

CHEP2023 : infrastructures et services

Michel Jouvin

31/5/2023

Impressions générales

- Enorme plaisir de retrouver en présentiel tant de collègues
 - Les “anciens” bien sûr
 - Les nouveaux : tous ceux connus dans des reunions Zoom depuis 3 ans...
- Participation des “grands CHEP” : ~650
- Tracks bien fournies
 - Peu de “présentations de rupture” (à part sur le ML/IA peut être...)
 - Beaucoup de travail pour creuser les idées discutées les dernières fois...
- Accroissement de l’ouverture aux communautés autres que HEP
 - Forte présence de la communauté HENP avec EIC : forte intégration avec HEP
 - Astrocosmo : une session de keynotes, de multiples présentations
- Synthèse: tracks Data Management (1), Distributed Computing (4), Facilities & Virtualization (7)
 - Pas une vue exhaustive de chaque

L'émergence du problème de l'énergie...

- Suite à la crise énergétique, une préoccupation partout...
 - Driver : difficulté à assumer les coûts dans des budgets constants
 - Mais aussi la question de la responsabilité dans la question climatique (CO2)
 - 2 questions dont la corrélation varie en fonction de la géographie...
- Plusieurs présentations sur des estimations de l'empreinte carbone de sites
- Discussions sur empreinte carbone durant la phase d'exploitation vs. fabrication/démantèlement
- Plusieurs rénovations de datacenters pour en améliorer l'efficacité énergétique
 - Un facteur important dans les économies d'énergies : moyenne des PUE WLCG ~1,55
 - PUE assez agressifs dans les nouveaux projets, ex: CERN Preveissin ~1,1 (VirtualData ~1,25)

... L'émergence du problème de l'énergie

- Keynote de S. Campana une « étude holistique des besoins énergétiques de WLCG »
 - <https://indico.jlab.org/event/459/contributions/11499/attachments/9236/14205/WLCGEnergyNeedsCHEP2023.pdf>
 - Projection moins pessimiste que par le passé : les derniers processeurs AMD ont amélioré l'efficacité énergétique des processeurs (HS06 / Watt)
 - GPUs apportent un gain d'efficacité significatif quand ils sont fortement utilisés : extension de leur utilisation dans les fermes online, encore marginal dans le offline computing
 - Espoir pour l'avenir : l'architecture ARM64 aujourd'hui ~30% plus efficace, des marges d'amélioration plus grandes que AMD/Intel
 - Role important du logiciel et des computing models reflété dans une nouvelle unité : GWh/fb-1
 - Amélioration d'un x10 entre Run 1 et (prévision) Run 5 (2035)
 - Dans le scénario optimiste, +10% d'électricité en 2035 par rapport à 2010 pour x400 luminosité
 - Cache aussi qu'on améliore l'efficacité, pas vraiment la consommation totale...

Hétérogénéité des infrastructures distribuées

- Toutes les expériences (présentes à CHEP) utilisent des infrastructures de calcul distribuées
 - Besoin de maximiser les points communs entre les expériences : généralement plusieurs expériences par site fournisseur de calcul
- Cloud (commerciaux) et HPC sont devenus des first-class citizens
 - Plus uniquement des ressources d'appoint opportunistes
 - Malgré les défis techniques, « retour sur investissement » des efforts demandés
 - Adoption de frameworks communs par les expériences (ex: PanDA, DIRAC) facilitent l'intégration pour les différentes expériences
 - Pas toutes les workloads partout : nécessité de prendre en compte le type de tâche que peut exécuter une ressource
 - HPC : augmentation importante du nombre de workloads qui peuvent tirer parti efficacement du parallélisme et/ou des GPUs d'un calculateur HPC, solutions possibles aux restrictions de connectivité externe (cache)

