## Setting Up Your Hortonworks Hadoop Sandbox

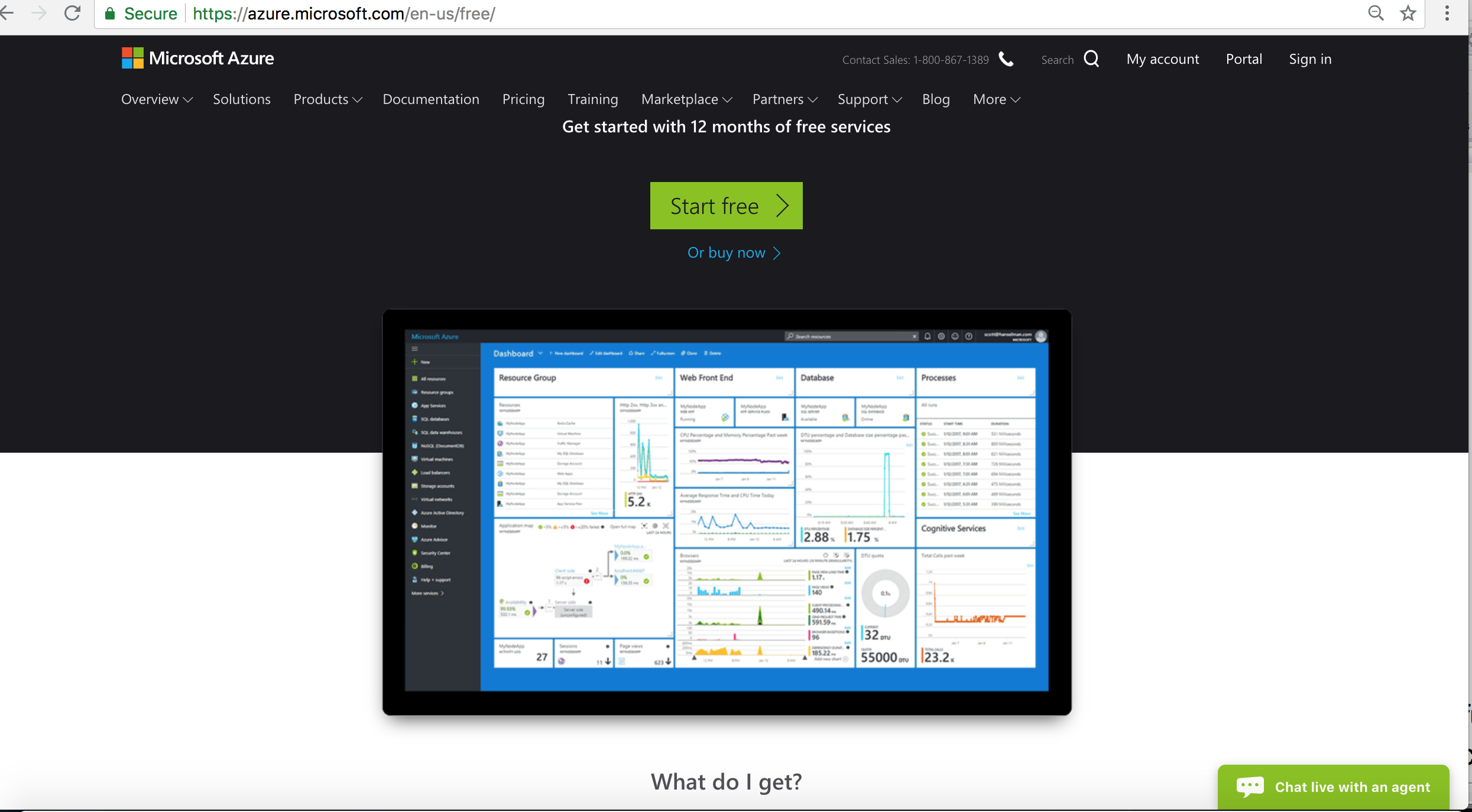
Setting up a Hadoop sandbox environment of your own (provided by Hortonworks) on the Microsoft Azure cloud is a necessary prelude for much of the remainder of this course. While some of this process may seem unfamiliar to you, the instructions Hortonworks provides are quite detailed. Moreover, if you have completed assignment #2 you already have the tools and most of the configuration available to connect to Azure.

Before you proceed to the Hortonworks documentation, however, it is important that you carefully read and apply the below supplementary notes. If you encounter any issues with the following material check the Hortonworks instructions. But feel free to contact me for help.

