Token based AAI & WLCG Data Challenge 2024

Christoph Wissing (DESY)

D-CMS Computing Team Meeting, November 2023

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Disclaimer

Both topics are complex (and partly interlinked)

For Token based AAI there is a WLCG working group

- WLCG Resource Trust Evolution Task Force: twiki link
- Group started more than 5 years ago

For the upcoming WLCG Data Challenge 2024 (DC24) there was a 2 days workshop

- Preparations for DC24 actually started already beginning of this year

Today only a few aspects are being addressed

Token based AAI

AAI has been realized in WLCG with X509 certificates and VOMS extensions for groups & roles

X509 certificates considered complex and not attractive by other communities

Token based authentication based on web login technology

- Typically employ OpenID Connect
  - Implemented using OAUTH2 frame work with Java Web Tokens (JWT)

Using the same technology stack for access to distributed computing and storage is challenging

If you were not happy with X509 certs, you will not like tokens either!

Tokens in WLCG

WLCG converged on its own solution

- Identity provider: Indigo IAM
  - Also provides VOMS features as legacy feature
- A special set of “WLCG Tokens” based on SciTokens
  - Very complex scheme:
  - Capability based authorization: scope, e.g.
    - `storage.read:/` or `storage.create:/store/temp`
    - `compute.create`

Storage and compute middleware needed special implementations to support WLCG tokens

- Most features implemented in HTCondor, XrootD, dCache, StoRM, EOS…

WLCG aims to use only tokens by HL-LHC

- Run 3 will be a mix of X509 and increasing use of tokens

Tokens in CMS

Compute Infrastructure

- Global pool already migrated to tokens
- Tokens used to access HTCondor CEs
- Transparent for users (who still submit with X509 proxy)

Storage

- Initial *(minimal*) tokens setup
- Reference configurations for all major storage middleware
  dCache (worked out by Dmitry Litvintsev and Christian Voss)
- ETF/SAM tests: Token tests already succeeding for KIT and DESY

Rucio and FTS

- FTS-pilot instance at CERN successfully tested with hand crafted FTS jobs
- Integration with Rucio soon (first suited release scheduled still in November)

Recap: Data Challenges for HL-LHC

- WLCG has been mandated to execute data challenges for HL-LHC
  - Demonstrate readiness for expected HL-LHC data rates
  - Increasing volume/rates
  - Increase complexity (e.g. additional technology)
  - A data challenge roughly every two years

- DOMA is the coordination and execution platform
  - Agreements across the LHC experiments and beyond
    - Suited dates
    - Reasonable targets
    - Functionalities
  - Help in orchestration

- Dates and high level goals always approved by WLCG MB

Recap of (initial) modelling & resulting rates

ATLAS & CMS T0 to T1 per experiment

350PB RAW, taken and distributed during typical LHC uptime of 7M seconds

- 50GB/s or 400Gbps

Another 100Gbps estimated for prompt reconstruction data tiers (AOD, other derived output)

1Tbps for CMS and ATLAS summed

ALICE & LHCb T0 Export

100 Gbps per experiment estimated from Run-3 rates

WLCG data challenges for HL-LHC - 2021 planning
https://zenodo.org/records/5532452

Minimal Model

Sum (ATLAS,ALICE,CMS,LHCb)*2(for bursts)*2(*overprovisioning*) = **4.8Tbps for the expected HL-LHC bandwidth needs**

Flexible Model

Assumes reading of data from above for reprocessing/reconstruction in 3 month (about 7M seconds)
Means doubling the Minimal Model: **9.6Tbps for the expected HL-LHC bandwidth needs**
However data flows primarily from the T1s to T2s and T1s!

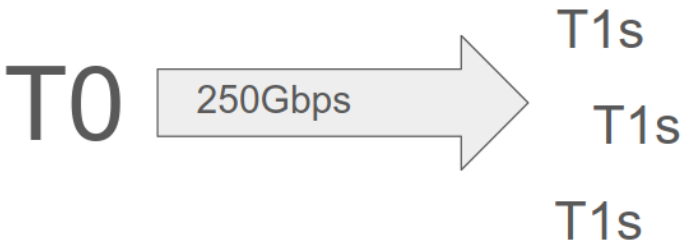
Data Challenges target: **50% filling of expected** HL-LHC bandwidth needs

DC24 Timetable

- Dates: **February 12th (Mon) to February 23rd (Fri)**
  - Main goal is a **25% of HL-LHC** challenge
  - DC21 was a 10% challenge
- Proposal to distribute different exercises over the challenge days
  - Note: All DC exercises run **on top of ongoing production**
  - Common program mostly for ATLAS and CMS
  - Data taking scenario (T0 to T1s)
  - Production like scenarios
  - Ramp up to reach target of flexible scenario
  - Squeeze in dedicated technical tests (e.g. special TCP setup at selected sites)
  - Some contingency to repeat something
- Even with 2 weeks the schedule is tight

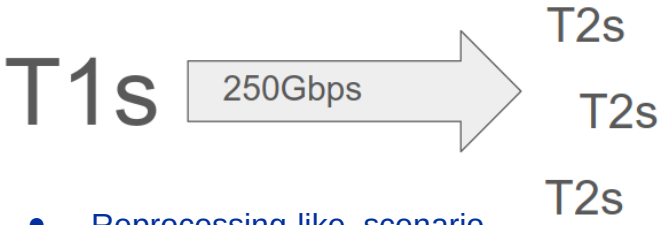
CMS - Main Scenarios

1. “T0 export”



