# 一、mo平台训练离线任务说明文档

## 1.环境与资源

### 1.1 平台环境

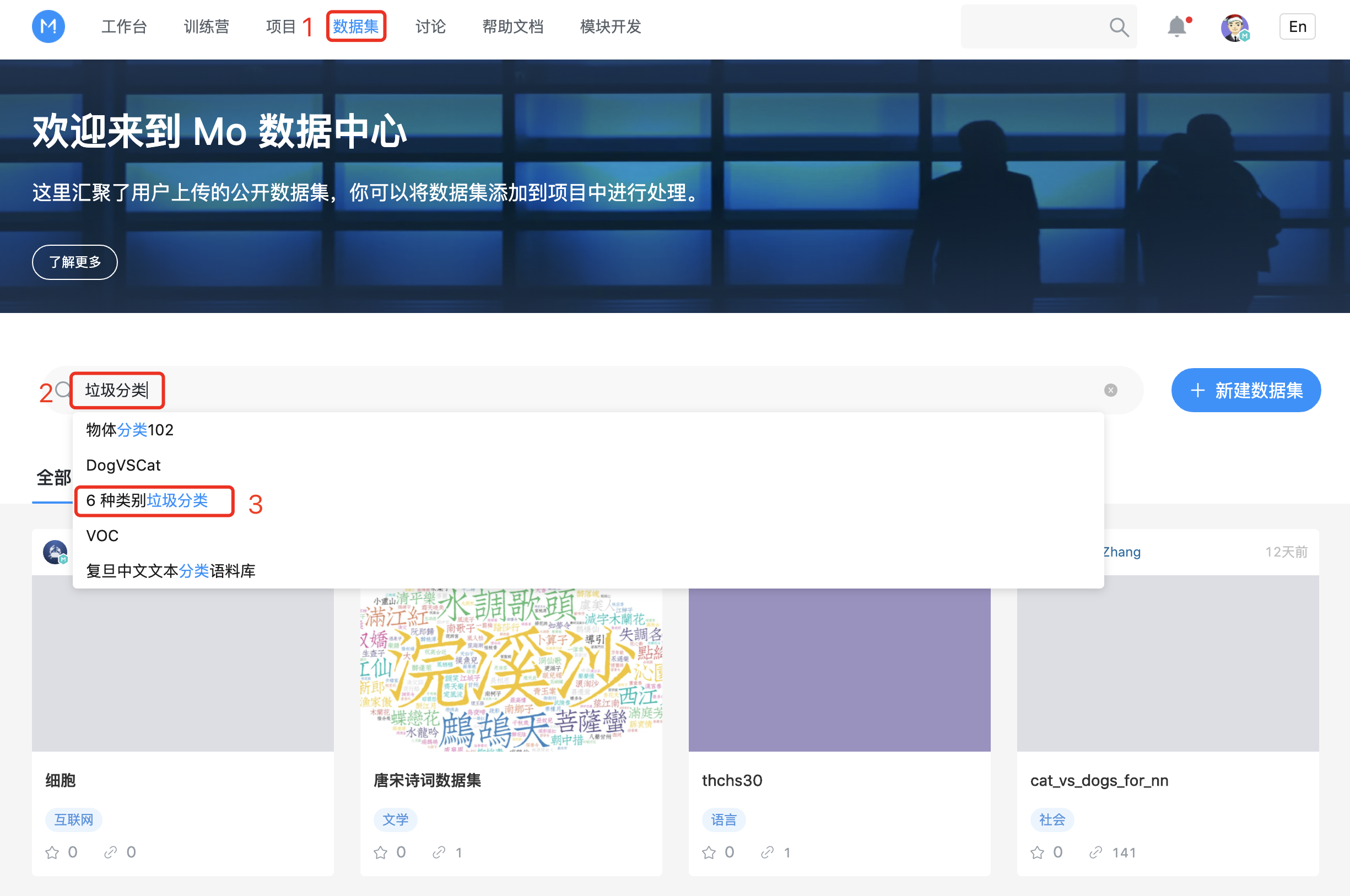
平台提供的是 **Python 3.6.5** 的开发环境，暂时不支持 Python 其它版本更换。

