

在OpenPAI上训练 对联数据

秦婷婷
微软亚洲研究院

第一步：申请账号

在已经部署好的OpenPAI平台上申请集群账号
获得：

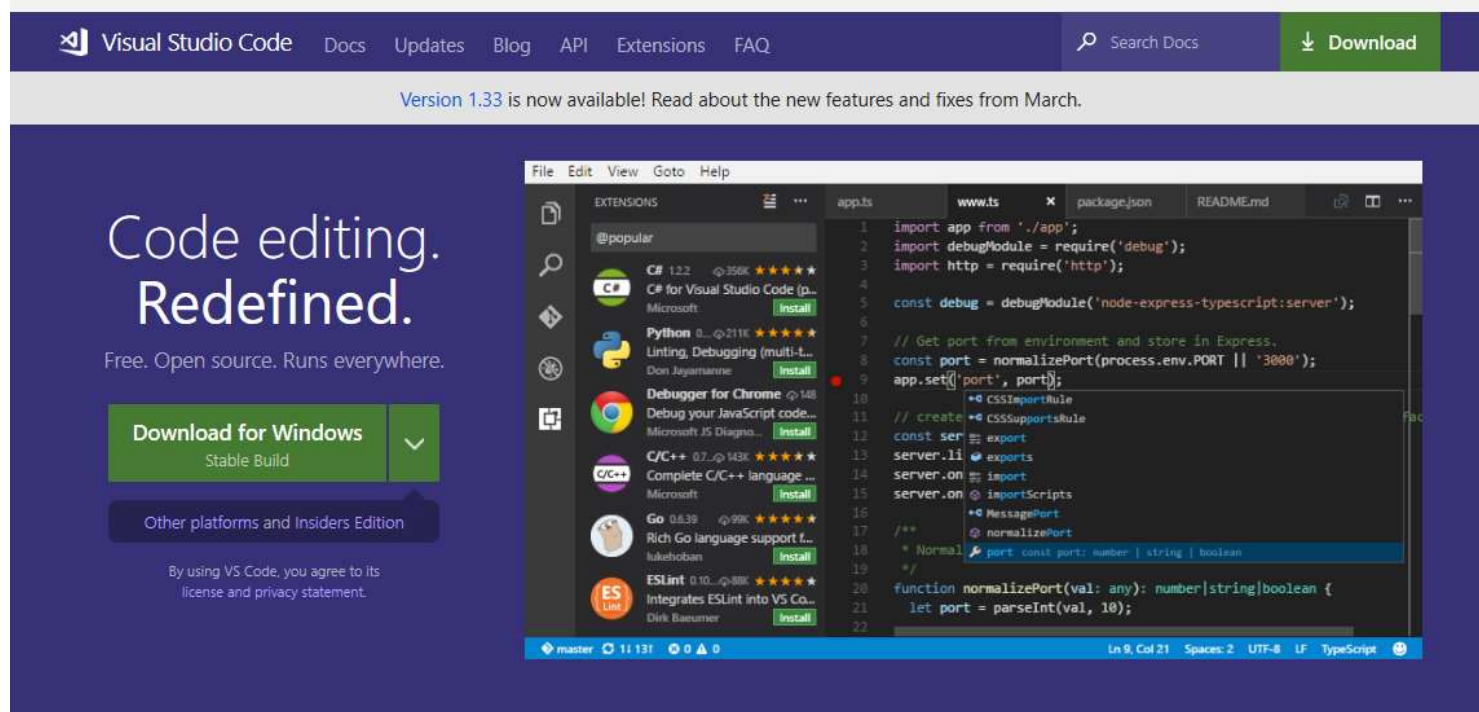
集群IP

用户名

登录密码



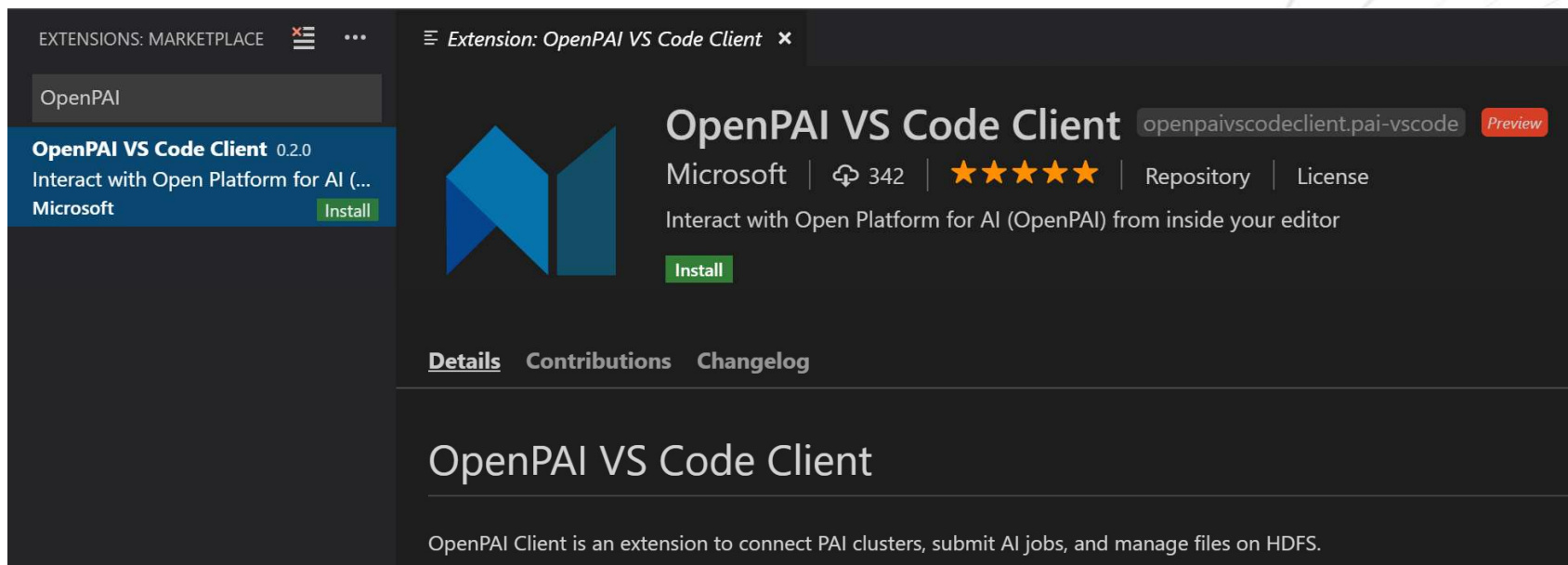
第二步：下载Visual Studio Code



下载地址：<https://code.visualstudio.com/>

