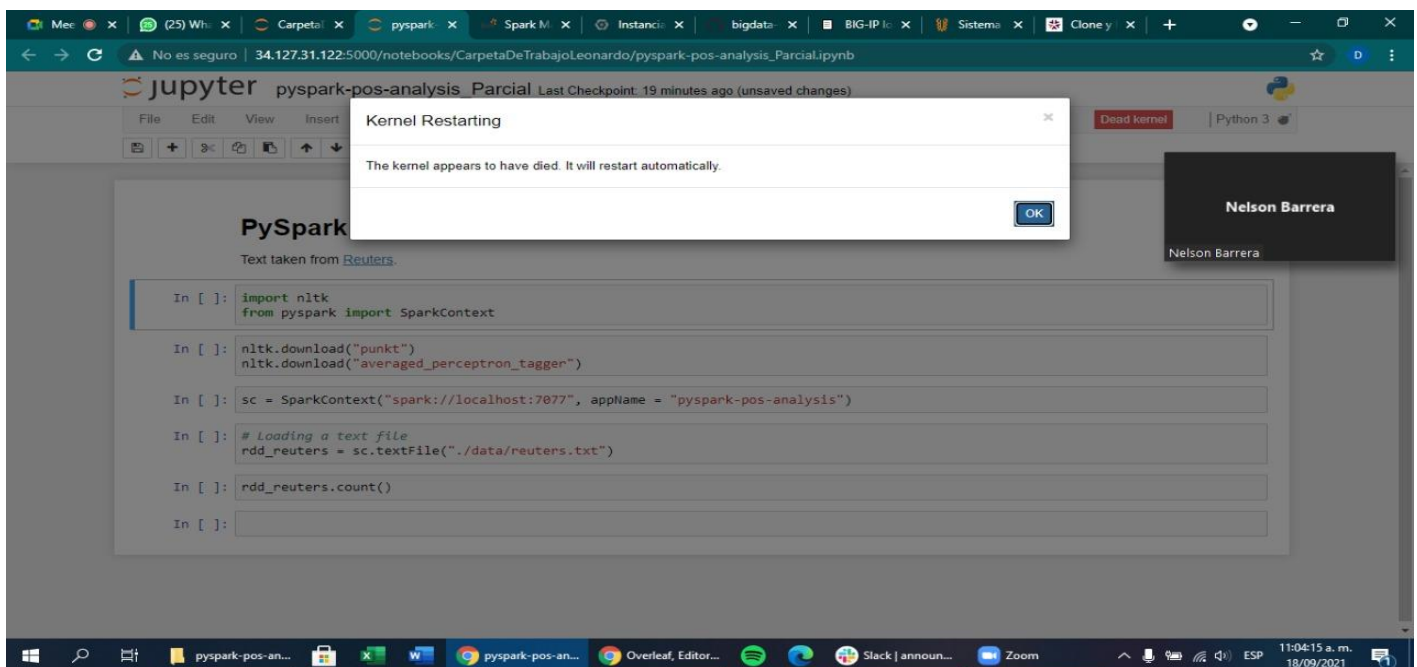
