Tarea 4 – Almacenamiento y Consultas Datos Big Data

Lindsay Andrea Quintero Hernández

202016911A_2034: Big Data

Grupo 24

Tutor. Alvaro Javier Gomez

Programa. Ingeniería de Sistemas
Universidad Nacional Abierta y a Distancia -UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI
Zona Centro Oriente - ZCORI (CEAD Bucaramanga)
Noviembre del 2024

Objetivos

Objetivo general

"Aplicar técnicas de análisis y visualización a grandes conjuntos de datos para obtener información útil" (UNAD, 2025).

Objetivo especifico

- Realizar implementación, manipulación y análisis de datos de Base de Datos MongoDB.
- Realizar un ensayo comparando los diferentes tipos de bases de datos NoSQL.
- Discutir las ventajas e inconvenientes de cada tipo, así como sus casos de uso más apropiados.
- Diseñar base de datos NoSQL con la selección del caso de uso presentado por cada estudiante.
- Implemenatar MongoDB en creación y consulta de base de datos no relacional.
- Publicar avances y dudas durante el desarrollo en el foro de discusión para fomentar la participación y el avance en la actividad grupal.

Fase 1. Presentación de Ensayos

Lindsay Quintero

A lo largo del tiempo, se ha recopilado y analizado datos para comprender mejor nuestro entorno, para almacenar, realizar cálculos de rendimientos y predecir, las bases de datos, o conjuntos de información para gestión de los datos, el tipo de gestores de datos más comunes son las bases de datos relacionales, se dice relacional porque los datos se organizan en tablas y sus relaciones (los datos mueven el mundo), "*en 1970, el señor Edgar Frank "Ted" definió el modelo relacional, forma de guardar la información y para evitar redundancia en la información*" (EDteam, 2022), sin embargo este concepto de organización generó problemas cuando se empezó a requerir más velocidad, escalabilidad y flexibilidad, lo que originó las bases de datos no estructuradas "*NoSQL o No solo SQL*" por Carlo Strozzi 1998 (Keith D. , 2018), "*garantizan la disponibilidad pero no la coherencia inmediata, el estado de la base de datos puede cambiar con el tiempo y finalmente se vuelve coherente*" (AWS, 2025).

Las bases de datos NoSQL almacenan los datos en distintas formas no estructuradas, "*incluyen otros tipos de almacenamiento para gran cantidad de volumen de datos no estructurados y semiestructurados con mayor flexibilidad y escalabilidad horizontal*" (Carrascal Porras & Varona Toborda, 2024), en un sistema de base distribuida 'múltiples computadoras' (Keith D. , 2018), se usa para modelos más flexibles, en tratamiento de datos, para redes sociales, categorías de productos, logs de aplicaciones, películas, videos, fotografías, entre otros; existen varios sistemas de bases de datos debido a variación en las que administran y almacenan los datos en esquemas, su uso de datos como documentos, pares clave-valor, columnas amplias y grafos, esta base de datos cumplirá con el teorema CAP (consistencia "datos consistentes", disponibilidad "siempre disponible", tolerancia a peticiones "aunque comunicación deje de ser fiable el sistema seguirá funcionando").

Bases de datos por columnas almacenan los datos por columnas en lugar de filas, fácil para tratamiento de datos masivos, optimizadas para análisis (mejorar tiempo de respuestas en comparación a base de datos relacionales) y consultas de agregación, "los datos son almacenados como secciones de las columnas de datos en lugar de filas de datos" (Hansel Gracia del & Osmel Yanes , 2012), entre sus inconvenientes rendimiento lento en transacciones, entre sus usos gran volumen de datos, registro de usuarios, ingresos a plataformas, se recomienda para escalabilidad

masiva, permitiendo gran cantidad de escrituras (series temporales o Big Data analítico); ejemplo de gestor de base de datos NoSQL son Cassandra o Hbase.

