Offre n°2025-09040

Doctorant F/H Communications réseau robustes, tolérantes aux pannes et à la volatilité, et adaptées aux applications d'intelligence artificielle

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

Niveau d'expérience souhaité : Jeune diplômé

A propos du centre ou de la direction fonctionnelle

Le centre Inria de l'université de Bordeaux est un des neuf centres d'Inria en France et compte une vingtaine d'équipes de recherche. Le centre Inria est un acteur majeur et reconnu dans le domaine des sciences numériques. Il est au cœur d'un riche écosystème de R&D et d'innovation : PME fortement innovantes, grands groupes industriels, pôles de compétitivité, acteurs de la recherche et de l'enseignement supérieur, laboratoires d'excellence, institut de recherche technologique…

Contexte et atouts du poste

Le poste s'inscrit dans le cadre d'une collaboration entre l'entreprise Hivenet et l'équipe TOPAL. La personne retenue rejoindra l'équipe TOPAL, basée à Inria Bordeaux.

À propos de TOPAL :
Chez TOPAL, nous relevons les défis émergents à l'intersection du calcul haute performance (HPC), de la simulation numérique et de l'apprentissage automatique. Alors que les plateformes de calcul deviennent de plus en plus vastes et complexes — avec des millions de cœurs et une diversité de matériels comme les GPU — nous mettons à profit notre expertise de longue date dans les systèmes d'exécution dynamiques pour exploiter au mieux ces ressources. Notre travail permet de s'adapter à des charges de travail imprévisibles et d'optimiser l'ordonnancement des tâches sans s'appuyer sur une planification rigide. Nous étendons aujourd'hui ce savoir-faire à des applications émergentes comme l'entraînement de réseaux de neurones profonds, qui imposent des contraintes particulières en matière de calcul et de mémoire.

Parallèlement, nous répondons à l'urgence de réduire la consommation énergétique et l'empreinte carbone du HPC. Cela implique de repenser les algorithmes, les déplacements de données et l'usage du matériel pour construire des systèmes plus durables. Une priorité pour nous est la gestion efficace des données, car leur déplacement et leur stockage deviennent souvent plus coûteux que le calcul lui-même. En combinant notre expertise en algèbre linéaire, en ordonnancement des ressources et en optimisation algorithmique avec ces nouveaux enjeux, nous contribuons à façonner l'avenir du calcul scientifique sur les plateformes de prochaine génération.

À propos de l'entreprise Hivenet :
Hivenet façonne l'avenir du cloud computing en tirant parti de la capacité de calcul inutilisée pour proposer une alternative décentralisée, écologique et centrée sur l'utilisateur aux services cloud traditionnels.

Mission confiée

Contexte

Hivenet est une entreprise qui propose aux particuliers comme aux entreprises de mettre à disposition leurs ressources informatiques inutilisées. Hivenet propose ainsi une offre de stockage de données, **HiveDisk**, qui utilise les espaces de stockage accordés par les contributeurs à HiveDisk. Cela permet aux utilisateurs de HiveDisk de profiter d'un stockage géo-distribué et répliqué. De la même façon, Hivenet souhaite pouvoir partager, via **HiveCompute**, les ressources de calcul (surtout des GPU) inutilisées pour réaliser principalement des tâches d'entraînement et d'inférence d'applications d'intelligence artificielle. L'utilisateur peut, via une interface web, demander l'allocation d'un certain nombre de GPU répartis sur différentes machines, puis y accéder pour exécuter ses calculs. Dans un premier temps, les GPU alloués seront sur des machines appartenant à un même réseau local (par exemple le réseau d'un site d'entreprise ou un conteneur PoliCloud), mais l'objectif à terme est de pouvoir utiliser des GPU situés sur différents réseaux d'entreprises ou de communautés à différents endroits (par exemple tous les sites d'une entreprise à l'échelle d'un pays) [1].

Ce projet présente de nombreux défis, principalement parce que l'environnement visé diffère des environnements HPC traditionnels. D'un point de vue matériel, les machines sont moins puissantes, hétérogènes, et sont interconnectées par un réseau classique, moins performant et moins fiable qu'un réseau HPC. Il faut également prendre en compte que les ressources de calcul ne sont pas disponibles en permanence (par exemple, les machines sont moins disponibles en journée car les employés les utilisent) et qu'elles sont plus susceptibles de disparaître à tout moment. De plus, utiliser des machines appartenant à différents sites géographiques crée un réseau aux performances hétérogènes : la latence pour communiquer entre deux sites est bien plus élevée qu'au sein d'un même site.

