

WLCG et le Cloud

Michel Jouvin, CNRS/LAL

Journée LCG France

Décembre 2015

Différents Types de Cloud

- Infrastructure as a Service (IaaS) : fournit des machines virtuelles qui peuvent être utilisées en lieu et place d'une machine physique
 - Même besoin de les configurer et de les administrer mais permet un contrôle de l'utilisateur sur l'environnement d'exécution
 - Démarrage à la demande à travers une API : dominé par les API propriétaires (Amazon EC2, Open Stack Nova, MS Azur...) même si des APIs « standard » existent (OCCI, CIMI)
 - **Pas de notion de jobs**
 - Le type de cloud dont nous allons parler aujourd'hui
- Platform as a Service (PaaS) : fournit des (web) services permettant de construire des applications
 - Les services peuvent s'exécuter ou non dans un cloud IaaS
 - Exemple : Google Application Engine
- Software as a Service (SaaS) : typiquement une application web
 - Les applications peuvent elle-même utiliser des web services ou bien s'exécuter dans un cloud IaaS
 - Exemple : gmail

Intérêt de IaaS pour WLCG...

- Contrôle de l'environnement d'exécution
 - Pas la raison majeure : toutes les expériences utilisent le même environnement, WN grille dominé par les requirements WLCG
 - Peut permettre d'éviter la synchronisation entre les expériences lors des changements de version majeure d'OS (ex: EL6 -> EL7)
- Allocation dynamique des ressources HLT en fonction des besoins
 - Fermes HLT \geq T1
 - Partage des ressources entre activité trigger et online
 - Mise en œuvre par ATLAS et CMS (ALICE?)
 - LHCb : stratégie différente pour faire la « reconstruction finale » au niveau du online entre les fills
 - Peut être étendu à d'autres types de ressources
- Accès à de nouvelles ressources de computing fournies par des sites non grille
 - Aucun requirement particulier pour le site
 - Accès à des providers commerciaux : nécessite éventuellement la capacité de tirer parti des spot instances (arrêtables à tout moment en fonction du prix de la ressource)

