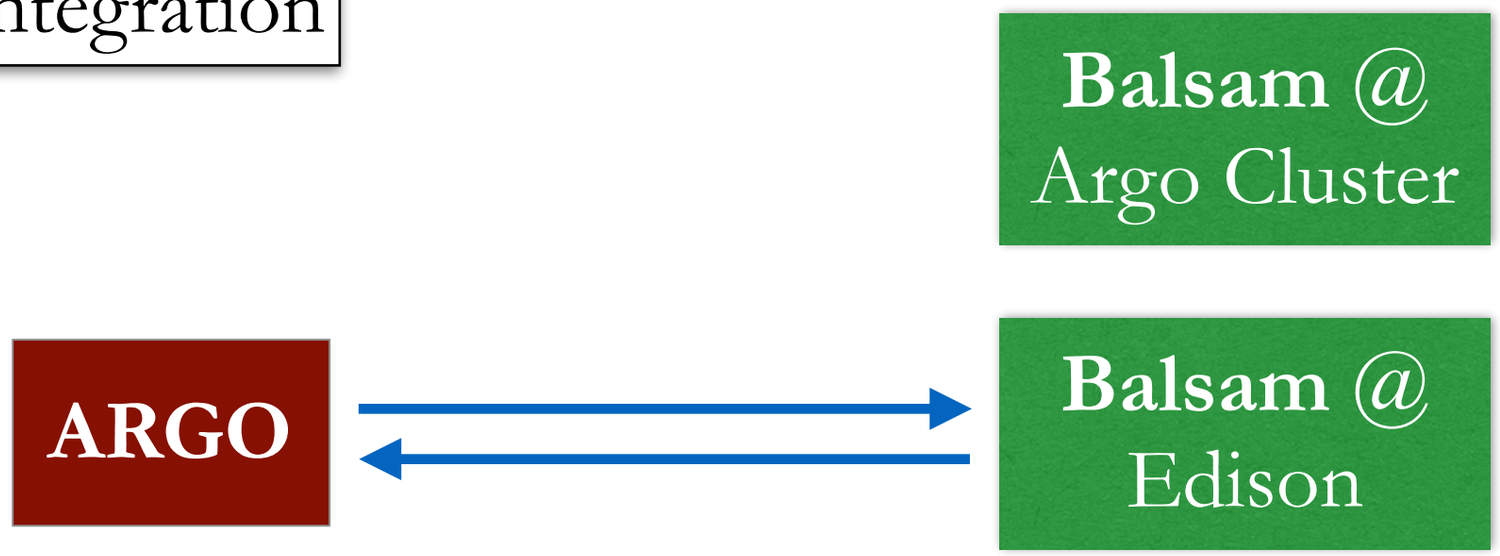
Sherpa Workflow with ARGO+Balsam

Step 1: Generate Code



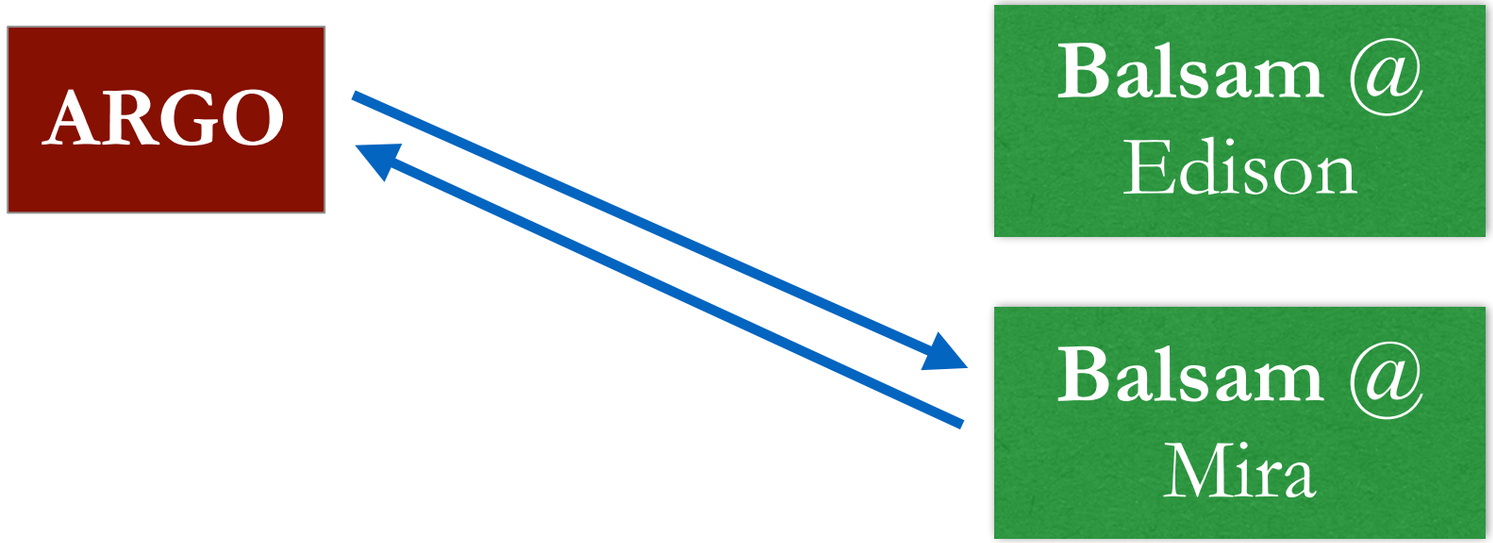
Sherpa Workflow with ARGO+Balsam

Step 2: Generate Integration



Sherpa Workflow with ARGO+Balsam

Step 2: Generate Events

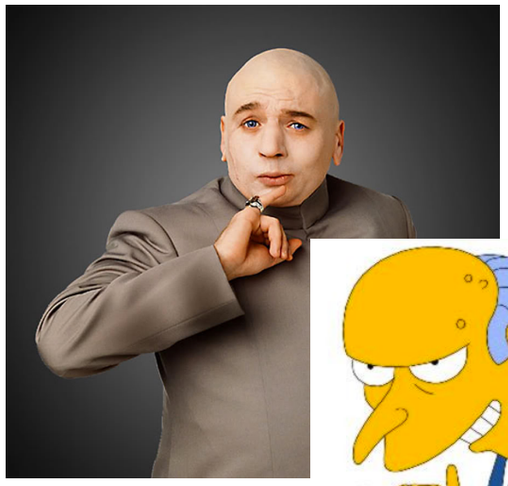


Integration with PanDA

- Danila came to Argonne at the end of August
- He implemented the pilot to talk to ARGO and submit incoming jobs
- Local Tests have been successful (see his talk)
- Tests submitted with PanDA are still having key/cert issues that we are working out
- All tests performed on Argo Cluster since ASCR forbid running Grid jobs on Mira

Integration with PanDA: Security Plan

Typically, users need two-step authentication to access these resources so it is important we are not exposing these resources to extra risks.