### 1.2 训练模型框架及其版本号

平台支持 **TensorFlow=1.13.1**、**Keras=2.3.1**、**Pytorch=1.4.0+cpu** 三种框架，请尽量使用这三种框架完成作业；如果使用其它框架，请联系我们。

### 1.3 数据集搜索与下载

* 全局搜索框搜索【6种类别垃圾分类】，点击后进入垃圾分类数据集页面



* 在数据集详情页点击【下载】按钮下载数据集到本地

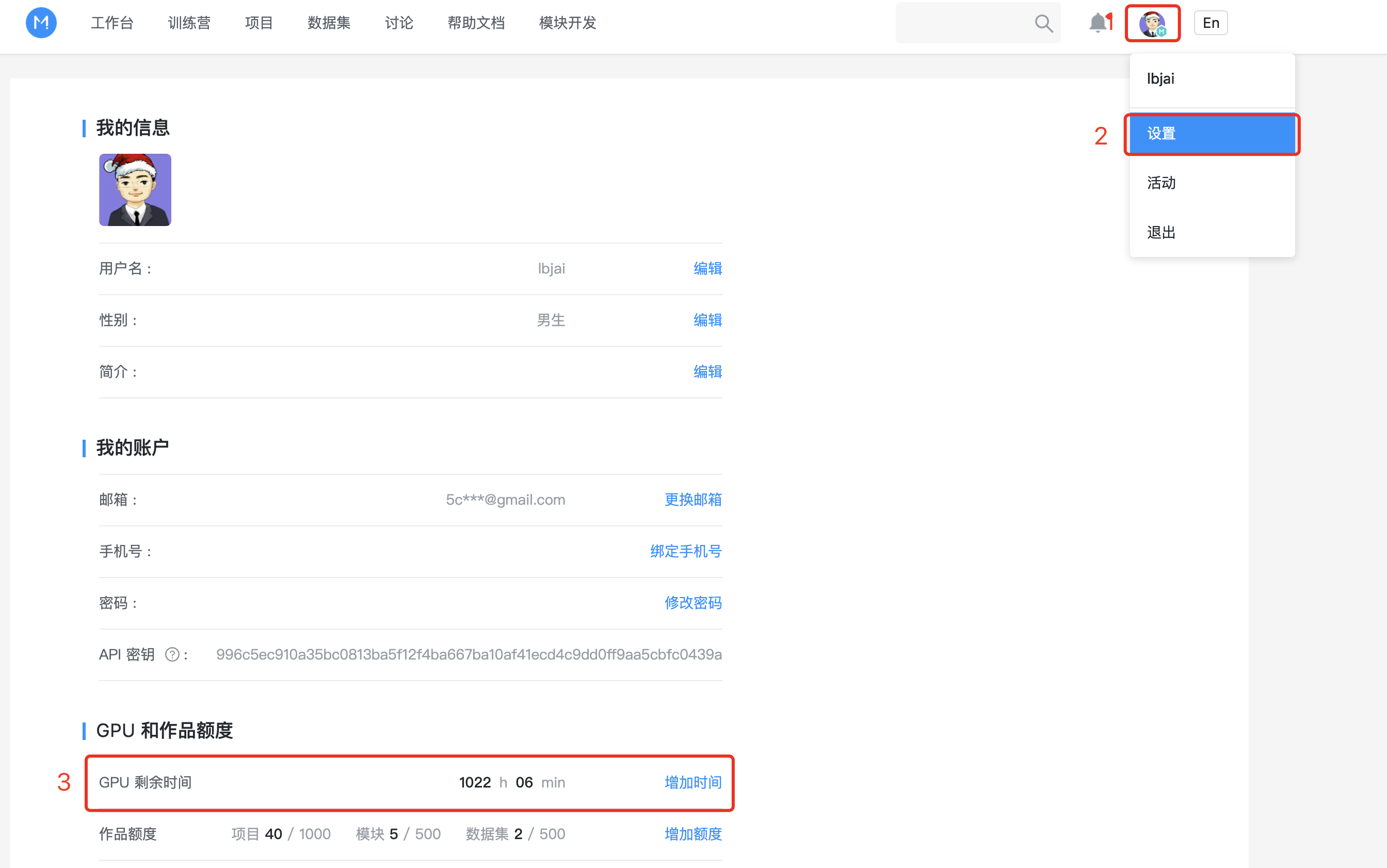


### 1.4 查看 GPU 剩余时长

* 在工作台查看



* 点击头像 --> 设置 --> GPU和作品额度 --> 查看 GPU 时长



* 在创建任务界面查看剩余时长



## 2.创建 job 流程与常见问题

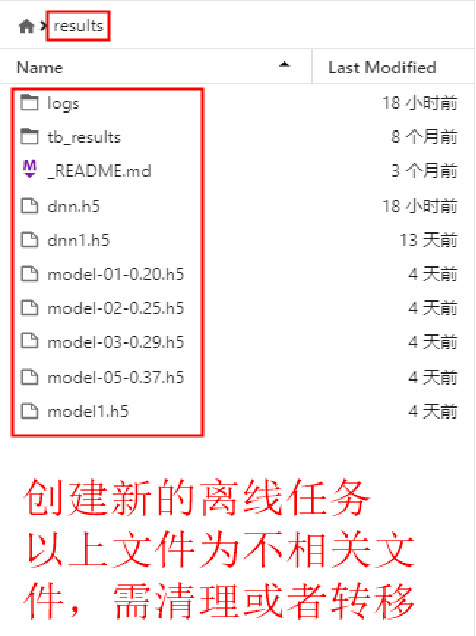
### 2.1 创建离线任务（job）流程与注意事项

该部分可以参考[在 GPU 或 CPU 资源上训练机器学习模](user_cancel)型：<https://momodel.cn/docs/#/zh-cn/%E5%9C%A8GPU%E6%88%96CPU%E8%B5%84%E6%BA%90%E4%B8%8A%E8%AE%AD%E7%BB%83%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%E6%A8%A1%E5%9E%8B>

