

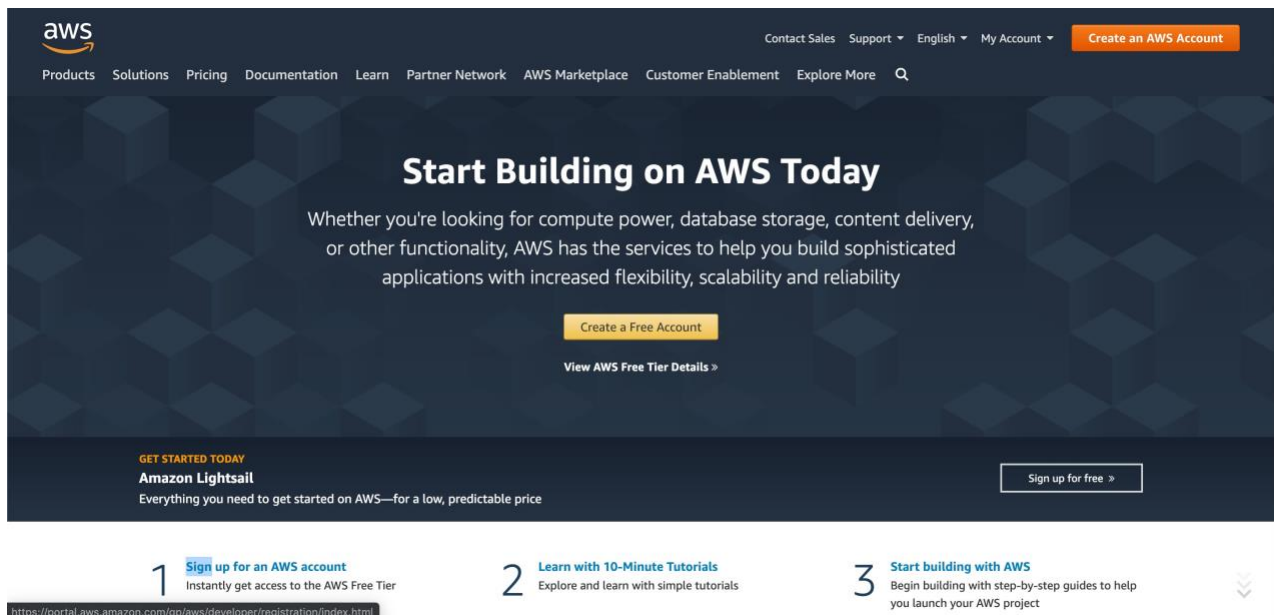
Setting Up Your Amazon Web Services Account

This semester we are allowing all students to explore cloud computing as offered by Amazon's Web Services. Using the instructions below one can establish a service at AWS. Once established, you will be able to run the map reduce job for homework #3 on an AWS instance.

1. Sign up for AWS

To sign up for AWS, you need a credit card. If you do not have one, buy a \$25 American Express Gift card at Ralphs or other grocery store.

To sign up go to:



and click on **Create a Free Account**. Follow the instructions to create your AWS account using the "AWS Free Usage Tier".

Note: Please select Account Type: Personal when setting up the account.

After you are signed up, from the drop down next to your name, select **My Account**. In the *Account Settings* section, you will notice the *Account Id*. This is the **AWS Account ID** to be used next when signing up for AWS Educate.

Please note that many of the URLs listed from now on will only be available if you are

signed up to AWS.

2. Sign up for AWS Educate

To sign up for AWS Educate and get a \$100 credit (USC is a member of AWS Educate) go to:

<http://aws.amazon.com/education/awseducate>

Click on the *Join AWS Educate* button.

On Step 1, click on the arrow in *Student* button.

On Step 2, fill out the form appropriately. **University of Southern California** will auto-complete. Leave the **Promo Code** field empty and click **NEXT**.

On Step 3, Scroll to view the complete Term and Conditions. Select the checkbox **I Agree**.

Click **SUBMIT** and finish the sign-up process.

After your application is reviewed and approved, you will receive a welcome e-mail from AWS Educate Support, which includes details to set your password, as shown below:

Congratulations!

Your AWS Educate application has been approved. As a member of the AWS Educate program, you will gain access to the benefits listed below:

AWS Educate Student Portal

The AWS Educate Student Portal is the hub for AWS Educate students around the world to find AWS content to help with classwork, connect to self-paced labs and training resources.

