Cherenkov Telescope Array Observatory report

L. Arrabito[1] on behalf of CTAO DPPS WP

K. Kosack, N. Benekos (DPPS WP coordinators)
F. Gillardo CTAO (DPPS Integration infrastructure)
M. Fussling CTAO (SUSS WP coordinator)

DPPS WMS team
L. Arrabito[1], N. Pigoux[1], A. Faure[1]

DPPS BDMS team
S. Gallozzi[2], G. Zacharis[2], S. Anwar Ul Hasan[3], R. Walter[4], E. Lyard[4], H. Li[4]
[1]LUPM CNRS-IN2P3, France
[2]INAF-Astronomical Observatory of Rome
[3]ETH Zurich
[4]University of Geneva, ISDC

DIRAC & Rucio for CTAO

- CTAO has been using DIRAC for Monte Carlo simulations and processing since more than 10 years
  - Massive productions in the design phase to optimize CTAO array configuration and layout
  - User-specific simulations and analysis
- LUPM team leading this effort
  - Operating a dedicated CTADIRAC instance accessing EGI resources
  - Developing a CTADIRAC extension and contributing to vanilla DIRAC (e.g. Production System, TS)

DIRAC & Rucio for CTAO

- CTAO involvement in Rucio started in the context of the ESCAPE project WP2 DIOS ('data lake') (2019-2023)
  - First Rucio server installation deployed at PIC
  - Put in operation a full CTAO data lake instance on a distributed system with first large-scale tests of data management (data ingestion on a remote site (La Palma), replication to off-site (PIC, then to CNAF) and data transfer via 'intercontinental link', data removal
  - Simple DIRAC/Rucio integration tests performed
  - Summary document can be found here
  - Several contributors from CTAO (G. Hughes, N. Neyroud, K. Kosack, N. Benekos, G. Hughes, M. Fussling), LAPP, PIC, LUPM
- Today the DPPS BDMS team (INAF/Rome, ETH Zurich, Univ. Geneva) is working on Rucio for the development of the Bulk Data Management System of CTAO

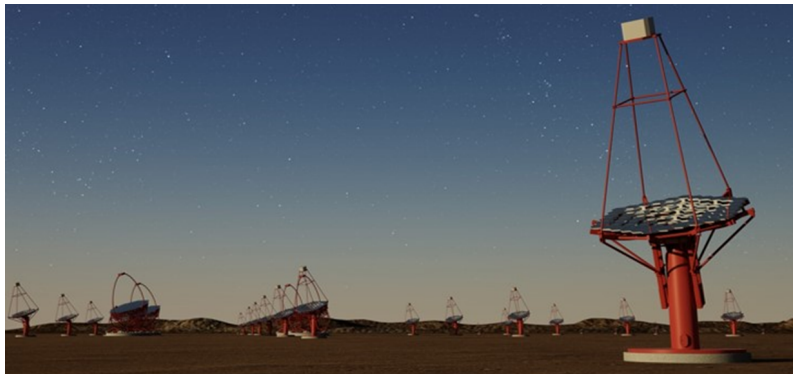
DIRAC & Rucio for CTAO

- CTAO is currently under construction and major infrastructure development will start in 2024
- During the construction phase, DPPS will start operations with increasing functionalities over time
- The plan is to use DIRAC and Rucio as core components of the future DPPS
- Rucio is also a candidate technology for the CTAO Science Archive (work has not started, and it would come with additional requirements for Rucio)
- CTAO will continue participating in the ESCAPE collaboration and in particular in ESCAPE WP2 DIOS (data lake) to collaborate with CERN and SKAO on Rucio-related topics

CTAO data rates and volumes

CTAO-North, Canarian island

- 4 Large Size Telescopes
- 9 Mid Size Telescopes



CTAO-South, Chile

- 14 Mid Size Telescopes
- 37 Small Size Telescopes

- 15% of observation time each year (dark time available per year)
- 20% monitoring and service data