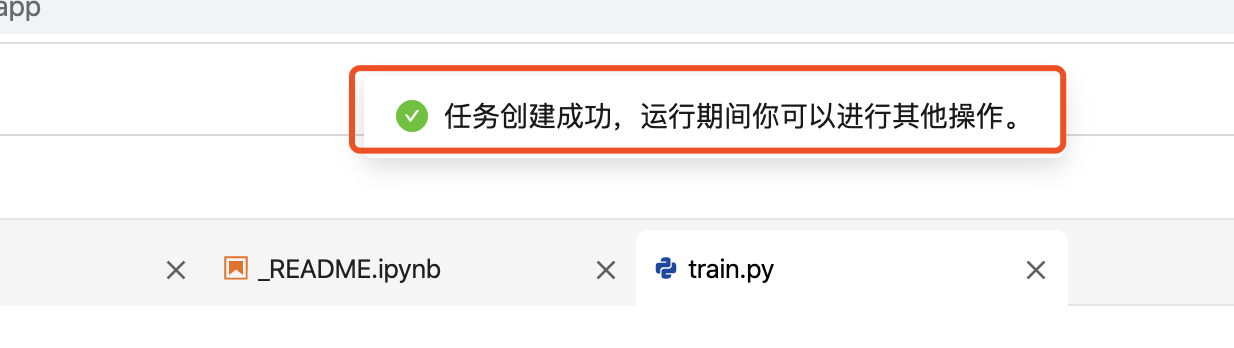


创建离线任务过程：

* 在 notebook 中调试好代码，并转化为 py 文件
* 打开需要运行的 py 文件，点击 运行按钮
* 选择 job 运行环境
  + 选择【CPU 机器】，任务完成后在 results 文件夹查看结果
  + **选择【官方 GPU 机器】，任务完成后在 results 文件夹查看结果**
  + Option选择【矩池云 GPU 机器】，需要去对方网页注册，会默认赠送 6 小时，任务完成后在 matpool-job-res 文件夹中查看结果文件
* 【可选】补充创建 job 名称和参数
  + job 名称：可以自己命名，如果未填写，会有默认值
  + job 参数：遵循 argparse/args 使用方式
* 勾选需要的文件（夹）
  + 一般必须勾选 results 文件夹（返回结果的文件夹，矩池云 GPU 也需要勾选）
  + 一般必须勾选需要的 py 文件和数据集文件夹（datasets）
  + 清理被勾选文件夹里不相关的文件 (e.g. 如勾选 results 文件夹，可以先清除旧模型日志等)



