

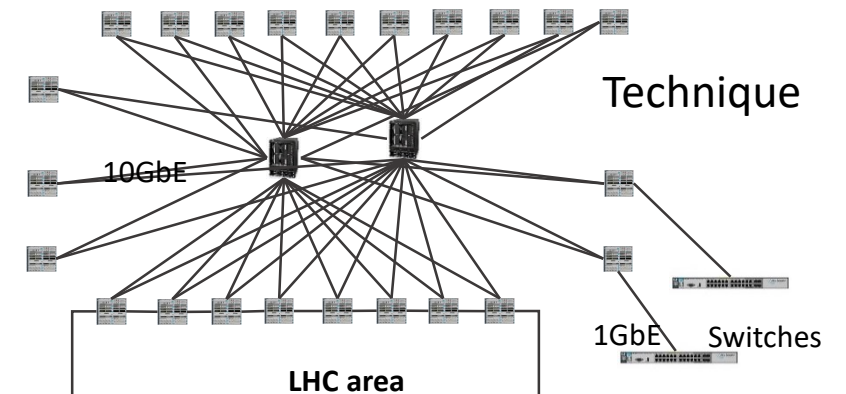
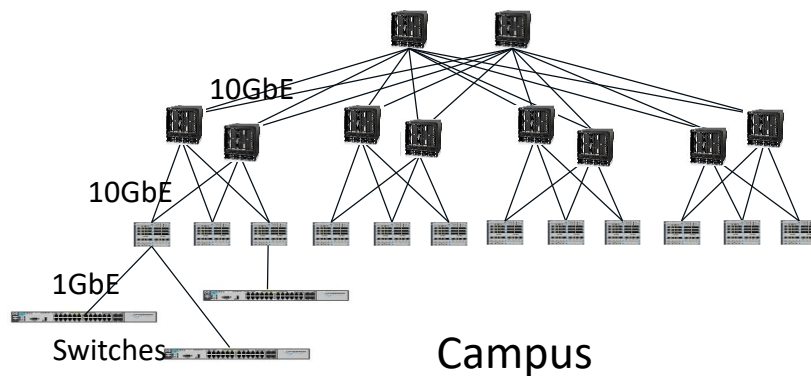
RESEAU ET SECURITE

▶ 4 Sessions CERNoise sur l'évolution de leurs infrastructures réseau

- Upgrade des réseaux du CERN.
- Evolution de l'infrastructure wifi du CERN.
- Evolution de l'outil de management de l'infrastructure réseau.
- Connectivité des objets communicants au CERN.

▶ CERN upgrade des réseaux campus et technique

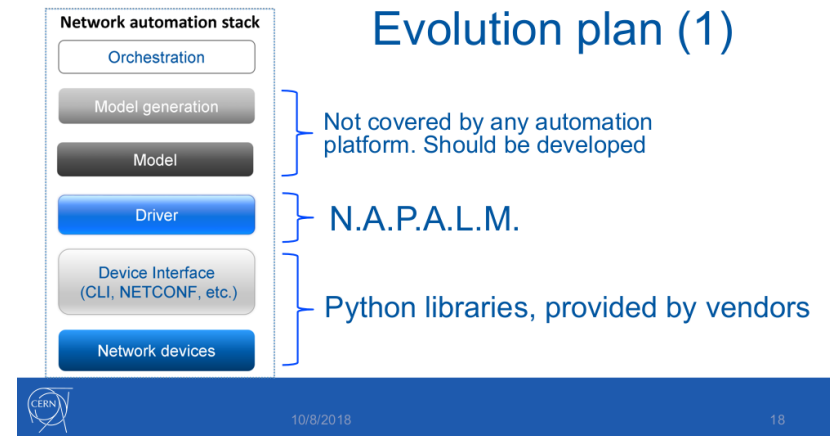
- Réseau campus : 2 sites, 80 000 prises réseau. Dédié aux équipements enduser (postes, wifi, imprimantes,...) et aux expériences.
- Réseau technique : Fournir la connectivité pour tout ce qui concerne les accélérateurs



- Souhaites faire évoluer les deux infrastructures pour plus de débit et plus de redondance (notamment pour le réseau technique)
 - Envisagent des solutions du type : Juniper QFX10002-36 (port 100 et 40 Gb/s mais pas de 1 Gb/s) et Juniper EX9208
 - LS2 : 2019-2020 : Remplacement des routeurs du réseau technique et mise en place de routes redondantes. Renouvellement complet (router et switch) du réseau de campus
 - LS3 : Renouvellement des switchs du réseau technique.
- ▶ **Projet WiSE : Wi-Fi Service Enhancement project**
- Proposer le wifi sur l'ensemble du CERN : intérieur et extérieur.
 - Support du standard 802.11ac, du roaming et accessible aussi pour les visiteurs.
 - Basé sur une solution ARUBA (HP).
 - Aujourd'hui : 3 700 points accès déployés/ 300 switches , ce qui permet de couvrir 180 bâtiments (reste ~20)
 - Configuration centralisé

▶ Network configuration management at CERN

- 4400 équipements réseau de marques diverses.
- CFMGR : Configuration management tool, actuellement en service et basé sur une DB et un ensemble de scripts en PERL.
- Souhaite évoluer vers une solution moins perl, compatible avec les nouveaux équipements et plus uniforme.
- N.A.P.A.L.M. (Network Automation and Programmability Abstraction Layer with Multivendor support)



▶ Challenges for connecting Internet of Things devices at CERN

- Présentation pour mettre en évidence les différents éléments à prendre en compte (Sécurité, Séparation des trafics, QoS, passage à l'échelle,...)
- Appel à partager des expériences et expertises.

▶ Evolution des réseaux LHCONe, LHCOPN

- La France fait partie des bons élèves pour ce qui est du passage à IPV6, les US sont à la traîne.
- IPV6 représente 25 % des usages FTS dans WLCG, les transferts xrootd ne sont pas instrumentés IPV4 vs IPV6

