



Introduction to Research on AWS Hands-On Lab

Getting Started with S3 and Cloud9

Workshop Overview

This lab simulates a research workflow. You will create an S3 bucket to store data, launch a Cloud9 IDE as a virtual workstation, install R, download and process a sample dataset, and upload the resulting graph to S3 for permanent storage.

Create a Bucket in S3

To upload your data (photos, videos, documents etc.) to Amazon S3, you must first create an S3 bucket in one of the AWS Regions. You can then upload your data objects to the bucket. Every object you store in Amazon S3 resides in a bucket. You can use buckets to group related objects in the same way that you use a directory to group files in a file system.

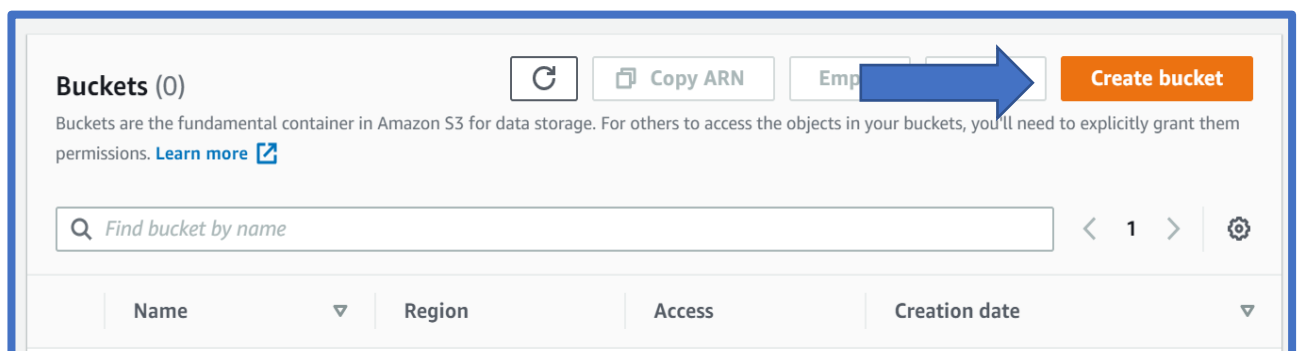
Note: You are not charged for creating a bucket; you are only charged for storing objects in the bucket and for transferring objects in and out of the bucket.



Whilst S3 does qualify for the Free Tier, data transfer to the internet may incur a small charge.

Note: Transfer IN does not incur a charge, but Transfer OUT does. Data transfers are free if you are within the same region and within the same availability zone, and use a private IP address. Data transfers to other regions or services will have a cost associated with them. See <https://aws.amazon.com/s3/pricing/> for more details.

1. Sign into the AWS Management Console and open the Amazon S3 console at <https://console.aws.amazon.com/s3>.
2. Click **Create Bucket**. The **Create a Bucket** wizard will open.



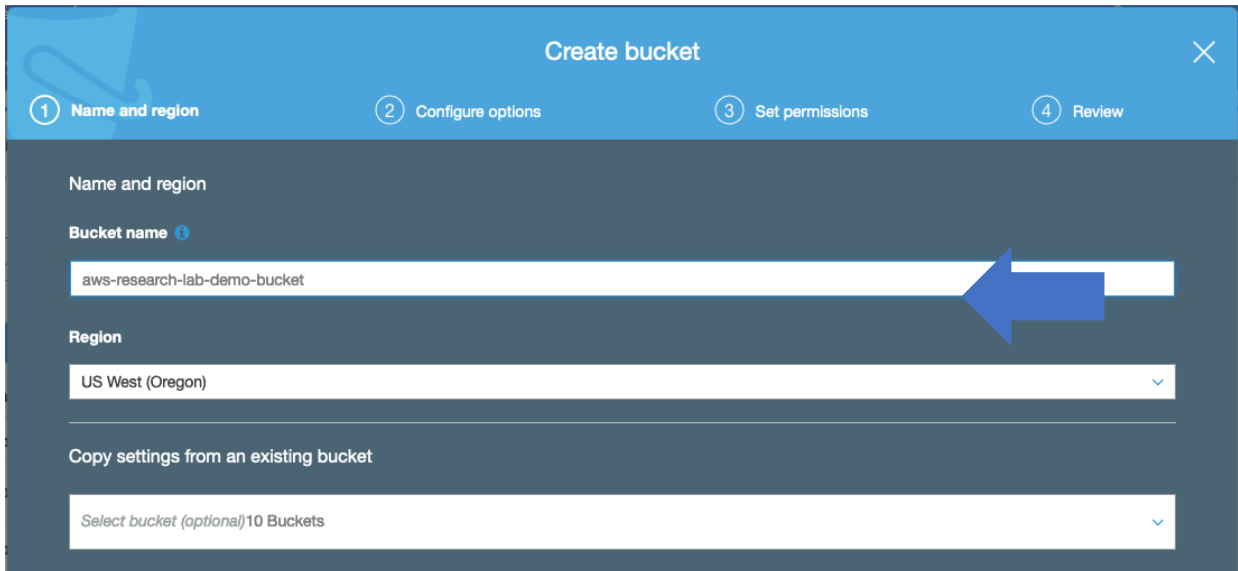
Note: Enter a bucket name in the Bucket Name field. Since S3 is a global service, the bucket name you choose must be unique across all existing bucket names in Amazon S3. One way to do that is to prefix your bucket names with your organization's name.

