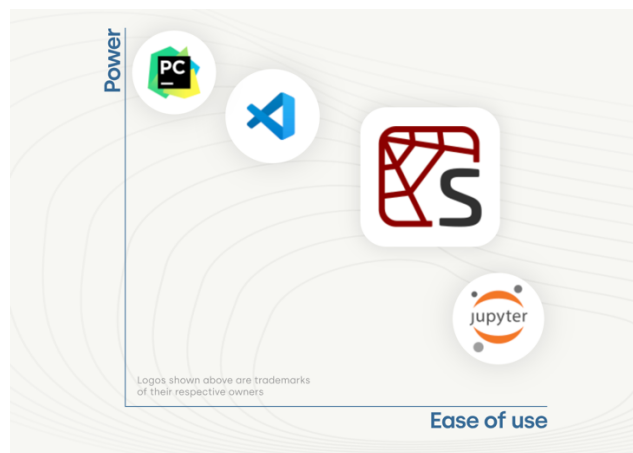


CODECON

Featurize&JupyterNote 服务器租借创建虚拟环境与使用指南

Featurize 是一个服务器租借平台，以较为便宜的价格、提供服务器租借、云存储服务。当上一个用户停止租借服务器时，所有存在本地的文档，包括环境配置都会被全部删除（除非保存在只有你个人可以访问的 work 文件夹中）。因此，租借一个 Featurize 服务器相当于开启了一个新的环境。

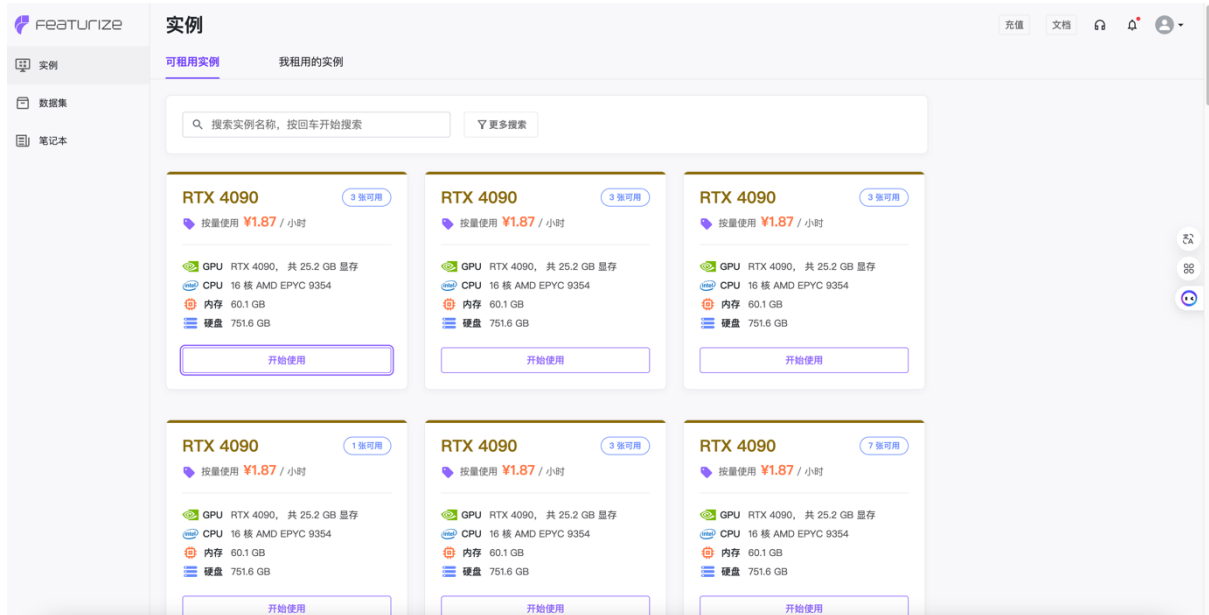
JupyterNote 是一个可交互的集成开发环境（IDE）。可以逐行输入代码后，分代码块显示结果。基于 JupyterNote 的课 JupyterLab 为很多服务器提供了便于使用的操作界面。



