tnc25
Brighton, UK | 9-13 JUNE 2025
BRIGHTER TOGETHER

Token Based Authorisation

The key to the future of High Energy Physics computing

Berk Balci (CERN), Hannah Short (CERN), Thomas Dack (STFC-RAL), Maarten Litmaath (CERN)

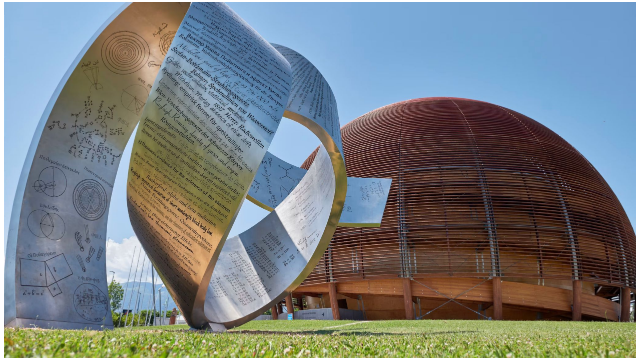
Co-funded by
the European Union

GÉANT

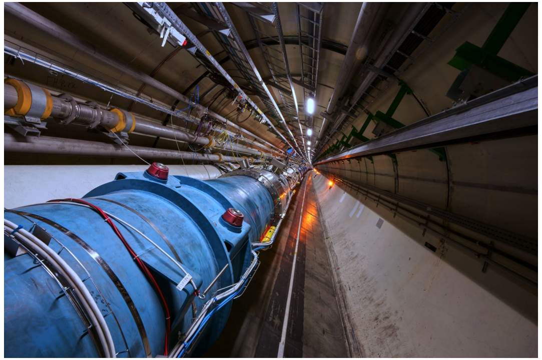
Content

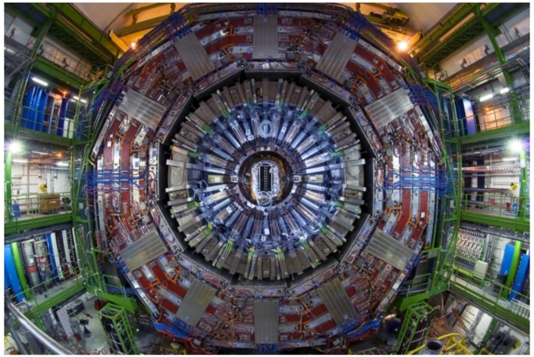
- CERN - LHC
- WLCG
- WLCG AAI History
- Token Transition
- Deployment at CERN
- AARC
- WLCG Workflow and Challenges
- Performance
- EOSC

CERN - LHC

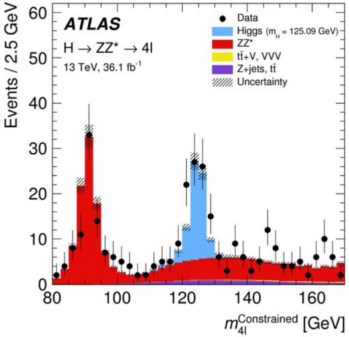


LHCb
ATLAS
CERN Meyrin
CERN Prévessin
SPS 7 km
ALICE
SUISSE
FRANCE
CMS
LHC 27 km





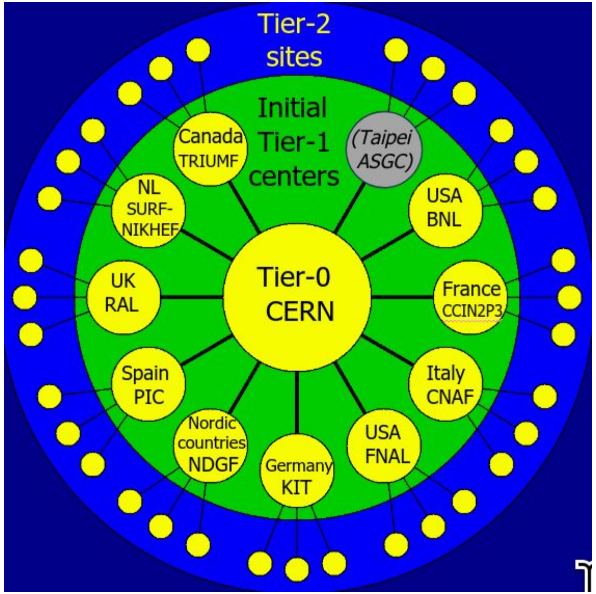
?