Bucket names must comply with the following requirements. The bucket name:

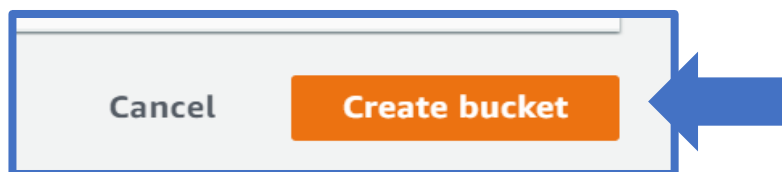
- Must be unique across all of Amazon S3
- Must be between 3 and 63 characters long
- Must not contain uppercase characters
- Can contain only lower-case characters, numbers, periods, and dashes
- Must Start with a lowercase letter or number
- Must Not contain underscores, end with a dash, have consecutive periods, or use dashes adjacent to periods.
- Cannot be formatted as an IP address ([198.51.100.24](#)).

Note: There might be additional restrictions on bucket names based on the region your bucket is in or how you intend to access the object. Once you create a bucket, you cannot change its name. In addition, the bucket name is visible in the URL that points to the objects stored in the bucket. Make sure the bucket name you choose is appropriate.


1. In the **Region** drop-down list box, select a region. Choose a Region close to you to minimize latency and costs and address regulatory requirements. Objects stored in a Region never leave that Region unless you explicitly transfer them to another Region




2. By default, all buckets are set to “block public access”. At this time, leave the default setting of block all public access. [Learn More](#)
3. Under Advanced settings, you have the option to enable Object Lock. Object Lock stores objects using a write-once-read-many (WORM) model to prevent objects from being deleted or overwritten for a fixed amount of time or indefinitely. For today’s lab, we will leave this disabled. [Learn More](#)
4. Click Create Bucket to create your bucket.



5. Once your bucket is created, you will be able to see it in your Buckets list. You can also see the region your bucket is in along with the access type.

S3 buckets  [Discover the console](#)


All access types 







+ Create bucket

Edit public access settings

Empty

Delete

11 Buckets 4 Regions 

<input type="checkbox"/> Bucket name 	Access  	Region 	Date created 
<input type="checkbox"/>  aws-research-lab-demo-name	Bucket and objects not public	US West (Oregon)	Aug 4, 2020 2:52:57 PM GMT-0700

Well done – you have created your first bucket in Amazon S3!

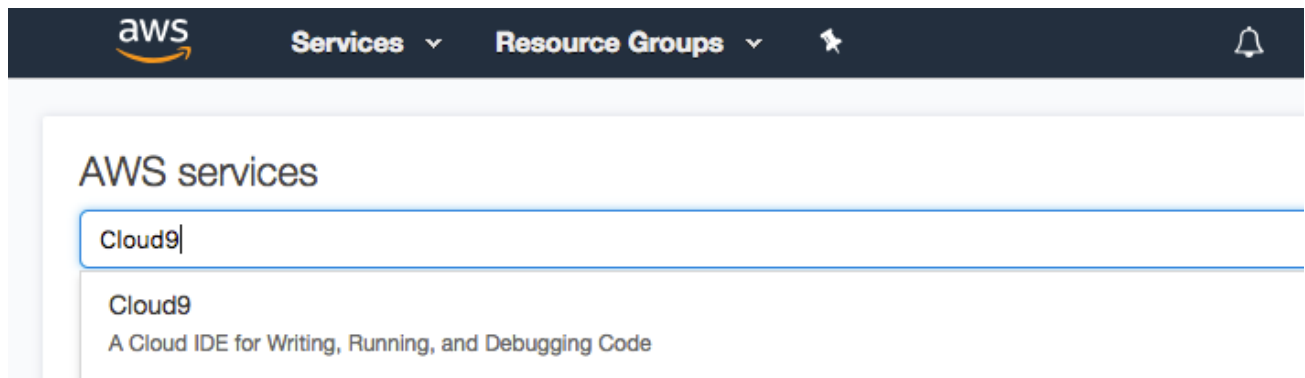
Launch a Cloud9 Environment

For this lab we will be using AWS Cloud9, a cloud-based integrated development environment (IDE) that lets you write, run, and debug your code with just a browser. Each AWS Cloud9 IDE is backed by an EC2 instance. Cloud9 is an easy way to get access to the compute power of EC2, and will automatically shut down when not in use.