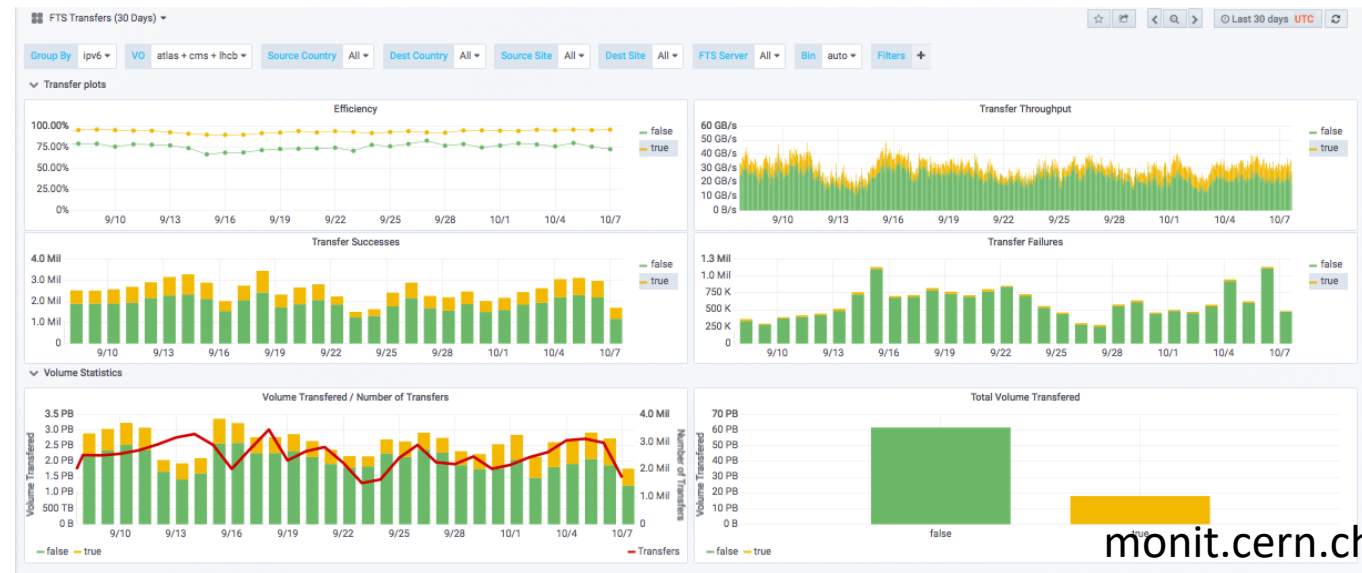
- GGEANT a déployé des instances IPV6 sur LHCONe

: [ici](#)

- De nouveaux dashboard grafana sont disponibles : [ici](#)

▶ Sécurité

- RGPD
- Vulnérabilités Meltdown, Spectre et dérivés mais aussi dans les BMC (ILO HP)



STOCKAGE ET FS

▶ Présentation des solutions de stockage au CERN

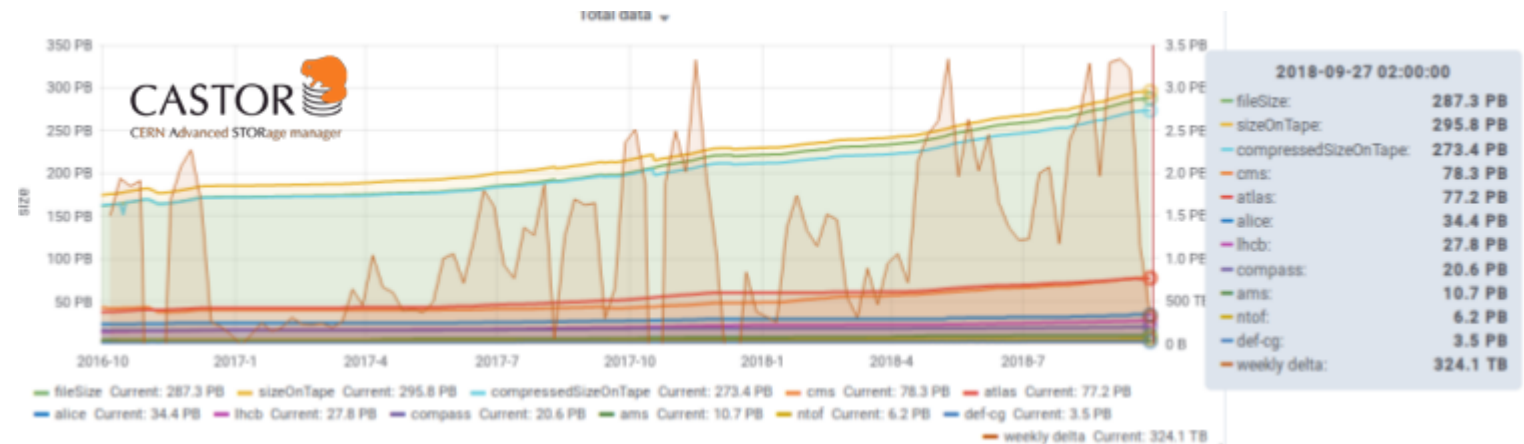
◦ Solutions pour le stockage des données scientifiques

• EOS



- Instances : 5 pour LHC, 6 CERNBox (incl EOSHOME), EOSMEDIA (photo/video), EOSPUBLIC (non LHC & OpenData) , EOSBACKUP (backup for CERNBox), 3 for various tests.
- Les difficultés sont au niveau de la gestion de l'espace de nommage

• CASTOR (CTA : Cern Tape Archive)



◦ Autres

- CERNBOX, CVMFS, Ceph/RBD for OpenStack, CephFS, NFS filers, S3

▶ Evolution des outils au CERN

◦ FTS

- Release 3.8.0
- Réutilisation des session (minimisation des authentications,...) Support des Scitoken et Macaroon en HTTP



◦ DPM

- ~110 Instances pour 90 PB managés
- Code entièrement revu en 2018 -> DOME

◦ XRRootD

- A venir Release 4.9.0

◦ EOS

- Intégration de CTA
- eoasxd : file système Posix sur EOS.

▶ Changement des bases de données (Oracle)

- Durant le shutdown LS2 (2019-2020) renouvellement du hardware
- Evolution de version vers 12.1 (12.2)

- ▶ Plusieurs présentations autour des tapes
- ▶ Une session BOF autour des exercices Carrousel
- ▶ ATLAS Data Carrousel R&D
 - Exercice fait sur l'ensemble des Tier1 pour mesurer le comportement des solutions de stockage sur bande avec un workflow bien spécifique (venant de ATLAS)
 - Tester le workflow : Rucio-FTS-Site avec un jeu de données ~200TB et des fichiers de taille importante (2-3 GB).
 - Tests fait en parallèle des activités quotidiennes.
 - Les objectifs étaient :
 - Apporter aux sites un moyen d'identifier les goulots d'étranglements à l'usage des bande
 - Orienter les sites vers des pistes d'amélioration.
 - Leurs donner un outil pour "bencher" leurs infrastructures.

