
Laboratorio No. 2

Data Transformation - **INDIVIDUAL**

I. Modalidad y fecha de entrega

- a) El laboratorio debe hacerse individualmente
- b) Debe ser enviado antes de la fecha límite de entrega: Martes 30 de enero a las 23:59 horas a más tardar
- c) Luego de la fecha y hora límites se restará 10 puntos por cada hora de atraso en la entrega

II. Objetivo y descripción de la actividad

Esta actividad tiene como propósito principal familiarizar a los estudiantes con el entorno de MongoDB, principalmente de MongoDB Shell y sus comandos, así como también el funcionamiento general de índices.

Instrucciones generales y observaciones

El laboratorio se compone de dos incisos en donde pondrá en práctica los dos conceptos importantes vistos durante la semana. Deberá trabajar con MongoDB y poner en práctica las herramientas que vienen añadidas sobre la base de datos (MongoDB Database Tools). De igual manera, estará apoyándose de un lenguaje de programación para la generación de datos aleatorios con el objetivo de probar el rendimiento de los índices.

El primer ejercicio involucra una colección de información de usuarios en una plataforma, donde necesitamos realizar operaciones avanzadas utilizando índices en MongoDB. El objetivo es diseñar índices que optimicen consultas específicas y que se enfrenten a situaciones más complejas. El segundo ejercicio, contempla los comandos en MongoDB Shell y cómo va a interactuar con esta consola. Adicionalmente, se pretende hacer un par de consultas un poco más complejas con el objetivo de explorar algunas otras funciones de MongoDB.

Se deberá entregar un documento PDF que muestre evidencia de cada instrucción ejecutada, su resultado y el script que contenga todo. En este caso **sí es necesario** documentar el *step by step* de cada ejercicio.

Ejercicio 1 - Rendimiento de Índices

Para este ejercicio, podrá usar Atlas o su instalación de MongoDB de manera local. La idea no es qué motor de Mongo utilice, sino más bien lo importante será el análisis y la discusión derivadas de cada instrucción de este ejercicio.

1.1. Generación de Data Random: Apoyándose del comando **insert**, cree una colección de datos llamada **usuarios** con al menos **100,000 documentos** que tenga los siguientes campos (se sugiere leer primero las consultas antes de ponerse a generar la data):

- nombre (string)
- email (string)
- fecha_registro (date)
- puntos (int)
- historial_compras [array conformado por documentos embebidos con datos de productos (string) y fecha (date)]. Al generar estos datos asegúrese de generar suficiente información del producto llamado "Producto 1". Le servirá más adelante.
- dirección [documento embebido con datos de calle (string), ciudad (string), código_postal (int)]
- tags (array de strings). Al momento de generar estos valores, asegúrese de incluir suficientes valores con la etiqueta "tag2". El resto podrá quedar a su discreción.
- archivo (boolean)

- notas (string)
- visitas (entero)
- amigos (arrays de enteros). Genere algunos usuarios que tengan más de mil amigos
- preferencias [documento embebido con datos de color (string), idioma (string) y tema (string)]

1.2. Realice las siguientes consultas:

- a. Usuarios activos con más de 500 puntos
- b. Usuarios que han comprado el producto "Producto 1" en la última semana.
- c. Usuarios con la etiqueta "tag2" y que tienen más de 100 visitas.
- d. Usuarios con preferencias de color "azul" y que tienen entre 1000 y 2000 amigos.

1.3. Realice una evaluación del rendimiento de cada consulta. Apóyese del comando visto en clase (executionStats). Explique qué puede destacar de cada consulta realizada y el rendimiento asociado a la misma.

1.4. Diseño de Índices:

- a. Cree un índice compuesto sobre esta colección en los campos activo y puntos.
- b. Cree un índice compuesto sobre los campos embebidos producto (historial_compras) y fecha (historial_compras)
- c. Cree un índice compuesto sobre los campos de tags y visitas.
- d. Cree un índice compuesto sobre el campo embebido color (preferencias) y amigos.

1.5. Realice una evaluación del rendimiento de cada consulta ahora con los índices creados. Explique y argumente qué cambios observó comparando los resultados y su análisis del inciso 1.3.

1.6. Vuelva a ejecutar su script realizado en el inciso 1.1, insertando al menos 50,000 documentos más y vuelva a ejecutar el inciso 1.5. ¿Nota algo diferente en cuanto al rendimiento de las consultas? Explique sus observaciones.

Ejercicio 2 - Carga, transformación y exportación de datos.

Para la realización de este ejercicio, deberá utilizar **obligatoriamente** los comandos en MongoDB shell y no deberá apoyarse de ningún IDE.

2.1 Carga inicial:

- a. Crear base de datos: **laboratorio2**.
- b. Importar la data del archivo vehiculos.json a la colección **vehiculos**.
- c. Importar la data del archivo peliculas.json a la colección **peliculas**.
- d. Consultar la información de todas las colecciones (db.getCollectionInfos()).

2.2 Transformando los resultados de la colección **Vehículos**:

- a. Renombrar los campos de la colección a: _id, brand, model, year & price. Apóyese de la instrucción **\$rename**.
- b. Calcular la cantidad de modelos y precio promedio por año por marca. Apóyese de la instrucción **aggregate**. Exporte su resultado a un JSONArray llamado *vehiculos_brand_stats.json*.
- c. Buscar los 20 vehículos más caros de la década de los 90s. Apóyese de la instrucción **aggregate**. Exportar a CSV llamado *vehiculos_top_models.csv*.

III. Temas a reforzar

- Operaciones CRUD
- Índices
- MongoDB Database Tools

IV. Documentos a entregar

1. Un documento PDF por persona, correctamente identificado que contenga:
 - a. La instrucción ejecutada para cada inciso de cada ejercicio.
 - b. Pantallazos de cada instrucción ejecutada y su resultado. En algunos casos deberá incluir explicaciones y todos los argumentos necesarios para complementar su respuesta.
2. Script (*.js) debidamente documentado con todos los incisos, así como también los archivos exportados para el inciso 2.

V. Evaluación

- Ejercicio #1: 70 puntos (10 puntos por los incisos del 1 al 4, 15 puntos para incisos 5 y 6)
 - Ejercicio #2: 30 puntos (para el 2.1, 3 puntos para cada inciso; para el 2.2, 6 puntos cada inciso)
- Total: 100 puntos**

VI. Referencias

1. Operaciones CRUD: <https://www.mongodb.com/docs/manual/crud/>
2. MongoDB Atlas: <https://www.mongodb.com/es/cloud/atlas/register>
3. Documentación completa de comandos de mongo: <https://www.mongodb.com/docs/manual/reference/command/>