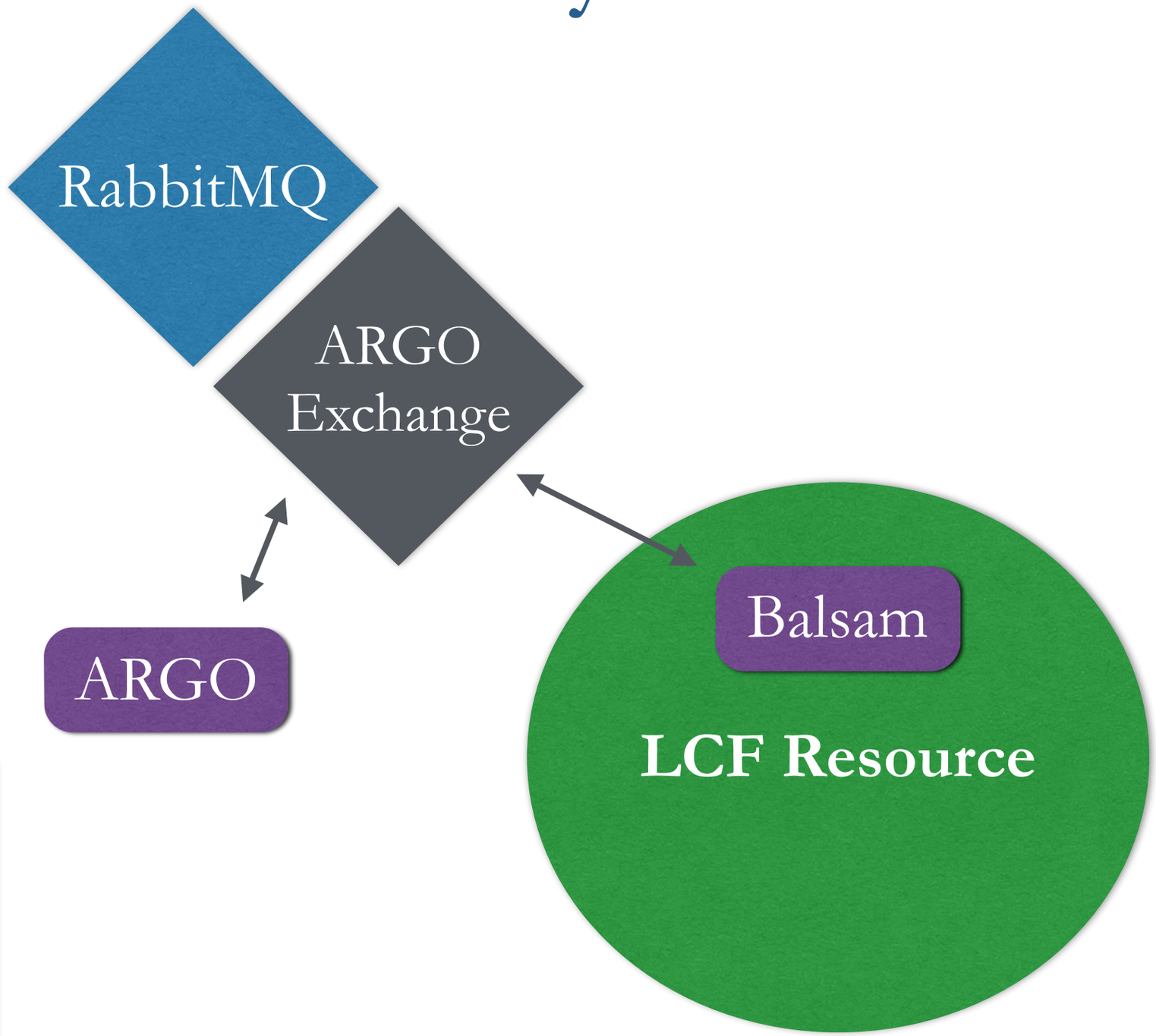


LCF Resource

Integration with PanDA: Security Plan

The Balsam service runs on the resource.

Integration with PanDA: Security Plan



Balsam sends and receives text messages to/from ARGO via the RabbitMQ service running on an outside resource.

