

Processus de sélection - AAP générique 2018 étape 2

PYTHIA

Coordinateur du projet

Nom : **Trahay**
Prénom : **François**

Aucun droit de réponse n'a été enregistré par le coordinateur de projet.

QUALITE ET AMBITION SCIENTIFIQUE

- CLARTÉ DES OBJECTIFS ET DES HYPOTHÈSES DE RECHERCHE
- CARACTÈRE NOVATEUR, ORIGINALITÉ, PROGRÈS PAR RAPPORT À L'ÉTAT DE L'ART
- FAISABILITÉ NOTAMMENT AU REGARD DES MÉTHODES, GESTION DES RISQUES SCIENTIFIQUES

Le projet pose des objectifs concrets qui sont clairs. Il s'agit de développer une méthodologie dans laquelle un outil va prendre des traces d'exécution d'un programme HPC (en général déterministe), puis les analyser, en vue de prédire des solutions en terme de ressources (ex, placement des threads, prefetch des I/O) pour de futures exécutions du même programme. Cela semble donc de bon sens, bien que novateur, et l'outil est réalisable, sans prise de risques trop importante. Dans l'état de l'art, des traces sont collectées, puis analysées post mortem, puis, c'est au programmeur de tenter manuellement d'optimiser son code pour les futures exécutions. Avec Pythia, ce sera l'outil qui informera le runtime sous jacent des bonnes décisions à prendre automatiquement, phase par phase de l'exécution si cela est nécessaire, et ce, non pas en fonction d'heuristiques, mais au vu des traces collectées en amont qui seront vues comme un oracle (qui prédira le futur en se basant sur le fait d'avoir construit un automate à partir des traces). Pour parvenir à ses fins, le projet expérimentera sur des petits cas, puis passera à des applications réelles. Il partira d'un outil de collecte de traces, EZTrace développé précédemment dans l'équipe, et l'enrichira.

ORGANISATION DU PROJET ET MOYENS MIS EN ŒUVRE

- COMPÉTENCE, EXPERTISE ET IMPLICATION DU COORDINATEUR SCIENTIFIQUE ET DES PARTENAIRES
- QUALITÉ ET POTENTIEL DU JCJC ET DE SON ÉQUIPE
- ADÉQUATION AUX OBJECTIFS DES MOYENS MIS EN ŒUVRE ET DEMANDÉS

Le porteur sera très impliqué, et est compétent, de par ses travaux antérieurs sur la prise et l'analyse de traces d'exécution. Il fait partie d'une équipe reconnue, de plus, pour ses compétences en optimisation de ressources à l'exécution. Point mineur à propos du risque encouru en tâche 1, sur la difficulté de la génération de traces OTF2 par EZTrace: le risque n'est il pas au niveau du format mais plutôt au niveau de la sémantique de ce que de telles traces "racontent" (selon la manière dont les mesures sont faites?). Ne faudrait il pas un travail amont pour bien décider quels genres de traces vous devez faire remonter à l'outil et quel outil (le vôtre ou d'autres) se prête le mieux à la construction de l'oracle ? Le budget demandé, y compris en terme d'investissement dans une machine NUMA, est bien en phase avec la description des tâches et leurs objectifs. Les partenaires hors Telecom Sud Paris, ont déjà des collaborations avérées avec le porteur. La demande d'un doctorant et de deux stagiaires de master 2 est tout à fait justifiée. L'ensemble des personnes qui va donc travailler sur Pythia est donc de bonne facture.

IMPACT ET RETOMBÉES DU PROJET

- CAPACITÉ DU PROJET À RÉPONDRE AUX ENJEUX DE RECHERCHE DÉFINIS DANS LE PLAN D'ACTION
- IMPACT POTENTIEL DANS LES DOMAINES SCIENTIFIQUE, ÉCONOMIQUE, SOCIAL OU CULTUREL
- CAPACITÉ DU PROJET À FAVORISER LE DÉVELOPPEMENT D'UNE THÉMATIQUE OU D'UNE ÉQUIPE PROPRE À LA JEUNE CHERCHEUR OU AU JEUNE CHERCHEUR

Le projet rentre bien dans les défis liés aux logiciels mettant en oeuvre des applications gourmandes en temps d'exécution. Si la méthodologie expérimentée s'avère pertinente comme prévu, il est effectivement bien possible qu'elle s'applique à d'autres applications très répandues telles celles de l'analyse de données. L'impact peut donc être fort. Et il est clair que si l'oracle aide à prédire les besoins en ressources d'applications, leur temps d'exécution optimisé a un impact économique et pratique évident. Et en effet, le porteur via ce projet, qui inclut aussi un volant de collaboration nationale et un volant de collaboration internationale, pourra accélérer son rayonnement scientifique, autour de la thématique qu'il porte et qui sera renforcée grâce à ce projet ANR.

AVIS GENERAL

Proposition bien construite, objectifs clairs et réalisables, impact espéré très pertinent. Compétences du porteur et de son équipe actuelle cadrent très bien avec le projet de recherche proposé.

QUALITE ET AMBITION SCIENTIFIQUE

- CLARTÉ DES OBJECTIFS ET DES HYPOTHÈSES DE RECHERCHE
 - CARACTÈRE NOVATEUR, ORIGINALITÉ, PROGRÈS PAR RAPPORT À L'ÉTAT DE L'ART
 - FAISABILITÉ NOTAMMENT AU REGARD DES MÉTHODES, GESTION DES RISQUES SCIENTIFIQUES

+ The general idea of the project is clearly described. The intuition that gathering and extracting information about the previous executions of an HPC application could be used to optimize the performance of next executions at runtime definitely makes sense.

+ I am not aware of other work adopting the same approach in the context of HPC applications. As mentioned in the proposal, several work on data prefetching try to predict the future based on the recent past, but I do not know about work that would also take into account information gathered during previous executions to make predictions.

+ The proposal is based on extending tools already developed by the coordinator of the proposal. Taking this point into account, the objectives of the project seem realistic.

- The proposal does not provide any concrete examples of applications that could benefit from the proposed approach. The risk, as mentioned in the document, is that only few applications, if any, could take advantage of the proposed approach.

- I am concerned about the fact that in practice the proposed solution would not provide any benefit compared to existing solutions. For instance, the papers cited as motivation for thread placement [37, 40] are considering rather uncommon and outdated architectures. It is unclear that such problems still exist in recent systems. Regarding improvement of memory accesses locality, it is not clear that the solution proposed by the authors can be better than existing state-of-the-art solutions (eg [1]).

- The impact section does not provide any estimation of the performance gain that is expected thanks to the new framework that the proposal is describing.

- The proposal provides a large amount of details about the techniques that will be used to gather information about running application. The techniques that will be used to extract relevant information from the traces and to build oracles for thread placement and I/O prefetching are much more vague. Although I understand that this is part of the research to be conducted during the project, I would have liked to see some possible directions described in the proposal.

- COMPÉTENCE, EXPERTISE ET IMPLICATION DU COORDINATEUR SCIENTIFIQUE ET DES PARTENAIRES
- QUALITÉ ET POTENTIEL DU JCJC ET DE SON ÉQUIPE
- ADÉQUATION AUX OBJECTIFS DES MOYENS MIS EN ŒUVRE ET DEMANDÉS

- + The resources requested to conduct the project with respect to man-power seem adapted and sufficient to make the project successful
- + The coordinator of the project has a high expertise in the domains the proposal is dealing with (execution traces, thread placement, I/O prefetching)
- + The team of the coordinator has the required expertise to run the project.
- Regarding the test-beds used for experiments, the proposal only mention a large scale NUMA machine. I am worried that this would not be enough to run a proper evaluation in an HPC context. No solutions are mentioned about the way to evaluate I/O prefetching techniques.

IMPACT ET RETOMBÉES DU PROJET

- CAPACITÉ DU PROJET À RÉPONDRE AUX ENJEUX DE RECHERCHE DÉFINIS DANS LE PLAN D'ACTION
- IMPACT POTENTIEL DANS LES DOMAINES SCIENTIFIQUE, ÉCONOMIQUE, SOCIAL OU CULTUREL
- CAPACITÉ DU PROJET À FAVORISER LE DÉVELOPPEMENT D'UNE THÉMATIQUE OU D'UNE ÉQUIPE PROPRE À LA JEUNE CHERCHEUR OU AU JEUNE CHERCHEUR

- + As HPC systems are moving towards extreme scale, optimizing resource utilization becomes a major concern. As such, the proposal aims at addressing a timely issue.
- + The proposal perfectly matches the goals of "Défi 7 - Axe 2"
- + The coordinator of the project has demonstrated his commitment to open source. The results of the project (software, traces) will be made available to the community through appropriate means.
- + Already initiated collaborations with other scientist and companies increase the potential impact of the project.
- The proposal does not aim at starting a new research topic for the coordinator. The proposed project is in the line of his current work. On the other hand, it increases the probability of success of the project.

AVIS GENERAL

The project addresses a timely issue. The main idea of the project is clear and I am not aware of other projects adopting the same approach. Considering the background of the coordinator, his collaborations, and the proposed working plan, the project seems realistic with respect to scientific objectives and deliverables.

The proposal however does not include examples of applications that could clearly benefit from the proposed approach. Also, the expected performance improvements that could be achieved compared to the existing situation is not provided. Finally, the means that will be

used to evaluate the proposed solutions are not detailed enough.