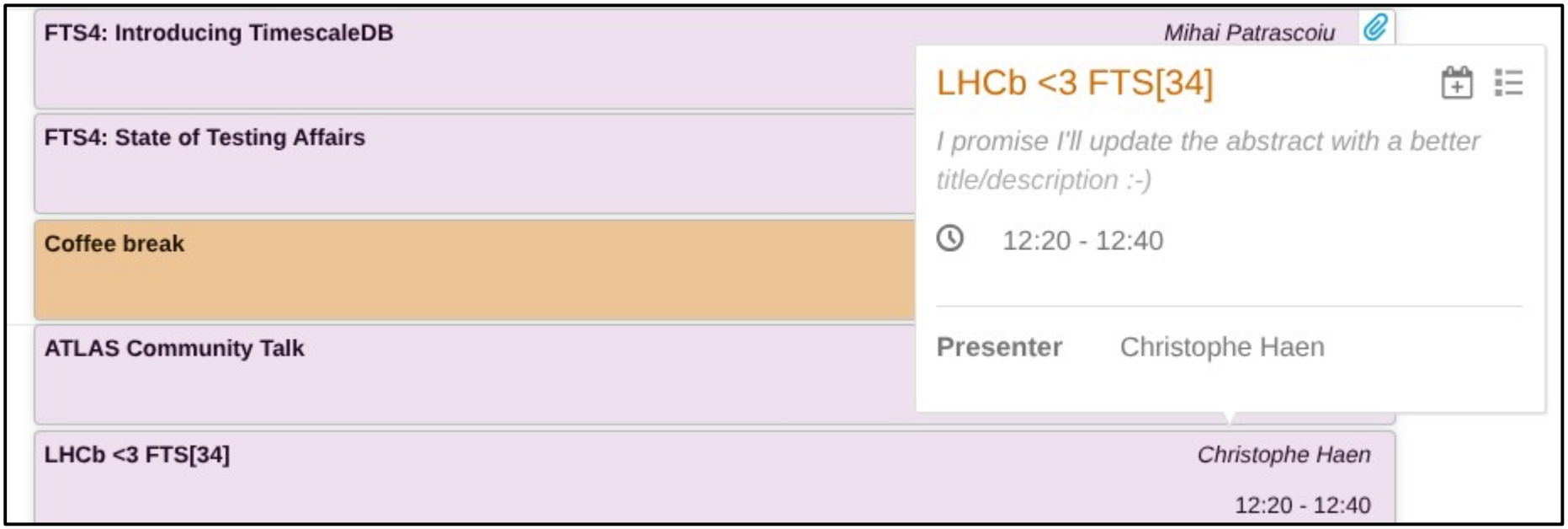
FTS & XRootD Workshop: LHCb

15/09/2026

Christophe Haen

Sorry Maria !


FTS4: Introducing TimescaleDB
Mihai Patrascoiu
FTS4: State of Testing Affairs
Coffee break
ATLAS Community Talk
LHCb <3 FTS[34]
Christophe Haen
12:20 - 12:40
LHCb <3 FTS[34]
I promise I'll update the abstract with a better title/description :-)
12:20 - 12:40
Presenter
Christophe Haen

