

## **Assignment 0**

Version 0.10

### Cloud Computing Infrastructure Tutorial

Due Date: Friday February 1<sup>st</sup>, 2019 @ 11:59 pm, tentative

#### **Objective**

The purpose of assignment 0 is to establish AWS accounts, and gain experience with technologies used to provide distributed computing infrastructure to support future TCSS 558 programming assignments.

We will leverage the AWS Educate program for education credits from Amazon Web Services (AWS) to provide cloud computing resources for TCSS 558 projects. We will create virtual machines, known as elastic compute cloud (EC2) instances to host individual nodes of our distributed systems. To support working with VMs to host our distributed applications, we will harness the Docker-Machine tool to automatically create and configure VMs. We will use Docker containers then to deploy code (our nodes) onto VMs.

Assignment 0 provides a tutorial on the use of Cloud Computing Infrastructure. Specifically assignment 0 walks through the use of EC2 instances, docker, docker-machine, and haproxy for load balancing.

#### **Use of a Linux environment is recommended for assignment 0.**

For Windows 10 users, there is now a Ubuntu “App” that can be installed onto Windows 10 directly. This provides a Ubuntu Linux environment without the use of Oracle Virtualbox. Alternatively, Windows users can install Oracle Virtual Box to enable creating virtual machines under Windows 10, and then install a Ubuntu 18.04 virtual machine.

Windows 10 Ubuntu “App” instructions:

[https://msdn.microsoft.com/en-us/commandline/wsl/install\\_guide](https://msdn.microsoft.com/en-us/commandline/wsl/install_guide)

Windows Oracle Virtual Box & Ubuntu VM instructions:

There are a number of blogs and YouTube videos that walk through installing Oracle VirtualBox on Windows 10, and how to then install Ubuntu 18.04 LTS on Virtual Box. Search using google.com or video.google.com to find blogs and/or videos to help.

Oracle VirtualBox can be downloaded from:

<https://www.virtualbox.org/wiki/Downloads>

## Task 0 - Getting an AWS account

### **DO NOT SELECT THE “STARTER PACK” OPTION WHEN APPLYING FOR AN AWS ACCOUNT.**

If you do not presently have an AWS account, the best option that offers the largest pool of free credits is to apply for the GitHub Student Developer pack. This program provides up to \$150 in usage credits:

Apply using your UW email id:

<https://education.github.com/pack>

If you already have an AWS account created on your own, not using a UW email, then try applying for the GitHub program as a new user under your UW email, and creating a new “uw.edu” AWS account.

If you already have an AWS account created using your UW email, either uw.edu, u.washington.edu, or washington.edu then you may try to apply for a new account using the “other” domain name: u.washington.edu, uw.edu, or washington.edu. Occasionally GitHub student developer pack applications are denied because student status can not be confirmed. In some cases using an alternate ID has been sufficient to resolve the issue.

If denied AWS credits through the GitHub program, alternatively you can try to apply directly to AWS educate using your UW email. This program provides up to \$75 in credits. **DO NOT SELECT THE “STARTER PACK” OPTION WHEN APPLYING FOR AN AWS ACCOUNT.** The starter pack provides a limited AWS account that is designed for online tutorials, and has restricted access to many AWS services.

**awseducate.com**

If this doesn't work, contact the instructor. Provide your AWS account ID (if available), and your UW email address an account was created with. The instructor will follow-up with the UW AWS Educate representative as soon as possible. **Please note it may take a few days to receive a response from Amazon. Please contact the instructor ASAP if needing credits for an existing AWS account.** If credits are not available, it is possible to complete assignment 0 using only free tier resources (e.g. t2.micro instances).

---

## Task 1 - AWS account setup

Once having access to AWS, task #1 involves creating AWS account credentials to work with Docker-Machine, if you have not already done so.

From the AWS services home page, locate the “IAM” Identity Access Management link, and select it:



## Security, Identity & Compliance

IAM

Inspector

Once in the IAM dashboard, on the left hand-side select “Users”:

- Dashboard
- Groups
- Users**
- Roles
- Policies
- Identity providers
- Account settings
- Credential report

Provide a user account name. Here I am using “TCSS558” as an example:

### Add user

1

Detail

#### Set user details

You can add multiple users at once with the same access type and permissions. [Learn more](#)

User name\*

tcss558

[+ Add another user](#)

#### Select AWS access type

Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. [Learn more](#)

Access type\*

☒ **Programmatic access**  
Enables an **access key ID** and **secret access key** for the AWS API, CLI, SDK, and other development tools.

☐ **AWS Management Console access**  
Enables a **password** that allows users to sign-in to the AWS Management Console.

Be sure to select the “Programmatic access” checkbox.

Then click the “Next: Permissions” button...

For simplicity, you can simply select the button:



Using the search box, search, find, and select using the checkbox the following policy:

\* AmazonEC2FullAccess

If you plan to use this user account to explore additional Amazon's services, then I recommend also adding:

\* AdministratorAccess

This will allow you, via the CLI, to explore and do just about everything with this AWS account.

Now click the "Next: Review" button, and then select "Create user".

You'll now see a screen with an Access key ID (grayed out below), and a Secret access key. You can copy both the Access key, and the secret access key to a safe place, or alternatively, click the "Download .csv" button to download a file containing this information.