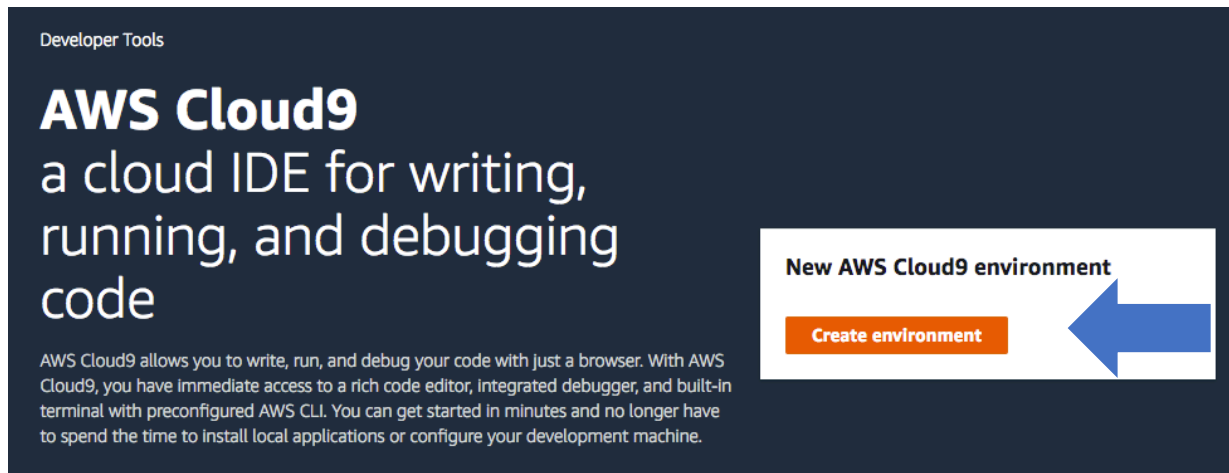
Upon logging into your AWS Console, you should ALWAYS check which region you are operating in. This can be found in the top right of your Console window.

1. Sign into the AWS Management Console and On the AWS Console home page, type Cloud9 into the service search bar and select it.

