

Actividad Bases de Datos NoSQL

Las bases de datos no relacionales o NoSQL nos permiten almacenar información en situaciones donde las bases de datos relaciones pueden tener problemas de escalabilidad y rendimiento. Estas bases de datos están diseñadas para modelos de datos específicos y tienen esquemas flexibles.

En este ejercicio se pide modelar el caso de uso propuesto para una de las BDNosQL vistas en el módulo: Cassandra, InfluxDB, Neo4J, o MongoDB.

Caso de uso

Por ejemplo, suponga que queremos modelar el esquema de una base de datos sencilla para almacenar la información de **terremotos registrados a nivel mundial**. Si utilizásemos una base de datos relacional, almacenaríamos los datos en diferentes tablas: una para los **eventos sísmicos**, otra para las **características geográficas**, otra para los **parámetros del fallo tectónico**, y otra para la **evaluación del riesgo de tsunami**. Estas tablas estarían relacionadas entre sí mediante **claves primarias y foráneas**.

El modelo relacional está diseñado para garantizar la **integridad referencial** entre las distintas entidades (por ejemplo, asegurar que un evento sísmico esté asociado a un país o zona sísmica existente) y para **reducir la redundancia** de la información mediante la **normalización**.

Sin embargo, en el caso de los datos sísmicos, el **volumen**, la **variedad** y, especialmente, la **velocidad de generación** de la información (nuevos registros de sensores, actualizaciones de magnitudes, alertas de tsunami, datos geoespaciales y de profundidad) hacen que un enfoque relacional resulte poco eficiente. Las bases de datos relacionales presentan limitaciones al escalar horizontalmente y gestionar grandes volúmenes de inserciones o lecturas en tiempo real.

En este ejemplo, si pensáramos en un esquema relacional tendríamos las siguientes tablas:

- countries

country_id (PK)

name

- regions

region_id (PK)

country_id (FK → countries.country_id)

name

- seismic_zones

zone_id (PK)

name

- earthquakes

eq_id (PK)

event_time

region_id (FK → regions.region_id)

zone_id (FK → seismic_zones.zone_id)

latitude

longitude

depth_km

mw

mb

ms

energy_joule

intensity_mmi

casualties_est

- fault_parameters

eq_id (PK, FK → earthquakes.eq_id)

strike

dip

rake

- tsunami_assessments

eq_id (PK, FK → earthquakes.eq_id)

potential

tsunami_height_m

distance_to_coast_km

Por ejemplo, un registro típico en JSON podría ser, en la que cada documento representa un terremoto significativo.

```
{
  _id: ObjectId,
  eq_id: String,
  date: ISODate,
  location: {
    country: String,
    region: String,
    latitude: Number,
```

```
longitude: Number
},
magnitude: {
  mw: Number,
  mb: Number,
  ms: Number
},
depth_km: Number,
energy_joule: Number,
intensity_mmi: Number,
tsunami_risk: {
  potential: Boolean,
  tsunami_height_m: Number,
  distance_to_coast_km: Number
},
fault_parameters: {
  strike: Number,
  dip: Number,
  rake: Number
},
seismic_zone: String,
casualties_est: Number,
updated_at: ISODate
}
```

Y registros de ejemplo, como los del dataset de Kaggle:

[Global Earthquake-Tsunami Risk Assessment Dataset](#)

Tareas

Las tareas que se piden son las siguientes:

1. Explicar las ventajas y desventajas de la BD NOSQL seleccionada, en resumen, el por qué.
2. Definir el esquema, y las sentencias de creación de este, ya sean generación de tablas, grafos o documentos...

3. Definir las sentencias de inserción para 100 registros del dataset. Si es necesario utilizar un script para adaptar la inserción desde el csv o usar un API (por ejemplo, de Python), adjuntar el script y explicar el por qué y cómo se ejecutaría.
4. Definir (si se puede) sentencias de modificación para dos de los registros, convertir el nombre de una zona sísmica a mayúsculas:
5. Definir las siguientes consultas:
 - a. **Consulta por un terremoto específico**
 - i. Filtrando por el **año del evento mayor de 2015**
 - ii. Filtrando que el **país donde ocurrió empiece por “Japa...”** (por ejemplo, *Japan*)
 - b. **Consulta por un país concreto donde se han producido terremotos**
 - i. Mostrar todos los **terremotos con magnitud superior a 7.0**
 - ii. Incluir sólo los **que presenten riesgo potencial de tsunami**

Nota: Para cada consulta ofrecer una tabla de tiempos de ejecución sobre la BD elegido, indicando las conclusiones que os aporte, por ejemplo, si redefinierais algo del esquema o incluiríais (si se puede) conceptos como índices.

Opción practica libre sobre BD vectoriales

****** Opcionalmente si se quieren usar para la práctica las bases de datos de vectores, deberá definirse el caso de uso en el que se utilizan, y un ejemplo del uso.

Entrega

Los documentos (documento en el que se especifiquen las tareas anteriores como secciones) se deberán enviar como tarea en ALUD y por email a antorre@deusto en un zip, con nombre y asunto: [NombreAlumno]_actividadBDNOSQL_ATBD.zip, además si se usa código se dejará disponibles en repositorios privados de Github/Gitlab, y se indicara claramente en la documentación la dirección y requisitos de acceso para visualizarlos dentro de la memoria en una sección dedicada “Código”.

Fecha entrega limite en ALUD