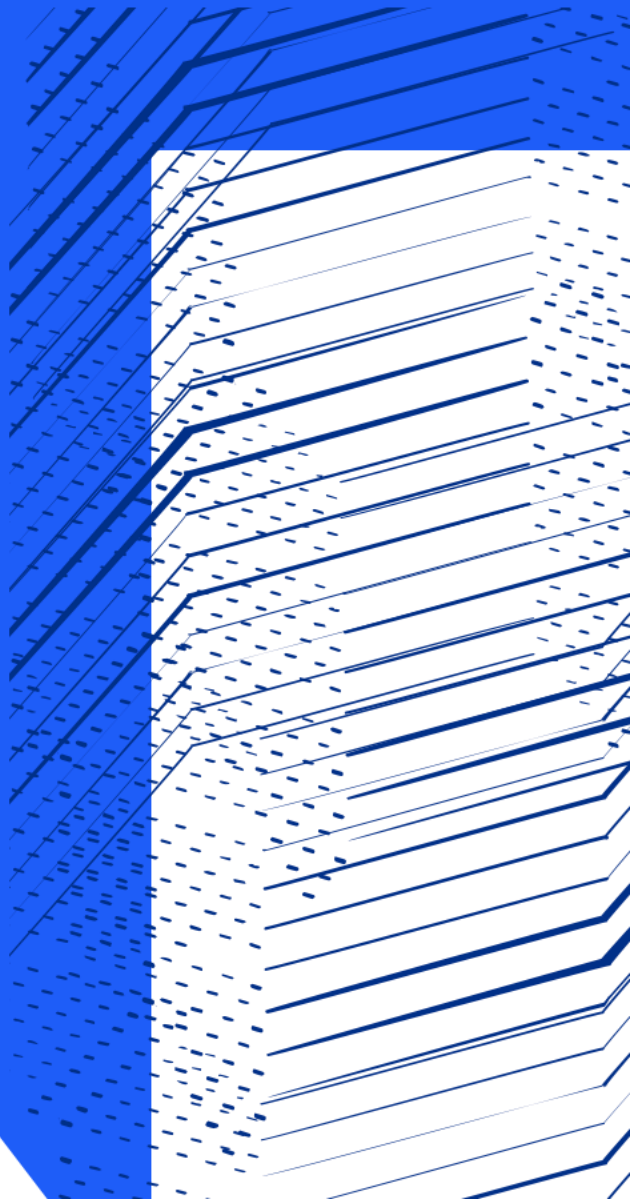
Science and Technology Facilities Council

# FTS as a part of the SKA data movement pipeline

Rose Cooper
rose.cooper@stfc.ac.uk



# Square Kilometer Array (SKA)

- Large scale radio telescope
- Expected to produce data on the exobyte scale
- Will utilize globally distributed storage and compute infrastructure