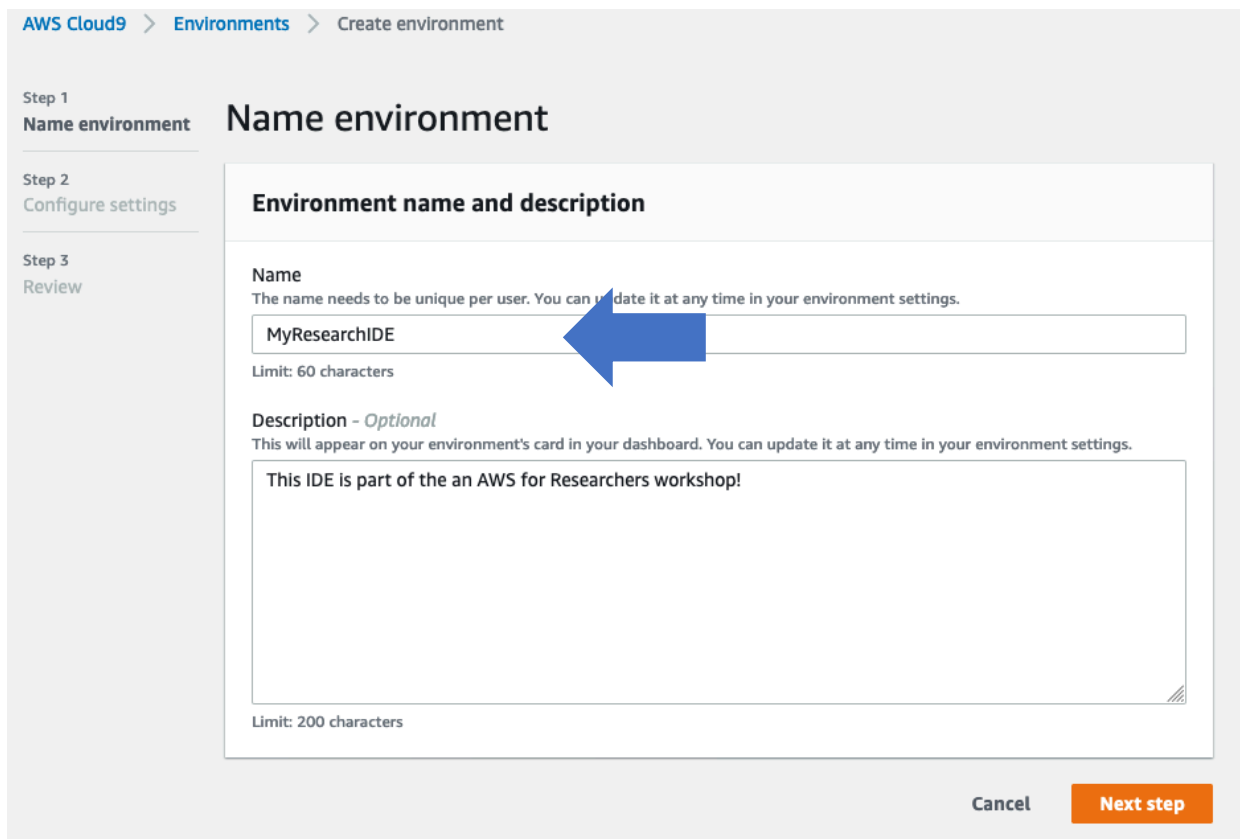


2. In the upper-right corner of the AWS Management Console, confirm you are in the desired AWS region (e.g., Oregon).

3. Click **Create Environment** on the Cloud9 home page:



4. Name your environment **MyResearchIDE** with any description you'd like, and click Next Step:

A screenshot of the "Name environment" form in the AWS Cloud9 console. The breadcrumb trail at the top reads "AWS Cloud9 > Environments > Create environment". On the left, a sidebar shows three steps: "Step 1 Name environment" (active), "Step 2 Configure settings", and "Step 3 Review". The main heading is "Name environment". Below it, the section "Environment name and description" contains two fields. The "Name" field is a text input with the value "MyResearchIDE" and a blue arrow pointing to it from the right. Below the input is the text "Limit: 60 characters". The "Description - Optional" field is a text area with the value "This IDE is part of the an AWS for Researchers workshop!". Below the text area is the text "Limit: 200 characters". At the bottom right, there are two buttons: "Cancel" and "Next step".

5. Leave the Environment settings as their defaults and click **Next Step**:

Environment settings

Environment type [Info](#)
Choose between creating a new EC2 instance for your new environment or connecting directly to your server over SSH.

☒ **Create a new instance for environment (EC2)**
Launch a new instance in this region to run your new environment.

☐ **Connect and run in remote server (SSH)**
Display instructions to connect remotely over SSH and run your new environment.

Instance type

☒ **t2.micro (1 GiB RAM + 1 vCPU)**
Free-tier eligible. Ideal for educational users and exploration.

☐ **t2.small (2 GiB RAM + 1 vCPU)**
Recommended for small-sized web projects.

☐ **m4.large (8 GiB RAM + 2 vCPU)**
Recommended for production and general-purpose development.

☐ **Other instance type**
Select an instance type.
t2.nano

Cost-saving setting
Choose a predetermined amount of time to auto-hibernate your environment and prevent unnecessary charges. We recommend a hibernation settings of half an hour of no activity to maximize savings.

After 30 minutes (default)

IAM role
AWS Cloud9 creates a service-linked role for you. This allows AWS Cloud9 to call other AWS services on your behalf. You can delete the role from the AWS IAM console once you no longer have any AWS Cloud9 environments. [Learn more](#)

AWSServiceRoleForAWSCloud9

► **Network settings (advanced)**

Cancel Previous step Next step



If you wanted a more powerful instance, you can select different instances from the "Other instance type" dropdown.

6. Click Create Environment:

Environment name and settings

Name
MythicalMysfitsIDE

Description
This IDE will be used to create the Mythical Mysfits website as part of an AWS tutorial.

Environment type
EC2

Instance type
t2.micro

Subnet
subnet-f9bd479e

Cost-saving settings
After 30 minutes (default)