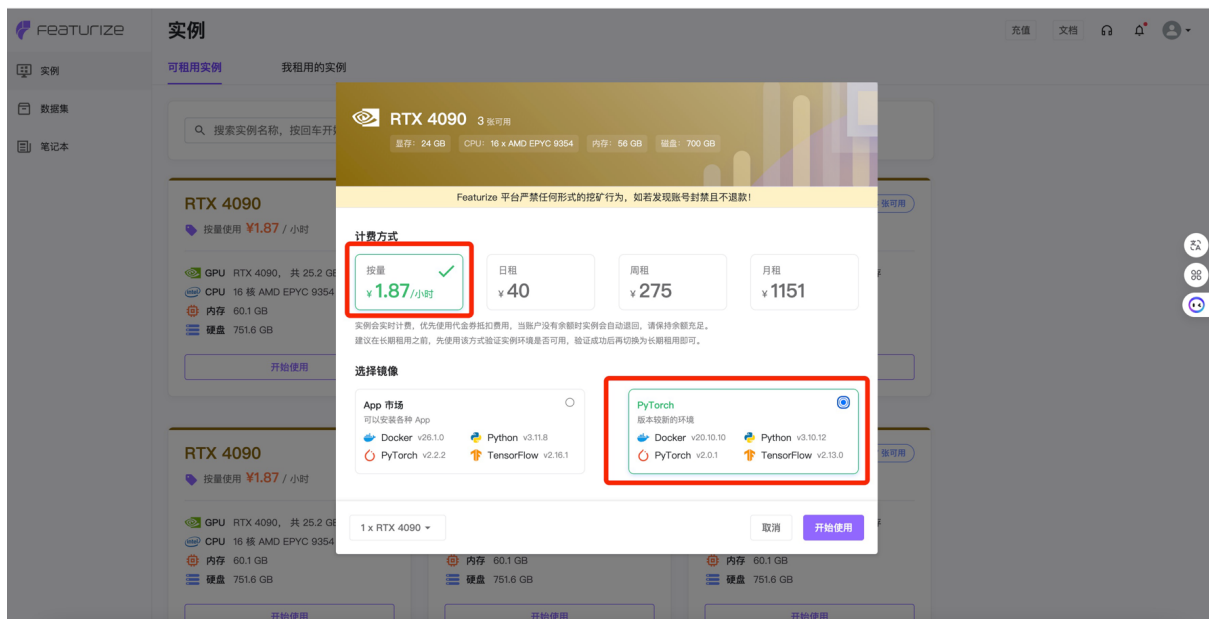
常用 IDE 介绍

本教程仅主要聚焦环境配置部分。Featurize（包括云储存功能等）与 JupyterNote 的具体使用方法网络上有丰富的教学资源，请自行检索。

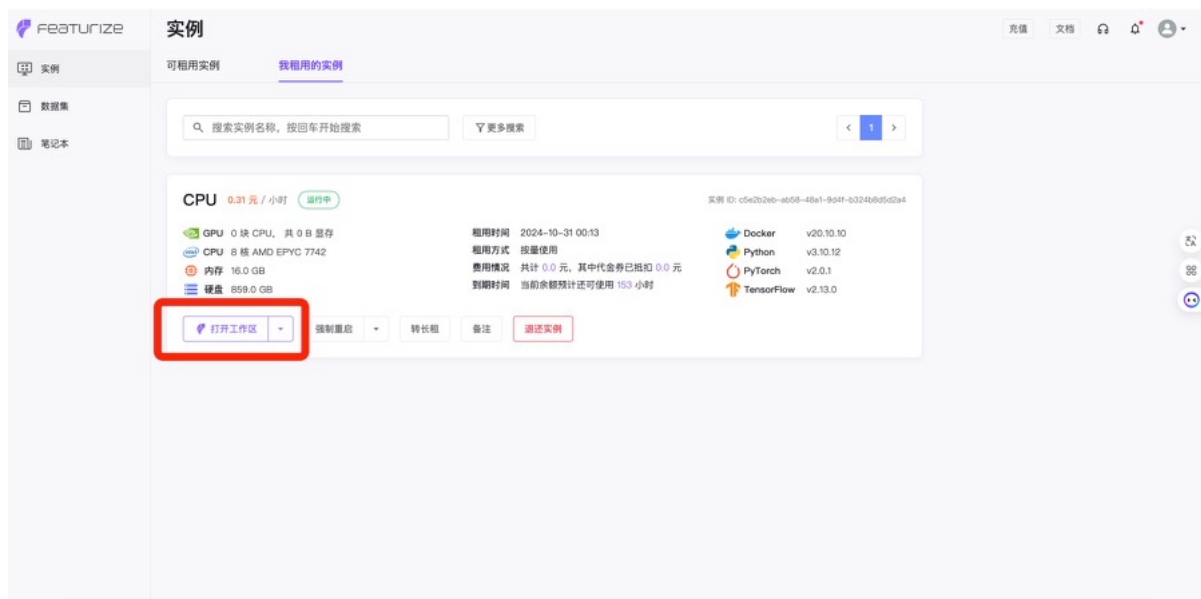
1. 搜索 Featurize 网站(<https://featurize.cn/vm/available>)并注册账号,充几块钱(新注册会送代金券)。
2. 点击左侧实例，选择可租用实例，集合自己的需要选择含有 GPU 芯片的服务器（基本上序号越高越贵，训练的速度也越快）。本教学说明以租借 4090 为例。