Bases de datos clave-valor, almacena los datos como colección "clave-valor, siendo la clave la identificación única, datos almacenados como objetos binarios" (Carrascal Porras & Varona Toborda, 2024), es muy rápida para lectura y escritura, escalable y facilidad de uso, entre sus desventajas de uso, su simplicidad arquitectónica y la capacidad limitada (operaciones limitadas a get), entre sus usos se encuentra carritos de compra, sesiones de usuario, almacén temporal, ideal para acceso rápido o un rendimiento máximo; ejemplo de gestor de base de datos NoSQL son Redis o DynamoDB.

Bases de datos de documentos "formato de modelo de documento objetos JSON o BSON, flexibles, semiestructurados y naturaleza jerárquica, son las más versátiles, ya que permiten consultas más avanzadas además de las consultas clave-valor" (Carrascal Porras & Varona Toborda, 2024), gran rendimiento en consultas permite almacén semiestructurado, escritura flexible, entre sus desventajas, el manejo ineficiente de datos relacionados complejos, la aplicación debe manejar la integridad del esquema o de los datos "los tipos de datos ingresados son adecuados", entre sus casos de uso cuando gestión de contenido, aplicaciones móviles (para catálogos) y redes sociales, son ideales para estructuras jerárquicas; ejemplo de gestor de base de datos NoSQL son MongoDB o Firebase.

Bases de datos de grafos, "*diseñadas para almacenar relaciones y navegar por ellas, se usan nodos, sus tipos y dirección, experiencia de búsqueda más eficiente, usando estructura de datos para consultas semánticas, representan los datos de forma de nodos, bordes y propiedades*" (Carrascal Porras & Varona Toborda, 2024), más conocidos en el uso de redes sociales, análisis ruta conexión, relaciones, entre sus desventajas se centran en la complejidad en la curva de aprendizaje, ineficiencia para tarea o analíticas masivas, ideal cuando relaciones son complejas, existen relaciones; ejemplo de gestor de base de datos NoSQL son Neo4J

Otros tipos de datos pueden ser bases las bases de datos orientadas a objetos "datos representados en forma de objetos" (Carrascal Porras & Varona Toborda, 2024) y las bases de datos NoSQL tabulares "soluciones hibridas aceleran las consultas de bases de datos en columnas" (Carrascal Porras & Varona Toborda, 2024)

En conclusión, las bases de datos no relacional o NoSQL, experimentaron una gran implementación debido a su flexibilidad y optimización de sus diseños según la problemática, la orientación en organización de los datos y tipos de datos a gestionar; BD NoSQL se convirtieron en elemento fundamental en proyectos de Big Data, lo que permiten la gestión de datos no estructurados y semiestructurados, esto quiere decir que No solo se usa SQL, también otra estructura para organizar y manipular los datos, entre los gestores de base de datos NoSQL, Cassandra, MongoDB, DynamoDB, entre otros. Para la implementación de los distintos tipos, ya sea por clave valor, por documentos, grafos o columnas, sus ventajas, desventajas y casos prácticos (uso), cada una con sus diversos casos de uso e implementaciones según el tipo de dato y su funcionalidad esperada, la estructura de los datos, la naturaleza de las consultas y la escalabilidad, por ejemplo en redes sociales se recomendaría uso de grafos para relaciones sociales, para aplicativos móviles estructuras de documentos, para gran volumen de datos: logs; y eventos se recomendaría uso de columnas; para compras web y datos temporales: clave-valor; finalmente entre las desventajas de implementar NoSQL es el pago por el uso de licencia en los distintos gestores de base de datos NoSQL.

Fase 2. Implementación de MongoDB

Lindsay Quintero.

Selección de base de datos: utilizaré catalogo de productos o suministros tecnológicos (5000 documentos), se identifica con item, descripción, ubicación de tienda, método de compra; lo que permitirá realizar análisis de ventas y rankings por medio del SGBD; el conjunto de datos se ha descargado de Kaggle, he seleccionado esto debido a que los documentos pueden ser gestionados por el SGBD MongoDB, ya que la BD contienen una estructura similar al JSON lo que permitirá realizar la practica por medio de MongoDB.

Definición:

Colecciones: son colecciones de documentos similares, en este caso se creará colección con los documentos.

Documentos: estructura de datos tipo JSON, cada documento contiene información de cada suministro.