Objectifs

Cette thèse a pour objectif d'explorer les problématiques liées aux communications réseau dans un tel contexte. Pour cela, il sera nécessaire de considérer l'ensemble des bibliothèques de communication existantes (telles que

PCCL [3], MPI [2] ou libp2p), afin de définir celle ou celles qui seront ensuite adaptées à l'environnement considéré. Une fois ce modèle de communication établi, on s'intéressera, étant donné un ensemble de machines et leur topologie, aux adaptations nécessaires des schémas de communication des applications d'apprentissage pour minimiser le coût des communications : par exemple en utilisant des algorithmes de routage et une répartition des calculs et des données plus adaptés au réseau connectant les machines. Il faudra également être en mesure de détecter la disparition et l'ajout possible de machines et s'adapter en conséquence, par exemple en ignorant les contributions des machines perdues dans le cas d'un parallélisme de données, ou bien en redistribuant les données et les calculs.

Dans un second temps, on envisagera la gestion de l'occupation du réseau dans le cas où HiveDisk et HiveCompute sont présents simultanément sur les mêmes réseaux et les mêmes machines, afin de conserver des performances satisfaisantes pour les deux services et d'adapter dynamiquement les paramètres de qualité de service en fonction des conditions du réseau et des exigences des utilisateurs.

Bibliographie

[1] N. T. Karonis, B. de Supinski, I. Foster, W. Gropp and E. Lusk, "A Multilevel Approach to Topology-Aware Collective Operations in Computational Grids." arXiv preprint cs/0206038, 2002

[2] L. Shalev, H. Ayoub, N. Bshara, and E. Sabbag, "A Cloud-Optimized Transport Protocol for Elastic and Scalable HPC", *IEEE micro*, *40*(6), 67-73.

[3] M. Keiblinger, M. Sieg, J. Min Ong, S. Jaghouar and J. Hagemann, "Prime Collective Communications Library -- Technical Report", *arXiv preprint arXiv:2505.14065*, 2025

Principales activités

Le doctorant mènera des recherches originales sur le sujet décrit ci-dessus et collaborera avec les membres de l'équipe TOPAL ainsi qu'avec les partenaires de Hivenet.

Les activités incluent, sans s'y limiter : une synthèse bibliographique, des travaux de recherche, des développements logiciels, la présentation des résultats lors de conférences, la participation à des écoles de recherche, etc.

Compétences

Compétences techniques et niveau requis :

- Solide compréhension des communications réseau (sockets, TCP/IP) ;

- Maîtrise de la programmation système (C/C++) et d’un langage de haut niveau (Python) ;

- Une expérience avec les frameworks d’intelligence artificielle (PyTorch, TensorFlow, ...) et la mesure de performance logicielle est un atout ;

- Un intérêt pour les systèmes distribués et le calcul haute performance (HPC) est également un plus.

Langues :

Bonnes compétences en communication en anglais (le français est un plus).

Compétences relationnelles :

Capacité à travailler en collaboration dans un environnement mêlant recherche académique et secteur industriel.

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

Rémunération

La rémunération sera de 2200€ brut par mois, puis 2300€ brut par mois à compter de 2026.

Informations générales

- **Thème/Domaine :** Calcul distribué et à haute performance
  Calcul Scientifique (BAP E)
- **Ville :** Talence
- **Centre Inria :** Centre Inria de l'université de Bordeaux
- **Date de prise de fonction souhaitée :** 2025-09-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2025-09-30

Contacts

- **Équipe Inria :** TOPAL
- **Directeur de thèse :**
  Herault Thomas / thomas.herault@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'e?orce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

Se sentir à l'aise dans un environnement de dynamique scientifique, aimer apprendre et se confronter à des problématiques réelles, avoir une approche expérimentale, sont des qualités essentielles pour réussir cette mission.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Si vous êtes intéressés, merci de candidater via le site jobs.inria avec les documents suivants :

- cv
- lettre de motivation
- lettre de recommandation

Sécurité défense :
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.