

---

# **Image Webs Cloud Tools Documentation**

***Release 0.1***

**Kyle Heath**

July 07, 2015



<b>1</b>	<b>Contents</b>	<b>3</b>
1.1	Tutorial . . . . .	3
1.2	Manual . . . . .	13
1.3	Guides . . . . .	13



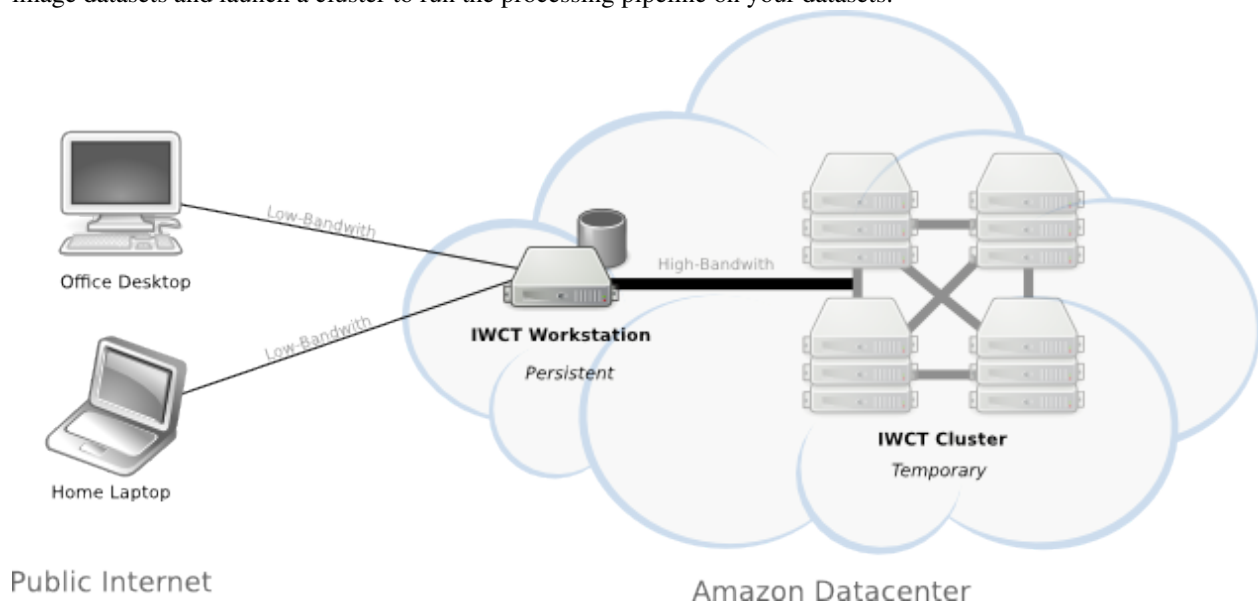
The documentation explains how to build [Image Webs](#).

### Who is this for?

The [tutorial](#) section is written for anyone who wants to visualize their own large-scale image collection as an image web. No programming is required. The [manual](#) section is for users who want to hack on the code of the matching pipeline used to build the image webs itself. All of the code is open source and available from [github](#).

### What is this?

The Image Webs Cloud Toolkit (IWCT for short) creates a computing environment on Amazon's EC2 cloud with the Image Webs software pipeline installed and ready to use. This environment consists of two elements: the workstation and the cluster. You connect to the workstation from your local machine. From the workstation, you can prepare your image datasets and launch a cluster to run the processing pipeline on your datasets.



### *IWCT Computing Environment*

**workstation** A persistent remote workstation setup for you in the AWS cloud.

**cluster** A temporary cluster setup for you in the AWS cloud on-demand.

### How much does it cost?

The IWCT software is opensource and free to use. However, you pay amazon for whatever resources you use. Roughly \$1 per day to run the workstation and \$6 per hour to run a typical cluster. See the [estimating costs](#) section for details.



---

## Contents

---

## 1.1 Tutorial

### 1.1.1 Setup

You must install NX client on any computer from which you wish to work. Also because IWCT launches computer resource on the public Amazon Web Services cloud, you'll need to create an AWS account.

---

**Note:** You'll need to provide a credit card and phone number to create an AWS account.

---

#### Install NX Client

The [NX client](#) allows you to use another computer remotely as if it were your local machine. It provides a fast full-screen GUI session even across poor connections. Please install the NX client for your OS:

- [Linux](#)
- [Mac OSX 10.3-10.6](#)
- [Mac OSX 10.7](#)
- [Windows](#)

#### Create AWS Account

1. Go to <http://aws.amazon.com>
2. Click the **Sign Up** button
3. Follow the on-screen instructions

---

**Note:** If you are a student or researcher, you can [request a grant](#) for free AWS credits. When your grant credit is exhausted, your credit card will be charged once again.

---

**Warning:** You are responsible for any charges you accrue for use of AWS resources. When you complete a work session, always login to your [AWS management console](#) and check that your workstation instances are stopped and your cluster instances terminated to avoid accidentally wasting resources.

### 1.1.2 Launch workstation

The IWCT workstation console allows you to start, stop, and connect to your workstations from any internet connected computer.

---

**Note:** Make sure *NX Client is installed* on the computer you are using and that you can login to the [AWS Management Console](#).

---

**To start and connect to your workstation:**

1. Go to the IWCT workstation console: <http://iwct.dasgizmo.net>
2. Click **login** and enter your username and password (or click **register** if you don't have an account).
3. Click the **connect** button next to a workstation workstation (or the **create new workstation** button if no workstations are listed).

Use `ctrl-alt-f` to use nx-client in fullscreen mode. Here are some more handy [NX Client keyboard shortcuts](#).

---

**Note:** After login, you'll be prompted for your AWS Access Key. Follow the guide here to [get your AWS Access Key](#).

---

### 1.1.3 Prepare dataset

---

**Note:** Perform the following steps from your *IWCT Workstation* (i.e. *connect using the NX client*). You may want to open this page in a browser running on the IWCT Workstation for convenience.

---

You can either use an default example image dataset or provide your own image files.

#### Optional - Prepare your own dataset

To use your own image collection, you'll need to place jpeg format image files in an archive file (tar or zip). Any directory structure is fine, non-JPEG files will be ignored.

---

**Note:** For convenience, place your .zip or .tar file on a webserver so you can easily download it to your workstation in the next step.

---

#### Import Data

Open a terminal (right click on desktop and select **Open in terminal** from the context menu) and enter the following command.

```
iwct_import_dataset
```

At the prompt `Please enter url to dataset archive:`, provide the path to your tar or zip image archive in URL format (or use the URL for the example dataset).

---

**Note:** If your file is copied locally on the workstation use the URL prefix `file://` instead of `http://`.

---

At the prompt `Please enter name for dataset (no spaces) :`, provide a name for your dataset (no spaces - only letters, numbers, underscores).