Preliminary Results

- Throughput

Site	Tape Drives used	Average Tape (re)mounts	Average Tape throughput	Stable Rucio throughput	Test Average throughput
[1]BNL	31 LTO6/7 drives	2.6 times	1~2.5GB/s	866MB/s	545MB/s (47TB/day)
FZK	8 T10KC/D drives	>20 times	~400MB/s	300MB/s	286MB/s (25TB/day)
INFN	2 T10KD drives	Majority tapes mounted once	277MB/s	300MB/s	255MB/s (22TB/day)
PIC	5~6 T10KD drives	Some outliers (>40 times)	500MB/s	[2] 380MB/s	400MB/s (35TB/day)
[1]TRIUMF	11 LTO7 drives	Very low (near 0) remounts	1.1GB/s	1GB/s	700MB/s (60TB/day)
CCIN2P3	[3]36 T10KD drives	~5.33 times	2.2GB/s	3GB/s	2.1GB/s (180TB/day)
SARA-NIKHEF	10 T10KD drives	2.6~4.8 times	500~700MB/s	640MB/s	630MB/s (54TB/day)
[4]RAL	10 T10KD drives	n/a	1.6GB/s	2GB/s	1.6GB/s (138TB/day)
[5]NDGF	10 IBM Jaguar/LTO-5/6 drives, from 4 sites	~3 times	200~800MB/s	500MB/s	300MB/s (26TB/day)

[1] dedicated to ATLAS

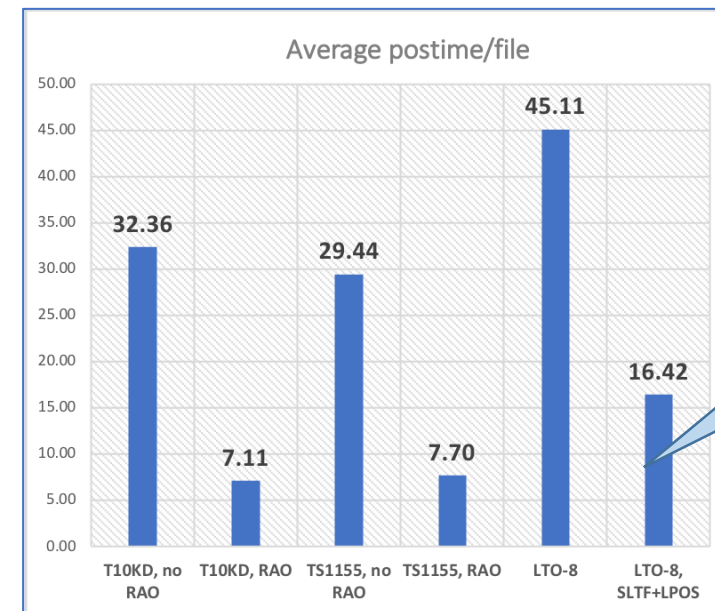
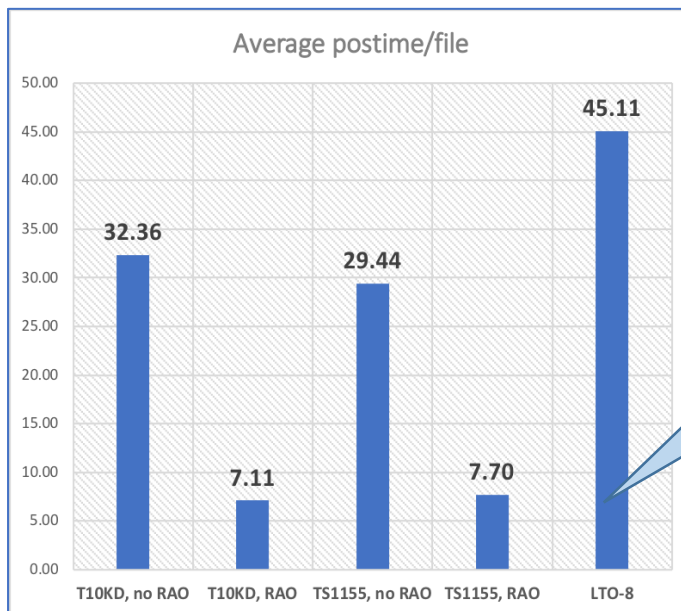
[2] with 5 drives, later increased to 6 drives

[3] 36 is the max number of drives, shared with other VOs who were not using them during the test

[4] 8 drives dedicated to this test. Will have 22 shared with other VOs in production.

[5] federated T1, 4 physical sites have tapes

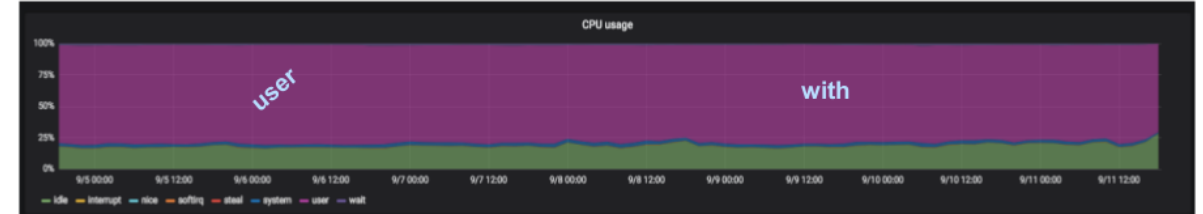
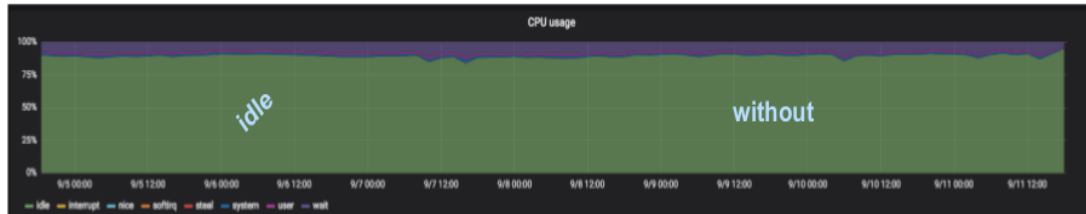
- ▶ LTO Performance
- ▶ Présentation par le CERN sur comment augmenter les performances des solutions LTO pour s'approcher de celles des gammes Enterprise.
- ▶ Approche basée sur le fait de rajouter/construire une information sur la localisation des fichiers sur la bande dans la solution LTO qui par spécificité technique est moins précise que les solutions Enterprise.



BATCH ET COMPUTING

▶ BEER at CERN

- Utilisation des machines du stockage EOS pour y faire tourner des containers afin d'étendre les capacités batch du CERN.



▶ Benchmarking group

- Pas vraiment de nouvelles
- Les HS17 (Specint 2017) n'apportent rien de nouveau par rapport au HS06 et surtout ne permettent de mieux prendre en compte les architectures nouvelles dans la mesure du bench.
- L'idée est d'aller vers des containers qui "embarqueront" de vrais applicatifs pour faire une mesure la plus réaliste possible.

