

## OFFRE DE STAGE

<b>Titre</b>	Etude et extension des mécanismes d'orchestration des pipelines de traitement de flot de données scientifiques pour la plateforme MUST
<b>Niveau</b>	Master 2, Ingénieur de préférence en spécialité Informatique
<b>Date de début/fin</b>	De 4 à 6 mois. Démarrage dès que possible
<b>Ville, Pays</b>	Annecy, France
<b>Laboratoire</b>	Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC)
<b>Description du sujet</b>	<p><b>Multidisciplinaire</b> : développement, recherche</p> <p><b>Mots clés</b> : Ordonnancement, workflows scientifiques, environnements distribués</p> <p>Le sujet de ce stage se situe dans le cadre du projet de recherche NORMAL (in NetwORk Machine Learning). Ce projet a pour objectif de développer des outils méthodologiques et informatiques permettant de simplifier la tâche de déploiement des workflows scientifiques dans un environnement distribué. L'objectif de ce stage est d'étudier et d'évaluer des systèmes de gestion des flux de données scientifiques (SGFDS) existants comme la plateforme LUIGI, AIRFLOW, KUBERNETES ou HTCCondor et d'étendre ces plateformes en y implémentant certaines nouvelles fonctionnalités pertinentes aux besoins des usagers de la plateforme de calcul distribuée MUST. Une fonctionnalité clé des systèmes SGFDS consiste à fournir une interface permettant aux utilisateurs de définir facilement, par le biais d'un langage de description simple, les flux de traitement de données intégrant des dépendances de données complexes. Le SGFDS doit ensuite être capable de valider la description de ces workflows et de planifier leur exécution sur un ensemble de ressources matérielles distribuées. Enfin, le système SGFDS doit permettre de superviser l'exécution des workflows et générer leurs résultats et leurs statistiques de performances (le temps d'exécution, le taux de consommation d'énergie, etc). Ces résultats peuvent être utilisés afin de développer des mécanismes permettant d'optimiser des critères d'exécution comme la consommation d'énergie et les ressources matérielles utilisées (comme les GPUs et l'espace de RAM).</p> <p>En résumé, les résultats attendus de ce stage sont les suivants :</p> <ol style="list-style-type: none"><li>1. Effectuer une étude comparative des modèles des plateformes de déploiement des workflows scientifiques existants comme LUIGI, AIRFLOW, KUBERNETES et HTCCondor.</li><li>2. Déterminer les fonctionnalités clés d'un SGFDS nécessaires pour l'exécution et</li></ol>

	<p>l'optimisation des performances des workflows scientifiques sur la plateforme MUST.</p> <p>3. Proposer et implémenter une extension (ou sur-couche) de l'architecture du méso-centre MUST [2] permettant de prendre en compte les fonctionnalités clés définies et la tester en utilisant des exemples de workflows scientifiques existants.</p> <p><b>Références :</b></p> <p>[1] Bouselmi, K., Brahmi, Z., &amp; Gammoudi, M. M. (2017). DR-SWDF: A dynamically reconfigurable framework for scientific workflows deployment in the cloud. <i>Scalable Computing: Practice and Experience</i>, 18(2), 177-193.</p> <p>[2] MUST, <a href="https://must-datacentre.fr/">https://must-datacentre.fr/</a></p>
<b>Gratification</b>	Stage rémunéré
<b>Compétences requises</b>	Connaissances en systèmes distribués, ordonnancement Informatique : C/C++, Linux, python
<b>Tuteurs/ Contacts</b>	Khadija Arfaoui (khadija.arfaoui@univ-smb.fr), Hervé Verjus, Kavé Salamatian