	User	Access key ID	Secret access key
▶	✓ tcsc558	AKIAI44QH8DHBEXAMPLE	***** Show

Once you've downloaded these keys, be sure to **never** publish these key values in a source code repository such as github where your account credentials could be exposed. **Protect these keys as if they were your credit card or wallet!**

## **Task 2 - Working with Docker, creating Dockerfile for Apache Tomcat**

Next, let's launch a virtual machine on Amazon to support working with Docker/Docker-Machine. You will want to have access to a computer with the ssh/sftp tools. It is best to have access to a local computer with Ubuntu installed either natively, or on Oracle Virtualbox. It is possible to download putty, an "SSH" client and also an "SFTP" client, for Windows, but not recommended.

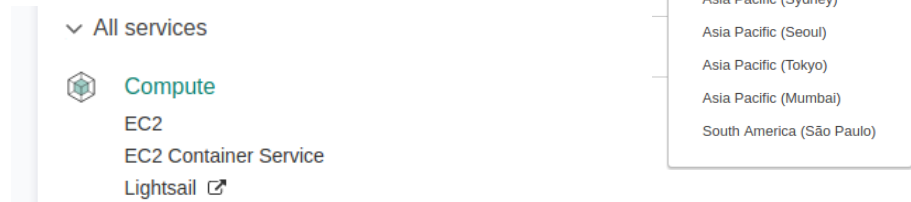
First, let's choose the "region" that you'll work in. Recommended options are "US East (N. Virginia)" known as "us-east-1" via the CLI, or "US West (Oregon)" known as "us-west-2".

The region can be set using the dropdown in the upper-right hand corner. Selecting the region configures the entire AWS console to operate in that region as shown to the right

----->

For assignment 0, we will use “t2.micro” instances. Every user is allowed up to 750-hours/month of instance time for FREE using the t2.micro type.

From the AWS menu, under Compute services, select “EC2”:



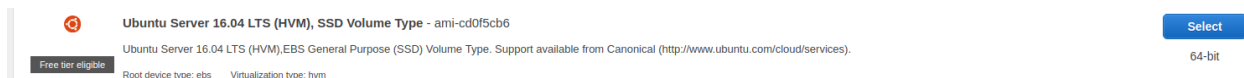
Next, click the Launch Instance button:

### Create Instance

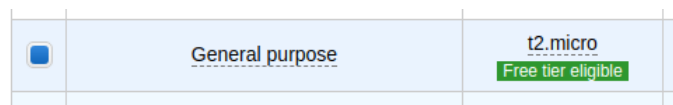
To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

[Launch Instance](#)

Select Ubuntu:



Specify t2.micro as the instance type, and click the “Next: Configure Instance Details” button,



Next, specify the following instance details:

### Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot

Number of instances	<input type="text" value="1"/>	<a href="#">Launch into Auto Scaling Group</a>
Purchasing option	<input type="checkbox"/> Request Spot instances	
Network	<input type="text" value="vpc-4f8db929 (default)"/>	<a href="#">Create new VPC</a>
Subnet	<input type="text" value="No preference (default subnet in any Availability Zone)"/>	<a href="#">Create new subnet</a>
Auto-assign Public IP	<input type="text" value="Use subnet setting (Enable)"/>	
IAM role	<input type="text" value="None"/>	<a href="#">Create new role</a>
Shutdown behavior	<input type="text" value="Stop"/>	

Network: choose “(default)” for the Virtual Private Cloud (VPC).

Subnet: choose an availability zone such as us-east-1e

Auto-assign Public IP: choose “Use subnet setting (Enable)”. This will provide a public IP address to enable connecting to your instance.

Shutdown behavior: Choose “Stop”

Next, click “Next: Add Storage”.

Then, keep defaults and click “Next: Add Tags”.

Then, keep defaults and click “Next: Configure Security Group”.

Choose the option:

☒ Select an existing security group

And then mark the option for “default VPC security group”.

As we go along, apply all security changes to the default security group for your default VPC. This way the rule changes will persist as you come back to AWS for future work sessions.

Then click “Review and Launch”.

Review the details and if everything looks ok, click “Launch”.

The very first time you’ll be prompted to create a new RSA private/public keypair to enable logging into your instance.

Select an existing key pair or create a new key pair

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about [removing existing key pairs from a public AMI](#).

Create a new key pair

Key pair name

Download Key Pair

You have to download the **private key file** (\*.pem file) before you can continue. **Store it in a secure and accessible location.** You will not be able to download the file again after it's created.

Cancel

Launch Instances

The instance should launch and be visible by clicking “Instances” on the left-hand side of the EC2 Dashboard. Locate the IPv4 Public IP:

**Throughout the tutorial, Linux commands are prefaced with the “\$”.**

**Comments are prefaced with a “#”.**

First, from the Linux CLI change permissions on your keyfile:

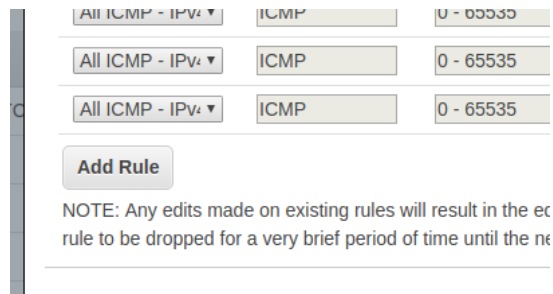
```
$chmod 0600 <key_file_name>.pem
```

Before you can SSH into the instance, the default security group used by your instance must be modified to allow SSH (port 22) access from your computer.