This creates a new dataset directory `/home/ubuntu/Desktop/datasets/dataset_name` containing a PERT file `imageid_to_image.pert` containing your dataset images.



### 1.1.4 Launch cluster

Login or register a new account at <http://www.mapr.com/Log-yourself-in-mapr>

MapR Technologies  
EASY. DEPENDABLE. FAST.

MapR Blog | Support | Contact Us | Login/Register |

[DOWNLOAD NOW](#)

SOLUTIONS PRODUCTS SERVICES SUPPORT PARTNERS Search...

HOME > LOGIN > LOGIN

**Login**

Email

Password

☐ Remember me

[LOGIN](#)

[Forgot password?](#)

OR

If you are not currently a member, please register here.

First Name:

Last Name:

Company:

Email:

Country:

State:

Zip:

Phone:

Password:

Verify Password:

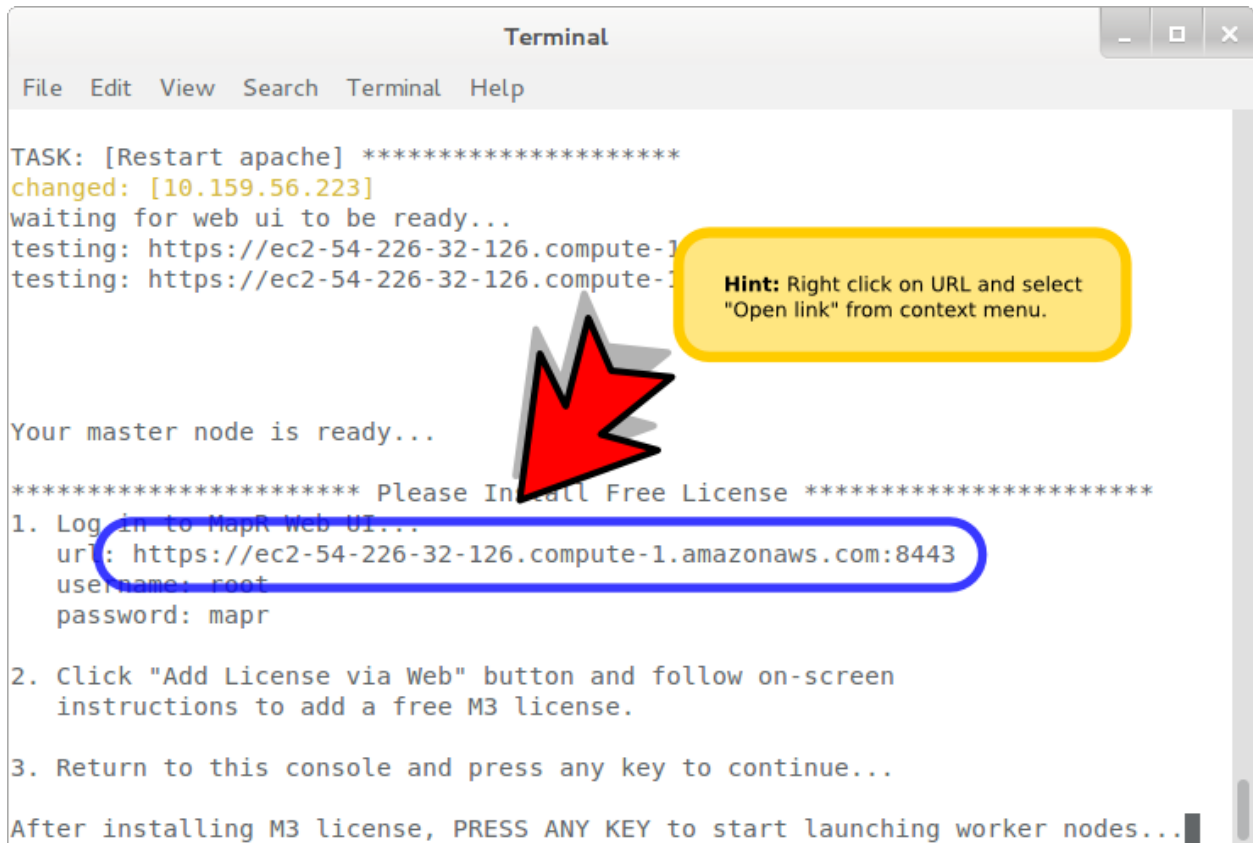
[REGISTER](#)

Required field

Open a terminal, and start a cluster with 5 nodes

```
cirrus_cluster_cli create 5
```

When the prompt shows this message, open the indicated URL in a browser.



The image shows a terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output includes the following text:

```
TASK: [Restart apache] *****
changed: [10.159.56.223]
waiting for web ui to be ready...
testing: https://ec2-54-226-32-126.compute-1.amazonaws.com:8443
testing: https://ec2-54-226-32-126.compute-1.amazonaws.com:8443

Your master node is ready...

***** Please Install Free License *****
1. Log in to MapR Web UI...
   url: https://ec2-54-226-32-126.compute-1.amazonaws.com:8443
   username: root
   password: mapr
2. Click "Add License via Web" button and follow on-screen
   instructions to add a free M3 license.
3. Return to this console and press any key to continue...

After installing M3 license, PRESS ANY KEY to start launching worker nodes...
```

Annotations in the image include a red arrow pointing to the URL in step 1, a yellow callout box with the text "Hint: Right click on URL and select 'Open link' from context menu.", and a blue oval highlighting the URL and username fields.