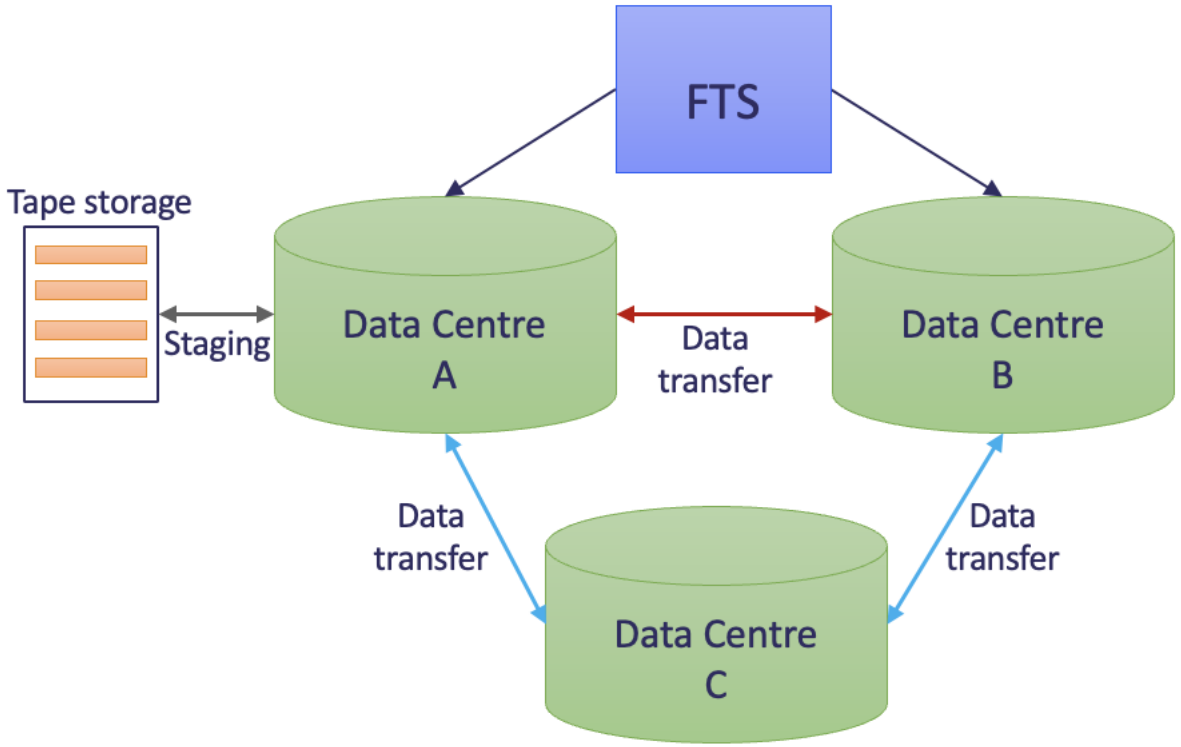
*https://www.skao.int/en/about-us/91/history-ska-project*

Also see talks by Ian Collier (24/10 13:30) and James Walder (24/10 16:45)

Science and Technology Facilities Council

# What is the FTS?

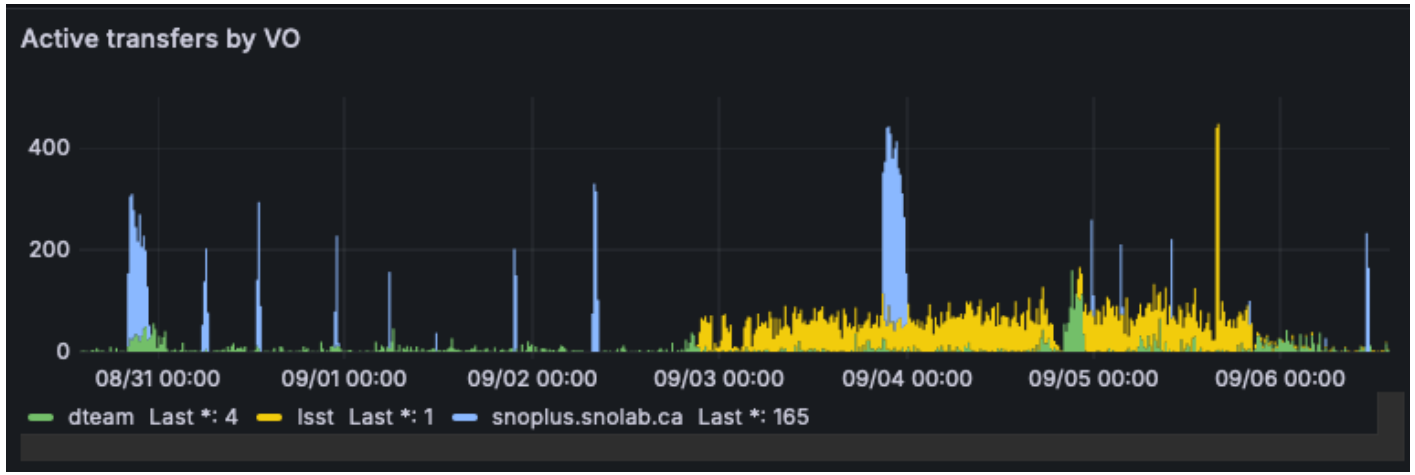
- Bulk data movement service
- Efficiently schedules data transfers, maximizes use of available network & storage resources whilst respecting any limits
- Developed at CERN
- Critical for globally distributing for WLCG