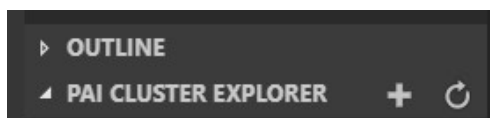
第三步：下载OpenPAI VS Code Client插件

- 打开Visual Studio Code, 在插件搜索栏中搜索：OpenPAI VS Code Client
- 点击 “Install” 安装插件

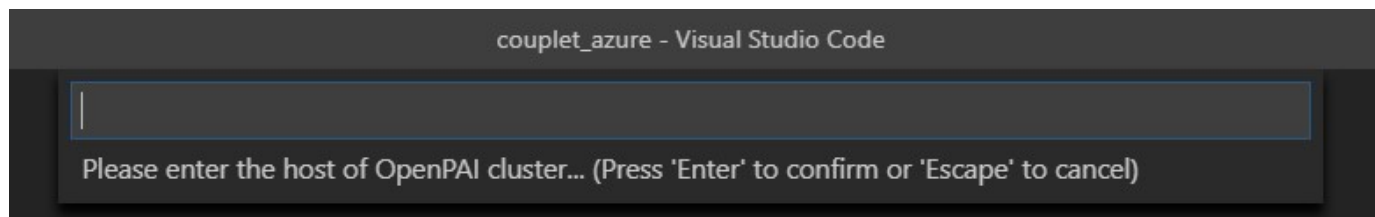


第四步：添加集群

点击左侧边栏  图标，回到Explorer界面，左下角出现PAI CLUSTER EXPLORER”



点击 “+” 按钮，添加集群。在VS Code界面上部的弹出窗，填写集群IP并回车。



第四步：添加集群（续）

更新配置文件

- 填写用户名和密码
- 点击“Finish”保存配置

```
----// PAI cluster name
----"name": "Azure-PAI",

----// PAI cluster user name
----"username": "██████████",

----// PAI cluster password
----"password": "██████████",

----// PAI rest api endpoint
----"rest_server_uri": "██████████rest-server",

----// PAI hdfs endpoint
----"hdfs_uri": "hdfs://██████████9000",

----// PAI webhdfs endpoint
----"webhdfs_uri": "██████████webhdfs/api/v1",

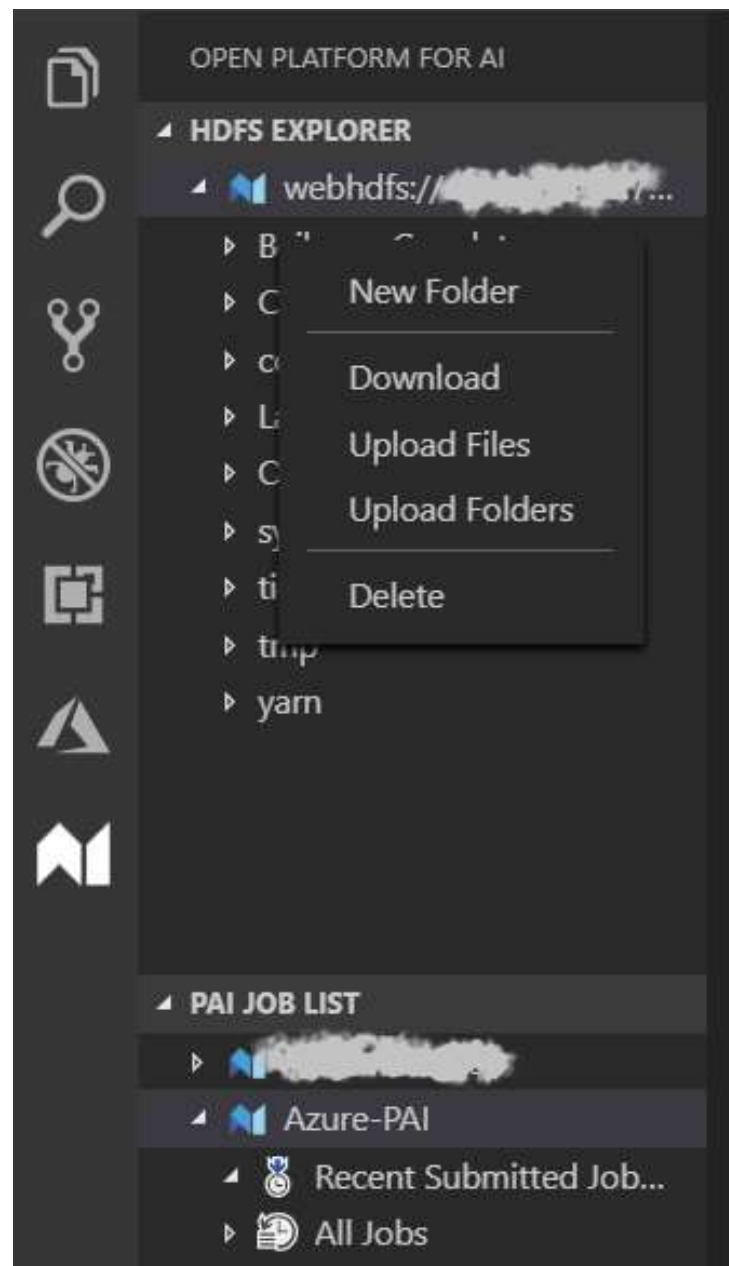
----// PAI grafana dashboard endpoint
----"grafana_uri": "██████████grafana",

----// PAI kubernetes dashboard endpoint
----"k8s_dashboard_uri": "██████████kubernetes-dashboard",

----// PAI web portal endpoint
----"web_portal_uri": "██████████"
```

第五步：上传数据

- 在当前集群，双击“Open HDFS”，打开HDFS存储目录。
- 在HDFS Explorer中，右键单击
- 点击Upload Files
- 上传对联数据文件到指定路径。



第六步：编写程序

- 创建文件夹
- 在Visual Studio Code中打开该文件夹
- 在文件夹中添加文件并编写训练脚本和相关代码
 - 添加HDFS操作支持文件 (hdfshelper.py, __init__.py)
 - 添加训练脚本文件 (train.sh)
 - 添加主文件 (run_samples.py)