Campos: se identifica los atributos o características de cada documento como por ejemplo su identificación, fecha venta, item, ubicación de tienda, uso de cupón y método de compra

Enlace de mi selección: https://www.kaggle.com/datasets/siwachtoprasert/mongodb-supplies/data

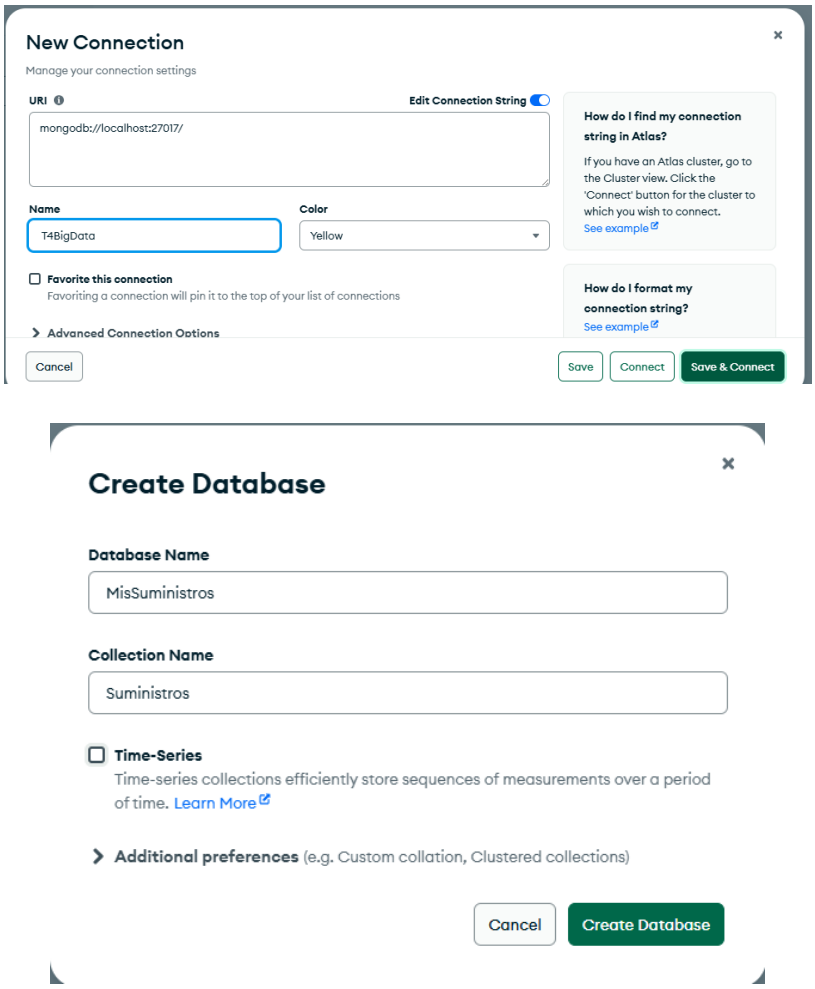
Descripción del caso de uso: el catálogo de productos incluye 5000 suministros el objetivo de este análisis es identificar la cantidad de artículos que aplican a cupón y contar el número de artículos totales de la tienda para realizar una estimación de productos.

