



ENERGY-EFFICIENT STRAGGLER MITIGATION FOR BIG DATA APPLICATIONS ON THE CLOUDS

le 30 novembre 2017 09h00

ENS Rennes, Salle du conseil

**Soutenance de thèse de Tien-Dat Phan (équipe KerData - IRISA / Inria Rennes)
mention informatique**



La consommation d'énergie est une préoccupation importante pour les systèmes de traitement Big Data à grande échelle, ce qui entraîne un coût monétaire énorme. En raison de l'hétérogénéité du matériel et des conflits entre les charges de travail simultanées, les stragglers (i.e., les tâches qui sont relativement plus lentes que les autres tâches) peuvent augmenter considérablement le temps d'exécution et la consommation d'énergie du travail. Par conséquent, l'atténuation des stragglers devient une technique importante pour améliorer les performances des systèmes de traitement Big Data à grande échelle. Typiquement, il se compose de deux phases: la détection de stragglers et la manipulation de stragglers. Dans la phase de détection, les tâches lentes (par exemple, les tâches avec une vitesse ou une progression inférieure à la moyenne) sont marquées en tant que stragglers. Ensuite, les stragglers sont traités en utilisant la technique d'exécution spéculative. Avec cette technique, une copie du straggler détecté est lancée en parallèle avec le straggler dans l'espoir qu'il puisse finir plus tôt, réduisant ainsi le temps d'exécution du straggler. Bien qu'un grand nombre d'études aient été proposées pour améliorer les performances des applications Big Data en utilisant la technique d'exécution spéculative, peu d'entre elles ont étudié l'efficacité énergétique de leurs solutions.

Dans le cadre de cette thèse, nous commençons par caractériser l'impact de l'atténuation des stragglers sur la performance et la consommation d'énergie des systèmes de traitement de Big Data. Nous observons que l'efficacité énergétique des techniques actuelles d'atténuation des stragglers pourrait être considérablement améliorée. Cela motive une étude détaillée de ses deux phases: détection de straggler et traitement de straggler. En ce qui concerne la détection de straggler, nous introduisons un cadre novateur pour caractériser et évaluer de manière exhaustive les mécanismes de détection de straggler. En conséquence, nous proposons un nouveau mécanisme énergétique de détection de straggler. Ce mécanisme de détection est implémenté dans Hadoop et se révèle avoir une efficacité énergétique plus élevée par rapport aux mécanismes les plus récentes. En ce qui concerne le traitement de straggler, nous présentons une nouvelle méthode pour répartir des copies spéculatives, qui prend en compte l'impact de l'hétérogénéité des ressources sur la performance et la consommation d'énergie. Enfin, nous introduisons un nouveau mécanisme éconergétique pour gérer les stragglers. Ce mécanisme fournit plus de ressources disponibles pour lancer des copies spéculatives, en utilisant une approche de réservation dynamique de ressources. Il est démontré qu'elle améliore considérablement l'efficacité énergétique en utilisant une simulation.

Mots-clés

Big Data, MapReduce, Cloud, Efficacité énergétique, Performance, Atténuation des stragglers, Détection, Exécution speculative.

ABSTRACT

Energy consumption is an important concern for large-scale Big Data processing systems, which results in huge monetary cost. Due to the hardware heterogeneity and contentions between concurrent workloads, stragglers (i.e., tasks performing relatively slower than other tasks) can severely increase the job's execution time and energy consumption. Consequently, straggler mitigation becomes an important technique to improve the performance of large-scale Big Data processing systems. Typically, it consists of two phases: straggler detection and straggler handling. In the detection phase, slow tasks (e.g., tasks with speed or progress below the average) are marked as stragglers. Then, stragglers are handled using the speculative execution technique. With this technique, a copy of the detected straggler is launched in parallel with the straggler with the expectation that it can finish earlier, thus, reduce the straggler's execution time. Although a large number of studies have been proposed to improve the performance of Big Data applications using speculative execution technique, few of them have studied the energy efficiency of their solutions.