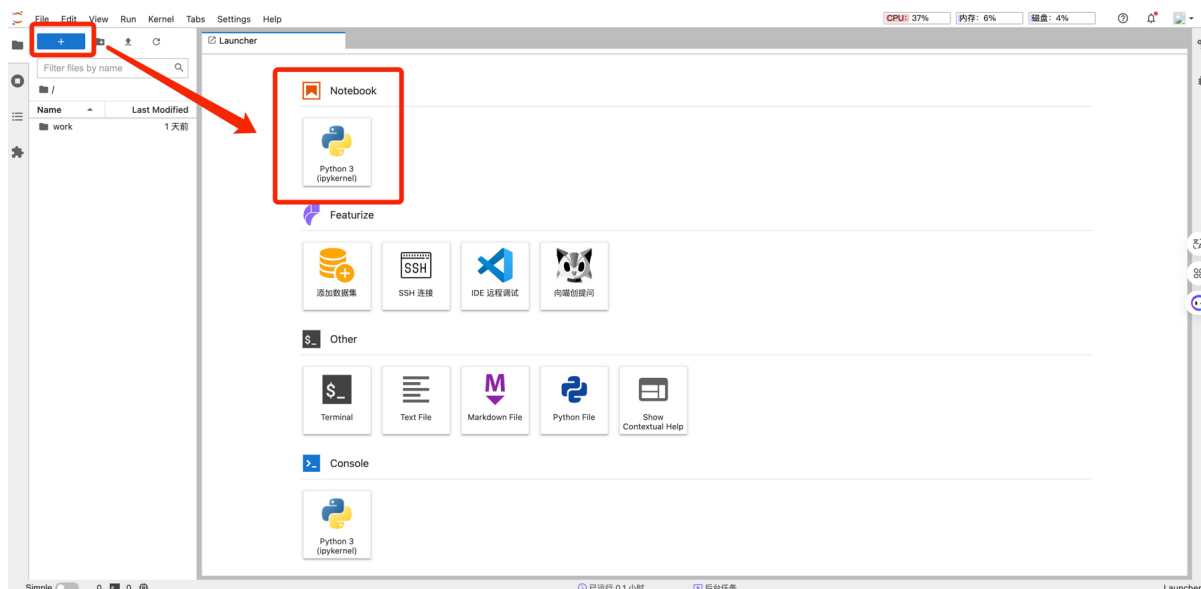
3. 点击按小时租借 RTX 4090 一枚(1.87/小时)，选择选择镜像 ('PyTorch' 版本较新的环境)，点击开始使用



4. 稍等片刻，服务器租借完成后可以在我租用的实例中找到，点击打开工作区



5. 此时进入的是 JupyterLab 界面。点击左上角蓝色按钮，点击菜单页`Jupyter Notebook`后新建 JupyterNote。自动进入 Jupyternote 界面，且左侧会出现对应 JupyterNote 文件。