Código UI MongoDB Compas

Creación base de datos

Figura 1

Creación de base de datos con UI MongoDB

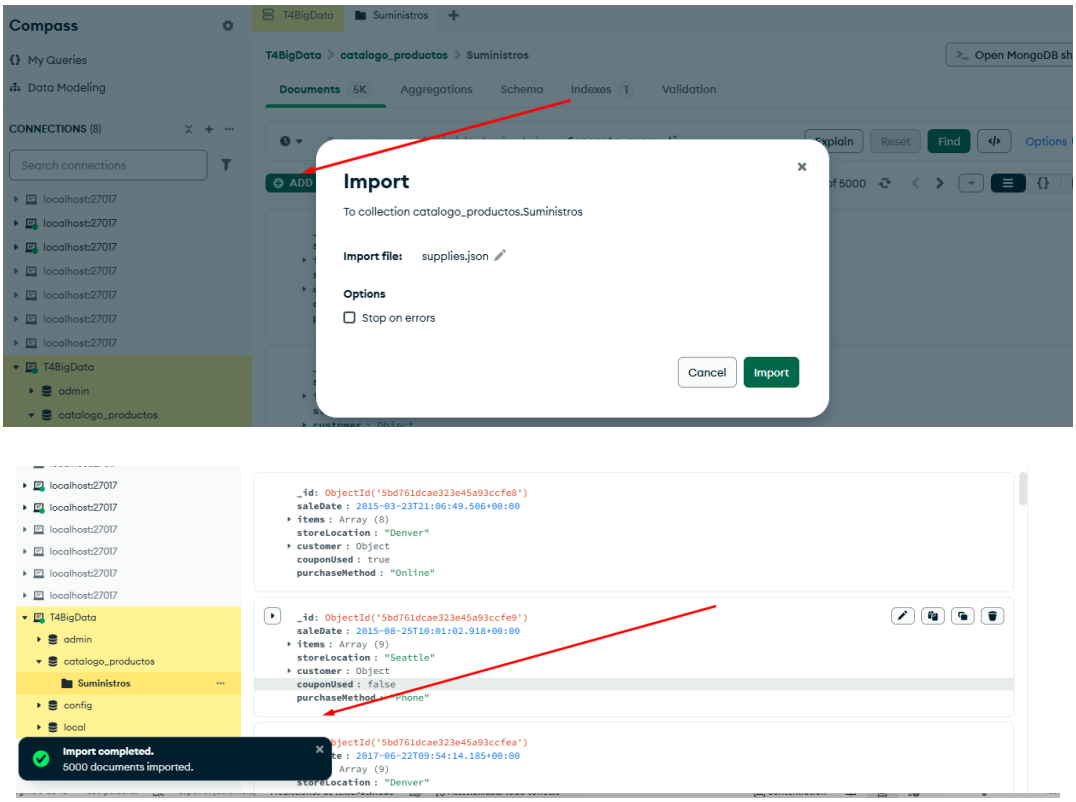


Nota. La imagen “captura de pantalla” creación base de datos, MongoDB, (2025). Interfaz UI [Software].

Insertar caso de uso(datos)

