

XS3210 Sistemas de Información - Investigación Grupal 1

Luis Fernando Solano Coto¹, Amanda Cedeño Guzmán², Keyla Zúñiga Martínez³, Luis Alonso Madrigal Ramírez⁴, James Araya Cerdas⁵ y Luis Felipe Montero González⁶

¹Escuela de Estadística, LUIS.SOLANOCOTO@ucr.ac.cr

²Escuela de Estadística, AMANDA.CEDENO@ucr.ac.cr

³Escuela de Estadística, KEYLA.ZUNIGAMARTINEZ@ucr.ac.cr

⁴Escuela de Estadística, LUISALONSO.MADRIGAL@ucr.ac.cr

⁵Escuela de Estadística, JAMES.ARAYACERDAS@ucr.ac.cr

⁶Escuela de Estadística, LUISFELIPE.MONTERO@ucr.ac.cr

24 de septiembre de 2023

1. Descripción del problema

Manejar grandes volúmenes de datos no estructurados dentro de las aplicaciones de redes sociales que requieren ser recolectados y analizados en **tiempo real**.

2. Objetivo del proyecto

Crear una plataforma que sea capaz de recolectar y analizar datos de redes sociales como *tweets*, publicaciones e interacciones entre usuarios en **tiempo real**. Dicha plataforma permitirá: identificar temas en tendencia y medir las interacciones de los usuarios dentro de la plataforma.

3. Tipo de bases de datos a utilizar y justificación

Las plataformas de redes sociales generan grandes cantidades de datos no estructurados en la forma de posteos, comentarios, *likes*, y perfiles de usuarios. La naturaleza de estos datos es altamente variable e impredecible, lo cual lo hace difícil de manejar con bases de datos relacionales. NoSQL permite manejar justamente datos no estructurados y provee la flexibilidad necesaria para sostener una red social *Real-World NoSQL Database Use Cases: Examples and Use Cases for Developers* (2023).

Dentro de NoSQL se encuentra un sistema llamado *Key-Value*, el cual, a diferencia de las bases de datos tradicionales relacionales, ofrece flexibilidad (es capaz de almacenar texto, números, objetos y datos binarios, entre otros) y escalabilidad, haciéndola ideal para proyectos con grandes volúmenes de datos que están en constante evolución Patil (2023). En los modelos de datos de *key-value* se tendría una *key* que referencie diversos valores almacenados en un archivo tipo *json*.

4. Softwares a utilizar para el tipo de base de datos seleccionada

4.1. Redis

Redis es un producto open source en el cual la estructura de los datos almacenada en la memoria es utilizada como una base de datos, caché, intermediario de mensajes y *streaming engine*. La razón por la cual se decide hacer uso de este producto es porque permite crear potentes, interactivas y dinámicas *dashboards*, haciendo uso de una pequeña cantidad de comandos Ramuka (2022).

4.2. MongoDB

MongoDB es un producto que almacena datos en colecciones que son un grupo de documentos BSON (JSON binario) donde cada uno de ellos es esencialmente construido desde una estructura *field-value*. Este producto logra almacenar eficientemente documentos con múltiples esquemas e indexar campos adicionales para búsquedas aleatorias, esto lo convierte en un almacenamiento de datos *key-value*. Sus valores permiten no solo acceder a datos por medio de una llave en un sentido global, sino también manipular datos asociados con las llaves dentro de documentos y al indexar, permite una recuperación rápida de estas llaves secundarias. Se decide hacer uso de este producto ya que MongoDB ofrece una solución de almacenando de datos *field-value* mientras permite la creación de objetos complejos para ser formados a conveniencia y crear múltiples queries *Key-Value Databases* (2023).

Referencias

Key-Value Databases. (2023). (<https://www.mongodb.com/databases/key-value-database/>)

Patil, P. (2023). *Key-Value Databases: Building Efficient and Scalable Data Solutions.* (<https://www.redswitches.com/blog/key-value-databases/>)

Ramuka, M. (2022). *Build a real time analytics with Redis.* (<https://medium.com/google-cloud/build-a-real-time-analytics-with-redis-b94e502864f9/>)

Real-World NoSQL Database Use Cases: Examples and Use Cases for Developers. (2023). (<https://www.datastax.com/guides/nosql-use-cases#/>)