ATLAS
H → ZZ* → 4l
13 TeV, 36.1 fb^{-1}
● Data
Higgs (m_H = 125.09 GeV)
ZZ*
$t\bar{t}$+V, VVV
Z+jets, $t\bar{t}$
Uncertainty
Events / 2.5 GeV
0
10
20
30
40
50
60
80
100
120
140
160
$m_{4l}^{\mathrm{Constrained}}$ [GeV]

Worldwide LHC Computing Grid (WLCG)

- Large Hadron Collider (LHC) is the world's largest particle collider, famous for discovering the Higgs boson
- With great power comes great responsibility:
  - LHC experiments produce 200+ PetaByte / year
  - ~1 million computer cores are needed for scientists' analysis
  - Access needed for 12 000 physicist around the world
- With great responsibility comes great power - WLCG -
  - 147 sites in 42 countries
  - 2+ ExaBytes of storage
  - 2+ million processing tasks (jobs) / day
- With great power comes again great responsibility - IAM -
  - Identity management for granting access to scientists
  - Access management for LHC data and computing resources

WLCG
Worldwide LHC Computing Grid



WLCG AAI History

- From early 2000s authentication has been done with X.509 certificates and VOMS
- VOMS extends X.509 certificates by
  - Creating short-lived proxies of user certificates
  - And adding user roles and group memberships to them for authorization capability

Workflow:

- Users get personal X.509 certificates from IGTF-trusted Certificate Authorities
- WLCG sites and services trust IGTF certificates
- Each experiment uses VOMS to assign users to roles and groups
- Users use VOMS CLI tool to:
  - Create a short-lived proxy certificate
  - Add VOMS authorisation info
- These proxies carry both authentication and authorisation and are used to submit jobs and access data across the WLCG.

Token Transition - Theoretical Benefits

- Certificates
  - Hard to manage
  - Poor usability and portability
  - Poor interoperability with other research infrastructures
  - Weaker security
    - Long-lived proxies
    - Lacking access control granularity
- Tokens
  - Easier for users to manage
    - Minimal interaction required from end users
    - Tools deal with them under the hood
  - Reduced risk if compromised
    - Fine-grained access control
    - Typically short-lived
  - Designed for modern infrastructures
    - Increased interoperability with other infrastructures