第七步：提交任务

在当前集群中，双击Create Job Config...创建配置文件。更改配置并保存

```
{
  "jobName": "test_couplet_001",
  "image": "tobeygin/pai.build.cu80:my",
  "codeDir": "$PAI_DEFAULT_FS_URI/$PAI_USER_NAME/$PAI_JOB_NAME",
  "dataDir": "$PAI_DEFAULT_FS_URI/Data/$PAI_JOB_NAME",
  "outputDir": "$PAI_DEFAULT_FS_URI/Output/$PAI_JOB_NAME",
  "taskRoles": [
    {
      "name": "test_001",
      "taskNumber": 1,
      "cpuNumber": 2,
      "gpuNumber": 1,
      "memoryMB": 8192,
      "command": "pip3 --quiet install future && cd $PAI_JOB_NAME && chmod
+x train.sh && python3 run_samples.py"
    }
  ]
}
```

第七步：提交任务（续）

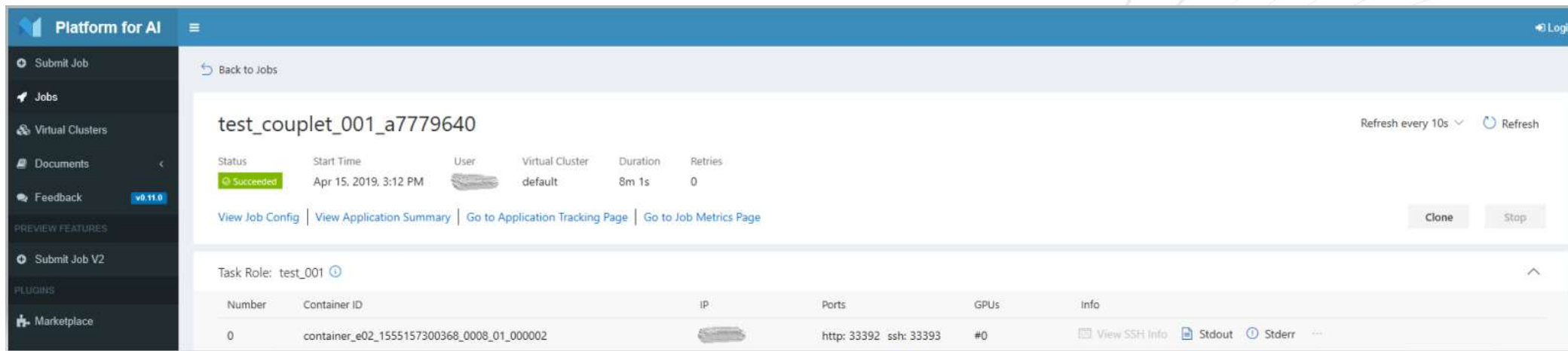
在配置文件中，单机鼠标右键，
选择Submit Job to PAI Cluster，
提交任务

```
{
  "jobName": "test_couplet_001",
  "image": "tobeyqin/pai.build.cu80:my",
  "codeDir": "$PAI_DEFAULT_FS_URI/$PAI_USER_NAME/$PAI_JOB_NAME",
  "dataDir": "$PAI_DEFAULT_FS_URI/Data/$PAI_JOB_NAME",
  "outputDir": "$PAI_DEFAULT_FS_URI/Output/$PAI_JOB_NAME",
  "taskRoles": [
    {
      "name": "test_001",
      "taskNumber": 1,
      "cpuNumber": 2,
      "gpuNumber": 1,
      "memoryMB": 8192,
      "command": "pip3 --quic"
    }
  ]
}
```

Change All Occurrences	Ctrl+F2
Format Document	Shift+Alt+F
Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
Command Palette...	Ctrl+Shift+P
Simulate PAI Job Running	
<u>Submit Job to PAI Cluster</u>	