By now, you know the LHCb drill

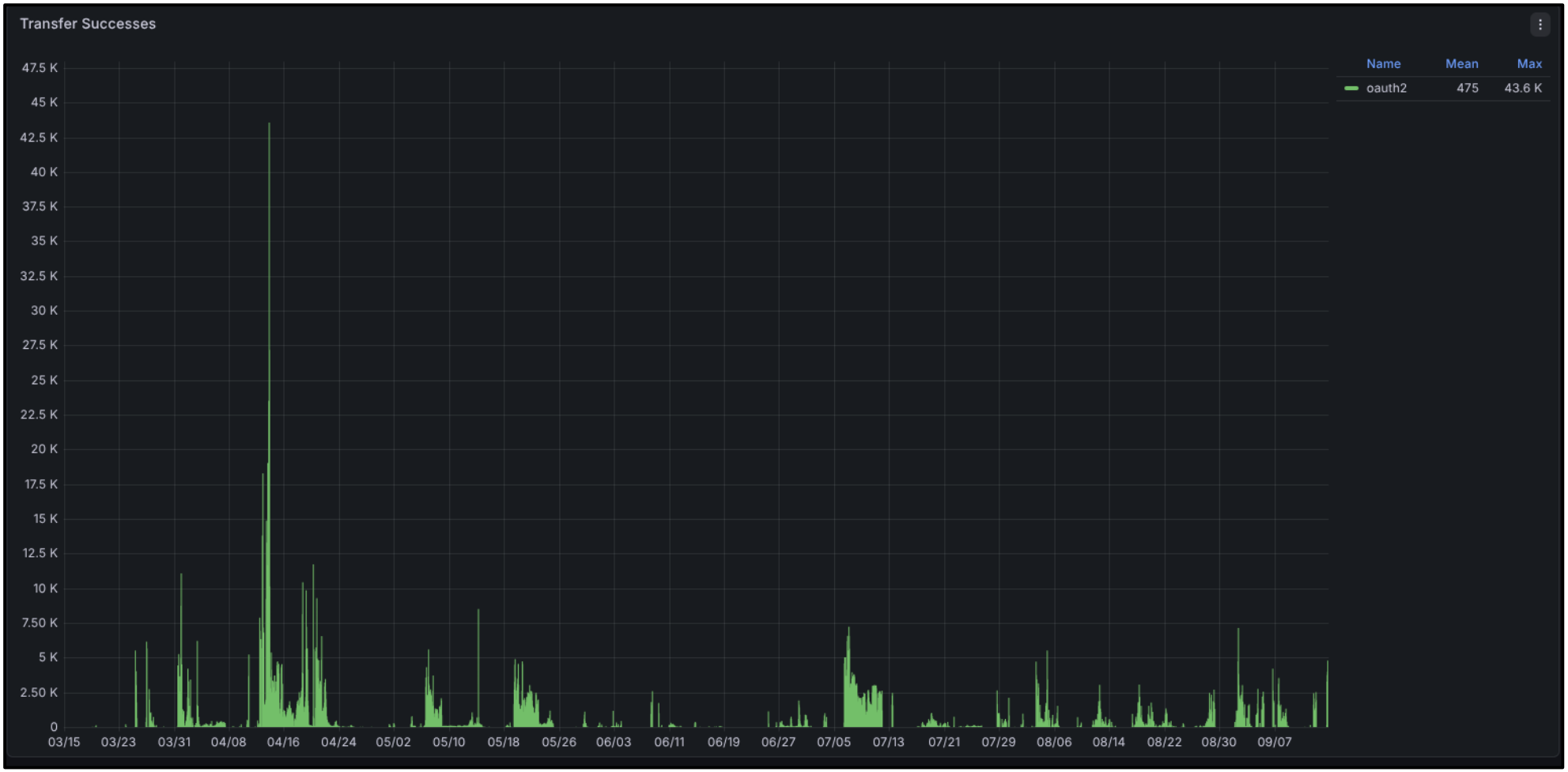
- All Third Party Copy transfers of LHCb go through FTS
  - Transfers dominated by Real Data flows
  - Jobs upload data to their final destination (i.e. no rebalancing campaign)
- Favor LAN over WAN for jobs and staging
- HTTPs absolutely everywhere for TPC
- Everything orchestrated by DIRAC (and diracx in the future)
- As usual, follow the KISS principle
  - Little to no use of “fancy features”
- Reliability, performance & stability are our main requirements

This is what I aim for

Submitted	Active	Staging	S.Active	Archiving	Finished	Failed	Cancel	Rate (last 1h)
221463	1632	0	0	20866	38093	0	0	100.00 %

No artificial intelligence was harmed achieving this. Just dummy humans.