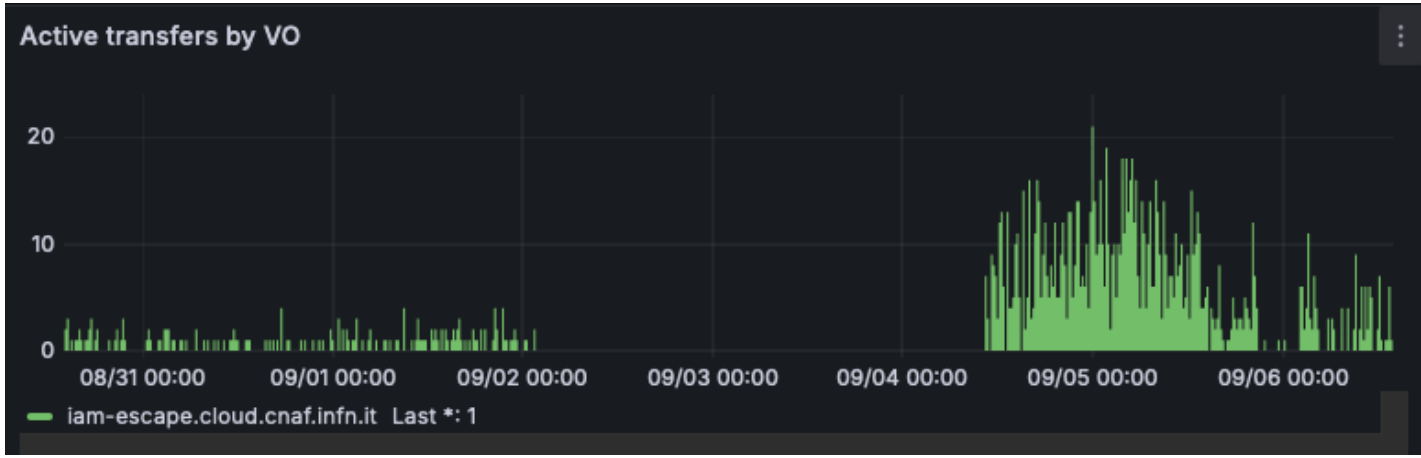



Science and Technology Facilities Council

# Overview of FTS at RAL

- Primarily used by snoplus, dteam, CMS, LSST and SKA
  - Separate instance for SKA
- Average 1-1.5M files per week
- LSST submit large volumes of smaller jobs

