本作业分为两个难度，体验作业（低难度）和进阶作业（中等难度），推荐大家先完成体验作业，再尝试进阶作业

进阶作业相比于体验作业，需要额外完成前置的finetune步骤，即不再提供tnews ckpt文件，而是提供bert\_base ckpt，通过完成finetune，生成tnews ckpt文件，后续的evaluation过程与体验作业一致。创建华为云账号，申请公测、创建OBS桶与体验作业一致，如已完成，可以跳过

**进阶作业：完成finetune+Evaluation流程**

体验作业提供：bert\_base文件、tnews数据集、源代码

体验作业流程：

1. 申请华为云账号、申请ModelArts华为云昇腾集群服务公测

2. 从链接下载bert\_base ckpt文件、tnews数据集和源代码

3. 创建obs桶

4. 上传bert\_base ckpt文件、tnews数据集和源代码到obs桶

5. 提交ModelArts 训练作业

6. 等待、查看结果，得到tnews ckpt文件

7. evaluation流程，与体验作业一致

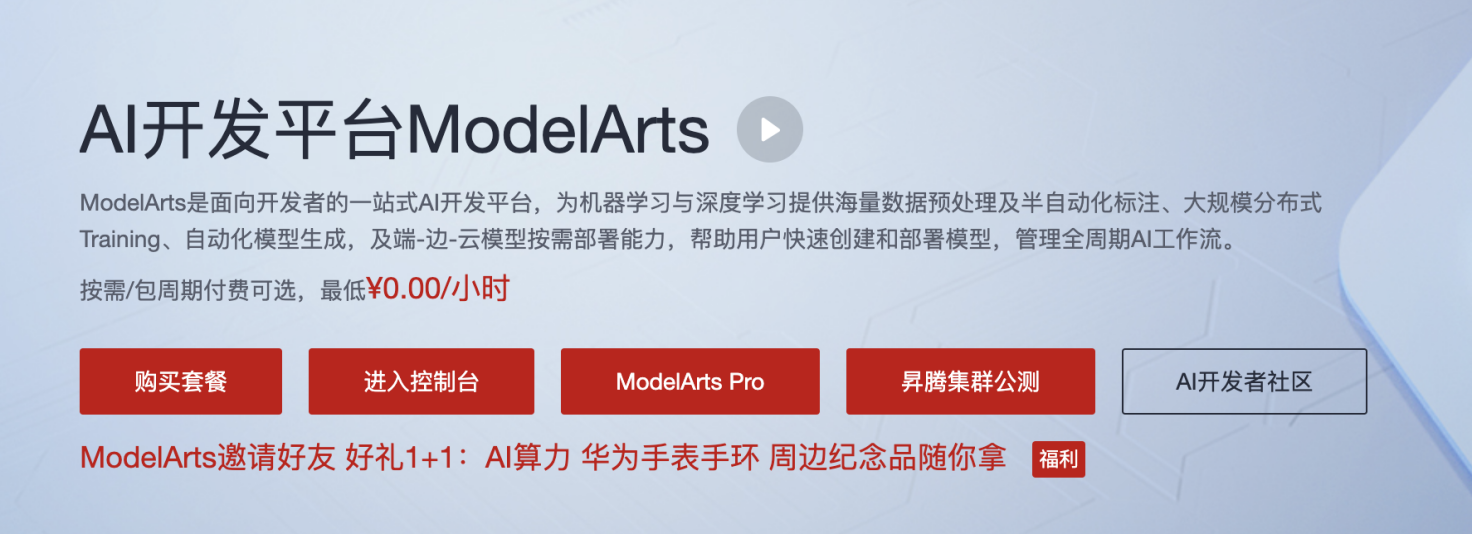
**进阶作业详细步骤**

1. 申请华为云账号、申请ModelArts华为云昇腾集群服务公测（如已完成，可跳过）



<https://www.huaweicloud.com/>

注册完毕之后，进入到ModelArts界面，点击昇腾集群公测，完成申请流程



<https://www.huaweicloud.com/product/modelarts.html>

1. 完成申请之后，开始下载所需文件，

包括数据集文件、bert\_base ckpt文件和源代码文件（红色标出的文件需要重新下载，另外领个可以直接用体验作业中下载的文件）

**tnews数据集：**

https://e-share.obs-website.cn-north-1.myhuaweicloud.com?token=

提取码:

123456

**bert源代码：**

<https://21days-bert.obs.cn-north-4.myhuaweicloud.com:443/bert.zip?AccessKeyId=M7KX8KLMT0ZL1P8QWXZ5&Expires=1634522579&Signature=LY3ZdN9gtoxx8ULDQRkkPuOp1nk%3D>

源代码下载完成后，在本地解压缩

**Bert\_base ckpt:**

https://21days-bert.obs.cn-north-4.myhuaweicloud.com:443/bert\_base.ckpt?AccessKeyId=M7KX8KLMT0ZL1P8QWXZ5&Expires=1634522231&Signature=1CSNQY%2BK%2BmQxRBkqBwBHMbf0%2BE4%3D

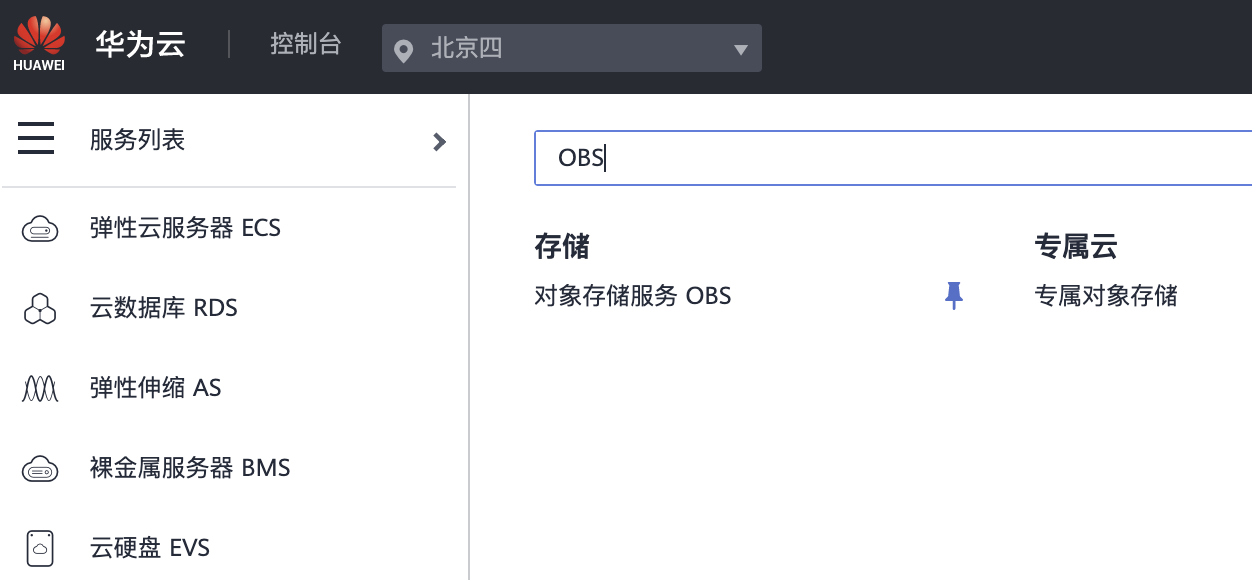
这个文件比较大，请大家耐心下载

1. 创建OBS桶(如已完成，可跳过，直接使用之前创建的桶)

进入官网界面https://www.huaweicloud.com/product/obs.html，点击右上角的控制台按钮



在控制台界面中，在左上角选择北京四节点，同时在左边的弹出式菜单的搜索框中，搜索OBS，并选择“对象存储服务OBS”



此时我们已经进入了OBS对象存储服务的界面，点击右上角“创建桶”按钮



按照以下模板创建桶



数据冗余存储策略选择“单AZ存储”