Addressing this lack, we conduct an experimental study to fully understand the impact of straggler mitigation techniques on the performance and the energy consumption of Big Data processing systems. We observe that current straggler mitigation techniques are not energy efficient. As a result, this promotes further studies aiming at higher energy efficiency for straggler mitigation. In terms of straggler detection, we introduce a novel framework for comprehensively characterizing and evaluating straggler detection mechanisms. Accordingly, we propose a new energy-driven straggler detection mechanism. This straggler detection mechanism is implemented into Hadoop and is demonstrated to have higher energy efficiency compared to the state-of-the-art mechanisms. In terms of straggler handling, we present a new speculative copy allocation method, which takes into consideration the impact of resource heterogeneity on performance and energy consumption. Finally, an energy efficient straggler handling mechanism is introduced. This mechanism provides more resource availability for launching speculative copies, by adopting a dynamic resource reservation approach. It is demonstrated, via a trace-driven simulation, to bring a high improvement in energy efficiency.

Keywords

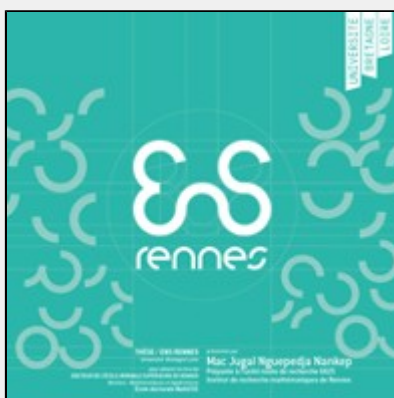
Big Data, MapReduce, Cloud, Energy efficiency, Performance, Straggler mitigation, Detection, Speculative execution.

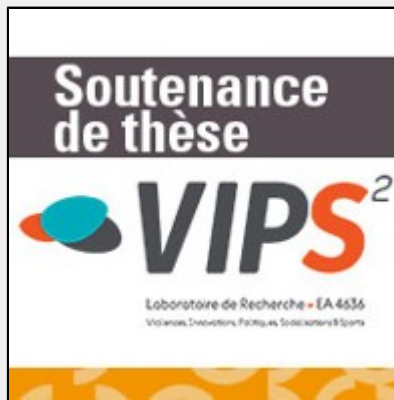
THÉMATIQUE(S)

Débouchés, Recherche - Valorisation, Vie de l'École

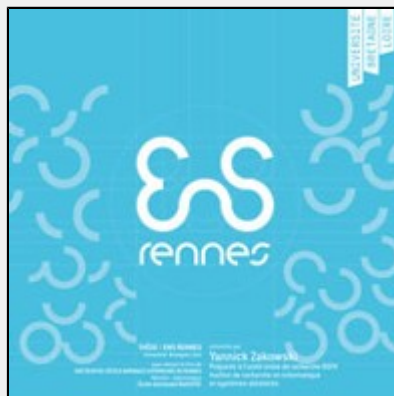
Mise à jour le 23 janvier 2018

À LIRE AUSSI





La mode du ping-pong de 1901 à 1939 : d'un jeu de salon mondain à un sport moderne



Verification of a concurrent garbage collector

JURY

Mme PEREZ Maria / rapportrice
Professeur des universités, Universidad Politécnica de Madrid

M. LEGRAND Arnaud / rapporteur
Chargé de Recherche, Laboratoire Informatique de Grenoble

M. PIERSON Jean-Marc / examinateur
Professeur des universités, Université Paul Sabatier de Toulouse

M. BOUGÉ Luc / directeur de thèse
Professeur des universités, ENS Rennes

M. IBRAHIM Shadi / co-directeur de thèse
Chargé de Recherche, INRIA Rennes

M. ANTONIU Gabriel /co-encadrant de thèse
Directeur de Recherche, INRIA Rennes