

FloydHub

FloydHub 是一个在云端训练和部署深度学习模型的平台。借助该平台，你无需再启动自己的云端实例和配置环境。例如，FloydHub 将自动利用 TensorFlow、整个 Python 数据科学工具包以及 GPU 设置 AWS 实例。然后，你可以在实例上运行你的脚本或 Jupyter notebook。实际上，借助 FloydHub，你只需输入不超过 3 个命令就能在 GPU 上训练你的网络。

开始使用

在 **FloydHub** 上注册帐号（可以免费注册！）。注册完毕后，就可以在 FloydHub 上运行你的深度学习网络了。首先，你需要安装命令行工具

```
pip install -U floyd-cli
```

安装完毕后，你可以输入以下命令，登录 FloydHub：

```
floyd login
```

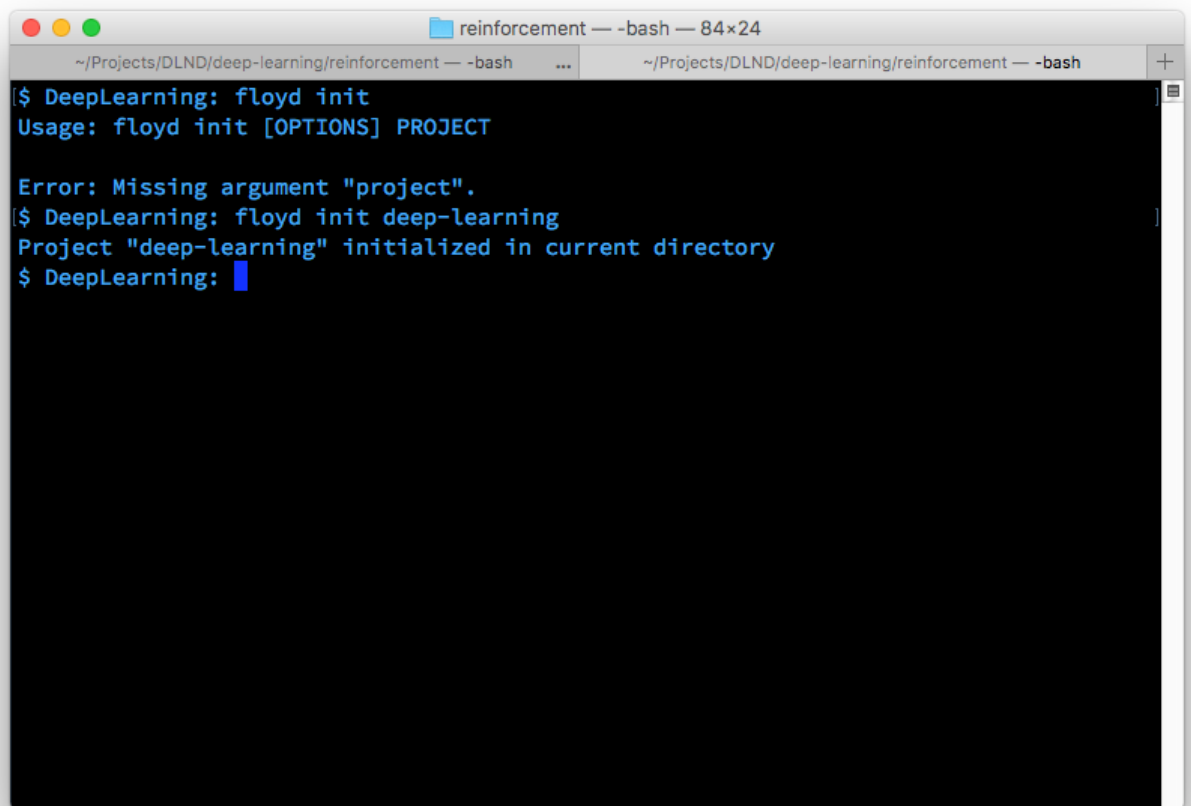
系统会要求你输入身份验证令牌（authentication token），你可以从 FloydHub 网站获取令牌。系统会提示“身份验证令牌页面现在将在您的浏览器中打开，要继续吗？