### Step 1) Establish a Microsoft Azure cloud account

Use an existing Azure account or establish a new one as follows:

1. Go to: <https://azure.microsoft.com/en-us/free/>

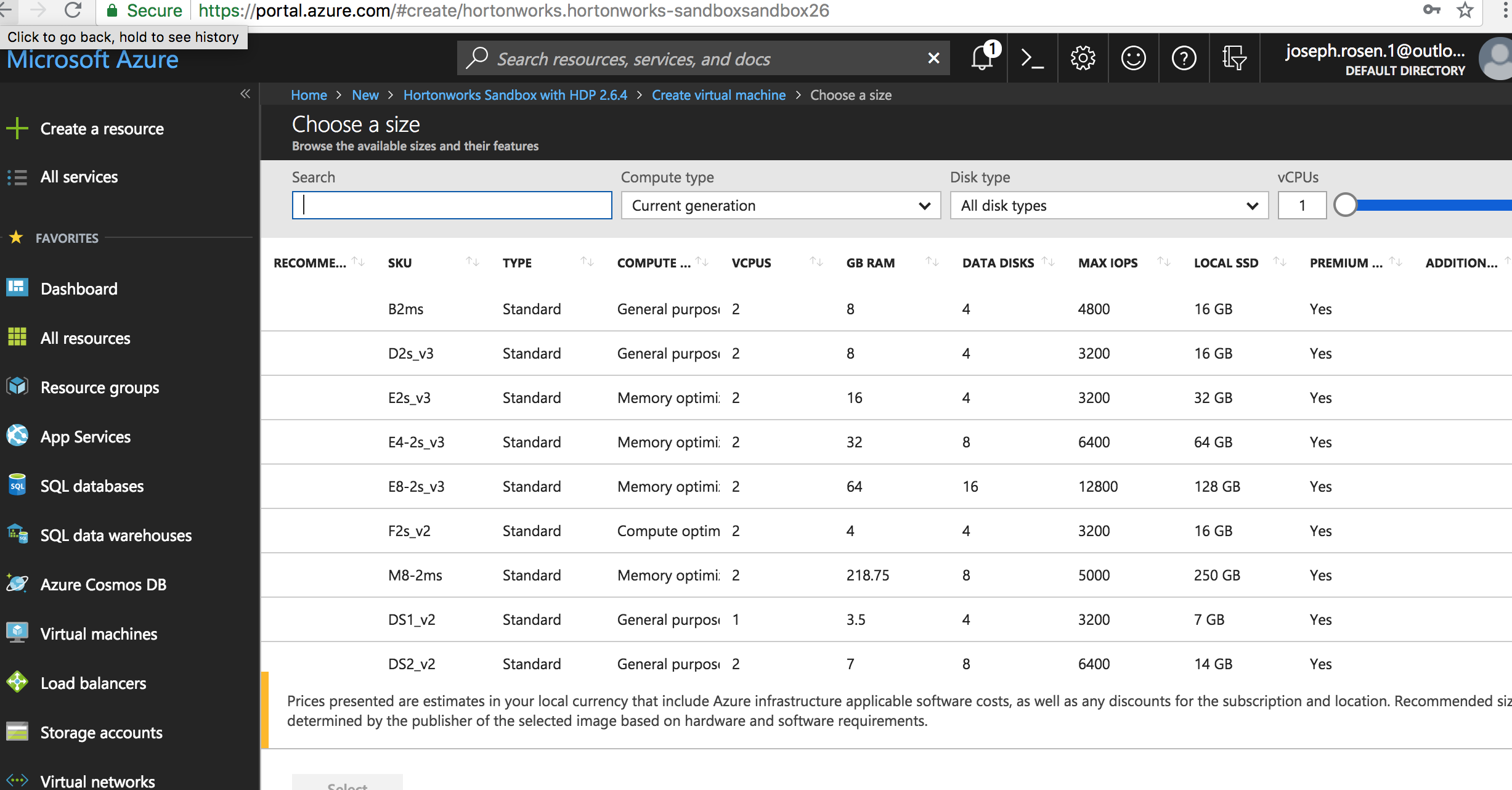


1. Set up a new account as per instructions. You will need a credit card. If you are careful the first 30 days of use will be free. The following months of our course should each cost some small amount (perhaps $10 per month) AS LONG AS YOU STOP THE SANDBOX VIRTUAL MACHINE (VM) WHEN IT IS NOT IN USE. Instructions for starting and stopping your sandbox are provided below.