Tokens for all disk disk transfers



Token transfers

- Per file token for both source and destination
- Credential flow
- Long lived (5 days)
- IAM high availability required
- So far so good !

New since last time

- Added scitag support
  - One to One mapping with the activity.
  - Not actually sure what to do with it so far :-)
- Started relying more on the “retryable” flag
  - Very useful
  - Varying degrees of success
  - *The code is the documentation* ™

```
86
87      // If the following strings appear in the error
88      const char *msg_imply_do_not_retry[] = {
89          "proxy expired",
90          "with an error 550 File not found",
91          "File exists and overwrite",
92          "No such file or directory",
93          "SRM_INVALID_PATH",
94          "The certificate has expired",
95          "The available CRL has expired",
96          "SRM Authentication failed",
97          "SRM_DUPLICATION_ERROR",
98          "SRM_AUTHENTICATION_FAILURE",
99          "SRM_NO_FREE_SPACE",
100         "digest too big for rsa key",
101         "Can not determine address of local host",
102         "Permission denied",
103         "System error in write into HDFS",
104         "File exists",
105         "checksum do not match",
106         "CHECKSUM MISMATCH",
107         "The source file is not ONLINE",
108         NULL
109     };
110
111     if (findSubstring(message, msg_imply_do_not_ret
112         return false;
```

FTS4 shopping list

- Structured configuration file with API
  - YAML file with Pydantic model stored on a gitlab repo pulled by all the VO FTS instance ?
- Reproducer of a transfer:
  - Helps me and sites debug
- Smooth API transition for submission/monitoring
- Conditional deletion of multi-hop intermediate file
- Helm chart integration ?
- File based / metadata based search ?

From last year's workshop

- Structured configuration file with API
  - YAML file with Pydantic model stored on a gitlab repo pulled by all the VO FTS instance ?
- Conditional deletion of multi-hop intermediate file

- Could even help share FTS configuration between VO !
- From Xuantong talk !

- **Request To FTS Dev Team (Alternatives):**

  A: If a multihop intermediate is also a normal transfer's destination, could the jobs share that transfer?

  OR B: With a dedicated relay directory, could a switch delete the intermediate file after hop 2 succeeds?

DIRAC Follow-Up: after hop 1 failed, downstream files stayed **NOT_USED** and were missed by DIRAC retries. Reported to DIRAC for a fix; not an FTS issue.

- Helm chart integration ?

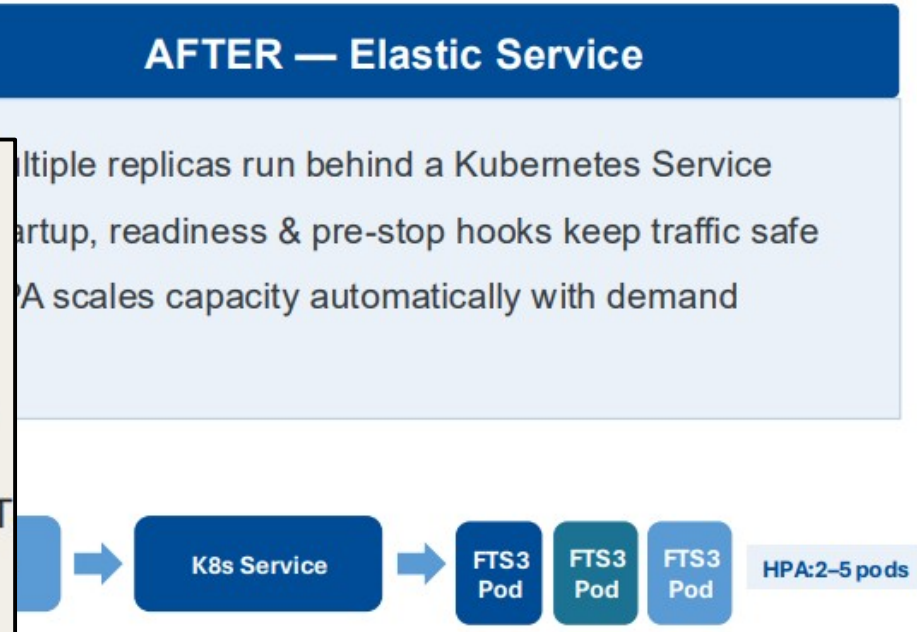
```
$ kubectl --context usdf-fts3 -n fts3 get pod
NAME                          READY   STATUS    RESTA
fts3-activemq-667989f59c-szxxk  1/1     Running   13
fts3-mon-7989476bc9-sq6w9       4/4     Running   0
fts3-rest-fd799b6b9-7mb6c       4/4     Running   5 (21
fts3-server-5dd5967f7-9t6bv     1/1     Running   0
fts3-server-5dd5967f7-hvnt8     1/1     Running   0          43d   ← Easy t
fts3-server-5dd5967f7-wxsrt     1/1     Running   0          43d

$ kubectl --context usdf-fts3 -n fts3-db get pod
NAME                  READY   STATUS    RESTARTS   A
fts3-db-cluster-solo-0  1/1     Running   0          38d
```

