

Conceptos básicos y modelos NoSQL

Ejercicio 1 (20%)

A partir de la lectura de los apuntes (transcripciones de los vídeos) de los temas I y II se pide responder con tus propias palabras a las siguientes preguntas de manera concisa (una página y media en total), **justificando la respuesta** mediante lo que has leído en los materiales. En **cada justificación deberá indicarse la cita de los apuntes, vídeo o libros en la que se sustenta. No serán válidas las respuestas que no se justifiquen o no tengan cita.**

1. Explica qué implicaciones tiene el concepto de agregado a nivel operativo para la base de datos.
2. Explica la razón por la que una base de datos orientada a agregados puede favorecer la escalabilidad horizontal.
3. Explica algunas características de la estructura interna de los agregados en el modelo documental.
4. Explica si es adecuado o no el uso de bases de datos relacionales para el procesamiento de flujos de datos.

Los criterios de evaluación para evaluar este ejercicio son los siguientes:

No logrado (C-)	Mínimamente logrado (C+)	Logrado (B)	Logrado de forma sobresaliente (A)
No se responde a la pregunta planteada o la respuesta es incompleta.	La respuesta es correcta y es mayoritariamente completa, pero no está justificada o la justificación no es coherente con los contenidos vistos en clase. Indica las referencias de los apuntes en los que se basa.	La respuesta es correcta, completa y está correctamente justificada. Indica las referencias de los apuntes en los que se basa.	La respuesta es correcta, está correctamente justificada y cumple los requisitos formales (concreción - ocupa menos de una página y media). La argumentación es excelente, aborda los aspectos relevantes e indica las referencias de los apuntes en los que se basa.

Ejercicio 2 (20%)

A partir de la lectura de los apuntes de los temas I y II indica si te parecen ciertas o falsas las siguientes afirmaciones.

Para cada una de las afirmaciones indica si es cierta o falsa, **justificando la respuesta** mediante lo que has leído en los materiales. En **cada justificación deberá indicarse la cita de los apuntes, vídeo o libros en la que se sustenta. No serán válidas las respuestas que no se justifiquen o no tengan cita**. Se valorará la concisión (una página y media para las 5 afirmaciones como máximo).

Afirmación 1

La normalización es un proceso del que se benefician las lecturas cuando se realiza un diseño adecuado de los agregados de una base de datos.

Afirmación 2

Una de las diferencias entre los modelos de datos relacionales y el modelo en grafo está en la representación de las relaciones. En el modelo en grafo, las relaciones se representan implícitamente mientras que en el modelo relacional se representan explícitamente.

Afirmación 3

Una de las razones por las que la estrategia *one size fits all* no siempre funciona es por la necesidad en las aplicaciones modernas de realizar operaciones simples sobre modelos de datos complejos.

Afirmación 4

Dos características presentes en el modelo documental son: a) Un documento puede ser recuperado mediante su clave o mediante el valor que toman sus atributos, y b) Es necesario definir explícitamente el esquema que seguirán los datos de los documentos.

Afirmación 5

El responsable de mantener las restricciones de integridad en un conjunto de agregados es el sistema de gestión de bases de datos, el cual se ocupará de comprobar que cualquier operación mantiene la consistencia de la información.

Los criterios de evaluación para evaluar este ejercicio son los siguientes:

No logrado (C-)	Mínimamente logrado (C+)	Logrado (B)	Logrado de forma sobresaliente (A)
La respuesta es incorrecta o no está justificada/referenciada.	La respuesta es incorrecta, pero la justificación es correcta, demuestra conocimiento sobre el tema y está referenciada.	La respuesta es correcta y está convenientemente justificada/referenciada.	La respuesta es correcta, está correctamente justificada/referenciada y es concisa (ocupa menos de una página y media).

Ejercicio 3 (30%)

Considerar una base de datos documental del ámbito médico donde se almacena información acerca de datos particulares de los pacientes, las dolencias que padecen y los tratamientos que reciben para las dolencias.