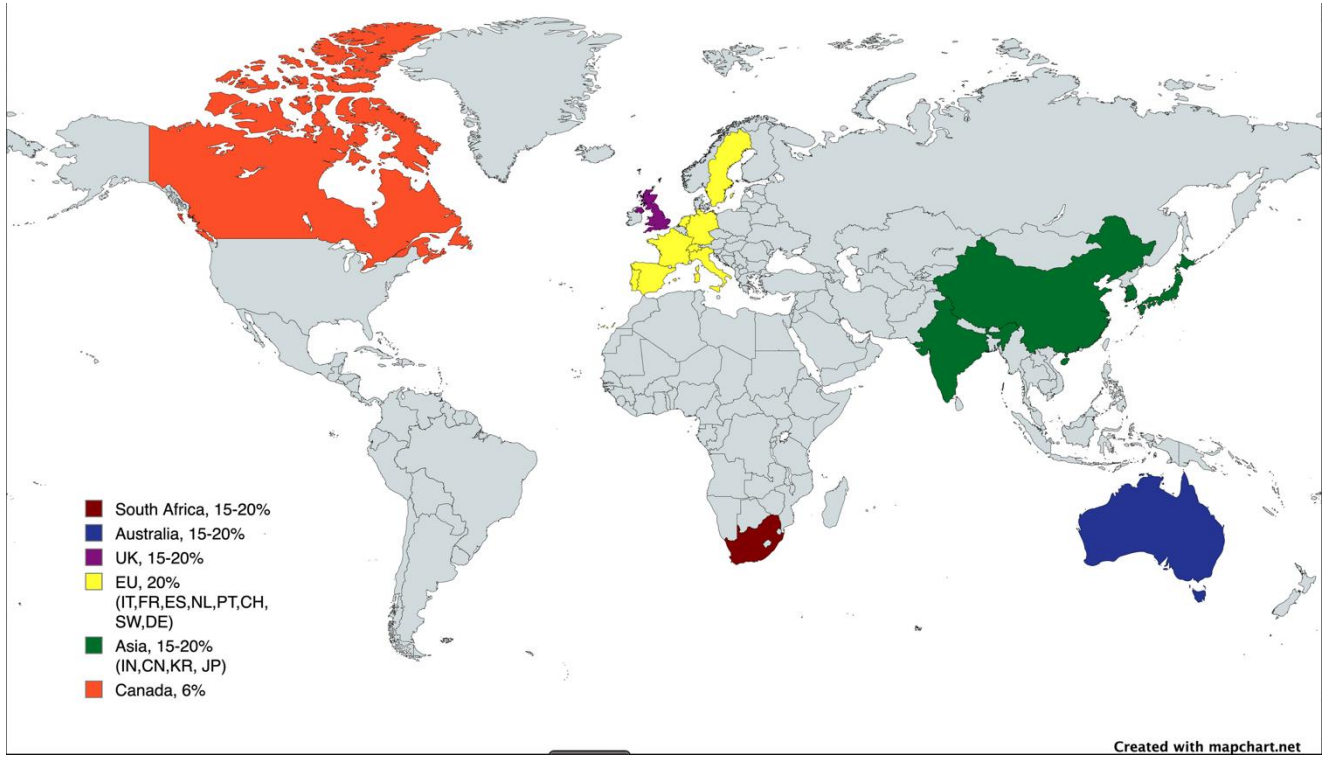



Science and Technology Facilities Council

# Data movement considerations

- Expecting potentially up to 1 petabyte a day
- Will require two copies of each data product within the system
- Potential for data products to be very large




Science and Technology Facilities Council

# Data management software stack selection

- Investigated using Rucio & FTS, as well as storage inventory

- Highlighted several differences:

| Rucio | Storage Inventory |
| --- | --- |
| Centrally managed data replication rules | Locally managed data subscriptions |
| Scalable across multiple storage systems and transfer protocols | Designed with astronomical data discovery in mind |

Science and Technology Facilities Council

# FTS and SKA

- Rucio & FTS have been selected as one of the data movement tools for SKA
- Currently run a prototyping/preproduction FTS instance at RAL for SKA
  - Will transition to production
- Exclusively using tokens, not X.509 certificates

Science and Technology Facilities Council

# Deploying FTS for SKA

- Repurposed existing test hosts to be used by SKA
  - Utilize existing infrastructure
- Integrated hosts with SKA-IAM prototype
  - Required hosts use latest FTS version 3.13
  - Also included migration to Rocky 9
- Integrated with SRCNet Rucio instance
  - Had previously been using CERN pilot FTS & escape IAM