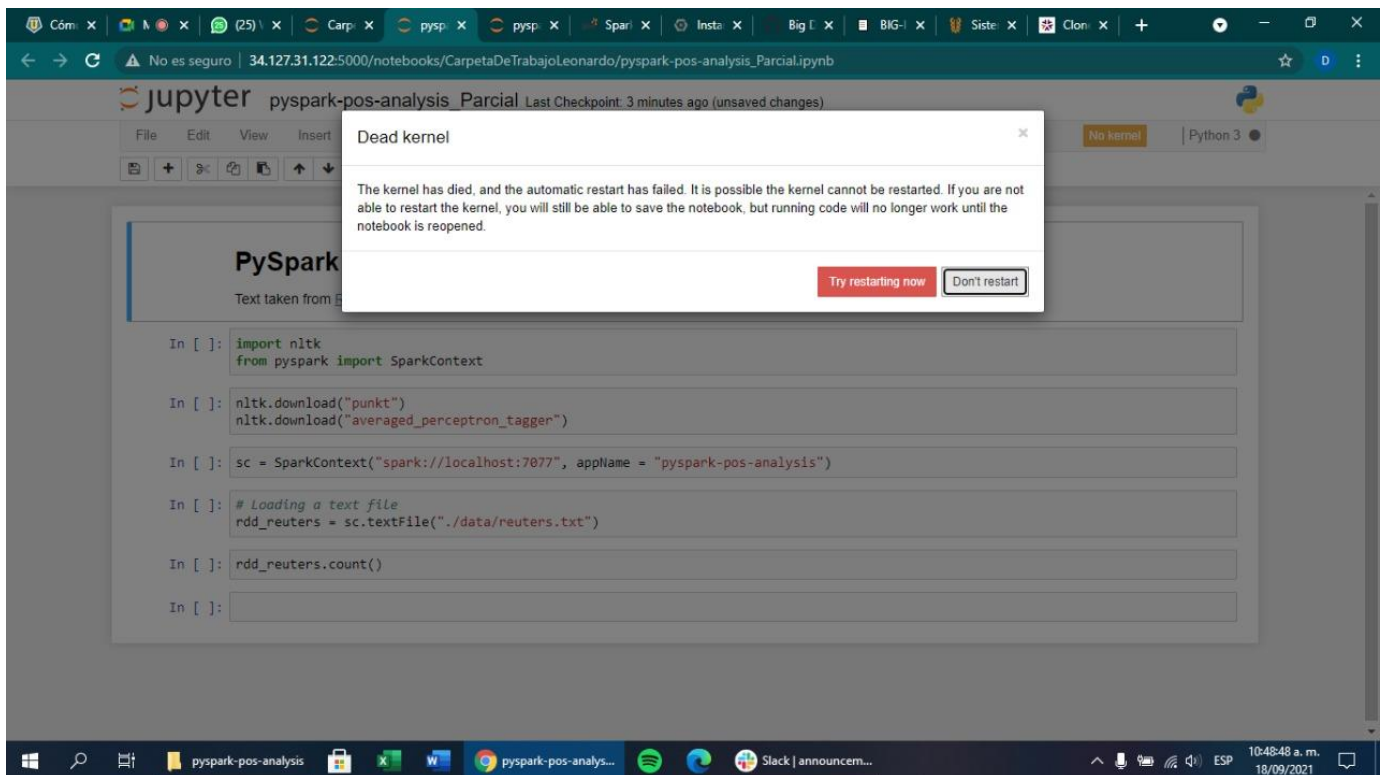