In the Amazon management console., under instances, look at the detailed instance information and click on “**default**” next to “Security groups”:

Availability zone	us-east-
Security groups	<b>default.</b>
Scheduled events	No sche

Click the “Inbound” tab, and then the “Edit” button.  
Scroll down and click the “Add Rule” button at the bottom of the dialog box:



Add a “SSH” Rule with the following settings:

Protocol = TCP

Port Range = *will automatically be set to 22*

Source = My IP

Then “Save” the security change.

Then connect using ssh:

```
$ssh -i <key_file_name>.pem ubuntu@<IPv4 Public IP>
```

Say yes, when the following message is displayed:

```
The authenticity of host '107.21.193.159 (107.21.193.159)' can't be
established.
ECDSA key fingerprint is SHA256:0cy2eP8Q15zmBThAqTq9z1Tw00+MS0ldKi1SmPZhkE0.
Are you sure you want to continue connecting (yes/no)? Yes
```

Note, the actual IP address will be different than “107.21.193.159”.

Linux tracks every machine you ever ssh to. The very first time it hashes the public key and places it into a file at /home/<user\_user\_id>/.ssh/known\_hosts

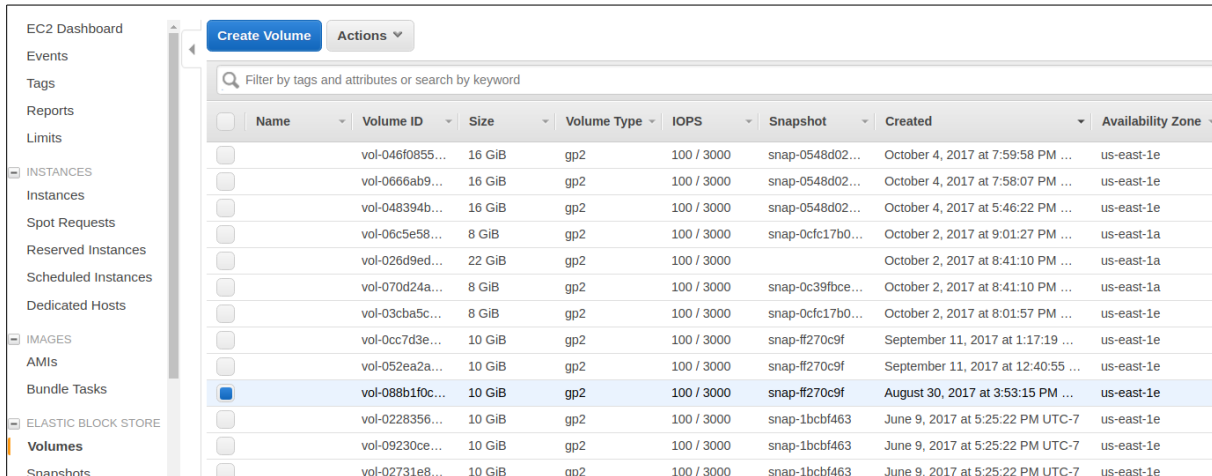
The idea is when you reconnect to the VM again, there is a possibility that someone masquerades as the VM. To prevent someone from masquerading as the VM you’re trying to connect to, ssh tracks the identity of each host and alerts the user **every time** there is a change. Sometimes the changes are expected, such as when you launch a new VM to replace an old one. The idea is to help notify the user if the VM’s identity changes unexpectedly.

## Stopping, and backing up your VM on Amazon:

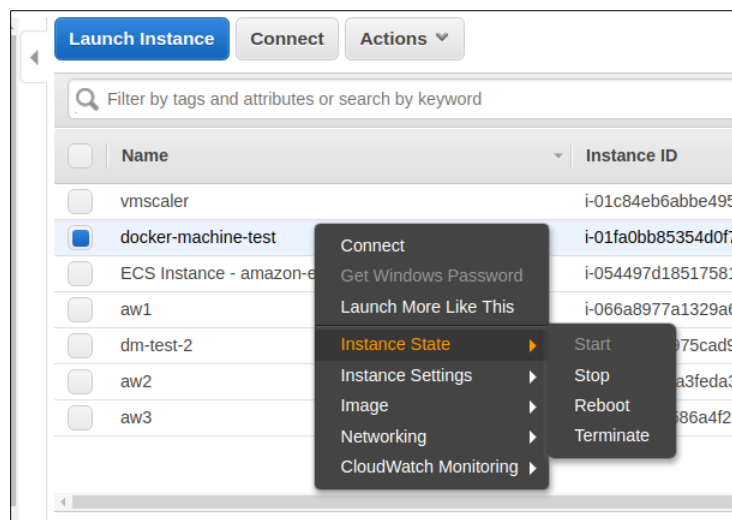
By default, the t2.micro is an “EBS-backed” instance. The t2.micro instances make use of remotely hosted elastic blockstore virtual hard disks for their “/root” volume. “EBS-backed” instances can be paused at any time. This allows you to stop your work, and come back later. Billing is paused, but storage charges for your EBS disk are ongoing 24/7. Every user is allowed 30GB of EBS disk space for free (free tier 1<sup>st</sup> year only). Beyond this, the price for storage is 10 cents per GB, per month for standard “GP2-General Purpose 2” EBS storage. A second 30GB (total of 60GB) will cost \$36/year in credits. In the console, any volumes listed under “Elastic Block Store | Volumes” will count towards this 30GB quota. Snapshots, under “Elastic



Block Store” represent compressed copies of EBS volumes that are stored using Amazon Simple Storage Service (S3), aka blob storage. Standard pricing for S3 storage is 2.3 cents per month per GB. If not using a VM for a considerable time, a cost effective way to preserve the data is to “snapshot” the EBS volume and create an AMI. Then delete the VM and live EBS volume.

