

EJERCICIO CHECKPOINT



Fons Social Europeu

L'FSE inverteix en el teu futur

Sumario

CheckPoint.....	2
-----------------	---

CheckPoint

Para poner el practica el funcionamiento de los puntos de control o CheckPoint, vamos a crear un RDD simple por ejemplo: `rdd = 1,2,3,4,5`

A partir de este RDD vamos a iterar y cada iteración vamos a sumar 1 a cada elemento del RDD, usaremos la función `map` para hacer esto.

Vamos a ir subiendo el número de iteraciones hasta que la memoria ya no de más de sí y nos de error.

Para solucionar este error usaremos Check Points

1.- Lo primero que haremos será crear un RDD de 5 elementos:

```
rdd = sc.parallelize([1,2,3,4,5])
```

2.- Iteramos y cada iteración le sumamos 1 a cada elemento:

```
for i in range(200):  
    rdd = rdd.map(lambda myInt: myInt + 1)  
    #print("Iteration",i)  
rdd.collect()
```

3. Vamos subiendo el número de iteraciones hasta que nos de error

```
for i in range(800):  
    rdd = rdd.map(lambda myInt: myInt + 1)  
    #print("Iteration",i)  
rdd.collect()
```

```
for i in range(800):  
    rdd = rdd.map(lambda myInt: myInt + 1)  
    #print("Iteration",i)  
rdd.collect()
```

```
-----  
RecursionError                                Traceback (most recent call last)  
C:\apps\spark-3.1.1-bin-hadoop2.7\python\pyspark\cloudpickle\cloudpickle_fast.py in dump(self, obj)  
    539         try:  
--> 540             return Pickler.dump(self, obj)  
    541         except RuntimeError as e:
```

4. Vamos a usar checkpoints para así hacer que no tenga que estar guardando toda la información y ahorrarle espacio, almacenando en disco esa información lo primero que haremos será creamos nuestro directorio de checkpoints

```
sc.setCheckpointDir("checkpoints")
```

5.- Añadimos a nuestro bucle que cada 10 iteraciones por ejemplo haga un checkpoint:

```
rdd = sc.parallelize([1,2,3,4,5])
```

```
for i in range(50):
```

```
    rdd = rdd.map(lambda myInt: myInt + 1)
```

```
    print("Iteration",i)
```

```
    # Cada 10 iteraciones hacemos un checkpoint
```

```
    if i % 10 == 0:
```



















```
        print("Checkpoint")
```

```
        rdd.checkpoint() #Hacemos checkpoint
```

```
        rdd.count() #Hacemos un count ya que hasta que no se hace una acción no se ejecuta el checkpoint
```

6.- Ahora el bucle no fallará, podemos consultar la carpeta que hemos creado y podremos ver que se han almacenado ahí los checkpoints .

equipo > Documentos > Ejercicios Spark > Sesión 2 > checkpoints

Nombre	Fecha de modificación	Tipo
 1b16a5bb-412c-4eed-8133-2a7fab25fe23	18/05/2021 13:44	Carpeta de ar
 2af74d44-79ad-43f1-9a8a-01ff315512c6	18/05/2021 12:56	Carpeta de ar
 4b10be59-3ae3-48fc-95da-a983195c7061	18/05/2021 12:55	Carpeta de ar
 5a35440a-bb3e-4684-9f34-d8a98c0bbb3c	18/05/2021 12:56	Carpeta de ar
 73ef92be-0b44-4042-8765-70be6a51097a	18/05/2021 12:54	Carpeta de ar
 80a5a66e-6e51-4286-a8cf-7375195b3d88	18/05/2021 12:53	Carpeta de ar
 102f3ab8-e885-4255-96ab-10798f2447b0	18/05/2021 13:35	Carpeta de ar
 328acdda-375e-41de-bd22-122ea3f30daa	18/05/2021 12:55	Carpeta de ar
 373b4a3f-8633-49cf-a4ba-594910f22641	26/05/2021 20:04	Carpeta de ar
 805ba427-35da-458c-9cf6-8bb840feddb4	18/05/2021 13:05	Carpeta de ar
 988fefef-284f-472c-8d96-bb54e3c8690a	26/05/2021 20:02	Carpeta de ar
 2219e6d1-187a-482a-98bc-b03796c24186	18/05/2021 12:55	Carpeta de ar
 4626ed96-2d4b-4ee9-8789-df7f7cacf373	18/05/2021 13:34	Carpeta de ar
 95922158-5dca-40fe-b91b-c46eefc09e41	18/05/2021 12:50	Carpeta de ar
 bf363b27-c0d7-4222-ab7a-26e00a2526de	18/05/2021 13:05	Carpeta de ar
 c9b3fdd1-52d7-4532-a7a0-26445cf147fd	18/05/2021 12:53	Carpeta de ar
 ca7bfc4f-c4a4-4cae-87ba-d31280cdaa09	18/05/2021 13:05	Carpeta de ar
 d91805e0-ac4f-4740-a905-07b072ec693d	26/05/2021 20:00	Carpeta de ar
 e78e4fd5-d28a-4fdd-9325-51d00f1d38d4	18/05/2021 0:05	Carpeta de ar
 ece7a74f-df1a-48a1-8815-7ad7c5b7fd53	26/05/2021 20:05	Carpeta de ar