- Kubernetes deployment?

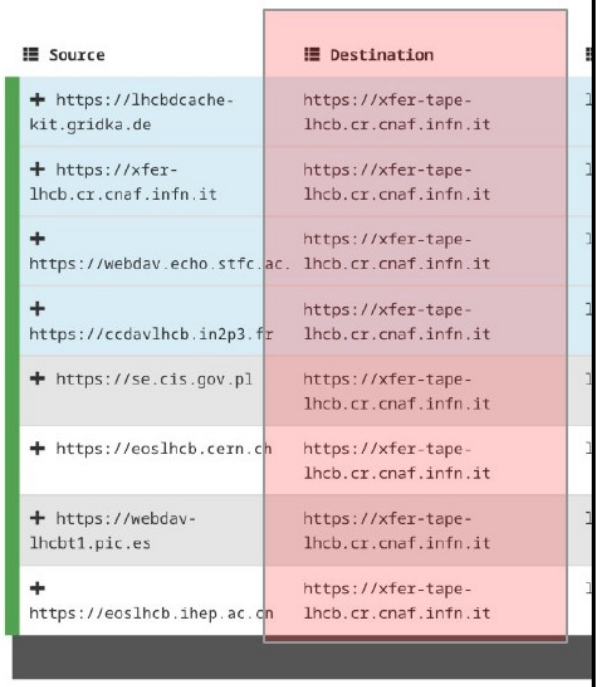
Science and Technology Facilities Council

- FTS + Indigo IAM hosted at STFC RAL
  (See Rose's talk for more details)
  - IAM to migrate to AWS shortly, with SKAO control
  - **No immediate planned changes for FTS:**
    - **Any centralised deployment will be k8s / helm**
- Integration environment already has helm charts for FT
  Used for developer testing



Another old request…

What would make my life a lot easier: diagnostics



Source	Destination
https://lhcbdcache-kit.gridka.de	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://xfer-lhcb.cr.cnaf.infn.it	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://webdav.echo.stfc.ac.	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://ccdavlhcb.in2p3.fr	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://se.cis.gov.pl	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://eoslhcb.cern.ch	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://webdav-lhcbt1.pic.es	https://xfer-tape-lhcb.cr.cnaf.infn.it
https://eoslhcb.ihep.ac.cn	https://xfer-tape-lhcb.cr.cnaf.infn.it

What would make my life a lot easier: diagnostics

- Having some levels of grouping would help
  - Source / Destination
  - Specific files
  - Storage nodes involved
  - Error cleanup

```
Resolved url: eoslhcb.cern.ch => eoslhcb-ns-ip563.cern.ch
Resolved url: se.cis.gov.pl => se.cis.gov.pl

   RemoteConnections: tcp:[2001:1458:301:6a::100:10]:8443
url
```