（Authentication token page will now open in your browser. Continue?）”，以便于你查看该令牌。

运行 FloydHub 实例

首先，你需要初始化 floyd，以在你的项目中使用。在包含项目代码的目录下，运行：

```
floyd init project_name
```



```
reinforcement — -bash — 84x24
~/Projects/DLND/deep-learning/reinforcement — -bash
[$ DeepLearning: floyd init
Usage: floyd init [OPTIONS] PROJECT

Error: Missing argument "project".
[$ DeepLearning: floyd init deep-learning
Project "deep-learning" initialized in current directory
$ DeepLearning: ]
```

如果你跟我一样，输入到一半忘了项目名称，系统会提示你。初始化后，你可以在 FloydHub 实例上用命令 `floyd run` 运行你的代码。你需要发出具体命令，例如 `floyd run "python test.py"`。

```
reinforcement — -bash — 84x24
~/Projects/DLND/deep-learning/reinforcement — -bash ...
~/Projects/DLND/deep-learning/reinforcement — -bash +
[$ DeepLearning: floyd run "python test.py"
Creating project run. Total upload size: 2.4MiB
Syncing code ...
RUN ID          NAME          VERSION
-----
e3bcgFaXVee4KcMHaD67oW  mcleonard/deep-learning:3  3

To view logs enter:
  floyd logs e3bcgFaXVee4KcMHaD67oW

$ DeepLearning: 
```

从上图中可以看出，系统给此次运行分配了一个 ID: `e3bcgFaXVee4KcMHaD67oW`。你可以使用该 ID 访问日志以及脚本可能会输出的任何数据（例如 TensorFlow 检查点）。查看日志：

```
reinforcement — -bash — 84x24
~/Projects/DLND/deep-learning/reinforcement — -bash ...
~/Projects/DLND/deep-learning/reinforcement — -bash +
|$ DeepLearning: floyd logs e3bcgFaXVee4KcMHaD67oW
2017-04-19 12:54:09,802 INFO - Preparing to run TaskInstance <TaskInstance: mcleonar
d/deep-learning:3 (id: ad3LADeEwiD74iYzfkBmr2) (checksum: f95ba36bea88c7efdff21f2443
0d74b7) (last update: 2017-04-19 12:54:09.794012) [queued]>
2017-04-19 12:54:09,813 INFO - Starting attempt 1 at 2017-04-19 12:54:09.806652
2017-04-19 12:54:10,523 INFO - Executing command in container: stdbuf -o0 sh command
.sh
2017-04-19 12:54:10,523 INFO - Pulling Docker image: floydhub/tensorflow:latest-py3
2017-04-19 12:54:11,763 INFO - Starting container...
2017-04-19 12:54:11,971 INFO -
#####

2017-04-19 12:54:11,971 INFO - Run Output:
2017-04-19 12:54:12,008 INFO - Hello world
2017-04-19 12:54:12,054 INFO -
#####

2017-04-19 12:54:12,054 INFO - Waiting for container to complete...
2017-04-19 12:54:12,434 INFO - [success] Finishing execution in 2 seconds for TaskIn
stance <TaskInstance: mcleonard/deep-learning:3 (id: ad3LADeEwiD74iYzfkBmr2) (checks
um: f95ba36bea88c7efdff21f24430d74b7) (last update: 2017-04-19 12:54:12.426257) [suc
cess]>

$ DeepLearning:
```

我在 `test.py` 中输出了 `"Hello world"`，你可以在运行输出结果中看到这一内容。

使用 GPUs

在训练深度学习网络时，你经常需要使用 GPU。借助 FloydHub，你可以非常轻松地使用 GPU，只需添加 `--gpu` 标记：

```
floyd run "python test.py" --gpu
```

Jupyter Notebooks

借助 FloydHub，你还可以轻松地在 FloydHub 实例上运行 Jupyter Notebook。

```
floyd run --mode jupyter
```

你也可以像之前一样添加 `--gpu` 标记，这样就可以使用 GPU 运行。

```
reinforcement — -bash — 84x24
~/Projects/DLND/deep-learning/reinforcement — -bash ...
~/Projects/DLND/deep-learning/reinforcement — -bash
[$ DeepLearning: floyd run --mode jupyter --gpu
Creating project run. Total upload size: 3.0MiB
Syncing code ...
Done=====] 3206395/3206395 - 00:00:00
RUN ID          NAME          VERSION
-----
bUNkL8oniF6BUbtYfuMSXi  mcleonard/deep-learning:6      6

Setting up your instance and waiting for Jupyter notebook to become available .....
.....

Path to jupyter notebook: https://www.floydhub.com:8000/RghRYFXuxjoPJUcFWYTCb6

To view logs enter:
    floyd logs bUNkL8oniF6BUbtYfuMSXi

$ DeepLearning: 
```

