

## EJERCICIOS DE PRESTO Y KAFKA

He creado el `text_topic2`, ya que el `text_topic` ya estaba creado:

```
spark@spark-virtualBox:~/kafka_2.11-2.1.0
Archive Editor Ver Buscar Terminal Ayuda
[2024-07-18 11:55:09,911] INFO Got user-level KeeperException when processing sessionid:0x10000871d870006 type:setData cxid:0x5 zxid:0x41a txntype:-1 reqpath:/ Error Path:/config/topics/text_topic2 Error:KeeperErrorCode = NoNode for /config/topics/text_topic2 (org.apache.zookeeper.server.PrepareRequestProcessor)
[2024-07-18 11:55:10,015] INFO Processed session termination for sessionid: 0x10000871d870006 (org.apache.zookeeper.server.PrepareRequestProcessor)
[2024-07-18 11:55:10,019] INFO Closed socket connection for client /127.0.0.1:53754 which had sessionid 0x10000871d870006 (org.apache.zookeeper.server.NIOServerCnxn)
[2024-07-18 11:56:34,736] INFO Accepted socket connection from /127.0.0.1:53760 (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2024-07-18 11:56:34,739] INFO Client attempting to establish new session at /127.0.0.1:53760 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-18 11:56:34,752] INFO Established session 0x10000871d870007 with negotiated timeout 30000 for client /127.0.0.1:53760 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-18 11:56:34,815] INFO Processed session termination for sessionid: 0x10000871d870007 (org.apache.zookeeper.server.PrepareRequestProcessor)
spark@spark-virtualBox:~/kafka_2.11-2.1.0
Archive Editor Ver Buscar Terminal Ayuda
score '_' ) could collide. To avoid issues it is best to use either, but not both.
Created topic "text_topic2".
spark@spark-virtualBox:~/kafka_2.11-2.1.0$ bin/kafka-topics.sh --list --zookeeper localhost:2181
__consumer_offsets
csv_topic
csv_topic2
first_topic
joa_json
joa_topic
joajson
joajson2
json_topic
new_joa_json
new_topic
prueba_topic2
prueba_topic3
text_topic
text_topic2
text_topic3
topic_app
wordcount_topic
spark@spark-virtualBox:~/kafka_2.11-2.1.0$
```

Mostrando los topics que contiene kafka:

```
spark@spark-virtualBox:~/kafka_2.11-2.1.0$ bin/kafka-topics.sh --list --zookeepe
r localhost:2181
__consumer_offsets
csv_topic
csv_topic2
first_topic
joa_json
joa_topic
joajson
joajson2
json_topic
new_joa_json
new_topic
prueba_topic2
prueba_topic3
text_topic
text_topic2
text_topic3
topic_app
wordcount_topic
```

Creamos el producer y vamos introduciendo en el topic mensajes, y vemos que se pueden ver en otra consola:

```
spark@spark-virtualBox:~/kafka_2.11-2.1.0$ kafka-console-producer.sh --broker-lis
st 127.0.0.1:9092 --topic text_topic2
>Hola mundo
>Hola mundo de nuevo
>Tercera frase de texto
>Cuarta y última

spark@spark-virtualBox:~/kafka_2.11-2.1.0$ bin/kafka-console-consumer.sh --boots
trap-server 127.0.0.1:9092 --topic text_topic2 --from-beginning
Hola mundo
Cuarta y última
Tercera frase de texto
Hola mundo de nuevo
```

Cambio las propiedades de kafka e inicio presto:

```
spark@spark-virtualBox:~/presto-server-0.213$ nano etc/catalog/kafka.properties
spark@spark-virtualBox:~/presto-server-0.213$ bin/launcher start
Started as 12707
```

las propiedades se pueden ver aqui:

```
GNU nano 2.9.3      etc/catalog/kafka.properties

connector.name=kafka
kafka.nodes=localhost:9092
kafka.table-names=text_topic2
kafka.hide-internal-columns=false
```

Lanzo el cliente de presto en otra consola:

```
spark@spark-virtualBox:~$ ./presto --catalog kafka --schema default
presto:default>
```

Vemos la tabla recién creada:

```
spark@spark-virtualBox:~$ ./presto --catalog kafka --schema default
presto:default> SHOW TABLES
      -> DESCRIBE text_topic2
presto:default> clear
      ->
presto:default> SHOW TABLES;
  Table
-----
text_topic2
(1 row)

Query 20240718_101816_00002_fkbjp, FINISHED, 1 node
Splits: 19 total, 19 done (100,00%)
0:00 [1 rows, 28B] [2 rows/s, 64B/s]

presto:default> DESCRIBE text_topic2;
  Column      | Type   | Extra | Comment
-----+-----+-----+-----
_partition_id | bigint |       | Partition Id
_partition_offset | bigint |       | Offset for the message within the partiti
_segment_start | bigint |       | Segment start offset
_segment_end   | bigint |       | Segment end offset
_segment_count | bigint |       | Running message count per segment
_message_corrupt | boolean |       | Message data is corrupt
_message       | varchar |       | Message text
_message_length | bigint |       | Total number of message bytes
```

