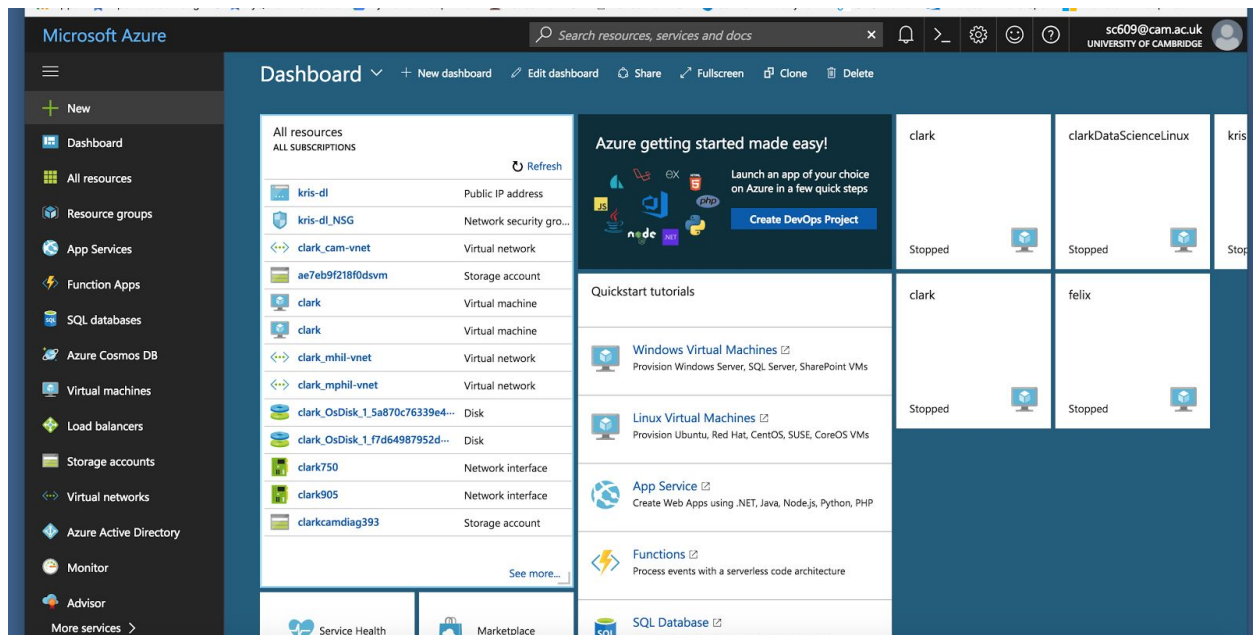


# How to Create an Individual Data Science VM

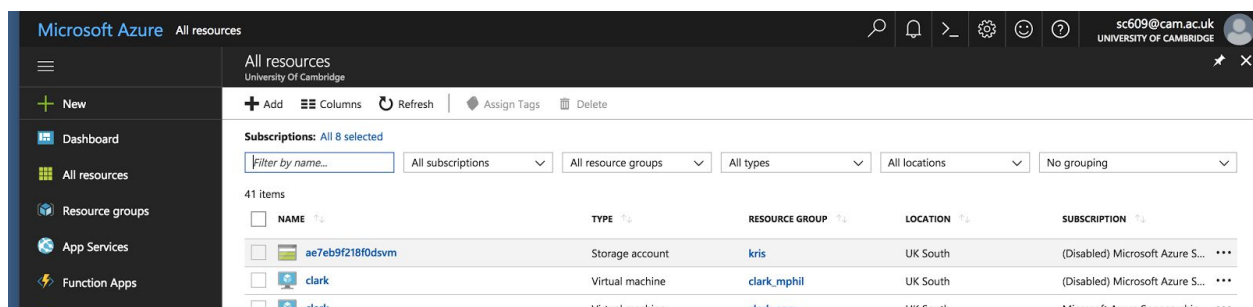
Log in to Azure

Go to [portal.azure.com](https://portal.azure.com) and log in via your institutional account, i.e. via Cambridge and your usual CRSid email and password. If you are having trouble logging in, you may need to enable your University Microsoft account first: <https://www.uis.cam.ac.uk/withdrawn/ees/ees-sign-up>