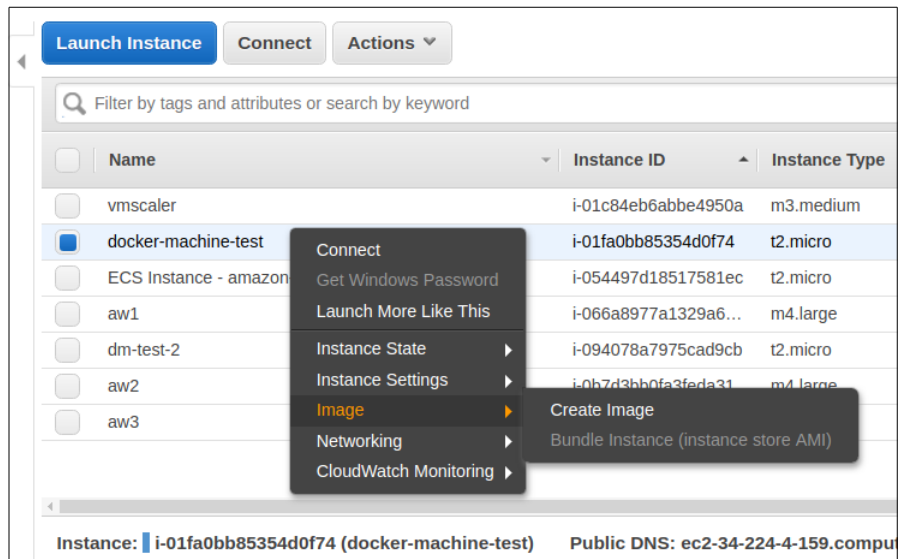


Name	Volume ID	Size	Volume Type	IOPS	Snapshot	Created	Availability Zone
	vol-046f0855...	16 GiB	gp2	100 / 3000	snap-0548d02...	October 4, 2017 at 7:59:58 PM ...	us-east-1e
	vol-0666ab9...	16 GiB	gp2	100 / 3000	snap-0548d02...	October 4, 2017 at 7:58:07 PM ...	us-east-1e
	vol-048394b...	16 GiB	gp2	100 / 3000	snap-0548d02...	October 4, 2017 at 8:46:22 PM ...	us-east-1e
	vol-06c5e58...	8 GiB	gp2	100 / 3000	snap-0cfc17b0...	October 2, 2017 at 9:01:27 PM ...	us-east-1a
	vol-026d9ed...	22 GiB	gp2	100 / 3000		October 2, 2017 at 8:41:10 PM ...	us-east-1a
	vol-070d24a...	8 GiB	gp2	100 / 3000	snap-0c39fbce...	October 2, 2017 at 8:41:10 PM ...	us-east-1a
	vol-03cba5c...	8 GiB	gp2	100 / 3000	snap-0cfc17b0...	October 2, 2017 at 8:01:57 PM ...	us-east-1e
	vol-0cc7d3e...	10 GiB	gp2	100 / 3000	snap-f270c9f	September 11, 2017 at 1:17:19 ...	us-east-1e
	vol-052ea2a...	10 GiB	gp2	100 / 3000	snap-f270c9f	September 11, 2017 at 12:40:55 ...	us-east-1e
	vol-088b10c...	10 GiB	gp2	100 / 3000	snap-f270c9f	August 30, 2017 at 3:53:15 PM ...	us-east-1e
	vol-0228356...	10 GiB	gp2	100 / 3000	snap-1bcbf463	June 9, 2017 at 5:25:22 PM UTC-7	us-east-1e
	vol-09230ce...	10 GiB	gp2	100 / 3000	snap-1bcbf463	June 9, 2017 at 5:25:22 PM UTC-7	us-east-1e
	vol-02731e8...	10 GiB	gp2	100 / 3000	snap-1bcbf463	June 9, 2017 at 5:25:22 PM UTC-7	us-east-1e



To “stop” your instance right-click on the row in the “Instances” view, select “Instance state”, and then “stop”. You may later resume the instance by selecting “start”. When restarting your instance, your public IPv4 address may be reassigned.

An image can be created by right-clicking on the instance row, and selecting “Image” and “Create Image”. This will temporarily shutdown your instance to create the image. Once the image has been created, the instance is restored to its online state. New images will be listed under “Images | AMIs” on the left-handside of the EC2 console. Sorting by Creation Date makes it easy to locate newly created images.



As you work throughout the course projects in TCSS 558, you will likely want to reuse your virtual machine from assignment 0 to help jump start development and testing of future projects.

### ***Next, let's install Docker on this VM.***

highlight the full text below including all spaces and linefeeds/newlines, then copy-and-paste directly to the VM. You may break this into separate commands by copying-and-pasting individual command separated by spaces to more carefully see what is happening:

```
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key
add -
sudo add-apt-repository "deb [arch=amd64]
https://download.docker.com/linux/ubuntu $(lsb_release -cs) stable"

# refresh sources
sudo apt-get update

# install packages
apt-cache policy docker-ce

sudo apt-get install -y docker-ce

#verify that docker is running
sudo systemctl status docker
```

The “Docker Application Container Engine” should show as running.

When working with Docker directly on your local VM, we will preface docker commands with “sudo”, so the commands run as the superuser.

## **Create a docker image for Apache Tomcat**

The “Docker Hub” is a public repository of docker images. Many public images are provided which include installations of many different software packages.