▶ Evaluation des processeur EPYC de AMD

- Les EPYC sont une nouvelle gamme de processeurs "serveurs" basés sur l'architecture ZEN de AMD.
- Disponible depuis été 2017 et de plus en plus proposés par les revendeurs

How Does EPYC Differ From Skylake-SP?

Intel's Skylake-SP Xeon x86_64 server CPU line also released in 2017

Both Skylake-SP and EPYC CPU dies manufactured using 14 nm process

Skylake-SP introduced AVX512 vector instruction support in Xeon

AVX512 not available in EPYC

HS06 official GCC compilation options exclude autovectorization

Stock SL6/7 GCC doesn't support AVX512

Support added in GCC 4.9+

Not heavily used (yet) in HEP/NP offline computing

Both have models supporting 2666 MHz DDR4 memory

Skylake-SP

6 memory channels per processor

3 TB (2-socket system, extended memory models)

EPYC

8 memory channels per processor

4 TB (2-socket system)

3

How Does EPYC Differ From Skylake (Cont)?

Some Skylake-SP processors include built in Omnipath networking,
or FPGA coprocessors

Not available in EPYC

Both Skylake-SP and EPYC have SMT (HT) support

2 logical cores per physical core (absent in some Xeon Bronze models)

Maximum core count (per socket)

Skylake-SP – 28 physical / 56 logical (Xeon Platinum 8180M)

EPYC – 32 physical / 64 logical (EPYC 7601)

Maximum socket count

Skylake-SP – 8 (Xeon Platinum)

EPYC – 2

Processor Interconnect

Skylake-SP – UltraPath Interconnect (UPI)

EPYC – Infinity Fabric (IF)

PCIe lanes (2-socket system)

Skylake-SP – 96

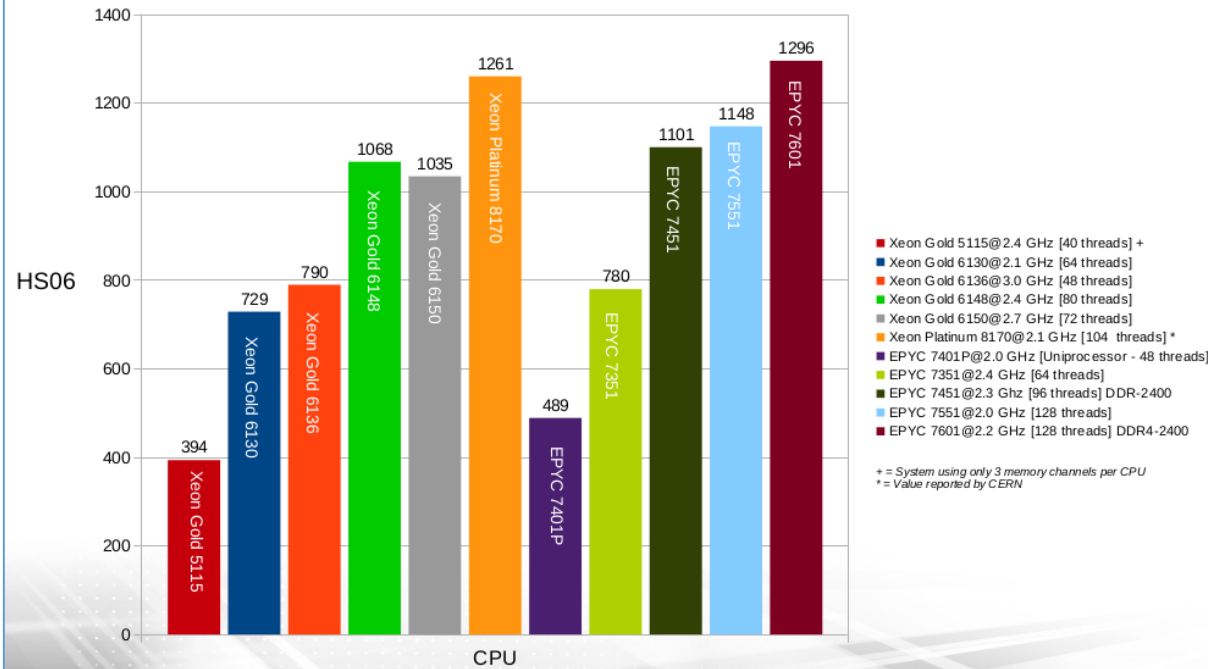
EPYC – 128 (some used by SoC functionality)

Same number available in single socket configuration

4

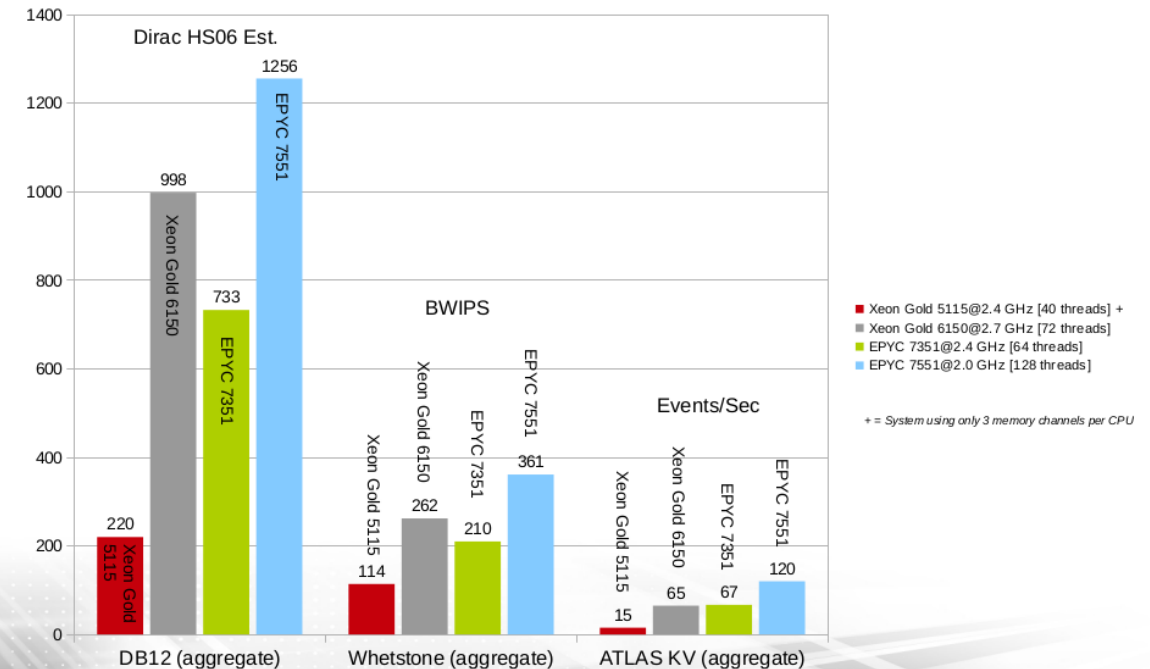
- ▶ Comparatif effectué avec des benches : HS06, DB12, Whetstone, ATLAS KV
 - La configuration des canaux de mémoire à un impact significatif sur les performance des l'EPYC