[Click here](#) to set your password and log in to the AWS Educate Student Portal. After logging in, click AWS Account at the top of the page to choose how you would like to access AWS services.

Bookmark the AWS Educate Student Portal for easy access, or [click here](#) to sign in directly.

You can access a video walk-through of the AWS Educate Student portal [here](#).

Free AWS Essentials Training

To access our foundational AWS Cloud Practitioner Essentials online learning class for free and find other self-paced labs, you must have either an AWS account or an Amazon ID.

- If you have an AWS account, sign in and [click here](#) to receive these benefits.
- If you do not have an AWS account, [click here](#) and follow the instructions to create an Amazon ID to access these benefits.

Once you access the Training and Certification portal, click "Learning Library" and search for "AWS Cloud Practitioner Essentials" to easily locate and enroll in AWS Cloud Practitioner Essentials on-line training. You can access AWS training any time after setting up your account by clicking [here](#).

Thank you again for participating in AWS Educate and we hope you enjoy the program!

Good luck with your continued studies,

The AWS Educate Team

2.2 Issues Sign up for AWS Educate

If you are having issues signing up for AWS Educate, and your initial application is rejected, you can create a Support Case at <https://console.aws.amazon.com/support>, where you describe the problem and attached a copy of the front and back of your USC ID. Alternatively you could also contact AWS Educate Support directly at:

<http://aws.amazon.com/education/awseducate/contact-us>

3. Apply the \$100credit

Once you have completed signing up for both AWS and AWS Educate, login to your AWS Educate account.

Then go to the URL:

<https://www.awseducate.com/students/s/awssite>

Click on **AWS Account** in the toolbar. Select "I would like to use a personal AWS Account". Enter the *AWS Account ID* to get the **Promo Code** which will be displayed in a box saying, "Show my AWS Promotional Credit".



How to access AWS Services

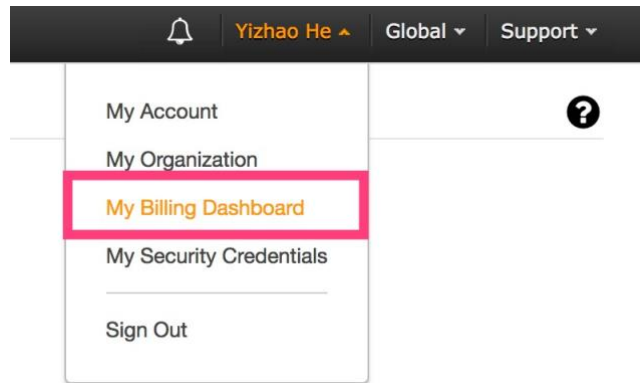


Choose an option to get started!

- ☐ I would like to use a personal AWS Account
- ☐ I would like to use an AWS Educate Starter Account