Data Transfers

Messages

Integration with PanDA: Security Plan

RabbitMQ

ARGO Exchange

ARGO

Balsam

LCF Resource

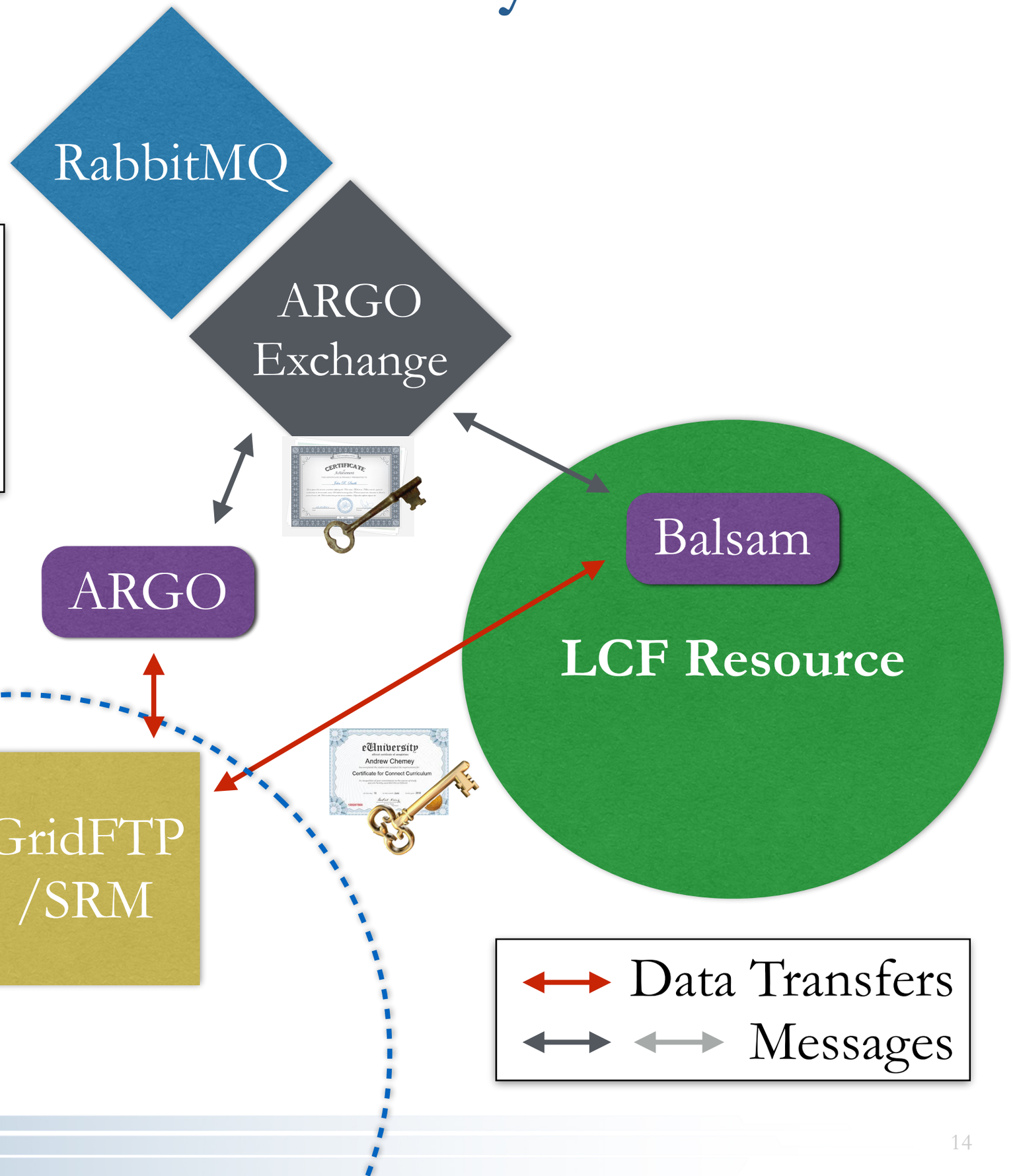
These messages are isolated inside the RabbitMQ "ARGO Exchange".
A Key/Certificate pair is required to send/receive messages in this exchange.
Only ARGO/Balsam can use this exchange.

