



Universidad Carlos III

Arquitectura de Datos
Curso 2024-25

Práctica 1.2: Diseño de arquitectura del clúster

Grupo 05

Ingeniería Informática, 4º Curso

Luis Otero Gómez (NIA: 100472104, 100472104@alumnos.uc3m.es)

Paula Morales García (NIA: 100472240, 100472240@alumnos.uc3m.es)

Alejandro Calderón Hernández (NIA:100472087, 100472087@alumnos.uc3m.es)

Profesora: Lourdes Moreno López

Grupo: 81

Índice

Introducción	3
Descripción de los criterios	3
AreaRecreativa	3
Juego	4
RegistroClima	4
IncidenteSeguridad	5
Incidencia	6
Usuario	7
Mantenimiento	7
EncuestaSatisfacción	8

Introducción

La fragmentación es un proceso esencial para poder expandir el rango de alcance de un proyecto sin disminuir el rendimiento y la disponibilidad del sistema. Este proceso consiste en dividir los datos de un sistema de información y repartirlos entre los nodos que forman la red, con el objetivo de conseguir escalabilidad horizontal.

En este documento se justifican las decisiones tomadas para expandir el proyecto a una escala más amplia, tal y como indica el enunciado de la práctica 1. Por ello, se diseña un clúster con diferentes estrategias de fragmentación y replicación, adaptado a distintos escenarios que puedan existir.

Descripción de los criterios

Esta sección trata sobre las decisiones tomadas para los procesos de fragmentación y replicación. Por ello, se muestran diversos supuestos sobre los accesos de lectura y escritura sobre las colecciones en la base de datos para determinar qué acción realizar sobre cada una de ellas en función de su función en los agregados creados. A continuación, se detallan sus características:

AreaRecreativa

Esta colección contiene todas las áreas recreativas. Se utiliza para crear el agregado AreaRecreativa_Clima, por lo que la fragmentación se debe centrar en mejorar las operaciones relacionadas con Juego, RegistroClima, IncidenteSeguridad y EncuestaSatisfaccion.

Supuestos:

- **Tráfico:**
 - Alta frecuencia de lecturas debido a la gran cantidad de consultas realizadas sobre el estado de las áreas.
 - Actualizaciones poco frecuentes ya que solo ocurren al cambiar el estado de las áreas.
- **Patrones de uso:**
 - Se realizan mayormente operaciones de lecturas sobre el estado de las áreas, y además se realizan pocas actualizaciones sobre el estado de estas.

Clave de partición:

Se ha decidido que la mejor clave de partición es COD_POSTAL debido a que es un atributo que comparte con las colecciones Juego y RegistroClima. Esto ayuda a optimizar las consultas, dado que toda la información que esté relacionada según su zona será almacenada cerca en el mismo shard.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting con códigos postales conocidos, permitiendo una distribución equilibrada de los datos desde el inicio.

- **Zonas de Sharding:** Se definen zonas de sharding para agrupar ciertos rangos de códigos postales en un mismo shard, optimizando especialmente las actualizaciones de los datos en cada zona definida.

Juego

Esta colección contiene todos los juegos disponibles entre todas las áreas recreativas. La colección se utiliza para los agregados AreaRecreativa_Clima y Juego. Por lo tanto, para optimizar la distribución de sus datos para las consultas y actualizaciones, se debe centrar en operaciones con las colecciones AreaRecreativa y Mantenimiento.

Supuestos:

- **Tráfico:**
 - Alta frecuencia de lecturas debido a la gran cantidad de consultas realizadas sobre el estado de los juegos en ambos agregados.
 - Actualizaciones poco frecuentes ya que solo ocurren al cambiar el estado de los juegos.
- **Patrones de uso:**
 - Se realizan mayormente operaciones de lecturas sobre el estado de los juegos, y además se realizan pocas actualizaciones sobre el estado de estos.

Clave de partición:

Se ha decidido que la mejor clave de partición es COD_POSTAL debido a que es un atributo que comparte con la colección AreaRecreativa. Esto nos ayuda a optimizar las consultas, dado que toda la información que esté relacionada según su zona será almacenada cerca en el mismo shard.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting con códigos postales conocidos, permitiendo una distribución equilibrada de los datos desde el inicio.
- **Zonas de Sharding:** Se definen zonas de sharding para agrupar ciertos rangos de códigos postales en un mismo shard, optimizando especialmente las actualizaciones de los datos en cada zona definida.