Sites : mutualisation des solutions

- Beaucoup de sites exposés à une multitude de demandes : impossible de déployer des solutions ad-hoc pour chaque besoin
- Computing : HTCondor, cloud, Kubernetes
 - HTCondor : le système batch le plus répandu, ressources HW généralement dédiées
 - Cloud / Kubernetes : principalement provisioning de ressources dynamique pour des besoins autres, Kubernetes partout !
 - CVMFS s'est élargi de la distribution d'applications à la distribution de containers
 - Cas du CERN : unification de la gestion des ressources HTCondor (batch) avec celles du cloud en utilisant le service de provisioning de ressources « bare-metal »
- Stockage : importance de xrootd, forte présence dCache, croissance de Ceph
 - Xrootd (/EOS) et dCache restent les composants dominants de la gestion de données distribués
 - Ceph : utilisation croissante soit directement (S3), soit comme backend de Xrootd (CephFS)

Data management

- Rucio (+ FTS) everywhere !
 - Adopté par toutes les grandes collaborations scientifiques astro & cosmo ou NP, en plus de HEP
- Intégration avec les clouds commerciaux
 - S3 est le protocole principal
 - Optimisation des coûts de transfert de/vers le cloud
- Cache
 - Peuvent jouer un rôle important pour les performances (latency hiding)
 - Un service déployé par OSG (USA) : StashCache, basé sur xrootd
 - Analyse/simulation/modélisation de l'efficacité des caches
 - Peut-on faire plus efficace que de l'éviction LRU: le problème de la prédiction (ML)

Authentification globale

- La problématique de l'authentification (et de la gestion des autorisations) est au cœur des infrastructures distribuées
 - La plupart des communautés y est donc confronté
- Un changement majeur en cours : remplacement des certificats par les « web tokens »
 - Certificats : extrêmement robuste mais un cauchemar pour les utilisateurs, essentiellement les expériences LHC
 - Web tokens : technologie très mature provenant des services internet permettant la fédération des identités (utilisation de vos credentials locaux) et la délégation entre service
 - Défis de l'intégration technique sont nombreux : bien avancé pour la partie soumission de jobs, encore en cours pour la partie accès aux données
 - Devrait beaucoup améliorer l'expérience utilisateur des infrastructures distribuées

Réseaux

- IPv6 : action de près de 10 ans dans WLCG pour que tous les services acceptent à la fois des connexions IPv4 et IPv6
 - Motivé par la pénurie d'adresses IPv4
 - Entre 95% et 100% des ressources des expériences LHC
 - Vers des sites seulement IPv6 : probablement quelque chose pour les années à venir
- Evolution de l'infrastructure de transfert de données LHCONE
 - Configuration spécifique du routage par les network providers pour optimiser les gros transferts et éviter/limiter l'impact sur les autres usages
 - IPv6 apporte de nouvelles fonctionnalités pour faire ça de façon plus souple
- Liens 400 Gb/s arrivent
 - USA/Canada les plus actifs
 - Validation que xrootd sait utiliser efficacement de tels liens

Nouveau Benchmark : HEPScore23

- Besoins d'une unité commune entre les expériences et les sites pour exprimer les besoins et les ressources fournies
 - Rendu complexe par l'utilisation des mêmes ressources par plusieurs expériences aux workloads diverses
- Approche HEP jusque là : version « customisée » du benchmark standard de l'industrie, SPEC
 - HEP SPEC 2000, HEP SPEC 2006 (HS06)
 - Décalage grandissant avec les applications réelles : corrélation toujours « par chance »
- Depuis le 1 avril, nouveau benchmark, HEPScore23
 - Fruit de plusieurs années de travail de définition et validation
 - Basé sur des applications réelles des expériences (LHC + Belle II), possibilité d'ajouts futurs
 - Architecture x86 mais aussi ARM64 et à l'avenir GPU

Conclusions

- CHEP reste une conférence exceptionnelle pour les disciplines d'IJCLab
 - Atout : toutes les problématiques de l'informatique représentées et discutées, une occasion unique pour favoriser les convergences ou élargir les perspectives
 - Au-delà des présentations, richesse des contacts informels du fait du nombre de personnes présentes
- Variété et richesse des présentations très vivifiantes
- IJCLab de mon point de vue sous-représenté
 - Même si un investissement important consenti par le laboratoire pour permettre la participation de 7 personnes...
 - Exploitation totalement absente (à part moi) et le online peu représenté (une personne à la frontière du online) : renforcer ces composantes à l'avenir
- Prochain CHEP : 21-25/10/2024, Cracovie