14 GB/s → 67 PB/year

7.5 GB/s → 36 PB/year

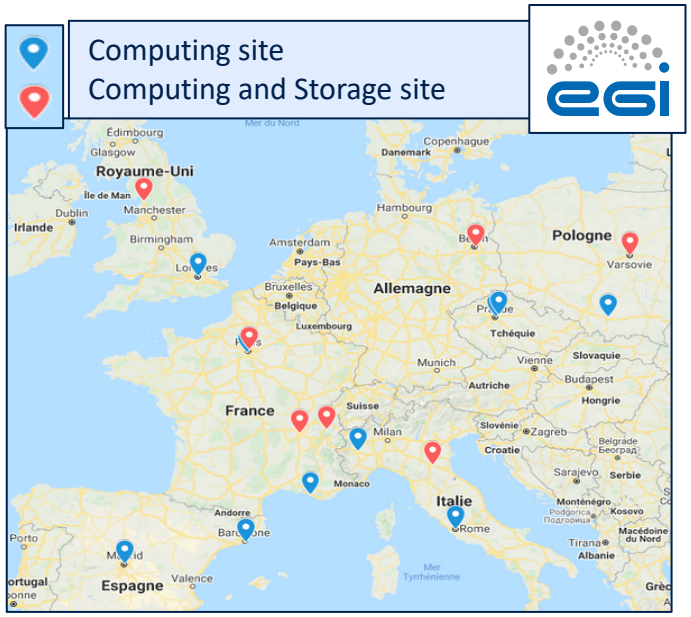
1-2 Gbps capacity **-> Required Data Volume Reduction ratio after 5 years: 50**

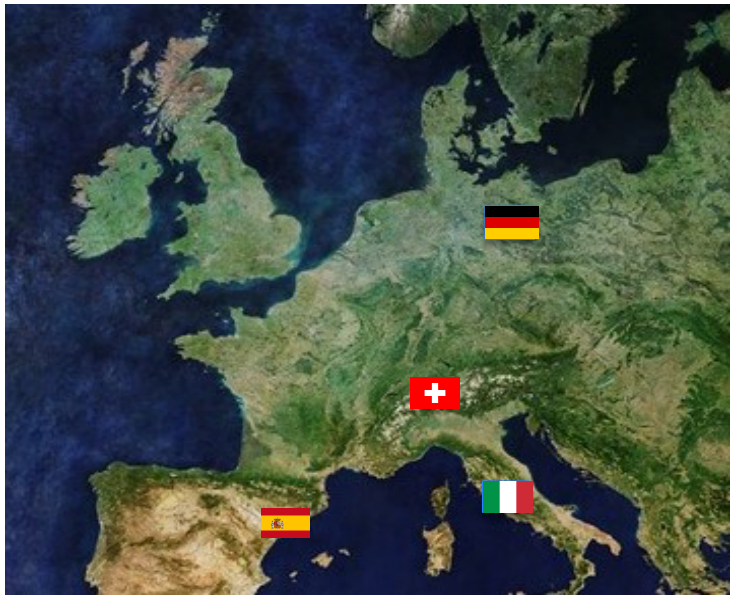
TO BE TRANSFERRED OFF-SITE, PROCESSED, REPLICATED → 3 PB/year New data

22.5 PB Simulation data

CTAO Computing model

- Moving from a grid model with ~ 10 sites providing ‘opportunistic’ resources to a computing model with 4 Data Centers and pledged resources
- 4 Data Centers sharing processing and storage : PIC, CSCS, DESY, Frascati





DIRAC for CTAO

CTAO DIRAC instances

1. **CTADIRAC 'Prototype'** running DIRAC v.8.0
   - Configured to access EGI grid resources (vo.cta.in2p3.fr VO)
   - Used in production for MC simulations and processing
   - Deployed on 6 VMs at PIC, CC-IN2P3, DESY

2. **CTADIRAC Test** instance running DIRAC v.8.0
   - Deployed on 2 VMs at the CTAO test infrastructure at DESY
   - Not meant for scalability, just for functional and integration tests with external systems (Rucio, etc.)
   - Configured to access the resources of the 4 CTAO Data Centers

DIRAC functionalities in use

- Resources accessed through the WMS
  - HTCondor and ARC CEs
  - Tests done in the past with SSH CE to access standalone clusters
  - Tests done with VMDIRAC to access Cloud resources (2017)
- Data Management System
- DIRAC File Catalog (Replica and Metadata Catalog)
  - More than 26 million replicas
  - About 20 meta-data defined to characterize CTA datasets
  - Using *datasets* to expose data selections to users and as input to transformations/productions (currently 1080 defined datasets)
- Transformation System
  - For processing workflows and data management operations (FTS as backend)
  - Using the TSCatalog interface (CTAO contribution)
- Production System (CTAO contribution)
  - For processing workflows composed of several transformations
- Monitoring System with ElasticSearch backend
  - Component Monitoring/WMS history/RMS monitoring
  - Not yet for Job Parameters

Main CTADIRAC developments since last workshop

- User interface (YAML-based) for the Production System (A. Faure)
- Initial support of CWL for describing transformations (A. Faure)
- Transformation failure recovery (N. Pigoux)

User interface for the Production System

- In vanilla DIRAC the Production System can be used only through the Python API and a limited CLI, see doc
- For our convenience we have developed a YAML-based user interface to configure and submit productions
  - For now it's specific to CTAO but we plan to generalize it and port it to vanilla DIRAC