Science and Technology Facilities Council

# Regular data movement tests with SKA-FTS

| Source | Destination | VO | Submitted | Active | Staging | S.Active | Archiving | Finished | Failed | Cancel | Rate (last 1h) | Thr. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| davs://webdav.grid.surfsara. | davs://canfar.shao.ac.cn | ska-iam.stfc | – | – | – | – | – | – | 25 | – | 0.00 % | – |
| davs://storm.srcdev.skao.int | davs://storm.srcdev.skao.int | ska-iam.stfc | – | – | – | – | – | 4 | – | – | 100.00 % | – |
| davs://storm.srcdev.skao.int | davs://xrootd01.uksrc.rl.ac. | ska-iam.stfc | – | – | – | – | – | 4 | – | – | 100.00 % | – |
| davs://storm.srcdev.skao.int | davs://dcachetest.grid.surfs | ska-iam.stfc | – | – | – | – | – | 5 | – | – | 100.00 % | – |
| davs://storm.srcdev.skao.int | davs://xrootd.dev.skach.org | ska-iam.stfc | – | – | – | – | – | 4 | – | – | 100.00 % | – |
| davs://storm.srcdev.skao.int | davs://webdav.grid.surfsara. | ska-iam.stfc | – | – | – | – | – | 6 | – | – | 100.00 % | – |
| davs://storm.srcdev.skao.int | davs://xrootd-01.swesrc.chalmers.se | ska-iam.stfc | – | – | – | – | – | 4 | – | – | 100.00 % | – |
| davs://xrootd.dev.skach.org | davs://webdav.grid.surfsara. | ska-iam.stfc | – | – | – | – | – | 5 | – | – | 100.00 % | – |

Science and Technology Facilities Council

# Token integration with monitoring

| File ID | File State | File Size | Throughpu | Remainin | Start Time | Finish Time | Staging Start | Staging End | Archiving Start | Archiving End | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| + 29112684 | FINISHED | 976.56 KiB | 1.03 MiB/s | – | 2024-10-17T12:31:2 | 2024-10-17T12:31:3 | – | – | – | – | Log |

davs://xrootd-01.swesrc.chalmers.se:80/data/testing_functional/d1/93/1000KB_171024T12.31.16

davs://tank-04.ira.inaf.it:80/space5/xrootd/deterministic/testing_functional/d1/93/1000KB_171024T12.31.16

```
INFO    Thu, 17 Oct 2024 13:31:29 +0100; Setting Gfal2 configuration: DEFAULT_COPY_MODE=3rd pull
INFO    Thu, 17 Oct 2024 13:31:29 +0100; Transfer accepted
INFO    Thu, 17 Oct 2024 13:31:29 +0100; Source token: {"aud":["fts","https://wlcg.cern.ch/jwt/v1/
any"],"exp":1729168477,"iat":1729164877,"iss":"https://ska-iam.stfc.ac.uk/","nbf":1729164877,"scope":"offline_access profile
storage.create:/ storage.modify:/ storage.read:/ wlcg.groups","wlcg.ver":"1.0"}
INFO    Thu, 17 Oct 2024 13:31:29 +0100; Destination token: {"aud":["fts","https://wlcg.cern.ch/jwt/v1/
any"],"exp":1729168477,"iat":1729164877,"iss":"https://ska-iam.stfc.ac.uk/","nbf":1729164877,"scope":"offline_access profile
storage.create:/ storage.modify:/ storage.read:/ wlcg.groups","wlcg.ver":"1.0"}
INFO    Thu, 17 Oct 2024 13:31:29 +0100; VO: ska-iam.stfc.ac.uk
```

Sign in with your IdP

Last time you choose the following IdP:

**Science and Technology Facilities Council**

Proceed?

Sign in with IdP

Search again

Back to login page

Science and Technology Facilities Council

# Data movement challenges for SKA

- Currently in the process of determining what tests would be needed
- Planned several data transfer test campaigns for 2025
  - Network & storage performance
  - Stress and scalability
- SKA-FTS will be used for data movement for these campaigns
  - Can test token workflow, config settings, scalability, load, etc.

Science and Technology Facilities Council

# Consideration for future

- Integration with S3 storage endpoints?
  - Can take advantage of FTS multi-hop for S3-S3 transfers if needed
- Run FTS in containers?
  - Scalability of service
  - Can deploy additional hosts as needed

Science and Technology Facilities Council

# Summary

- FTS is proven as an essential component for data movement in WLCG community
- FTS instance for SKA has been deployed and is being used to move data between SKA resource center sites
- Plan to test the data moving capabilities in data movement campaigns in 2025

Science and Technology Facilities Council


Science and
Technology
Facilities Council

# Questions?