How to get the data?

By Jingbo Shang (**shangjingbo**)

(edited for Round 2 by walrus71)

# Preface

Downloading data for the SpaceNet Challenge (Round 2) will be different from how Topcoder usually provides you data for a match. That being said, it will be a straightforward and painless process if you follow this guide. In total, if you already have a linux, mac machine, or AWS CLI for Windows, the process should take you ~10 minutes to get to the data!!!

Note that downloading the minimal data set that is necessary to participate in the contest will cost you about $7. The data is $0.09 per GB to download and we have about ~80GB of data to download.

Note that for “Data Transfer OUT From Amazon S3 To Internet” gives you:

* First 1 GB / month $0.000 per GB
* Up to 10 TB / month $0.090 per GB
* If you are a new AWS customer, for up to 12 months following your AWS sign-up, you are given a free usage tier of 15GB. Once your usage expires or your application use exceeds the free usage tiers, you simply get pay-as-you-go service rates (see each service page for full pricing details). Restrictions apply; see offer terms for more details.

In the future, Topcoder will run more SpaceNet Challenges with larger datasets that will follow this same download process.

# AWS account

If you have an Amazon shopping account, you already have an AWS account implicitly.

[Amazon Web Services](https://aws.amazon.com/) (AWS) , a subsidiary of Amazon.com, offers a suite of cloud-computing services that make up an on-demand computing platform.

In the top-right corner, there is a “**Sign In to the Console**” button as the following.



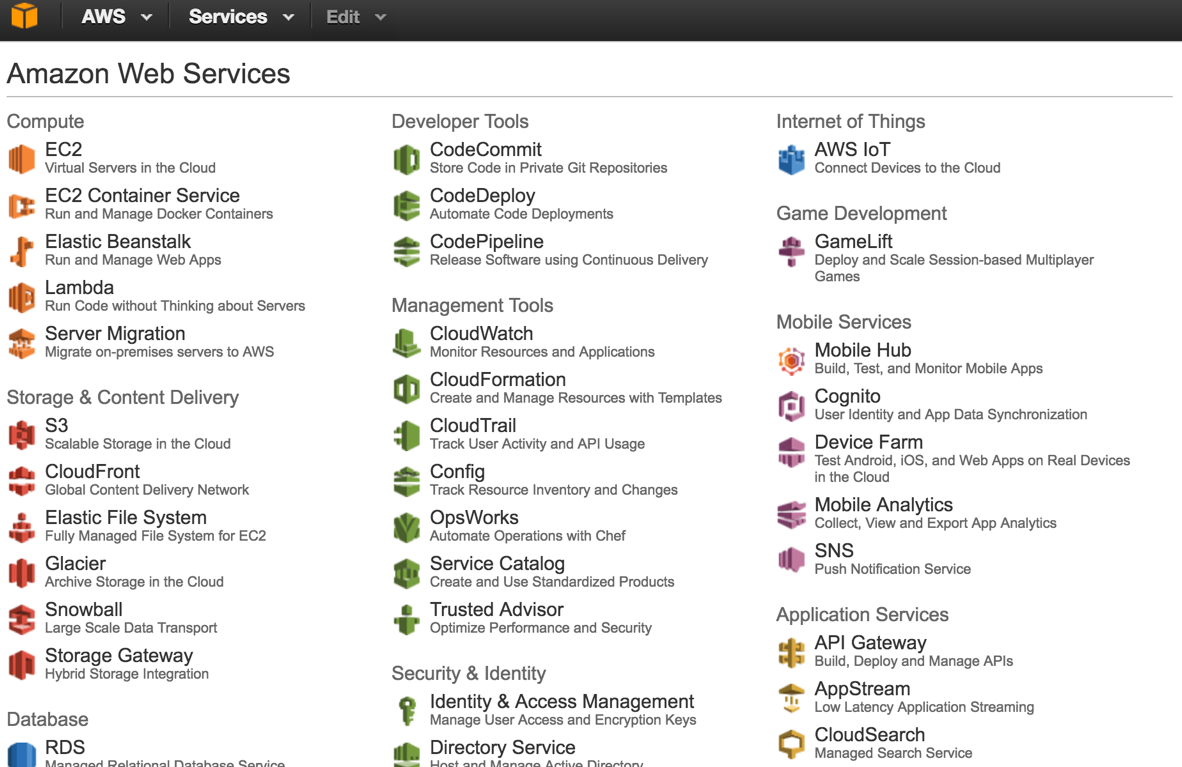
Amazon Web Services uses information from your Amazon.com account to identify you and allow access to Amazon Web Services. Therefore,

