

UNIVERSIDAD NACIONAL DEL ALTIPLANO PUNO

FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

Implementación del procesamiento masivo de datos para grupos de empresas eléctricas

Esther Milagros Bautista Peralta

October 16, 2022



Resumen

El crecimiento en el volumen de datos generados por diferentes sistemas y actividades cotidianas en la sociedad ha forjado la necesidad de modificar, optimizar y generar métodos y modelos de almacenamiento y tratamiento de datos.[Hernández-Leal et al., 2017] Por ello el procesamiento ha evolucionado hacia el paralelismo, asociado con multiprocesamiento en hardware y concurrencia en software que tiene objetivos múltiples[De Giusti et al., 2000]. Los datos comerciales de la industria de la energía eléctrica también mostraron un rápido crecimiento y el tamaño de los datos está en continua expansión. Por lo tanto, los sistemas de gestión de grupos de empresas eléctricas se encontraron con algunas deficiencias en el rendimiento del procesamiento de datos masivo.[Wu et al., 2016]

La metodología a emplear para la exploración incluye la aplicación de un ejemplo usando la base de datos de Electro Puno S.A.A. Puno mediante Hadoop que es un marco de software de código abierto desarrollado para la informática y el almacenamiento distribuidos fiables, escalables y eficientes, y es ampliamente utilizado en el procesamiento masivo de datos[Ponce et al., 2018]. Basado en el estudio de la propuesta de procesamiento masivo de datos de negocios de la industria eléctrica. Se presentará una arquitectura novedosa, que se basa en el ecosistema de Hadoop, para los sistemas de gestión de grupos de empresas eléctricas para optimizar el procesamiento. En primer lugar, se analiza la lógica de negocio y los productos de procesamiento del sistema de gestión. Luego se introduce en detalle una nueva arquitectura con un modelo de procesamiento distribuido basado en Hadoop.[Wu et al., 2016]

Como resultado los escenarios de negocio reales y los datos probarán que la arquitectura puede resolver eficazmente los cuellos de botella actuales de rendimiento del procesamiento masivo de datos empresariales, y mejorará en gran medida el rendimiento y la eficiencia del sistema de gestión de grupos de empresas eléctricas.[Wu et al., 2016]

El tratamiento de grandes volúmenes de información no está exento de numerosos problemas y retos, algunos de los cuales se tratarán de mejorar. Dentro de las posibilidades actuales debemos tener en cuenta la evolución que han tenido los sistemas durante los últimos años y las oportunidades de mejora que existan en cada uno de ellos.[Herrera Hernández, 2020]

Palabras claves: =1Big Data; procesamiento paralelo; análisis de datos; Hadoop

Bibliography

- [De Giusti et al., 2000] De Giusti, A. E., Naiouf, M., De Giusti, L. C., and Tarrío, D. F. (2000). Análisis y evaluación de algoritmos paralelos sobre diferentes modelos de arquitecturas multi-procesador. In *II Workshop de Investigadores en Ciencias de la Computación*.
- [Hernández-Leal et al., 2017] Hernández-Leal, E. J., Duque-Méndez, N. D., and Moreno-Cadavid, J. (2017). Big data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, 20(39):17–24.
- [Herrera Hernández, 2020] Herrera Hernández, J. (2020). *Optimización de arquitecturas distribuidas para el procesamiento de datos masivos*. PhD thesis, Universitat Politècnica de València.
- [Ponce et al., 2018] Ponce, A. G., Rosas, R. M. V., Marcial-Romero, J. R., and Eleuterio, R. A. (2018). Entornos de trabajo para procesamiento de datos masivos y aprendizaje automático. *Res. Comput. Sci.*, 147(5):225–237.
- [Wu et al., 2016] Wu, L., Shen, Z., and Feng, X. (2016). Implementation of massive data processing architecture for electric enterprise groups. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pages 1–6.