



# ESCAPE

European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures

## Dress Rehearsal Experiment review : Rubin Observatory / LSST

Bastien Gounon

on behalf of CC-ESCAPE team:

Ghita Rahal, Fabio Hernandez, Paul Musset, Gino Marchetti, Adrien Georget

# Rubin/LSST Plans for the DR

- Goal: ingest data at realistic LSST data rates
  - Expected data rates: ~20TB/day, which translates to ~230MBps
- Dataset: HSC RC2 from the Subaru Telescope
  - 45248 x 18MB FITS images (~800GB)
  - No replication
  - CC-IN2P3's dCache backend
  - davs:// protocol



# Execution environment

- Initial plan

- Use CC-IN2P3 batch farm
- Use a shared Python venv for Rucio client configuration
- Run ~100 single-core ingestion jobs in parallel

=> Issue with the batch farm right before FDR starts \o/

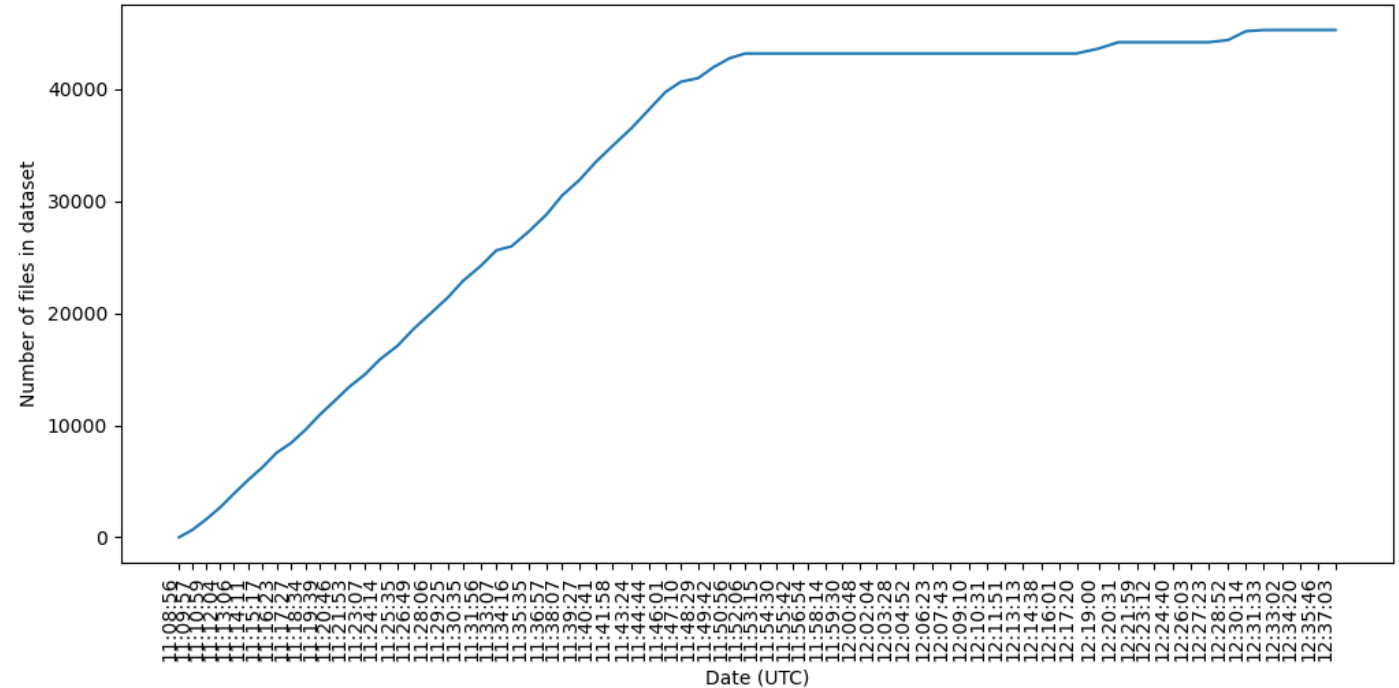
- Backup plan

- Use a Singularity container to package Rucio client configuration
- Run 64 « interactive » ingestion jobs split among 4 LSST DTNs