The “sudo docker search” command enables searching the repository to look for images.

Let’s start by downloading the “ubuntu” docker container image:

Note that docker commands are prefaced as “sudo”.

They must be run as superuser.

```
sudo docker pull ubuntu
```

Verify that the image was downloaded by viewing local images:

```
sudo docker images -a
```

Next, make a local directory to store files which describe a new docker image.

```
mkdir docker_tomcat  
cd docker_tomcat
```

Now, download the Java application that we will deploy into the Docker container:

```
wget
```

```
http://faculty.washington.edu/wlloyd/courses/tcss558/assignments/a0/fibo.war
```

Using a text editor such as vi, vim, pico, or nano, edit the file “Dockerfile” to describe a new Docker image based on ubuntu that will install the Apache tomcat webserver:

```
nano Dockerfile
```

```
# Apache Tomcat Dockerfile contents:  
FROM ubuntu  
RUN apt-get update  
RUN apt-get install -y tomcat8  
COPY fibo.war /usr/share/tomcat8/webapps/  
COPY entrypoint_tomcat.sh /  
RUN mkdir /usr/share/tomcat8/logs  
RUN mkdir /usr/share/tomcat8/temp  
RUN ln -s /var/lib/tomcat8/conf /usr/share/tomcat8  
ENTRYPOINT ["/entrypoint_tomcat.sh"]
```

Next, create a script called “entrypoint\_tomcat.sh” under your docker\_tomcat directory as follows:

```
#!/bin/bash
# tomcat daemon - runs container continually until tomcat exits
/usr/share/tomcat8/bin/startup.sh
echo "tomcat daemon up..."
sleep 3
while :
do
    tomcatstatus=`ps aux | grep tomcat8 | grep java`
    if [ -z "$tomcatstatus" ]
    then
        #exit
        echo "tomcat down"
    fi
    sleep 1
done
```

You’ll need to change permissions on this file.  
Give the owner execute permission:

**chmod u+x entrypoint\_tomcat.sh**

Next, build the docker container:

**sudo docker build -t tomcat1 .**

Check that the docker image was build locally:

**sudo docker images**

Next launch the container as follows:

**sudo docker run -p 8080:8080 -d --rm tomcat1**

Check that the container is up

**sudo docker ps -a**

Now, you’ll need to open port 8080 in the default security group in the Amazon management console.

Under instances, look at the detailed instance information and click on “default” next to “Security groups”:

Availability zone	us-east-
Security groups	default.
Scheduled events	No sche

Click the “Inbound” tab, and then the “Edit” button.  
Scroll down and click the “Add Rule” button at the bottom of the dialog box:

All ICMP - IPv4	ICMP	0 - 65535
All ICMP - IPv4	ICMP	0 - 65535
All ICMP - IPv4	ICMP	0 - 65535

**Add Rule**

NOTE: Any edits made on existing rules will result in the rule to be dropped for a very brief period of time until the new rule is added.

Add a “Custom TCP Rule” with the following settings:

Protocol = TCP

Port Range = 8080

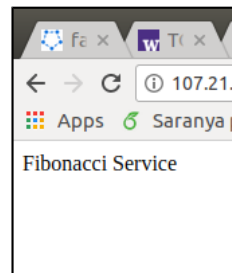
Source = My IP

Then “Save” the security change.

Now, using your browser, point at the http GET endpoint for the web application:

http://<IPv4 Public IP of instance>:8080/fibo/fibonacci

You should see a web page as follows:



Now, test the fibonacci web service deployed onto this container on your EC2 instance using the testFibPar.sh script.

Download the script here:

<http://faculty.washington.edu/wlloyd/courses/tcss558/assignments/a0/testFibPar.sh>

This script uses a Linux utility known as GNU parallel to coordinate separate threads to support parallel client sessions with Apache Tomcat.

If not already installed, you’ll need to install GNU parallel in your Ubuntu (Linux) environment:

```
sudo apt-get install parallel
```

Near the top of the script, you’ll see parameters for host and port:

```
host=34.232.53.152
```

```
port=8080
```

Update the host to match the public IPv4 Public IP for your EC2instance.

Now try exercising your web service using this script.

The first parameter is the total number of service requests to perform.

The second parameter is the number of concurrent threads to use.  
Since we just have one docker container hosting the service, try just one thread:

```
./testFibPar.sh 10 1
```

Run this script 3 times.

The first and second runs may feature slower times reflecting “warm-up” of the infrastructure: VM, container, JVM...

```
Setting up test: runsperthread=10 threads=1 totalruns=10
run_id,thread_id,json,elapsed_time,sleep_time_ms
1,1,{"number":50000},258,.74200000000000000000
2,1,{"number":50000},300,.70000000000000000000
3,1,{"number":50000},306,.69400000000000000000
4,1,{"number":50000},390,.61000000000000000000
5,1,{"number":50000},274,.72600000000000000000
6,1,{"number":50000},288,.71200000000000000000
7,1,{"number":50000},279,.72100000000000000000
8,1,{"number":50000},356,.64400000000000000000
9,1,{"number":50000},317,.68300000000000000000
10,1,{"number":50000},328,.67200000000000000000
```

By the 3<sup>rd</sup> run, performance should be fairly consistent and stable.