第八步：查看状态

提交任务后，右下角提示Open Web Portal，
点击可打开OpenPAI的web portal，查看任务的运行状态。

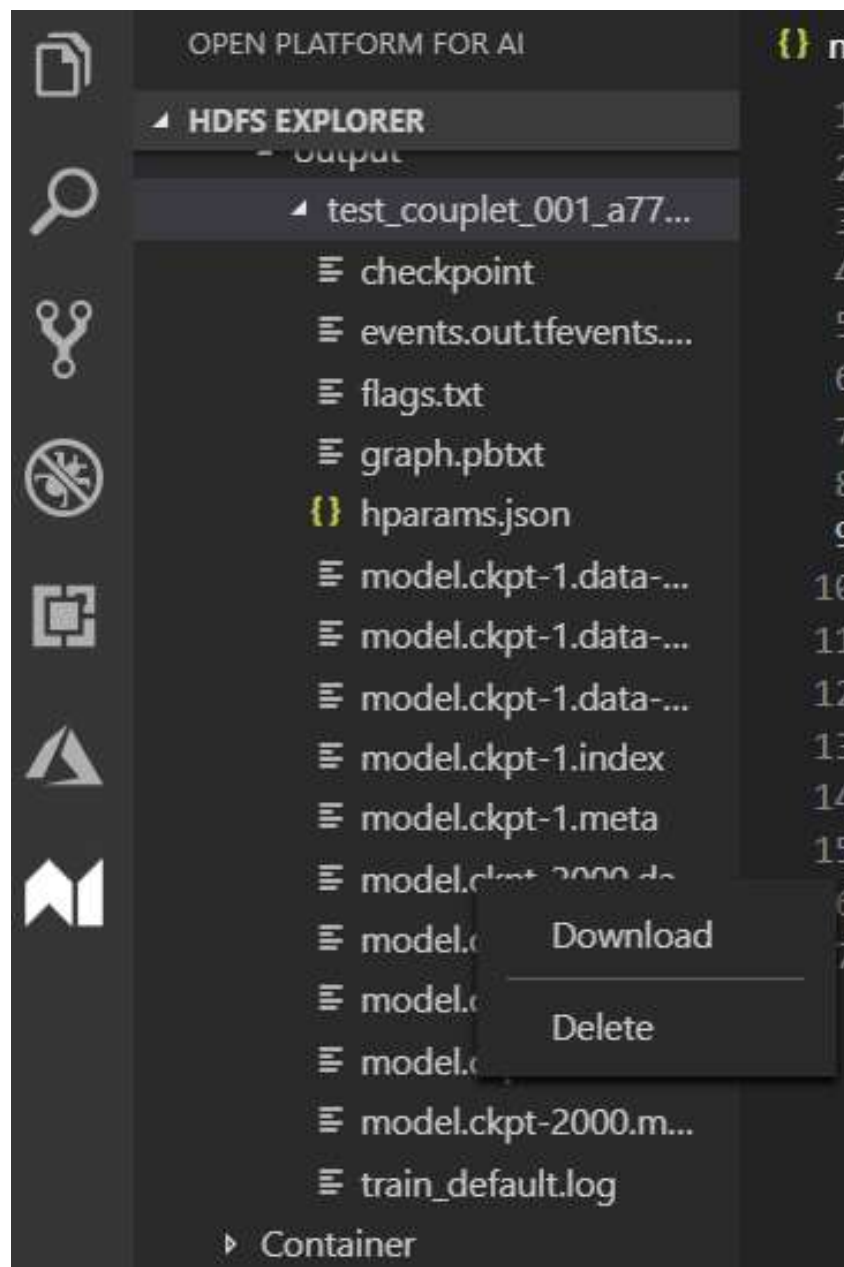


The screenshot displays the OpenPAI Platform for AI web portal. The left sidebar contains navigation links: Submit Job, Jobs, Virtual Clusters, Documents, Feedback (v0.11.0), PREVIEW FEATURES, Submit Job V2, and PLUGINS (Marketplace). The main content area shows the job details for 'test_couplet_001_a7779640'. The status is 'Succeeded', and the start time is 'Apr 15, 2019, 3:12 PM'. The virtual cluster is 'default', and the duration is '8m 1s'. Below the job details, there are links for 'View Job Config', 'View Application Summary', 'Go to Application Tracking Page', and 'Go to Job Metrics Page'. A table lists the task role 'test_001' with one container. The table has columns for Number, Container ID, IP, Ports, GPUs, and Info.

Number	Container ID	IP	Ports	GPUs	Info
0	container_e02_1555157300368_0008_01_000002		http: 33392 ssh: 33393	#0	View SSH Info Stdout Stderr ...



- 点击PAI CLUSTER EXPLORER
- 当前集群中双击Open HDFS, 打开HDFS explorer。
- 下载训练好的模型数据



第十步：编写应用程序，完成推理

- 编写前端应用程序
- 调用模型进行推理
- 完成对联自动生成





<https://github.com/Microsoft/ai-edu>

创新程 AI 未来

“中国高校人工智能人才国际培养计划” 2019国际人工智能专家论坛
暨2019微软新一代人工智能开放科研教育平台合作论坛

THANKS