EPYC vs Skylake-SP: HEP SPECC06



11

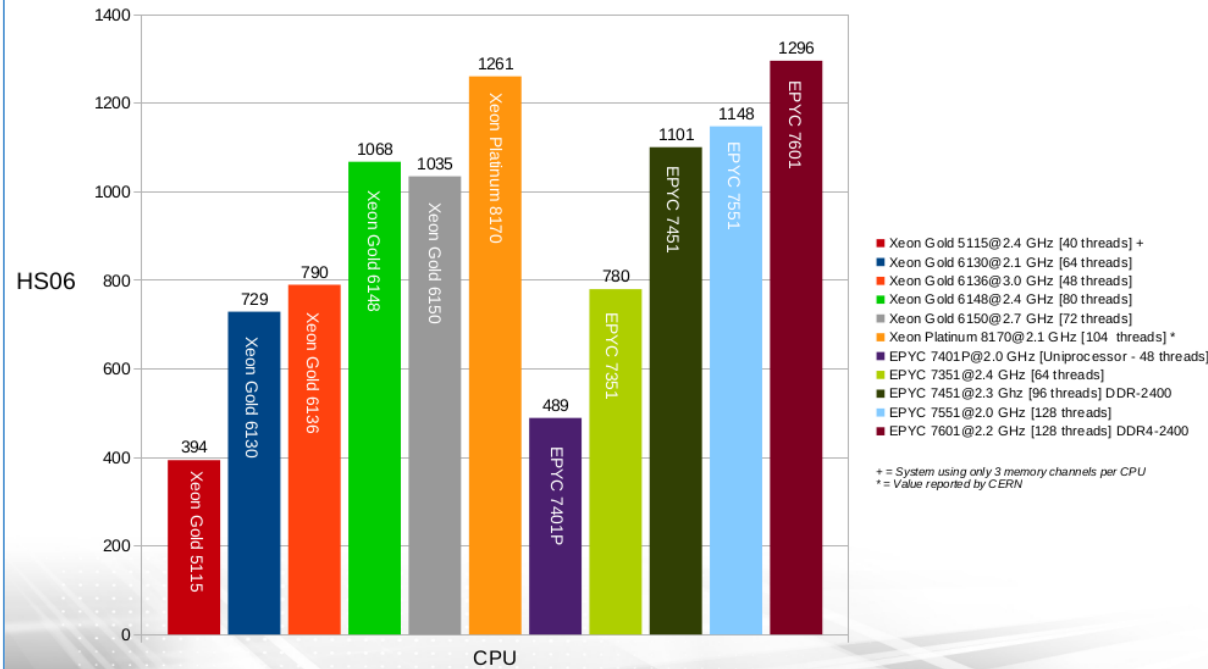
EPYC vs Skylake-SP: CERN Cloud Benchmarks



13

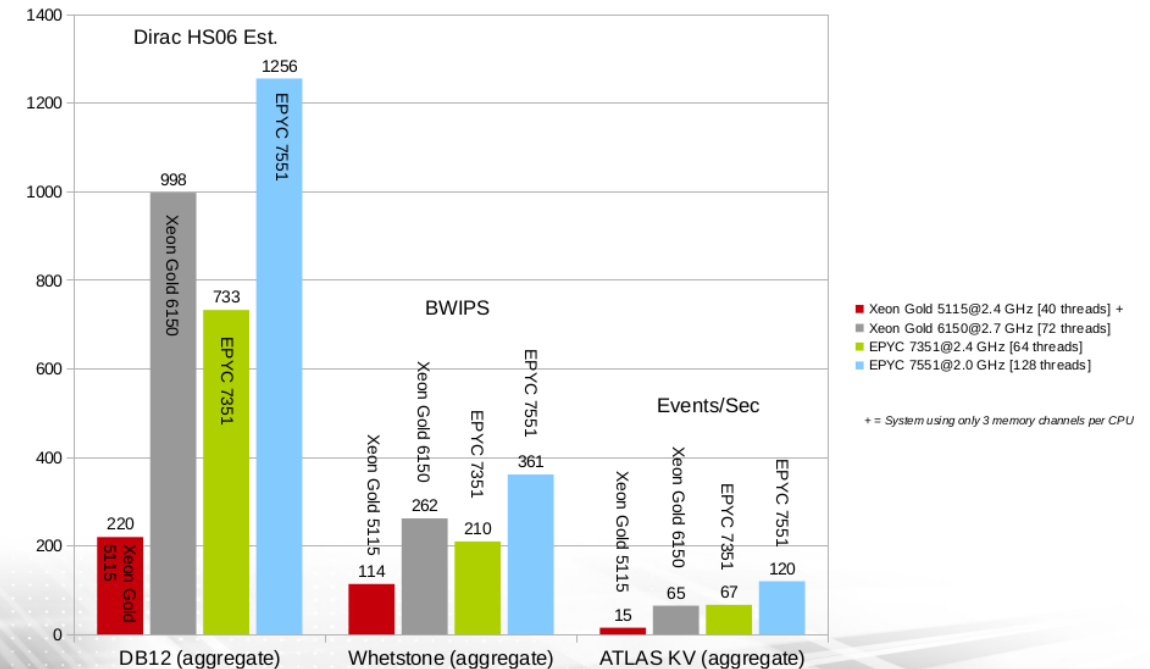
- ▶ Finalement peu de différences sur les mesures de performance.
- ▶ Peu de différence aussi en terme de consommation

