

Offre de stage 2021-2022

Titre	Conception d'une architecture ETL générique pour l'analyse vidéo en temps réel
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 : Video analytics, Big Data, Redis, Kafka, MUST</p> <p>Contexte. Le sujet de ce stage se situe dans le cadre d'un projet de recherche visant à proposer une nouvelle architecture pour l'acquisition, le traitement et l'analyse de données massives (Big Data). En effet, de nos jours les sources et les types de données se multiplient au sein de l'entreprise : de fichiers plats, de données opérationnelles, de nouveaux services internet, de différents réseaux sociaux, de nouvelles applications de l'internet des objets (IOT), en particulier les capteurs. Cette révolution informationnelle a généré une grande masse de données, dite « Big Data ». Le Big Data est caractérisé par le grand « <i>Volume</i> » de données collectées par l'entreprise, la « <i>Variété</i> » de ces données, qui peuvent être structurées, semi-structurées ou non structurées et aussi par la fréquence de l'arrivée des données « <i>Vitesse</i> » qui devrait être prise en considération. Dans ce stage nous nous intéressons à l'acquisition, le traitement et l'analyse des données vidéos en temps réel, c'est-à-dire les « video analytics » [2], [3].</p> <p>Par ailleurs, autour du Big Data une bonne variété de technologies dédiées est apparue, tels que l'écosystème d'Hadoop (HDFS, Map Reduce, Yarn, etc), Flink, Kafka, Elasticsearch, Kibana, etc. En particulier nous trouvons ceux qui s'adressent à l'ingestion et au traitement de données produites en streaming (les vidéos), tels que : Kafka, Apache Samza, Spark Streaming, Flume, Storm et d'autres.</p> <p>Objectifs du stage. L'objectif du stage sera de développer les composants fondamentaux d'une architecture générique pour l'extraction, le traitement, le chargement (ETL) et l'analyse des vidéos de trafic routier collectées en temps réel. En particulier, l'étudiant étudiera comment optimiser la localisation de l'état des applications au sein de l'architecture de stockage utilisée par les pipelines ETL pour améliorer les performances d'analyse vidéo. L'étudiant s'appuiera sur un prototype déjà développé au sein du LISTIC qui utilise un schéma de routage dit Bloom Filter-based Routing « BFR » [4] sur un cache distribué pour la découverte des contenus. Détails supplémentaires sur l'architecture à implémenter seront définies avant le début du stage dans le cadre d'un sujet de thèse annexe.</p> <p>Pour l'exécution des tâches de traitement sur les flux de vidéo collectés et le stockage des données, le candidat aura accès durant la période du stage à la plateforme MUST, mésocentre de stockage et de calcul scientifique mutualisée ouverte sur la grille de recherche européenne utilisée par les chercheurs des différents laboratoires de l'USMB.</p> <p>En résumé, les résultats attendus de ce stage sont les suivants :</p> <ol style="list-style-type: none"> 1. Acquérir de compétences sur les technologies Big Data dédiées au traitement de flux vidéo. 2. Implémenter l'architecture générique proposée pour l'extraction, de traitement, le chargement, et l'analyse de vidéo. 3. Validation du système avec un exemple de flux de vidéo en temps réel, en vue de détecter par exemple les matricules de voitures circulant sur la route.



	<p>4. Réaliser des tests de performance de l'architecture implémentée sur la plateforme MUST.</p> <p>Références.</p> <p>[1] Thibeault, J. (2020). Streaming Video Fundamentals. SMPTE Motion Imaging Journal, 129(3), 10-15. DOI: 10.5594/JMI.2020.2976257</p> <p>[2] J. Jiang, G. Ananthanarayanan, P. Bodik, S. Sen, and I. Stoica. Chameleon: scalable adaptation of video analytics. In Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, pages 253–266, 2018.</p> <p>[3] Rachuri, S. P., Bronzino, F., and Jain, S. (2021, October). Decentralized modular architecture for live video analytics at the edge. In Proceedings of the 3rd ACM Workshop on Hot Topics in Video Analytics and Intelligent Edges (pp. 13-18).</p> <p>[4] Marandi, A., Braun, T., Salamatian, K. and Thomos, N., 2020, June. Network Coding-based Content Retrieval based on Bloom Filter-based Content Discovery for ICN. In ICC 2020-2020 IEEE International Conference on Communications (ICC) (pp. 1-7). IEEE.</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 Francesco Bronzino – fbronzino@univ-smb.fr Asma Dhaouadi – asma.dhaoudai@univ-smb.fr Kavé Salamatian – kave.salamatian@univ-smb.fr