

Resumen

Este estudio resalta que la empresa "Casa de Incentivos Casintour" enfrenta el reto de procesar y analizar grandes volúmenes de datos mediante un proceso de extracción, transformación y carga (ETL) para llevarlos a un repositorio centralizado. A medida que la cantidad de datos aumenta, este proceso se vuelve más lento debido a su naturaleza secuencial, sin aprovechar el procesamiento paralelo disponible en entornos de Big Data.

La optimización del proceso ETL en Big Data es esencial para "Casa de Incentivos Casintour"

debido a la ineficiencia de recursos al no aprovechar completamente el hardware, lo que puede requerir inversiones adicionales sin mejoras notables en el rendimiento y la pérdida de

oportunidades comerciales al demorar la disponibilidad de datos procesados, lo que socava la toma de decisiones basadas en datos en tiempo real, crucial en un entorno competitivo

empresarial.

Además, este estudio destaca la necesidad de abordar la falta de paralelismo que afecta directamente el rendimiento, eficiencia, escalabilidad y oportunidades comerciales.

Resolver este problema es fundamental para mejorar el procesamiento de datos, acelerar la disponibilidad de la información y aprovechar los recursos en entornos de Big Data. Para abordar los desafíos mencionados y garantizar el éxito en la optimización del proceso ETL en "Casa de Incentivos Casintour", se adoptará la metodología ágil SCRUM para la ejecución de este proyecto de tesis.

Palabras clave: hilos de programación, procesos, rendimiento, iso 25010, paralelismo.

Abstract

This study highlights that the company "Casa de Incentivos Casintour" faces the challenge of processing and analyzing large volumes of data through an extract, transform and load (ETL) process to bring it to a centralized repository. As the amount of data increases, this process becomes slower due to its sequential nature, without taking advantage of the parallel processing available in Big Data environments.

Optimizing the ETL process in Big Data is essential for the company "Casa de Incentivos Casintour" due to the inefficient use of resources in Big Data, due to the inefficiency of resources by not fully leveraging hardware, which may require additional investments without noticeable performance improvements and the loss of business opportunities by delaying the availability of processed data, which undermines real-time data-driven decision making, crucial in a competitive environment.

In addition, this study highlights the need to address the lack of parallelism that directly affects performance, efficiency, scalability and business opportunities. Addressing this issue is essential to improve data processing, accelerate information availability, and maximize resource utilization in the Big Data environment.

To address the above challenges and ensure success in optimizing the ETL process at "Casa de Incentivos Casintour", the agile SCRUM methodology will be adopted for the execution of this thesis project.

Key words: programming threads, processes, performance, iso 25010, parallelism.