---

**Hint:** Hover and right click on the URL and select **Open link in browser** from the context menu.

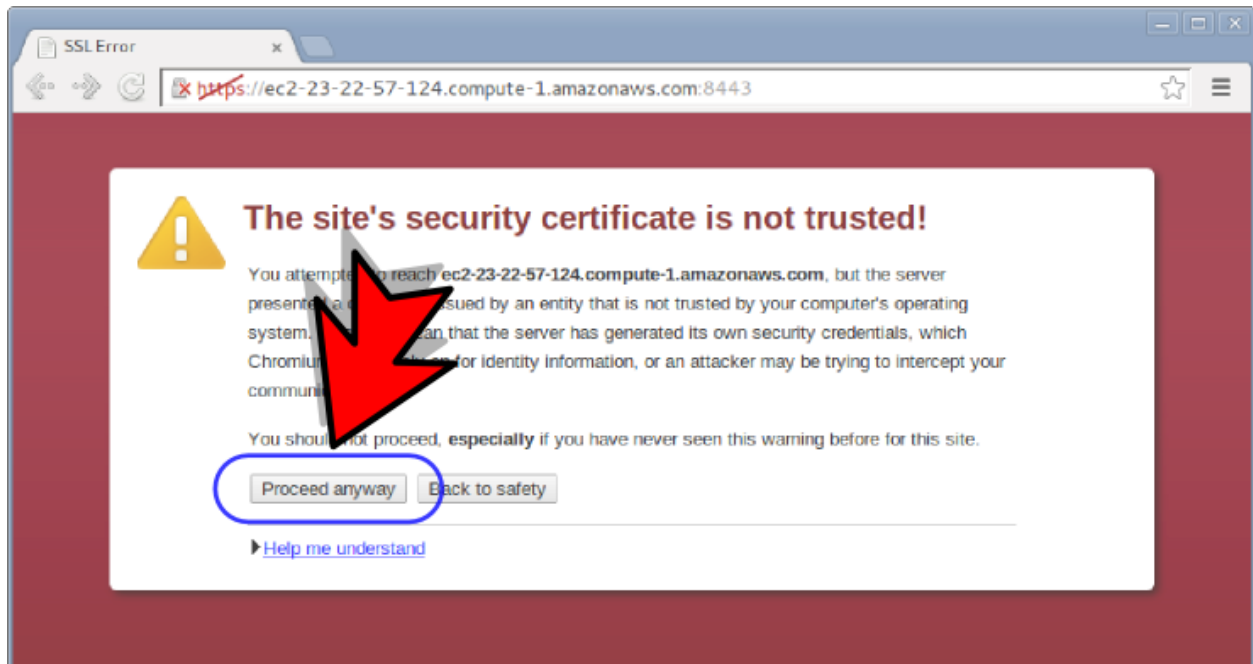
---

Click the **Proceed anyway** button.

---

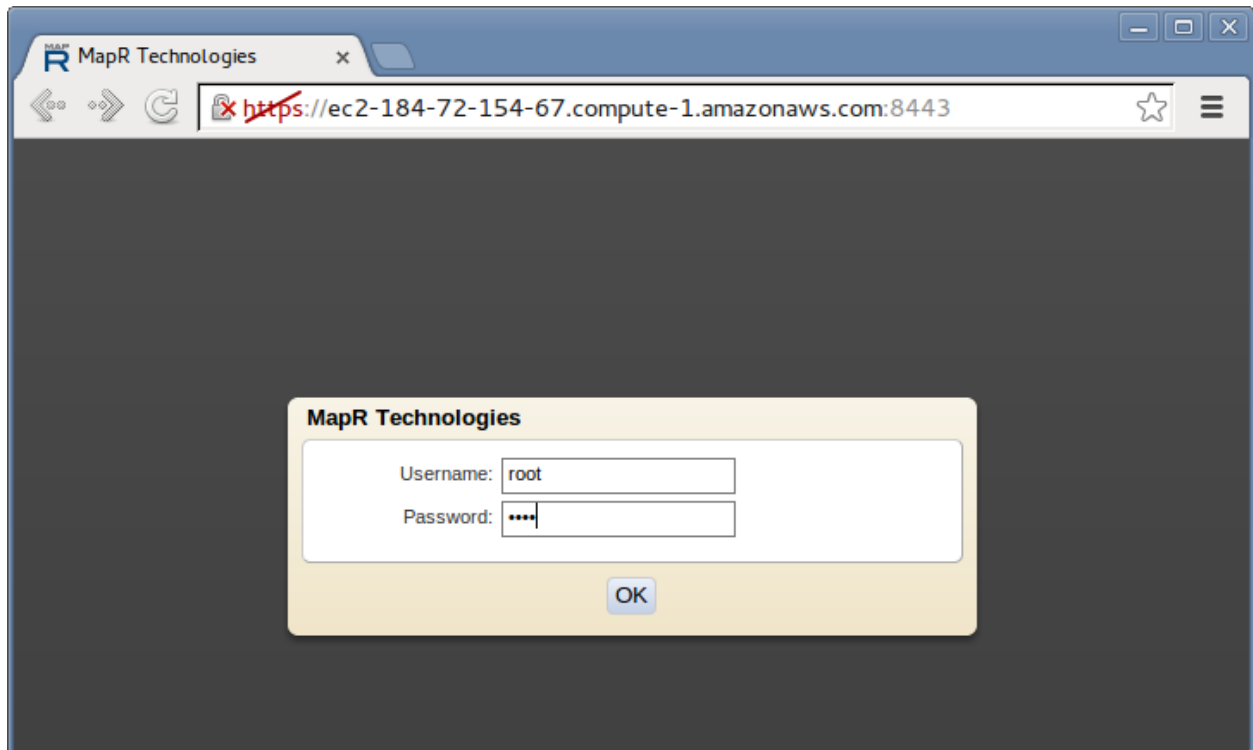
**Note:** You can safely ignore the scary looking SSL warning.

