

Offre de stage 2021-2022

Titre	Dynamic Load balancing of Cloud servers based on Machine Learning techniques
Niveau du stage	Master 2ème année / Ingénieur 5ème année
Date de début/ fin	De février-mars 2022 au juillet 2022 (4-6 mois)
Ville, Pays	Annecy-le-Vieux, France
Laboratoire	LISTIC - Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance - http://www.polytech.univ-savoie.fr/LISTIC
Description du sujet de stage	<p>Multidisciplinaire : développement, recherche</p> <p>Mots clés: Cloud Computing, load balancing, Reinforcement Learning, MUST</p> <p>Contexte du stage:</p> <p>Le Cloud Computing (CC) offre de multiples avantages pour l'hébergement et l'exécution des applications complexes tels que les applications liées aux domaines scientifiques, appelée aussi les applications e-Science (comme l'astronomie, la biologie, la physique, ...). Ces dernières sont caractérisées, principalement, par leurs calculs intensifs et traitent des données massives. En particulier, le cloud fournit pour l'e-Science une infrastructure informatique évolutive et scalable, comme des services de bonne qualité (QoS) et un environnement informatique personnalisable [1]. Pour répondre à la demande persistante de ressources, le CC permet d'héberger une énorme quantité de serveurs contenant chacun un nombre variable de machines virtuelles (VM) à hautes performances. Les consommateurs de cloud peuvent soumettre un nombre variable de tâches à exécuter sur une VM. Pour mieux gérer l'attribution des tâches aux différentes machines disponibles et assurer une utilisation efficace des ressources, les fournisseurs du Cloud ont souvent recours la technique d'équilibrage de charge (Load Balancing ou LB). Cette technique consiste à attribuer les différentes tâches des utilisateurs aux VMs disponibles et vérifier la charge de la VM dans l'environnement CC. L'équilibrage de charge peut être statique ou dynamique. L'équilibrage de charge dynamique (Dynamic Load Balancing ou DLB) est une solution incontournable pour remédier au comportement imprévisible des applications, permettant l'évolutivité de ses simulations. Cependant, les solutions diffèrent selon les applications, même si elles visent à résoudre le même problème, à savoir la planification des tâches dans des machines parallèles. De plus, étant donné que certains aspects du comportement de l'application peuvent varier au cours de l'exécution (par exemple, les coûts de communication et de calcul de chaque tâche), une solution unique de DLB ne peut pas être en mesure de toujours servir une application de manière efficace. Ainsi, la sélection d'un algorithme LB qui s'adapte au contexte d'utilisation des VMs n'est pas une tâche triviale.</p> <p>Dans ce projet, nous nous intéressons aux approches d'équilibrage dynamiques adaptées au contexte d'exécution dans le CC en se basant sur les algorithmes d'apprentissage automatique (Machine Learning ou ML). En effet, ML est devenu un outil commun pour modéliser le comportement d'interactions complexes entre les applications, les systèmes et les plateformes.</p> <p>Objectifs du projet :</p> <p>Les performances d'applications scientifiques irrégulières peuvent être facilement affectées par une répartition inégale de charges entre les ressources informatiques. Dans ce contexte, l'équilibrage de charge (LB) est l'une des solutions les plus importantes pour améliorer l'utilisation des ressources. De nombreux algorithmes heuristiques et méta-heuristiques ont été proposés dans la littérature pour l'équilibrage de charges dynamique des VMs dans le CC tels que [2][3][4][5][6][7]. Cependant, choisir l'algorithme d'équilibrage de charge le plus performant pour une application donnée n'est pas une tâche triviale surtout dans les situations où les applications ont un comportement dynamique ou inconnu. Ainsi, notre objectif est de proposer une solution d'équilibrage de charge dynamique qui permet de choisir un algorithme LB approprié pour un état donné de l'application et assurer une utilisation efficace des ressources du CC. Nous nous intéressons particulièrement aux algorithmes d'apprentissage automatique supervisé tels que l'apprentissage par renforcement (Reinforcement Learning).</p> <p>Méthodologie de résolution du projet :</p> <p>Afin de résoudre ce problème, nous allons procéder comme suit :</p> <ul style="list-style-type: none"> - Tâche 1 : Formuler le problème d'équilibrage de charge dynamique sous la forme d'un problème d'optimisation multi-objectif



	<ul style="list-style-type: none">- Tâche 2 : Résoudre le problème formulé en utilisant un algorithme d'apprentissage par renforcement permettant d'optimiser les objectifs définis dans la tâche 1 et s'adapter aux besoins applications dynamiques.- Tâche 3 : Effectuer les expérimentations nécessaires pour évaluer la solution proposée en utilisant une solution de simulation tels que Cloud Analyst. <p>Références :</p> <p>[1] Yang, X., Wallom, D., Waddington, S., Wang, J., Shaon, A., Matthews, B., ...&Vasilakos, A. V, 2014. Cloud computing in e-Science: research challenges and opportunities. The Journal of Supercomputing, 70(1), 408-464.</p> <p>[2] Jena, U.K., Das, P.K. and Kabat, M.R., 2020. Hybridization of meta-heuristic algorithm for load balancing in cloud computing environment. Journal of King Saud University-Computer and Information Sciences.</p> <p>[3] Negi, S., Rauthan, M.M.S., Vaisla, K.S. and Panwar, N., 2021. CMODLB: an efficient load balancing approach in cloud computing environment. The Journal of Supercomputing. 1-53.</p> <p>[4] Semmoud, A., Hakem, M., Benmammar, B. and Charr, J.C., 2020. Load balancing in cloud computing environments based on adaptive starvation threshold. Concurrency and Computation: Practice and Experience. 32 (11), e5652.</p> <p>[5] I. Banicescu, F. M. Ciorba, and S. Srivastava, Scalable Computing: Theory and Practice. John Wiley&Sons, Inc., 2013, ch. Performance Optimization of Scientific Applications using an Autonomic Computing Approach, pp. 437–466.</p> <p>[6] A. Boulmier, I. Banicescu, F. M. Ciorba, and N. Abdennadher, “An autonomic approach for the selection of robust dynamic loop scheduling techniques,” in Proceedings of the International Symposium on Parallel and Distributed Computing (ISPD). Innsbruck, Austria: IEEE, 2017, pp. 9–17.</p> <p>[7] H. Menon, N. Jain, G. Zheng, and L. Kal'e, “Automated load balancing invocation based on application characteristics,” in Proceedings of the International Conference on Cluster Computing (CLUSTER). Beijing, China: IEEE, 2012, pp. 373–381.</p> <p>[8] Kaushik, M. and Majhi, S.K., 2021. A binary Bird Swarm Optimization based load balancing algorithm for cloud computing environment. Open Computer Science. 11 (1), 146-160.</p> <p>[9] Kumar, P., and Kumar, R., 2019. Improved Active Monitoring Load-Balancing Algorithm in Cloud Computing. In Proceedings of 2nd International Conference on Communication, Computing and Networking, Springer, Singapore. 1033-1040.</p>
Compétences requises	Connaissances en Réseau et Systèmes Distribués Informatique : Linux, Python, Redis, Spark Streaming, Kafka, container (Docker et/ou Kubernetes)
Gratification	Selon législation en vigueur
Tuteurs / Contacts	Khadija Arfaoui – khadija.arfaoui@univ-smb.fr Sébastien Monnet sebastien.monnet@univ-smb.fr