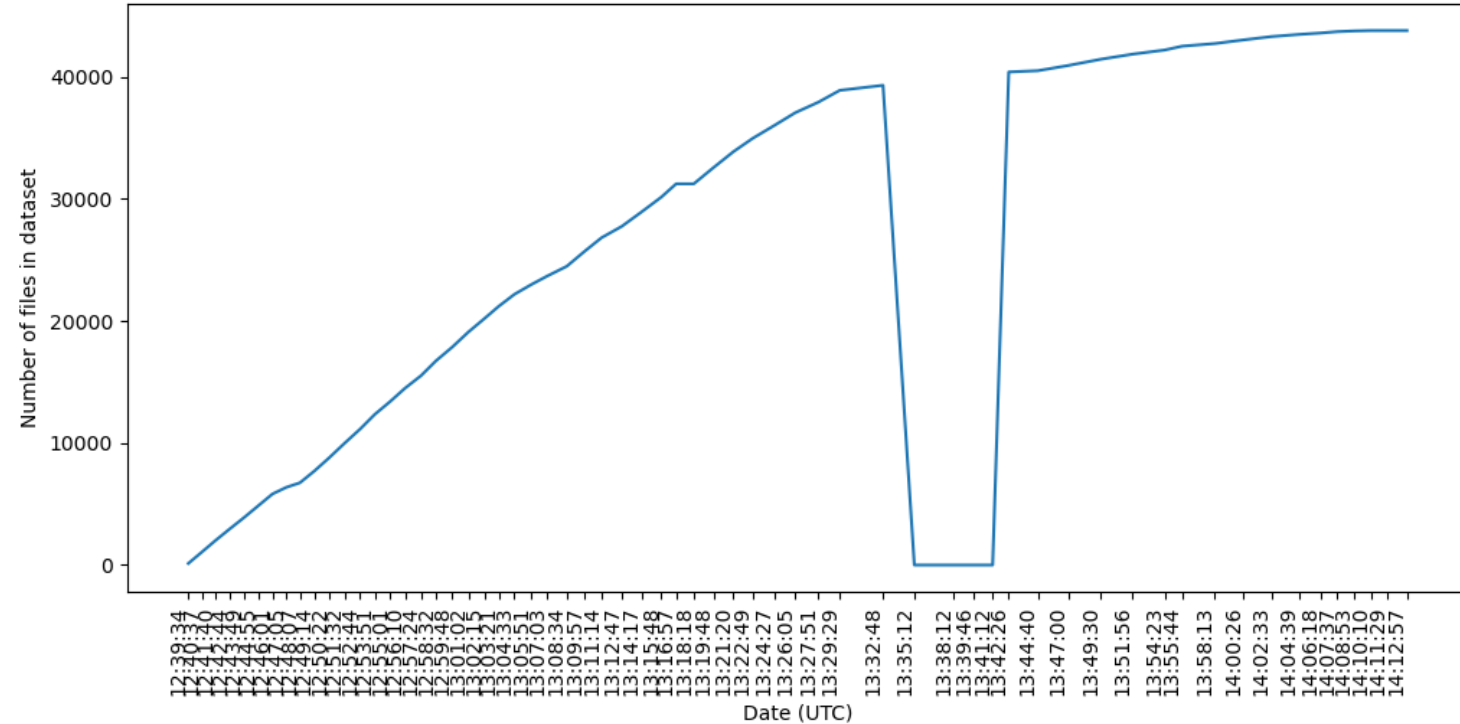
# Details from Run1

- 43151 successful uploads
- 2097 failed uploads
- small number of very long-runners
- 95% of the dataset uploaded after 42min, but almost 150min to completion



# Details from Run2

- 43811 successful uploads
- 1437 failed uploads
- drop linked to a crash on the K8s platform
- slow uploads after coming back up, maybe high load
- 97% of the dataset uploaded after 90min