EPYC vs Skylake-SP: HEPSPREC06



11

EPYC vs Skylake-SP: CERN Cloud Benchmarks

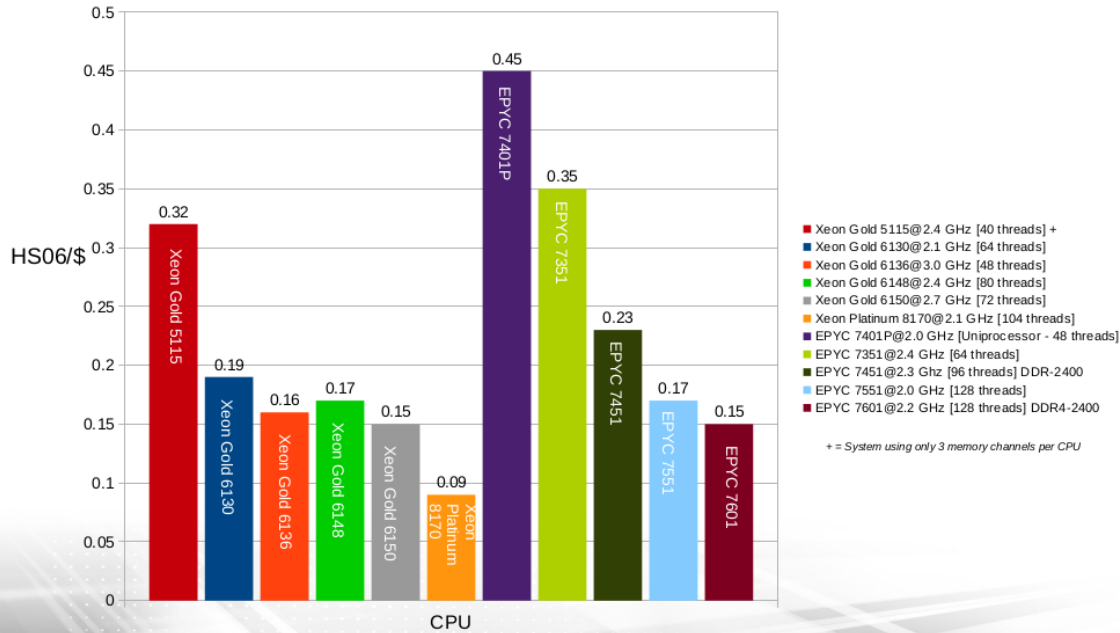


13

- ▶ Au premier ordre peu de différences sur les mesures de performance.
 - Sur les bench plus “maisons”, les EPYC s’en sortent plutôt bien (effet nombre de core)
- ▶ Peu de différences aussi en terme de consommation

EPYC vs Skylake-SP: CPU HS06/Dollar

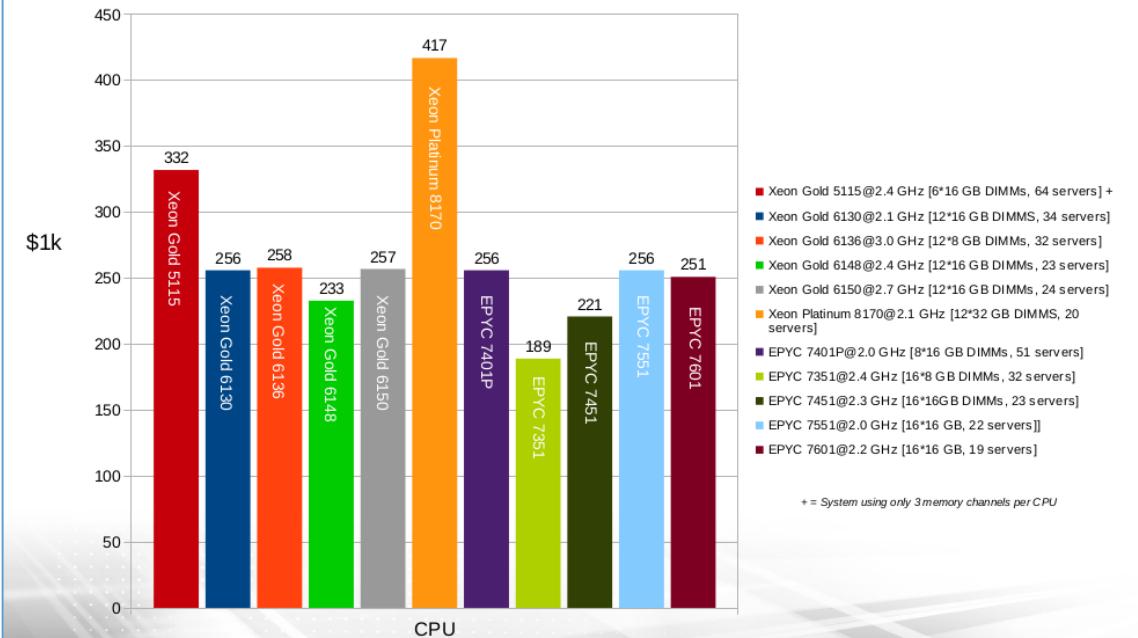
Only retail CPU cost accounted for in calculations
Does not represent reality given memory and base server pricing



15

EPYC vs Skylake-SP: Estimated 25kHS06 Cost

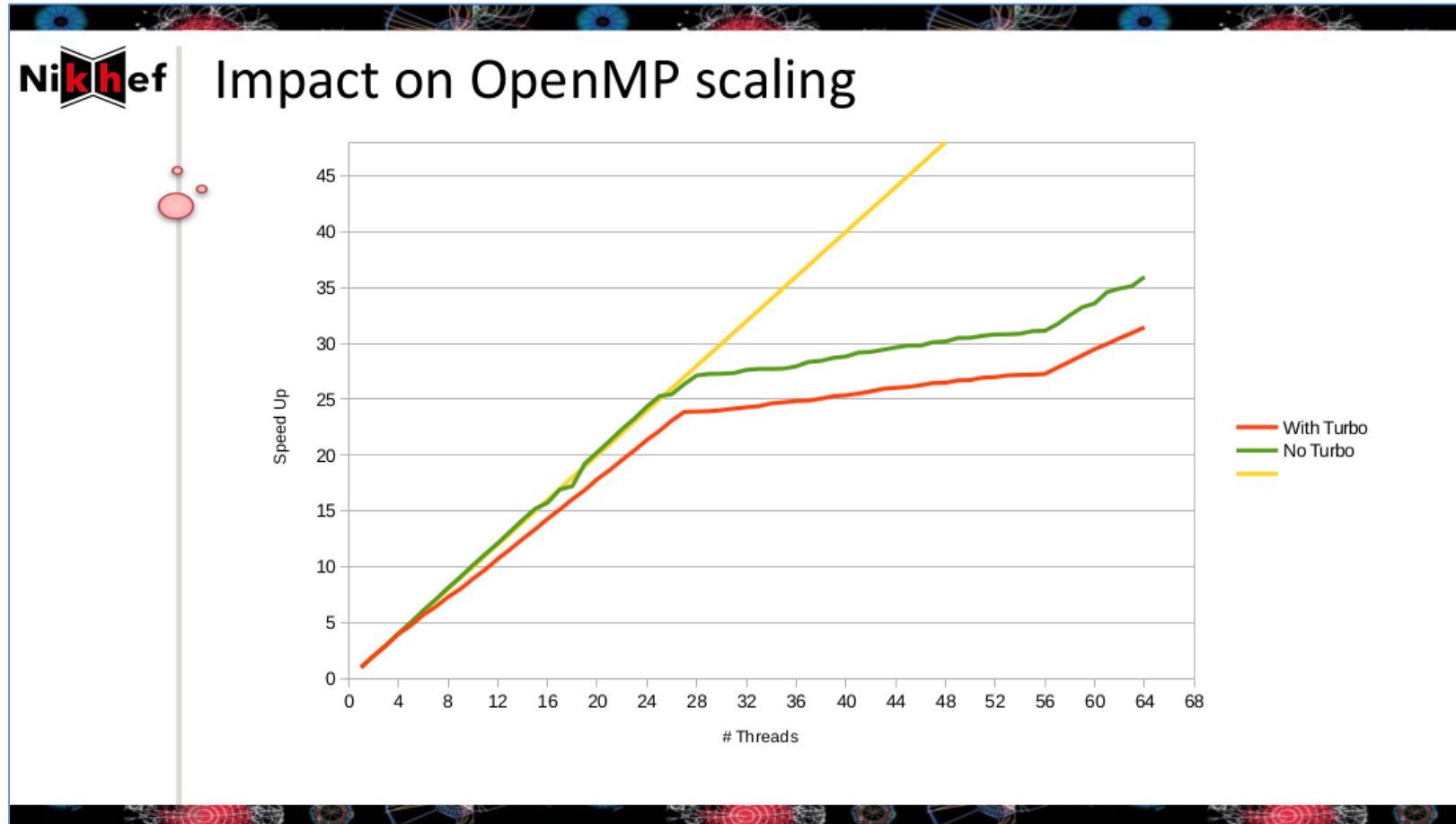
Estimated total cost of 25kHS06 +-500HS06
Assuming \$1,500 irreducible server cost, and retail CPU/memory pricing