Consulta el topic en presto y veo los mensajes:

```
presto:default> SELECT _message FROM text_topic2;
_message
-----
Hola mundo
Cuarta y última
Tercera frase de texto
Hola mundo de nuevo
(4 rows)

Query 20240718_103002_00007_fkbjp, FINISHED, 1 node
Splits: 19 total, 19 done (100,00%)
0:01 [4 rows, 67B] [5 rows/s, 98B/s]
```

Meto don nuevas frases en topic:

```
spark@spark-virtualBox:~/kafka_2.11-2.1.0$ kafka-console-producer.sh --broker-list 127.0.0.1:9092 --topic text_topic2
>Hola mundo
>Hola mundo de nuevo
>Tercera frase de texto
>Cuarta y última
>Quinta
>Sexta a ver si aparece en presto
>
```

Y vemos que aparece en presto...

```
presto:default> SELECT _message FROM text_topic2;
      _message
-----
Hola mundo
Cuarta y última
Hola mundo de nuevo
Quinta
Tercera frase de texto
Sexta a ver si aparece en presto
(6 rows)

Query 20240718_103156_00008_fkbjp, FINISHED, 1 node
Splits: 19 total, 19 done (100,00%)
0:00 [6 rows, 105B] [18 rows/s, 317B/s]
```

Creo un nuevo topic llamado csv\_topic1

```
spark@spark-virtualBox:~/kafka_2.11-2.1.0$ kafka-topics.sh --zookeeper 127.0.0.1:2181 --topic csv_topic1 --create --partitions 3 --replication-factor 1
WARNING: Due to limitations in metric names, topics with a period ('.') or under score ('_') could collide. To avoid issues it is best to use either, but not both.
Created topic "csv_topic1".
```

Edito de nuevo el fichero kafka.properties, para que pueda usar el nuevo topic creado:

```
GNU nano 2.9.3      etc/catalog/kafka.properties

connector.name=kafka
kafka.nodes=localhost:9092
kafka.table-names=text_topic2, csv_topic1
kafka.hide-internal-columns=false
```

He tenido que reiniciar presto para que asuma los cambios y me va:

```
spark@spark-virtualBox:~/presto-server-0.213$ bin/launcher stop
Stopped 12707
spark@spark-virtualBox:~/presto-server-0.213$ bin/launcher start
Started as 16820
spark@spark-virtualBox:~/presto-server-0.213$ cd $HOME
spark@spark-virtualBox:~$ ./presto --catalog kafka --schema default
presto:default> SHOW TABLES
      ;
      Table
-----
csv_topic1
text_topic2
(2 rows)

Query 20240718_105325_00001_nd86h, FINISHED, 1 node
Splits: 19 total, 19 done (100,00%)
0:02 [2 rows, 55B] [0 rows/s, 26B/s]
```

Vemos el mensaje desde presto:

```
presto:default> SELECT _message FROM csv_topic1;
      _message
-----
"JOSE ANTONIO","38","75.5","C/LIMONERO 39, 1-C"
(1 row)

Query 20240718_105839_00003_nd86h, FINISHED, 1 node
Splits: 17 total, 17 done (100,00%)
0:01 [1 rows, 47B] [1 rows/s, 56B/s]
```

Definimos la tabla dentro de etc en la carpeta kafka1, en el fichero csv\_topic1.json:

```
spark@spark-virtualBox:~/presto-server-0.213$ mkdir etc/kafka
mkdir: no se puede crear el directorio «etc/kafka»: El archivo ya existe
spark@spark-virtualBox:~/presto-server-0.213$ mkdir etc/kafka1
spark@spark-virtualBox:~/presto-server-0.213$ nano etc/kafka1/csv_topic1.json
spark@spark-virtualBox:~/presto-server-0.213$ nano etc/kafka1/csv_topic1.json
spark@spark-virtualBox:~/presto-server-0.213$ bin/launcher start
Started as 17634
```

\*\*\*\*\*PROBLEMAS AL VISUALIZAR LAS TABLAS ASI QUE AVANZO\*\*\*\*\*

//////VUELVO

Me estaba dando fallos ya que lo habia metido en una carpeta llamada kafka1, por lo que no me lo reconoció, ya que TIENEN QUE ESTAR EN LA CARPETA "/kafka"