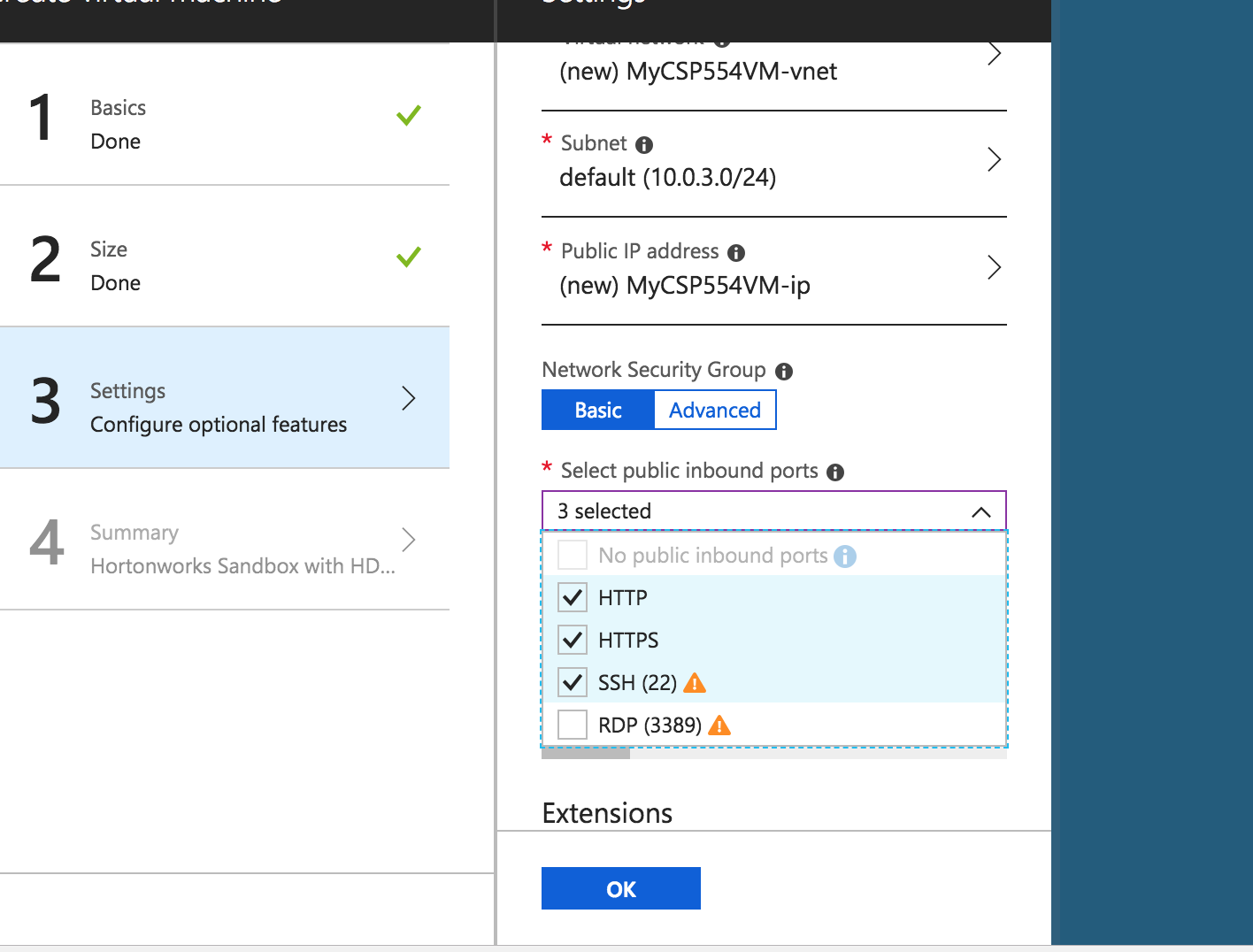
### Step 2: Read and apply these additions to the Hortonworks instructions

And don’t look at those instructions (I will tell you where to find them soon) until after you read the below.

1. In the Hortonworks instructions you will encounter a section called “SIZE FORM,” the screen image in the Hortonworks instructions may differ from the one with which you are presented. The screen could appear as follows:



1. In the section called “SIZE FORM,” **do not** select a cheaper but less powerful machine such as “A5 standard.” If you try your sandbox might never start or take an hour to start. Rather, look for a machine with 2 Cores or more, 14 Gb of RAM or more, and 28GB or more of local disk. This could be “D3\_V2, D11\_V2, D12\_V2” or similar.
2. In the section called “SETTINGS FORM,” set the public inbound ports as indicated below to allow HTTP, HTTPS and SSH:



1. Once the offer is submitted by selecting **Purchase**, the sandbox will take several minutes to set up and deploy. But even after the sandbox VM appears to have started, **the Hadoop software takes an additional 10 minutes to initialize** and respond to requests to log on and accept commands. Each time you stop and then start the sandbox, it will again take up to 10 minutes for the Hadoop software to initialize. If you try to connect to a sandbox, via SSH, before the Hadoop software is fully initialized your logon attempt may not be acknowledged or may return an error. So be patient.
2. In the Hortonworks instructions you will encounter a section called “Using SSH,” if you have completed assignment #2 then you have a config file in your “.ssh” directory. Save a copy in case of emergency to “config.old” Now edit the original “config” as indicated in the Hortonworks instructions. Ignore the section “USING PUTTY.” Now you can access your Hadoop sandbox as before by establishing an ssh session to the maria\_dev account (remember to set up SSH tunneling first in a separate terminal window by entering ‘ssh azureSandbox’).

