

Sujet de Stage de M2 - Elastic load balancing between exascale simulation and in situ analysis

Ce sujet de stage est proposé dans le cadre du projet Exa-DoST du PEPR NumPEX, qui regroupe l'ensemble de la communauté de recherche française du CEA, du CNRS, de l'Inria, des universités et des écoles d'ingénieurs autour des thématiques de gestion de données HPC. Exa-DoST vise à concevoir les méthodes et des outils qui permettront notamment de tirer parti du premier supercalculateur exascale français qui sera installé au CEA/TGCC à l'horizon 2025. Le stage se déroulera à la Maison de la Simulation (MdlS), un laboratoire HPC commun au CEA, au CNRS, à l'UVSQ et à l'université Paris-Saclay. Le stage sera co-encadré avec une équipe de la MdlS et du CEA.

Encadrants:

Benoît Martin, MdlS, bmartin@cea.fr, tel: 01 69 08 87 71

Julien Bigot, MdlS, julien.bigot@cea.fr, tel: 01 69 08 01 75

Laurent Colombet, DAM Ile de France, laurent.colombet@cea.fr, tel: 01 69 26 43 19

Contexte

Les trois premiers supercalculateurs exascales viennent d'être déployés. L'explosion de la puissance de calcul avec l'entrée dans cette nouvelle ère promet des simulations produisant des données à des échelles jamais envisagées auparavant. L'analyse de ces masses de données nécessite l'utilisation de méthodes statistiques ou d'IA de plus en plus avancées. Historiquement, ces analyses étaient exécutées *post-hoc*, les données brutes étant stockées sur disque au cours de la simulation et les analyses exécutées par la suite. Depuis plusieurs années, l'augmentation de la performance du stockage (*rapidité d'accès et volume*) ne suit pas celle exponentielle du calcul ; un gouffre se creuse et le disque devient le nouveau goulot d'étranglement de performances. Pour contourner cette limitation, une nouvelle approche consiste à analyser les données *in situ*, pendant leur production, pour ne stocker que les résultats de ces analyses.

Cette méthode rend accessible l'exécution de simulations à des échelles extrêmes, mais elle exacerbe les problèmes d'équilibrage de charge. Historiquement, il s'agissait d'adapter la vitesse de simulation aux ressources de calcul disponibles, et donc faire varier le débit de génération de données. Avec l'analyse *in situ*, il devient, en plus, nécessaire de s'adapter à des quantités de données à analyser qui varient au cours du temps, avec un coût d'analyse qui peut lui aussi varier. Sans solution spécifique, les ressources matérielles à provisionner pour gérer

les pics de besoins de l'analyse risquent d'exploser inutilement en laissant ces ressources inutilisées hors pics. Il est donc impératif de concevoir et de mettre en œuvre des propositions d'adaptation et d'équilibrage de charge innovantes pour rendre l'approche viable et pouvoir effectivement tirer parti des supercalculateurs exascale.

Dans le domaine du *cloud computing*, la notion d'élasticité répond à ce besoin. Il s'agit de provisionner les ressources matérielles dynamiquement au cours de l'exécution, en fonction des besoins. Des travaux existants ont tenté d'adapter ce concept au calcul intensif, mais ils se sont confrontés à des problèmes conceptuels liés à la moindre flexibilité des plateformes et à la plus forte adhérence au matériel nécessaire pour en tirer les performances maximales.

L'arrivée de nouvelles formes de stockage éphémères sur les supercalculateurs, la flexibilité apportée par la désagrégation, et l'utilisation d'intergiciels issus du *cloud computing* pour l'analyse *in situ*, rebattent les cartes et ouvrent de nouvelles possibilités.

Sujet

L'objectif de ce stage est de concevoir et de mettre en œuvre une solution permettant de déclencher dynamiquement des analyses *in situ* avancées en fonction des résultats d'analyses préliminaires.