Figura 2

Insertar datos, caso de uso a base de datos con UI MongoDB



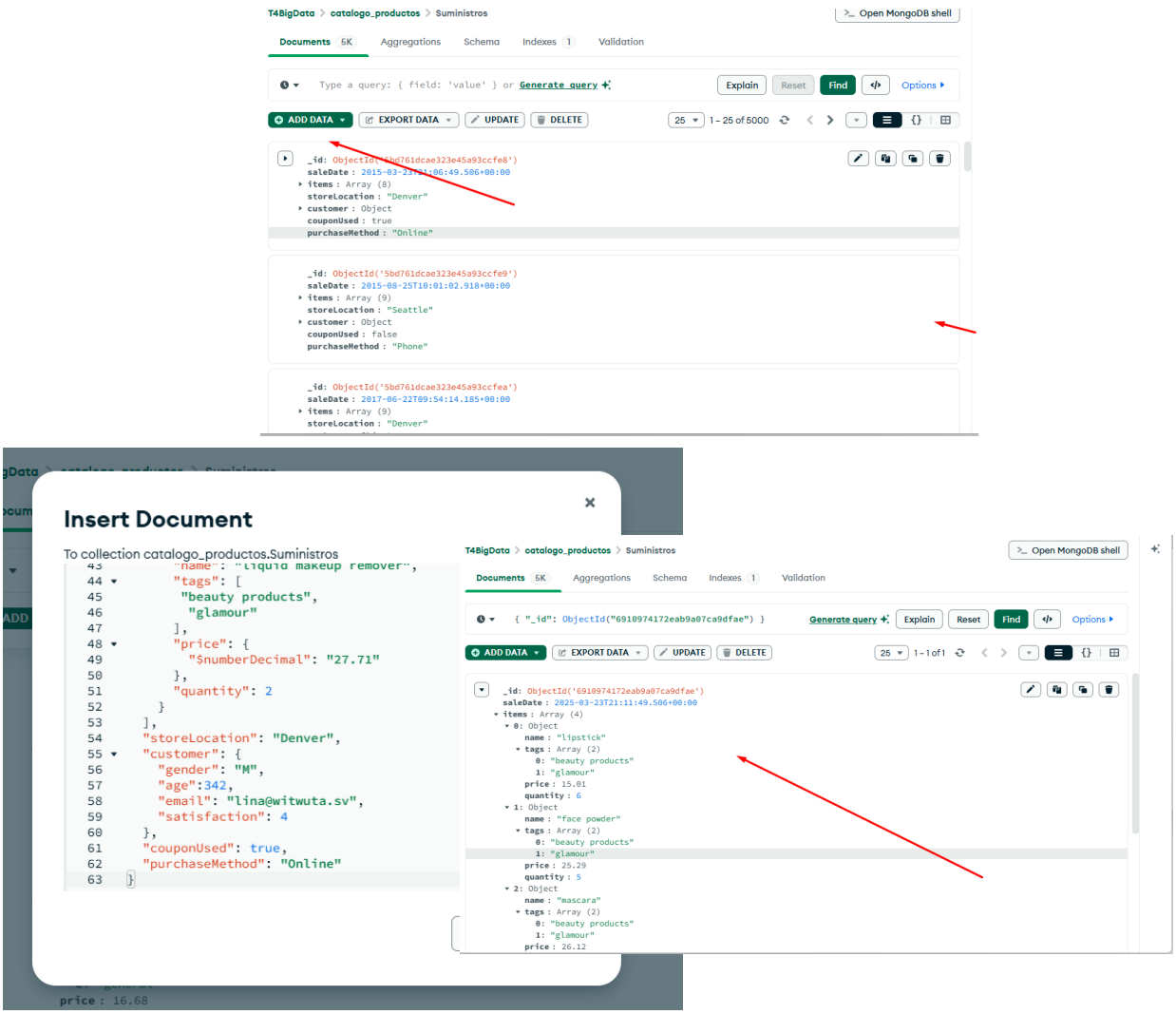
Nota. La imagen “captura de pantalla” insertar caso de uso, MongoDB, (2025). Interfaz UI [Software].

Explicación de códigos

Inserción de datos

Figura 3

Insertar datos, análisis de información y explicar codigo con UI MongoDB

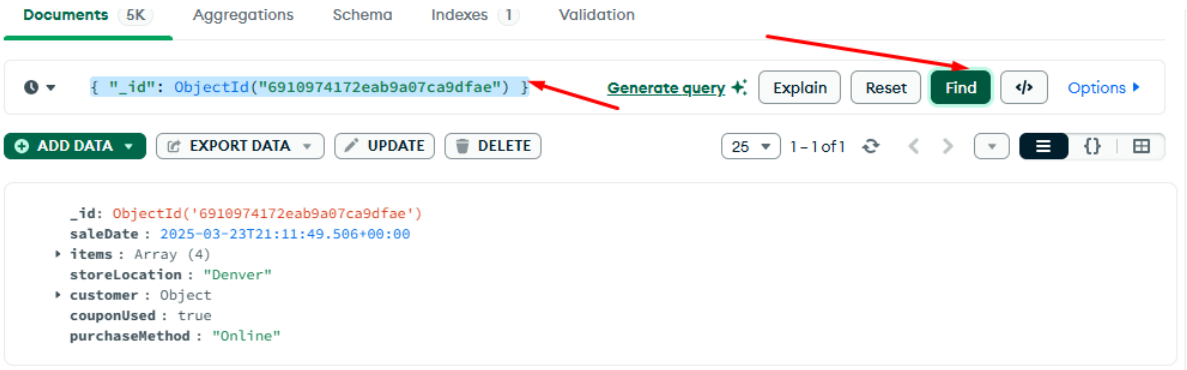


Nota. La imagen “captura de pantalla” insertar datos, MongoDB, (2025). Interfaz UI [Software].