# Details from Run2

```
ConnectionError: ('Connection aborted.', RemoteDisconnected('Remote end closed connection without response',))
```

```
ConnectionError: HTTPSConnectionPool(host='escape-rucio.cern.ch', port=32300): Max retries exceeded with url: /dids/LSST_CCIN2P3_GOUNON-test/HSC-0036808-075.fits_013/meta?plugin=DID_COLUMN (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x7f4d66d06128>: Failed to establish a new connection: [Errno 111] Connection refused',))
```

```
ConnectionError: HTTPSConnectionPool(host='escape-rucio.cern.ch', port=32300): Max retries exceeded with url: /dids/LSST_CCIN2P3_GOUNON-test/HSC-0036808-075.fits_013/meta?plugin=DID_COLUMN (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x7f4d4cb0c860>: Failed to establish a new connection: [Errno 111] Connection refused',))
```

```
2020-11-17 15:35:12,178 ERROR Cannot connect to the Rucio server.
```



# Details from Run2

2020-11-17 15:16:53,922 DEBUG utils.py

\_\_call\_\_

The requested service is

not available at the moment.

Details: An unknown exception occurred.

Details: TRANSFER ERROR: Copy failed with mode streamed, with error: HTTP 500 :  
Unexpected server error: 500

2020-11-17 15:16:54,533 WARNING Upload attempt failed

2020-11-17 15:16:54,533 INFO Exception: RSE does not support requested operation.

Details: The requested service is not available at the moment.

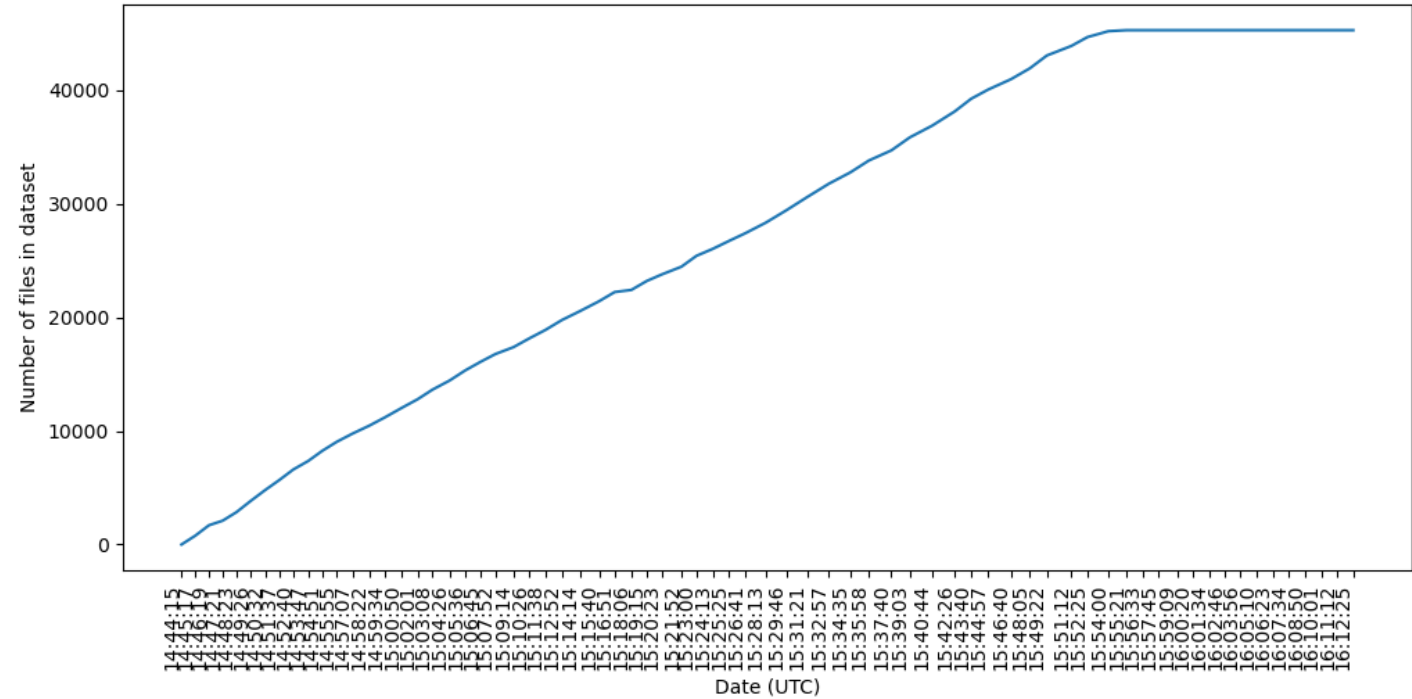
Details: An unknown exception occurred.