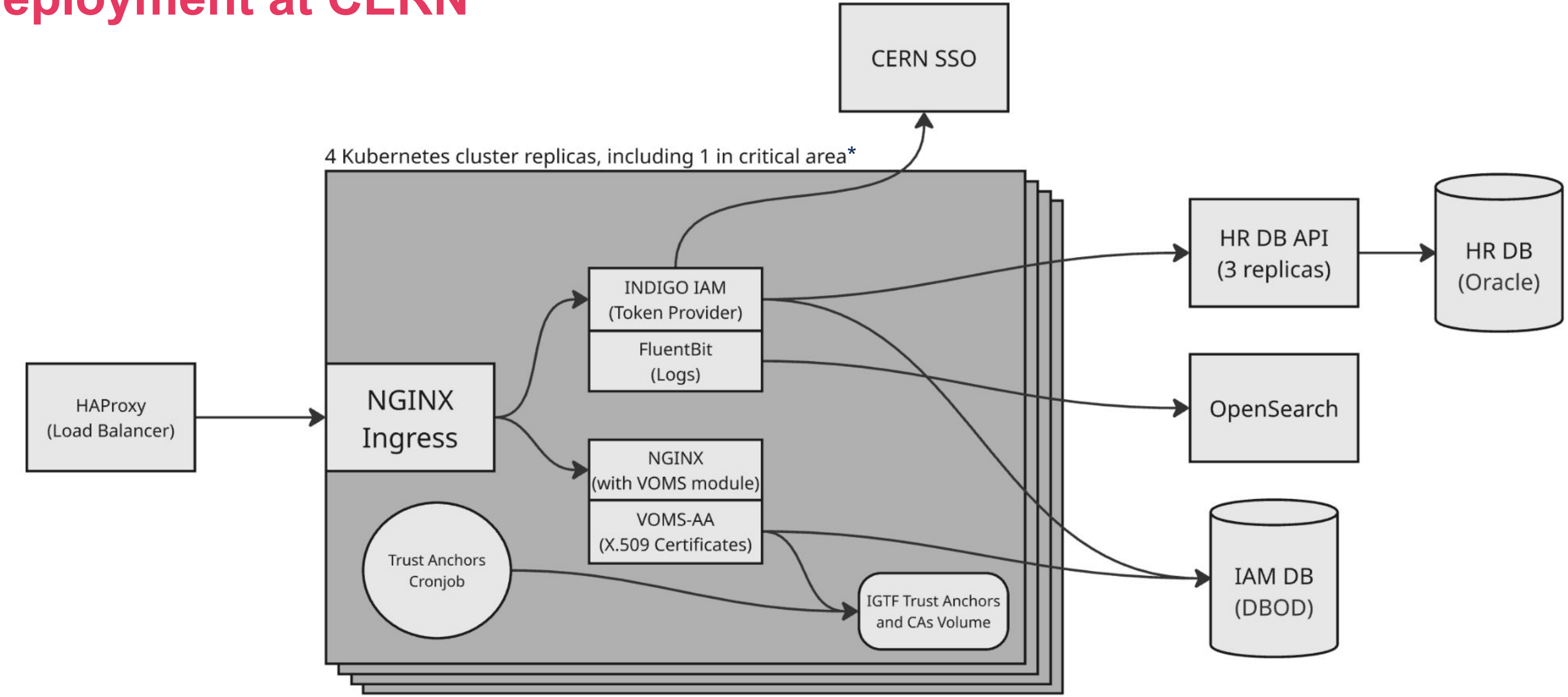
Token Transition - Timeline

- 2017 – WLCG Authorisation Working Group created to lead the token transition
- 2019 – WLCG JWT Profile v1.0 released
- 2019 – Decided to adopt INDIGO IAM
  - Developed by INFN, opensource
  - Has VOMS-AA for backwards compatibility
- 2020 – First IAM instances in production
- 2024 – Phaseout of the legacy VOMS services completed
- 2025 – Migration to HA Kubernetes for better performance and reliability
- Now – Already: ~25% of ATLAS file transfers use tokens
- 2025 – Tokens begin production use for data access by computing jobs
- 2028 – Completion of the X509 / VOMS phaseout

Deployment at CERN

- Multi-cluster HA deployment on Kubernetes
- Separate instances for each Virtual Organization:
  - 5 LHC experiments (ALICE, ATLAS, CMS, LHCb, MoEDAL)
  - 5 other experiments / projects (AMBER, CALICE, COMPASS, FCC, ILC)
  - 2 Operations (WLCG Ops, DTeam)