Example workflow.yml

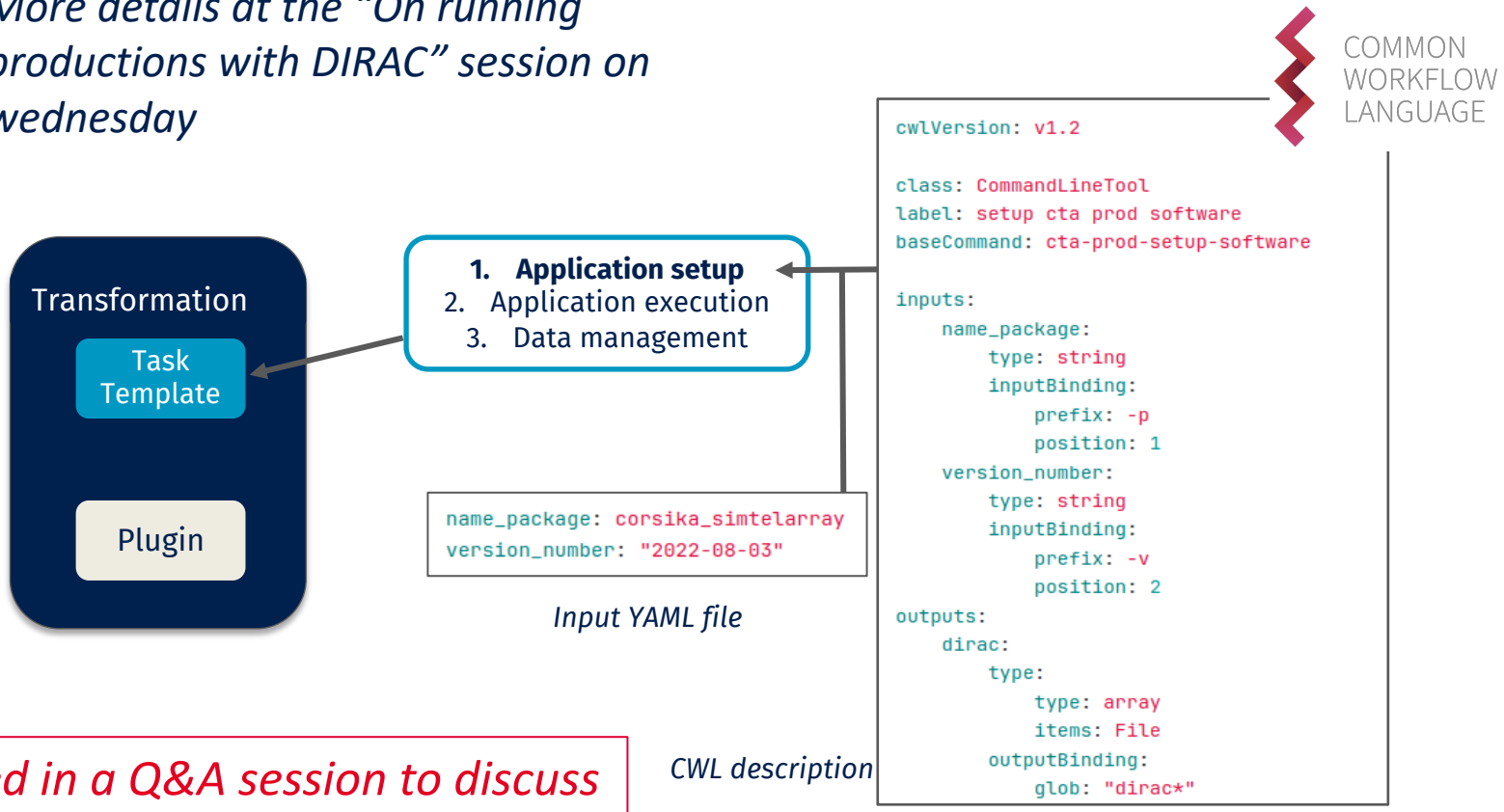
```yaml
ProdSteps:
  - ID: 1
    input_meta_query:
      parentID:
      MCCampaign: PROD5
      array_layout: AlphaNectarcam
      site: Paranal
      particle: gamma-diffuse
      split: train_en
      thetaP: 20.0
      phiP: 0.0
      analysis_prog: ctapipe-merge
      analysis_prog_version: v0.19.3
      data_level: 2
      outputType: Data
      configuration_id: 8
      merged: 2
      moon: dark
    job_config:
      type: CtapipeTrainEnergy
      version: v0.19.3
      options: -c v3/train_energy_regressor.yml
```