Selección de datos

Figura 4

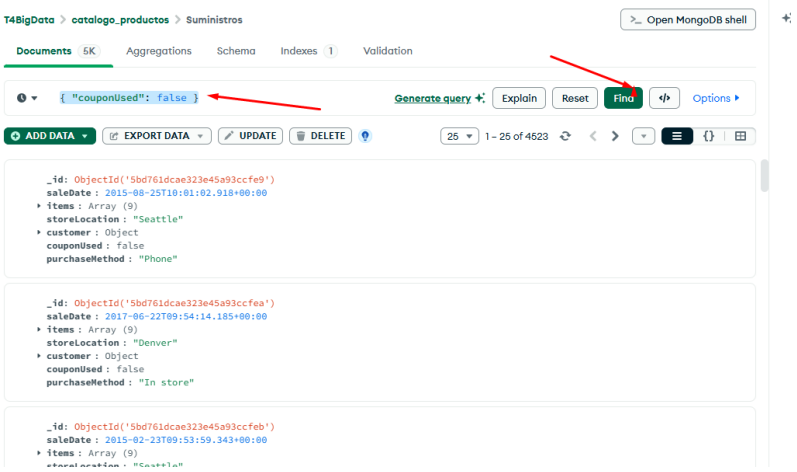
Selección de datos, explicación de datos con UI MongoDB



Nota. La imagen "captura de pantalla" selección de datos, MongoDB, (2025). Interfaz UI [Software].

Figura 5

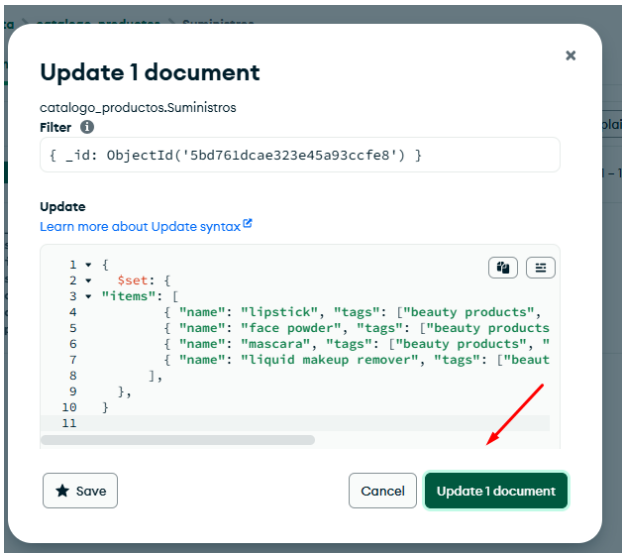
Selección de datos, explicación de datos, selección de suministro cumple con cupón, UI MongoDB



Nota. La imagen "captura de pantalla" selección de datos, MongoDB, (2025). Interfaz UI [Software].

Figura 5

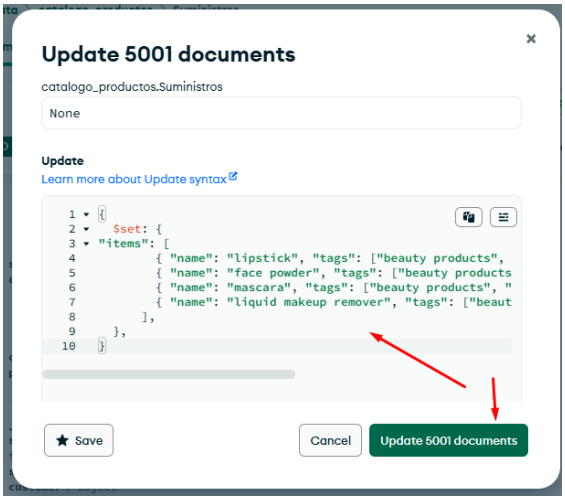
Actualizar un solo documento, UI MongoDB



Nota. La imagen "captura de pantalla" actualizar datos, MongoDB, (2025). Interfaz UI [Software].