RegistroClima

La colección RegistroClima contiene toda la información relacionada con los registros climáticos en cada zona donde se localizan los juegos y las áreas recreativas. Esta colección se utiliza para el agregado AreaRecreativa_Clima. Dado que se actualiza continuamente la información de los registros climáticos, se debe centrar en las operaciones con la colección AreaRecreativa.

Supuestos:

- **Tráfico:**
 - Alta frecuencia de lecturas debido a la gran cantidad de consultas que se realizan sobre los datos climáticos.
 - Las actualizaciones son bastante frecuentes por la constante inserción de datos climáticos.
- **Patrones de uso:**
 - Principalmente, se realizan operaciones de actualización por la constante recogida de datos climáticos, también se realizan con bastante frecuencia consultas sobre estos.

Clave de partición:

Se ha decidido que la mejor clave de partición es COD_POSTAL debido a que es un atributo que comparte con la colección AreaRecreativa. Esto ayuda a optimizar las consultas, dado que toda la información que esté relacionada según su zona será almacenada cerca en el mismo shard.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting con códigos postales conocidos, permitiendo una distribución equilibrada de los datos desde el inicio.
- **Zonas de Sharding:** Se definen zonas de sharding para agrupar ciertos rangos de códigos postales en un mismo shard, optimizando especialmente las actualizaciones de los datos en cada zona definida.

IncidenteSeguridad

Esta colección contiene registros sobre los incidentes de seguridad reportados de las áreas recreativas. Solo está involucrada en el agregado AreaRecreativa_Clima. Además, hace referencia a la colección raíz AreaRecreativa mediante el atributo AREA_RECREATIVA_ID, por lo que se busca optimizar las consultas sobre esta colección.

Supuestos:

- **Tráfico:**
 - Alta frecuencia de lecturas debido a la gran cantidad de consultas que se realizan sobre los incidentes de seguridad.
 - Las actualizaciones son poco frecuentes por el reducido registro de incidencias.
- **Patrones de uso:**
 - Principalmente, se realizan operaciones de lectura por la alta consulta de detalles sobre los sucesos, también se realizan con menor frecuencia actualizaciones sobre estos.

Clave de partición:

Se ha decidido que la mejor clave de partición es AREA_RECREATIVA_ID debido a que es un atributo que comparte con la colección AreaRecreativa. Esto permite obtener consultas más rápidas y eficientes de todas las incidencias para un área específica, mejorando el rendimiento en zonas de alta actividad.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting con los IDs de las áreas recreativas, permitiendo una distribución balanceada de los datos desde el inicio.
- **Zonas de Sharding:** Se crean zonas de sharding para asociar las incidencias de seguridad a áreas específicas, garantizando que las incidencias de un área permanezcan en shards alineados.

Incidencia

Esta colección contiene las incidencias reportadas por los usuarios en las áreas recreativas, con información como el tipo de incidencia, su estado y la fecha de reporte. Esta colección es necesaria para el agregado Juego e Incidencia, incluyendo referencias a las colecciones Mantenimiento y Usuario mediante MANTENIMIENTO_ID y USUARIOS_ID.

Supuestos:

- **Tráfico:**
 - Alta frecuencia de lecturas y actualizaciones debido a la gran cantidad de consultas y modificaciones que se realizan sobre los estados de las incidencias e incidencias reportadas.
- **Patrones de uso:**
 - Las lecturas y actualizaciones son muy frecuentes para gestionar las incidencias reportadas.

Clave de partición:

Se ha seleccionado como clave de partición MES_ANIO. Esta clave es un campo derivado de la fecha de reporte (FECHA_REPORTE), que contiene el mes y el año de la incidencia en formato “YYYY-MM”. Este campo permite agrupar las incidencias por periodos de tiempo, lo cual facilita las consultas para reportes y análisis temporales.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting en Incidencias para períodos de alto tráfico, como las estaciones con mejor clima (primavera y verano), para distribuir la carga de manera uniforme desde el inicio y evitar hotspots.

- **Zonas de Sharding:** Se definen zonas específicas para los meses de mayor volumen de reportes, como los meses de verano, asegurando que estos se distribuyan de manera equilibrada entre diferentes shards.

Usuario

Esta colección contiene toda la información personal sobre los usuarios que han reportado incidencias o han proporcionado retroalimentación. Esta colección se utiliza en el agregado Incidencia. Para poder establecer una relación en la distribución de su información con la colección Incidencias, esta incluye el NIF de los usuarios que reportaron cada incidencia.