Deployment at CERN



* Only the 4 big LHC experiments have a 4th replica in the critical area

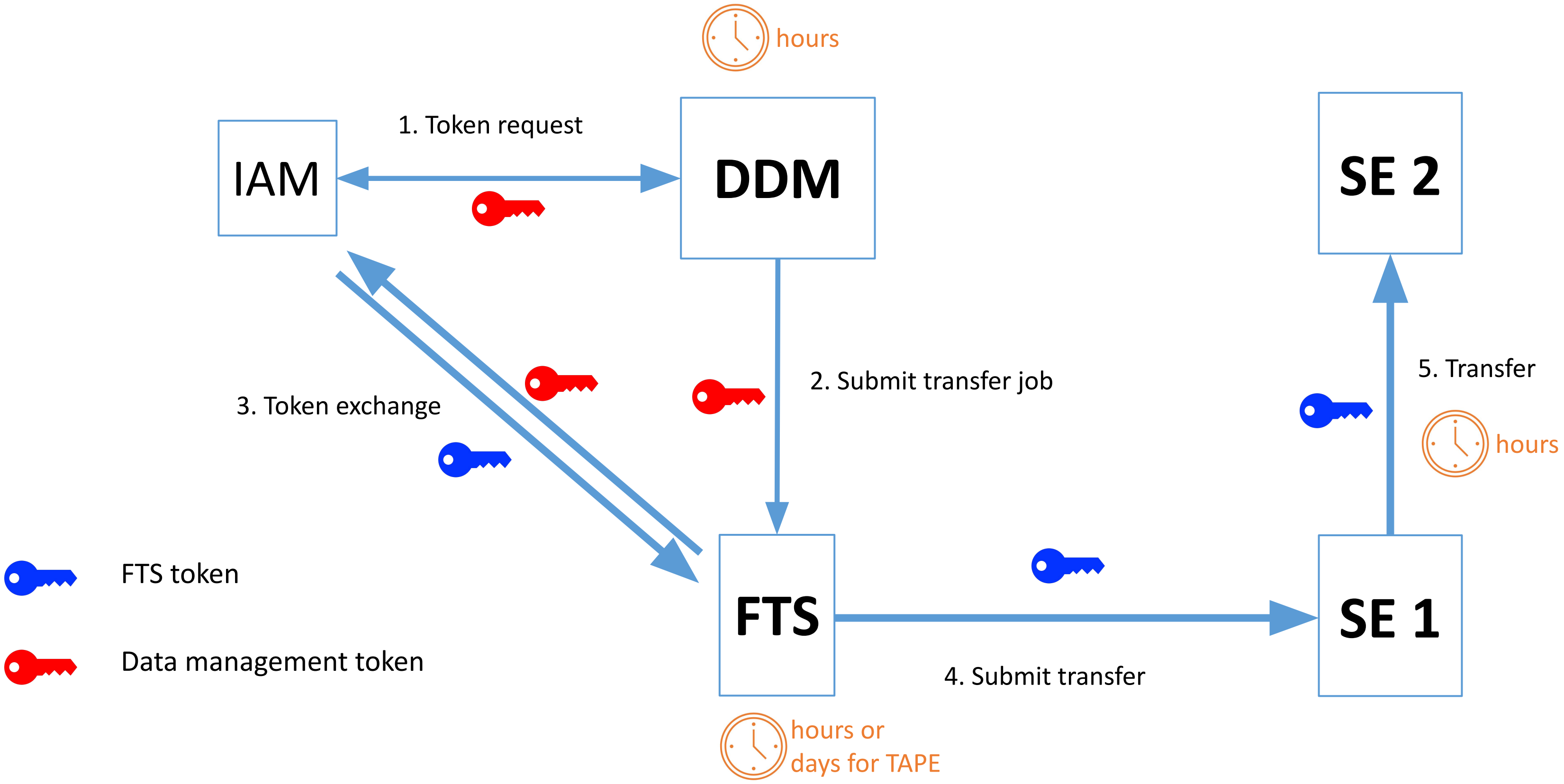
AARC (Authentication and Authorisation for Research and Collaboration)

- AARC provides guidelines and architectures to enable interoperable access across research infrastructures in Europe.
- CERN participates in the AARC TREE project
  - To stay aligned with evolving standards and best practices
  - To support interoperability among European research communities, which is essential for future collaborations requiring cross-community access to services.
- AARC BPA (Blueprint Architecture) is adopted as a reference model
- GUT (Grand Unified Token) profile work has been started to define a single token profile that aligns SciTokens, the WLCG token profile, and AARC guidelines.
- WLCG has unique requirements due to its scale
  - Full alignment with AARC can be challenging – e.g., token lifetime, proxied token validation