VO	Count	Reason
lhcb	14	TRANSFER [5] TRANSFER ERROR: Copy failed (3rd pull). Last attempt: Transfer failure: Error when interacting with local filesystem: Failure when closing file handle: Unable to store file - file has been cleaned because the stored file does not match the provided targetsize /eos/lzu/lhcb/lhcb/LHCb/Collision25/SL.DST/00294967/0005/00294967_00050800_1.sl.dst; input/output error (code=5)
lhcb	13	TRANSFER [5] TRANSFER ERROR: Copy failed (3rd pull). Last attempt: Transfer failure: Error when interacting with local filesystem: Failure when closing file handle: Unable to store file - file has been cleaned because the stored file does not match the provided targetsize /eos/lzu/lhcb/lhcb/LHCb/Collision25/QEE.DST/00294787/0000/00294787_00000733_1.qee.dst; input/output error (code=5)
lhcb	13	TRANSFER [5] TRANSFER ERROR: Copy failed (3rd pull). Last attempt: Transfer failure: Error when interacting with local filesystem: Failure when closing file handle: Unable to store file - file has been cleaned because the stored file does not match the provided targetsize /eos/lzu/lhcb/lhcb/LHCb/Collision25/IFT.DST/00294787/0000/00294787_00000998_1.ift.dst; input/output error (code=5)
lhcb	13	TRANSFER [5] TRANSFER ERROR: Copy failed (3rd pull). Last attempt: Transfer failure: Error when interacting with local filesystem: Failure when closing file handle: Unable to store file - file has been cleaned because the stored file does not match the provided targetsize /eos/lzu/lhcb/lhcb/LHCb/Collision25/SL.DST/00294967/0004/00294967_00049437_1.sl.dst; input/output error (code=5)
lhcb	13	TRANSFER [5] TRANSFER ERROR: Copy failed (3rd pull). Last attempt: Transfer failure: Error when interacting with local filesystem: Failure when closing file handle: Unable to store file - file has been cleaned because the stored file does not match the provided targetsize /eos/lzu/lhcb/lhcb/LHCb/Collision25/B2OCLOW.DST/00294967/0005/00294967_00050717_1.b2oclow.dst; input/output error (code=5)

Monitoring is trendy !

Consolidate it ?

Make it user facing ?

Details Matter

Building up monitoring and ability to cross-correlated across multiple sources

Large scale transfers; several hours; correlating monitoring across sources.
Telemetry from the intra-transfer perfmarker outputs in monitoring would be great.

Single File transfer progress

Checksum stage timeout - under investigation

FTS – Monit (single) point

Transfer Throughput

300 MB/s
200 MB/s
100 MB/s
0 B/s

FTS Transfer log scraping

HTTPs transfer throughput — 2.13 TB in 24.0 h (avg 0.197 Gbps)