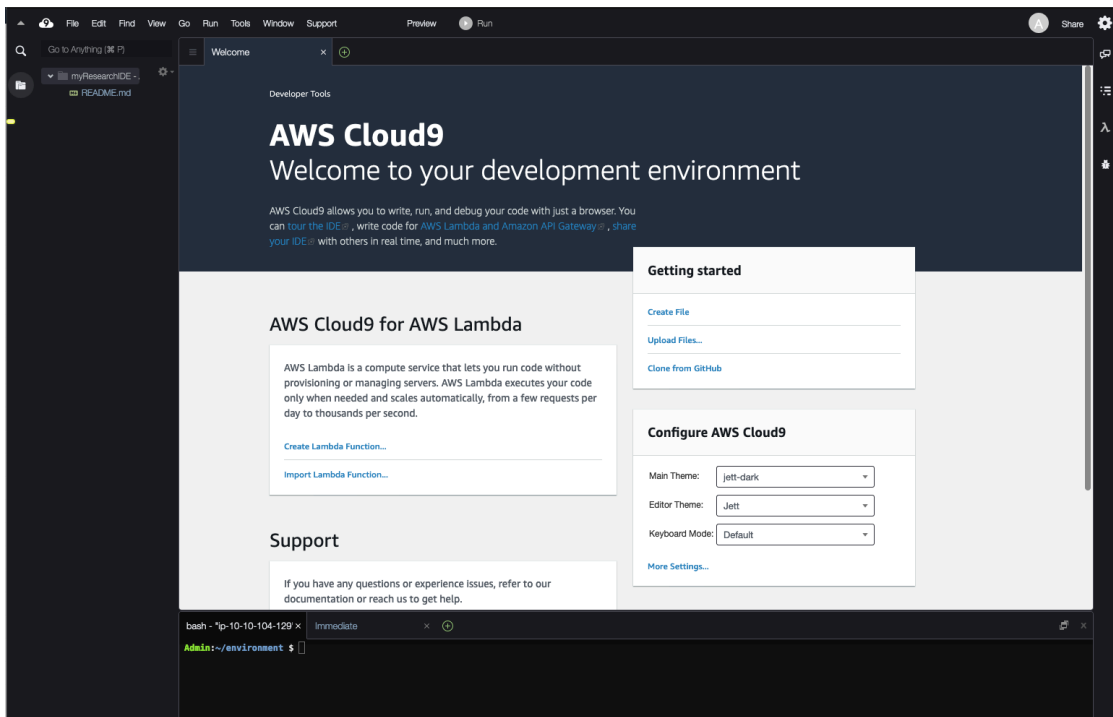
IAM role
AWSServiceRoleForAWSCloud9 (generated)

We recommend the following best practices for using your AWS Cloud9 environment

- Use **source control and backup** your environment frequently. AWS Cloud9 does not perform automatic backups.
- Perform regular **updates of software** on your environment. AWS Cloud9 does not perform automatic updates on your behalf.
- **Turn on AWS CloudTrail** in your **AWS account** to track activity in your environment. [Learn more](#)
- Only share your environment with **trusted users**. Sharing your environment may put your AWS access credentials at risk. [Learn more](#)

Cancel Previous step Create environment

7. When the IDE has finished being created for you, you'll be presented with a welcome screen that looks like this:



Install R on our Cloud9 IDE

In this example, we will launch an Amazon Linux 2 instance, bootstrap Apache/PHP, and install a basic web page that will display information about our instance.

In the bottom panel, you will see a terminal command line open and ready to use.

1. Enter the following in the console to install R.

```
Admin:~/environment $ sudo yum -y install R
```

2. Start R as root to install libraries

```
Admin:~/environment $ sudo -i R
```

3. Install the readr and ggplot2 packages with

`install.packages(c("readr", "ggplot2"))` When prompted for a mirror, choose
58: USA (OR) [https]

```
> install.packages(c("readr", "ggplot2"))
Installing package into '/usr/lib64/R/library'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
Secure CRAN mirrors

1: 0-Cloud [https]
2: Australia (Canberra) [https]
3: Australia (Melbourne 1) [https]
<snip>
57: USA (OH) [https]
58: USA (OR) [https]
59: USA (TN) [https]
60: USA (TX 1) [https]
61: Uruguay [https]
62: (other mirrors)

Selection: 58
```