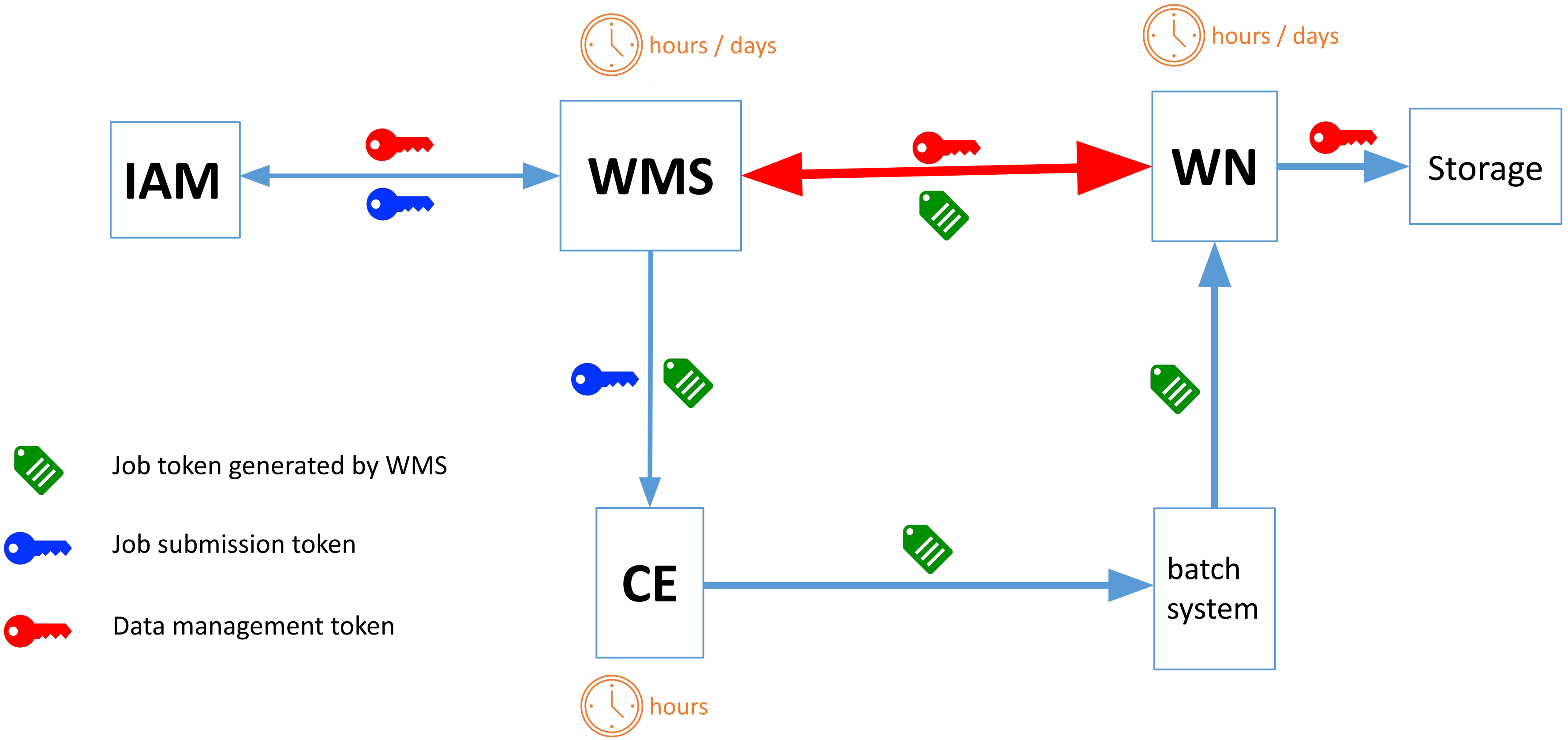
Challenges

- While tokens offer many advantages, they also introduce new operational challenges that need to be addressed:
    - The risks associated with service centralisation
    - Token lifetimes and rates
    - Scope granularity
- These challenges are mainly due to the fact that most software was not built to run at the scale required by WLCG experiments, especially in certain workflow models.
    - O(5M) file transfers per day needing tokens for reading source file and writing destination file
    - O(millions) jobs per day needing tokens for reading input and uploading output data

WLCG File Transfer Workflow - One model



WLCG Job Submission Workflow



Centralisation — A New Single Point

- In the X.509 model, validation is decentralized by trusting IGTF CAs
- In the token model, validation relies on the central IAM server
- This introduces a single point of failure: if IAM is unavailable, token validation – and therefore usage – fails.
- Increased operational responsibility:
  - High availability
  - Redundancy
  - Monitoring
- Two methods of token validation:
  - Offline validation - sites validating with the public key served by central IAM server
    - Allows to mitigate operational risks by sites caching the key and IAM serving the key in distributed manner by using static fallback
  - Online Validation - Sites validating with the introspection endpoint of central IAM server
    - AARC recommends this to allow Proxied Token Introspection
    - This model leads to unpredictable and possibly unacceptable loads on the issuer at WLCG scale