More details at the "On running productions with DIRAC" session on wednesday

Support of CWL for describing transformations

- Initial support of transformations described in CWL

More details at the "On running productions with DIRAC" session on wednesday



Interested in a Q&A session to discuss potential CWL support in DIRAC(X)

Transformation failure recovery

- On-going development to automatically recover incomplete transformations (*TransformationFailure Agent*)
- Transformations rarely complete successfully because of different kinds of failures
- Various manual operations may be needed to recover failed jobs and thus complete transformations
- The *DataRecovery Agent* already addresses this issue but it's not really applicable to our use case
- It's crucial for CTAO to automatize the processing as much as possible

We propose a Q&A session on this topic to share our approach to recover incomplete transformations and get feedback from other communities

Other activities

- Tested CTADIRAC deployment on K8 (N. Pigoux)
  - Essentially based on previous work done by LHCb
  - Proof of concept OK but not tested for a full installation and not used in production
  - We prefer to wait for DIRACX which seems more adapted to K8 deployment
- CTADIRAC tests and CI improvement

Feedback on DIRAC from CTAO

- What is your biggest frustration with DIRAC ?

  -> No big frustration, however :
  - DIRAC deployment is not that simple
    - We have developed a set of scripts to simplify full server installation (N. Pigoux)
  - Scalability issue with the Transformation System, maybe due to our installation (see next slides)
- Very good communication in the project
  - Many opportunities to discuss and raise topics/issues, DIRAC team very reactive

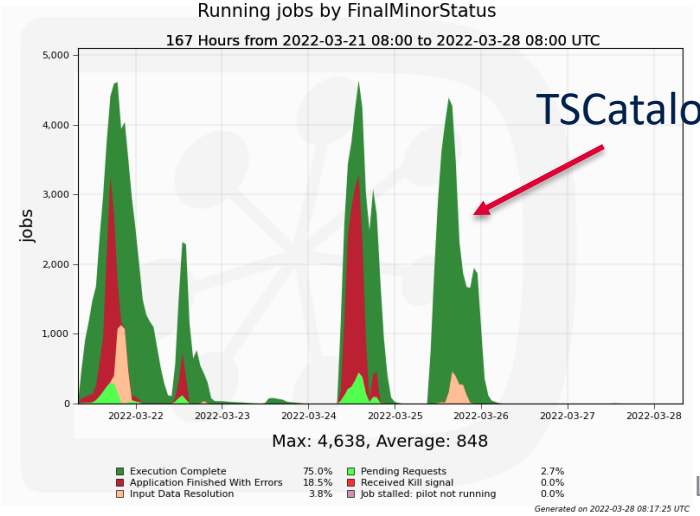
Operations incidents in the last year

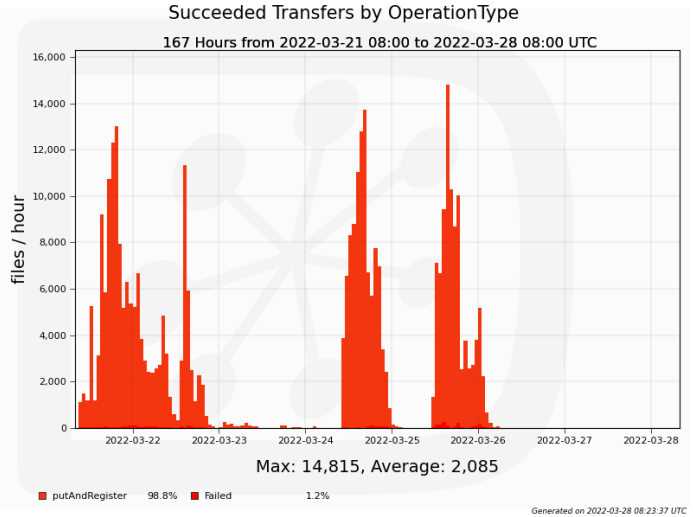
Slides from last Workshop

- Frequent connection timeouts to the Transformation Manager service during productions if using the TSCatalog

dirac-jobexec WARN: Issue getting socket: <DIRAC.Core.DISET.private.Transports.M2SSLTransport.SSLTransport object at 0x2b0b575a98b0> : ('dips', 'ccdcta-server04.in2p3.fr', 9131, 'Transformation/TransformationManager') : timed out:SSLTimeoutError('timed out')

- With TSCatalog enabled
  - Each file is registered in DFC is also attempted to be registered in the TSCatalog (addFile)
  - Idem for setMetadata