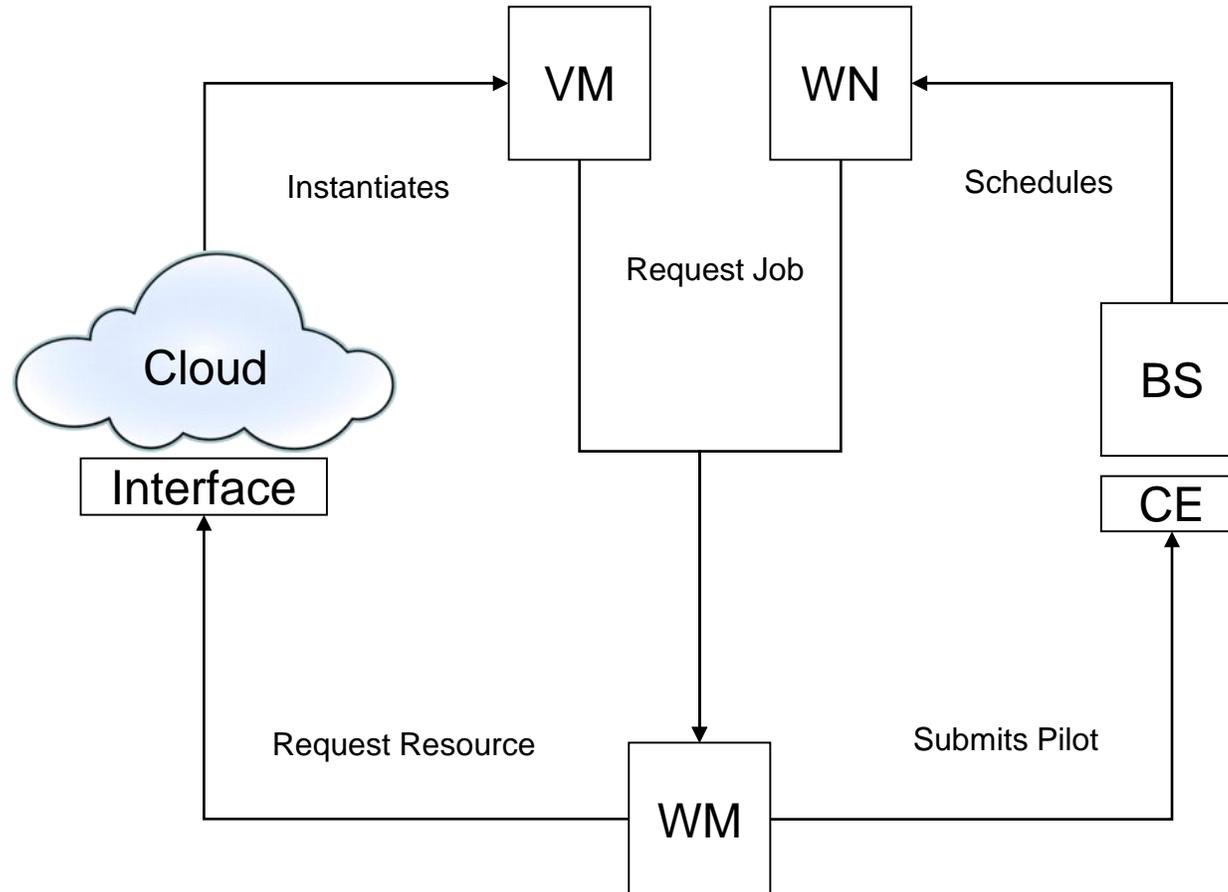
... Intérêt de IaaS pour WLCG

- Utilisation d'une méthode standard de provisionning de ressources de computing diminuant le coût d'opération de WLCG
 - Cloud everywhere
 - L'infrastructure préférentielle des autres communautés scientifiques
 - Des middleware clouds open source maintenus par des communautés larges (ex: OpenStack)
 - Plus besoin de maintenir ni d'opérer un CE, quelque soit sa technologie
- Overhead négligeable par rapport à des machines réelles
 - A condition d'un tuning correct de l'infrastructure : peut être difficile avec des clouds commerciaux...
 - CERN a démontré à grande échelle un overhead < 5%
 - Les containers pourraient diminuer significativement cet overhead

Atouts de WLCG pour Utiliser des Clouds

- Pilot job frameworks
 - VM provisioning caché à l'utilisateur
 - VM image générique pour chaque expérience, non par job : essentiellement une machine qui se contacte à la pilot factory
 - Possibilité d'exécuter plusieurs jobs (d'une même VO)dans une VM pour limiter l'overhead : même problématique que les multi-users pilot jobs
 - Possibiliter d'implémenter du fair share avec une approche comme vcycle : exécution de la VM générique de l'expérience en dessous de son share lorsqu'un slot se libère
- CVMFS : rend très simple l'accès au SW de l'expérience
- Chaque expérience a déjà une expérience des machines virtuelles hors cloud
 - Projet μ CERNVM : un (petit) bootstrap, VM image dans CVMFS
- Une infrastructure de monitoring dans lequel le cloud est facilement intégrable

Cloud vs. Batch Provisionning



From Laurence Fiels,
Laurence.Field@cern.ch

Problèmes posés par IaaS pour WLCG...

- Tâches sysadmin déportées du côté de l'expérience
 - Minimisation possible avec CERNVM... mais une tâche supplémentaire
- IaaS ne connaît pas les jobs : éviter de remettre un batch system... atout des pilot factories
- Complexification du suivi des incidents de sécurité
 - La traçabilité requiert la collaboration site/VO car le site ne voit plus les payloads : déjà en partie vrai avec les pilot jobs
 - Plus grande importance du monitoring réseau pour pouvoir faire la corrélation avec le comportement des VMs
 - Une VM sur laquelle l'utilisateur peut être root n'est pas trustable par nature... Dans le cadre WLCG, aucune nécessité de root pour l'utilisateur (seulement la pilot factory)
- Corrélation entre accounting VO et accounting site
 - Le site voit des VMs mais pas les payloads qu'elles font tourner, la VO voit les jobs qu'elle soumet
 - Possibilité d'un accounting CPU-based dans les clouds ? Besoin de passer à un accounting WC ?
 - Nécessité d'un benchmark de chaque VM : une VM n'est pas liée à un type HW prédéfini
 - Besoin d'un fast benchmark (~1-2 mn) alors que HS06 nécessite plusieurs heures...