4. Quit R by typing “`quit()`”

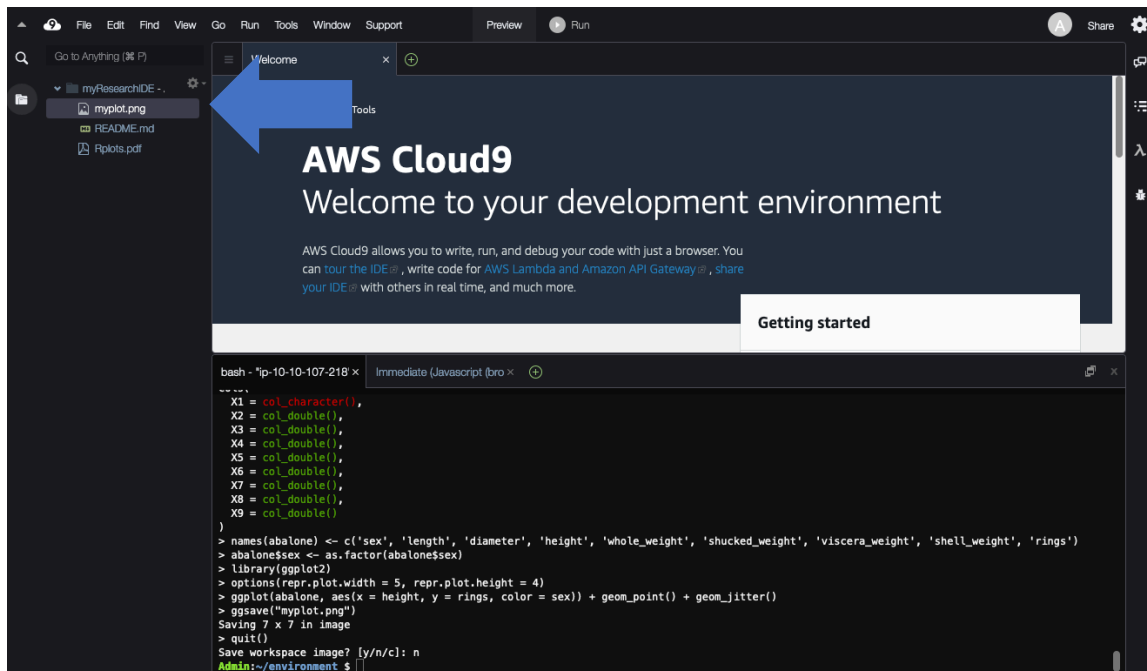
Run a computation and save the resulting plot

We will use the [Abalone](#) dataset from the [UCI Machine Learning Repository](#), and generate a plot of abalone ring count by height and sex.

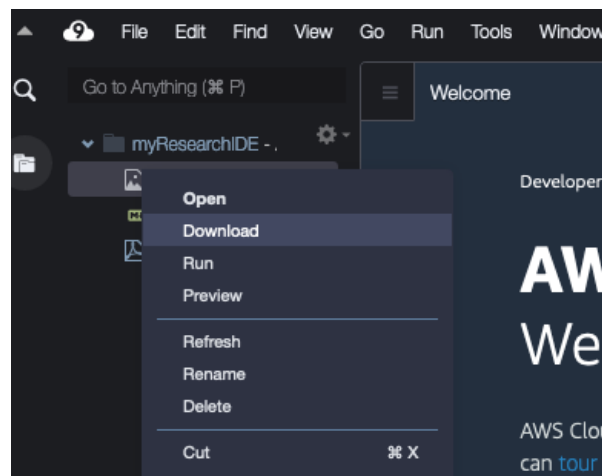
1. In the command line, type “R” to start R.
2. Copy and paste the following code to import the dataset and create a plot. Then quit R with `quit()`.

```
library(readr)
data_file <- 'http://archive.ics.uci.edu/ml/machine-learning-
databases/abalone/abalone.data'
abalone <- read_csv(file = data_file, col_names = FALSE)
names(abalone) <- c('sex', 'length', 'diameter', 'height', 'whole_weight',
'shucked_weight', 'viscera_weight', 'shell_weight', 'rings')
abalone$sex <- as.factor(abalone$sex)
library(ggplot2)
options(repr.plot.width = 5, repr.plot.height = 4)
ggplot(abalone, aes(x = height, y = rings, color = sex)) + geom_point() +
geom_jitter()
ggsave("myplot.png")
```

3. You will see “myplot.png” in the file browser on the right.



4. You can download that file by right clicking on the file.



5. We can also copy that file to our S3 bucket using the CLI. In the command line enter “aws s3 cp myplot.png s3://<Bucket name>” to copy to the S3 bucket.

```
python3.6 - "ip-10-10-107 x Immediate (Javascript (bro x +)
```

```
Admin:~/environment $ aws s3 cp myplot.png s3://aws-research-lab-demo-name
upload: ./myplot.png to s3://aws-research-lab-demo-name/myplot.png
Admin:~/environment $
```

6. Go to the S3 console and you can verify that the file was uploaded successfully.

aws-research-lab-demo-name

Overview

Properties

Permissions

Management

Access points

Q Type a prefix and press Enter to search. Press ESC to clear.

