

Outils d'analyse et de visualisation de traces d'exécution

Contexte

Une plateforme de calcul à haute performance peut être vue comme un ensemble de machines de calcul (très puissantes) reliées par un réseau (à très faible latence et très forte bande passante). On se sert de ces plateformes en soumettant des tâches (exécutions de processus parallèles distribués) à un gestionnaire de tâches, qui est chargé de décider où et quand exécuter les tâches sur l'ensemble des machines en optimisant différentes métriques et en assurant une certaine équité entre tâches et utilisateur-ice-s (via un algorithme d'ordonnancement). Différentes traces d'exécution sur ces plateformes sont disponibles en libre accès, notamment sur la [Parallel Workloads Archive](#).

L'analyse et la visualisation de traces d'exécution est primordiale pour faire des expériences scientifiques utilisant les traces réelles. Il est en particulier souvent nécessaire générer des traces synthétiques car évaluer le *bon* fonctionnement d'un algorithme d'ordonnancement nécessite d'utiliser des entrées variées, il faut donc une manière de d'analyser et de visualiser que les traces générées aient les propriétés attendues.

Objectifs du projet

Le but de ce projet est de développer un ensemble d'outils d'analyse et de visualisation de traces d'exécutions. Il est attendu que différents formats de trace (SWF, sorties Batsim...) soient supportés. Vous reproduirez d'abord des visualisations similaires à celles disponibles sur la [Parallel Workloads Archive](#), puis vous ajouterez différentes analyses, en particulier pour approfondir l'analyse du comportement des utilisateurs. Il est attendu que l'ensemble d'outils développés soit réutilisable avec de très fortes possibilités de *customisation*, il est donc attendu que les outils développés soient des bibliothèques utilisant de la composition fonctionnelle de données ou de graphiques (e.g., abstractions de [ggplot2](#) ou [plotnine](#)). Vous écrirez enfin un rapport sur ce qui a été fait sous la forme d'un mini-article décrivant le problème, l'état de l'art sur le problème, les implémentations réalisées et une analyse des résultats.

Licences, droit d'auteur, propriété intellectuelle

Les implémentations réalisées seront faites sous licence libre ; en Apache-2.0 si possible ou dans une autre licence libre si le projet d'origine en impose déjà une. Les données issues de l'expérience, les documentations, rapports et articles seront écrits sous licence libre CC-BY. Un document de cession de droit d'auteur à l'UT sera signé pour pérenniser la liberté d'accès et de modification de vos productions.

Candidature

Pour maximiser vos chances de candidature, contactez-moi par mail avec les informations suivantes.

- Très courte motivation par rapport au sujet (2 phrases)
- Bulletins de notes (master et licence)
- CV court (parcours d'études, expériences professionnelles, compétences)

Si vous candidatez en groupe à un projet, merci de ne m'envoyer qu'un seul mail pour tout le groupe.