---

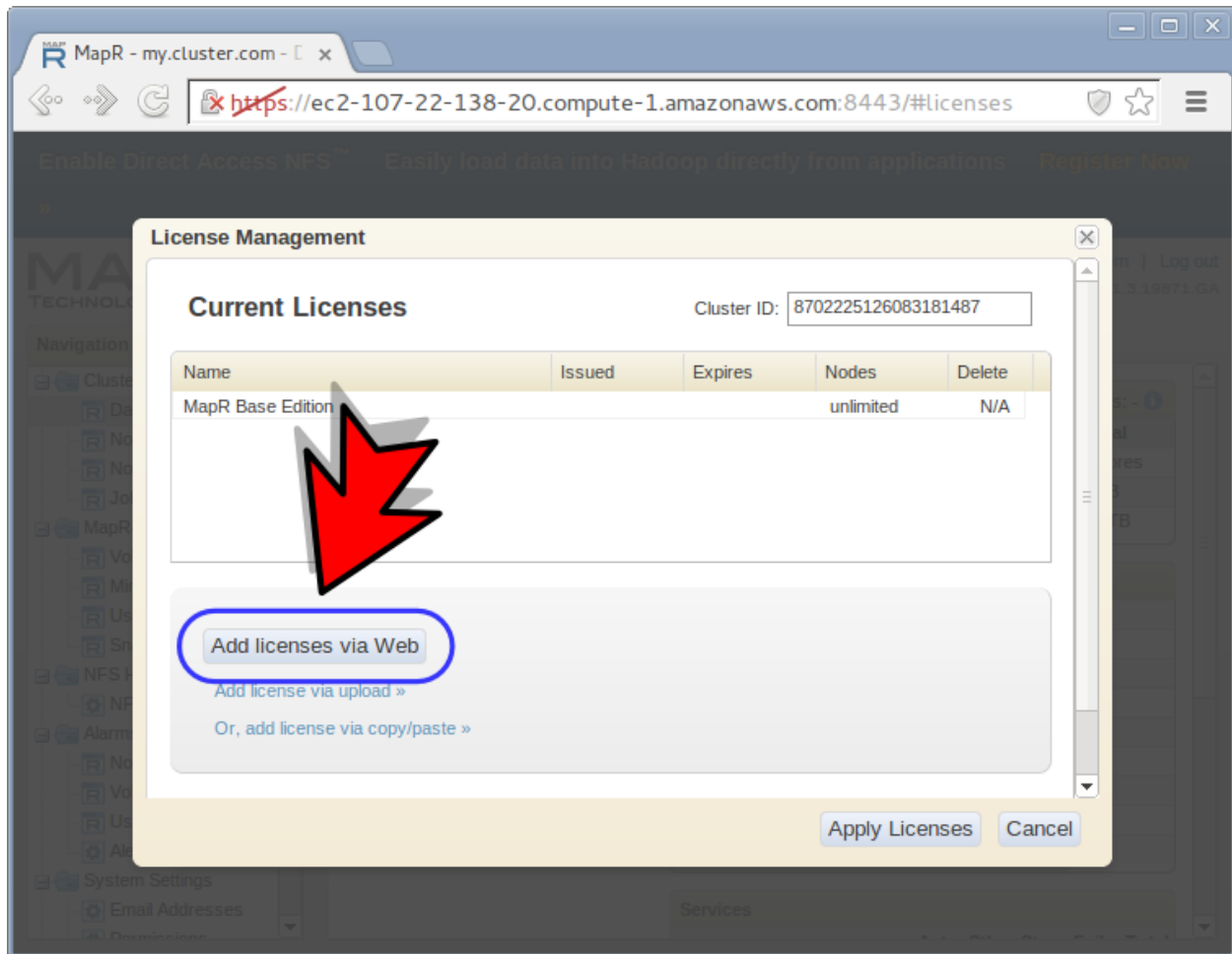


Login as the root user.

username	root
password	mapr



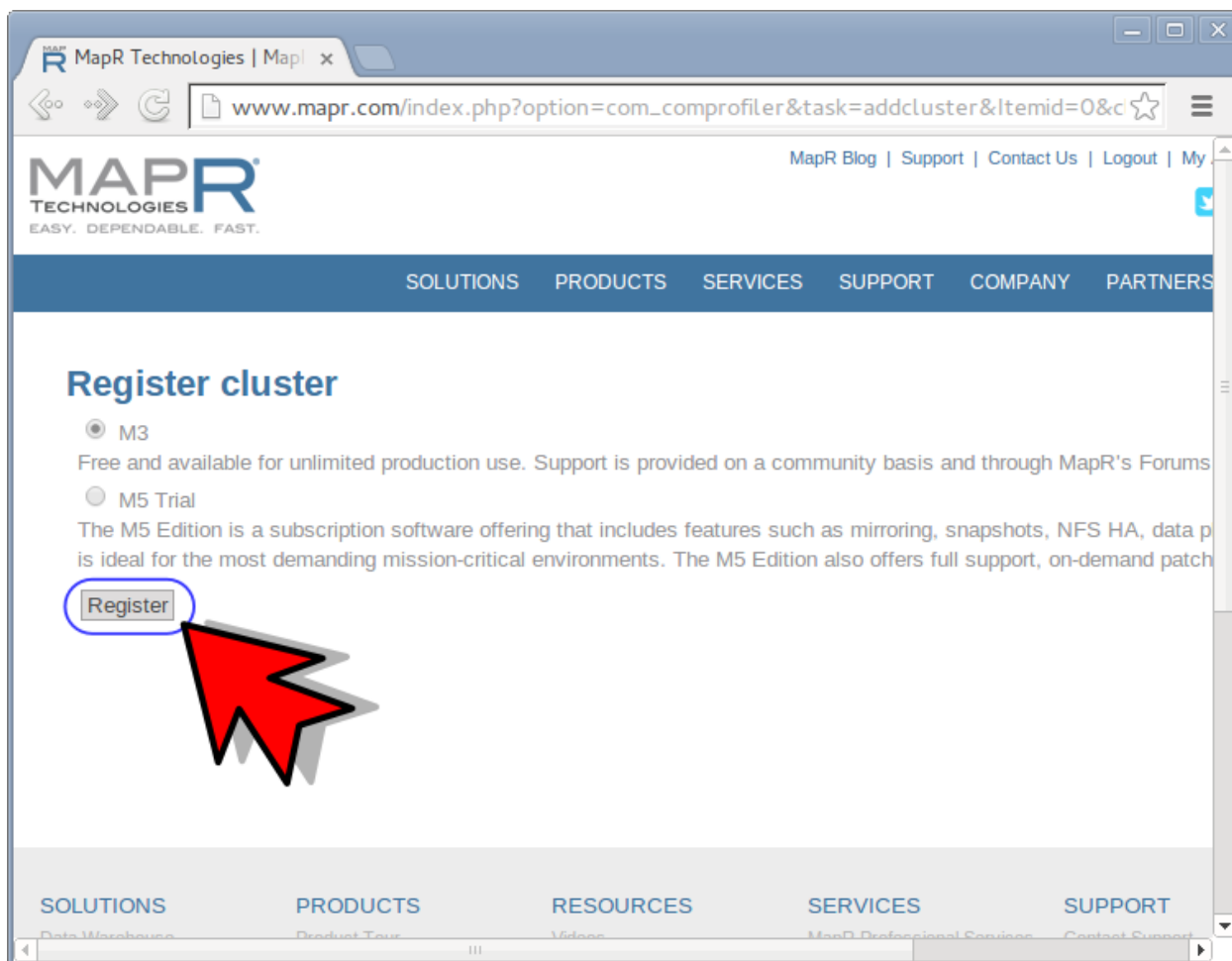
Click the **Add license from web** button.



If not yet done, create a MapR account (set the remember me option to skip this step next time).

