

Ejercicio 6.2: Pig & Cassandra

Versiones del software utilizado:

- jre1.7.0_71
- hadoop-2.6.0
- pig-0.13.0
- cassandra-2.1.2

Explicación del ejercicio

Para elegir el modelo de datos se ha asumido lo siguiente:

- La mayoría de los datos que se desean calcular son invariantes una vez ha finalizado el mes, mientras el mes no haya finalizado los datos del mes en curso son susceptibles de cambiar con cada nuevo dato que se añade a la serie (entendiendo aquí por serie, la serie de datos para cada empresa)
- Los datos que se desean calcular tienen una finalidad principal de lectura
 - La única serie que se desea capturar es `daily(open)-daily(close)/avg_month(daily(open)-daily(close))` de ahora en adelante la llamaremos resultado
- Cada barra es única para cada empresa y día, así que la tupla (empresa, día) puede actuar como PRIMARY KEY

Con estos supuestos, el modelo óptimo para almacenar esos datos es:

```
{empresa: string, dia: datetime, resultado: double}
```

Debido a que los datos (una vez ha finalizado el mes) solo es necesario calcularlos una vez por día, los calculamos en Pig y almacenaremos en cassandra el modelo de datos anterior.

Si se desean actualizar los datos solo sería necesario calcular el último mes en curso con los nuevos datos recibidos, pero no es el objetivo de esta práctica.

NOTA: este ejercicio no se pudo realizar en la máquina virtual usada en las clases porque no se podía disponer de ella, en su lugar se ha ejecutado en una máquina local con linux con la siguiente descripción:

```
Linux X 3.16.0-25-generic #33-Ubuntu x86_64 x86_64 x86_64 GNU/Linux
```

El código utilizado se puede encontrar [aquí](#)