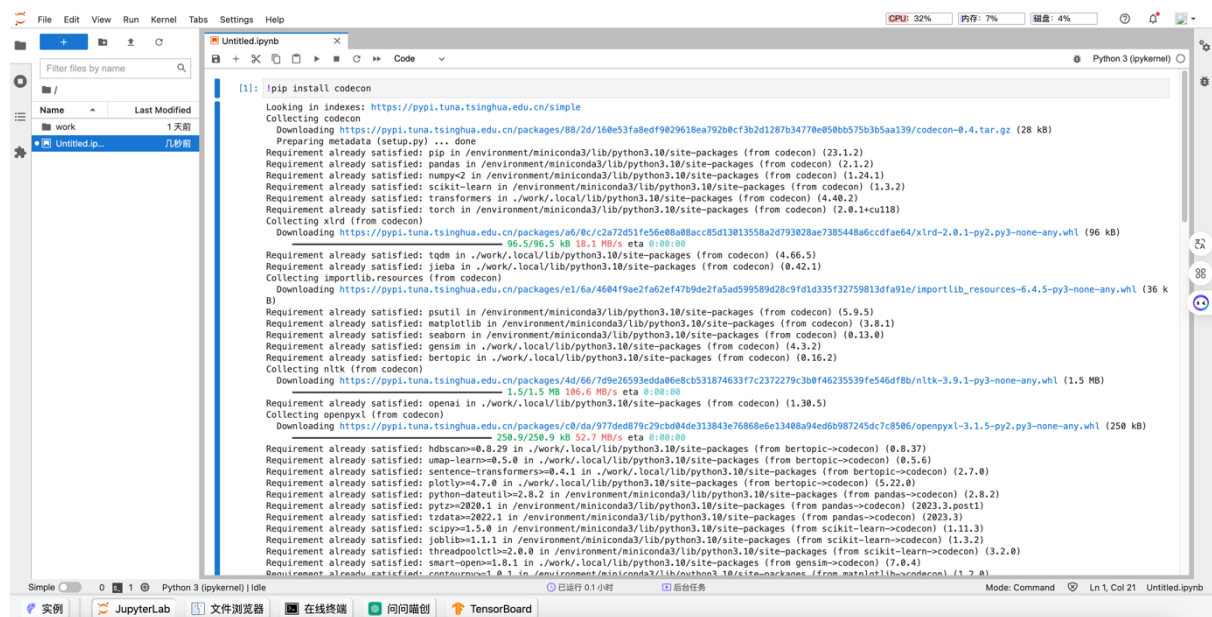


JupyterLab 左侧是文件列表，小文件(<200M)可直接拖拽到空白处自动上传，右键该文件可选择压缩或者解压缩。大文件建议先上传到数据集后（支持断点上传）再从服务器终端下载到你的服务器上

6. - 在第一行敲下(注意加感叹号)