### Step 3: Now follow the Hortonworks sandbox installation instructions (as modified by the above notes)

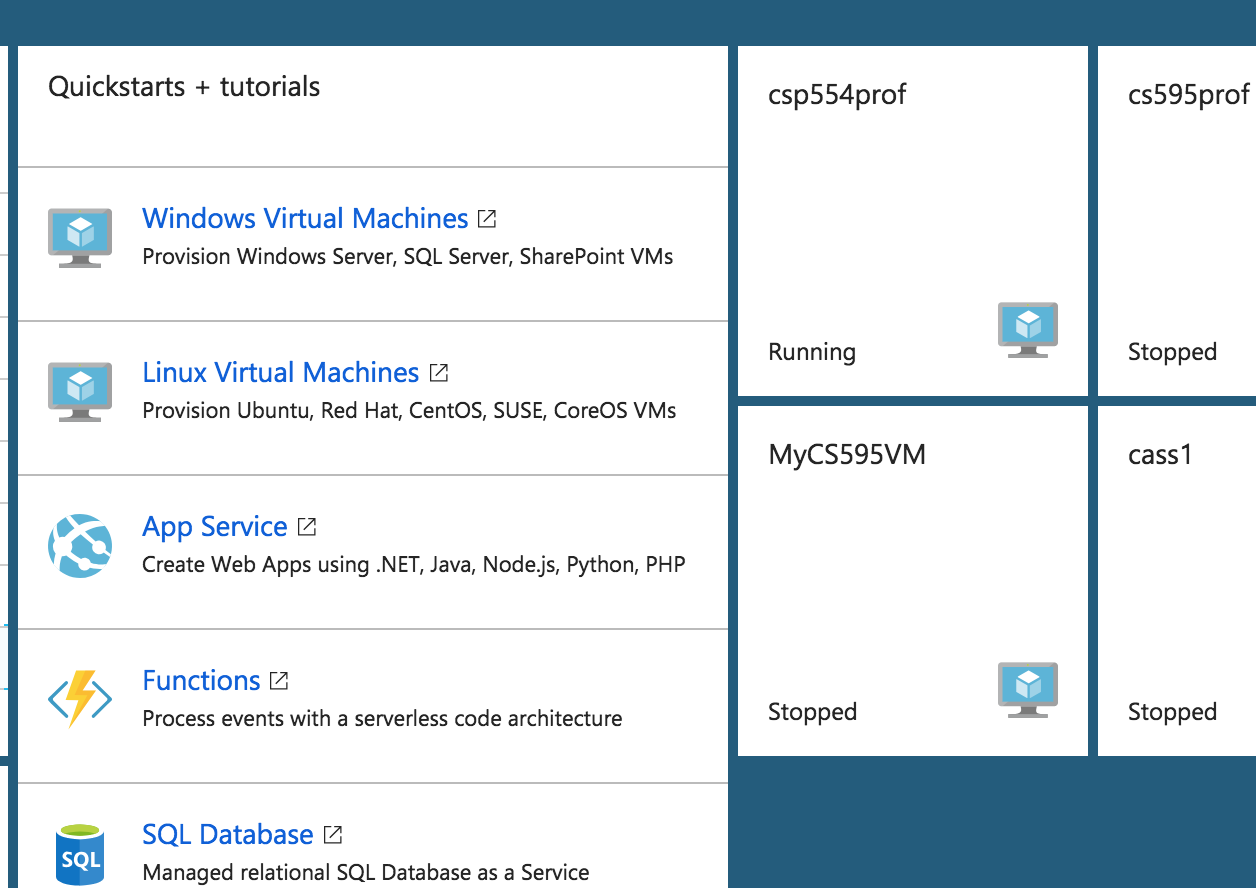
They are on the Blackboard in the “Free Books and Chapters” section as “Deploying Hortonworks Sandbox on Microsoft Azure” or go directly to:

<https://hortonworks.com/tutorial/sandbox-deployment-and-install-guide/section/4/>

### Step 4: Stopping and starting your sandbox

Never let you sandbox run overnight! Always stop it after you complete your assignments and then start it again as needed (recall it takes 10 minutes to initialize).

Click on the tile associated with your sandbox similar to the below, in this example, the one called ‘cs554prof:



You should then see a form like the following. At the top of the form select either “Stop” to stop your sandbox or “Start” to start it again:

