

**Pig** es una plataforma de alto nivel en la que se utiliza el lenguaje Pig Latin para abstraer la programación en MapReduce y hacerla similar a consultas SQL. No obstante no almacena bases de datos persé.

El mapeo inicial sería bastante similar al que se realiza con MapReduce, partir el CSV en líneas y quedarnos con los campos que nos interesan. Pig Latin facilita mucho las siguientes partes. En dos órdenes (GROUP y AVG) cumple la funcionalidad del primer reducer, un código bastante más extenso y complejo. El segundo mapeo sería mucho más sencillo en Pig gracias a su función JOIN que permite unir el resultado de la etapa anterior con los datos que nos interesan del CSV de películas (en este caso los títulos). MapReduce por su parte no ofrece esta función y hay que hacerlo manualmente. Una vez tengamos todos los títulos con sus respectivas medias sólo habría que filtrar por nota, igual que en los otros casos Pig ofrece una función FILTER que permite simplificar lo que en MapReduce se haría a mano. Además, Pig ofrece la opción de definir funciones a los usuarios (UDF) e integrarlas fácilmente para ampliar la funcionalidad si fuese necesario.

**Hive** es un sistema de almacenamiento de datos en bases de datos. No se basa en programación si no en consultas sobre estas bases de datos mediante el lenguaje HiveQL, muy similar a SQL.

Habría que comenzar creando las bases de datos, tanto para notas como para películas, sobre las que se realizarán las consultas. Una vez estructurados los datos lo único que habría que hacer es realizar las pertinentes consultas. Primero una SELECT del id\_movie agrupando por película y aplicando la media a las notas de cada grupo. Para obtener el nombre de la película se haría un JOIN de las dos bases de datos quedándonos con el movie\_name y la media calculada anteriormente. Finalmente, para tener las películas de un cierto rango se haría una SELECT condicional (WHERE media > x). Es un proceso mucho más simple, corto y manejable.