16

- ▶ A l'arrive peu de différences de coûts.
- ▶ Exercice à faire au cas par cas.

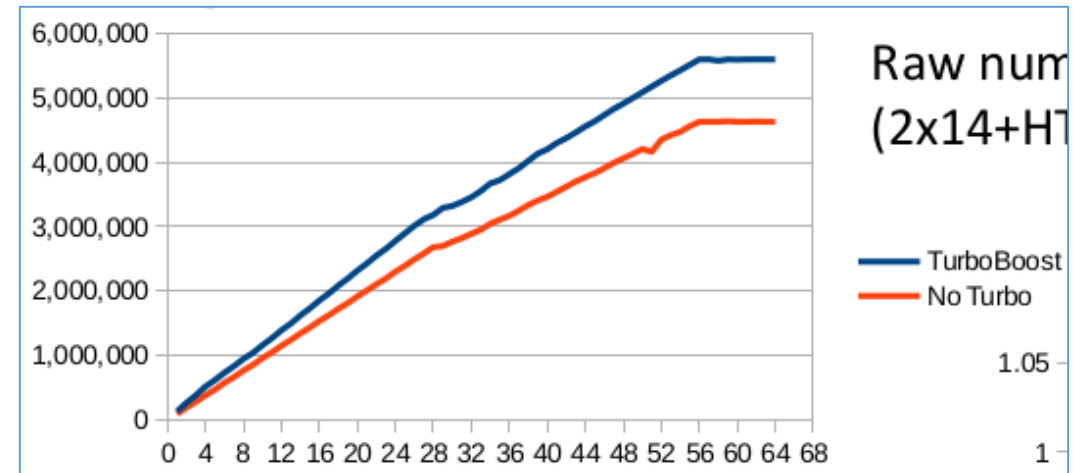
- ▶ Improving OPENMP scaling with openssl
 - Le constat de départ est que les solutions implémentées par les fondeurs pour gérer la fréquence des cores (turbo xxxxx) biaise le scaling de openMP



- L'idée c'est utiliser openssl pour anticiper le comportement de openmp dans un environnement où les fonctions turbo xxx sont actives/désactivées

- **Avantage de openssl**

- Scale bien avec le nombre de core (cf ci-contre)
- Disponible en espace utilisateur
- Ne nécessite que quelques secondes de calcul
- Ne change quasiment plus (code gelé ou presque)



- **Conclusions**

- OpenMP scalability can be tested in userspace without having to turn off features like Turbo Boost or Hyperthreading
- Using a set of 'openssl speed' commands a correction factor can be calculated and applied to adjust for Turbo Boost frequencies
- The same set of 'openssl speed' commands can be used to quickly determine the configuration and performance of an (unknown) CPU in userspace

