



36018173. CSIFC90. MP5074. Sistemas de big data. 2023-2024. (Grupo A)

Parcial JUNIO24

Nombre completo y DNI

Leer antes de empezar.

- La duración del examen es de 2 horas con 5 minutos de cortesía.
- Al acabar el examen debes firmar en la hoja de entrega.
- Ni se permitirá la firma en la hoja de entrega ni se recogerán exámenes fuera de tiempo.
- Durante esta prueba escrita puedes utilizar libremente todo tu material escrito y digital.
- No estará permitido comunicarse con el resto de alumnado por ningún medio.
- La prueba escrita debe permanecer inmutable por lo que es obligatorio usar bolígrafo permanente y no escribir en lápiz, usar corrector o cualquier otro método que pueda poner en duda una alteración posterior.
- Es necesario usar lenguaje técnico.
- Se puede solicitar hojas extra si se necesitan.
- El espacio disponible es suficiente para contestar cada pregunta.



1.- La aplicación Whatsapp envía un “heartbeat” cada 5 segundos a sus servidores para mantener actualizado el momento de su última conexión. Indica que base de datos recomendarías para almacenar exclusivamente esta información y justifica tu respuesta con argumentos técnicos.

RESPUESTA: se esperan muchas conexiones por lo que necesitaremos muchos servidores y por tanto la bd debe ser tolerante a la partición. No es importante la consistencia hasta el extremo de que si se pierde un heartbeat y hay que esperar a otro dentro de 5 segundos no pasa nada. En cambio sí debe ser disponible, quiero respuesta aunque sea antigua. Pienso que no estoy interesado en la persistencia de los datos así que una base de datos que trabaje en memoria no estaría mal. Por ello de las bases de datos vistas elijo Redis. La respuesta debe justificar CAP y algún aspecto de ACID.

2.- “MyHeritage” es una plataforma online enfocada a la creación de árboles genealógicos donde los usuarios pueden investigar su historia familiar y conectar con familiares lejanos. Indica que base de datos recomendarías para almacenar y consultar los árboles genealógicos de esta plataforma.

RESPUESTA: en este escenario las relaciones entre personas son importantes tanto a nivel de escritura de información como de consulta. La base de datos adecuada es de tipo grafos.

3.- Explica con tus propias palabras en qué consiste el teorema CAP poniendo ejemplos con la base de datos MongoDB.

RESPUESTA: El teorema CAP les dice a los diseñadores de sistemas distribuidos que solo se puede garantizar simultáneamente 2 de estas 3 características: consistencia (misma imagen), disponibilidad (siempre se responde) y tolerancia a particiones (sistema sigue en funcionamiento incluso con fallos en comunicaciones).

Consistencia-Disponibilidad: sería un MongoDB en local. No hay distribución por lo que hay una única imagen y si el servidor funciona siempre hay respuesta.

Disponibilidad-Tolerancia: Mongo distribuido en varios servidores con réplicas configuradas para que los secundarios respondan a las lecturas y que el primario escriba sin esperar confirmación de escritura de ningún otro nodo.

Consistencia y Tolerancia: Mongo distribuido en varios servidores con réplicas configuradas para que las escrituras exijan confirmación de todos los nodos o que las lecturas solo se hagan en el primario.

4.- Explica con tus propias palabras que es y en qué consiste el sistema de versionado que incluye HBase.



RESPUESTA: HBase es una bd nosql columnar. En cada nuevo registro se inserta automáticamente un campo timestamp. Con este campo se puede determinar qué actualización de un determinado registro es el más actual y por tanto, devolver la más reciente como resultado de su lectura.

5.- ¿Cuál de las siguientes afirmaciones es correcta? Las bases de datos NoSQL...

- ☐ Siempre garantizan consistencia completa en sistemas distribuidos.
- ☐ No pueden manejar grandes volúmenes de datos ni transacciones concurrentes.
- ☐ **Son ideales para aplicaciones que requieran flexibilidad en el esquema y escalabilidad horizontal.**
- ☐ No soportan operaciones ACID.

6.- ¿Cuál de las siguientes afirmaciones es incorrecta? Las bases de datos NoSQL...

