

## Proyecto Intermedio 1

### Procesamiento Distribuido con Map-Reduce

**Utilizando el dataset de “house prices”\*, crear y ejecutar en AWS EMR los respectivos scripts map-reduce que permitan realizar las siguientes consultas sobre esos datos.**

1. Número de casas vendidas por año  
ej:  
1995 , 203  
1996, 101
2. Precio promedio de las casas en cada ciudad
3. Casa con menor precio por ciudad.
4. Listado de precios de menor a mayor en la ciudad de STAMFORD para 2015
5. Mes con mayor número de ventas en cada año  
ejemplo  
1995 03  
1996 07
6. Conteo de condados con el respectivo número de ciudades. Ej:

No_Condados	No_ciudades	
2	5	(hay 2 condados con 5 ciudades)
4	2	(hay 4 condados con 2 ciudades)
3	8	(hay 3 condados con 8 ciudades)

Se debe entregar:

- Scripts en python para cada caso
- Resultados en S3
- Descripción del cluster utilizado(número de nodos, tipos de instancia, etc)
- Tiempo de ejecución para cada caso

\*<https://bigdata20192.s3.amazonaws.com/uk-housing-prices-paid.zip>