桶名称自己取一个不重复的名称即可

桶策略采用公共读

其他单策略使用默认单即可

点击创建按钮之后，OBS桶就创建完毕了

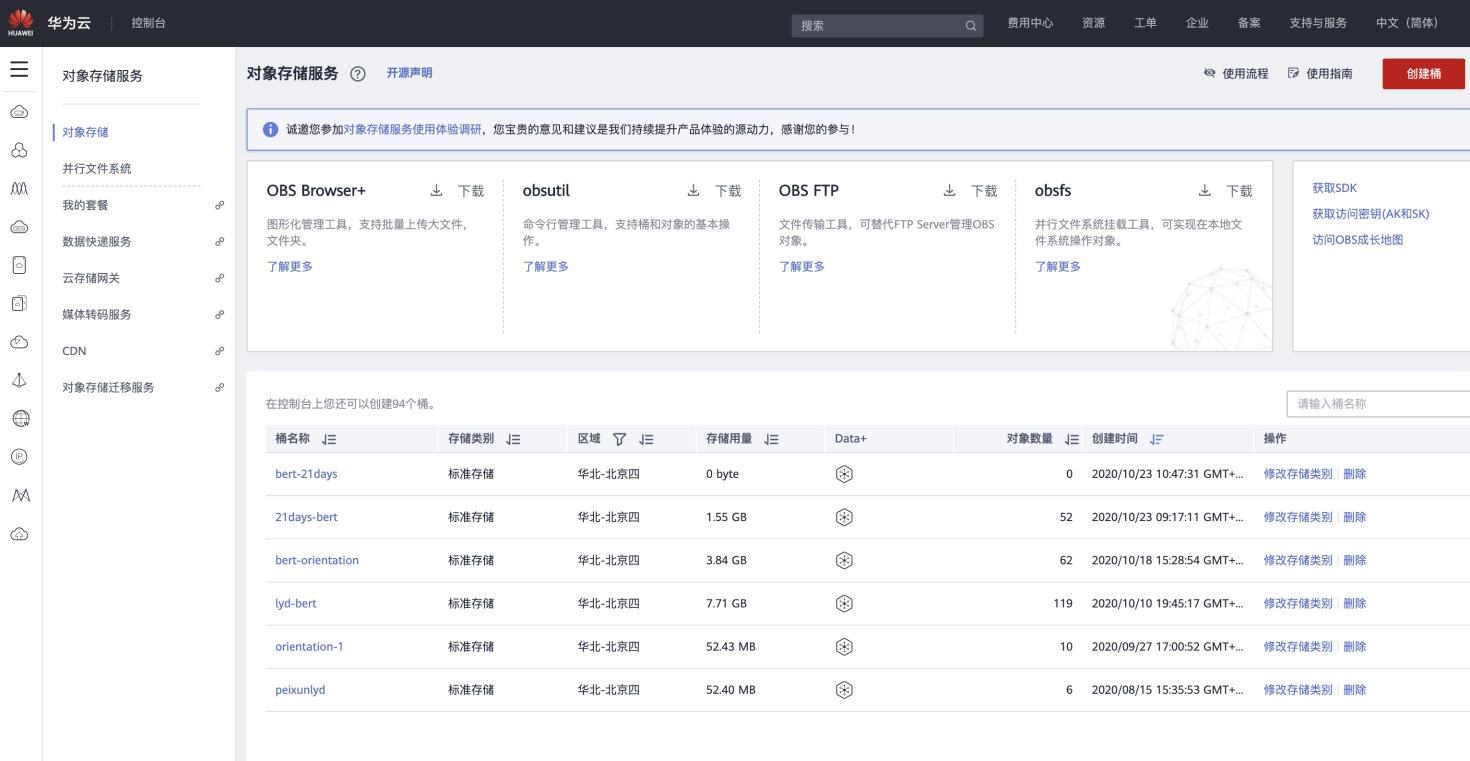
1. 上传文件

这一步与体验作业的区别仅在于进阶作业不上传tnews ckpt，而是上传bert\_base ckpt到代码根目录下

如果之前已经完成了体验任务，我们需要把之前上传的tnews ckpt文件删除

这一步我们需要在刚刚创建的OBS桶内新建一些文件夹，并且上传我们刚刚下载好的文件

在OBS主页面点击进入到刚才创建出来的桶



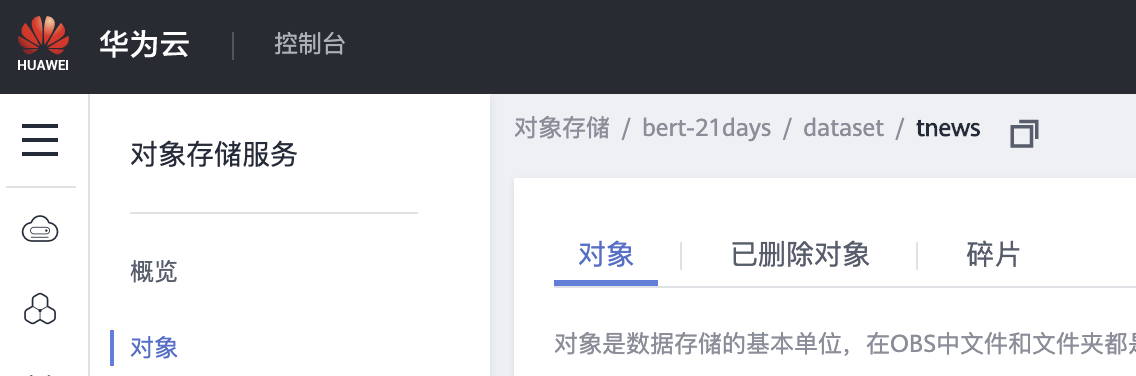
在左侧的弹出菜单中，点击“对象”，然后在右边点击新建文件夹,在根目录下新建两个文件夹：bert和dataset



点击进入dataset文件夹，再创建一个二级文件夹tnews，并且进入tnews文件夹，点击“上传对象”按钮开始数据集的上传，可以采用拖拽的方式，直接将文件拖拽进入下图的红框中



在完成了数据集的上传之后，回到桶的根目录，并且点击进入bert文件夹，同样的方法，开始上传解压缩之后的源代码文件，这里推荐使用拖拽的方式，把解压之后的文件和文件夹全选，然后拖拽进入上图的框中



最后上传bert\_base ckpt文件，这里我们直接把bert\_base ckpt文件上传到bert目录即代码根目录下，由于bert\_base ckpt这个文件比较大，所以上传时间会略微长一些

