Automatisation de la structuration des logs pour le Cloud Computing

Arthur Vervaet*, Raja Chiky**, Mar Callau-Zori*

*3DS Outscale,1 rue Royale 92210 Saint-Cloud
**LISITE-ISEP, 10 rue de Vanves, 92130 Issy-les-Moulineaux
arthur.vervaet@outscale.com, raja.chiky@isep.fr

Résumé. Les registres de *logs* sont une composante fondamentale des systèmes informatiques modernes. Ils permettent aux équipes d'analyse et de surveillance de comprendre les comportements anormaux ou malveillants ayant pu survenir. Cependant, l'augmentation permanente du volume de *logs* générés par ces systèmes a rendu impossible l'inspection manuelle et pose un véritable défi d'automatisation du processus. Afin de traiter automatiquement ces données, plusieurs solutions de structuration des *logs* ont vu le jour. Dans cet article, nous analysons les capacités de deux d'entre elles à répondre aux enjeux du *Cloud Computing* en termes d'efficience et d'efficacité. Nos travaux se concentrent sur l'impact des paramètres et du prétraitement sur les performances de ces méthodes, deux étapes importantes, car nécessitant une intervention humaine incompatible avec l'automatisation du processus de structuration des *logs*.

1 Introduction

Les plateformes de *Cloud Computing* mettent à disposition de leurs clients différentes ressources informatiques à la demande. Cette externalisation rend les fournisseurs garants de la haute disponibilité et de la qualité de leurs services. La gestion d'un parc de ressources mutualisées en croissance constante demande de minimiser l'intervention humaine afin de suivre le changement d'échelle des infrastructures et d'éviter les erreurs. Pour atteindre cet objectif, on doit pouvoir se servir de toutes les informations à disposition afin de développer des outils autonomes servant à contrôler et assurer le respect de la qualité de service.

La journalisation des événements ou *logs*, consiste à enregistrer de manière détaillée des informations relatives à un programme pendant son exécution. Ces *logs* constituent une source d'information précieuse, utilisée pour retracer les différentes étapes d'un processus à la recherche de l'origine d'erreurs ou de pannes, mais également pour identifier des anomalies de performance ou analyser les statistiques d'utilisation (Zhu et al., 2019).

Malgré toutes les possibilités offertes par les *logs* et l'information qu'ils contiennent, les traiter efficacement est une tâche complexe. Les développeurs ayant très peu de contraintes dans l'écriture des macros destinées à produire les messages *logs*, ceux-ci suivent un format semi-structuré. Par exemple, le protocole RFC 5424 pour rsyslog (Network Working Group, 2009) fixe un format en trois parties : un en-tête (HEADER), une partie optionnelle pour des