Upload


Create folder

Download

Actions

US West (Oregon) ↻

Viewing 1 to 1

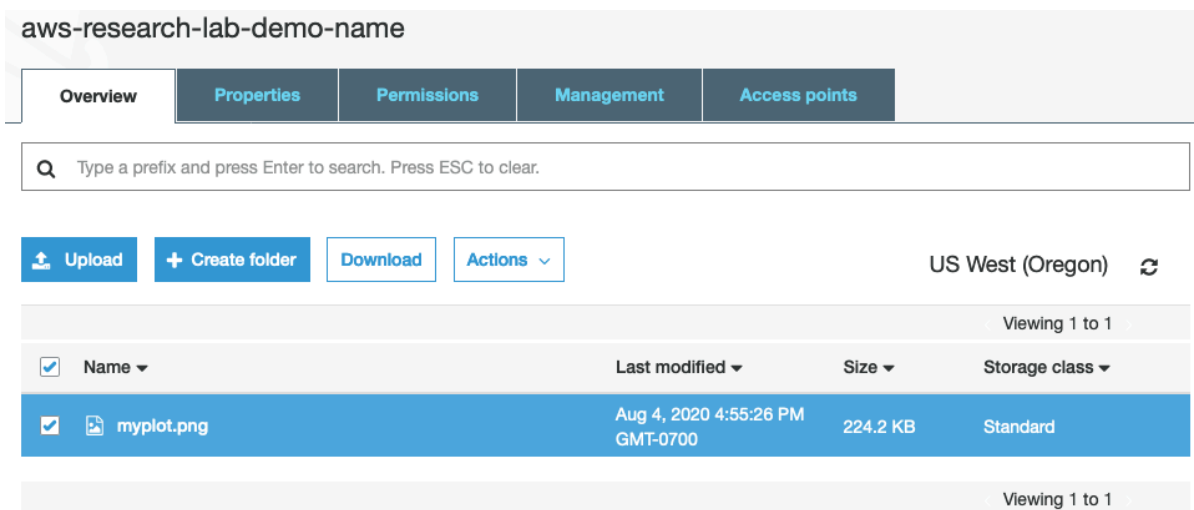
<input type="checkbox"/>	Name ▾	Last modified ▾	Size ▾	Storage class ▾
<input type="checkbox"/>	 myplot.png	Aug 4, 2020 4:55:26 PM GMT-0700	224.2 KB	Standard

Viewing 1 to 1

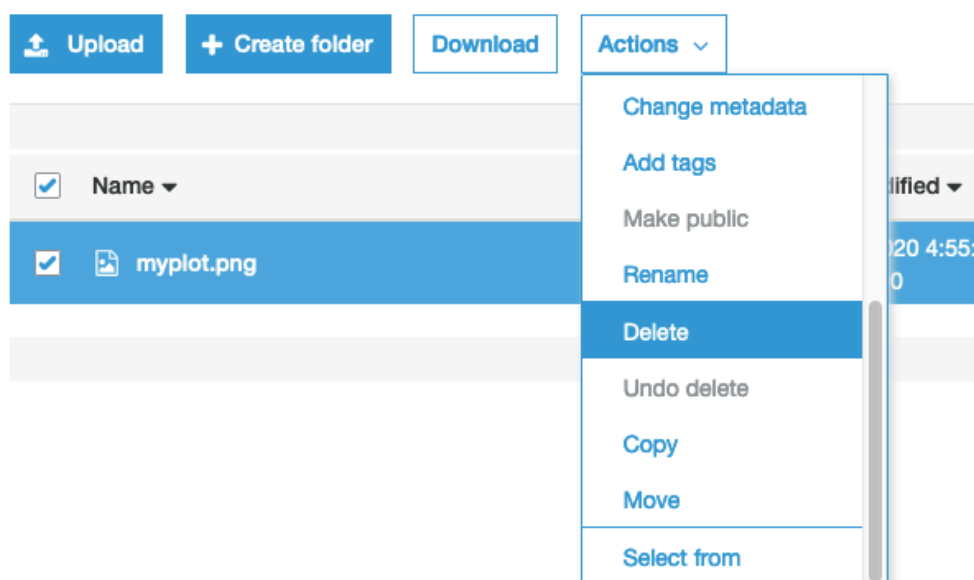
Delete an Object and Bucket

You have added an object to a bucket. Now, you can delete it and the bucket it is in. If you no longer need to store the objects you uploaded and moved while going through this guide, you should delete them so you do not incur further charges on those objects.

1. In the Amazon S3 console, click on the link representing the bucket containing the object(s) you want to delete. Then select the checkboxes for the object(s) you would like to delete.



2. Click the **Actions** button. Then select **Delete**. To confirm the action in the **Delete objects** dialogue, click **Delete**.



3. Navigate back to the S3 console and select the bucket icon of the bucket you want to delete (not the link to its right), and at the top of the page, click **Delete**. Confirm the deletion by typing its name verbatim at the **Delete bucket** prompt.