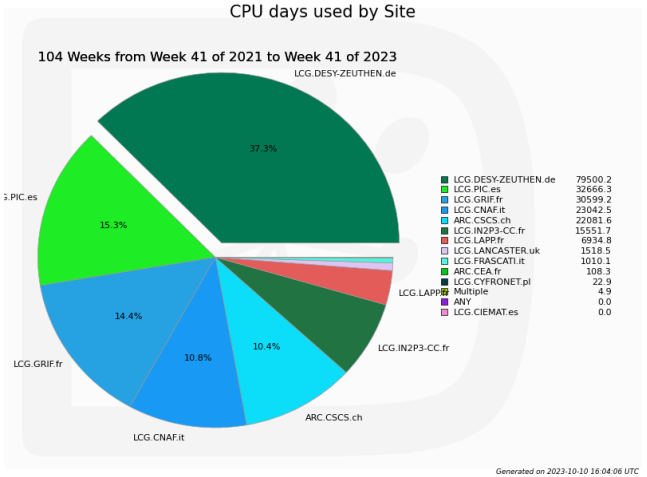
Operations incidents in the last year

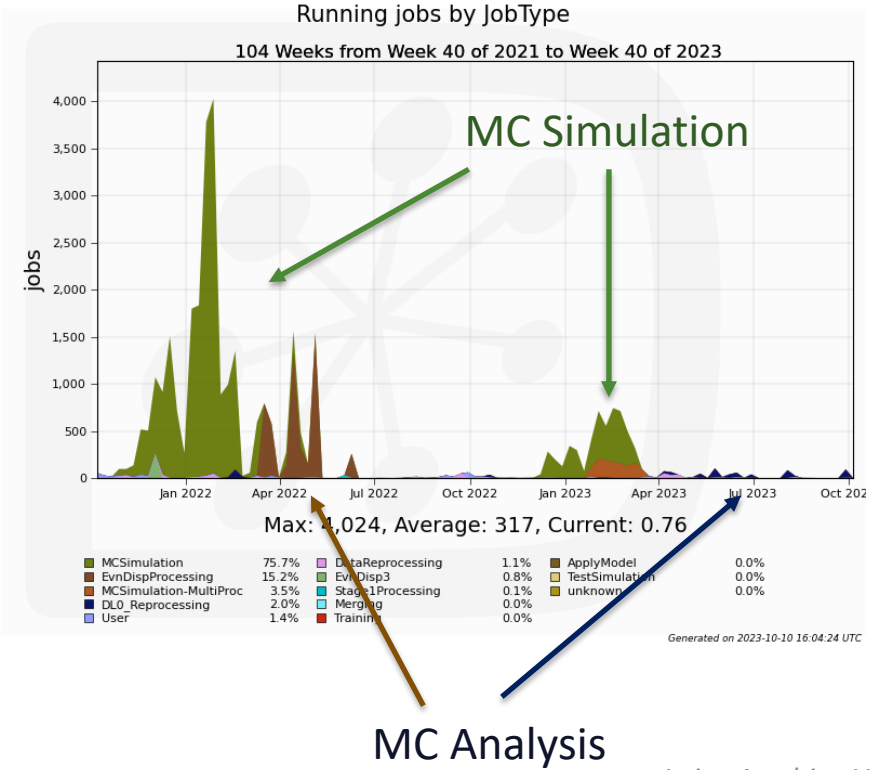
Slides from last Workshop

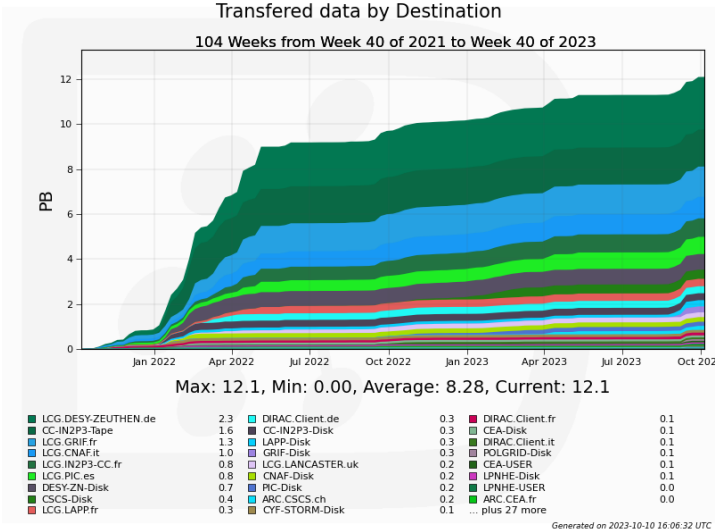
- Investigations done with DB and sys admins at CC-IN2P3
  - No particular CPU of I/O load observed on Maria DB cluster but found errors reading communication packets
  - No load observed on the server running the TSManager service
  - Relaxing some connection timeout parameters on the Maria DB server did not help
  - Tried to tune network parameters on the DIRAC server did not help
- See discussion