!pip install codecon

安装成功后会在结尾显示 **Successfully installed codecon-0.4 et-xmlfile-2.0.0 importlib.resources-6.4.5 nltk-3.9.1 openpyxl-3.1.5 xlrd-2.0.1**。之后可继续使用参考说明文档使用 codecon。



```
[1]: !pip install codecon
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting codecon
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/88/2d/160e53fa8edf9029618ea792b0cf3b2d1287b34770e058b575b3b5aa139/codecon-0.4.tar.gz (28 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: pip in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (23.1.2)
Requirement already satisfied: pandas in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (2.1.2)
Requirement already satisfied: numpy<2 in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (1.24.1)
Requirement already satisfied: scikit-learn in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (1.3.2)
Requirement already satisfied: transformers in /work/.local/lib/python3.10/site-packages (from codecon) (4.40.2)
Requirement already satisfied: torch in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (2.0.1+cu118)
Collecting nltk (from codecon)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/a6/8c/c2a72d51fe56e08a08acc85d13813558a2d793028ae7385448a6ccdfae64/xlrd-2.0.1-py3-none-any.whl (96 kB)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/a6/8c/c2a72d51fe56e08a08acc85d13813558a2d793028ae7385448a6ccdfae64/xlrd-2.0.1-py3-none-any.whl (96 kB)
Requirement already satisfied: tqdm in /work/.local/lib/python3.10/site-packages (from codecon) (4.66.5)
Requirement already satisfied: jieba in /work/.local/lib/python3.10/site-packages (from codecon) (0.42.1)
Collecting importlib.resources (from codecon)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/e1/6a/4604f9ae2fa62ef47b9de2fa5ad59589d28c9fd1d335f32759813dfa91e/importlib_resources-6.4.5-py3-none-any.whl (36 kB)
Requirement already satisfied: psutil in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (5.9.5)
Requirement already satisfied: matplotlib in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (3.8.1)
Requirement already satisfied: seaborn in /environment/miniconda3/lib/python3.10/site-packages (from codecon) (0.13.0)
Requirement already satisfied: gensim in /work/.local/lib/python3.10/site-packages (from codecon) (4.3.2)
Requirement already satisfied: bertopic in /work/.local/lib/python3.10/site-packages (from codecon) (0.16.2)
Collecting nltk (from codecon)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/4d/66/7d9e26593edda06e8cb531874633f7c2372729c3b0f46235539fe546df8b/nltk-3.9.1-py3-none-any.whl (1.5 MB)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/4d/66/7d9e26593edda06e8cb531874633f7c2372729c3b0f46235539fe546df8b/nltk-3.9.1-py3-none-any.whl (1.5 MB)
Requirement already satisfied: openai in /work/.local/lib/python3.10/site-packages (from codecon) (1.30.5)
Collecting openpyxl (from codecon)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/c8/da/977de0879c29c0b04de313843e76868e6e13488a94ed6b987245dc7c8586/openpyxl-3.1.5-py3-none-any.whl (250 kB)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/c8/da/977de0879c29c0b04de313843e76868e6e13488a94ed6b987245dc7c8586/openpyxl-3.1.5-py3-none-any.whl (250 kB)
Requirement already satisfied: hdbscan==0.8.29 in /work/.local/lib/python3.10/site-packages (from bertopic->codecon) (0.8.37)
Requirement already satisfied: umap-learn==0.5.0 in /work/.local/lib/python3.10/site-packages (from bertopic->codecon) (0.5.6)
Requirement already satisfied: sentence-transformers==0.4.1 in /work/.local/lib/python3.10/site-packages (from bertopic->codecon) (2.7.0)
Requirement already satisfied: plotly==4.7.0 in /work/.local/lib/python3.10/site-packages (from bertopic->codecon) (5.22.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /environment/miniconda3/lib/python3.10/site-packages (from pandas->codecon) (2.8.2)
Requirement already satisfied: pytz==2020.1 in /environment/miniconda3/lib/python3.10/site-packages (from pandas->codecon) (2023.3.post1)
Requirement already satisfied: tzdata==2022.1 in /environment/miniconda3/lib/python3.10/site-packages (from pandas->codecon) (2023.3)
Requirement already satisfied: scipy==1.5.0 in /environment/miniconda3/lib/python3.10/site-packages (from scikit-learn->codecon) (1.11.3)
Requirement already satisfied: joblib==1.1.1 in /environment/miniconda3/lib/python3.10/site-packages (from scikit-learn->codecon) (1.3.2)
Requirement already satisfied: threadpoolctl==2.0.0 in /environment/miniconda3/lib/python3.10/site-packages (from scikit-learn->codecon) (3.2.0)
Requirement already satisfied: smart-open==1.8.1 in /work/.local/lib/python3.10/site-packages (from gensim->codecon) (7.0.4)
Requirement already satisfied: contourpy==1.0.1 in /environment/miniconda3/lib/python3.10/site-packages (from matplotlib->codecon) (1.2.0)
```

注意，在 JupyterNote 中训练模型虽然方便，但是在关闭浏览器页面，或者网络不稳定时，很容易无法持续跟踪训练进度，当然，在任务完成后结果仍会正常保存。建议自主学习如何：

- 在服务器终端创建 tmux 或 screen
- 运行 python
- 任意时刻返回 tmux 或 screen 查看 python 代码运行进度

再次提醒，当你停止租借服务器时，所有存在本地的文档都会被删除（除非存在 work 文件夹中）。