Provisioning Azure Data Science VMs

Bookmark this page

The Microsoft Data Science Virtual Machine is an Azure virtual machine (VM) image pre-installed and configured with several popular tools that are commonly used for data analytics and machine learning.

Before you can create a Microsoft Data Science Virtual Machine, you must have the following:

- An Azure subscription: To obtain one, see Get Azure free trial.
- An Azure storage account. Alternatively, the storage account can be created as part of the process of creating the VM if you do not want to use an existing storage account.

In this section, we will go through the provisioning of Azure Data Science VM based on Linux (Ubuntu), configuring CNTK and Jupyter notebook, and tunneling from your local machine to the notebook server.

Provisioning Azure Data Science Virtual Machine

1. Go to https://portal.azure.com and sign in using your account.
2. Click **New** and type **Data Science** on the Search box.
3. Select **Data Science Virtual Machine for Linux (Ubuntu)**.
4. Provide the requested details and click **OK**.
5. Select a size and click **Select**.
   If you have an Azure subscription (not a trial account), you can go for the NC series which has one or more GPU (they need to have HDD as the disk type).
6. Accept the default options and click **OK**.
7. Click **Purchase** to provision your VM.
8. When the VM is Provisioned, the VM dashboard will open.
9. Click Connect and note the command to connect to the VM using SSH. It will be something like `ssh username@123.123.123.123`.

Now you will connect to the VM to configure CNTK and run the Jupyter notebook server. You will then tunnel to the VM using the correct port to open the Jupyter notebook.

Connecting to the Azure Data Science VM using SSH

You can use any SSH client. In this example, we will use PuTTY, developed originally by Simon Tatham for the Windows platform.

1. Download PuTTY and run it.
2. Enter the Host Name. This is from the command to connect you noted earlier, minus the ssh in the beginning. It will be something like `username@123.123.123.123`.
3. Expand **SSH** in the **Category** and select **Tunnels**.
4. Enter **8000** for Source port, **localhost:8888** for Destination, and click **Add**.
5. Select **Session** in the **Category**, enter a name for **Saved Sessions** and click **Save**.
6. Click **Open** to connect to the VM and click Yes on the dialog box.
7. Enter the password.
8. You are now connected to the Azure Data Science VM you provisioned earlier.

Configuring CNTK on the Azure Data Science VM

1. On the VM command prompt, enter the following command:
   `source activate py35`
   This will activate the py35 environment.

2. On the py35 environment, enter the following command:
   `pip install <url>`
   Where `<url>` is the corresponding wheel (.whl) file to install CNTK.
   You must select the correct installation from the list here, and substitute `<url>` during the installation.

3. So for a GPU installation with Python 3.5, the command should be as follows:

```
pip install https://cntk.ai/PythonWheel/GPU/cntk-2.0-cp35-cp35m-linux_x86
_64.whl
```

Or if you need super user permission:

```
sudo /anaconda/envs/py35/bin/pip install https://cntk.ai/PythonWheel/GPU/
cntk-2.0-cp35-cp35m-linux_x86_64.whl
```

NOTE: You might need to upgrade pip before installing CNTK.

```
pip install --upgrade pip
```

Or if you need super user permission:

```
sudo /anaconda/envs/py35/bin/pip install --upgrade pip
```

Running Jupyter Notebook Server

1. Navigate to the local directory that you want to be the root folder for the Jupyter notebook.
2. Enter the following command:
   `jupyter notebook`
3. Type **A**, and then type **Q**. The notebook server is now running. Note the URL that is displayed. Now, let's connect from your local machine using a browser.
4. On your local machine, open a browser.
5. Enter the URL noted previously, replace the port 8888 with 8000. It will be something like `http://localhost:8000/?token=9cd2f689e8d8e289e63c38b5ac05166d5d906173c2aa98cb`.
6. You can open and run any notebook (.ipynb) file or create a new notebook. The notebook is running on the Azure Data Science VM, but you can view it from your local machine.
7. To create a new notebook, click **New** and select a language.
8. You can now run the newly created notebook.

Discussion

Hide Discussion

Topic: Mod1-3 Environment Setup / Provisioning Azure Data Science VMs

Add a Post

All Posts

Installation problem - version 8.1.2 to upgrade to 9.0.1

discussion posted 26 days ago by **darabaf**

I am getting "You are using pip version 8.1.2, however version 9.0.1 is available. You should consider upgrading via the 'pip install --upgrade pip' command." However, that command and without pip putting the URL of the other whl files and paths won't work. I don't know a) if I should ignore that remark b) if not, how to upgrade so that I don't get that warning. Any help appreciated.

This post is visible to everyone.

Add a Response

1 response

jonsan21 (Staff)
26 days ago

pip install --upgrade pip command will upgrade pip. You can run this prior to cntk installation.

I still get the same error, @josan21, when you say "prior to cntk installation", does that mean that I have somehow uninstall it? and could you let me know if there are specific steps I have to follow to overcome this?

posted 25 days ago by **darabaf**

no need to uninstall.

if the system asks you to upgrade "pip", then do so before you use "pip" to install cntk.

posted 22 days ago by **jonsan21** (Staff)

As I mentioned, @jonsan21, I am afraid I am still unable to do as you say. I get the following first: pip install --upgrade pip Collecting pip Using cached pip-9.0.1-py2.py3-none-any.whl Installing collected packages: pip Found existing installation: pip 8.1.2 Uninstalling pip-8.1.2: Then it goes into Exception handling with messages: Exception: Traceback (most recent call last): File "/anaconda/envs/py35/lib/python3.5/shutil.py", line 538, in move os.rename(src, real_dst) PermissionError: [Errno 13] Permission denied: '/anaconda/envs/py35/bin/pip' -> '/tmp/pip-khkzucc3-uninstall/anaconda/envs/py35/bin/pip'

During handling of the above exception, another exception occurred:

Traceback (most recent call last): File "/anaconda/envs/py35/lib/python3.5/site-packages/pip/basecommand.py", line 215, in main status = self.run(options, args) File "/anaconda/envs/py35/lib/python3.5/site-packages/pip/commands/install.py", line 317, in run prefix=options.prefix_path, File "/anaconda/envs/py35/lib/python3.5/site-packages/pip/req/req_set.py", line 736, in install requirement.uninstall(auto_confirm=True) File "/anaconda/envs/py35/lib/python3.5/site-packages/pip/req/req_install.py", line 742, in uninstall paths_to_remove.remove(auto_confirm) File "/anaconda/envs/py35/lib/python3.5/site-packages/pip/req/req_uninstall.py", line 115, in remove renames(path, new_path) File "/anaconda/envs/py35/lib/python3.5/site-packages/pip/utils/**init**.py", line 267, in renames shutil.move(old, new) File "/anaconda/envs/py35/lib/python3.5/shutil.py", line 553, in move os.unlink(src) PermissionError: [Errno 13] Permission denied: '/anaconda/envs/py35/bin/pip' And finally says: You are using pip version 8.1.2, however version 9.0.1 is available. You should consider upgrading via the 'pip install --upgrade pip' command.

posted 22 days ago by **darabaf**

seems like you dont have permission to do what you need to do.. try using sudo?

If you're not familiar with linux, you can also use the Windows based Data Science VM.

Alternatively, you can also use the Azure Notebook service.

posted 22 days ago by **jonsan21** (Staff)

OK I got it now, The full command should be like follows as the pip is not in the PATH env. variable: sudo env "PATH=$PATH" pip install --upgrade pip All is set then. Thanks for the hint @jonsan21

posted 21 days ago by **darabaf**

Add a comment

Showing all responses

Add a response:

Preview

Submit

English