4. Provide Credit Card or Gift Card Information

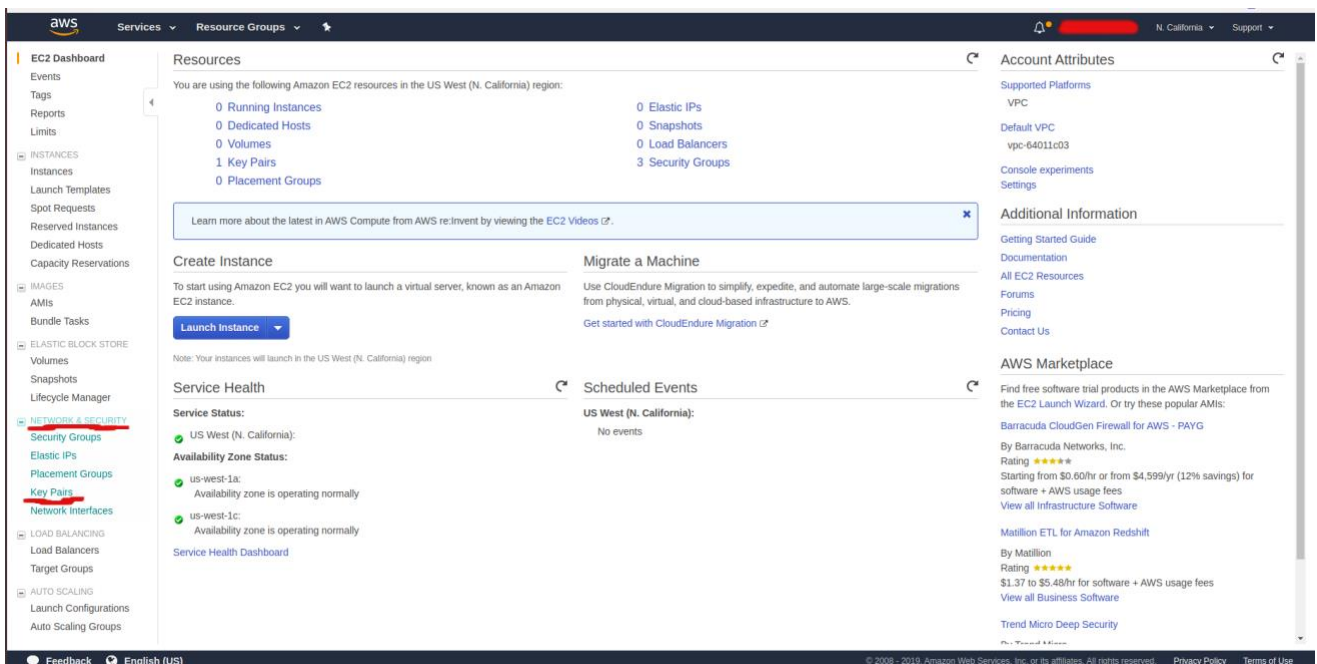
- ☐ In the top menu, click on your name and select **My Billing Dashboard** in the drop-down menu.



- ☐ In the left menu select **Payment Methods**
- ☐ Click on **Add a Card** button
- ☐ Provide your card information (Credit Card Number, Name of Cardholder, Expiration Date)
- ☐ Optionally Edit your billing address.
- ☐ The \$100 promotional credit will be available in your billing section inside the "Credits" tab.

5. Create a Security Key-Pair for secure SSH communication

- ☐ Go to AWS Dashboard by clicking the AWS logo on left top.
- ☐ Search for ec2 service.
- ☐ On the left side toolbar, look for **Network and Security** -> Select **Key Pairs**.



- ☐ Click on **Create Key Pair**. Give any suitable name to it. You will be prompted to save the private key, but that is not necessary.

6. Create a Data Storage – s3 Bucket

- To store the input data, output data and jar file of java program, we will need a data storage. On AWS, it is called as s3 (Simple Scalable Storage) bucket.
- Go to **AWS Dashboard** by clicking the **AWS logo** on top left corner.
- Search for **s3** service.
- Click on **Create Bucket**.
- Inside write a **bucket name**. Bucket name should be globally unique. Then Select the **Region**. For region, you can check the top right corner on **AWS Dashboard** or at the end of URL, there is a region. For USC, region is **US West (N. California)**.
- Keep all other settings as default and click on **Create** to create the bucket.
- Now go to the following URL and upload the input folder.
<http://csci572.com/2020Spring/hw3/xxxxxxxxxxxxx>
It might take some time because of size of data. You can see the progress bar at the bottom of the page.
- Upload in the same way the jar file of code you have written to the same bucket.

7. Create a Map-Reduce job

- To create a map-reduce job, we will use the **AWS EMR (Elastic Map Reduce) Service**.
- For that, go to **AWS dashboard** and search for **EMR** service.
- Then click on **Create Cluster**. Inside give your cluster a name. Keep the **Logging** checked. At **s3 folder** option, choose the s3 bucket you have created. For **Launch Mode**, choose **Cluster**. Inside the **Software Configuration**, we need to select the version of Hadoop we want to use. So choose the one, which you have used while writing the map reduce code. Select the **Applications** as **Core Hadoop**. (Note: Make sure the version of Hadoop that you select agrees with the version of Hadoop that you used to create the jar file.)
- In **Hardware Configuration**, we need to choose the type of machines we need for this job. AWS offers large variety of machines and charges you based on the machine you select. For this homework **m3.xlarge** is enough. Then in **Number of Instances**, keep the count as **3**. It specifies the number of machines we need for the Map-Reduce job. In case of 3, there will be 1 master machine and 2 slave machines. If we increase the number of machines, then the speed of the job will increase as well as the cost.
- In **Security and access**, for **EC2 key pair**, choose the key pair you have created before. Keep the permissions to default.
- Click on the **Create cluster**.