如果之前已经完成了体验任务，我们需要把之前上传的tnews ckpt文件删除



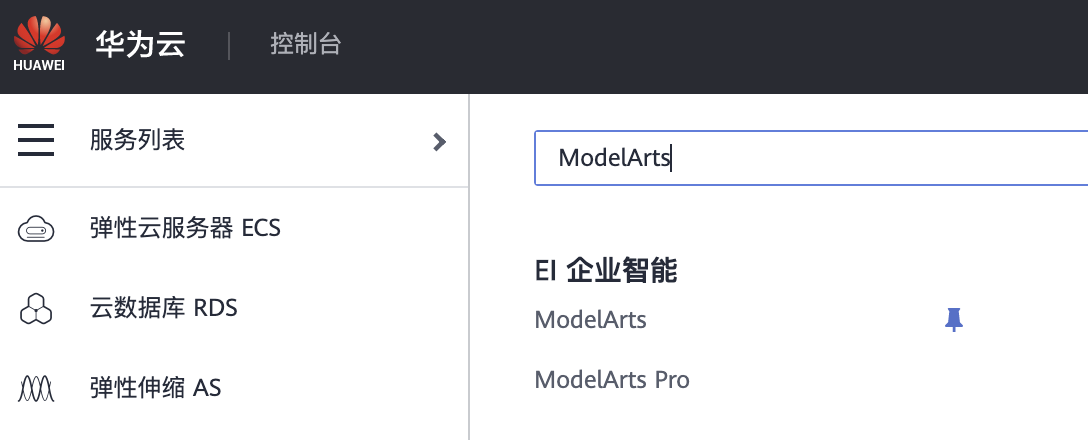
上传全部完成之后，截图如下：



至此，上传部分结束

1. 这一步与体验作业类似，在参数输入方面有小部分区别

提交ModelArts训练作业，类似我们进入OBS的方式，在左侧服务列表的弹出菜单中，在搜索框中输入modelarts，并且选择下面的“ModelArts”



进入ModelArts界面之后，点击左侧“训练管理”，在下拉菜单中点击“训练作业”，进入训练作业界面后，点击中上方的“创建”按钮



在创建训练任务的界面中，名称同样的，自己取一个不重复的名称即可



下面的参数配置中，以下几个参数需要配置：