Se pide diseñar los tipos de documentos más eficientes para las siguientes tres consultas que se desean realizar:

Consulta 1: Para cada tipo de dolencia se quiere conocer su nombre, número de pacientes que la han padecido, número de pacientes que fallecieron debido a esta dolencia, número de pacientes que superaron la enfermedad y una lista de los tratamientos utilizados. Para cada tipo de tratamiento utilizado contra esta dolencia se quiere conocer el nombre del tratamiento, el número de pacientes que siguieron exclusivamente este tratamiento, el número de pacientes que sufrieron reacciones adversas, el número de pacientes que superaron la enfermedad usando exclusivamente el tratamiento, y los médicos que usaron estos tratamientos. Para cada médico se desea conocer su nombre y el hospital en el que trabaja.

Consulta 2 : Se desea obtener un listado de los tratamientos, donde se quiere conocer su nombre, la empresa que lo fabrica, el número de pacientes que utilizaron el tratamiento, el número de pacientes en los que tuvo efectos positivos sobre la dolencia padecida, el número de pacientes en los que no tuvo éxito y una lista de las enfermedades sobre las que se utilizó. Para cada enfermedad se quiere conocer el nombre de la enfermedad, el número de pacientes que mejoraron usando el medicamento, el número de pacientes que sufrieron reacciones adversas al usar el tratamiento, el número de pacientes que tuvieron que cambiar de fármaco para tratar

la enfermedad, y un listado de otros tratamientos alternativos considerados. Para cada tratamiento alternativo se desea conocer su nombre y la empresa que lo fabrica.

Consulta 3: Para cada hospital se quiere obtener una lista de enfermedades que ha tratado y, para cada una de ellas, la lista de los médicos que tratarán esa enfermedad (nombre, departamento y especialidad) agrupados por el medicamento que utilizarán para tratar la enfermedad.

Cada consulta requiere un tipo concreto de documento. Para cada pregunta, se pide indicar:

- Una representación gráfica del documento propuesto utilizando un diagrama de cajas anidadas como el que se explica en los apuntes sobre "Diseño de agregados" proporcionado en el aula virtual.
- Una breve explicación de la estructura del documento y una justificación del porqué de la estructura propuesta.
- Un documento JSON que ejemplifique el resultado de una potencial consulta a cada agregado.

Los criterios de evaluación para evaluar este ejercicio son los siguientes:

No logrado (C-)	Mínimamente logrado (C+)	Logrado (B)	Logrado de forma sobresaliente (A)
El agregado propuesto es incorrecto. Entendemos que un agregado no es correcto cuando no permite almacenar la información solicitada.	El agregado propuesto es correcto pero la justificación de su estructura es insuficiente y no existe un ejemplo de documento JSON.	El agregado propuesto es correcto, su estructura está correctamente justificada y se ejemplifica mediante un documento JSON.	El agregado propuesto es correcto, tiene una estructura eficiente y está correctamente justificado y ejemplificado.

Ejercicio 4 (30%)

Considerad el desarrollo de un sistema de generación de alarmas ante terremotos a nivel estatal que deberá procesar la información sísmica que se recibe de forma continuada de la red de estaciones sísmicas repartidas por todo el territorio peninsular. La aplicación debe ser capaz de procesar, en tiempo real, toda la información y buscar patrones que pudieran indicar el inicio de un terremoto y, en caso de confirmarse, deberá calcular el epicentro del movimiento sísmico mediante una triangulación de la información recibida, así como calcular otras magnitudes como la velocidad de propagación de las ondas y la intensidad entre otras. Así mismo, se sabe que la información recibida en muchos casos contiene datos erróneos, datos ausentes y que la codificación de los datos puede ser diferente dependiendo del fabricante de cada estación sísmica. También se conoce que periódicamente se cambian las estaciones sísmicas por otras modernas en las que probablemente se utilicen otras codificaciones de la información.