Operations in last 2 years

- About 85 millions CPU HS06 hours
- Used about 6 PB distributed in 7 sites



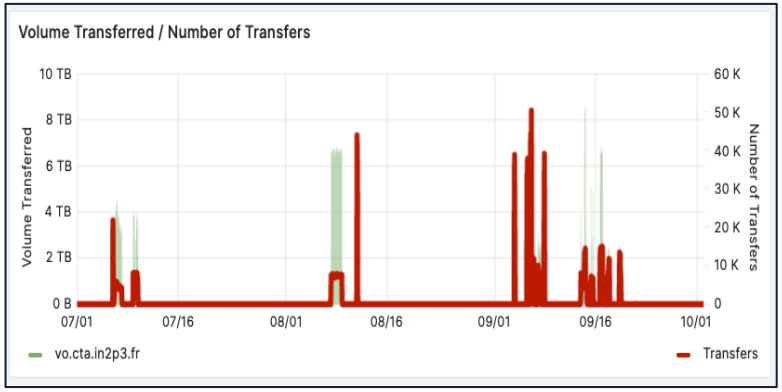


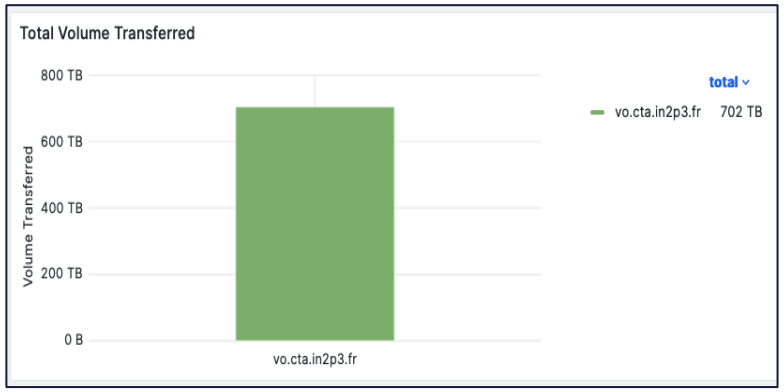


Operations in last 2 years

- Massive migration due to decommissioning of a few SEs ~ 700 TB
- Using the Transformation System and FTS

