



escola
britânica de
artes criativas
& tecnologia

Módulo | Computação em Nuvem III

Caderno de **Exercícios**

Professor [André Perez](#)

Tópicos

1. Introdução;
 2. Apache Spark;
 3. Data Wrangling com Spark.
-

Exercícios

1. Apache Spark

Replique as atividades do item 2.1 e 2.2 para instalar e configurar um cluster Apache Spark na máquina virtual do Google Colab.

2. Data Wrangling

A base de dados presente neste [link](#) contem dados macroeconômicos sobre o Reino Unido desde o século 13.

2.1. Data

Faça o download dos dados utilizando a máquina virtual do Google Colab com o código abaixo.

In []:

```
!wget -q  
"https://raw.githubusercontent.com/cluster-apps-on-docker"  
"/spark-standalone-cluster-on-docker/master/build/workspace"  
"/data/uk-macroeconomic-data.csv"  
-O "uk-macroeconomic-data.csv"
```

2.2. Wrangling

Processe os dados para que a base de dados final apresente os valores da taxa de desemprego (`Unemployment rate`) e população (`Population (GB+NI)`) estejam ordenados por ano decrescente:

```
csv
year,population,unemployment_rate
...,...,...
```

Para isso, utilize:

- Pandas

In []:

```
...
```

- PySpark

In []:

```
...
```