Con ese supuesto en mente, os animamos a realizar la lectura del artículo: “Data management in cloud environments: NoSQL and NewSQL data stores” que se puede encontrar en la siguiente dirección web: <https://journalofcloudcomputing.springeropen.com/articles/10.1186/2192-113X-2-22>

Tomando como base la lectura del artículo, así como los apuntes de los temas vistos, se pide responder de forma breve y concisa, pero justificada, a las siguientes preguntas acerca del supuesto planteado:

1. De acuerdo a las características del sistema que se plantea construir, ¿Qué tipo de base de datos sería mejor (NoSQL/Relacional/ NewSQL/Otros tipos de bases de datos)?
2. Describe breve y conceptualmente cómo serían las características del sistema en cuanto a disponibilidad, escalabilidad, consistencia de los datos, y distribución geográfica.

Los criterios de evaluación para evaluar este ejercicio son los siguientes:

Pregunta	No logrado (C-)	Mínimamente logrado (C+)	Logrado (B)	Logrado de forma sobresaliente (A)
Pregunta 1	La respuesta es incorrecta o no está justificada.	La respuesta es correcta y está justificada, pero sólo identifica alguna de las	La respuesta es correcta, está correctamente justificada, referencia correctamente el	La respuesta es correcta e indica, por cada tipo de base de datos, su adecuación y su justificación. Se referencia

		posibles bases de datos a utilizar.	artículo y los apuntes e identifica todas las posibles bases de datos a utilizar.	correctamente el artículo y los apuntes para justificar la respuesta.
Pregunta 2	La propuesta realizada no está justificada o la justificación no es coherente.	La respuesta es correcta y está bien justificada en la mayoría de las características del sistema (disponibilidad, escalabilidad, consistencia de datos y distribución geográfica).	La respuesta es correcta y está bien justificada en casi todas las características del sistema (disponibilidad, escalabilidad, consistencia de datos y distribución geográfica).	La respuesta es correcta y está bien justificada para todas las características del sistema (disponibilidad, escalabilidad, consistencia de datos y distribución geográfica).

Criterios de valoración

Los ejercicios tienen un peso de 20%, 20%, 30%, 30% respectivamente. Se valorará, para cada apartado, la validez de la solución y la claridad de la argumentación de acuerdo a la rúbrica facilitada en cada ejercicio. Muchos de los ejercicios planteados no tienen una única respuesta, por lo que es importante justificar la respuesta propuesta adecuadamente. Cualquier solución no justificada se considerará incompleta.

Formato y fecha de entrega

Tenéis que enviar la PEC al buzón de Entrega y registro de EC disponible en el aula (apartado Evaluación). El formato del archivo que contiene vuestra solución puede ser .pdf, .odt, .doc y .docx. Para otras opciones, por favor, contactar previamente con vuestro profesor colaborador. El nombre del fichero debe contener el código de la asignatura, vuestro apellido y vuestro nombre, así como el número de actividad (PEC1). Por ejemplo apellido1_nombre_nosql_pec1.docx. La fecha límite para entregar la PEC1 es el **2 de abril**.

Propiedad intelectual

Al presentar una práctica o PEC que haga uso de recursos ajenos, se tiene que presentar junto con ella un documento en que se detallen todos ellos, especificando el nombre de cada recurso, su autor, el lugar donde se obtuvo y su estatus legal: si la obra está protegida por el copyright o se acoge a alguna otra licencia de uso (Creative Commons, licencia GNU, GPL etc.). El estudiante tendrá que asegurarse que la licencia que sea no impide específicamente su uso en el marco de la práctica o PEC. En caso de no encontrar la información correspondiente tendrá que asumir que la obra está protegida por el copyright.

Será necesario, además, adjuntar los ficheros originales cuando las obras utilizadas sean digitales, y su código fuente, si así corresponde.