- ☐ Pueden ser categorizadas en varios tipos como de clave-valor, documentos, columnas y grafos.
- ☐ Diseñadas para proporcionar escalabilidad horizontal, lo que facilita el manejo de grandes volúmenes de datos distribuidos.
- ☐ **No ofrecen soporte para operaciones de lectura ni escritura en tiempo real y son más adecuadas para el análisis de datos por lotes.**
- ☐ Permiten esquemas flexibles y dinámicos que facilita la modificación de la estructura de datos sin interrumpir la aplicación.

7.- ¿Cuál de las siguientes afirmaciones es correcta? Neo4j...

- ☐ Es una base de datos relacional diseñada para manejar grandes volúmenes de datos estructurados.
- ☐ Es una base de datos de grafos que utiliza el modelo de datos de clave-valor para almacenar información
- ☐ Es una base de datos de grafos que utiliza el lenguaje SQL para consultas y manipulación de datos.
- ☐ **Es una base de datos de grafos que utiliza el lenguaje de consultas Cypher y está optimizada para almacenar y consultar relaciones entre datos de manera eficiente.**

8.- ¿Cuál de los siguientes lenguajes de consultas pertenece al modelo relacional?

- ☐ MongoDB Query Language (MQL)
- ☐ Cypher
- ☐ **SQL**
- ☐ GQL



9.- ¿Cuál de las siguientes afirmaciones es correcta? PowerBI

- ☐ Es una plataforma de visualización de datos desarrollada por Microsoft que permite a los usuarios crear informes interactivos a partir de datos de diversas fuentes.
- ☐ Es un sistema de gestión de bases de datos relacionales utilizado para el almacenamiento y recuperación eficiente de grandes volúmenes de datos.
- ☐ Es una biblioteca de desarrollo de aplicaciones móviles.
- ☐ Es una librería Python usada para implementar y lanzar aplicaciones distribuidas en Hadoop.

10.- ¿Qué función principal tiene Power Query dentro de Power BI?

RESPUESTA: permite transformar los datos para poder elaborar informes interactivos además provee de mecanismo para automatizar las transformaciones ante actualizaciones o nuevos datos.

11.- Explica con tus propias palabras qué diferencias hay entre MongoDB y Neo4j según el modelo lógico de datos que usan y la forma en que almacenan físicamente los datos.

RESPUESTA: En mongoDB se usa el modelo de datos de documentos compuesto por parejas de clave valor en documentos de tipo BSON. En Neo4j se usa el modelo de datos de grafos almacenando nodos y relaciones.

12.- Explica con tus propias palabras que significa que Neo4j tenga transacciones ACID.

RESPUESTA: indica que las secuencias de operaciones que se dan en la base de datos cumple con 4 propiedades. Atomicidad (o se cumple totalmente con éxito o no se hace ninguna), Consistencia (después de la operación la base de datos seguirá en estado válido), Aislamiento (las operaciones no son visibles hasta que no están completadas) y durabilidad (una vez que la operación se confirma los cambios son permanentes)

13.- Explica con tus palabras la ventaja que consideres más importante sobre el almacenamiento físico que hacen las bases de datos columnares.

RESPUESTA: guardar los registros por columnas hace que recorrer la base de datos cuando solo se busca un compo (o una familia) sea más rápido. También resulta más fácil y rápido comprimir los datos porque los tipos son iguales para toda la columna.

14.- Explica con tus propias palabras qué es GQL.

RESPUESTA: Es el estándar de lenguaje de consultas para bases de datos nosql basadas en el modelo de grafos.

15.- Una empresa soporta picos de conexiones para escrituras y lecturas en determinados momentos del mes. Su base de datos ya ha crecido en tamaño lo suficiente como para notar lentitud en las operaciones por lo que se plantean migrar a una solución en la nube o bien en local. Indica alguna ventaja e inconveniente objetivo de ambas opciones y justifica tu selección personal.

RESPUESTA: Ventajas de la nube: sin mantenimiento, escalable según necesidad, reduce costes de instalación.

Desventajas de la nube: Pago por uso, no control de ubicación de datos.

Ventajas de local: propiedad de los datos, control. Desventajas: coste de hardware y mantenimiento puede ser caro..

Escalabilidad limitada

16.- Una startup comenzó su negocio usando una base de datos NoSQL en la nube, Amazon Neptune. Después de un tiempo