“算法来源” 选择“常用框架”，

“AI引擎”选择“Ascend-Powered-Engine”和“MindSpore-0.5-python3.7”

“代码目录”选择之前在obs建立的bert目录即代码根目录

“启动文件”选择代码根目录下的finetune.py

“数据来源”选择“数据存储位置”

在“数据存储位置”中之前在obs中建立的dataset文件夹下的tnews文件夹

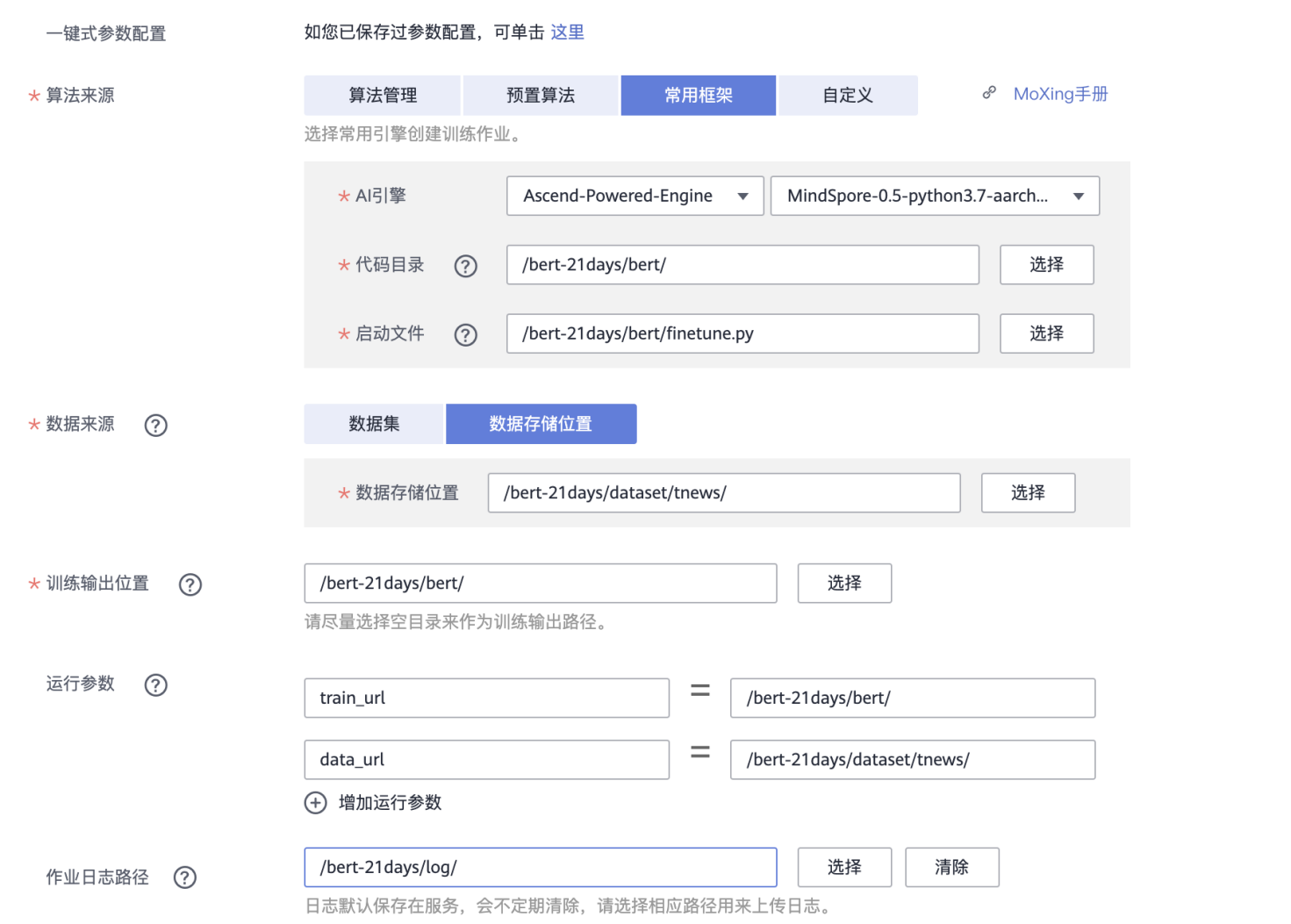
“训练输出位置”这里我们直接把输出位置选择为bert目录下即代码根目录下，直接把训练好的tnews模型输出到代码根目录，方便我们做evaluation，这样做evaluation时就不再需要把tnews ckpt文件从output文件夹拷贝过来了