你将看到 notebook 服务器的 URL。在浏览器中打开该 URL，就可以开始使用 notebook 了。

停止运行

通常，运行完毕后，实例就会自动关闭。但是，Jupyter 服务器将一直运行，除非你手动关闭。为了避免浪费时间和金钱，你需要手动停止实例。你可以使用命令 `floyd stop run_ID`。如果你忘记了运行 ID，可以用这个命令找到：`floyd status`。

```
reinforcement — -bash — 84x24
~/Projects/DLND/deep-learning/reinforcement — -bash
$ DeepLearning: floyd status
RUN ID          CREATED          STATUS    DURATION(s)  NAME
-----
INSTANCE        VERSION
-----
Xe6otckWwe67LCb3mVgdnT  a minute ago  running   0            mcleonard/deep-learn
ning:7  gpu          7
bUNkL8oniF6BUbtYfuMSXi  36 minutes ago  shutdown  1192         mcleonard/deep-learn
ning:6  gpu          6
vnfFqn88nsfhPmUD7APZRC  38 minutes ago  shutdown  70           mcleonard/deep-learn
ning:5  gpu          5
SgAPi24EpoYt3AV9L5oAUU  44 minutes ago  shutdown  129          mcleonard/deep-learn
ning:4  gpu          4
e3bcgFaXVee4KcMHaD67oW  2 hours ago     success   3            mcleonard/deep-learn
ning:3  cpu          3
fTVoZ6WEvfhdLHSfU8LksV  2 hours ago     success   6            mcleonard/deep-learn
ning:2  cpu          2
YGfCJ93hj2gy52GvTn2RgE  2 hours ago     success   3            mcleonard/deep-learn
ning:1  cpu          1
$ DeepLearning:
```

可以看到，列表中的第一个实例正在运行，ID 是 `x6otckwwe67LCb3mVgdnT`。你可以通过输入以下命令停止该实例：`floyd stop x6otckwwe67LCb3mVgdnT`。此外，如果你想详细了解该实例，输入命令：`floyd info x6otckwwe67LCb3mVgdnT`。

```
reinforcement — -bash — 84x24
~/Projects/DLND/deep-learning/reinforcement — -bash  ...  ~/Projects/DLND/deep-learning/reinforcement — -bash
$ DeepLearning: floyd info Xe6otckWwe67LCb3mVgdnT
-----
Run ID      Xe6otckWwe67LCb3mVgdnT
Name        mcleonard/deep-learning:7
Created     3 minutes ago
Status      running
Duration(s) 0
Output ID   NQEaQUPNb2Bu9RcqUNA69C
Instance    gpu
Version     7
Mode        jupyter
Url         https://www.floydhub.com:8000/NQEaQUPNb2Bu9RcqUNA69C
-----
$ DeepLearning: 
```

环境

FloydHub 包含了大量热门的深度学习框架，例如 TensorFlow、Keras、Caffe、Torch 等等。你可以通过设置环境指定要使用哪个框架。这是 [FloydHub 支持的环境列表](#)，并且将会支持更多环境！

要指定环境，使用 `--env` 标记，然后是环境名称（名称请参阅环境列表）。例如，如果你想使用 TensorFlow 1.0、Jupyter notebook 和 GPU，则输入：

```
floyd run --mode jupyter --gpu --env tensorflow-1.0
```

FloydHub 和数据

对于 FloydHub，你是从你的机器上将数据上传到 FloydHub 的远程实例上。强烈建议不要在脚本中上传大型数据集，例如 CIFAR。你应该始终下载 FloydHub 实例上的数据，而不是从你自己的机器上上传。

公共数据集

FloydHub 提供了公共数据集，所以你不用每次运行时都下载大型数据集。你可以[在此处](#)找到公共数据集列表。例如，如果你使用的是 CIFAR10 数据集，你可以使用 `--data` 标记将其附加到你的实例上：

```
floyd run "python train.py" --data diSgciLH4WA7HpCHNasP9j
```

这里，`diSgciLH4WA7HpCHNasP9j` 是该数据集的 ID，你可以从上述列表中获取其他数据集的 ID。数据集将位于 `/input` 文件夹下，所以 CIFAR10 数据集将位于 `/input/CIFAR10` 下。