---

### Task 3 - Creating a Dockerfile for haproxy

Haproxy is a TCP load balancer that is capable of distributing client requests to a very large number of server hosts. We will next create a Docker image for our haproxy load balancer deployment.

```
mkdir docker_haproxy
cd docker_haproxy
```

First, download the sample haproxy config file:

```
wget
http://faculty.washington.edu/wlloyd/courses/tcss558/assignments/a0/haproxy.c  
fg
```

Using a text editor such as vi, pico, or nano, edit the file “Dockerfile” to describe a new Docker image based on ubuntu that will install the Apache tomcat webserver:  
`$nano Dockerfile`

```
# haproxy Dockerfile contents:
FROM ubuntu
RUN apt-get update
RUN apt-get install -y haproxy
COPY entrypoint_haproxy.sh /
COPY haproxy.cfg /etc/haproxy/
ENTRYPOINT ["/entrypoint_haproxy.sh"]
```

Next, create a script called “entrypoint\_haproxy.sh” under your docker\_haproxy directory as follows:

```
#!/bin/bash
# haproxy daemon - runs container continually until haproxy exits
service haproxy start
echo "haproxy daemon up..."
sleep 3
while :
do
    haproxystatus=`ps aux | grep haproxy-systemd | grep cfg`
    if [ -z "$haproxystatus" ]
    then
        #exit
        echo "haproxy down"
    fi
    sleep 10
done
```

You’ll need to change permissions on this file.

Give the owner execute permission:

```
chmod u+x entrypoint_haproxy.sh
```

Now, let’s update the haproxy configuration file (haproxy.cfg) using your favorite text editor. As provided the haproxy configuration file will perform round-robin load balancing against 3 nodes:

```
server web1 54.210.51.9:8080
server web2 54.210.51.9:8081
server web3 54.210.51.9:8082
```

So far, we have just one Apache Tomcat server in one container, let’s comment out the bottom two entries by using the “#” character:

```
server web1 54.210.51.9:8080
#server web2 54.210.51.9:8081
#server web3 54.210.51.9:8082
```

Now, update the IP address (here 54.210.51.9) to match your public IPv4 address of your ec2instance. Also, instead of using port 8080, change this port to 8081.

We will need to destroy your existing tomcat container which is presently using port 8080 and change this to port 8081. First destroy the old container:

```
sudo docker ps -a
```

Locate the “tomcat1” docker instance. The CONTAINER ID will be the left most column. Using this ID, stop the container:

```
sudo docker stop <CONTAINER ID>
```

Now, relaunch the Apache Tomcat container mapping container port 8080 to the host port 8081:

```
sudo docker run -p 8081:8080 -d --rm tomcat1
```

Now, we’re ready to build the docker container:

```
$sudo docker build -t haproxy1 .
```

Check that the haproxy docker image was built:

```
sudo docker images
```

Now let’s launch the haproxy container. Haproxy will direct incoming traffic to port 8080 to port 8081 which will map to Apache Tomcat:

```
sudo docker run -p 8080:8080 -d --rm haproxy1
```

Now, using the testParFib.sh script, retest that you’re still able to access your webservice, but this time through the haproxy load balancer server.

```
./testFibPar.sh 10 1
```

If this works, then all of the pieces are ready to be deployed across different Docker hosts and containers to complete assignment 0.

---

## Task 4 - Working with Docker-Machine

We will use docker-machine to support working with multiple docker hosts and EC2 instances. Docker-machine makes it very easy to create and destroy instances, and deploy code using Docker containers to multiple VMs on Amazon.

***Before we begin, please stop all containers created for Task 2 and Task 3.***

Search using “sudo docker ps -a”, and use the “sudo docker stop <CONTAINER ID>” command to stop ALL running containers.

<p><b><u>Sudo vs. non-sudo:</u></b> When using docker-machine, docker commands run on remote hosts are not prefaced with “sudo”.</p>
--



Let's start by installing the Amazon Web Services Command Line Interface onto your VM (AWS CLI):

```
sudo apt update
sudo apt install awscli
```

Next configure the AWS CLI using your access credentials created earlier:

```
# configure aws cli
aws configure
```

Next install docker-machine onto your EC2 instance:

```
#to install Docker-Machine:
```

```
# Download the application
curl -L https://github.com/docker/machine/releases/download/v0.12.2/docker-machine-Linux-x86\_64 >/tmp/docker-machine
```

```
# Make it executable
chmod a+x /tmp/docker-machine
```

```
# Copy it into an executable location in the system PATH
sudo cp /tmp/docker-machine /usr/local/bin/docker-machine
```

```
# verify the version
docker-machine version
```

Check and note the Docker machine version.

For further information on Docker Machine see documentation here:  
<https://docs.docker.com/machine/overview/>

Now, let's create a virtual machine to serve as a docker host.

A single command creates the EC2 instance of the specified type, installs the latest version of docker, and prepares the instance for hosting docker containers !!!

Below, I've specified "m4.large" an EC2 instance with 2 virtual CPUs. **Note this is not free. See discussion below on using the free t2.micro if needed.** We will launch this instance as a "spot" instance with a maximum bid of 17 cents per hour:

```
docker-machine create --driver amazonec2 --amazonec2-region us-east-1
--amazonec2-instance-type "m4.large" --amazonec2-spot-price ".17"
--amazonec2-request-spot-instance --amazonec2-zone "e" --amazonec2-open-port
8080 --amazonec2-open-port 8081 --amazonec2-open-port 8082 --amazonec2-
open-port 8083 aw1
```

Note that I've specified availability zone "e". Please set your availability zone accordingly. It will be best to consolidate your instances into the same availability zone for project work in TCSS 558.

