

**Máster universitario de Ciencia de Datos**

**Prueba de Evaluación Continua – PEC1**

**Arquitecturas de bases de datos no tradicionales – Conceptos básicos y modelo de datos.**

Autor:

Mario Ubierna San Mamés

|  |
| --- |
| Índice de Contenido |

[Índice de Contenido 2](#_Toc84760295)

[1. Enunciado 3](#_Toc84760296)

[1.1. Ejercicio 1 3](#_Toc84760297)

[1.1.1. ¿Qué significa que las bases de datos NoSQL orientadas hacia agregados favorecen los esquemas de crecimiento horizontal (escalabilidad horizontal)? 3](#_Toc84760298)

[2. Bibliografía 4](#_Toc84760299)

|  |
| --- |
| Enunciado |

## Ejercicio 1

*A partir de la lectura de los apuntes (locuciones de los vídeos) de los temas I y II se pide responder de manera concisa (una página y media en total) a las siguientes preguntas:*

### ¿Qué significa que las bases de datos NoSQL orientadas hacia agregados favorecen los esquemas de crecimiento horizontal (escalabilidad horizontal)?

Lo primero de todo es determinar qué es la escalabilidad horizontal, la podemos definir como la capacidad que tiene nuestro sistema de aumentar la capacidad de cómputo y almacenamiento añadiendo más nodos u ordenadores.

Con las bases de datos NoSQL se busca priorizar la disponibilidad de los datos frente a la consistencia de los mismos, es decir, se tiene una gran cantidad de datos que están altamente distribuidos, es por ello que la bases de datos NoSQL orientadas hacia agregados se comportan mejor en estos casos, ya que el agregado se puede replicar y distribuir por los diferentes nodos.

### Explica las ventajas del modelo de procesamiento map-reduce.

*MapReduce* es un *framework* que permite procesar de forma eficiente información de los agregados que hay en la base de datos, gracias a la programación paralela. Esto lo hace dividiendo una consulta compleja en otras más simples, de tal forma que cada consulta simple se ejecuta en cada nodo que contenga datos útiles y luego se unen para convertir los datos en información.

Esto tiene dos ventajas, la primera, minimiza el tiempo de respuesta (al dividir las consultas de forma paralela es como si tuviéramos diferentes hilos ejecutándose de un mismo programa, de tal forma que cada hilo nos proporciona la información que queremos. Si esta tecnología fuera secuencial, esta ventaja no la obtendríamos). La segunda, minimiza el transporte de datos innecesarios (como las consultas se ejecutan de forma paralela, accedemos a los datos de forma local en cada nodo, entonces a la hora de transmitir la información por la red solo se envía lo que se ha solicitado).

### Explica qué es la persistencia políglota y cómo se refleja en un proyecto.

La persistencia políglota no es más que el uso de diferentes tecnologías de bases de datos para hacer frente a diferentes problemáticas, cuando hablamos de persistencia políglota no nos estamos refiriendo a hacer uso exclusivo de una sola tecnología, sino que dentro de una misma organización y proyecto se pueden usar varias.

Dentro de un proyecto puede verse reflejada esta persistencia cuando se usan diferentes tecnologías de almacenamiento de datos, por ejemplo si somos *Twitter*, podemos tener una base de datos orientada a grafos para almacenar los datos generados en la interacción de usuarios, y por otro lado tener una base de datos relacional, en la cual se almacena la gestión de usuarios, la gestión de compañías publicitarias…

### Explica las razones por las que un modelo relacional no es buena opción para procesar flujos de datos.

Los sistemas que trabajan con un gran flujo de datos se caracterizan porque prima la disponibilidad a la consistencia de los mismos, es decir, características que son opuestas en los modelos relacionales.

Una de las razones por lo que un modelo relacional no es buena idea, es por el hecho de que estos modelos necesitan primero almacenar los datos en la base de datos para luego ser consultados, mientras que los modelos que tienen un elevado flujo de datos permiten consultar los datos antes de almacenarlos en la base de datos, esto se debe a que los puede almacenar en la memoria, con esto quiero decir que hay casos en los que los datos se tienen y necesitan en tiempo real (mercado bursátil, control de tráfico, líneas de producción…). Otra razón, es que a veces no es necesario almacenar todo el flujo de datos, es decir, hay una estructura variable, y lo modelos relacionales no soportan dicha variación.

### ¿Diferencias entre el modelo relacional y el modelo en grafo en cuanto a la representación de las relaciones?

La principal diferencia entre el modelo relacional y el modelo en grafo respecto a las relaciones es que, en el modelo relacional las relaciones representan conceptos (equipo de fútbol) y las filas son instancias de los conceptos (Real Madrid), es decir, son construcciones distintas. Sin embargo, en el modelo en grafo los conceptos y las instancias se representan de la misma forma, con los nodos.

## Ejercicio 2

|  |
| --- |
| Bibliografía |

a