# Rucio @ ATLAS

*Mario.Lassnig@cern.ch*

**4th Rucio Community Workshop**
2021-09-28 to 2021-10-01
https://indico.cern.ch/event/1037922/

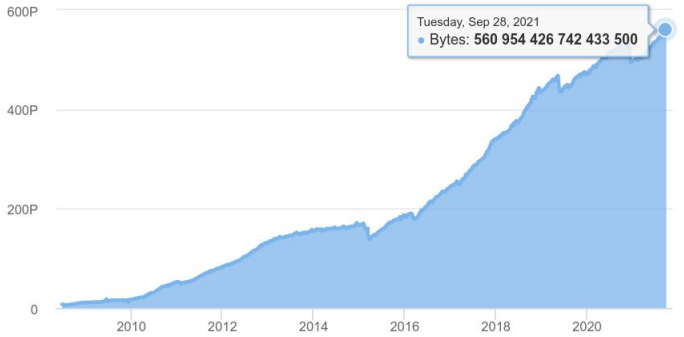
# Overview

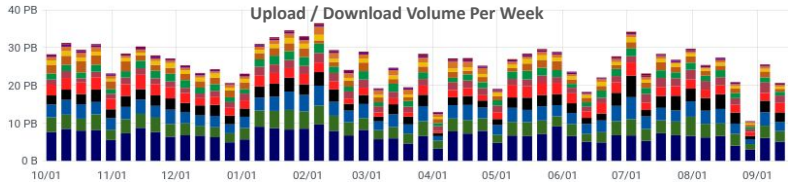
## Rucio is working very well for ATLAS

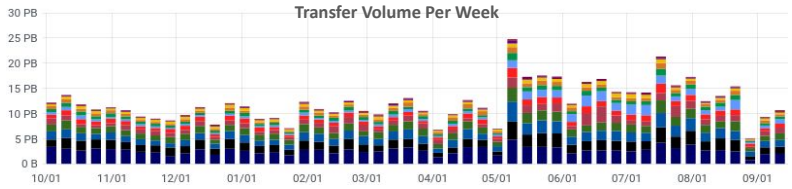
- 1B+ files, 500+ PB of data, 400+ Hz interaction
- 350+ RSEs including HPC & cloud
- 500 Petabytes/year transferred & deleted
- 2.5 Exabytes/year uploaded & downloaded

## Operational challenges

- Drastically reduced operations personpower
- Constant struggle for disk space / occupancy
  - Storage is practically always full, with only 10% margin
  - Forced data deletion based on "lifetime model"
- Large site decommissioning and consolidation effort
  - Highly time-consuming manual process involving the sites
  - Often reveals inconsistencies and other problematic states
  - Working towards automating this as much as possible



# Major things

Most of the ATLAS-specific work in the last year has been on ensuring LHC Run-3 readiness
- More deployment and commissioning work than development

Getting Data Carousel in production (*cf. Paul's talk*)
- Continuous exchange of ideas with PanDA WMS and DPA
- Development, deployment, and testing

Oracle upgrade from 11c to 19g
- New setup for unit tests and CI/CD (Oracle is not happy about distributing its database)
- Three step production upgrade (AWR cleanup, DBMS upgrade, Optimizer upgrade)

Move from AGIS to CRIC, CASTOR to CTA, Puppet to Kubernetes, …

Protocol transition

Participation in many R&D projects, many already geared towards HL-LHC / DOMA

# The move to Kubernetes

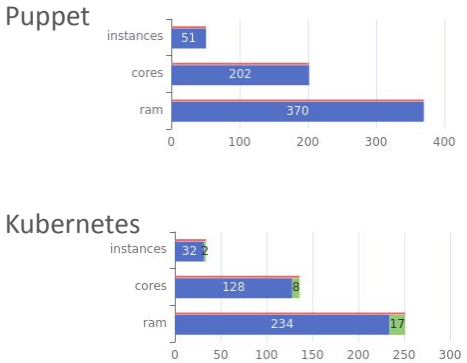
Hosted on CERN OpenStack cloud infrastructure

Puppet-managed VMs, with sometimes *dubious* manifests
- Handcrafted hostname exceptions, configuration inconsistencies,
- *Puppet run* vs. *Service restart*

Decommission service on Puppet, move quota to K8S

Cluster setup for high availability
- *atlas-rucio-prod-01*, *atlas-rucio-prod-02*, *atlas-rucio-int-01*
- Anticipate Kubernetes multi-master for added resilience

Configuration centralised with helm and flux, thanks to Gitlab with a straightforward WebIDE
- Uses standard published images
- Added support for pod hotpatching



# Protocol transition

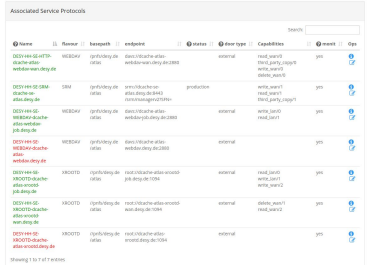
Move away from SRM and gsiftp to HTTP-TPC