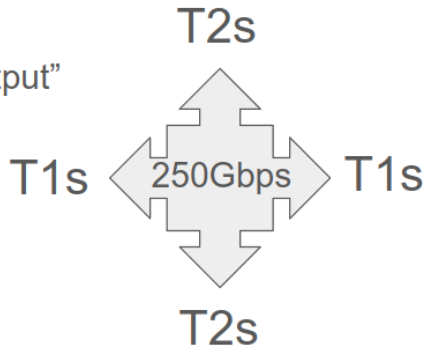
- Rather well modelled
- Numbers derived from DAQ TDR and LHC uptime assumptions

2. “T1 export”



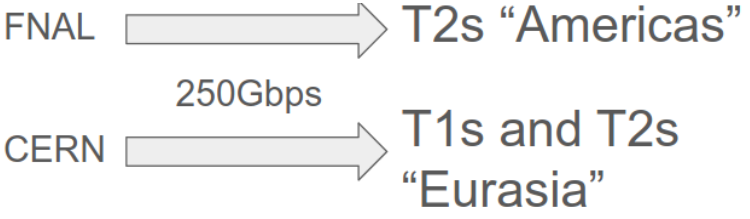
- Reprocessing-like scenario
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

3. “Production output”



- MC & derived data scenario
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

4. “AAA”



- Unscheduled remote reads via Xrootd
  - Main traffic presently MC premixing served from CERN and FNAL
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

Impact on German Sites

ATLAS DC24 transfer rates

Final T2 ingress/egress depends on number of participating T2 sites and might b

rows in red color: sites must explicitly ask be included in DC24 (details will be sent t

(preliminary version: 20231103)

Deletion rates are calculated from ingress bandwidth assuming 3GB average filesize)

Table: DC24 (src: ingress / egress)			Site WAN (Gb/s)		DC24 minimal scenario				DC24 flexible scenar		
			Total (Gb/s)	Usable by ATLAS	T0 Export	Total Gb/s & bandwidth		Space [TB/24h] (deletions/hour)	T0 Export	Total Gb/s & bandwidth	
Site	Tier	Cloud				∑ ingress	∑ egress			∑ ingress	∑ egress
CERN-PROD	T0	CERN	2100	911	270.0	27.9	291.3	0 (0k)	270.0	93.1 - 112.2	363.1
T0 summary					270.0	27.9	291.3		270.0	93.1 - 112.2	363.1
BNL-ATLAS	T1	US	400	400	60.0	82.2	60.0	764 (11k)	60.0	107.5 - 119.6	120.0
FZK-LCG2	T1	DE	400	162	32.0	61.7	32.0	431 (6k)	32.0	86.3 - 100.3	64.0
MPPMU	T2	DE				2.6 - 3.3	1.3 - 1.3	10 (0k)		1.7 - 9.4	1.3 - 9.1
wuppertalprod	T2	DE	10	9		4.6 - 5.9	1.8 - 1.8	32 (0k)		9.9 - 10.0	6.4 - 8.8
DESY-ZN	T2	DE	40	40		6.4 - 8.4	1.9 - 1.9	48 (1k)		14.3 - 19.2	12.4 - 12.4
DESY-HH	T2	DE	100	100		9.1 - 10.0	1.9 - 1.9	49 (1k)		9.9 - 10.0	5.4 - 7.2
UNI-FREIBURG	T2	DE				2.6 - 3.3	1.7 - 1.7	9 (0k)		1.8 - 11.3	1.7 - 11.6

RSE	Ingress (GB/s)	Egress (GB/s)
T0_CH_CERN_Disk	0.000	31.250
T1_DE_KIT_Disk	5.588	4.472
T1_ES_PIC_Disk	2.333	1.887
T1_FR_CCIN2P3_Disk	5.576	4.494
T1_IT_CNAF_Disk	7.108	5.658
T1_RU_JINR_Disk	8.465	5.118
T1_UK_RAL_Disk	4.427	3.551
T1_US_FNAL_Disk	22.076	31.811

Tier 2 total rate goals

RSE	Ingress (GB/s)	Egress (GB/s)
T2_AT_Vienna	0.319	0.072
T2_BE_IIHE	2.624	0.714
T2_BE_UCL	1.114	0.280
T2_BR_SPRACE	1.158	0.286
T2_BR_UERJ	0.177	0.022
T2_CH_CERN	11.097	21.344
T2_CH_CSCS	2.311	0.398
T2_CN_Beijing	0.315	0.079
T2_DE_DESY	3.407	0.930
T2_DE_RWTH	1.611	0.429
T2_EE_Estonia	0.929	0.197

Integrated expected rates

- KIT: about 150Gbs ingress, 100Gbs egress
- DESY (HH): about 38Gbs ingress, 16Gbs egress
- RWTH: about 13Gbs ingress, 4 Gbs egress
- For KIT and DESY additional ingress of ~2Gbs from Belle-II (who also participates in DC24)

dCache

- dCache ready for DC24
- Token support is there
  - CMS just launched a deployment campaign
- Very recent Golden Release 9.2 is recommended
  - Enables new XRootD monitoring framework
- Discussion item:
  - Bulk deletion without SRM
    - Feature only provided via SRM presently
    - Use REST API?
    - Explore capabilities via HTTPS?

Summary

- dCache dev team sees DC24 as an opportunity for large-scale data transfer test (no tape-api tests ☹)
- Token support is available for all token capable protocols: HTTP, REST (TAPE-API), Xroot
- We recommend sites to run dCache version 9.2 to benefit from the latest developments
- We recommend sites to collect as much log information as possible for later analysis

2023-11-10 dCache for DC24 15/16

Bing Image Creator: "Worldwide LHC Computing Grid, Data Challenge Workshop, Happy Mood"

The End

(for today)

Bing Image Creator: "Worldwide LHC Computing Grid, Data Challenge Workshop, Serious Mood"