“作业日志路径”可以在obs根目录下创建一个“log”文件夹用来存放日志输出

“资源池”选择“公共资源池”，“类型”选择“Ascend”，“规格”选择“Ascend: 1 \* Ascend 910 CPU: 24核 96GiB”

其他参数选择默认即可

最终参数设置结果如下图所示：



然后就可以提交任务了

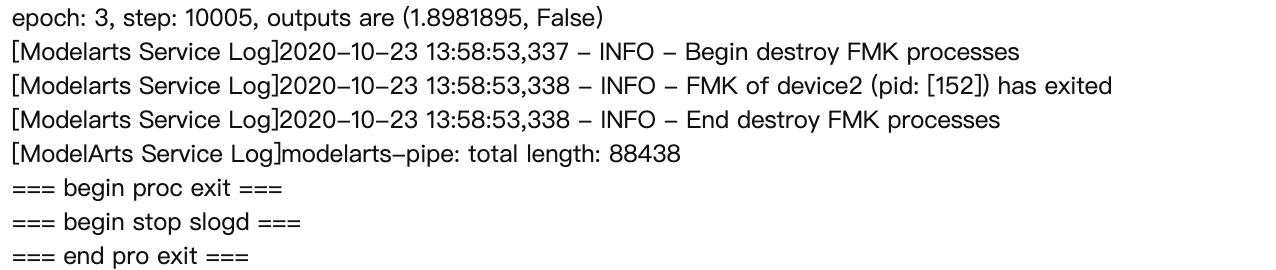
1. 等待、查看结果

finetune需要的时间比evaluation更长，约10分钟左右，同时等待的时间也主要取决于当前提交训练作业的人数，同样的，大家在提交作业之后定时回来查看状态即可

如果显示“运行成功”，即代表finetune运行作业已经完成，我们可以，可以点击作业名称，进入到作业内部详情，再点击“日志”，查看输出



在日志里上下拖动滑动条，找到epoch 3和end exit这两项，代表finetune完成



同时我们回到OBS，bert目录即代码根目录下面，检查是否生成了tnews ckpt文件



可以看到，finetune生成的tnews ckpt文件已经存储到了代码根目录下面，接下来的evaluation任务与体验作业是一样的，我们直接可以使用这个生成的tnews ckpt文件去创建evaluation任务并提交即可，参数也与之前一致。

学员可以截图finetune日志和obs桶里生成的tnews ckpt文件并上传完成作业