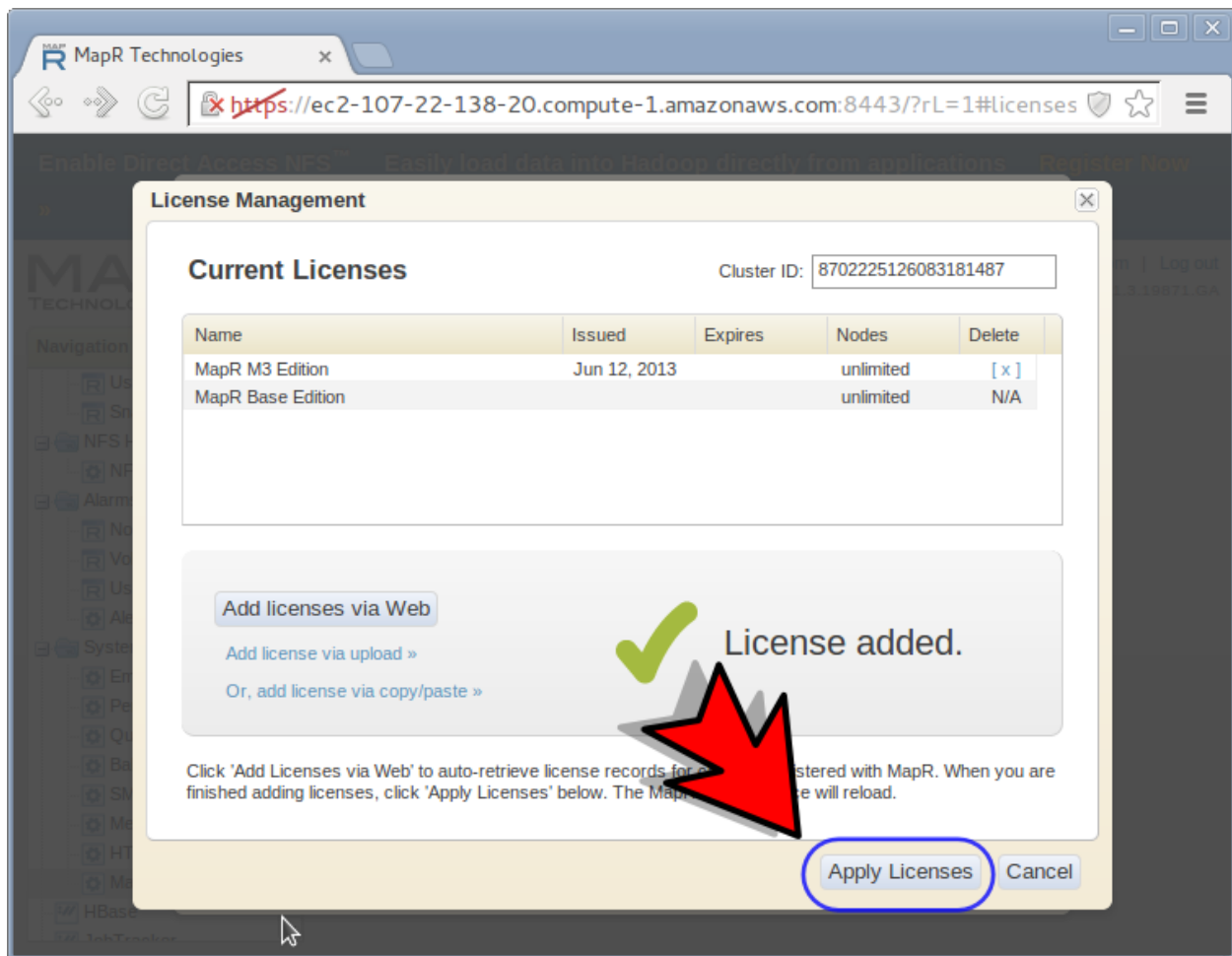
---

Select **M3** license and click the **Register** button.



Click the **Return to your MapR Cluster Ui** link.

Click the **Apply Licenses** button.



Return to the console and press ENTER key to continue...

```

Terminal
File Edit View Search Terminal Help

TASK: [Restart apache] *****
changed: [10.159.56.223]
waiting for web ui to be ready...
testing: https://ec2-54-226-32-126.compute-1.amazonaws.com:8443
testing: https://ec2-54-226-32-126.compute-1.amazonaws.com:8443

Your master node is ready...

***** Please Install Free License *****

1. Log in to MapR Web UI...
   url: https://ec2-54-226-32-126.compute-1.amazonaws.com:8443
   username: root
   password: mapr

2. Click "Add License via Web" button and follow on-screen
   instructions to add a free M3 license

3. Return to this console and press any key to continue...

After installing M3 license, PRESS ANY KEY to start launching worker nodes...

```

When the script finishes, the MapR control panel should look like this (one green square for each node you requested).

MapR - iwct - Dashboard

Cluster Name: iwct

Logged in as root | [Blog](#) | [Support](#) | [Help](#) | [mapr.com](#) | [Log out](#)

Manage Licenses v. 2.1.3.19871.GA

Navigation

- Cluster
- MapR-FS
- NFS HA
- Alarms
- System Settings
  - HBase
  - JobTracker
  - CLDB
  - Terminal
  - Nagios

Dashboard

Cluster Heat Map - 5 Nodes on 2 Racks [Show Legend](#) [Health](#)

1: /cldbonly  
 4: /data/us-east-1b/rack001

Alarms - No Alarms Raised

Cluster Utilization [Savings: 0%](#)

	%	Utilized	Total
CPU	1%	1 cores	160 cores
Memory	29%	83.9GB	292.6GB
Disk Space	1%	98.0GB	13.9TB

MapReduce

Running Jobs	0
Queued Jobs	0
Running Tasks	0
Running Map Tasks	0
Running Reduce Tasks	0
Map Task Capacity	120
Reduce Task Capacity	112
Map Task Prefetch Capacity	120
Blacklisted Nodes	0

Services

Actv	Stby	Stop	Fail	Total

### 1.1.5 Run pipeline

Open a terminal (right click on desktop and select “Open in terminal” from the context menu) and enter the following command.

```
iwct_run_pipeline
```

Next you will be prompted to select a dataset from the list of datasets you previously imported using the `iwct_import_dataset` command.

### 1.1.6 View Results

When the `iwct_run_pipeline` command completes, you will have a `results/html` directory with a self-contained static website that can be used to view the results. Simply copy the directory to any webserver to publish your image-graph viewer on the web.

An easy way to view the result locally is to start a python webserver for the html directory like this:

```
cd /home/ubuntu/Desktop/datasets/{dataset_name}/results/html
python -m SimpleHTTPServer
```

Now open a webbrowser to the url output by the python command (e.g. <http://localhost:8000>).

---

**Note:** The viewer inside the html directory is self-contained and portable. Simply copy the html directory to any webserver to share your image-graph viewer with others.

---

### 1.1.7 Shutdown

When the you are finished running the image webs pipeline, don’t forget to terminate the cluster. Otherwise, you may end up with a LARGE bill.

```
cirrus_cluster_cli destroy
```

**Warning:** All files on the `maprfs://` file system will be lost, be sure any data you wish to retain is copied to the IWCT workstation.

When you are finished using the workstation, click the “Turn Off” button in the IWCT Console. When your workstation is shut-off, you are not charged for instance-hours but continue to pay for retaining the instance storage (on an EBS volume). You can later return to the IWCT Console and click “Turn On” to start working where you left off. When you are done with the workstation and no longer wish to retain the storage, Open the **Options** menu and click **Destroy** to delete the workstation and storage.

