

BLOQUE SPARK SQL

EJERCICIO 1. Se nos proporcionan una serie de datos , (James", "Sales", 3000),

("Michael", "Sales", 4600),
("Robert", "Sales", 4100),
("Maria", "Finance", 3000),
("James", "Sales", 3000),
("Scott", "Finance", 3300),
("Jen", "Finance", 3900),
("Jeff", "Marketing", 3000),
("Kumar", "Marketing", 2000),
("Saif", "Sales", 4100)

Queremos aplicar **funciones de agregado** en SCALA SPARK

- Mostrar los datos por pantalla en forma de tabla
- Calcular la media de los salarios
- Saber cuántos departamentos y salarios distintos tenemos

EJERCICIO 2. . Repetir la misma operación pero ahora vamos a mostrar:

- Mostrar el primer elemento
- Mostrar el último
- Mostrar el salario máximo y mínimo

EJERCICIO3. JOINS

OPERACIÓN JOIN Se nos facilitan dos seq de datos con las que vamos a practicar las distintas estructuras de join de las que disponemos en SPARKsql

```
val emp = Seq((1,"Smith",-1,"2018","10","M",3000),  
              (2,"Rose",1,"2010","20","M",4000),  
              (3,"Williams",1,"2010","10","M",1000),  
              (4,"Jones",2,"2005","10","F",2000),  
              (5,"Brown",2,"2010","40","", -1),  
              (6,"Brown",2,"2010","50","", -1)  
)
```

```
val dept = Seq(("Finance",10),  
  ("Marketing",20),  
  ("Sales",30),  
  ("IT",40)  
)
```

Aplicar: Inner, Full, right y left join

EJERCICIO4 .Mostrar el contenido de un fichero CSV que habremos guardado en un dataframe y que mostraremos los campos. Mostrar los 5 primeros

EJERCICIO5 usando el csv anterior debemos mostrar los datos del ese CSV usando ahora sintaxis SQL

EJERCICIO6. Realizar la misma consulta de antes pero ahora filtrando los datos y limitando su salida

EJERCICIO7. Leer el contenido de un fichero JSON multilinea

EJERCICIO 8. Listar todo el árbol, mostrar el schema y un campo concreto del fichero json que se os facilita.