* 创建成功则弹窗如图：



**注意事项：**

* 在训练过程中如果 GPU 时间余额不足，不会影响当次训练任务。
* **请勿重复创建相同的离线任务！！！**
* 如果 GPU 额度不足，可联系助教添加时长。

### 2.2 在 notebook 环境中运行代码出现 kernel 状态从 busy -> restarting -> starting 或者 cell 处于 [\*]等卡住状态

* 采用 chrome 浏览器并用最新版本
* 检查代码是否存在使用内存过高（ > 4G）的情况
* 重启 Kernel：点击左上角 【Kernel -> Restart Kernel】
* 重启项目：点击右上角的【重启】按钮
* 重启 Chrome 浏览器
* 建议使用离线任务来运行该段代码

### 2.3 离线任务出现 kill, out of memory, oom等情况时

使用离线训练模型（CPU、GPU 环境） 的内存限制是 8G。超过内存限制会造成代码运行出现如上问题，所以请大家注意模型的大小，例如可减小 batch size。

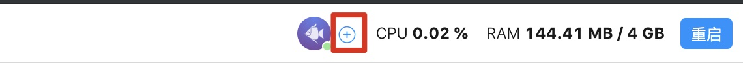
## 3.联系我们

### 3.1 网站讨论群 (zju\_ai\_2020\_ss)

### 3.2 钉钉讨论群

### 3.3 协作

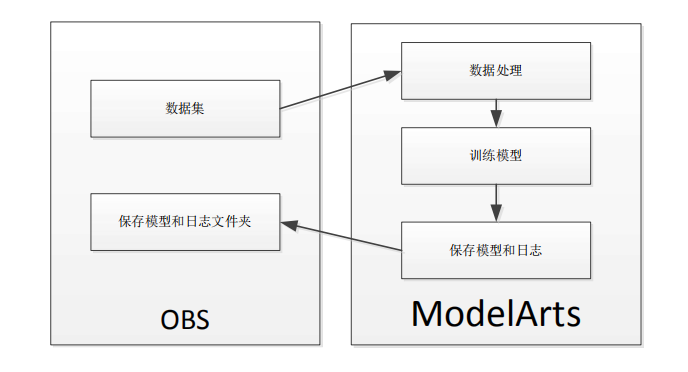
在项目中，点击头像右边的 “+”，搜索添加协作成员账号（助教账号：lbjai）。