Informe Parcial

Se realizo el intento mediante la utilización de la maquina previamente configurada en el taller 1.

Pero se obtuvo el siguiente error de ejecución al momento de proceder a ejecutar el Jupyter lab



- Posterior a esto se realizó la ejecución, del entorno creado por el docente pero se obtuvo errores al momento de ejecutar `sparkContext()`.

Spark Master at spark://spark-1.us-central1-a.c.unbosque-webprogramming.internal:7077

URL: spark://spark-1.us-central1-a.c.unbosque-webprogramming.internal:7077

Alive Workers: 1
Cores in use: 2 Total, 0 Used
Memory in use: 2.8 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 1 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory	Resources
worker-20210918125620-10.128.0.9-42683	10.128.0.9:42683	ALIVE	2 (0 Used)	2.8 GiB (0.0 B Used)	

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (1)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20210918194023-0000	pyspark-pos-analysisParcialDB	2	1024.0 MiB		2021/09/18 19:40:23	fabiancpl91	KILLED	2.8 min

The screenshot displays the JupyterLab environment. On the left, the file browser shows a directory structure with files like 'anaconda3', 'ejercicio', 'nltk_data', and 'ParcialPracticoDavidBarrera'. The main area on the right is a code editor showing a Python script. The script attempts to load a text file into an RDD using `sc.textFile`. A traceback is visible, indicating a `Py4JJavaError` due to a `FileNotFoundException` in the `textFile` method of the `Context` class. The error message is: `Py4JJavaError: Traceback (most recent call last): <ipython-input-42-c4e1e857a73f> in <module> 1 # Loading a text file 2 rdd_reuters = sc.textFile("./data/reuters.txt") ~/anaconda3/lib/python3.8/site-packages/pyspark/context.py in textFile(self, name, minPartitions, use_unicode) 644 ['Hello world!'] 645 646 minPartitions = minPartitions or min(self.defaultParallelism, 2) 647 return RDD(self._jsc.textFile(name, minPartitions), self, 648 UTF8Deserializer(use_unicode)) ~/anaconda3/lib/python3.8/site-packages/pyspark/context.py in defaultParallelism(self) 440 reduce tasks) 441 442 return self._jsc.sc().defaultParallelism() 443 444 @property ~/anaconda3/lib/python3.8/site-packages/py4j/java_gateway.py in __call__(self, *args) 1302 1303 answer = self.gateway_client.send_command(command) 1304 return_value = get_return_value(1305 answer, self.gateway_client, self.target_id, self.name) 1306 ~/anaconda3/lib/python3.8/site-packages/py4j/protocol.py in get_return_value(answer, gateway_client, target_id, name) 1307 1308 raise Py4JJavaError(An exception occurred. The error was: java.io.FileNotFoundException: /data/reuters.txt (No such file or directory))`



The screenshot shows a JupyterLab window with a file explorer on the left and a code editor on the right. The file explorer shows a directory named 'ParcialPracticoDavidBarrera' containing a 'data' folder, a 'pyspark-pos-analysisParcialDB1x.ipynb' file, and a 'README.md' file. The code editor shows the following code:

```
[32]: True

[43]: from pyspark import SparkContext
sc = SparkContext("spark://localhost:7077", appName = "pyspark-pos-analysisParcialDB1x")

-----
ValueError: Traceback (most recent call last)
<ipython-input-43-78ae1d4ee76a> in <module>
      1 from pyspark import SparkContext
----> 2 sc = SparkContext("spark://localhost:7077", appName = "pyspark-pos-analysisParcialDB1x")

~/anaconda3/lib/python3.8/site-packages/pyspark/context.py in __init__(self, master, appName, sparkHome, pyFiles, environment, batchSize, serializer, conf, gateway, jsc, profiler_cls)
    142     " is not allowed as it is a security risk.")
    143
--> 144     SparkContext._ensure_initialized(self, gateway=gateway, conf=conf)
    145     try:
    146         self._do_init(master, appName, sparkHome, pyFiles, environment, batchSize, serializer,
    147                        conf)
    148
~/anaconda3/lib/python3.8/site-packages/pyspark/context.py in _ensure_initialized(cls, instance, gateway, conf)
    340
    341     # Raise error if there is already a running Spark context
--> 342     raise ValueError(
    343         "Cannot run multiple SparkContexts at once; "
    344         "existing SparkContext(app=%s, master=%s)"

ValueError: Cannot run multiple SparkContexts at once; existing SparkContext(app=pyspark-pos-analysisParcialDB1x, master=spark://spark-1.us-central1-a.c.unbosque-webprogramming.internal:7077) created by __init__ at <ipython-input-33-5524e81a3d55>:2
```

The error message indicates that a SparkContext is already running with the same application name and master URL, preventing the creation of a new one. The bottom status bar shows 'Saving completed', 'Mode: Command', 'Ln 1, Col 1', and the file path 'pyspark-pos-analysisParcialDB1x.ipynb'.

- Error que cabe resaltar se intentó solucionar de miles de manera, pero obtuve como resultado que NO me corriera de manera exitosa.