# Bigdata

Para clonar o repositório fazer:

git clone https://github.com/pesobreiro/bigdata.git

Software recomendado: \* Visual Studio Code https://code.visualstudio.com/ \* git https://git-scm.com/ \* Pandoc https://pandoc.org/

## Comandos principais

Atualizar um repositório local (descarrega do git par ao computador) : git pull

Enviar alterações para o github:

1. Adicionar um ficheiro git add nomeFicheiro ou adicionar todos os ficheiros alterados git add .
2. Colocar na zona de “preparação” do computador git commit -m "nome ou descricao"
3. Colocar as alterações no git git push antes era git push origin master

# Instalar spark

Instalar o Java Runtime Environment  
testar com cmd  
java --version  
  
Se n├úo tiver instalado correr:  
conda install -c conda-forge openjdk

Vantagens de instalar o Java com o anaconda: Environment Management: Installing Java through conda keeps it nicely contained within a conda environment. This helps prevent conflicts if you need different Java versions for different projects.

1. conda create –name bigdate
2. conda activate bigdata
3. conda install -c conda-forge pyspark
4. conda install -c conda-forge findspark
5. conda install -c conda-forge jupyterlab

Uma alternativa ao anaconda mais r├ípido:

conda install mamba -n base -c conda-forge

Depois funciona tudo normalmente como com o conda. Se quiserem instalar o mamba no ambiente onde est├úo a trabalhar basta fazer:

conda install mamba

Se quiserem utilizar o code convem instalar o ipykernel

mamba install -c conda-forge ipykernel  
  
# se n├úo funcionar  
  
conda install -c conda-forge ipykernel