ESP->JP

FTS Transfer Monitoring System at IHEP

FTS Monitoring Flow

FTS3
fts-activemq Publisher

↓

ActiveMQ
DCI Container · Topics

↓

MonPipe
IHEP-Developed STOMP Consumer

↓

Elasticsearch
Computing Center Shared Service

↓

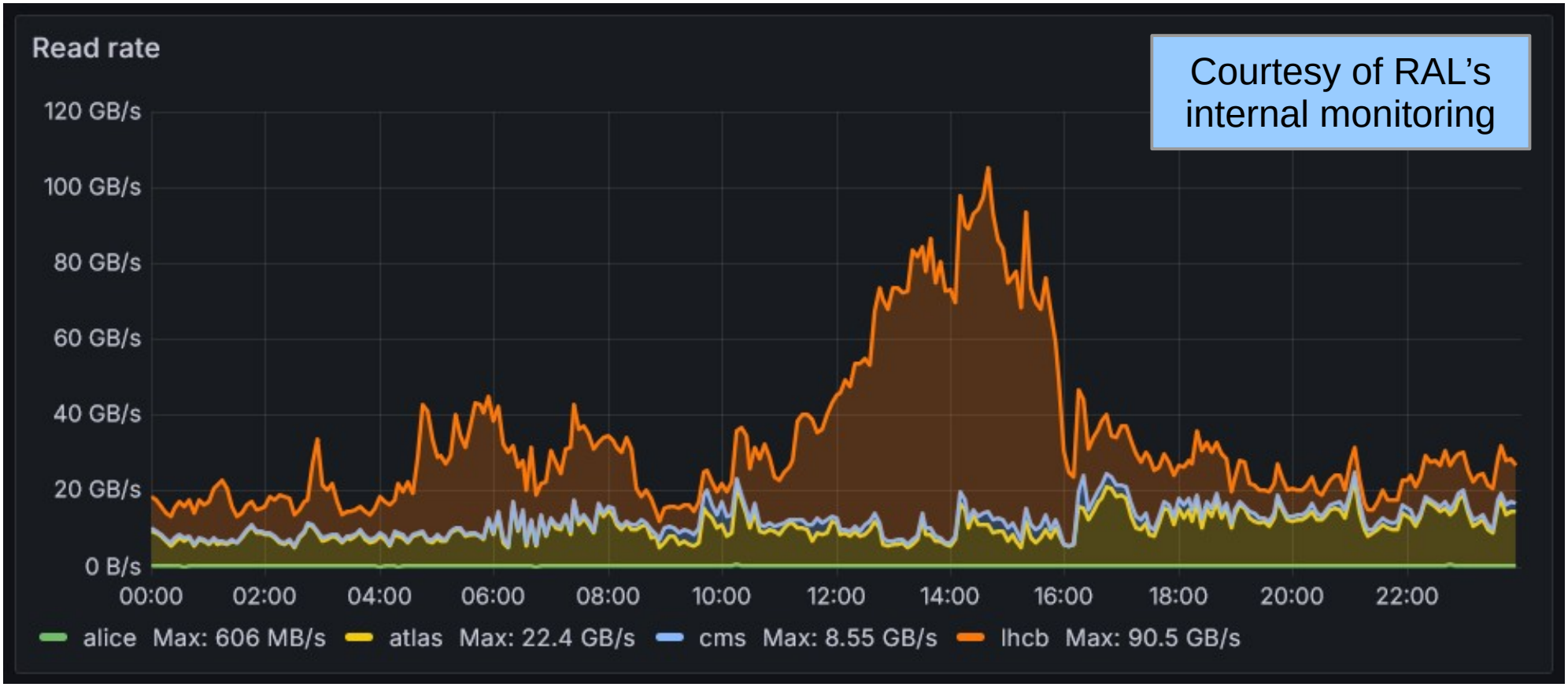
Grafana
IHEP-Designed Dashboards

From monitoring to operational intelligence

- **Improve observability**
  - Identify token-authenticated transfers by **experiment**
  - Expose **source and destination transfer errors**
  - Add custom **OpenTelemetry metrics**
  - Visualize through **Grafana**
- **AI-assisted operations**
  - Collect logs, metrics, incidents, and maintainer knowledge
  - Build a **RAG-based operational knowledge base**
  - Help operators investigate anomalies and previous incidents

Eager to know about FTS4

LHCb load comes from the jobs !