BASIC IT SERVICE

- ▶ Cern Megabus
- ▶ New CERN authentication and authorization

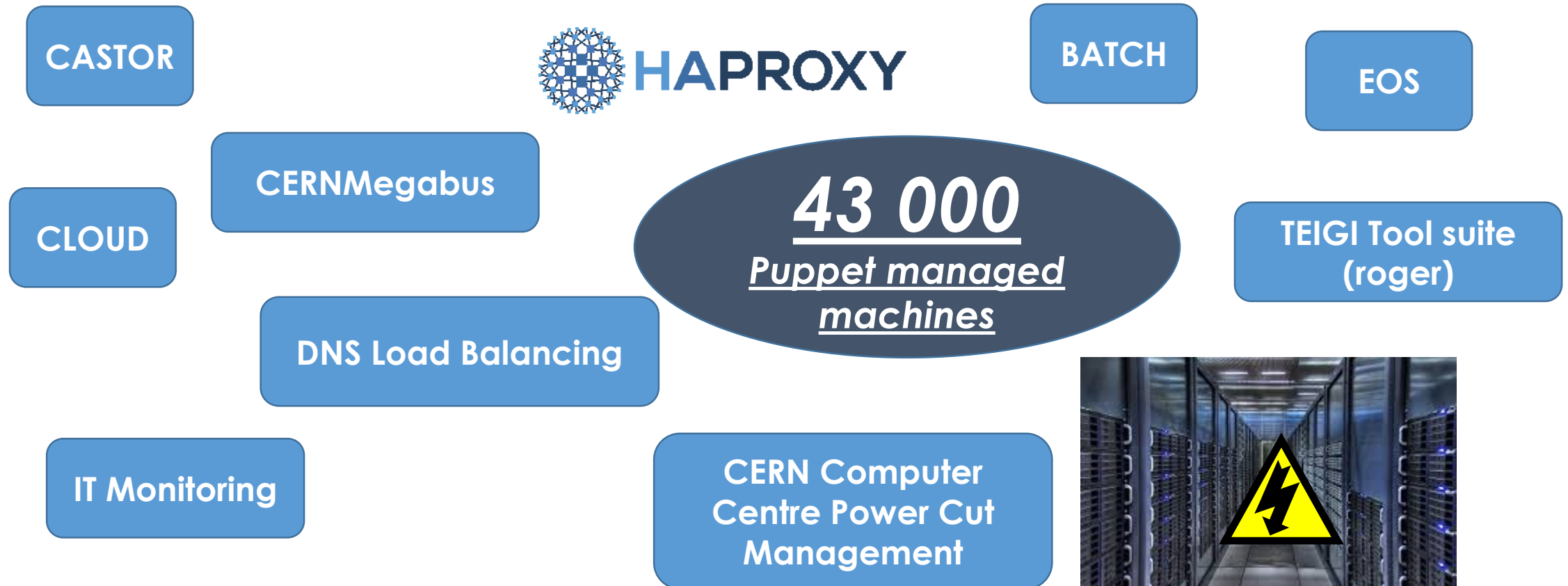
CERNMegabus at CERN IT

A service that provides for instant communication between services

CERNMegabus

- The CERNMegabus architecture is based on the **publisher-consumer model** and utilises the **CERN IT messaging infrastructure**
- The publisher and the consumer services comprises of building blocks
 - **configured with Puppet**
 - to use the **CERNMegabus python libraries**
- **Installed on all Puppet managed machines** in the **CERN CC**

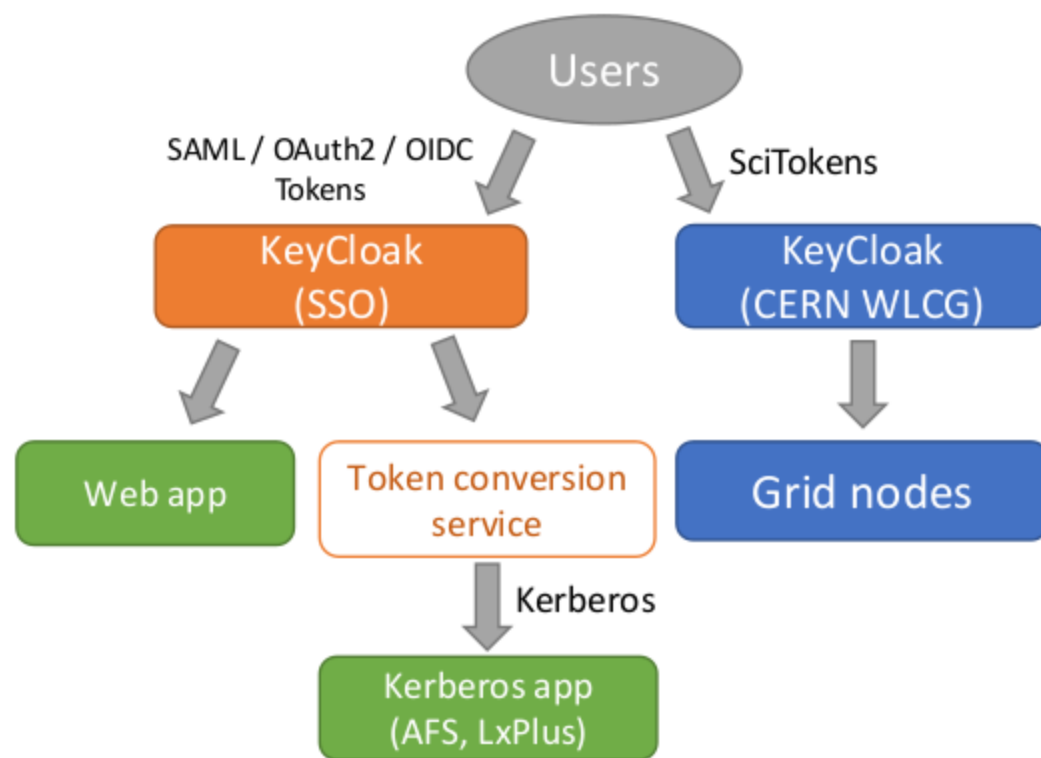
CERNMegabus at CERN IT



- ▶ **New CERN authentication and authorization**
- ▶ **Statut actuel**
 - Kerberos pour authentication (desktop/terminal, credential local)
 - Sso authentication (multifactor authentication, web application)
 - Autorisation basée sur des groupes (nécessite d'avoir des comptes locaux)

 - X509 pour les authentication WLCG
 - VOMS pour les autorisations
- ▶ **Volonté d'évoluer vers :**
 - Architectures plus généralistes (pas CERN centrique)
 - Des interfaces et façons de faire uniformes (notamment pour les utilisateurs).
 - Utiliser des mécanismes plus "récents".

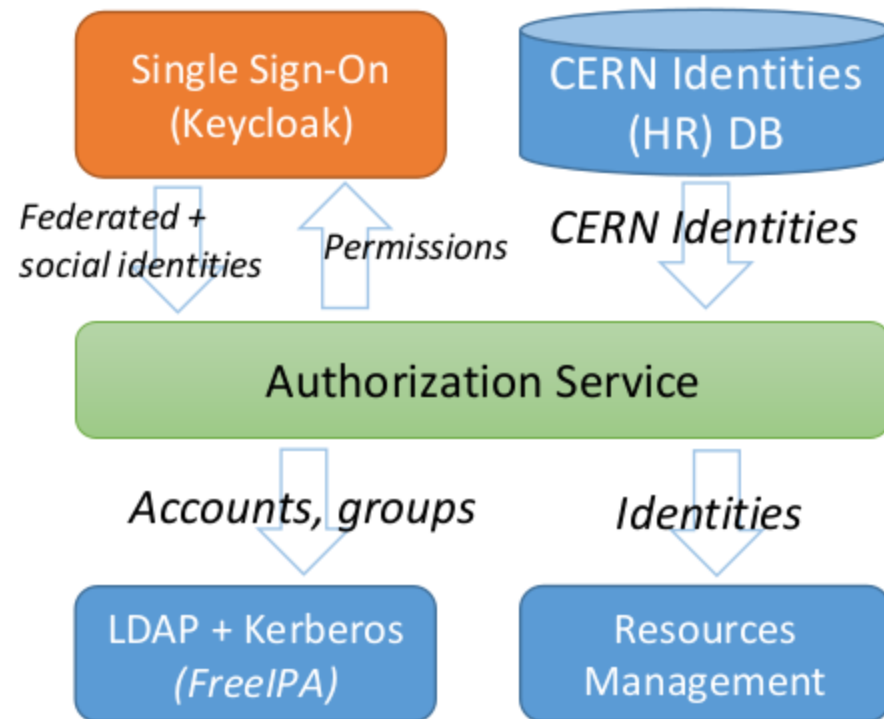
New authentication



- Tokens at the heart
 - WLCG alignment
- Single Sign-On for all
- Token conversion service



New authorization



Full federation support

Identities management

- Map account(s) to an identity

Application-specific roles

- Levels of Assurance, MFA
- Reduce privacy impact

