Log clustering tool

# Resumen

Contenido

[Resumen i](#_Toc519523891)

[1. Introducción 1](#_Toc519523892)

[2. Bibliografía 2](#_Toc519523893)

# 1. Introducción

Hoy en día se genera una gran cantidad de archivos de logs en campos como la industria y la ciencia, ya que estos son una fuente de información muy importante y útil en muchas situaciones, Sin embargo, a medida que aumenta la complejidad de los sistemas, el análisis de los archivos de logs es cada vez más exigente y dificultoso, ya que hay que realizar un gran esfuerzo para recopilar, almacenar e indexar una gran cantidad de logs que se agrava más cuando estos logs no son heterogéneos **(1)**.

Toda esta producción de logs hace, que, debido al tamaño de estos conjuntos de datos, que las soluciones de bases de datos convencionales no sean las adecuadas para el análisis de la información y en su lugar se consideren más apropiados bases de datos virtuales combinadas con sistemas de procesamiento distribuidos y paralelo.

Con todo esto, se puede suponer que el análisis de logs es un caso de uso de big data y por lo tanto es un gran desafío para el procesamiento de estos, su almacenamiento, su variedad y su gestión con los recursos disponibles.

Ccc

Los problemas de investigación de variedad de datos están relacionados con el manejo del volumen creciente de datos, la extracción de contenido significativo y la agregación y correlación de datos de transmisión de múltiples fuentes. Los problemas relacionados con el almacenamiento de datos se refieren a métodos eficientes para reconocer y almacenar información importante, técnicas para almacenar grandes volúmenes de información de forma que puedan recuperarse fácilmente y migrarse entre los centros de datos. El procesamiento de datos y la administración de recursos se relacionan con modelos de programación optimizados para transmisión de datos multidimensionales, motores capaces de combinar múltiples modelos de programación en una sola solución y uso de recursos y técnicas de optimización de consumo de energía.

ss (1)

# 2. Bibliografía

1. *Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark.* **Ilias Mavridis y Helen Karatza.** 2017, The Journal of Systems and Software, págs. 133-151.

2. **[En línea] https://spark.apache.org/.**