邀请协作后，协作双方可以同时进入项目，可以对同一个文件修改，系统会自动同步最新的文件。

# 二、华为云 ModelArts 平台训练模型说明文档

使用华为云训练模型主要需要了解 【对象存储服务 OBS】和 【ModelArts】，前者是放置数据和保存训练模型结果的地方，而后者是训练模型的地方。



流程：

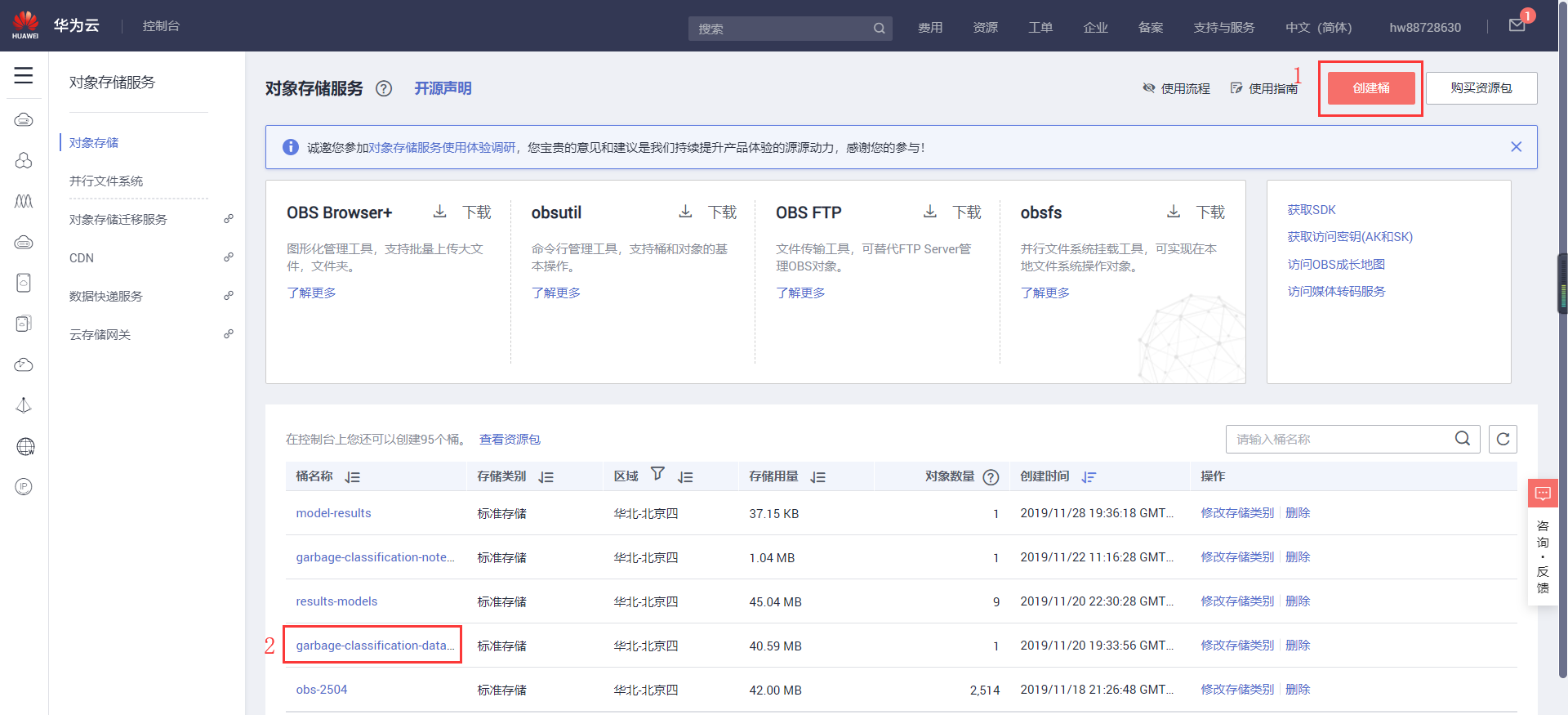
* 首先在 OBS 创建 2 个桶（文件夹），分别用来放置数据集和模型与日志，然后上传数据集；目前我们已经上传好数据集文件
* 进入 ModelArts 获取数据集，进行数据处理、模型训练和保存
* 然后将模型和日志回传到 OBS，然后下载到本地