The "aw1" refers to the name of the instance. This is the name that you'll use to interact with the VM using the docker-machine CLI. You can use any name desired.

Also please note that docker-machine automatically opens ports using "--amazonec2-open-port <port number>". This automatically adjusts the security-group to provide WORLD access to these ports. **\*\*This is not secure!\*\***, but ok, assuming your instances will not stay up for long.

Alternatively, you could use "FREE tier t2.micro" instances for your docker host(s). These instances will spend on your 750-hours/month FREE credits for t2.micro instances. These single-CPU instances are limited to an initial 30-minute burst of 100% CPU utilization. Once CPU credits are exhausted, the instance is down-throttled to 10% of one CPU core until credits are earned at a rate of 6-minutes @100% utilization per hour:

Instance type	Initial CPU credit*	CPU credits earned per hour	vCPUs	Base performance (CPU utilization)	Maximum earned CPU credit balance***
t2.nano	30	3	1	5%	72
t2.micro	30	6	1	10%	144
t2.small	30	12	1	20%	288
t2.medium	60	24	2	40% (of 200% max)**	576
t2.large	60	36	2	60% (of 200% max)**	864
t2.xlarge	120	54	4	90% (of 400% max)**	1296
t2.2xlarge	240	81	8	135% (of 800% max)**	1944

From: <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/t2-instances.html>

These t2 instances are not spot instances. They are considered full price where the first 750 hours is free. To create a t2.micro docker host:

```
docker-machine create --driver amazonec2 --amazonec2-region us-east-1
--amazonec2-instance-type "t2.micro" --amazonec2-zone "e" --amazonec2-open-
port 8080 --amazonec2-open-port 8081 --amazonec2-open-port 8082
--amazonec2-open-port 8083 aw1
```

Try listing docker-machine hosts:

```
docker-machine ls
```

NAME	ACTIVE	DRIVER	STATE	URL	SWARM	DOCKER
aw1	-	amazonec2	Running	tcp://34.232.53.152:2376		v17.09.0-ce

You should see something similar to the listing above, 1 remote docker host.

Now change your docker CLI to work against the remote host.

```
eval $(docker-machine env aw1)
```

Check “docker-machine ls” again. The host should be marked “ACTIVE”. The following command can also be used to show the active host:

```
docker-machine active
```

Next, we need provide the docker\_tomcat and docker\_haproxy containers locally on each host. While it is possible to use the “docker save” and “docker load” commands in conjunction with docker-machine to accomplish this, for simplicity we will simply rebuild the images on each host for assignment 0.

Try listing the container images known to this docker host:

```
docker images
```

There aren’t any!!! Now, go back into your docker\_tomcat directory on your local instance:

```
cd docker_tomcat
```

Rebuild the tomcat container, but this time because we ran the “eval” command above, the build occurs on the remote server:

```
docker build -t tomcat1 .
```

Now check the list of images:

```
docker images
```

Next rebuild the haproxy image on this remote host.

```
cd docker_haproxy
```

Before rebuilding, update the haproxy.cfg file.  
Please specify the IP address of the new docker-machine host that is listed using “docker-machine ls”. Specify port 8081.

After making these changes, build the haproxy image on the remote host:

```
docker build -t haproxy1 .
```

Now, create an apache tomcat docker container on the remote host.  
We will map apache-tomcat’s port 8080 to 8081 on the Docker Host.

```
docker run -p 8081:8080 -d --rm tomcat1
```

Next, create the haproxy docker container on the remote host.  
We will map haproxy’s port 8080 to 8080 on the Docker host.

```
docker run -p 8080:8080 -d --rm haproxy1
```

Now, by referring again to the IP address obtained from “docker-machine ls”. Using the testFibPar.sh script, update the host IP and test the service:

```
./testFibPar.sh 10 1
```

If your service works, then this certifies you’ve been able to deploy the service onto a docker host using both an apache-tomcat and apache-haproxy container. You’re now ready to tackle assignment 0’s deliverable (task 5).

---

## Task 5 - Completing Assignment 0

The objective for assignment 0 is to compare performance of running the Fibonacci web service using three different configurations.

For each configuration, you will point to the testFibPar.sh script host and port to the haproxy instance which load balances the containers. Please run testFibPar.sh 3 times, and copy the CSV output of the last, third test run into an Excel, Open Office, or Google Sheets spreadsheet.

Run the test script to perform 3 concurrent threads with 10 requests per thread:

```
./testFibPar.sh 30 3
```

In the spreadsheet, label the raw data for each configuration clearly by name. Please indicate the instance type used (e.g. m4.large, t2.micro) for the docker host(s) for the tests. You must use the same instance type for all of your configurations. In the spreadsheet, add a formula to calculate the “average” Fibonacci web service performance for the 30 test results for each of the 3 configurations. It will be something like: “=AVERAGE(D6:D35)”. It is nice to compute percentage differences between the configurations but not required.