Data Transfers

Messages

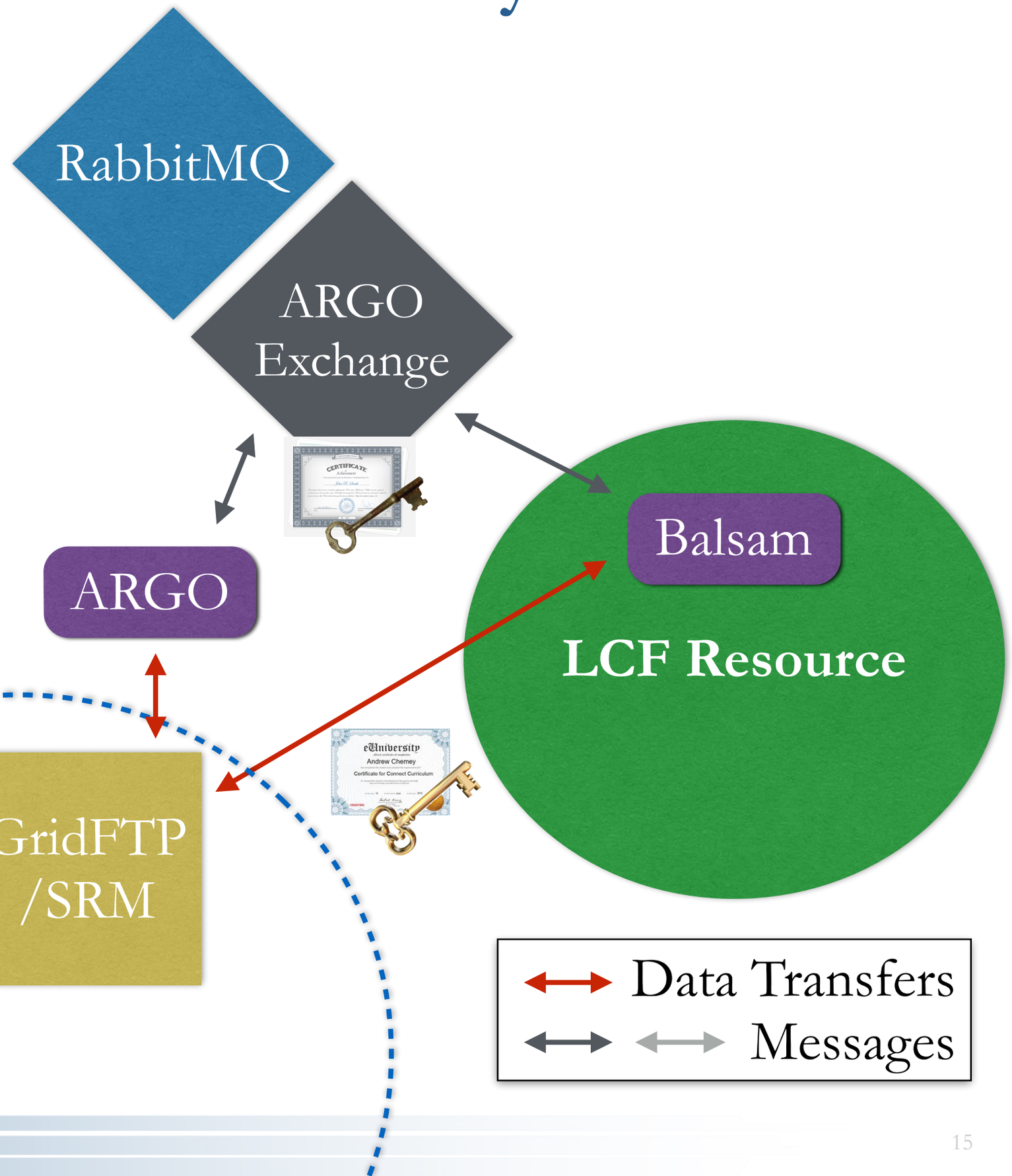
Integration with PanDA: Security Plan

ARGO messages direct Balsam with a GridFTP URL for handling input/ output data.