After successfully logging in, you should see a page which looks something like this (but without the VM boxes on the right):

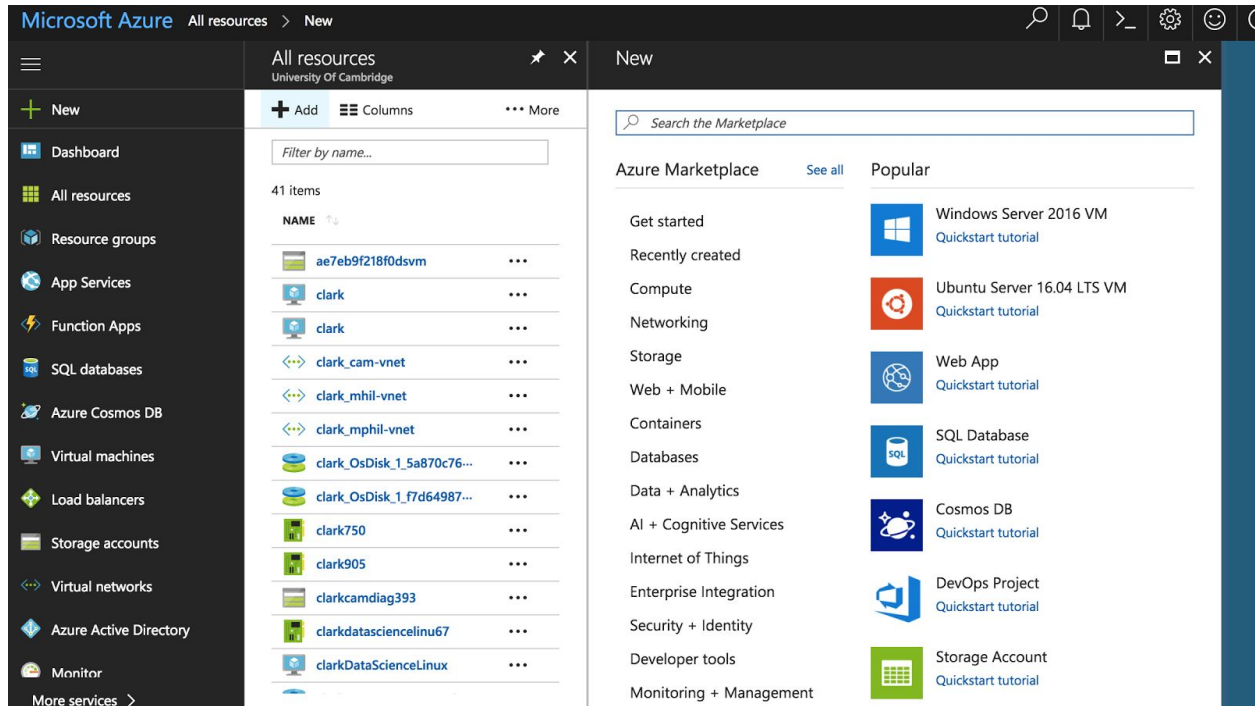


If you click on ALL SUBSCRIPTIONS (under All Resources under Dashboard, near to the top in second column from left) you should see a page that looks something like this:

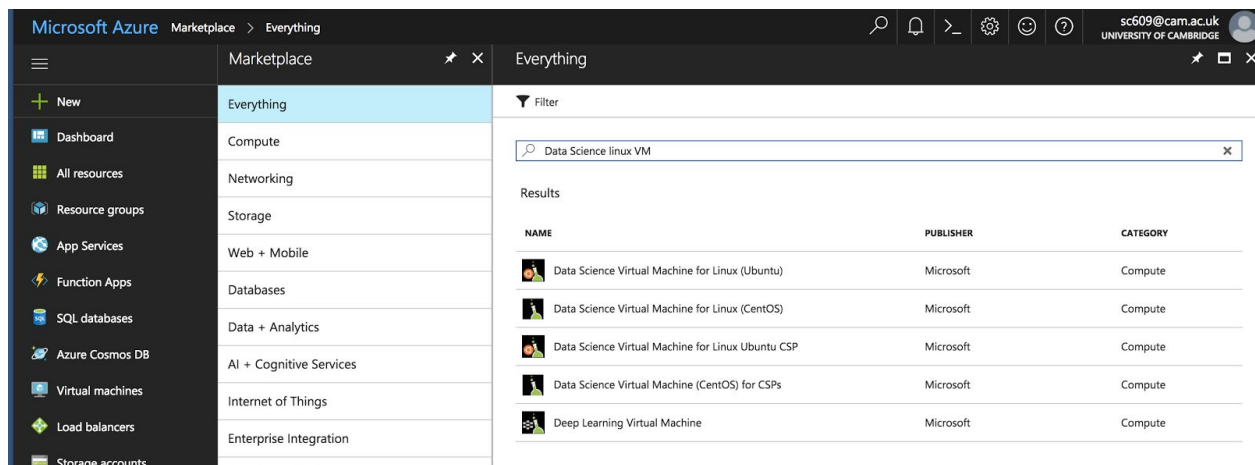


## Creating the VM

Now click on the +Add button, to the right of + New and to the left of Columns, in the bar near the top. There should now be a new panel that allows you to search the Azure marketplace:



In the search toolbar, type “Data Science Linux VM”:



Now click on Data Science Virtual Machine for Linux (Ubuntu). (Be sure to choose the right option.) The description of the VM should be as follows:

Microsoft Azure Marketplace > Everything > Data Science Virtual Machine for Linux (Ubuntu)

sc609@cam.ac.uk UNIVERSITY OF CAMBRIDGE

Data Science Virtual Machine for Linux (Ubuntu)

The Data Science Virtual Machine for Linux is an Ubuntu-based virtual machine image that makes it easy to get started with deep learning on Azure. The Microsoft Cognitive Toolkit, TensorFlow, MXNet, Caffe, Caffe2, Chainer, DIGITS, H2O, Keras, Theano, Torch, and PyTorch are built, installed, and configured so they are ready to run immediately. The NVIDIA driver, CUDA 8, and cuDNN 6 are also included. All frameworks are the GPU versions but work on the CPU as well. Many sample Jupyter notebooks are included.

The Data Science Virtual Machine for Linux also contains popular tools for data science and development activities, including:

- Microsoft R Server 9.2.1 with Microsoft R Open 3.4.1, MicrosoftML package with machine learning algorithms, RevoScaleR and revoscalepy for distributed and remote computing, and R and Python Operationalization
- Anaconda Python 2.7 and 3.5
- JupyterHub with sample notebooks
- [Apache Drill](#) for querying non-relational data using SQL
- Spark local 2.2.0 with PySpark and SparkR Jupyter kernels
- Single node local Hadoop
- Azure command-line interface
- Visual Studio Code, IntelliJ IDEA, PyCharm, and Atom
- H2O, Deep Water, and Sparkling Water
- Julia
- Vowpal Wabbit for online learning
- xgboost for gradient boosting
- SQL Server 2017
- Intel Math Kernel Library

Select a deployment model

Resource Manager

Create

	PUBLISHER	CATEGORY
(Ubuntu)	Microsoft	Compute
CentOS)	Microsoft	Compute
Ubuntu CSP	Microsoft	Compute
or CSPs	Microsoft	Compute
	Microsoft	Compute

Note that this particular VM has Python and Tensorflow installed, which we will be using for the practical. Hit the blue Create button at the bottom.

## Configure your VM

There are now 4 stages to go through to configure the Data Science VM.

Microsoft Azure Marketplace > Everything > Data Science Virtual Machine for Linux (Ubuntu) > Create virtual machine > Basics

New

Dashboard

All resources

Resource groups

App Services

Function Apps

SQL databases

Azure Cosmos DB

Virtual machines

Load balancers

Storage accounts

Virtual networks

Azure Active Directory

Monitor

Advisor

Create virtual machine

1 Basics  
Configure basic settings

2 Size  
Choose virtual machine size

3 Settings  
Configure optional features

4 Summary  
Data Science Virtual Machine f...

Basics

\* Name

VM disk type ⓘ  
SSD

\* User name

\* Authentication type  
SSH public key Password

\* SSH public key ⓘ

Subscription  
Amdla Mabona

\* Resource group ⓘ  
☐ Create new ☐ Use existing  
  
The value should not be empty.

