
Bases de Datos Clave-Valor: Riak

Máster en Business Analytics y Big Data



Asignatura: Bases de Datos No Convencionales

Módulo: Tecnologías de Big Data/Gestión de Datos

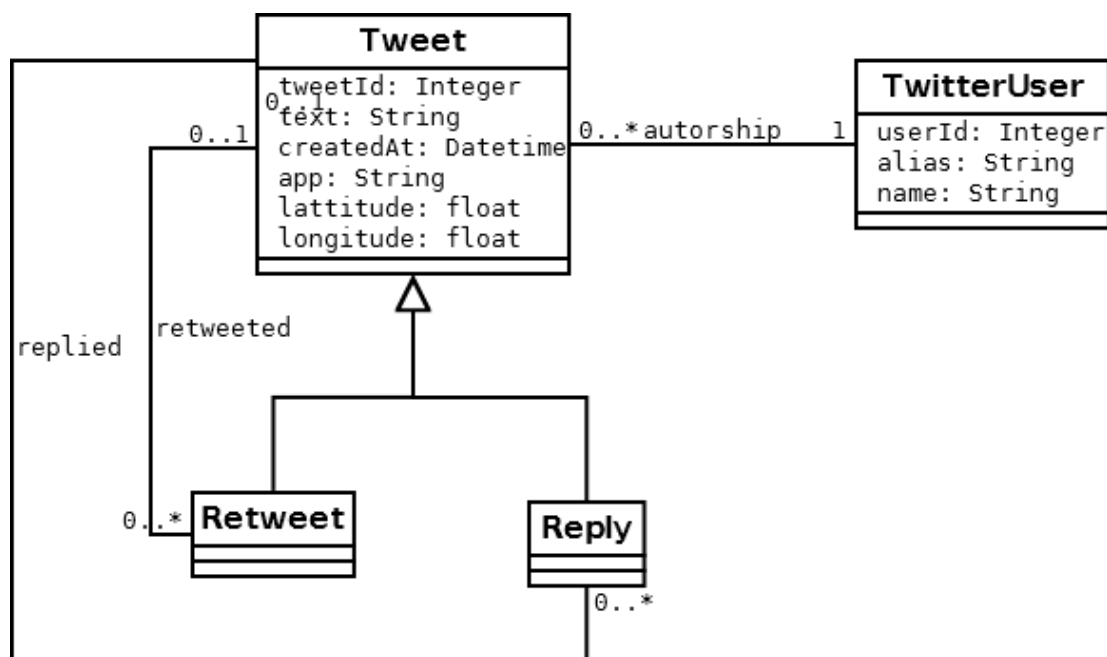
Coordinador: Elena García Barriocanal, elenagarcia@campusciff.net

Profesores: Jordi Conesa jordiconesa@campusciff.net

Elena García Barriocanal, elenagarcia@campusciff.net

Suponed que os encargáis de gestionar las bases de datos en una empresa y vuestro jefe ha oído hablar de la alta disponibilidad y versatilidad de Riak, de tal manera que os pide que hagáis el estudio para evaluar el almacenamiento de un conjunto de datos en dicha base de datos.

El conjunto de datos que se quieren almacenar tienen que ver con la extracción de información de Twitter. A continuación podéis ver el modelo conceptual de los datos a almacenar.



Como vemos, se quiere almacenar un conjunto de tweets y los usuarios que los escriben. Respecto a los tweets interesa almacenar su identificador, el texto del tweet, cuándo se creó, por qué aplicación y la ubicación desde la que se envió. También se quiere saber que tweets son replis de otros tweets, que tweets son reenvíos de otros tweets y los usuarios que realizan estas operaciones de reply y retweet. Evidentemente, no todos los tweets tendrán las mismas propiedades (o atributos). Por otro lado, interesa almacenar información sobre los usuarios y saber qué tweets envió cada usuario.

Una vez creados en la base de datos, los tweets nunca se modificarán. Cada noche se añadirán los nuevos tweets del día y las nuevas relaciones generadas (los nuevos replis y retweets).

Vuestro jefe os ha comentado que quiere que la base de datos ofrezca una alta disponibilidad y que no le importa que obtengamos algún tweet obsoleto de vez en cuando, pero eso sí, que los datos de usuarios de Twitter siempre deben mostrar un estado (o valor) consistente.

Teniendo en cuenta toda esta información se pide que:

- a) Hagáis un diseño de los datos en Riak: es decir, definir los buckets que crearíais en Riak, la clave de cada bucket y los posibles links que definiríais entre objetos.
- b) Evaluéis la conveniencia de usar Riak para el caso que os proponemos.
- c) Pobléis la base de datos con 4 o 5 objetos de cada tipo y las relaciones necesarias entre ellos. Una vez poblada la base de datos, realizad una consulta de un tweet por clave y una consulta de todos los tweets de un usuario a partir de la clave de usuario. Indicad las sentencias utilizadas, tanto para crear los objetos como para realizar las consultas solicitadas.