

使用集群进行神经网络的 训练

Anaconda 简介



Conda 是一个开源的软件包管理系统和环境管理系统，且支持Python、R、Ruby、Lua、Scala、Java、JavaScript、C/ C++、FORTRAN等多种语言。

下载Anaconda

1. 在集群的 `/public/sourcecode/anaconda/` 目录下有已经下载好的几个版本的anaconda 源文件, 如果版本符合需求可以直接对源码进行编译安装。
2. 也可以自行下载源文件到自己节点目录下, 可以使用的镜像站有:
 - Anaconda官方网站 : <https://repo.anaconda.com/archive/>
 - 清华镜像站 : <https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/>

安装Anaconda

1. 下载成功后，运行下载脚本进行安装：

```
sh conda_installer.sh -b -p ${HOME}/anaconda3
```

-b: 不允许其更改.bashrc, .bash_profile, .zshrc 等脚本

-p: 指定安装目录。

注意：

1. 用集群目录下的源文件安装时，把安装目录指定为自己节点下的目录，否则会导致安装失败.
2. 指定的安装目录在安装前不能存在。
3. 若是不加-b，直接安装会要求选择是否运行conda init, 不建议运行conda init.

2. 安装成功后，删除安装包

```
rm -f conda_installer.sh
```

安装Anaconda

3. 将环境变量写入:

```
echo "export PATH=${HOME}anaconda3/bin:\$PATH" >> ~/.bashrc
```

4. 执行更改后的脚本

```
source ~/.bashrc
```

注意：如果是曾经在自己节点中改用了其他shell解释器如：zsh, csh 等, 请在相应的配置脚本中更改环境变量。

创建虚拟环境

如果对于python的科学运算有不同的需求，建议在安装工具库时，先用conda创建相关的虚拟环境，保证工具库的有序和安全。

- 创查看已有的虚拟环境：`conda env list`
- 创建环境：`conda create -n env_name python=x.x pkg1 pkg2`
- 激活环境：`source activate env_name`
- 关闭环境：`conda deactivate env_name`
- 删除环境：`conda remove -n env_name --all`

注意：用 `source activate` 而不是 `conda activate`, 否则会要求启动 `conda init`

建议：可以把目前的环境转移到新环境中

- 得到旧环境的配置文件：`conda env export > my_environment.yml`
- 根据配置文件配置环境：`conda env create -f environment.yml`

安装计算包

1. 激活虚拟环境：`source activate env_name`
2. 使用 `conda install` 命令安装相关的python工具包
 - `conda install numpy`
 - `conda install matplotlib`
 - `conda install pytorch torchvision torchaudio cudatoolkit=10.2 -c pytorch`
3. 卸载安装包：`conda remove pkg_name`
4. 查看目前环境的包：`conda env list`

Anaconda换国内源

- 打开用户目录下的 `vim ~/.condarc`, 将配置文件写入：

- 清华源

```
channels:  
- https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkg/main/  
- https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkg/free/  
- https://mirrors.tuna.tsinghua.edu.cn/anaconda/cloud/conda-  
forge/  
ssl_verify: true
```

- 上海交通大学源

```
channels:  
https://mirrors.sjtug.sjtu.edu.cn/anaconda/pkg/main/  
https://mirrors.sjtug.sjtu.edu.cn/anaconda/pkg/free/  
https://mirrors.sjtug.sjtu.edu.cn/anaconda/cloud/conda-  
forge/  
ssl_verify: true
```


文件传输

- **sftp**

在终端中输入: `sftp -oPort=22112 username@remote ip(or remote host name)`, 例如: `sftp xiesj@10.15.22.111`

使用sftp的常用命令 `get` 和 `put` 来上传本地文件和下载远程文件

- **scp**

- 从本地复制到远程: `scp -P 22112 local_file remote_username@remote_ip:remote_folder`

- 从远程复制到本地: `scp -P 22112 username@remote_ip:remote_file local_file`

编写作业提交脚本

打开并编辑作业提交脚本: vim sub.pbs

```
#!/bin/bash
#PBS -N Top
#PBS -l nodes=1:ppn=1
#PBS -l walltime=24:00:00
#PBS -q sist
#PBS -I
cd $PBS_O_WORKDIR
echo 'your pbs file sub.pbs has been generated successfully'
source activate torch
python train_VAE.py
```

注意：在指定作业资源时可以将作业提交到特定的节点或者特定设备类型的节点上，如：

- #pbs -l nodes=1:ppn=1:centos7
- #pbs -l nodes=1:ppn=7:gpus=1:V100

编写完成后，使用脚本提交作业: qsub sub.pbs

查看提交结果

使用qstat 来查看作业状态

```
(base) [xiesj@HPC-login project]$ qstat
Job ID              Name              User              Time Use S Queue
-----
2788572.node1      Top              xiesj              0 Q sist
```

使用qdel 来取消作业

```
(base) [xiesj@HPC-login project]$ qdel 2788572
(base) [xiesj@HPC-login project]$ qstat
Job ID              Name              User              Time Use S Queue
-----
2788572.node1      Top              xiesj              00:00:00 C sist
```

打开 name.o_id, name.e_id 文件可以查看程序的输出和报错信息

参考链接

- 高性能平台使用手册：<https://it.shanghaitech.edu.cn/8851/list.htm>
- conda在线文档：<https://conda.io>
- Pytorch官方文档：<https://pytorch.org/docs/stable/index.html>