Figura 6

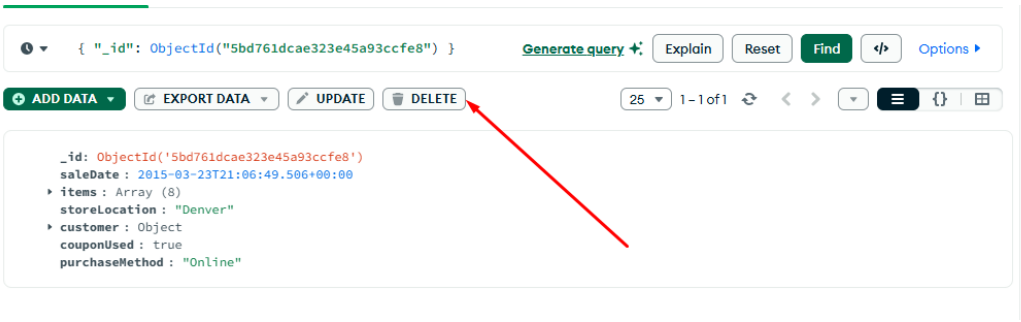
Actualizar varios documentos a la vez, UI MongoDB



Nota. La imagen "captura de pantalla" actualizar varios documentos a la vez, MongoDB, (2025). Interfaz UI [Software].

Figura 7

Eliminar elemento, UI MongoDB



Nota. La imagen "captura de pantalla" eliminar elemento, MongoDB, (2025). Interfaz UI [Software].

Análisis de resultados obtenidos

Se ha observado los documentos, se interpreta los Suministros con cada una de sus categorías de productos o ítems, estas contiene sus propios precios, nombre, categorías y cantidad, se ha identificado además que por cada localidad hay un total de elementos o productos para Austin 3827, London 4395, Seattle 6121, Denver 8442, New York 2758 y San Diego 1891; identificando así que el ranking uno en cantidad de productos es Denver (observe *figura 8*), así mismo se observó que el valor promedio por de los ítems (productos) por libros o suministros es de 5.4868 (observe *figura 9*).

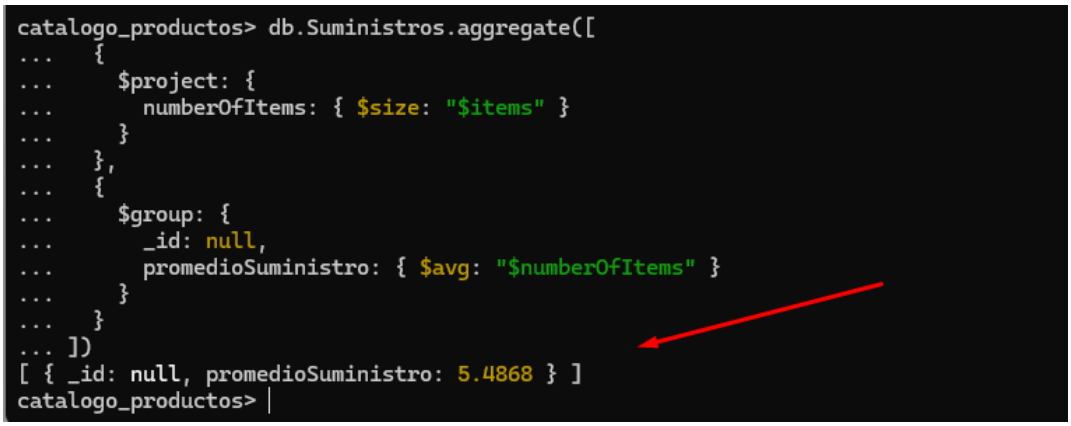
Figura 8

Suma de productos por localidad, MongoSh

Nota. La imagen "captura de pantalla" producto por localidad, MongoDB, (2025). MongoSh [Software].

Figura 9

Promedio del valor de productos por Suministro, MongoSh



Nota. La imagen "captura de pantalla" promedio de valor por item, MongoDB, (2025). MongoSh [Software].

Código MongoDB línea de comando

Código: el código se encuentra en el repositorio.