FTS transfers in summer 2023



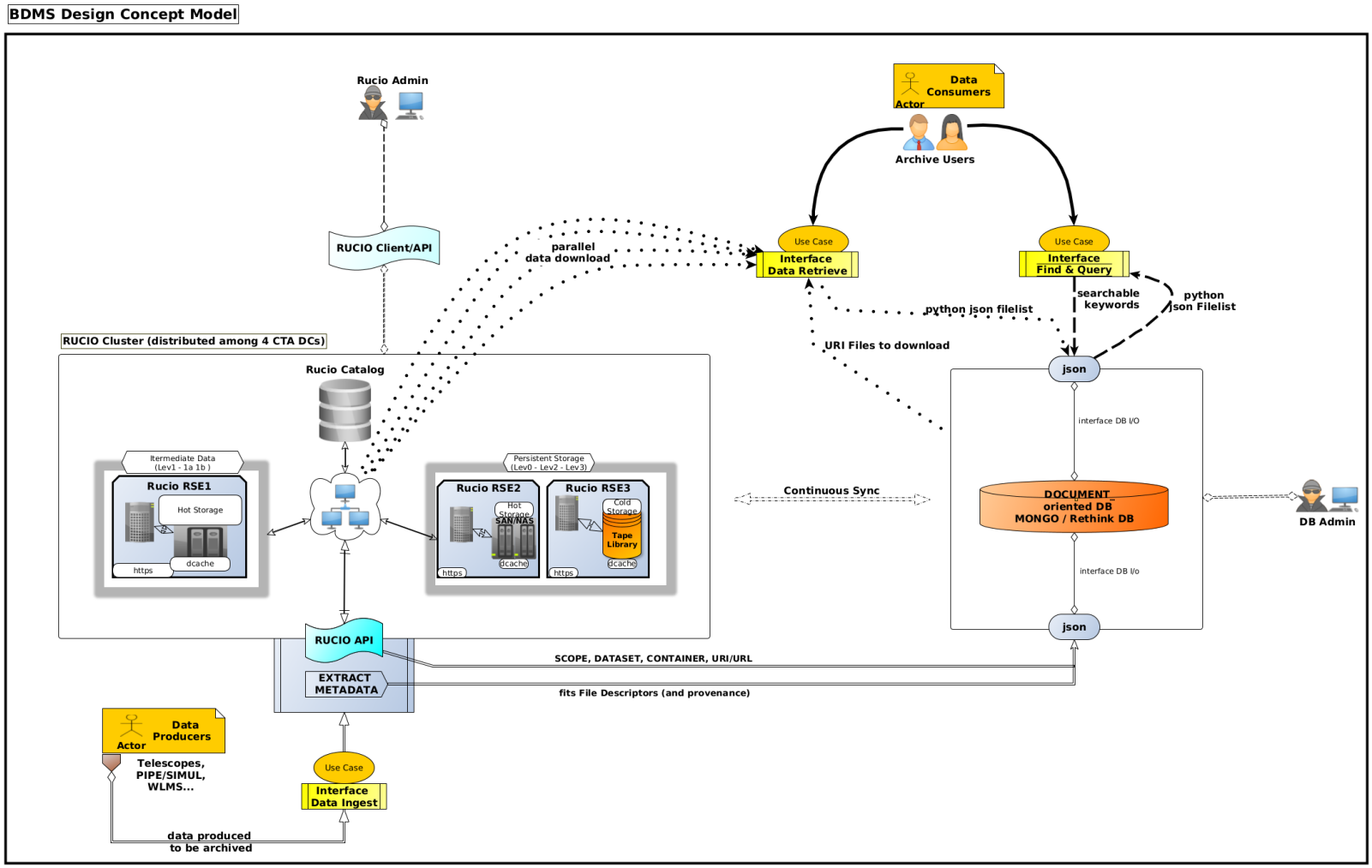


Expectations for DIRACX

- Easy deployment (e.g. on containers and K8)
- More based on standards
- Token support (IAM)
- Performance/scalability