aws Services Resource Groups

Create Cluster - Quick Options [Go to advanced options](#)

General Configuration

Cluster name:

☒ Logging ⓘ

S3 folder:

Launch mode: ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Release: ⓘ

Applications:

- ☒ Core Hadoop: Hadoop 2.8.5 with Ganglia 3.7.2, Hive 2.3.5, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.9.2
- ☐ HBase: HBase 1.4.10 with Ganglia 3.7.2, Hadoop 2.8.5, Hive 2.3.5, Hue 4.4.0, Phoenix 4.14.2, and ZooKeeper 3.4.14
- ☐ Presto: Presto 0.224 with Hadoop 2.8.5 HDFS and Hive 2.3.5 Metastore
- ☐ Spark: Spark 2.4.4 on Hadoop 2.8.5 YARN with Ganglia 3.7.2 and Zeppelin 0.8.1

☐ Use AWS Glue Data Catalog for table metadata ⓘ

Hardware configuration

Instance type:

Number of instances: (1 master and 2 core nodes)

Security and access

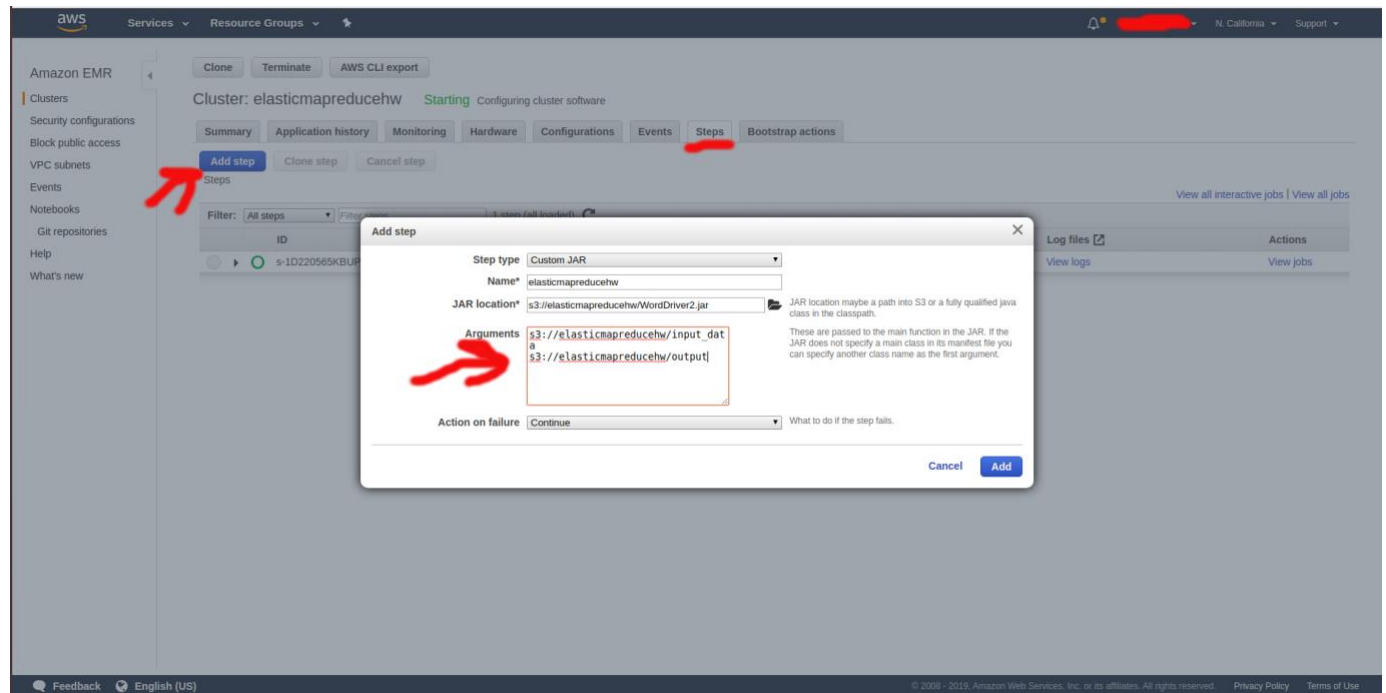
EC2 key pair: ⓘ [Learn how to create an EC2 key pair.](#)