输出

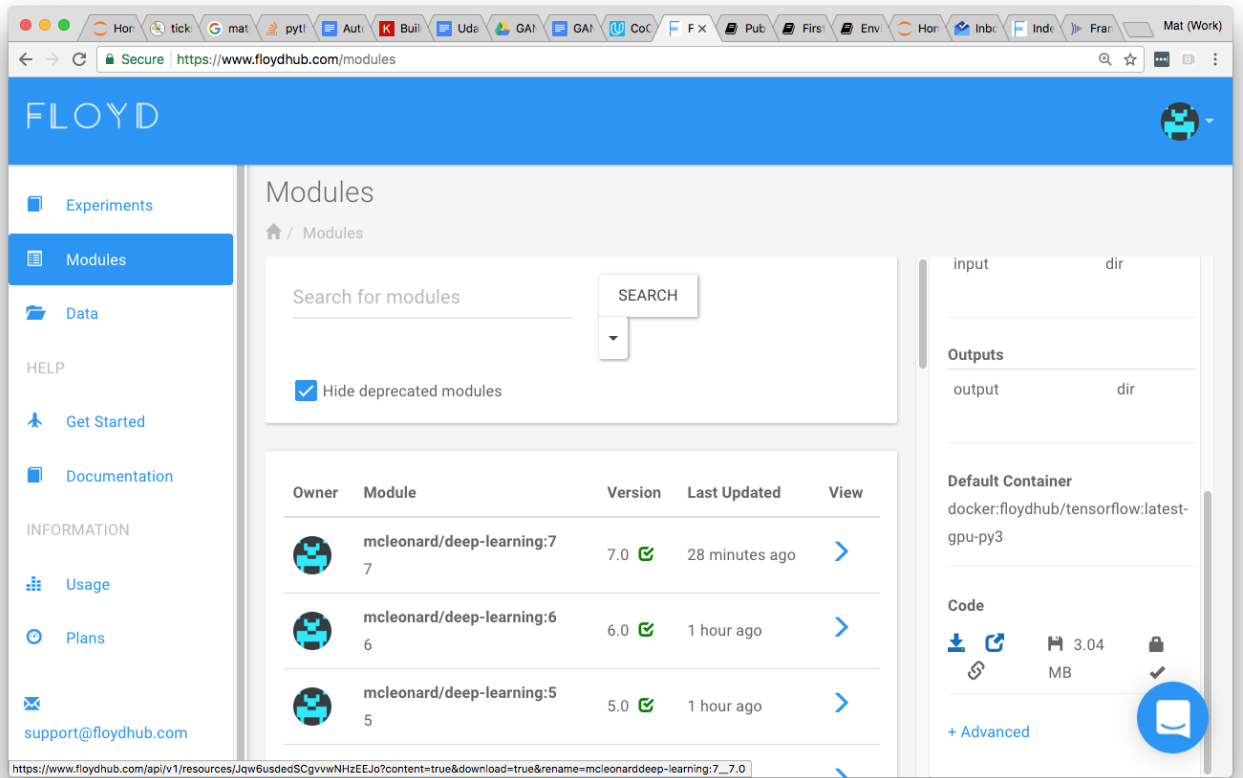
你经常会写出数据，例如 TensorFlow 检查点，或者更新 notebook。要获取这些文件，你可以使用以下命令来获取数据链接：

```
floyd output run_ID
```

该命令将打开一个 HTML 页面，其中包含实例存储空间的 `/output` 目录下的内容。要向文件中写入数据，并稍后能够获取这些数据，你需要将其保存到 `/output` 目录下。例如，你应该将 TensorFlow 检查点保存到 `/output/checkpoints` 下。Notebook 及你上传到实例的所有其他文件将自动保存到 `/output` 下。

即使实例被终止，这个方法仍然有用。所以你可以训练网络、保存检查点，并停止示例，然后稍后下载检查点。

你还可以通过控制面板将整个 `/output` 文件夹下载为 TAR 文件。在左侧侧栏中选择“模块 (Modules)”，然后选择你要从中获取数据的模块。在右侧侧栏中点击蓝色向右箭头即可获取模块信息。你将看到“代码 (Code)”，下方是下载 TAR 文件的链接。



更多资源

这篇文章应该可以帮助你顺利使用 FloydHub 了。如果你想进一步了解 FloydHub，请参阅他们的[入门指南](#)。你还可以查看一些其他相关指南，了解如何训练有趣的模型，例如[风格迁移模型](#)。