Enlace de repositorio de código (Lindsay Quintero):

https://github.com/122309/T4_AlmacenamietoConsultaBigDataNoSQLMongoDB.git

Conclusiones

Lindsay Quintero: La implementación de bases de datos no relacionales permite manejar documentos con datos semiestructurados y no estructurados. En el caso de MongoDB, estos documentos se representan en una estructura similar a JSON, lo que facilita el almacenamiento flexible de información; del mismo modo, este modelo, es posible realizar análisis de información masiva mediante consultas que permiten filtrar, sumar características, promediar valores y obtener estadísticas avanzadas. Para llevar a cabo estas operaciones es necesario comprender la estructura de cada documento y la naturaleza de sus campos; finalmente, por medio del framework de agregación, MongoDB también permite la implementación de operaciones estadísticas como conteos, sumas, promedios, máximos y mínimos, lo cual resulta fundamental para el análisis de datos en entornos de gran volumen.

Bibliografía

Carrascal Porras, F. L., & Varona Toborda, M. A. (2024). *Bases de Datos NoSQL*. Obtenido de Repositorio Institucional UNAD.: https://repository.unad.edu.co/handle/10596/62906

Contreras García, J. M., & Cabrera Lozano, R. (2024). *Visualización de datos con Power BI : Data Visualization with Power BI* . Obtenido de Epsilon: Revista de La Sociedad Andaluza de Educación Matemática.: https://research-ebsco-com.bibliotecavirtual.unad.edu.co/linkprocessor/plink?id=5307dbfa-9c36-3924-9b24-55868267b5bf.

EDteam. (2022). *¡La historia completa de las bases de datos SQL!* Obtenido de EDteam: https://ed.team/blog/la-historia-completa-de-las-bases-de-datos-sql-o-relacionales

Gutierrez Hernández, N. I., & Ibarra Limas, E. (2024). *Base de Datos para Proyectos de Big Data* . Obtenido de Congreso Internacional de Investigación Academia Journals: https://research-ebsco-com.bibliotecavirtual.unad.edu.co/linkprocessor/plink?id=b50cb8d0-5036-3c9c-a93d-baee11ae1d5e

Jiménez Beltrán , J. H. (2021). *Presentación de Datos en Power BI*. Obtenido de Repositorio Institucional UNAD.: https://repository.unad.edu.co/handle/1059

Manldonado, R. (11 de 04 de 2025). *¿Qué es el modelo ACID en bases de datos y por qué es tan importante?* Obtenido de KeepCoding: https://keepcoding.io/blog/que-es-acid-bases-datos/#:~:text=ACID%20es%20un%20acr%C3%B3nimo%20que,mantenga%20%C3%ADntegra%2C%20coherente%20y%20segura.

Miranda, A., Pilar, O., Mendoza, K., & Chango, W. (2023). *Evaluación comparativa de bases de datos NoSQL: clave/valor en un entorno de creaciones de aplicaciones . ESPOCH Congresses: The Ecuadorian Journal of S.T.E.A.M, 3(2), 129–142.* Obtenido de https://doi-org.bibliotecavirtual.unad.edu.co/10.18502/espoch.v4i1.15813

Picher Vera, D., Martínez , M. D., Soledad, M., & Bernal García, J. J. (2018). *Big Data en la universidad y Power BI como el software más óptimo para alumnos universitarios : Análisis, herramientas* . Obtenido de https://research-ebsco-

com.bibliotecavirtual.unad.edu.co/linkprocessor/plink?id=cf45b470-9ed2-3867-b25a-8b62692af273

Sarasa, A. (2016). *Introducción a las bases de datos NoSQL usando MongoDB* . Obtenido de https://research-ebsco-com.bibliotecavirtual.unad.edu.co/linkprocessor/plink?id=dd7df4c4-4b47-3d44-bc5e-d891f6b99503

Sarmiento, M. (28 de 06 de 2017). *Normalización de base de datos*. Obtenido de https://www.marcossarmiento.com/2017/06/28/normalizacion-de-base-de-datos/

UNAD. (2025). *Big Data - (202016911A_2034)*. Obtenido de Guía de aprendizaje - Tarea 4 - Almacenamiento y Consultas de Datos en Big Data: https://campus107.unad.edu.co/ses37/mod/resource/view.php?id=1990