Permissions: ☒ Default ☐ Custom

Feedback English (US)

© 2008 - 2018 Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

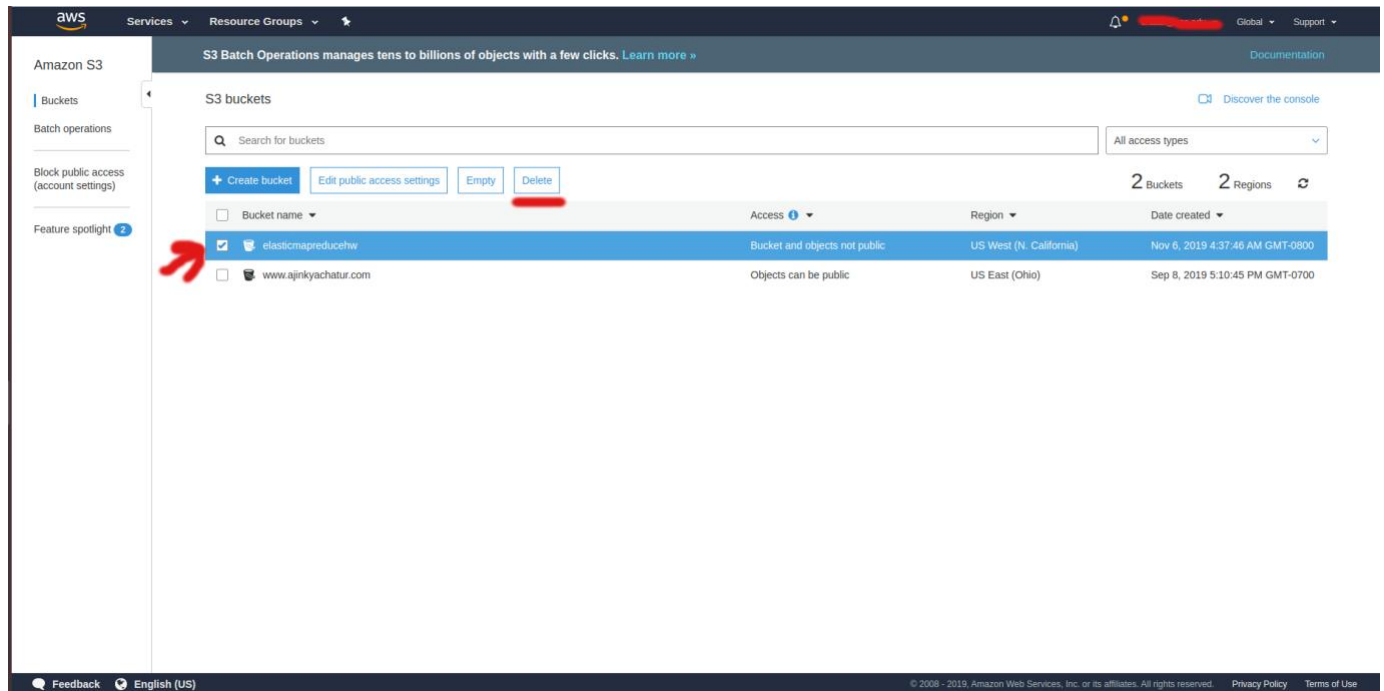
- Now wait for the cluster while **Master and Core** are in **Provisioning** state. (Note: This may take a while.)
- Once the state is changed to **Running**, go to **Steps** tab. This is where we will start our job.
- There click on **Add step**. Inside, keep the **type** as **Custom JAR**. Give a good name to step. Select path of jar file in **JAR Location** option. Then in the argument, we need to pass two arguments, first with our input directory and second with the output directory. Please follow the image below. (Note: Please make sure that there is a space between input and output directories.)