**Warning:** You are responsible for any charges you accrue for use of AWS resources. When you complete a work session, always login to your [AWS management console](#) and check that your workstation instances are stopped and your cluster instances terminated to avoid accidentally wasting resources.



## 1.2 Manual

### 1.2.1 Workstation

#### Add Storage

This section explains how to add storage space.

1. Login to the IWCT Console
2. Click the “Options” button next to your workstation
3. Click the “Add Storage” item from the options menu
4. Follow the on-screen instructions
5. Start and connect to verify workstation storage was increased

### 1.2.2 Cluster

- How to create, manager, destroy cluster

### 1.2.3 Estimating costs

The IWCT software is opensource and free to use. However, you pay amazon for whatever resources you use.

**workstation** The workstation costs \$0.14/hr when running (or about \$1 for typical workday). The workstation can be turned on and off as needed (and the data persists). The workstation storage costs < \$1 per month. When you are done using a workstation and no longer want to pay to keep the data, you can destroy it.

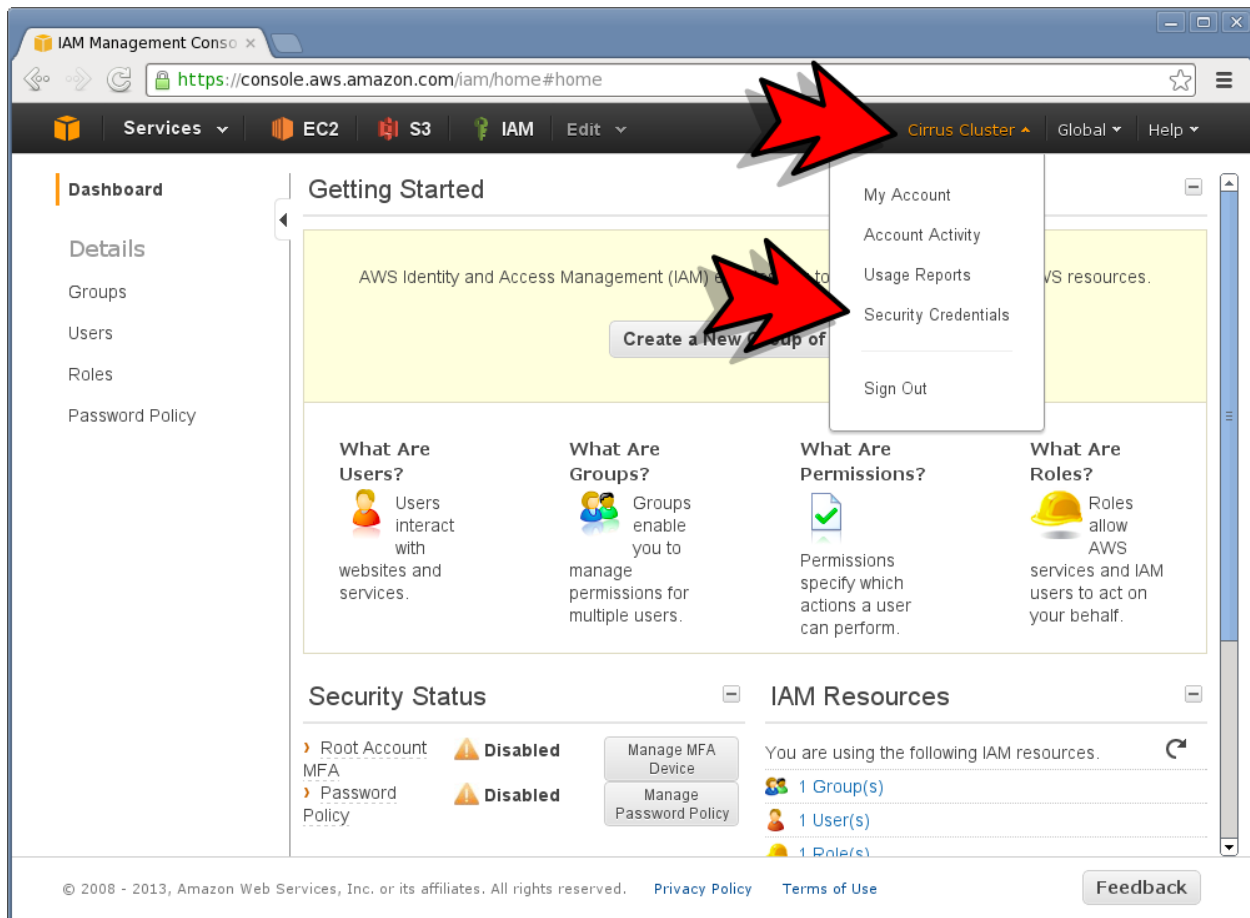
**cluster** You can start a private cluster from your workstation in minutes when you need to compute, copy the output from the cluster FS back to your remote desktop, and destroy the cluster. The cost of running a cluster depends depends on the number of nodes and the type of nodes.