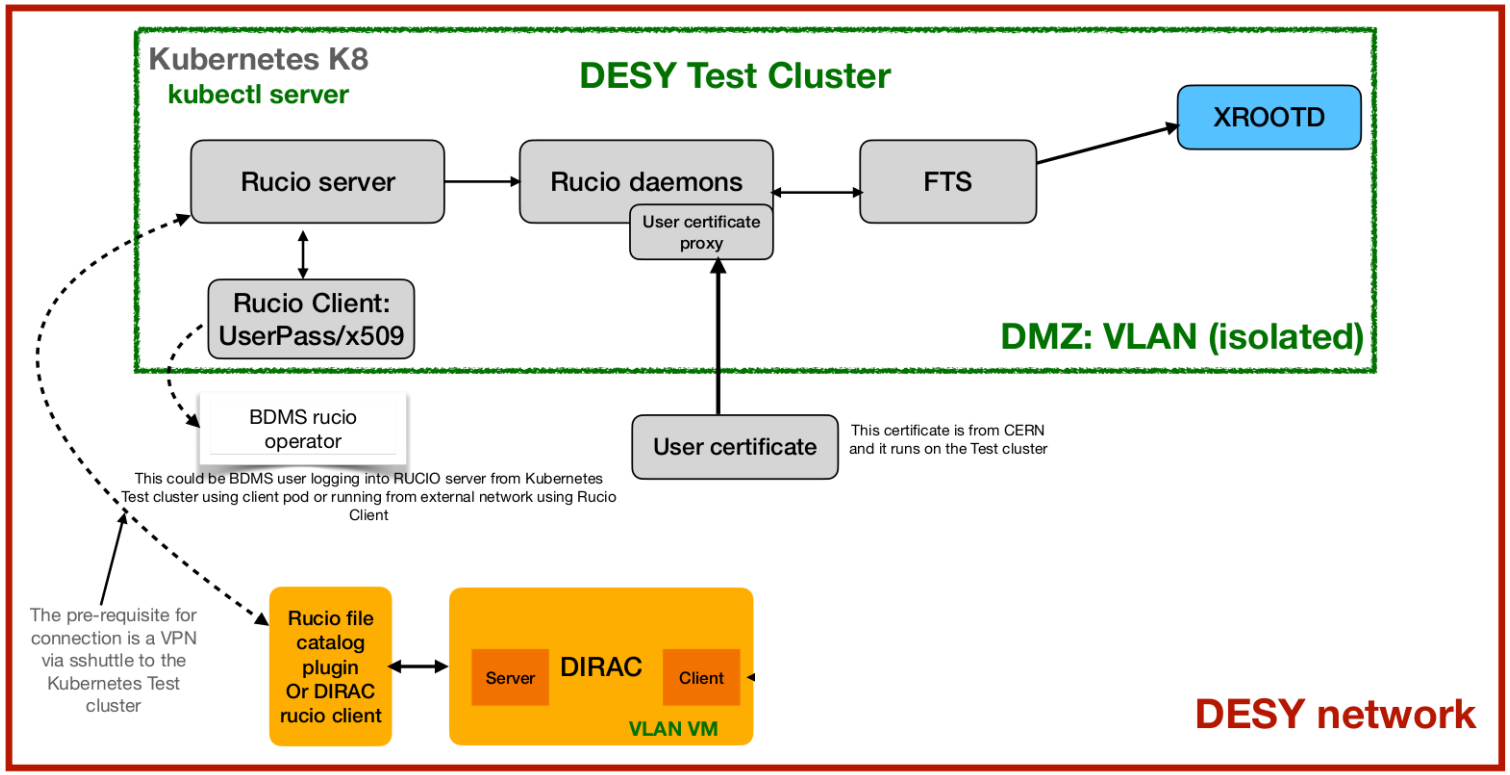
Rucio for CTAO

CTAO Bulk Data Management System and Rucio



Rucio installation for CTAO

Current: Status on BDMS-WMS integration for Release 0



6

Rucio experience, issues and suggested improvements from BDMS

- What is your biggest frustration with Rucio?
  - A centralized metadata catalog exposes two main threats
    1. It represents a SPOF and related issues in case of fault (i.e. disaster recovery issues)
    2. Query functions could be not efficient in case of billions of records

       -> We implemented an external metadata catalog (a distributed document oriented DB as a multi-node cluster with redundant replicas and shards -> SAAS + VirtImage)
- Any notable operations incidents in the last year?
  - Only related to K8 deployment
- Suggestions for improvement
  - Migrate Rucio catalog from Postgres to a noSQL approach and assign a noSQL node to any RSE as a service, failover, load balance and redundancy policy
- Other requirements
  - We would like to bypass rucio download functions/API with direct access to physical file locations
    - A https protocol for unidirectional high-performance download

Open questions about Rucio-IAM integration

Q&A session?

- Usage of tokens in Rucio (workflows, usage of scopes) :
  - How does it work? Is it production-ready? What needs to be further developed for full support? What is the interplay of roles/groups/scopes? Are scopes fully supported in Rucio?

- User authorization:
  - What is the vision where to manage and/or map the permissions of users to data (read, write, remove)? Is this part of the AAI (IAM) or is this part of a BDMS user metadata (Rucio)? This is still an open question for the architecture of CTAO (where we could have this centralized in IAM or distributed in the WMS, the DBMS, the Science Archive, etc.).

Rucio extensions needed for a Science Archive

Q&A session?

- Fine-grained authorization :
  - Datasets with fine-grained access permissions (proprietary vs. non-proprietary access, during proprietary period access by multiple non-overlapping teams with dynamic team composition (add/revoke individual user rights))
- Fast metadata catalogue with complex queries :
  - Metadata of the data separate from the data itself for fast and complex searching for data
- Data access :
  - Support for multiple data access methods, including CLI, APIs, web integration, data packaging
- Data hiding :
  - It should be possible to make data 'invisible' to users (independent of read/write access permissions

DIRAC-Rucio integration

- Restarted tests using the RucioFileCatalog plugin for basic operations (common work between LUPM and ETH teams with the help of F. Gillardo)
  - Put and Register File
  - Run jobs accessing Input Data registered in Rucio
- Essentially repeating the tests done in 2022 in the context of ESCAPE but using the CTAO test infrastructure
- Tests not yet fully successfull
  - We have to find out how to easily make changes to the Rucio server installation, e.g. add extract_scope_cta method, etc. without breaking the installation

- Very much appreciated help received from C. Serfon

What would you improve in DIRAC & Rucio interoperability?

- Using the same technologies (e.g. K8) for deployment it will help the administration of both services

- Support of metadata functionalities in the RucioFileCatalog plugin of DIRAC in order to use the DIRAC Transformation System with input data queries performed on Rucio

→ *Interested in a Q&A session on this topics*

Requirement for both DIRAC and Rucio

- IAM token support
  - CTAO intends to use DIRAC and Rucio with the CTAO AAI instance based on Indigo IAM
  - A test instance is available and is used for prototyping effort

Conclusions

- DIRAC and Rucio are two core software for Data Processing and Preservation System of CTAO currently under development
  - Interoperability between the two software is crucial for us
  - IAM-token support is also crucial
- We see DIRACX as challenging but promising for the sustainability of the DIRAC project
  - As CTAO DPPS WMS team we are willing to contribute
- **Thank you for organizing this common workshop.**
  Very helpful not only for learning from experts, but also for learning from other users communities