Details: TRANSFER ERROR: Copy failed with mode streamed, with error: HTTP 500 :  
Unexpected server error: 500



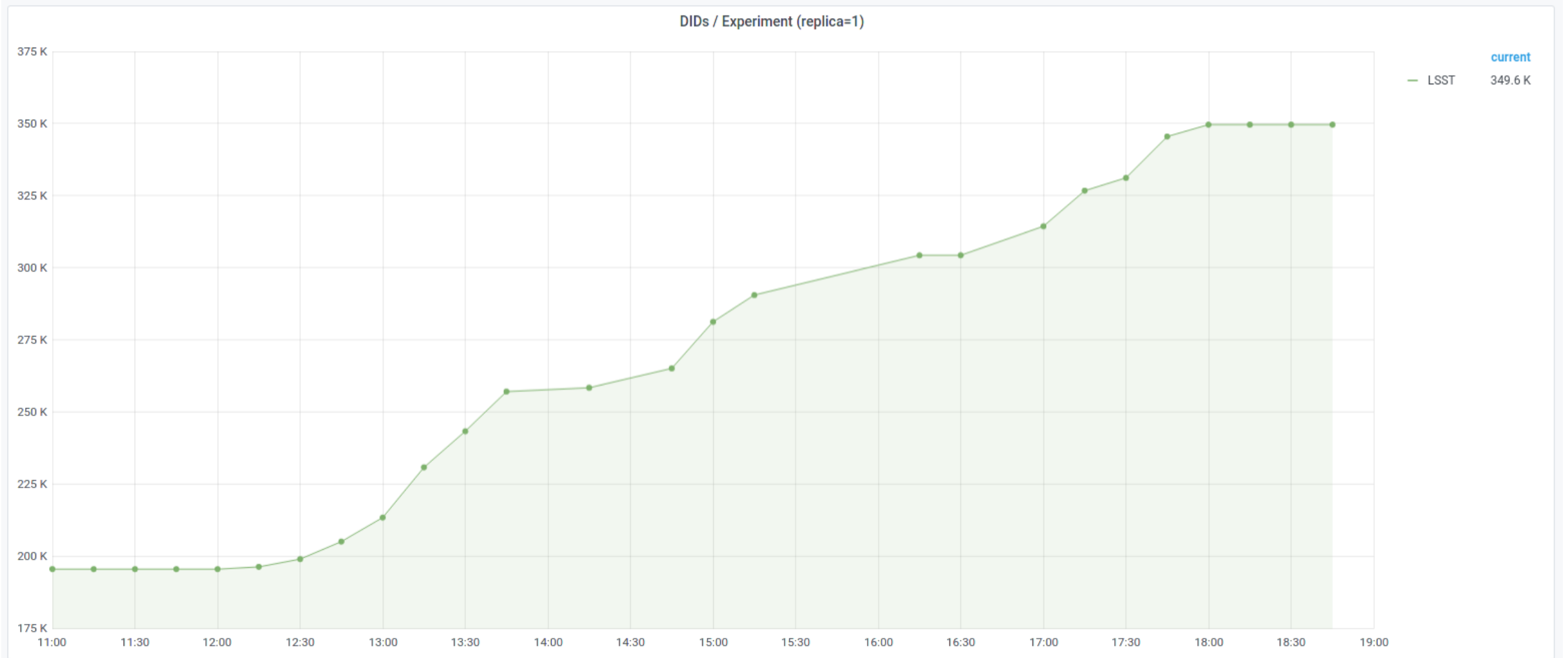
# Details from Run3

- 45248 successful uploads
- 0 failed uploads
- 100% of the dataset uploaded after 68min





# Overall Results

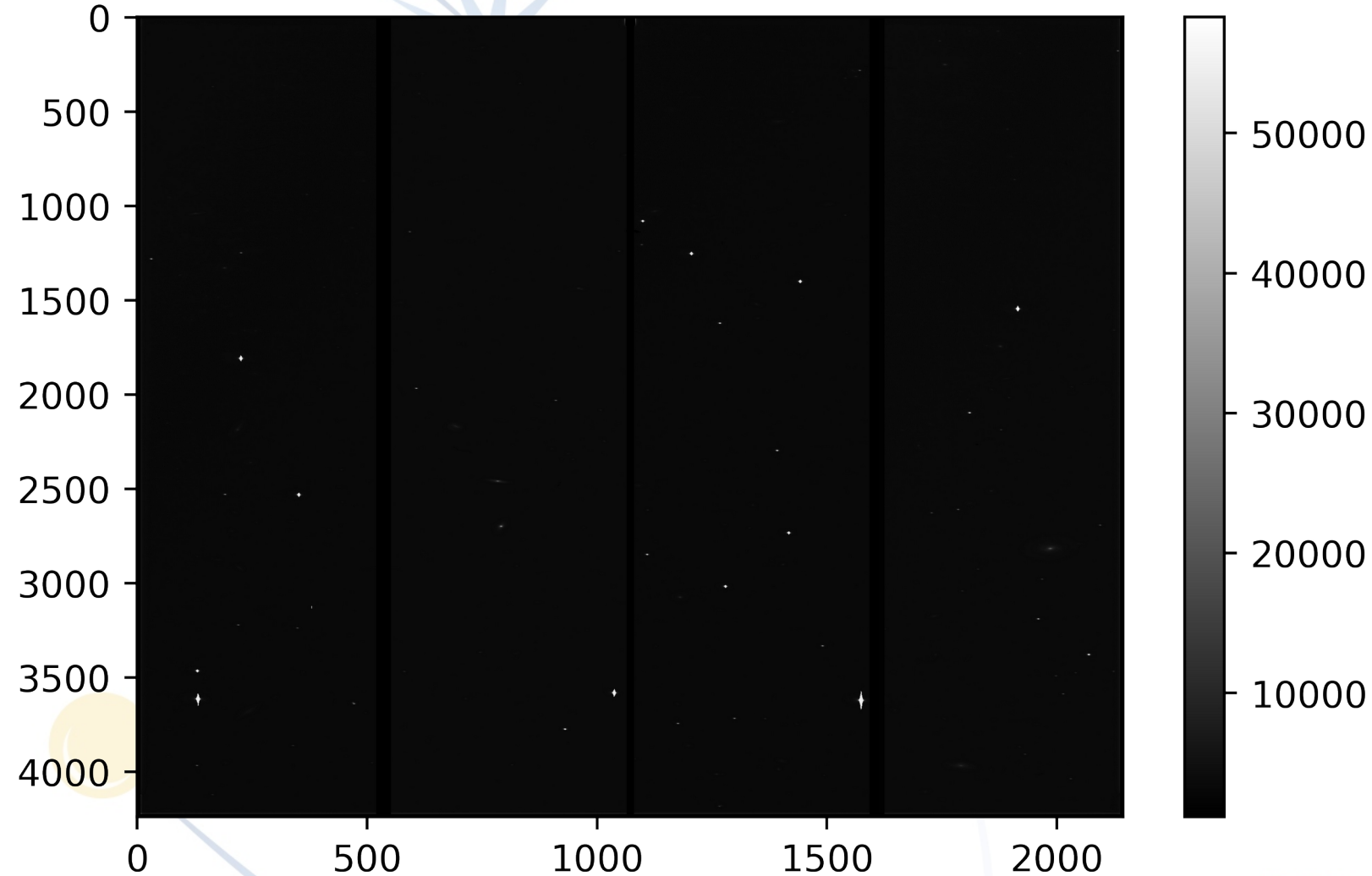


# Overall Results



# Overall Results

- Image from Run3
- Drawn within a Jupyter notebook, using Rucio Python API to get data



# Conclusion

- Although we had to change our data ingestion mechanism at the last minute, tests were successful
  - ~150K files and >2TB uploaded in total
  - Sustained, realistic data rates of over 300MBps were achieved
  - Singularity overhead for the rucio client was minimal, not an issue
  - Error rates stayed below 5%, and all uploads succeeded after a second try
  - We demonstrated uploaded data was viable for scientific analysis
- => This was a successful Dress Rehearsal from our point of view !



# Thank you !

# Questions ?