## 1.3 Guides

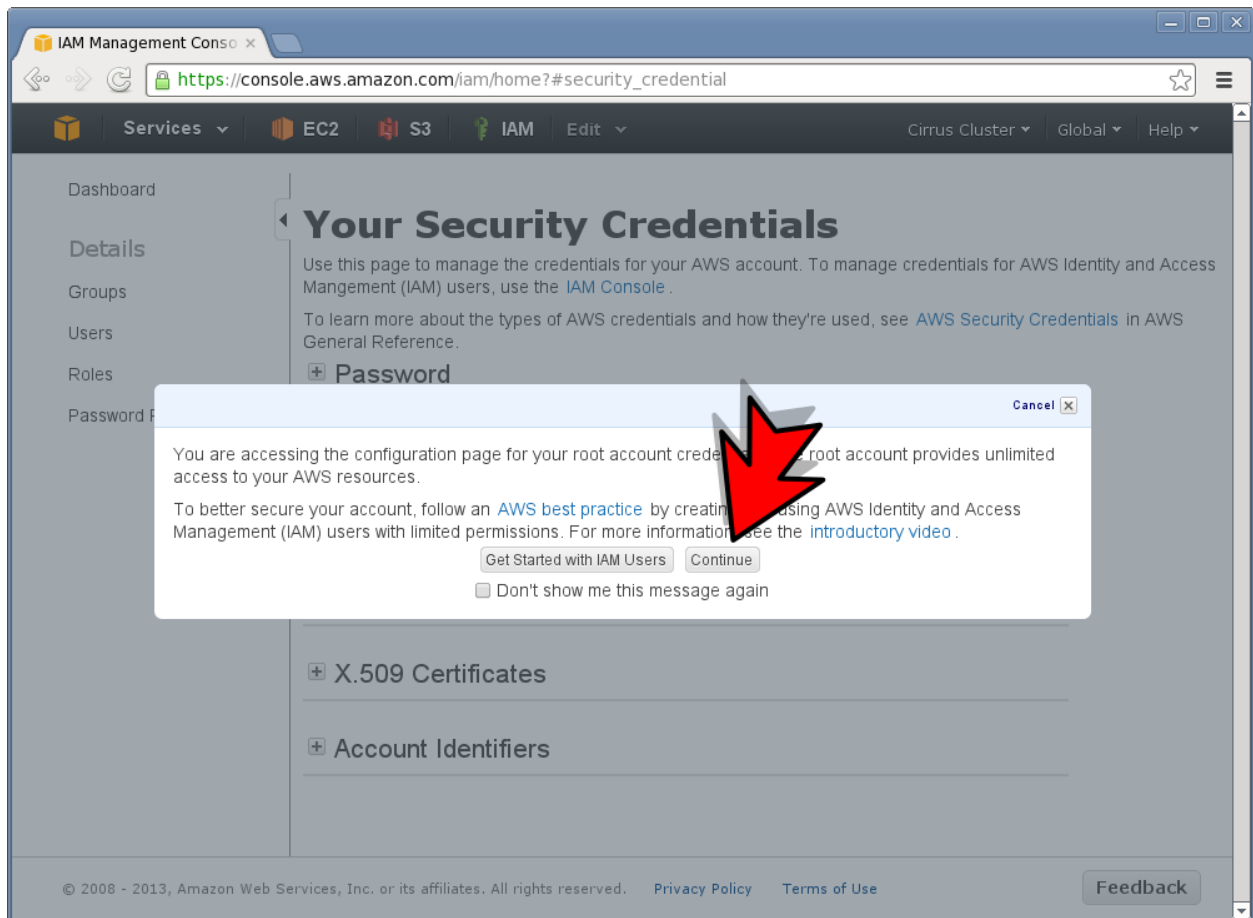
### 1.3.1 How to get your AWS Access Key

Go to <https://console.aws.amazon.com>

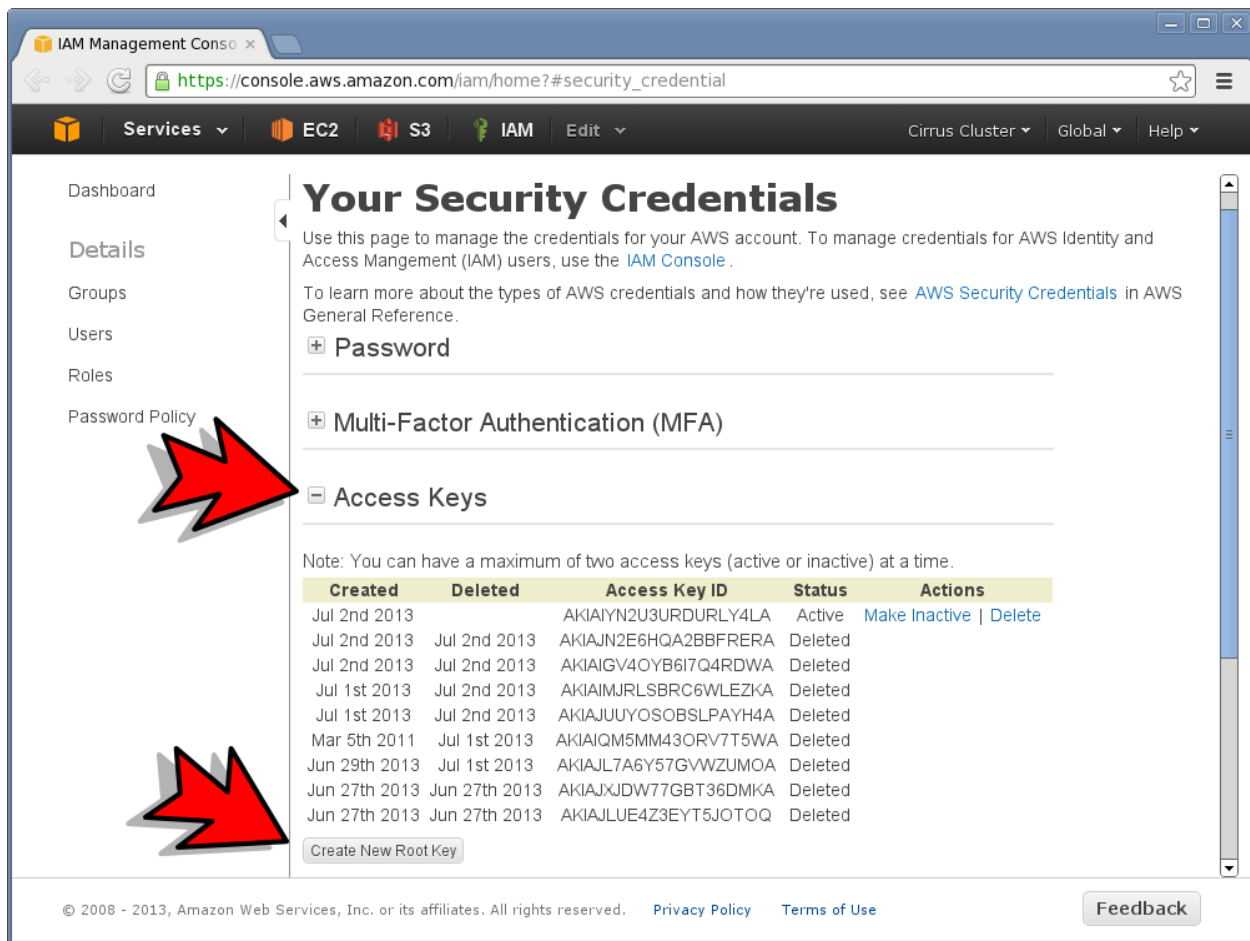
Click on your username in the top navigation bar and then click the **Security Credentials** option in the drop down menu.



Click the **Continue** button to dismiss the dialog.



Click the + button next to **Access Keys** and then click the **Create new root key** button.



**Your Security Credentials**

Use this page to manage the credentials for your AWS account. To manage credentials for AWS Identity and Access Management (IAM) users, use the [IAM Console](#).

To learn more about the types of AWS credentials and how they're used, see [AWS Security Credentials](#) in AWS General Reference.