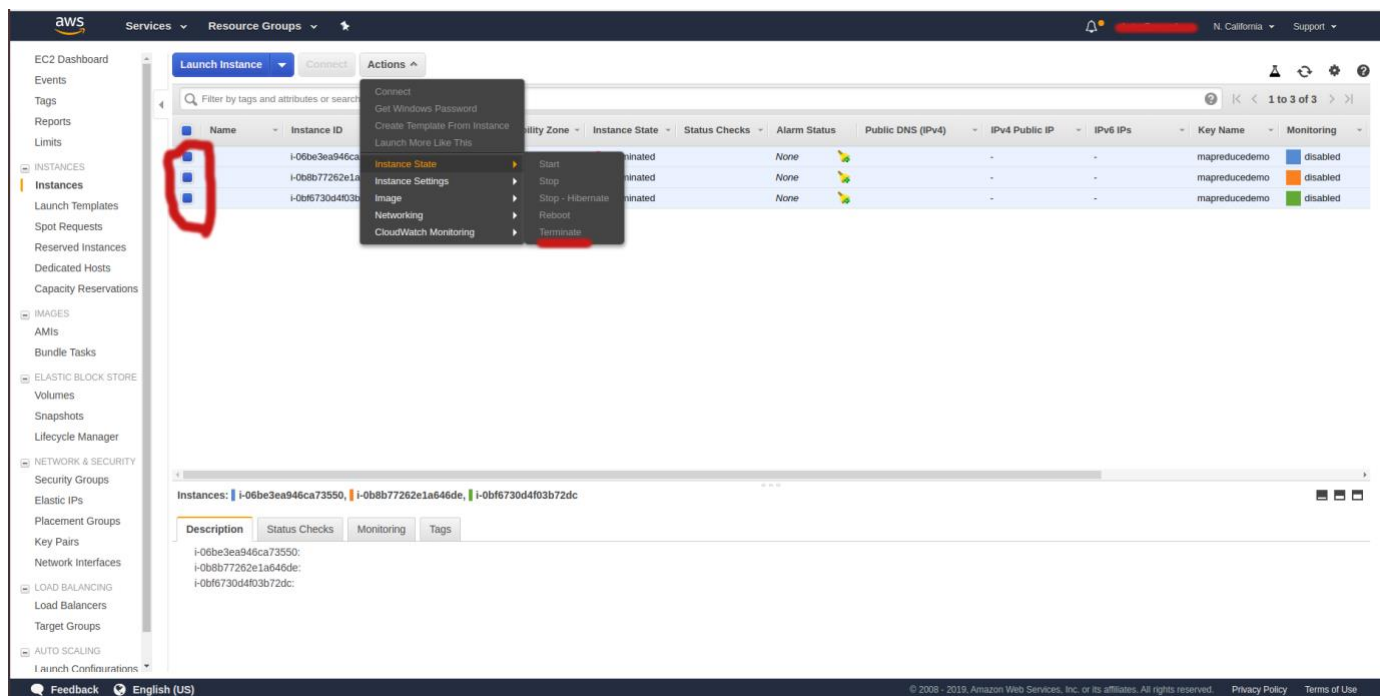
- Then click on **Add**.
- Wait for job to complete. If the **System Code** is **0** that means the job is successful, else we need to debug the problem. For debugging, you can visit the **Logs**.
- Once the job successfully completes, go to the **output** folder and download all the output.

8. Clean the resources.

- After successful completion of the job, it is very important to clean all the resources, otherwise **AWS** will keep charging you even if you are not using the services.
- For that, we will have to clean the s3 bucket and terminate all the ec2 instances of job.
- So, go to **AWS Dashboard** and search for **s3** service.
- After going there, select your s3 bucket and click on **Delete**.



- Now it is time to clean ec2 instances. While we executed the **EMR** service and created a **Map-Reduce job**, at that time AWS creates ec2 instances, one for each master and slave. So, it is necessary to close them.
- Go to **AWS Dashboard** and search for **ec2** service.
- Select all the instances there, which we used for this homework. Go to **Actions** dropdown. Go to **Instance State** and select **Terminate**.



Now you have the output and you can revert back and complete the assignment.

References :

For detailed understanding, you can refer all of the below resources :

1. <http://aws.amazon.com/s3/>
2. <http://aws.amazon.com/ec2/>
3. <http://aws.amazon.com/elasticmapreduce>
4. <http://hadoop.apache.org/>
5. https://www.youtube.com/watch?v=ce857_wtR-U
6. <https://www.youtube.com/watch?v=G6kQ14AAzXQ>
7. <https://www.youtube.com/watch?v=JDk-LYJMzEU>

Congratulations have fun and explore AWS !!