

---

# Caso Final: Persistencia polígloa y Bases de datos noSQL

Máster en Business Analytics y Big Data



---

**Asignatura:** Bases de Datos No Convencionales  
**Módulo:** Tecnologías de Big Data/Gestión de Datos  
**Coordinador:** Elena García Barriocanal, [elenagarcia@campusciff.net](mailto:elenagarcia@campusciff.net)  
**Profesores:** Jordi Conesa [jordiconesa@campusciff.net](mailto:jordiconesa@campusciff.net)  
Elena García Barriocanal, [elenagarcia@campusciff.net](mailto:elenagarcia@campusciff.net)

## Descripción del contexto y de los requisitos

En aras de acabar con la piratería, un conjunto de productores independientes ha creado un portal web para visualizar películas y series *online*. El portal, llamado *Veeme*, funciona mediante suscripción mensual.

En los últimos meses, el número de usuarios de *Veeme* ha aumentado exponencialmente gracias a su novedad, su publicidad y a su eficaz sistema de recomendación. Debido a este gran crecimiento, se debe renovar con la máxima celeridad su sistema de almacenamiento.

Nos piden, como conocedores de las tecnologías NoSQL, que hagamos una propuesta de sistema de almacenamiento para el nuevo sistema. El nuevo sistema de almacenamiento deberá satisfacer las siguientes necesidades:

1. Gestión de suscripciones: debido al alto número de usuarios concurrentes, en los pasados meses el portal ha tenido muchos problemas de disponibilidad que han provocado que, en determinadas ocasiones, los usuarios no hayan podido visualizar ninguna película ni serie aun teniendo sus suscripciones activas. Para solucionar este tema, se ha decidido utilizar un sistema de almacenamiento de alta disponibilidad que gestione la información de suscripciones.

La información a gestionar es, dado un usuario, saber si tiene la suscripción activa. La información de suscripciones no se actualiza muy a menudo, por lo tanto no es necesario que el sistema satisfaga consistencia fuerte, pero si deberá garantizar la máxima disponibilidad.

2. Catálogo de películas y series: el catálogo de películas y series disponibles debe ser accesible por todos los usuarios por igual, aunque sólo los usuarios con una suscripción activa podrán visualizarlas completamente. El visionado de las novedades en el catálogo supone una fuente de ingresos muy importante, por lo que mantener la integridad de datos en el catálogo es de vital importancia.

El catálogo está formado por el listado de las películas y series disponibles. En algunos casos, puede ser que estas sólo estén disponibles en determinados países, debido a sus derechos de autor y restricciones comerciales.

El catálogo también incluye los actores, y la relación de las películas/series donde han actuado. Para cada película/serie se deben indicar las películas/series similares, las categorías a la que pertenece (acción, suspense, comedia, etc.), el año de su estreno y su sinopsis. Aun teniendo disponibles la gran mayoría de películas y series de los últimos años, se tiene la certeza de que el sistema no tendrá la necesidad de escalar horizontalmente porque el número de nuevas películas, series, actores y directores esperado es bastante reducido.

Por otro lado, este sistema de almacenamiento sólo contendrá metadatos sobre las películas, las series y los actores. Las películas y series en sí se almacenarán en un sistema de streaming separado.

3. Análisis de usuarios: es de vital importancia analizar el comportamiento implícito de los usuarios, para ello el sistema debe almacenar datos sobre la navegación de los usuarios con el fin de extraer información sobre sus preferencias. A tal fin se almacenará información sobre cada conexión que el usuario realiza al portal, indicando la hora de la conexión, la página vista y el tiempo que ha permanecido en ella. Adicionalmente, a partir de la clave del usuario, se podrá acceder a información sobre todas sus conexiones.

*Veeme* está teniendo mucho éxito y, a pesar de contar ya con millones de usuarios activos, duplica el número de usuarios activos cada dos meses. Por tanto, para almacenar las conexiones de los usuarios es de vital importancia escoger un sistema

de almacenamiento que sea capaz de escalar horizontalmente acorde al crecimiento esperado. Por otro lado, al haber decenas/centenares de conexiones por segundo, el sistema elegido deberá garantizar una alta disponibilidad.

4. Sistema de recomendación ágil para los usuarios: se debe actualizar el sistema para realizar las recomendaciones a los usuarios. Los usuarios pueden valorar del 1 al 10 cada una de las películas o capítulos de las series que han visualizado y, de forma independiente, pueden puntuar también a sus actores.

Por otro lado, también interesa ofrecer recomendaciones a todos los usuarios de acuerdo con sus preferencias de forma rápida. El sistema que calcula las recomendaciones para cada usuario queda fuera del ámbito de este trabajo. Supongamos que ya está implementado, que obtiene los datos de las fuentes necesarias y recomienda películas/series para cada usuario de forma periódica. Aquí, sólo nos interesa definir la base de datos que almacene para cada usuario la información de sus votaciones, de las últimas películas y series que ha visto y de la lista de recomendaciones proporcionadas por el sistema (que será actualizada por el recomendador de forma periódica).

## Entregables

Tomando como marco de actuación el concepto de persistencia políglota, indique y argumente el modelo de datos a utilizar (agregado clave-valor, agregado documental, agregado por columnas, en grafo) que mejor satisfaga cada una de las necesidades expuestas en la anterior descripción del contexto y de los requisitos. Utilice tantos modelos de datos como considere necesario a fin de satisfacer lo mejor posible todas las necesidades.

Pueden tomarse como referencia las características técnicas de distribución, eficiencia e implementación de los productos estudiados para cada tipo de modelo: Riak, MongoDB, Cassandra y Neo4J. En caso de que se prefiera hacer referencia a cualquier otro producto, es necesario especificarlo en la solución propuesta.

Para cada propuesta, que deberá ocupar una página a lo sumo, se deberá ofrecer la siguiente información:

- Justificación: Explicación de por qué el modelo escogido se adapta adecuadamente al problema planteado. Justifique también por qué los modelos descartados no son tan adecuados.
- Diseño conceptual: Descripción de cómo será la estructura del agregado o del grafo a utilizar.

Una vez generado el documento con la información requerida, se deberá crear también una presentación en vídeo (máximo 5 minutos) donde se exponga la solución planteada al equipo directivo de Veeme.

La solución a la actividad puede realizarse en grupos de 2 o 3 estudiantes.