Supuestos:

- **Tráfico:**
 - Baja frecuencia de lecturas debido a que no se suelen realizar consultas sobre la información personal de los usuarios.
 - Las actualizaciones son bastante frecuentes por la constante inserción de reportes por parte de los usuarios.
- **Patrones de uso:**
 - Principalmente, se realizan operaciones de actualización por la constante participación de los usuarios en los comentarios sobre los sucesos ocurridos en las áreas.

Clave de partición:

Se ha decidido que la mejor clave de partición es NIF debido a que es un atributo que comparte con la colección Incidencias. Esto ayuda a optimizar las consultas, dado que toda la información que esté relacionada con las incidencias estará almacenada cerca en el mismo shard. Así, mejora especialmente las actualizaciones de usuarios en la base de datos.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting los NIFs de los usuarios, permitiendo una distribución equilibrada de los datos desde el inicio.
- **Zonas de Sharding:** Se definen zonas de sharding para distribuir las incidencias reportadas por los mismos usuarios entre varios shards. De esta manera se evita que las actualizaciones de incidencias se concentren en el mismo shard.

Mantenimiento

Esta colección contiene el historial de mantenimiento realizado sobre cada juego. Se utiliza en el agregado Juego, ya que se quiere consultar el historial para poder analizar el estado de cada juego. Contiene referencias a la colección Juego mediante el campo JUEGO_ID, por lo que se busca optimizar las operaciones relacionadas con esta colección.

Supuestos:

- **Tráfico:**
 - Lecturas frecuentes debido a las altas consultas sobre el historial de mantenimiento de los juegos.
 - Actualizaciones frecuentes al gestionar el mantenimiento de los juegos.
- **Patrones de uso:**
 - Son bastante frecuentes tanto las lecturas para consultar el historial, como las actualizaciones para gestionar el mantenimiento, pero se realizan mayoritariamente lecturas.

Clave de partición:

Se ha decidido que lo mejor es realizar una clave de partición compuesta por:

- **MES_ANIO:** Un campo derivado del atributo FECHA_INTERVENCION sobre el área recreativa (FECHA), que contiene el mes y el año de la encuesta en formato “YYYY-MM”. Este campo permite agrupar las encuestas por periodos de tiempo para mejorar el rendimiento en las consultas y actualizaciones.
- **JUEGO_ID:** Define zonas específicas para almacenar la información según los juegos disponibles, garantizando el guardado de la información en diferentes shards.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza pre-splitting en Mantenimiento para distribuir la información en periodos de tiempo, especialmente aquellos con un elevado tráfico, como los meses de verano. Esto permite distribuir la carga de manera uniforme y evitar hotspots.
- **Zonas de Sharding:** Se crean zonas de sharding específicas para los meses con mayor volumen de intervenciones realizadas sobre los juegos. Así, se asegura que los datos se distribuyan en diferentes shards.

EncuestaSatisfacción

Esta colección incluye datos sobre las encuestas de satisfacción de los usuarios respecto a las áreas recreativas y sus elementos. La colección EncuestaSatisfacción se aplica para el agregado AreaRecreativa_Clima. También contiene referencia a la colección AreaRecreativa mediante el campo AREA_RECREATIVA_ID, por lo que es necesario optimizar las operaciones relacionadas con esta colección.

Supuestos:

- **Tráfico:**
 - Alta frecuencia de lecturas debido a que se suelen realizar consultas sobre la satisfacción del usuario.
 - Las actualizaciones son también bastante frecuentes por la constante inserción de encuestas de satisfacción de los usuarios.
- **Patrones de uso:**
 - Se realizan tanto operaciones de lectura como de actualización por la constante participación de los usuarios en las encuestas sobre las áreas.

Clave de partición:

Se ha seleccionado una clave de partición compuesta que incluye:

- **MES_ANIO:** Un campo derivado del atributo FECHA de la encuesta de satisfacción sobre el área recreativa (FECHA), que contiene el mes y el año en formato “YYYY-MM”. Este campo permite agrupar las encuestas por periodos de tiempo para mejorar el rendimiento en las consultas y actualizaciones de las encuestas.
- **AREA_RECREATIVA_ID:** Define zonas específicas para almacenar la información según las áreas disponibles. Así, se garantiza que se guarda la información en diferentes shards.

Estrategia de Sharding:

Para balancear la carga y mejorar la eficiencia de las consultas, se aplican las siguientes estrategias de sharding:

- **Pre-Splitting:** Se realiza un pre-splitting en la colección EncuestaSatisfaccion, lo cual permite distribuir la carga de manera uniforme.
- **Zonas de Sharding:** Se definen zonas de sharding para distribuir las encuestas realizadas por cada área. Esto asegura que la información se distribuya en diferentes shards.