- ☒ Password
- ☒ Multi-Factor Authentication (MFA)
- ☒ Access Keys

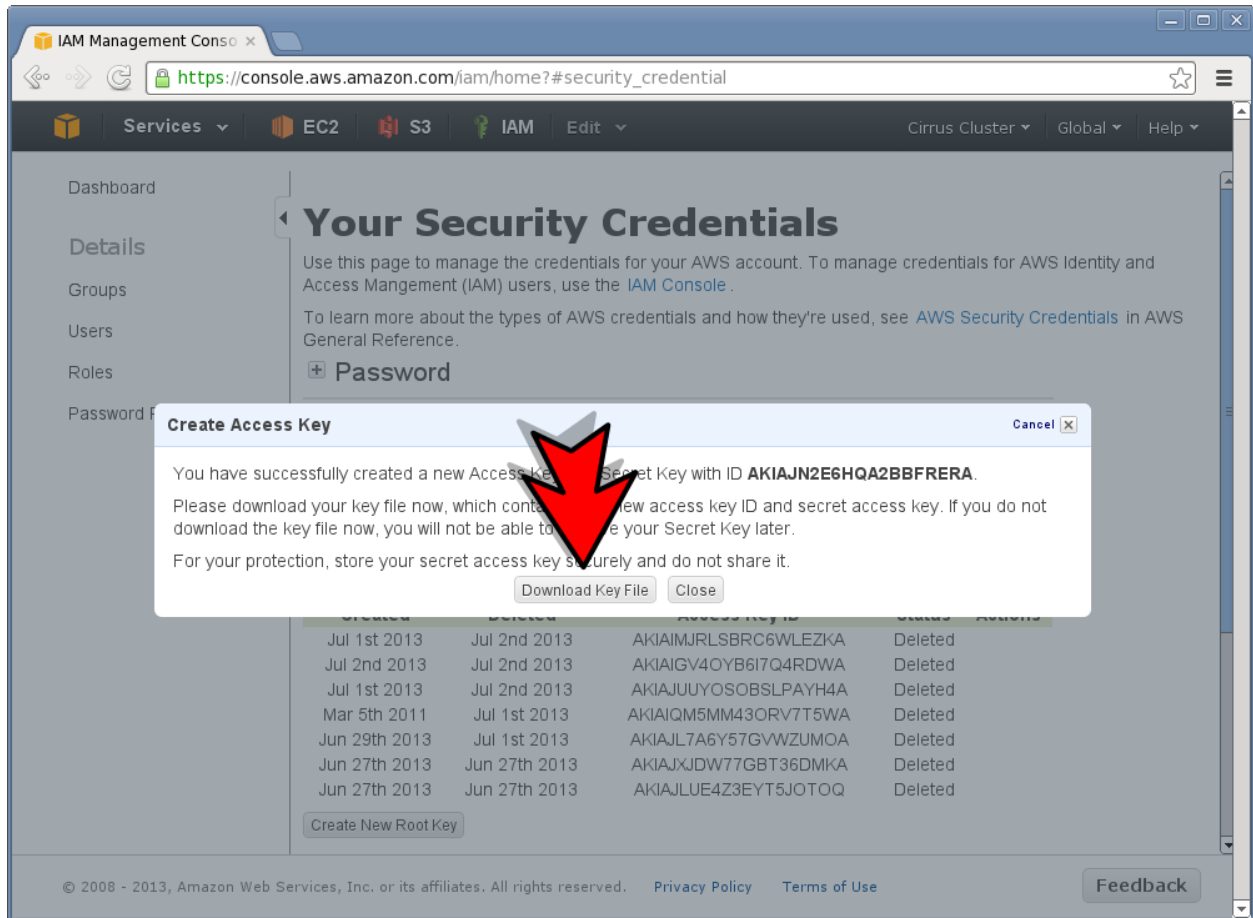
Note: You can have a maximum of two access keys (active or inactive) at a time.

Created	Deleted	Access Key ID	Status	Actions
Jul 2nd 2013		AKIAIYN2U3URDURLY4LA	Active	<a href="#">Make Inactive</a>   <a href="#">Delete</a>
Jul 2nd 2013	Jul 2nd 2013	AKIAJN2E6HQA2BBFRERA	Deleted	
Jul 2nd 2013	Jul 2nd 2013	AKIAIGV4OYB6I7Q4RDWA	Deleted	
Jul 1st 2013	Jul 2nd 2013	AKIAIMJRLSBR6WLEZKA	Deleted	
Jul 1st 2013	Jul 2nd 2013	AKIAJUUYOSOBSPAYH4A	Deleted	
Mar 5th 2011	Jul 1st 2013	AKIAIQM5MM43ORV7T5WA	Deleted	
Jun 29th 2013	Jul 1st 2013	AKIAJL7A6Y57GVWZUMOA	Deleted	
Jun 27th 2013	Jun 27th 2013	AKIAJXJDW77GBT36DMKA	Deleted	
Jun 27th 2013	Jun 27th 2013	AKIAJLUE4Z3EYT5JOTOQ	Deleted	

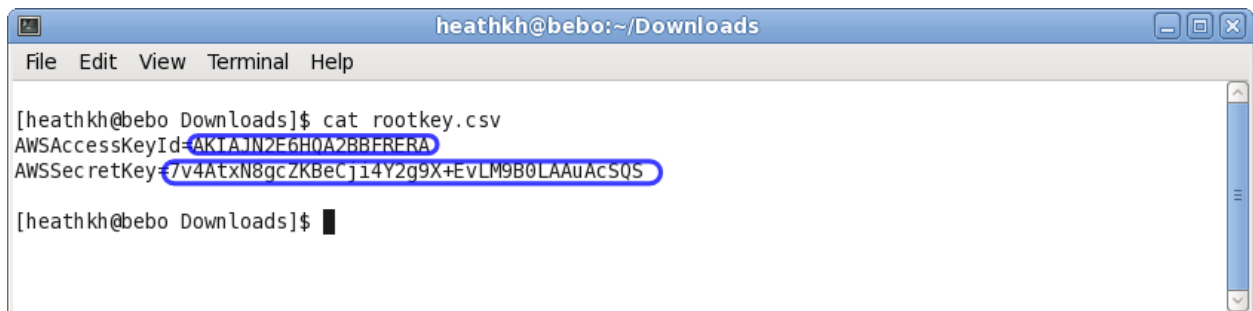
[Create New Root Key](#)

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#) [Feedback](#)

Click the **Download key file** button.



Open the downloaded file **rootkey.csv** to copy and paste your Access Key ID and Secret.



### 1.3.2 How to transfer files to and from IWCT Workstations

#### Connect To Server

Use the “Connect To Server” feature to easily drag and drop files from a remote server (via ssh, ftp, windows share). To do this, click **Places** in the bar at the top and click **Connect To Server...** from the drop down menu.

## SCP

Example to copy a file

```
scp username1@source_host:directory1/filename1 username2@destination_host:directory2/filename2
```

Example to copy a directory recursively

```
scp -r username1@source_host:directory1 username2@destination_host:directory2
```

---

**Note:** Use the **Options** menu in the *IWCT workstation console* to generate the SSH or SCP command to copy files to / from your workstation.

---