S3 buckets Discover the console

Search for buckets All access types

+ Create bucket Edit public access settings Empty Delete

11 Buckets 4 Regions

<input type="checkbox"/>	Bucket name	Access	Region	Date created
<input checked="" type="checkbox"/>	aws-research-lab-demo-name	Bucket and objects not public	US West (Oregon)	Aug 4, 2020 2:52:57 PM GMT-0700



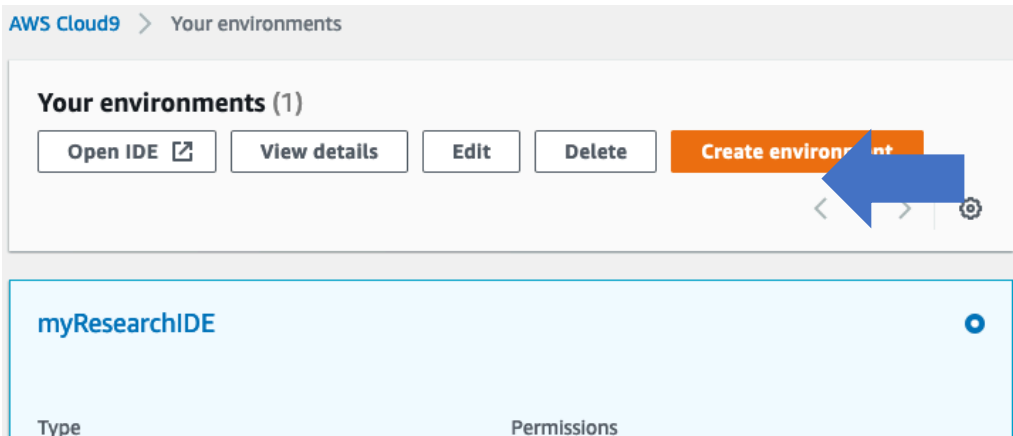
To delete a bucket, you must first delete all of the objects in it. If you haven't deleted all of the objects in your bucket, do that now.

Well done, your bucket is now deleted!

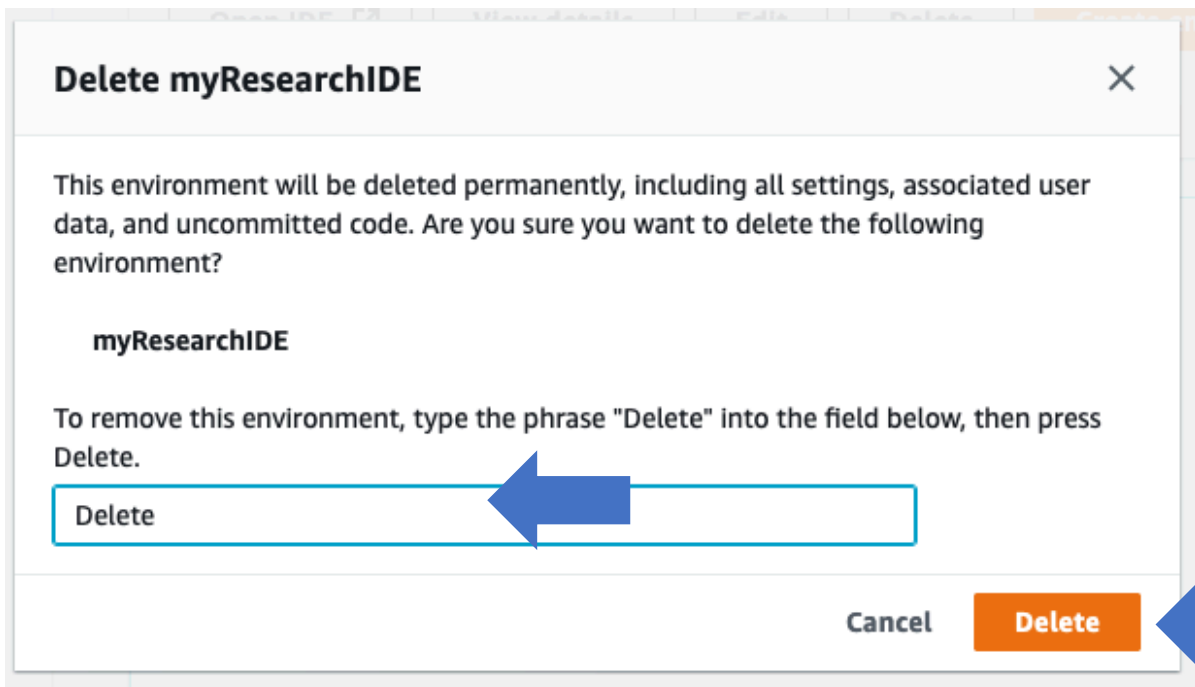
Delete a Cloud9 IDE

While the Cloud9 IDE will automatically shut down when it's not being used, it's good to completely delete the IDE if you don't intend to use it again. Even shut down, you will be billed for the storage that the Cloud9 IDE has provisioned.

1. Now we need to remove our AWS Cloud9 IDE. Navigate to the Cloud9 Dashboard, select your IDE and click the "Delete" button.



2. Type "Delete" in the pop-up and click the "Delete" button.



Congratulations you have deleted your Cloud9 IDE