* If you already have an amazon account (e.g., for online shopping), you can simply use the same.
* Otherwise, please register a new account through [registration website](https://aws.amazon.com/resources/create-account/).

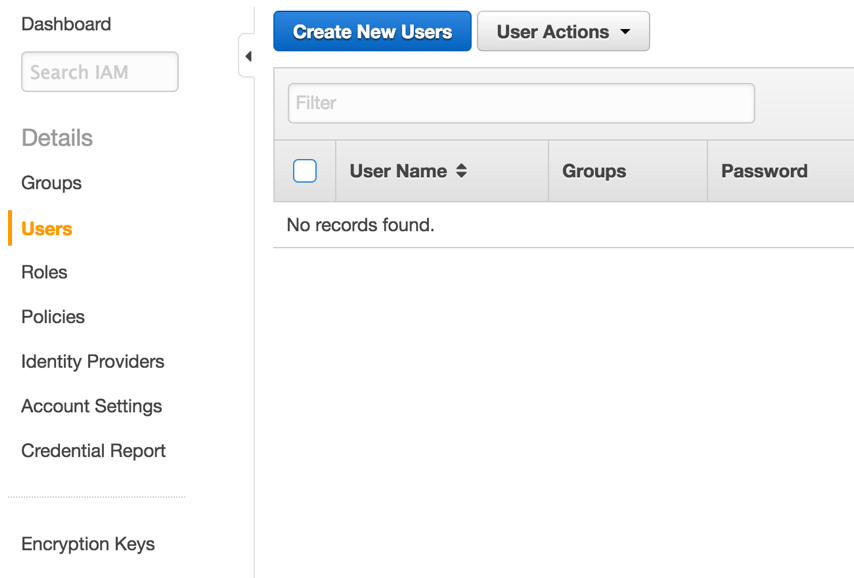
Note that you will need **a credit card associated with your account to download the data**.

# Set up a new user in your AWS account

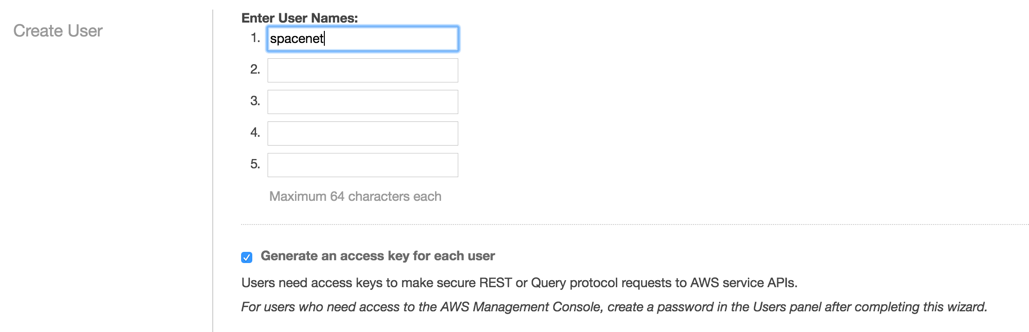
**Step 1.** Once you have an AWS account, you can sign in to your console. Your console looks like the following figure. Click on the “**Identity & Access Management**” as highlighted in the figure.