He realizado un mv desde donde lo tenia a /kafka,

Ahora si, dentro de presto he ejecutado y me sale bien lo insertado

```
presto:default> SHOW TABLES;
  Table
-----
csv_topic1
json_topic1
text_topic2
(3 rows)

Query 20240718_122709_00003_vqs2r, FINISHED, 1 node
Splits: 19 total, 19 done (100,00%)
0:03 [3 rows, 83B] [1 rows/s, 32B/s]
```

```
presto:default> DESCRIBE csv_topic1;
  Column      | Type   | Extra | Comment
-----+-----+-----+-----
nombre        | varchar |       | 
edad          | integer |       | 
peso          | double  |       | 
direccion     | varchar |       | 
_partition_id | bigint  |       | Partition Id
```

Hago los selects que me indica la practica y compruebo que estan bien:

```
presto:default> select * from csv_topic1;
 nombre | edad | peso | direccion | _partition_id | _partition_of
-----+-----+-----+-----+-----+-----
JOSE ANTONIO | 38 | 75.5 | C/LIMONERO 39, 1-C | 0 | 
(1 row)
```

```

presto:default> SELECT nombre, edad, peso, direccion FROM csv_topic1 where edad
= 38;
  nombre   | edad | peso |   direccion
-----+-----+-----+-----
JOSE ANTONIO |   38 | 75.5 | C/LIMONERO 39, 1-C
(1 row)

Query 20240718_123109_00011_vqs2r, FINISHED, 1 node
SpA Ayuda 17 total, 17 done (100,00%)
0:00 [1 rows, 47B] [3 rows/s, 172B/s]

```

## PARTE FINAL DE LA PRACTICA

Creo json\_topic1

```

spark@spark-virtualBox:~/kafka_2.11-2.1.0$ 0$ kaftopics.sh --zookeeper 127.0.0.1
:2181 --topic json_topic1 -create --partitions 3 --replication-factor 1
WARNING: Due to limitations in metric names, topics with a period ('.') or under
score ('_') could collide. To avoid issues it is best to use either, but not bot
h.
Created topic "json_topic1".

```

Metemos linea

```

spark@spark-virtualBox:~/kafka_2.11-2.1.0$ kafka-console-producer.sh --broker-li
st 127.0.0.1:9092 --topic json_topic1
>{"nombre":"JOSE ANTONIO","edad":38,"peso":75.5,"direccion":"C/ DEL LIMONERO 39,
PISO 1-A"}

```

Configuro el config de kafka y compruebo que sale bien:

```

presto:default> show tables
-> ;
  Table
-----
csv_topic1
json_topic1
text_topic2
(3 rows)

Query 20240718_123340_00012_vqs2r, FINISHED, 1 node
Splits: 19 total, 19 done (100,00%)
0:00 [3 rows, 83B] [7 rows/s, 215B/s]

```

Creo el csv dentro json\_topic1.csv

```

spark@spark-virtualBox:~/presto-server-0.213/etc/kafka1$ nano json_topic1.json
spark@spark-virtualBox:~/presto-server-0.213/etc/kafka1$ cat json_topic1.json
jq .
{
  "tableName": "json_topic",
  "topicName": "json_topic",
  "schemaName": "default",
  "message": {
    "dataFormat": "json",
    "fields": [
      {
        "name": "nombre",
        "mapping": "nombre",
        "type": "VARCHAR"
      },
      {
        "name": "edad",
        "mapping": "edad",
        "type": "INT"
      },
      {
        "name": "peso",
        "mapping": "peso",
        "type": "DOUBLE"
      },
      {
        "name": "direccion",
        "mapping": "direccion",
        "type": "VARCHAR"
      }
    ]
  }
}

```

Antes de eso he realizado los pasos de antes, he parado presto, he creado el csv dentro de /etc/kafka y he vuelto a reiniciar presto. Ahora compruebo si lo que he hecho se muestra en la tabla:

```
presto:default> DESCRIBE json_topic1;
```

Column	Type	Extra	Comment
-----+-----+-----+-----			
_partition_id	bigint		Partition Id
_partition_offset	bigint		Offset for the message within the partiti
_segment_start	bigint		Segment start offset
_segment_end	bigint		Segment end offset
_segment_count	bigint		Running message count per segment
_message_corrupt	boolean		Message data is corrupt
_message	varchar		Message text
_message_length	bigint		Total number of message bytes
_key_corrupt	boolean		Key data is corrupt
_key	varchar		Key text
_key_length	bigint		Total number of key bytes
(11 rows)			

No se por que aqui no me sale bien, he hecho todo como procedía.