At the bottom of the spreadsheet include a summary report. Include a ranking with place, average performance in ms, and % equivalence as follows:

Performance Ranking:			
1 <sup>st</sup> place	Configuration 2	300ms	100%
2 <sup>nd</sup> place	Configuration 1	400ms	133%
3 <sup>rd</sup> place	Configuration 3	500ms	166%

Test the following configurations:

### **Configuration #1 - Co-Located Servers No CPU Thresholds:**

Deploy three apache-tomcat containers on one Docker host Virtual Machine. Map the tomcat containers to use successive port numbers, and update the haproxy configuration accordingly to use these ports:

```
# launch 3 containers
docker run -p 8081:8080 -d --rm tomcat1
docker run -p 8082:8080 -d --rm tomcat1
docker run -p 8083:8080 -d --rm tomcat1
```

### **Configuration #2 - Co-Located Servers With CPU Thresholds:**

Deploy three apache-tomcat containers on one Docker host Virtual Machine, with 66% CPU allocations for m4.large, or 33% CPU allocations for t2.micro

```
# launch 3 containers - m4.large weights
docker run -p 8081:8080 -d --rm --cpus .66 tomcat1
docker run -p 8082:8080 -d --rm --cpus .66 tomcat1
docker run -p 8083:8080 -d --rm --cpus .66 tomcat1
```

### **Configuration #3 - Separate Servers No CPU Thresholds:**

Deploy three apache-tomcat containers on three separate Docker host Virtual Machines. This will require launching an additional two docker hosts using docker-machine. Map haproxy accordingly on the first host to load balance against the apache-tomcat containers running on the other remote hosts.

Use “docker-machine ls” to find the IP address of each host.

You will need to build the tomcat container separately for each new host.

```
# On each host, launch one apache-tomcat container
docker run -p 8081:8080 -d --rm tomcat1
```

The expected behavior is that each of these three configurations will perform differently. If this is not the case, please check your configuration to be sure haproxy has been reconfigured correctly each time for the appropriate hosts.

### **What to Submit**

To complete the assignment, upload your .xlsx spreadsheet file into Canvas under assignment 0.

### **Grading**

This assignment will be scored out of 24 points. (24/24)=100%  
Each cell in the summary spreadsheet is worth 2 points.

### **Teams (optional)**

*Optionally*, this programming assignment can be completed with two person teams.

If choosing to work in pairs, **only one** person should submit the team’s xlsx spreadsheet with results to Canvas.

Additionally, **EACH** member of a team should submit an **effort report** on team participation. **Effort reports** are submitted INDEPENDENTLY and in confidence (i.e. not shared) by each team member.

Effort reports are not used to directly numerically weight assignment grades.

**Effort reports** should be submitted as a PDF file named: “effort\_report.pdf”. Google Docs and recent versions of MS Word provide the ability to save or export a document in PDF format.

For assignment 0, the effort report should consist of a one-third to one-half page narrative description describing how the team members worked together to complete the assignment. The description should include the following:

1. Describe the key contributions made by each team member.
2. Describe how working together was beneficial for completing the assignment. This may include how the learning objectives of using EC2, Docker, Docker-machine, and haproxy were supported by the team effort.
3. Comment on disadvantages and/or challenges for working together on the assignment. This could be anything from group dynamics, to commute challenges, to faulty technology.
4. At the bottom of the write-up provide an effort ranking from 0 to 100 for each team member. Distribute a total of 100 points among both team members. Identify team members using first and last name. For example:

John Doe  
Effort 43

Jane Smith  
Effort 57

Team members may not share their **effort reports**, but should submit them independently in Canvas as a PDF file. Failure of one or both members to submit the **effort report** will result in both members receiving NO GRADE on the assignment...

#### Disclaimer regarding pair programming:

The purpose of TCCS 558 is for everyone to gain experience developing and working with distributed systems and requisite compute infrastructure. Pair programming is provided as an opportunity to harness teamwork to tackle programming challenges. But this does not mean that teams consist of one champion programmer, and a second observer simply watching the champion! The tasks and challenges should be shared as equally as possible.

## **Helpful Hints**

### To display all containers running on a given docker node:

**docker ps -a**

### To stop a container:

**docker stop <container-id>**

For example:

**docker stop cd5a89bb7a98**

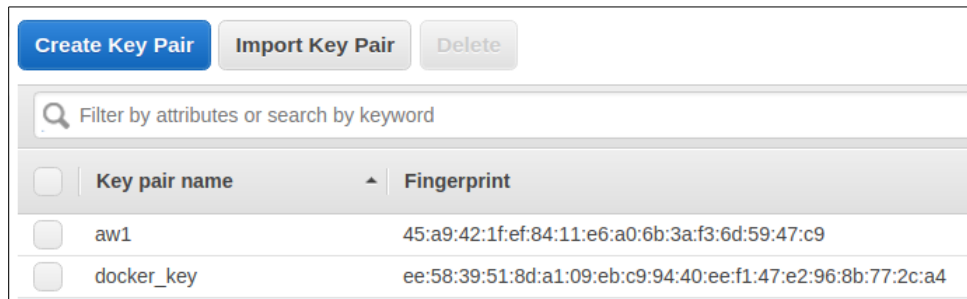
### Multiple docker hosts

When creating multiple docker VM hosts on amazon, each host is referred to by name. To see your hosts use the command:

**docker-machine ls**

The active host will be shown with a '\*'.

The hostname conveniently synced with the AWS keypair name, which is the SSH key used to interact with the virtual machine. If you should need to manually remove keys, this can be done via the EC2 console. On the left-hand side, see “Key Pairs” under “Network & Security”. Keys can be deleted if need be using the UI:



To use a specified remote docker host created by docker-machine:

```
eval $(docker-machine env <host-name>)
```

To unset the remote docker host, and work with your local docker:

```
# set docker back to the localhost
```

```
eval $(docker-machine env -u)
```

Remove a docker host

Once a host created by docker-machine is no longer needed it can be removed by name. This will destroy the VM and stop any associated charges.

```
$docker-machine rm aw2
```

## **Document History:**

v.10 Initial version.