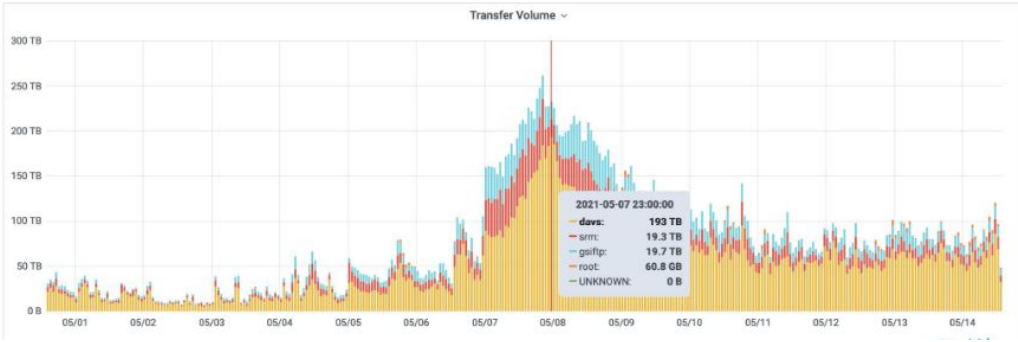
- Lots of commissioning effort and configuration updates in Rucio
- Dedicated Rucio conveyor instances for ongoing testing

By now, sans TAPE and a straggler site, ATLAS has fully moved to HTTP-TPC

ROOT will stay as data access protocol

Your typical complicated CRIC configuration



# Protocol transition

## For tape, we are taking a two-step approach

**Step 1** Enable SRM+HTTPS

**Step 2** New RESTful tape interaction API

## Actual protocol transition procedure

Setting up dedicated areas and endpoints at the sites, then configuring RSEs in CRIC

Major testing campaign ongoing with 100TB data samples

Site ticketing needed because operators need to check flush status, etc

## So far, it looks good™

No further development required on Rucio side

We believe we can have all ATLAS tape sites migrated before the end of the year

# Cache-aware brokering (Virtual Placement)

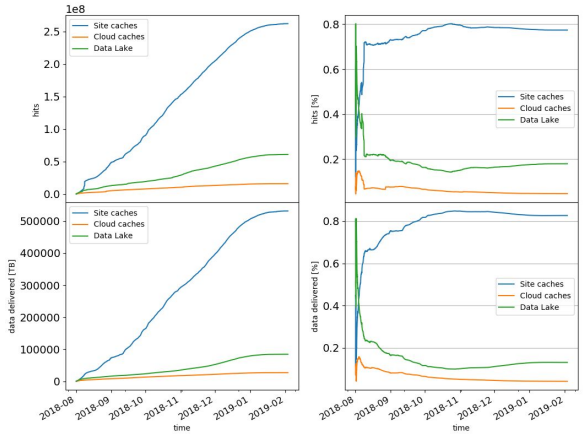
Demonstrate benefits of having small high-performance caches at ATLAS sites

Set up appropriate sized caches (100+ TB) at 7 sites that send heartbeats to dedicated VP service

VP receives heartbeats and content information from caches

When PanDA makes a request to find the location of a file

- *rucio/core/replica* —> augments the reply with "virtual location"
- Gives more flexibility for PanDA to schedule
- Job will retrieve file from cache not disk
- If statistics work out well, integrate feature into Rucio core



# Data Challenges

## Full-stack data transfer challenge, not only network challenge

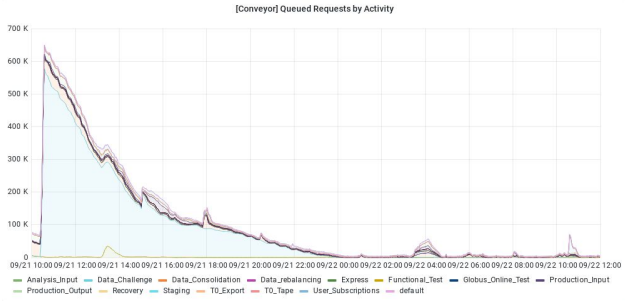
Series of bi-annual incremental tests (from 10% in 2021 to 100% in 2027) to get us to HL-LHC scale

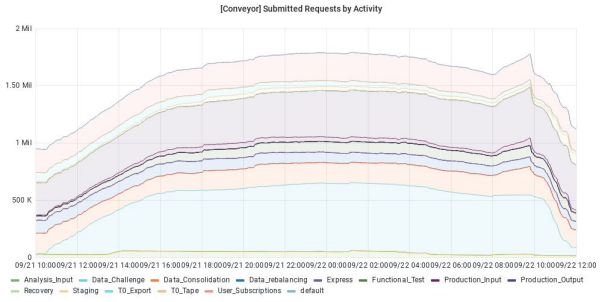
Running the challenges on our production infrastructure forces us to commission storage and network R&D

## Central "data challenge" framework

CMS & ATLAS benefit from common Rucio interface & FTS monitoring

## Mock challenge to test framework



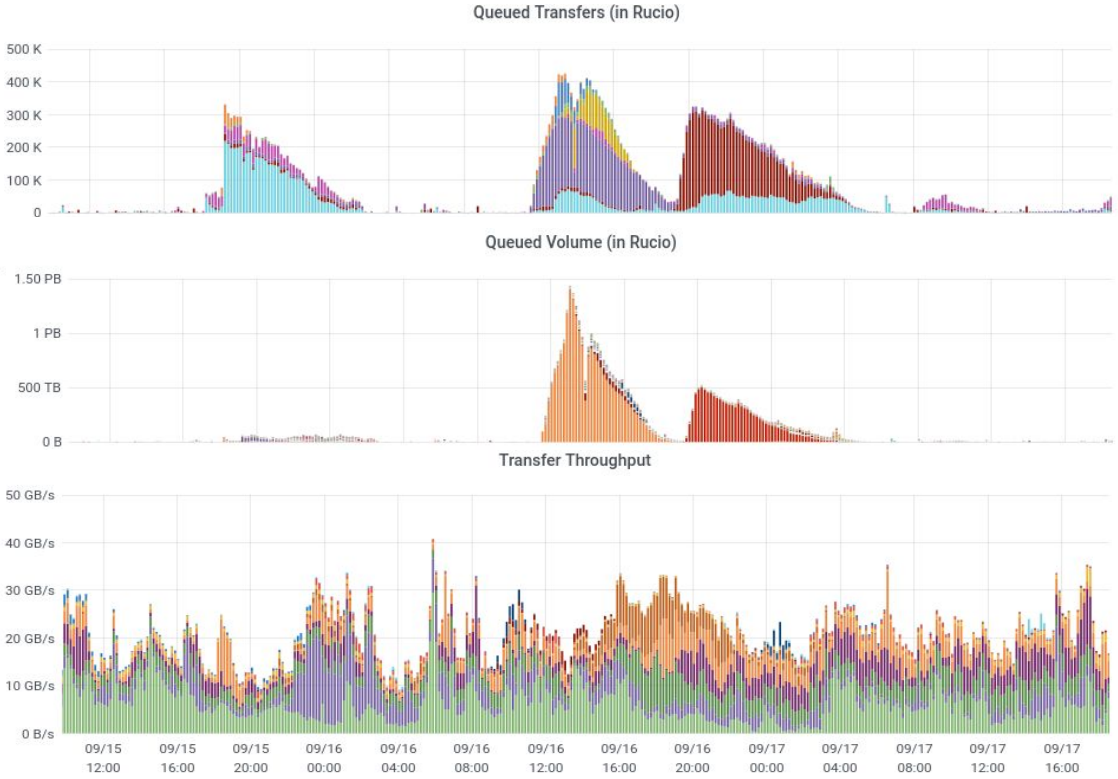

| T1 | Data Challenge target 2027 (Gbps) | Data Challenge target 2025 (Gbps) | Data Challenge target 2023 (Gbps) | Data Challenge target 2021 (Gbps) |
|---|---|---|---|---|
| CA-TRIUMF | 98 | 59 | 29 | 10 |
| DE-KIT | 312 | 187 | 94 | 31 |
| ES-PIC | 93 | 56 | 28 | 9 |
| FR-CCIN2P3 | 281 | 169 | 84 | 28 |
| IT-INFN-CNAF | 336 | 202 | 101 | 34 |
| KR-KISTI-GSDC | 25 | 15 | 7 | 2 |
| NDGF | 71 | 43 | 21 | 7 |
| NL-T1 | 94 | 56 | 28 | 9 |
| NRC-KI-T1 | 62 | 37 | 19 | 6 |
| UK-T1-RAL | 296 | 177 | 89 | 30 |
| RU-JINR-T1 | 52 | 31 | 15 | 5 |
| US-T1-BNL | 227 | 136 | 68 | 23 |
| US-FNAL-CMS | 454 | 273 | 136 | 45 |
| (atlantic link) | 681 | 408 | 204 | 68 |
| | | | | |
| Sum | 2400 | 1440 | 720 | 240 |

# Sharing data flow orchestration across experiments?

**By activity**
***(filtered)***

**By site**
***(filtered)***

**By activity**



# Summary

ATLAS is very happy with Rucio

Last many months were operationally intense

- Space occupancy, site decommissioning, protocol transition, deployment of R&Ds, …
- There were a lot of challenges ;-)
- Continuous improvement of our operational automations

Within Rucio, there were rather few ATLAS-specific developments

- Many are already community-driven and/or community-needed
- Strong focus on source code improvements and housekeeping instead

The R&D projects for HL-LHC will drive further ATLAS developments