Ejercicio 7

```
KeyboardInterrupt
>>> ProductSchema = StructType([
...     StructField("product_id", IntegerType(), True),
...     StructField("product_category_id", IntegerType(), True),
...     StructField("product_name", StringType(), True),
...     StructField("product_description", StringType(), True),
...     StructField("product_price", FloatType(), True),
...     StructField("product_image", StringType(), True)
...     ])
>>> products_df = (
...     spark
...     .read
...     .format('csv')
...     .options(inferSchema=True)
...     .schema(ProductSchema)
...     .load('/public/retail_db/products/part-00000')
... )
>>> products_df.write.format('hive').saveAsTable('products')
2022-09-25 23:46:12,454 WARN metastore.ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording the schema version 1.2.0
2022-09-25 23:46:12,639 WARN metastore.ObjectStore: Failed to get database default, returning NoSuchObjectException
2022-09-25 23:46:14,205 WARN metastore.HiveMetaStore: Location: file:/home/vagrant/spark-warehouse/products specified for non-external table:products
2022-09-25 23:46:18,058 ERROR hdfs.KeyProviderCache: Could not find uri with key [dfs.encryption.key.provider.uri] to create a keyProvider !!
2022-09-25 23:46:18,170 WARN metastore.ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
AttributeError: 'SparkContext' object has no attribute 'listTables'
>>> spark.catalog.listTables()
[Table(name=u'products', database=u'default', description=None, tableType=u'MANAGED', isTemporary=False)]
>>> spark.sql('SELECT * FROM products LIMIT 10').show()
+----------+-------------------+--------------------+-------------------+-------------+--------------------+
|product_id|product_category_id|        product_name|product_description|product_price|       product_image|
+----------+-------------------+--------------------+-------------------+-------------+--------------------+
|         1|                  2|Quest Q64 10 FT. ...|               null|        59.98|http://images.acm...|
|         2|                  2|Under Armour Men'...|               null|       129.99|http://images.acm...|
|         3|                  2|Under Armour Men'...|               null|        89.99|http://images.acm...|
|         4|                  2|Under Armour Men'...|               null|        89.99|http://images.acm...|
|         5|                  2|Riddell Youth Rev...|               null|       199.99|http://images.acm...|
|         6|                  2|Jordan Men's VI R...|               null|       134.99|http://images.acm...|
|         7|                  2|Schutt Youth Recr...|               null|        99.99|http://images.acm...|
|         8|                  2|Nike Men's Vapor ...|               null|       129.99|http://images.acm...|
|         9|                  2|Nike Adult Vapor ...|               null|         50.0|http://images.acm...|
|        10|                  2|Under Armour Men'...|               null|       129.99|http://images.acm...|
+----------+-------------------+--------------------+-------------------+-------------+--------------------+
```

```
>> spark.catalog.listTables()
Table(name=u'products', database=u'default', description=None, tableType=u'MANAGED', isTemporary=False)]
>>
```

Ejericio 8

```
>>> orders_schema = StructType([
...     StructField("order_id", IntegerType(), True),
...     StructField("order_date", DateType(), True),
...     StructField("order_customer_id", IntegerType(), True),
...     StructField("order_status", StringType(), True)
...     ])
>>> orders_df = (
...     spark.read
...     .format('csv')
...     .options(inferSchema=True)
...     .schema(orders_schema)
...     .load('/public/retail_db/orders/part-00000')
... )
```

```
>>> orders_df.write.format('hive').saveAsTable('orders')
2022-09-26 04:45:09,240 WARN metastore.HiveMetaStore: Location: file:/home/vagrant/spark-warehouse/orders specified for non-external table:orders
>>> query = """ SELECT COUNT(*) AS count, DATE_FORMAT(order_date, 'YYYYMM') AS month FROM orders GROUP BY DATE_FORMAT(order_date, 'YYYYMM')"""
>>>
```

```
>>> result = spark.sql(query)
>>> result.show(5)
+-----+------+
|count| month|
+-----+------+
| 5908|201401|
| 5467|201405|
| 5335|201312|
| 5335|201310|
|  557|201412|
+-----+------+
only showing top 5 rows

>>> result.write.option('compression', 'uncompressed').format('parquet').save('/user/vagrant/lab1/pregunta8/resultado')
>>>
```

```
[vagrant@10 ~]$ hdfs ls /result.write.option('compression', 'uncompressed').format('parquet').save('/user/vagrant/lab1/pregunta8/resultado')^C
[vagrant@10 ~]$ hdfs dfs -ls /user/vagrant/lab1/pregunta8/resultado
Found 15 items
-rw-r--r--   1 vagrant supergroup          0 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/_SUCCESS
-rw-r--r--   1 vagrant supergroup        350 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00000-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        643 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00043-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00051-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00053-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00071-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00073-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00091-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00121-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00128-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00156-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00157-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00163-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00176-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
-rw-r--r--   1 vagrant supergroup        659 2022-09-26 04:48 /user/vagrant/lab1/pregunta8/resultado/part-00190-0e87f894-3d68-4bfa-96f6-1985fa6ec5a3-c000.parquet
[vagrant@10 ~]$
```