

## UV « Sensibilisation à la recherche » OpenStack et schémas d'accès SQL

14 septembre 2017

<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>— Comprendre les traces JSON retournées par l'outil EnOS, outil utilisé pour l'évaluation des performances du système OpenStack ;</li><li>— Analyser ces traces de manière automatique afin de catégoriser les requêtes SQL réalisées par le système OpenStack ;</li><li>— De proposer comme résultat de cette analyse un ensemble de métriques (quantité de données manipulées par chaque catégorie, périodicité des requêtes, ...).</li></ul> <p>A plus gros grain, ce projet s'insère dans l'initiative Discovery pilotée par l'école. Le travail réalisé pourra donner lieu à une section dans un article de recherche.</p>
<b>Site Web</b>	<a href="http://beyondtheclouds.github.io">http://beyondtheclouds.github.io</a>
<b>Intitulé</b>	Analyse des accès SQL du système OpenStack

**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.

**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 IBM, 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 repondent pas à cette règle. A titre d'exemple, les travaux conduits à l'école ont identifié la base de données MySQL au sein du composant Nova. Bien qu'il soit possible comme nous l'avons montré de développer des prototypes au dessus des systèmes NoSQL [1] (ou plus récemment NewSQL), il est nécessaire de comprendre l'implication du modèle SQL au sein du code propre au système OpenStack.

### Objectif

L'objectif du travail proposé dans ce projet vise à étendre la phase d'analyse post-mortem de l'outil EnOS<sup>1</sup> [2] développé dans le cadre de l'initiative Discovery. Cet outil permet de deployer une infrastructure OpenStack au dessus d'une large variété de plateformes experimentales grâce à la technologie Docker et Ansible. Une fois le système déployé (phase 1), il est possible d'exécuter un ensemble de tests de performance (phase 2) et d'analyser les résultats de manière post-mortem (phase 3). L'objectif du projet est d'enrichir la phase d'analyse de manière à donner une vue synthétique des informations relatives aux accès SQL réalisés.

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'analyse des requêtes SQL contenues dans la trace JSON delivrée par EnOS (et plus précisément le module OSProfiler) ;
- La mise en place d'une vue synthétique (de type dashboard) permettant de parcourir les informations collectées.

### Méthodologie du projet

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

### Bibliographie :

[1] Revising OpenStack to Operate Fog/Edge Computing infrastructure, Lebre et al., In Proceedings of the IEEE Internation conference IC2E, Vancouver, April 2017 - <https://hal.inria.fr/hal-01273427/>.

---

1. <https://github.com/BeyondTheClouds/enos>

[2] ENOS : a Holistic Framework for Conducting Scientific Evaluations of OpenStack, Cherrueau et al., in Proceedings of the IEEE/ACM International Conference CCGRID, May 2017 - <https://hal.inria.fr/hal-01415522v2/>.