\* Location  
UK South

Fill in the options as follows:

Microsoft Azure Marketplace > Everything > Data Science Virtual Machine for Linux (Ubuntu) > Create virtual machine > Basics

Create virtual machine

Basics

- 1 Basics  
Configure basic settings
- 2 Size  
Choose virtual machine size
- 3 Settings  
Configure optional features
- 4 Summary  
Data Science Virtual Machine f...

\* Name  
StephenClarkTest ✓

VM disk type ⓘ  
HDD ▾

\* User name  
sc609 ✓

\* Authentication type  
SSH public key Password

\* Password  
..... ✓

\* Confirm password  
..... ✓

Subscription  
Stephen Clark ▾

\* Resource group ⓘ  
☒ Create new ☐ Use existing  
MPhilACS ✓

\* Location  
UK South ▾

OK

<https://portal.azure.com/#blade/HubsExtension/Resources/resourceType/Microsoft.Resources%2Fsubscriptions%2FresourceGroups>

Give the VM a sensible name. Be sure to choose HDD for the VM disk type. Use your CRSid username. Choose the password option, and set a password. Pick your own subscription. Create a new resource group, call it something sensible. Choose UK South for the location. Hit the blue OK button.

Now you choose a particular VM. Choose NC6 Standard:

Microsoft Azure Marketplace > Everything > Data Science Virtual Machine for Linux (Ubuntu) > Create virtual machine > Choose a size

### Create virtual machine

1 Basics  
Done

2 Size  
Choose virtual machine size

3 Settings  
Configure optional features

4 Summary  
Data Science Virtual Machine f...

### Choose a size

Browse the available sizes and their features

Prices presented are estimates in your local currency that include Azure infrastructure applicable software costs, as well as any discounts for the subscription and location. Recommended sizes are determined by the publisher of the selected image based on hardware and software requirements.

Supported disk type: HDD | Minimum vCPUs: 1 | Minimum memory (GiB): 0

★ Recommended | [View all](#)

NC6 Standard	NC12 Standard	NC24 Standard
6 vCPUs	12 vCPUs	24 vCPUs
56 GB	112 GB	224 GB
24 Data disks	48 Data disks	64 Data disks
8x500 Max IOPS	16x500 Max IOPS	32x500 Max IOPS
380 GB Local SSD	680 GB Local SSD	1440 GB Local SSD
1x K80 Graphics	2x K80 Graphics	4x K80 Graphics
Load balancing	Load balancing	Load balancing
711.44 GBP/MONTH (ESTIMATED)	1,422.87 GBP/MONTH (ESTIMATED)	2,846.30 GBP/MONTH (ESTIMATED)

On the next screen, just hit the OK button:

Microsoft Azure Marketplace > Everything > Data Science Virtual Machine for Linux (Ubuntu) > Create virtual machine > Settings

New

Dashboard

All resources

Resource groups

App Services

Function Apps

SQL databases

Azure Cosmos DB

Virtual machines

Load balancers

Storage accounts

Virtual networks

Azure Active Directory

Monitor

Advisor

More services >

Create virtual machine

1 Basics Done ✓

2 Size Done ✓

3 Settings Configure optional features >

4 Summary Data Science Virtual Machine f... >

Settings

High availability

\* Availability set ⓘ > None

Storage

Use managed disks ⓘ

No Yes

Network

\* Virtual network ⓘ > (new) MPhilACS-vnet

\* Subnet ⓘ > default (10.0.0.0/24)

