

## UV « Sensibilisation à la recherche » OpenStack et architecture multi-sites

16 septembre 2016

<b>Tuteur</b>	Adrien Lebre
<b>Projet</b>	<p>Fog/Edge Computing</p> <p>OpenStack est une suite logiciel permettant d'opérer des plateformes de Cloud Computing de type IaaS. Bien que le passage à l'échelle (le terme de scalabilité est généralement utilisé) soit un critère décisif lors du choix d'un système pour opérer ce genre de plateforme, de nombreuses briques fondamentales constituant OpenStack ne satisfont pas ce critère, notamment lorsque vient s'associer au problème du passage à l'échelle celui de géo-distribution. L'initiative Discovery pilotée par l'équipe ASCOLA étudie de manière approfondie le système OpenStack afin de déterminer les principales faiblesses et proposer des modifications permettant de pallier.</p> <p>Le travail proposé dans ce projet s'inscrit dans cette action. il consiste à :</p> <ul style="list-style-type: none"><li>— Prendre en main l'outil Kolla-G5K utilisé pour l'évaluation des performances du système OpenStack mono-site</li><li>— D'étendre ce même outil afin de permettre aux chercheurs d'étudier des déploiements multi-sites (<i>i.e.</i>, multi-sites)</li><li>— Valider les modifications proposées au dessus de la plateforme Grid'5000.</li></ul> <p>A plus gros grain, ce projet s'insère dans l'initiative Discovery pilotée par l'école.</p>
<b>Site Web</b>	<a href="http://beyondtheclouds.github.io">http://beyondtheclouds.github.io</a>
<b>Intitulé</b>	Extension de l'outil Kolla-G5K pour l'évaluation de performance d'OpenStack en contexte multi-sites

**Domaines**

- Systèmes distribués
- Cloud Computing (IaaS)
- Python

**Compétences requises**

- Fondamentaux Cloud Computing
- Programmation objets/composants

**Compétences à acquérir**

- Expertise sur l'architecture OpenStack
- Sensibilisation à l'évaluation de performances de systèmes complexes
- Contribution à un projet « open source »
- Programmation Python

**Contexte**

Afin de supporter la demande croissante d'informatique utilitaire (UC : Utility Computing) tout en prenant en compte les aspects énergétiques et financiers, la tendance actuelle consiste à construire des centres de données (ou centrales numériques) de plus en plus grands dans un nombre limité de lieux stratégiques. Cette approche permet sans aucun doute de satisfaire la demande tout en conservant une approche centralisée de la gestion de ces ressources mais elle reste loin de pouvoir fournir des infrastructures de calcul utilitaire efficaces et durables. L'initiative de recherche DisCoVery propose une approche radicalement différente permettant de prendre en compte la localité des demandes dès le départ. Pour ce faire, elle propose de tirer parti de tous les équipements disponibles sur l'Internet afin de fournir des infrastructures de calcul utilitaire qui permettront de part leur distribution de prendre en compte plus efficacement la dispersion géographique des utilisateurs et leur demande toujours croissante [1]. Un des aspects critique pour l'émergence de telles plates-formes de calcul utilitaire (peu à peu connues sous les termes Fog et Edge Computing) est la disponibilité de mécanismes de gestion appropriés exploitant les dernières contributions en terme d'algorithmiques distribués. Des premiers travaux menés à l'école des Mines ont permis de valider plusieurs mécanismes via la mise en œuvre de prototypes expérimentaux. Afin de passer du mode recherche à un mode plus industriel, nous avons commencé à regarder comment intégrer nos propositions au système OpenStack.

**Maîtriser OpenStack pour comprendre comment le modifier.**

**Problème : Les faiblesses dans la conception d'OpenStack**

OpenStack est une suite logiciel permettant d'opérer des plateformes de Cloud Computing de type IaaS. Elle est aujourd'hui la solution *Open Source* retenue et supportée par un grand nombre d'acteurs aussi bien académiques et industriels (850 institutions sont impliquées dans son développement, avec des acteurs de premier rangs tels que HP, RedHat ou encore plus récemment Google). Bien qu'une des règles fondamentales pour le développement du système a pour objectif d'adresser la problématique du passage à l'échelle de chaque composant, le système comprend plusieurs briques logicielles qui ne répondent pas à cette règle. A titre d'exemple, Jonathan Pastor, candidat à une thèse de doctorat à l'école, a identifié la base de données MySQL au sein du composant Nova et a proposé un pilote permettant de basculer sur un modèle NoSQL passant à l'échelle de manière native.

**Objectif**

L'objectif du travail proposé dans ce projet vise à étendre l'outil d'évaluation de performances Kolla-G5K<sup>1</sup> développé dans le cadre de l'initiative Discovery. Cet outil permet à ce jour de deployer une infrastructure OpenStack mono site dessus de la plateforme expérimentale Grid'5000 via la technologie Docker et Ansible. De part l'objectif visé par Discovery, à savoir la supervision d'infrastructures de type Fog/Edge Computing, il est primordiale d'étendre les fonctionnalités de l'outil actuel afin de permettre aux chercheurs du consortium d'évaluer la solution OpenStack dans un contexte multi-site et d'identifier les autres freins à la supervision d'architectures massivement distribuées.

Le projet sera articulé autour de trois volets (le dernier pouvant être vu comme un bonus) :

- La lecture et la compréhension d'articles scientifiques et de documentation technique (référéncés ci-après) ;
- L'extension de l'outil G5K afin de pouvoir décrire des configurations multi-sites et d'en assurer le déploiement.
- La mise en place d'un système d'analyse post-mortem des résultats collectés.

**Méthodologie du projet**

Ce projet sera réalisé suivant une méthode agile. Il devra impliquer une contribution au projet *Open Source* Kolla-G5K.

**Bibliographie :**

[1] The DISCOVERY Initiative - Overcoming Major Limitations of Tradi-

---

1. <https://github.com/BeyondTheClouds/kolla-g5k>

tional Server-Centric Clouds by Operating Massively Distributed IaaS Facilities, The DISCOVERY Consortium, Research Report - RR-8779, Mars 2016 - <https://hal.inria.fr/hal-01203648/file/RR-8779.pdf>.

[2] Using the EXECO toolbox to perform automatic and reproducible cloud experiments, Matthieu Imbert et al., in Proceedings of 1st International Workshop on UsiNg and building ClOud Testbeds (UNICO, colocated with IEEE CloudCom 2013, Dec 2013 - <https://hal.inria.fr/hal-00861886/>.

[3] OpenStack, Multisite deployment, Documentation technique - <http://docs.openstack.org/arch-design/multi-site.html>