

Nom et Prénom : AMEKRAZ ZOHRA

Date de soutenance : 02/07/2022

Directeur de Thèse : HADI MOULAY YOUSSEF

Sujet de Thèse :

La gestion de l'élasticité dans le Cloud Computing en utilisant des méthodes de Machine Learning et des méthodes statistiques

Abstract:

Cloud computing is a recent paradigm that offers interesting characteristics to its users, including elasticity. Elasticity allows Cloud users to scale their Cloud resources according to the demand changes. However, matching resource demands to applications is non-trivial due to the highly dynamic nature of workloads. One possible solution to this problem is to have prior knowledge of upcoming demand so that providers can perform proactive scheduling of resources. This thesis' first contribution consist of proposing some statistical and machine learning based accurate workload prediction models.

Elasticity is also supposed to guarantee cost efficiency, meaning that Cloud users should only pay for the resources used by their applications. Recently, Amazon have proposed a variable pricing model of cloud resources called "spot instances", that allows users to bid on unused resources at a discounted spot price. The users who get the spot instance are those with the bid price superior to the current spot price. If the spot price rises and exceeds the winning users bid prices, the instances will be terminated without notice. Thus, to avoid any interruptions, it is essential that users choose the bid price carefully. The second contribution of this thesis consists of a spot price prediction method based on Machine Learning to help users choose the right bid price.

Keywords : Cloud computing, elasticity, machine learning, time series analysis, resource management, workload prediction, spot instance.

Résumé:

Le Cloud Computing est un paradigme récent qui propose des caractéristiques intéressantes à ses utilisateurs, notamment l'élasticité. L'élasticité permet aux utilisateurs du Cloud d'adapter leurs ressources Cloud en fonction de l'évolution de la demande. Cependant, faire correspondre les demandes de ressources aux applications n'est pas trivial à cause de la nature très dynamique des charges de travail. Une solution possible à ce problème est d'avoir une connaissance préalable de la demande à venir afin que les fournisseurs puissent effectuer une planification proactive des ressources. La première contribution de cette thèse consiste à proposer des modèles de prédiction de charge de travail basés sur les statistiques et le Machine Learning.

L'élasticité est également censée garantir la rentabilité, ce qui signifie que les utilisateurs du Cloud ne doivent payer que pour les ressources utilisées par leurs applications. Récemment, Amazon a proposé un modèle variable de tarification des ressources Cloud appelé « les instances Spot ». Ces instances permettent aux utilisateurs d'enchérir sur des ressources inutilisées à un prix Spot réduit. Les utilisateurs qui obtiennent l'instance sont ceux dont le prix d'enchère est supérieur au prix Spot actuel. Si le prix Spot augmente et dépasse les prix d'enchères des utilisateurs gagnants, les instances seront résiliées sans préavis. Ainsi, pour éviter toute interruption, il est primordial que les utilisateurs choisissent le prix d'enchère soigneusement. La deuxième contribution de cette thèse consiste en une méthode de prédiction du prix Spot basée sur le Machine Learning afin d'aider les utilisateurs à choisir le bon prix d'enchère.

Mots-clés : Cloud computing, élasticité, Machine Learning, analyse de séries temporelles, gestion des ressources, prédiction de la charge de travail, instance Spot.