\* Public IP address ⓘ > (new) StephenClarkTest-ip

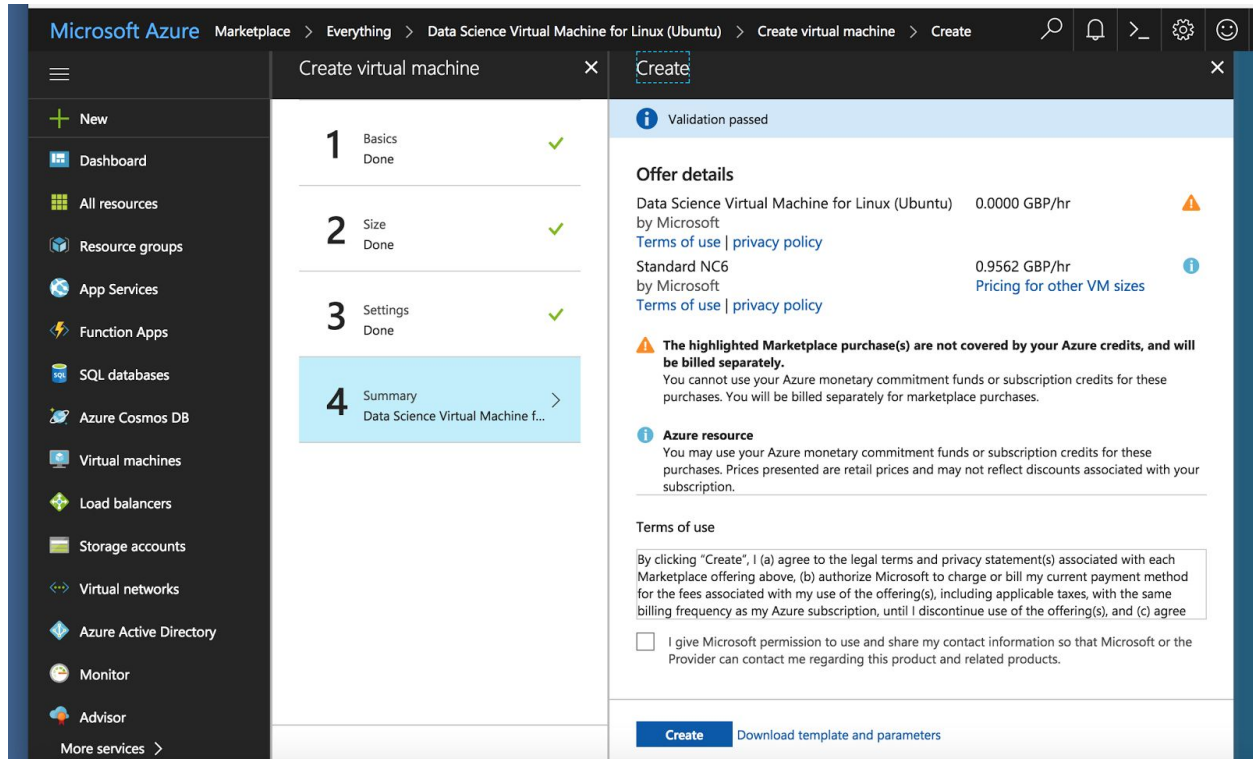
\* Network security group (firewall) ⓘ > (new) StephenClarkTest-nsg

Extensions

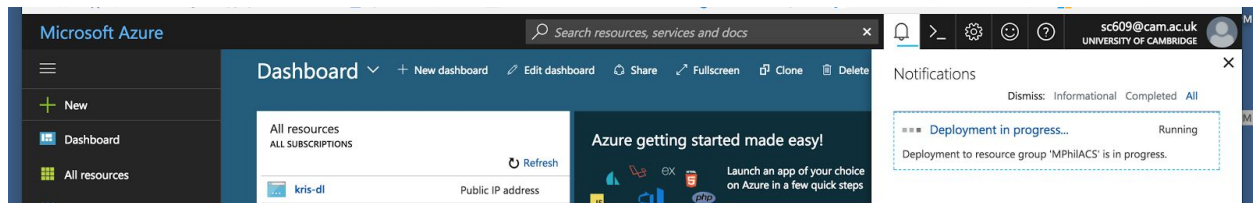
Extensions ⓘ > No extensions

OK

Finally hit the blue Create button, and you're done:



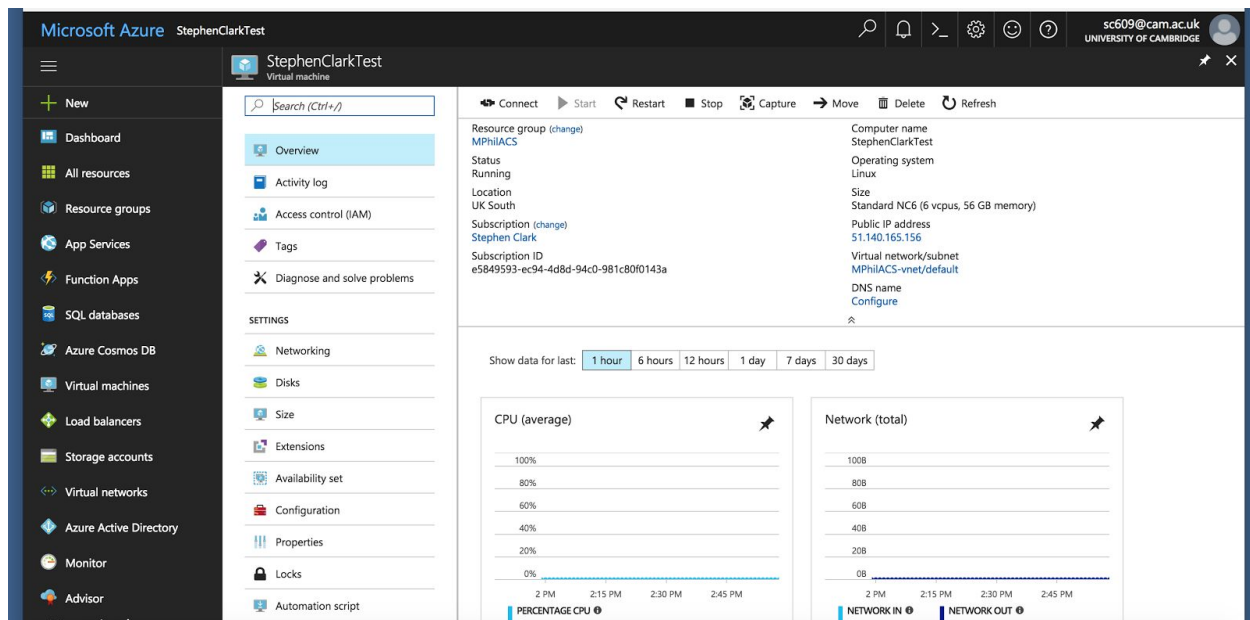
Creating the deployment will take a few minutes:



Starting the VM

You should now see a screen that looks like this:



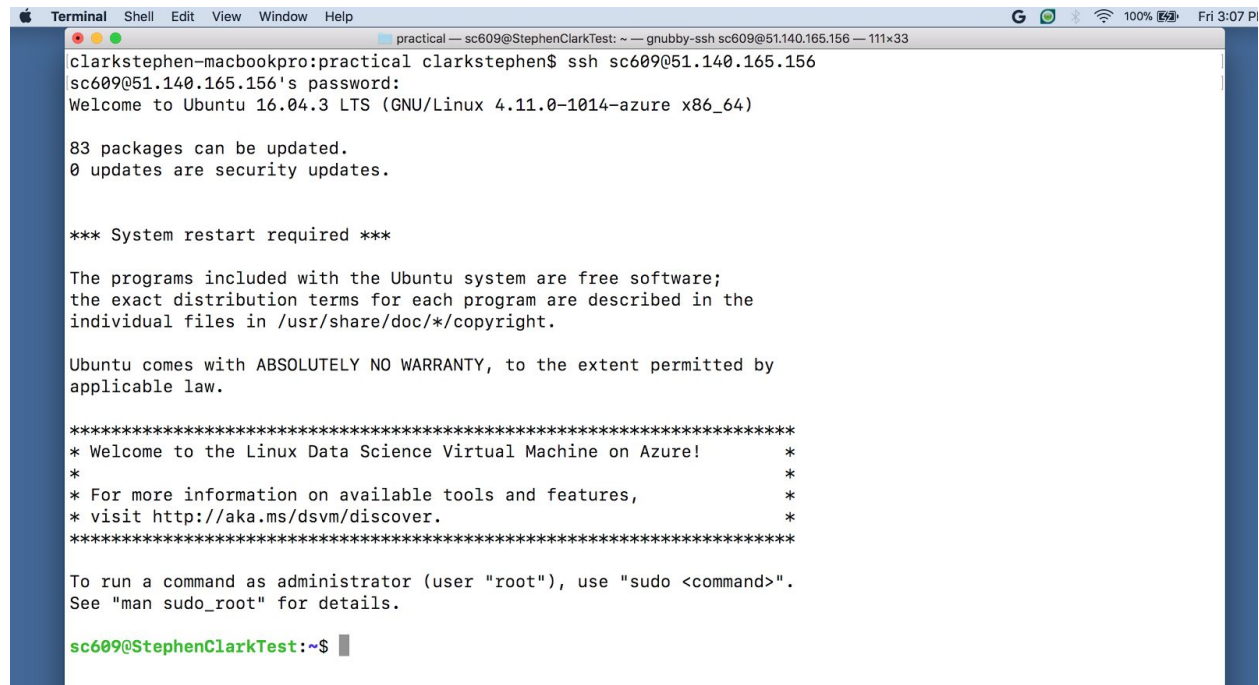


If you hit the start button in the bar at the top, this will fire up the VM (if you've just configured it, it may have been started up automatically). If you hit the connect button to the left of start, it will give you the ssh command that you can use from a terminal to ssh into your VM. Use the password you set up during the configuration stage to ssh in. (If you're not seeing the start or connect buttons as available, try a refresh on your browser.)

## Important

Remember to hit the stop button when you're done with a session. Otherwise the VM will keep churning away, eating up your subscription.

## SSH into the VM



A terminal window titled "practical — sc609@StephenClarkTest: ~ — gnuTTY-ssh sc609@51.140.165.156 — 111x33". The user "clarkstephen-macbookpro:practical clarkstephen" has executed "ssh sc609@51.140.165.156". The prompt is "sc609@51.140.165.156's password:". After entering the password, the user is welcomed to "Ubuntu 16.04.3 LTS (GNU/Linux 4.11.0-1014-azure x86\_64)". The terminal shows "83 packages can be updated. 0 updates are security updates." and a message "\*\*\* System restart required \*\*\*". It also displays the Ubuntu license and a welcome message for the "Linux Data Science Virtual Machine on Azure". The prompt is now "sc609@StephenClarkTest:~\$".

```
clarkstephen-macbookpro:practical clarkstephen$ ssh sc609@51.140.165.156
sc609@51.140.165.156's password:
Welcome to Ubuntu 16.04.3 LTS (GNU/Linux 4.11.0-1014-azure x86_64)

83 packages can be updated.
0 updates are security updates.

*** System restart required ***

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

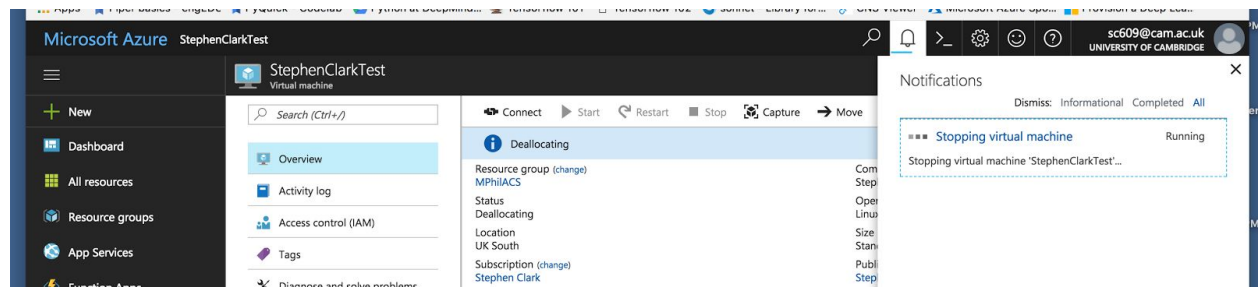
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

*****
* Welcome to the Linux Data Science Virtual Machine on Azure!          *
*                                                                      *
* For more information on available tools and features,                *
* visit http://aka.ms/dsvm/discover.                                  *
*****

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

sc609@StephenClarkTest:~$
```

Now you're in. Follow the instructions on the second instructions sheet in order to download the data and start the practical. And remember to stop when you're done:



## Checking your Balance

You can check your subscription balance at <https://www.microsoftazuresponsorships.com/Balance>  
You have \$400 to play with - use it wisely!