**Step 2.** Then, on the left side of the screen, you will find a list of details as follows. Please click on the “**Users**”. And then click on the “**Create New Users**” on the right.



**Step 3.** Then, enter **a new user name**, for example “spacenet”, with **“Generate an access key for each user” option selected** (it is selected by default), as shown in the following figure. Click the “**create**” button in the bottom-right corner.



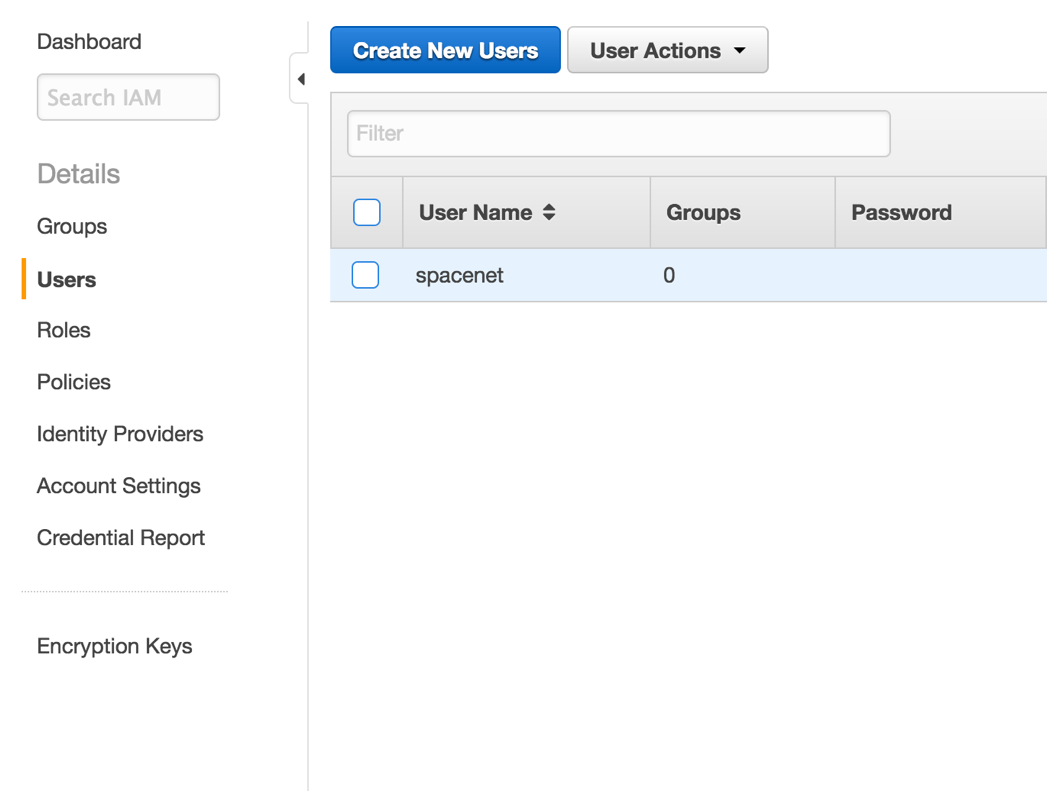
**Step 4.** Please save your credentials file and then close. To save this information, you can either click on “**Show User Security Credentials**” or “**Download Credentials**”. We will use this pair of “**Access Key ID**” and “**Secret Access Key**” later. Because of privacy issue, there is no figure instruction there ☺.

Now, we are going to change the permission of your user page. In our example, it is **IAM > Users > spacenet**. If you use a different username, the last click should be changed accordingly.

