

Base de Datos II (IC4302) – Semestre 2, 2023

Apuntes 5 – Clase 10-11-2023

David Suárez Acosta – 2020038304

Esta clase fue la explicación del Examen, por lo que se tomaron apuntes de las diferentes preguntas acerca de lo explicado por el profesor.

A continuación se explican los apuntes para cada sección específica de las indicaciones.

General

- Si se va a poner un pedazo textual de un documento web, hay que reconocer la fuente.

Pregunta 1

- Se trabaja con big data, es decir un montón de información, por esta razón no funcionaría con un script de Python, hay un tema del curso donde se habla acerca de trabajar con muchos datos.
- Cuando se refiere a relaciones es cualquier tipo de relación, como relacionar conceptos con otros artículos, se puede establecer o limitar la cantidad de relaciones que van a haber por ejemplo actores de una película.
- Ejemplo de lo que hay que hacer: se elige MongoDB, se puede poner el documento JSON con los campos que tendría.

1. ¿Qué tipo de base de datos (SQL o NoSQL) utilizaría para implementar la navegación entre distintos elementos de información? ¿Es necesario que este motor de base de datos contenga todo el elemento de información o solo palabras clave que permitan establecer relaciones? Justifique su respuesta mediante la elaboración de un pequeño modelo de datos, diagrama y las relaciones que establecería entre los diferentes elementos de información, lo más importante es garantizar una navegación y que permita descubrir relaciones. (10 pts)

--

2. ¿Qué motor de base de datos utilizaría para almacenar los elementos de información y garantizar full text search? Justifique su respuesta comentando: (10 pts)

a.

Capacidad del motor para implementar full text search.

- No funciona un PostgreSQL.

b.

Particionamiento o sharding de datos.

--

c.

Representación de elementos de información en la base de datos (tablas, documentos, collections, etc.)

--

d.

Localidad de datos.

--

3. Describa la forma en la cual combinaría los dos motores anteriores (navegación y full text search) para crear un sistema simple de búsqueda y navegación de información similar al que tiene el sitio Wikipedia donde se busca un elemento de información y nos podemos mover entre términos. (4 pts)

--

4. ¿De qué forma garantizaría alta disponibilidad de las bases de datos? (4 pts)

--

5. ¿Cómo podría garantizar que las búsquedas siempre tengan un tiempo de respuesta constante? (5 pts)

--

6. ¿Cómo el uso de caches y localidad podría mejorar el rendimiento del sistema? (4 pts)

--

7. ¿Describa como realizaría el procesamiento de los datos y su carga en las bases de datos? Tome en cuenta el volumen, la variedad y velocidad de carga de los datos (Big Data) (8 pts)

--

Pregunta 2

- Si se tiene algo como "Hola #Saludo" se pueden sacar todos los #Saludo para generar una historia.
- Entre más grande sea la base de datos, más se tarda haciendo búsquedas.
- Es importante asumir que es un motor MariaDB Standalone.
- Es importante que sea una base de datos que permita representar la información de forma eficiente para que, al construir las historias, sea rápido y no crezca con el tiempo.
- Hay que imaginarse como se pueden representar las tablas.
- "Memory Footprint" es cuanta memoria va a agarrar la base de datos.

Pregunta 3

- Es importante recordar que es para bases de datos relacionales.

Pregunta 4

- Hay diferentes "Data Tiers" en Elasticsearch, hay que justificarse con esto.
- "on-premise" es instalado en casa.