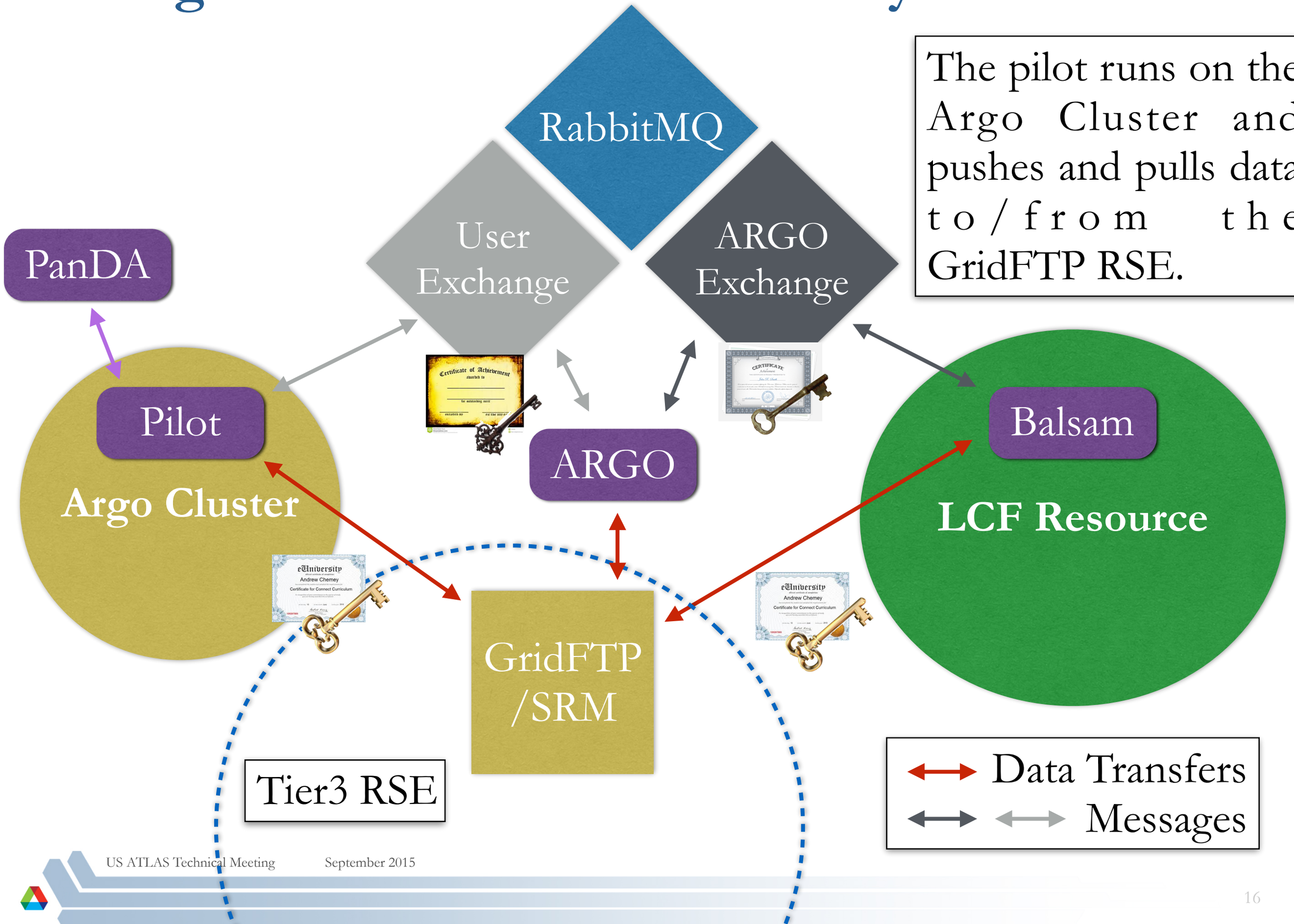
Integration with PanDA: Security Plan

Currently data is stored on a local GridFTP resource that is part of the Tier-3 at Argonne.
This includes the standard authentication scheme meaning one needs another Key/Certificate to upload/download data.