... Problèmes posés par IaaS pour WLCG

- Transmission sécurisée des identifiants
 - Identification de la VM vis-à-vis de la pilot factory
 - Identifiants utilisateurs : pris en charge par le pilot job
- Accès aux données
 - Quels protocoles : le but est de ne pas installer la stack grille dans la VM
 - WebDav ?
- Partage dynamique des ressources entre communautés/VOs/groupes
 - Dans un batch system, le partage dynamique (ex: fair share) est basé sur l'existence d'une queue de jobs en attente
 - Feature du cloud : absence de queue
 - Requête acceptée ou refusée immédiatement
 - Basé sur le principe de « l'infinité » des ressources (elasticity)
 - Un partage statique conduirait à un gaspillage de ressources hors des cas comme le T0

Quelques Expériences Existantes...

- Cluster HLT des expériences : principalement utilisés pour de la simulation actuellement
- CERN Agile Infrastructure : l'essentiel des services informatiques du CERN s'exécutent dans un cloud
 - Une grande fraction sont des grid WNs virtualisés, gérés par un batch système
 - Une utilisation directe du cloud croissante par les expériences : en particulier pour la partie T0
- Helix Nebula : collaboration européenne autour de l'utilisation de cloud commerciaux
 - Pas de vendor lock-in : nécessité pour les providers de fournir un des interfaces supportés par les expériences
 - Brokering inter-cloud via SlipStream
 - Simulation uniquement pour l'instant
- HNSciCloud : European PCP project autour de l'achat de « services novateurs » dans des clouds
 - Plusieurs fournisseurs (au moins 2)
 - Des tâches data-intensive
 - Démarre en janvier par une phase d'élaboration des requirements (6 mois) : CCIN2P3 participe

... Quelques Expériences Existantes

- **vcycle** : service permettant un partage dynamique des ressources dans un cloud
 - VM instancié par le service et non pas par la pilot factory de la VO
 - Le service contient une liste d'image par VO/groupe
 - Le service maintien un historique de la consommation de chaque VO/groupe et décide sur cette base quelle VM image instancier lorsqu'un slot se libère
 - Si la VO n'a rien en attente, elle arrête la VM et le slot est donné à une autre VO
 - Basé sur le concept Vac (Vacuum), développé par Andrew McNab (Manchester)
 - Implémentation existante pour OpenStack
- **Accounting** : APEL + portail accounting (version dev) supportent les machines virtuelles
 - Travail fait dans le cadre du EGI Federated Cloud
 - Parser disponible pour les principaux cloud MW open source

Les Containers

- Container = lightweight virtualization
 - Restriction : l'OS du container doit être de la même famille que celui de la machine physique
 - Utilise une fonctionnalité de l'OS (chroot) et nécessite la compatibilité binaire des interfaces du kernel, un feature Linux en particulier
 - Pas une full image mais seulement ce qui diffère par rapport à la machine de base (y compris le kernel éventuellement)
 - Pas de partage des points de montage, pas de possibilité d'interagir avec les daemons de l'image de base (ex: problème pour AFS)
 - Instanciation très rapide : quelques secondes
 - (presque) aucun overhead
 - Des repositories de containers tels Docker
- Des orchestrateurs de containers existent déjà, comme Mesos
- Cloud MW vont supporter les containers
 - OpenStack : support basic des Linux Containers, support à venir de Docker

Quelques Questions pour LCG France

- Quelle participation à la réflexion WLCG ?
 - Prochaine échéance : WLCG workshop de Lisbonne, 1-3 février
- Quel avenir pour nos clouds ?
 - Aujourd'hui ce sont des ressources très modestes par rapport à nos CEs grille
 - La taille est un élément clé d'un cloud pour qu'il tienne ses promesses
- Partage dynamique des ressources entre communautés utilisatrices
 - Quelle besoin/expérience dans les sites ?
 - Quels outils ?
- Virtualisation des WNs pour unifier l'infrastructure du site
 - Quels sites envisagent cette unification ?
 - Une approche pour augmenter la taille des clouds ?
 - Requiert la capacité d'ajuster dynamiquement le nombre de WN en fonction de la charge : impact sur le batch
- Accès distant aux données : un requirement pour les clouds
 - Performance limitée des I/O locales sur une machine virtuelle