Two modes of running

- The pilot downloads the files, the payload uses the local file
  - Done by DIRAC via gfal2 (https/xroot)
  - Latest versions (currently gfal 2.23.5, xrootd 6.1.1)
- The payload reads directly on the storage via remote read
  - LAN only
  - Any xrootd version down to v3 !!
- Bonus mode: anything goes with interactive use

LHCb input at the last xrootd OTF

LHCb Input

- The list of requests is more or less the same as the one presented in the last XRootD meeting:
  - Distribute wheels for xroot python bindings
  - Support gfal replacement
    - A good tool for manual tests, reproducing errors, etc.
    - Helps handling common IO operations via unified interface
  - Have a better monitoring
    - E.g. saving requests information on the client-side in a machine-readable format
  - Support Async IO in python bindings
  - Ensure backward compatibility of the clients where possible
    - Not always the case (e.g. clients v{5.1..5.3} are not compatible with TLS-aware xrootd servers due to a bug)
  - Dynamic timeouts for (Vector) read operations?

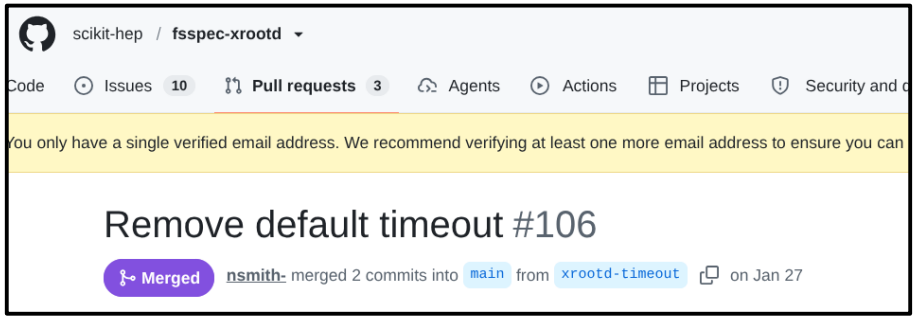
Updates

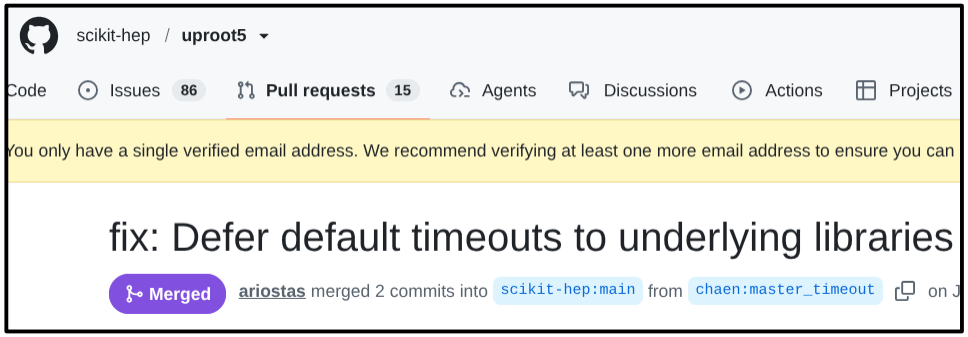
- Thanks for the wheels !
- Better client monitoring
  - Will hopefully test soon the Recorder plugin
  - Is the initialization order fixed ?
- Status of dynamic timeout for vector read ?

Debugging an XrootD client timeout

- Context: timeout problems remote-reading files at CERN from the UK
- OSError: Failed to open file: [ERROR] Operation expired
- XRD_LOGLEVEL,XRD_REQUESTTIMEOUT, XRD_STREAMTIMEOUT, …

```
root_file = uproot.open({auth(path): None}, **{"timeout": 1800})
```





Clients really are important…

- Python as first class citizen
  - fsspec is defacto in the picture
- Async is a must today
- A gfal like tool is still very much needed
  - binding/native for DIRAC/FTS/Rucio
  - CLI for sites
- Hiding everything in the xrootd client is probably not the right place

What is the xrootd file popularity ?

- Seen in a xrootd monitoring talk
  - supported by CMS ?
- Could be interesting, but what would it actually look like ?

Conclusions

- FTS & XRootD are absolutely paramount to LHCb operations
- We are very happy with it, even more with the people and the support they provide
- Maintaining this level of support and quality on the long term is essential for LHCb but also many other experiments