## 1.获取数据集

你可以在 OBS 处上传数据集，也可以使用我们已经上传的数据集；获取数据集的方式可以参考 1.3 处代码。

### 1.1 OBS

点击图中 1 处创建桶用来保存数据集或者模型等。



点击上图中 2 处，打开如下图的桶内容，然后点击【对象】可以【下载】相关文件。



### 1.2 ModelArts

* 点击开发环境下的开始体验，进入项目管理页面

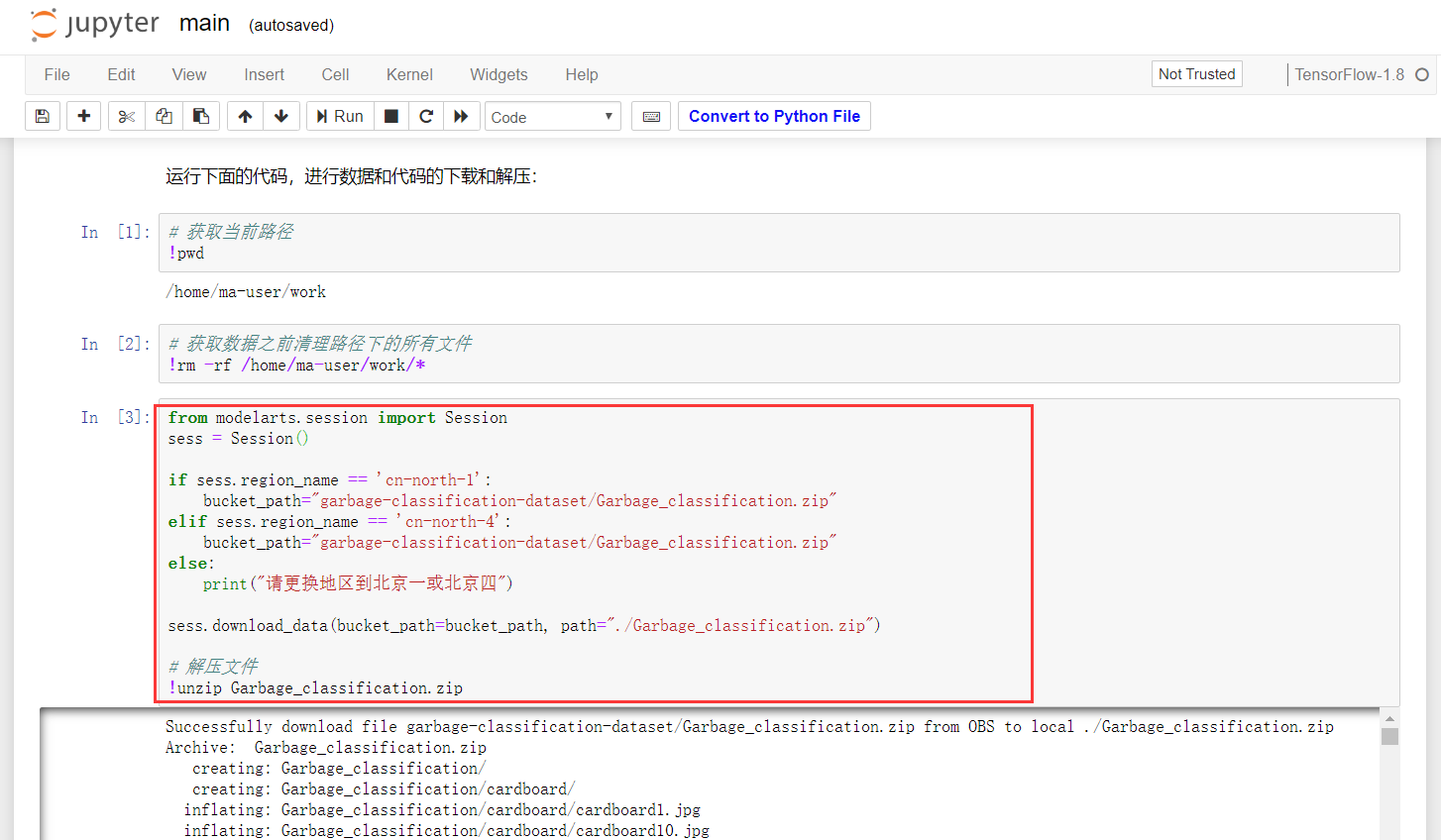


* 你也可以点击创建实现创建项目，也可以打开以前创建的项目



### 1.3 获取数据

获取数据集方式：



该处代码尽量在 notebook 中的 cell 中执行，如上图：

# 获取当前路径

!pwd

# 获取数据之前清理路径下的所有文件

!rm -rf /home/ma-user/work/\*

# 加载数据并解压

from modelarts.session import Session

sess = Session()

if sess.region\_name == 'cn-north-1':

bucket\_path="garbage-classification-dataset/Garbage\_classification.zip"

elif sess.region\_name == 'cn-north-4':

bucket\_path="garbage-classification-dataset/Garbage\_classification.zip"

else:

print("请更换地区到北京一或北京四")

sess.download\_data(bucket\_path=bucket\_path, path="./Garbage\_classification.zip")

# 解压文件

!unzip Garbage\_classification.zip

### 1.4 展示数据集



import os,glob

# 数据集路径

data\_path = "./Garbage\_classification"

# 获取数据名称列表

img\_list = glob.glob(os.path.join(data\_path, '\*/\*.jpg'))

# 打印数据集总量

print("数据集总数量:", len(img\_list))

print("数据路径和名称:",img\_list[0])

## 2.训练模型

在 notebook 环境中调试好模型代码并训练模型。

**注意：请定义好模型路径。**

## **3.把模型以及日志文件保存到 OBS**

我们将包含模型和日志文件（夹）回传至 OBS，然后进入对象存储服务 OBS 下载相关文件(比如模型等)，然后加载到 Mo 平台进行测试。

import moxing as mox

mox.file.shift('os', 'mox')

# 假如模型和日志保存在 ./model/ 文件夹

src = "./model/"

# 我们在 OBS 中创建 results-models 桶，用于接收 src 文件夹中的文件

dst = 's3://results-models'

mox.file.copy\_parallel(src, dst)