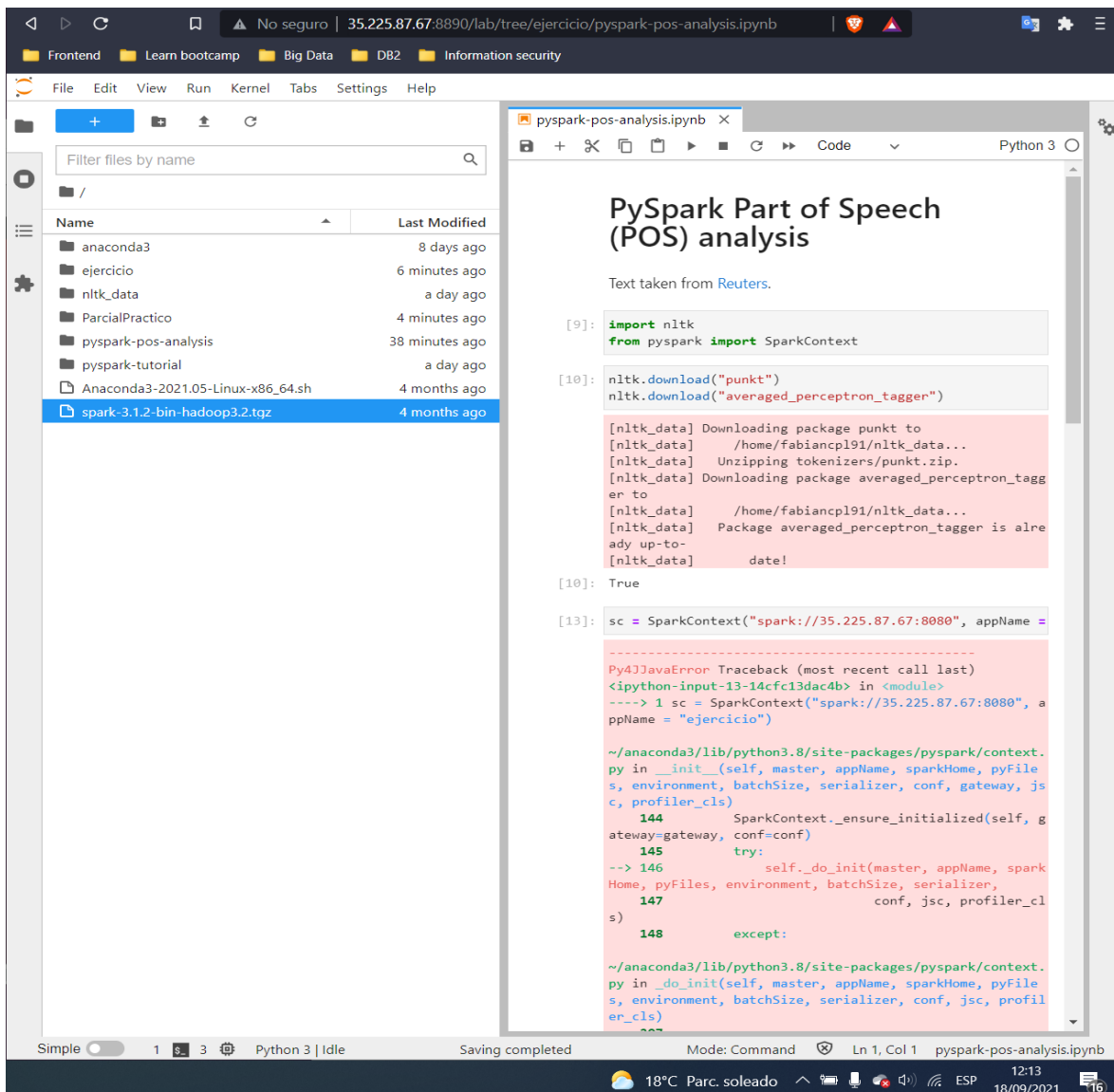


Parcial Big data

David Esteban Zamora Montero

1. El parcial se realizó en el cluster que dispuso el profesor sin embargo tras varios intentos no se logra obtener una respuesta favorable por ende adjunto capturas de el intento de conexión usando jupyter para su ejecución en la carpeta “ejercicio”.



The screenshot shows a Jupyter Notebook titled "PySpark Part of Speech (POS) analysis". The notebook is running in a browser at the URL `35.225.87.67:8890/lab/tree/ejercicio/pyspark-pos-analysis.ipynb`. The left sidebar shows a file explorer with a list of files and folders, including `anaconda3`, `ejercicio`, `nlk_data`, `ParcialPractico`, `pyspark-pos-analysis`, `pyspark-tutorial`, `Anaconda3-2021.05-Linux-x86_64.sh`, and `spark-3.1.2-bin-hadoop3.2.tgz`. The main area displays the notebook content, which includes the following code cells:

```
[9]: import nltk
from pyspark import SparkContext

[10]: nltk.download("punkt")
nltk.download("averaged_perceptron_tagger")

[nltk_data] Downloading package punkt to
[nltk_data] /home/fabianclp191/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] /home/fabianclp191/nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-date!

[10]: True

[13]: sc = SparkContext("spark://35.225.87.67:8080", appName = "ejercicio")

Py4JJavaError Traceback (most recent call last)
<ipython-input-13-14cfc13dac4b> in <module>
----> 1 sc = SparkContext("spark://35.225.87.67:8080", appName = "ejercicio")

~/anaconda3/lib/python3.8/site-packages/pyspark/context.py in __init__(self, master, appName, sparkHome, pyFile, s, environment, batchSize, serializer, conf, gateway, jsc, profiler_cls)
    144 SparkContext._ensure_initialized(self, gateway=gateway, conf=conf)
    145 try:
--> 146     self._do_init(master, appName, sparkHome, pyFiles, environment, batchSize, serializer, conf, jsc, profiler_cls)
    147 except:
    148     except:

~/anaconda3/lib/python3.8/site-packages/pyspark/context.py in _do_init(self, master, appName, sparkHome, pyFile, s, environment, batchSize, serializer, conf, jsc, profiler_cls)
```

The bottom status bar indicates the notebook is in "Simple" mode, with 1 cell selected, and the status is "Saving completed". The mode is set to "Command", and the cursor is at line 1, column 1. The bottom right corner shows the system clock as 12:13 on 18/09/2021.