

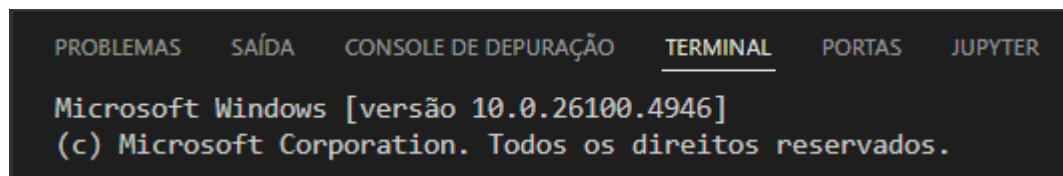
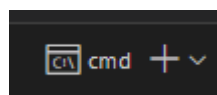
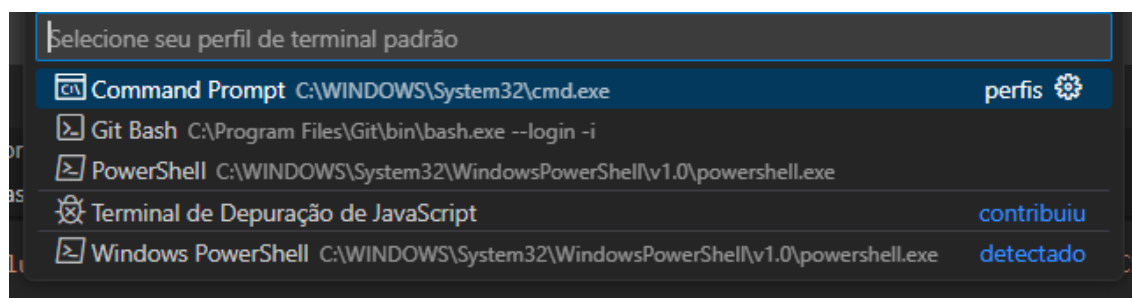
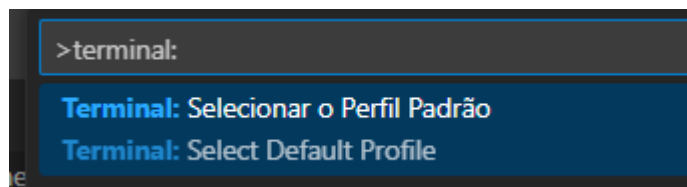
## Roteiro de instalação e configuração PySpark para VS Code e Anaconda

O roteiro a seguir deve ser feito com o **Anaconda** e **VS Code** instalado e utilizar o **Terminal CMD**

### Preparando o terminal do VS Code

Para correto funcionamento executar os comandos utilizando terminal **CMD** onde é possível deixar como padrão da seguinte forma:

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



### Instalação e Configuração

1) Criar o ambiente para evitar conflitos

**conda create -n ambiente\_fase3 python=3.11 -y**

**conda create** é o comando para criar o ambiente

**ambiente\_fase3** é o nome do ambiente e pode ser o nome que quiser

**python=3.11** é a versão do Python que vamos utilizar nesse ambiente

2) Selecionar o ambiente

**conda activate ambiente\_fase3**

3) Verificar o ambiente selecionado

**conda env list**

Onde estiver um \* é o ambiente selecionado

4) Instalar os pacotes Java e PySpark

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

5) Instalar Findspark

```
pip install findspark
```

6) Validar instalação

No terminal e com o ambiente criado selecionado (**conda activate ambiente\_fase3**) digitar **python**, pressione a tecla Enter e com o interpretador ativo (>>>) digite os seguintes comandos:

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

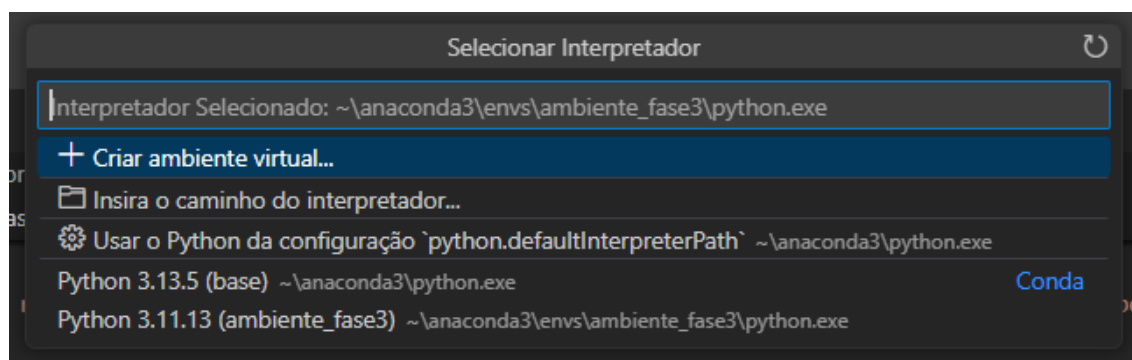
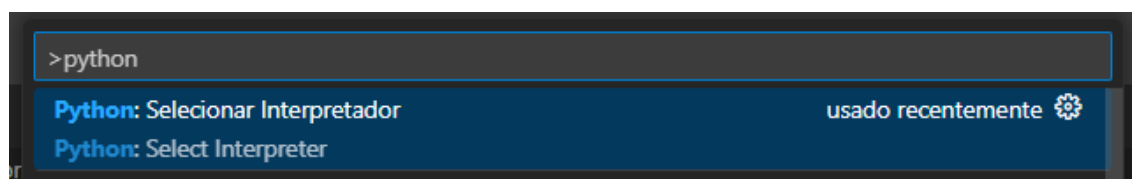
```
Python 3.11.13 | packaged by Anaconda, Inc. | (main, Jun 5 2025, 13:03:15) [MSC v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> import findspark
>>> findspark.init()
>>> from pyspark.sql import SparkSession
>>> spark = SparkSession.builder.getOrCreate()
WARNING: Using incubator modules: jdk.incubator.vector
25/08/13 19:49:37 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: java.io.FileNotFoundException: HADOOP_HOME and hadoop.h
bloms
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/08/13 19:49:39 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
>>> print(spark.version)
4.0.0
```

**Aviso Importante:** Durante a validação, é normal aparecer um aviso amarelo (WARN) sobre winutils.exe. Apenas ignore este aviso pois não impede o funcionamento do Spark

7) Selecionar o Interpretador e o Kernel

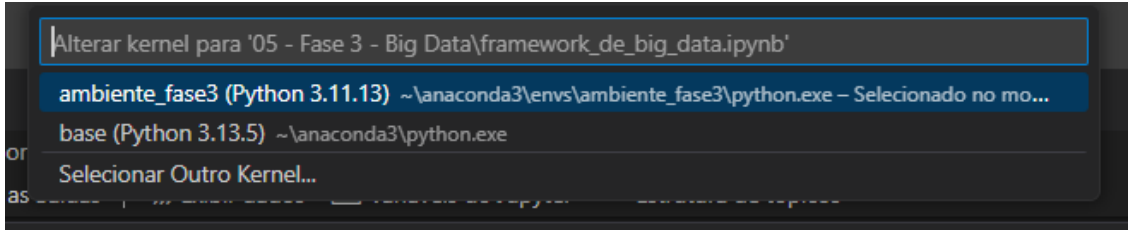
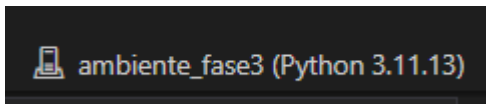
Selecionar o Interpretador através do caminho

**Ctrl+Shift+P → Python: Selecionar Interpretador → ambiente\_fase3**



E selecionar o Kernel com o arquivo **.ipynb** aberto

**Clica no canto superior direito → Selecionar o ambiente (ambiente\_fase3)**



**Remover o ambiente:**

1) Desativar o terminal

**conda deactivate**

2) Remover o ambiente

**conda env remove --name ambiente\_fase3 -y**

3) Verificar se foi removido

**conda env list**