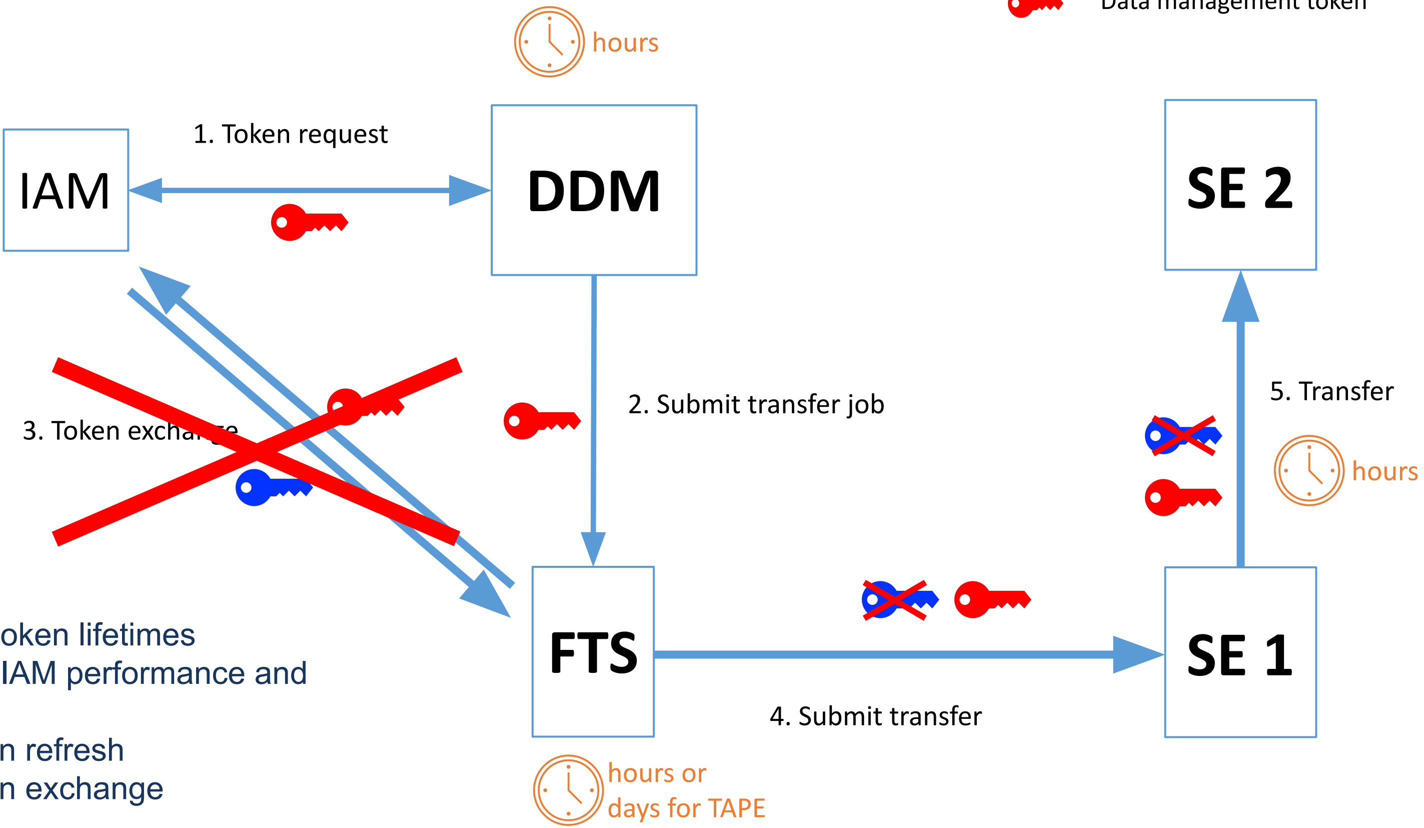
Token Lifetime — A Trade-Off

- X.509 proxies could last up to days.
- Tokens are short-lived by design – reduces impact if leaked.
- But this creates more frequent calls to the IAM server for refresh.
    - Higher load on IAM
    - Stronger dependency on uptime and scalability
- Token lifetime can be tuned for balancing security and operations:
    - Long lifetimes weaken security and reduce interoperability
    - Short lifetimes improve security but increase operational pressure at WLCG scale
        - More load on IAM
        - Increased need of high availability
- AARC guidelines recommend short lifetimes for better interoperability and security.
    - But WLCG is a special case: large-scale, long-running jobs and background services may need adjusted lifetimes.