Ce projet vise à simuler et à gérer un flux de travail complexe. Vous serez amené à explorer :

1. Les analyses de conditionnement dynamiquement : par exemple, un premier niveau d'analyse détecte un événement critique, déclenchant une analyse plus poussée.
2. L'adaptation dynamique des ressources : ajouter ou redimensionner les ressources dédiées à l'analyse en fonction des besoins émergents.
3. L'interaction entre la simulation et l'analyse : intégrer un mécanisme de contrôle de flux, permettant, par exemple, de ralentir ou mettre en pause la simulation lorsque l'analyse ne suit pas le rythme des données produites.

La solution s'appuiera sur l'approche d'analyse *in situ* Deisa[2] (conçue à la Maison de la Simulation), utilisant l'outil cloud Dask[3,4], et sera intégrée dans des environnements de calcul intensif. Elle sera validée sur des supercalculateurs de classe mondiale avec des applications de simulation telles que Gysela ou Coddex[5].

Ce travail permettra de créer une charge de travail complexe nécessitant des mécanismes avancés d'équilibrage de charge et de gestion des ressources, posant les bases d'une réflexion plus large.

Déroulement prévu

Au début du stage, vous effectuerez une phase d'analyse sur l'approche Deisa, qui exploite l'outil Dask pour offrir une grande flexibilité dans l'analyse des données.

Ensuite, vous concevrez un mécanisme permettant de déclencher automatiquement des analyses avancées en fonction des résultats d'analyses préliminaires. Vous intégrerez ces

fonctionnalités dans Deisa, en vous appuyant sur Dask pour gérer les dépendances entre tâches et orchestrer leur exécution.

Enfin, vous ajouterez des fonctionnalités permettant d'ajuster dynamiquement les ressources affectées à l'analyse en fonction de la charge (par exemple, à l'aide d'un déploiement adaptatif [7]) et un mécanisme permettant de réguler la simulation en cas de saturation de ressources dédiée à l'analyse, via un contrôle de flux (backpressure).

Le stage pourra également déboucher sur une thèse de 3 ans. Celle-ci visera à approfondir les concepts abordés durant le stage et à explorer de nouvelles approches pour améliorer l'élasticité des systèmes d'analyse *in situ*. En particulier, la thèse se concentrera sur l'intégration dynamique de nœuds Dask en cours de simulation, une fonctionnalité actuellement non prise en charge par l'approche Deisa. Cette avancée permettrait de répondre encore plus efficacement aux variations des besoins en ressources, renforçant ainsi la flexibilité et la performance des supercalculateurs dans des contextes de simulation et d'analyse à très grande échelle.

Compétences attendues

- Une bonne connaissance des systèmes distribués
- De bonnes compétences en programmation (Python, C/C++)
- Très bonnes capacités de communication en anglais, à l'oral et à l'écrit
- Ouverture d'esprit, fortes capacités d'intégration et esprit d'équipe.

Bibliographie

[1] <https://www.reactivemanifesto.org>

[2] A. Gueroudji, Distributed Task-Based In Situ Data Analytics for High-Performance Simulations, PhD Thesis, <https://www.theses.fr/2023GRALM019>

[3] dask-ml 0.1 documentation - dask ml.decomposition.IncrementalPCA. URL [modules/generated/dask_ml.decomposition.IncrementalPCA.html](https://dask.org/en/latest/modules/generated/dask_ml.decomposition.IncrementalPCA.html).

[4] Dask.distributed — Dask.distributed 2022.10.2 documentation, . URL <https://distributed.dask.org/en/stable/>.

[5] P. Lafourcade, Modélisation Multiéchelle du Comportement Mécanique d'un Matériau Energétique : Le TATB, PhD Thesis, <https://www.theses.fr/fr/2018ENAM0030>

[6] E. Dirand, L. Colombet, B. Raffin, "TINS: A Task-Based Dynamic Helper Core Strategy for In Situ Analytics", in Proceedings of Asian Conference on Supercomputing Frontiers, Singapore 2018.

[7] <https://docs.dask.org/en/latest/adaptive.html>