

# Guia de Configuração do PySpark no VS Code com Anaconda

Este guia apresenta as etapas para configurar o PySpark no VS Code, utilizando o Anaconda como gerenciador de ambientes. Certifique-se de que o VS Code e o Anaconda estão instalados antes de começar.

## Preparando o terminal no VS Code

Para garantir o funcionamento correto dos comandos, utilize o terminal CMD (Prompt de Comando):

**Atalho:** Ctrl + Shift + P → Terminal: Selecionar Perfil Padrão → Command Prompt

## Instalação e configuração

### 1. Criar um ambiente virtual

```
conda create -n ambiente_pyspark python=3.11 -y
```

### 2. Ativar o ambiente

```
conda activate ambiente_pyspark
```

### 3. Confirmar o ambiente ativo

```
conda env list
```

### 4. Instalar os pacotes necessários

```
conda install -c conda-forge openjdk=17 pyspark pandas -y
```

### 5. Instalar o Findspark

```
pip install findspark
```

### 6. Validar a instalação

```
python
import findspark
findspark.init()
from pyspark.sql import SparkSession
spark = SparkSession.builder.getOrCreate()
print(spark.version)
exit()
```

### 7. Selecionar o Interpretador e Kernel no VS Code

Ctrl + Shift + P → Python: Selecionar Interpretador → ambiente\_pyspark. Com o arquivo .ipynb aberto, selecione o Kernel no canto superior direito.

Durante a execução, pode aparecer um aviso sobre winutils.exe. Esse alerta é normal e não impede o funcionamento do Spark.

## **Removendo o ambiente**

Para excluir o ambiente criado:

1. Desativar o ambiente: `conda deactivate`
2. Remover o ambiente: `conda env remove --name ambiente_pyspark -y`
3. Verificar se foi removido: `conda env list`

Seu ambiente PySpark está configurado e pronto para uso no VS Cod.