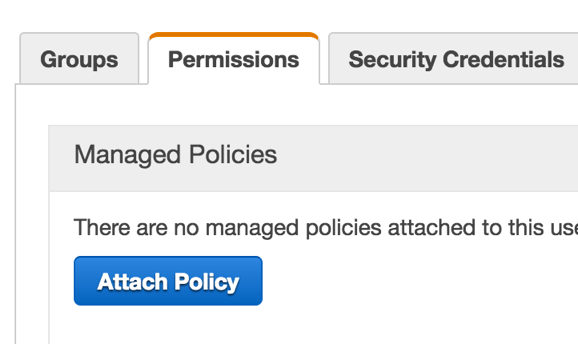
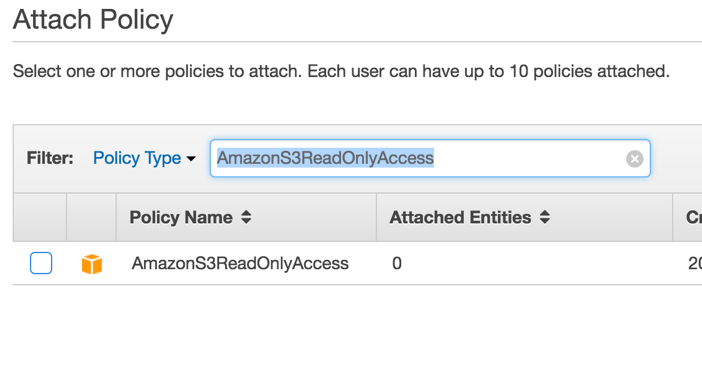
**Step 5.** First, click “**Services**” on the top of the web page, and click “**IAM**” as highlighted in the following figure:



**Step 6.** Click “**Users**” and select the new account (e.g., **spacenet**) you just created.



**Step 7.** Click “**Permissions**”, “**Attach Policy**”, find “**AmazonS3ReadOnlyAccess**”, check the box, and click “**Attach Policy**”.



# 

# Linux or Mac OS users

If you don’t have a Linux or mac OS, but would like to use this option, then [Virtual Box](https://www.virtualbox.org/wiki/Downloads) is a free way to get such an OS without purchasing a new machine. And [Ubuntu Desktop](https://www.ubuntu.com/download/desktop) is fairly easy to setup. There are many instructions on how to install Ubuntu in virtual box. One of the instruction could be from [Ask Ubuntu forum](http://askubuntu.com/questions/142549/how-to-install-ubuntu-on-virtualbox).

## Install AWS CLI

First, you need to install “**pip**”.

* On mac, just type “sudo easy\_install pip” in your terminal and press enter.
* In Linux, it depends on your exact version. Two useful references: [Ubuntu 16.04](https://www.rosehosting.com/blog/how-to-install-pip-on-ubuntu-16-04/) and [Ubuntu 10.10 and older](http://www.saltycrane.com/blog/2010/02/how-install-pip-ubuntu/).

Second, type **“sudo pip install awscli”** in your terminal.

# Windows users

Download the AWS CLI installer from [this page](https://aws.amazon.com/cli/) and follow the instructions.

Alternatively you can use Virtual Box to obtain an Ubuntu installation (see above at ‘Linux or Mac OS users’) then follow the instructions above.

# Configure AWS CLI

Next, configure aws cli with your credentials saved before. Type “**aws configure**” and then enter your keys:

* AWS Access Key ID [None]: XXXXXXXXXXXXXXXX
* AWS Secret Access Key [None]: XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
* Default region name [None]:
* Default output format [None]:

Last, test your connection with the following command.

aws s3api get-object --bucket spacenet-dataset --key manifest.txt --request-payer requester manifest.txt

This command will download a file “manifest.txt” in your current working directory. Note that this manifest file may not contain up to date information about the repository, nevertheless it’s good enough to test whether your process works.

# Download Files

See the contest’s problem statement or [this page](https://spacenetchallenge.github.io/) for a list of data files that you should download. To get them use the following command:

**Template:**

aws s3api get-object --bucket spacenet-dataset --key $PATH --request-payer requester $NAME

**Example:**

aws s3api get-object --bucket spacenet-dataset --key spacenet\_sample.tar.gz --request-payer requester spacenet\_sample.tar.gz