Integration with PanDA: Security Plan

The pilot runs on the Argo Cluster and pushes and pulls data to/from the GridFTP RSE.



Integration with PanDA: Security Plan

The Pilot submits jobs to ARGO through an additional "User Exchange".

RabbitMQ

User Exchange

ARGO Exchange

PanDA

Pilot

Argo Cluster

ARGO

Balsam

LCF Resource

GridFTP /SRM

Tier3 RSE

Data Transfers

Messages

Integration with PanDA: Security Plan

Using a different Exchange with a different key/certificate pair adds another layer of isolation from PanDA/Grid users.

RabbitMQ

User Exchange

ARGO Exchange

PanDA

Pilot

Argo Cluster

ARGO

Balsam

LCF Resource

GridFTP /SRM

Tier3 RSE

Data Transfers

Messages

Integration with PanDA: Security Plan

PanDA

RabbitMQ

User Exchange

ARGO Exchange

Pilot

ARGO

Balsam

Argo Cluster

LCF Resource

GridFTP /SRM

Tier3 RSE

Then there is the usual Grid Certificate authentication scheme.

Data Transfers

Messages

Integration with PanDA: Security Plan

PanDA

RabbitMQ

User Exchange

ARGO Exchange

Pilot

ARGO

Balsam

Argo Cluster

LCF Resource

GridFTP /SRM

Tier3 RSE

In addition, Balsam has a designated Path for binaries. If the binary doesn't exist in this path, the job will not be executed.

Data Transfers

Messages

Athena Building on PowerPC

Athena Building on PowerPC

- Well, before you can run Athena, you must have Gaudi installed…

Athena Building on PowerPC

- Well, before you can run Athena, you must have Gaudi installed…
- Before you can run Gaudi, you must have LCG installed…

Athena Building on PowerPC

- Well, before you can run Athena, you must have Gaudi installed…
- Before you can run Gaudi, you must have LCG installed…
- Meet Cooley: The ALCF Analysis machine
  - GPU-CPU x86 Hybrid
- Decided to begin by compiling everything on a non-Grid x86 before diving into non-x86 AND non-Grid Machine.
- LCGCMake (heptools-78root6) Took about 1 week of work to complete compilation on Cooley using gcc-4.8.1.
- Then began compiling LCGCMake on the Mira development machine Vesta.
- Login nodes are plain old powerpc64, not BGQ architecture, so I started by compiling for the login node, again trying to limit the complications.
- This has proven more complicated. I've made it about 25% (according to CMake) of the way through the build. Many packages need little tweaks here and there.
- When I fix a compile problem, I try to feed that back into the CMake configuration such that I can have a build-able LCGCMake when I'm done.

Athena Building on PowerPC

- The reasons I am bothering with this:
  - Zach is working toward a slimmed version of the Atlas Simulation
  - Vahko (et al.) have been successful deploying Yoda+Event Service on Titan.
- Together, these might make it worth while to run simulation on Mira.
- Also…
- Titan -> Summit (PowerPC) in 2017…

Summary & Final Thoughts

- Alpgen is still taking some of our time, but ATLAS wants it.
- Sherpa optimization is much more complicated
  - We've already identified many things we can do to improve it.
  - It is taking us longer because it is a currently developed code (Alpgen wasn't).
  - The development frequently answers our requests with patches that are then being included in the newest release. Sherpa 2.2.0 already has some of our updates, including the removal of a sleep statement we found!
  - This work is worth it because Sherpa will be around for the rest of the run and the processes get more computationally demanding at NLO.
- Sherpa production has begun on Edison & Mira
- Pilot development and integration is nearing completion (Danila's Talk)
- Security has always been something we have thought about as we build up the Edge Services that are ARGO/Balsam.
- LCG Build on PowerPC has begun.