WLCG File Transfer Workflow - Another model

~~FTS token~~

Data management token



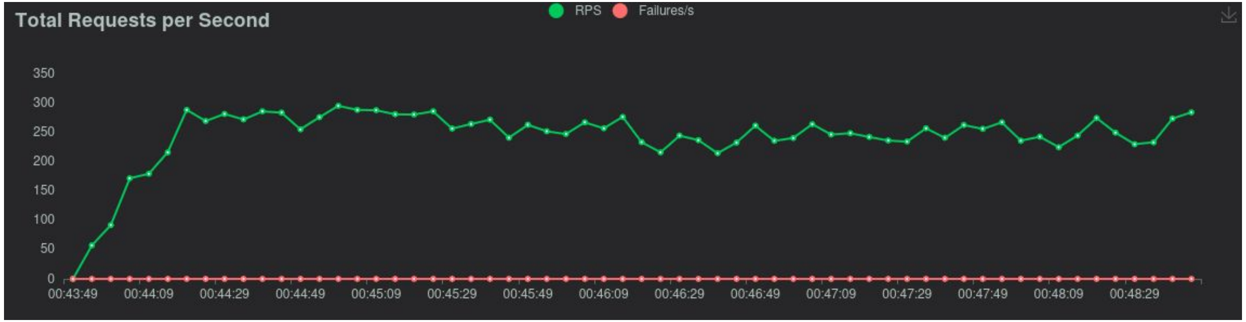
- Needs longer token lifetimes
- Relies less on IAM performance and availability
    - No token refresh
    - No token exchange

Scope Granularity — Limiting the Blast Radius

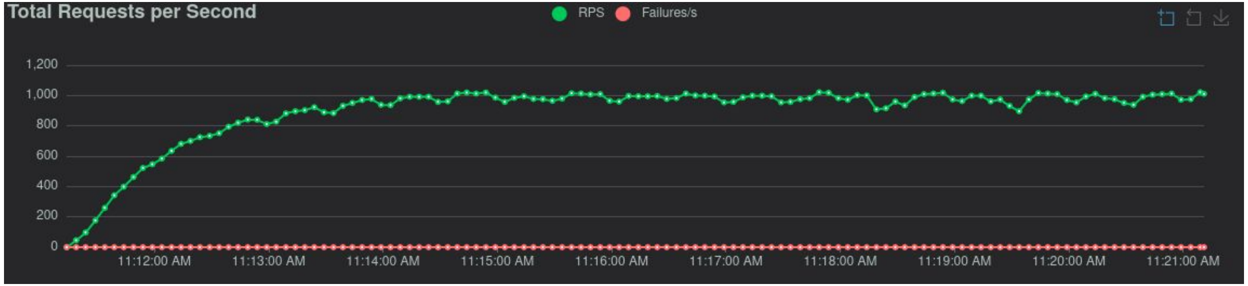
- X.509 proxies grant broad access — full access to all VO storage.
  - If compromised, the impact could be massive.
- Tokens can be scoped narrowly:
  - Access to a specific dataset, or specific file, possibly only at specific services named as the audience
  - Specific action (read, write, modify)
- This reduces the impact of leaked tokens:
  - An attacker can’t escalate privileges or move laterally
  - Better alignment with least-privilege principle
- With great power comes great responsibility:
  - Fine-grained tokens increases the number of tokens needed per activity
  - Increased performance demands on IAM (e.g. tokens/sec rate)
  - Stronger dependency on uptime and scalability

Performance

- Using single instance deployment on Openshift:



- After migration to multi-cluster HA Kubernetes deployment:



EOSC

- EOSC (European Open Science Cloud) aims to provide a federated environment for sharing research data and services across Europe
- CERN is planning to establish a CERN Node for EOSC to contribute to and benefit from the ecosystem
- EOSC AAI adopts AARC guidelines
  - Ensuring interoperability now prepares us for potential cross-community collaboration
  - If in the future WLCG IAM needs to serve as a community AAI within EOSC, it must comply with EOSC AAI
  - Further reinforces the relevance of aligning WLCG IAM with AARC guidelines

Thank you

Any questions?

berk.balci@cern.ch

Co-funded by
the European Union

GÉANT