



جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Laila BOUHOUC

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Mercredi 07 Février 2024 à 15h au Grand Amphi à l'ENSIAS

Intitulé de la thèse

**Efficient Management Of Big Data Applications Deployed In The
Cloud Computing**

Président :

Pr. Mohamed Dafir ECH-CHERIF EL KETTANI, PES, ENSIAS, Université Mohammed V de Rabat

Directeur de thèse :

Pr. Mostapha ZBAKH, PES, ENSIAS, Université Mohammed V de Rabat

Co-Encadrant :

Pr. Claude TADONKI, Professeur des Universités, Mines ParisTech, École des Mines de Paris, France

Rapporteurs :

Pr. Christophe CERIN, Professeur des Universités, Université Sorbonne Paris Nord, France

Pr. Mhsine ELEULDJ, PES, EMI, Université Mohammed V de Rabat

Pr. Mahmoud NASSAR, PES, ENSIAS, Université Mohammed V de Rabat

Examineurs :

Pr. Faissal EL BOUANANI, PES, ENSIAS, Université Mohammed V de Rabat

Pr. Houda BENBRAHIM, PES, ENSIAS, Université Mohammed V de Rabat



Résumé: Dans le monde digital actuel, la gestion des applications Big Data est cruciale en raison de l'explosion des données provenant de sources telles que les réseaux sociaux, l'IoT, les capteurs et les transactions. Ces mégadonnées nécessitent des solutions novatrices en raison de leur volume et de leur vitesse. Le Cloud Computing répond à ces besoins en offrant une scalabilité illimitée et un accès à la demande à des ressources flexibles, permettant ainsi une gestion efficace de charges de travail avec des données larges.

La gestion efficace des applications Big Data implique l'orchestration des mouvements de données, de leur stockage optimal, de leur réplication, ainsi que de l'optimisation de l'allocation des ressources et de la planification des tâches. Le but est de réduire les transferts de données, améliorer le temps d'exécution des tâches, économiser des coûts, accroître la tolérance aux pannes, optimiser les ressources et améliorer les performances.

Dans cette thèse, nous explorons la gestion des applications Big Data dans les environnements de type cloud. Nous commençons par fournir un aperçu sur le Big Data et sur le Cloud Computing, puis introduisons une nouvelle classification des techniques de placement des données conçues aux tâches intensives en données. Ensuite, nous proposons deux extensions innovantes du simulateur Cloudsim, pour la prise en compte des aspects liés aux données tels que la migration de fichiers, la migration de données et le processus de fragmentation de données. De plus, une stratégie dynamique de réplication et de placement de données est suggérée. Enfin, nous proposons une nouvelle approche pour la planification en ligne des tâches visant à optimiser les temps de réponse tout en considérant la migration et la réplication des données.

Pour évaluer les performances de nos stratégies, nous les avons comparées avec celles existantes dans la littérature. Nous avons utilisé nos extensions proposées du simulateur Cloudsim pour réaliser ces évaluations. Nos résultats indiquent que nos stratégies sont efficaces et pratiques pour les défis des applications Big Data dans le Cloud.

Mots-clés: Big Data ; Cloud Computing ; Cloudsim ; Fragmentation de données ; Mouvement de données ; Placement de données ; Planification de tâches en ligne ; Réplication de données.

Abstract: In today's digital world, managing big data applications is crucial because of a tremendous increase of data coming from sources like social media, IoT, sensors, and transactions. Handling big data's challenges like volume and speed requires novel solutions. Cloud computing has arisen as one solution to address the computational needs of big data applications. It offers benefits like unlimited scalability, on-demand access to flexible computing resources, and data availability, enabling organizations to effectively handle extensive data workloads in a cost-effective manner.

Efficiently managing big data applications involves orchestrating data movement, optimal data storage, data replication, and also optimizing resource allocation and task scheduling. Efficient data management aims to reduce



data transfers, enhance the execution time of tasks, save costs, increase fault tolerance, optimize resources, and improve system performance.

In this thesis, we investigate the management of big data applications in cloud environments. We start by providing an overview on big data and cloud computing areas then we introduce a new classification of data placement techniques tailored for cloud data centers with data-intensive tasks. Next, we propose two innovative Cloudsim extensions, aiming at facilitating the simulation and evaluation of data-related aspects like file migration, data migration, and data fragmentation. Furthermore, a dynamic data replication and placement strategy based on data movement is suggested for geographically distributed data centers. Lastly, we introduce a new algorithm for online scheduling of tasks aiming at optimizing task response times while taking into account data migration and replication. To evaluate the performance of our proposed approaches, we compared them with existing ones from the literature. We used our proposed extensions of the Cloudsim simulator to conduct these evaluations. Our results indicate that our strategies are effective and convenient when it comes to tackling big data challenges in the cloud.

Keywords: Big Data; Cloud Computing; Cloudsim; Data Fragmentation; Data Movement; Data Placement; Data Replication; Online Task Scheduling.