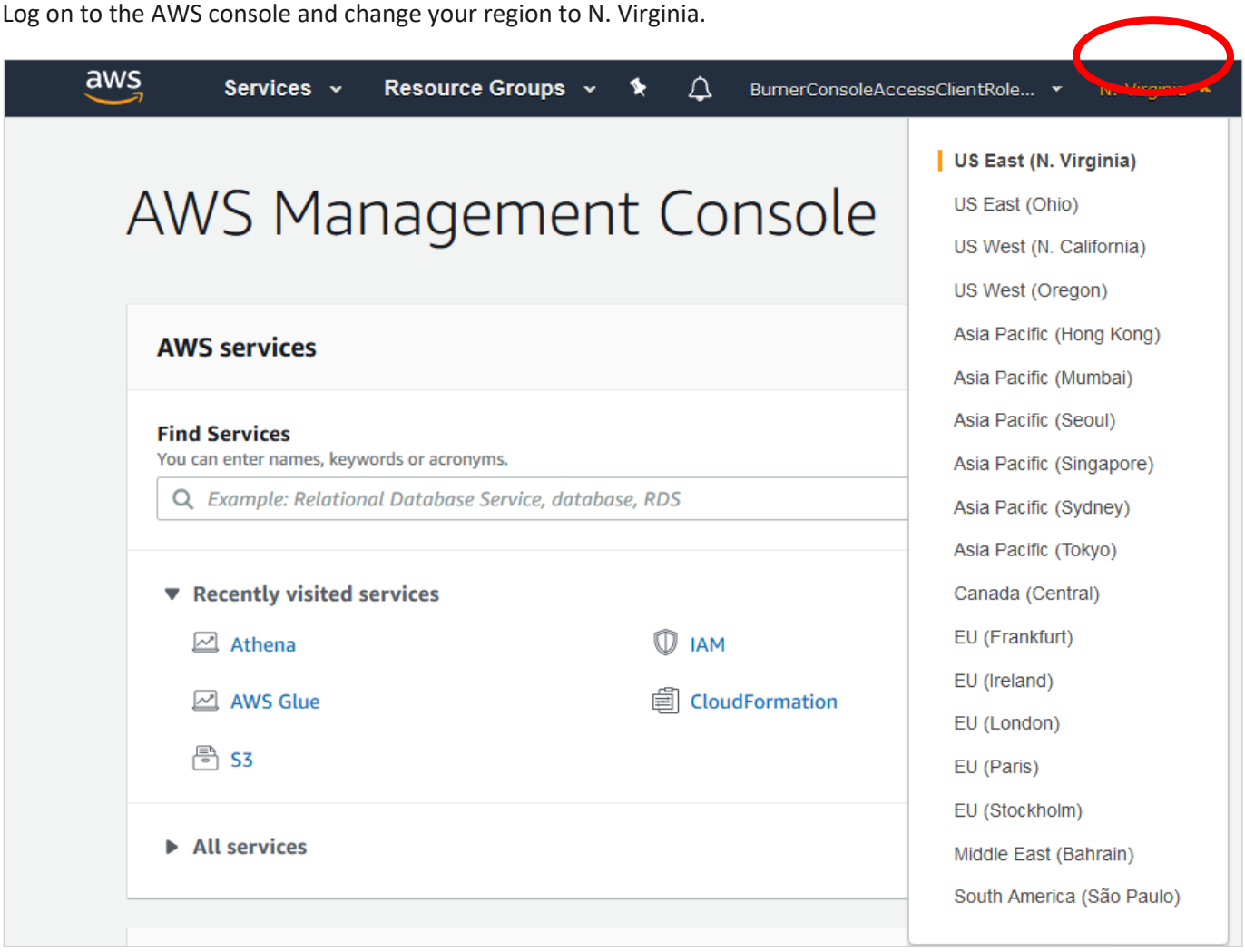
Section 1

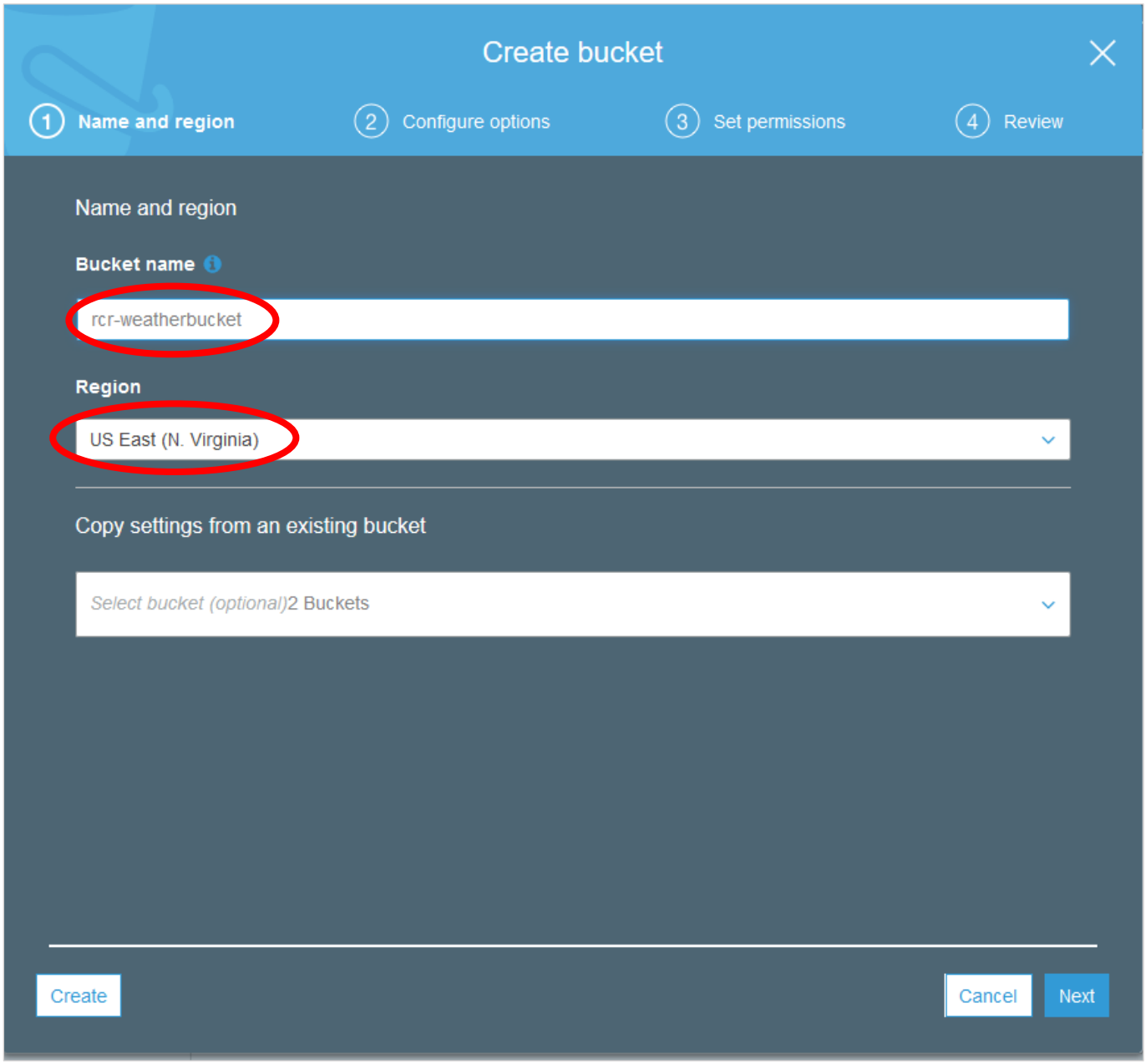
Create S3 bucket and subdirectories

In this section, you will create an S3 bucket and associated folders necessary to complete this lab. These folders will be used to provide a repository for the results from Athena, which you will be generating in the next section. You will also create a folder to hold a CSV file that lists cities you could potentially use for the event. This CSV will be "joined" to another CSV in an account that holds the temperature data.

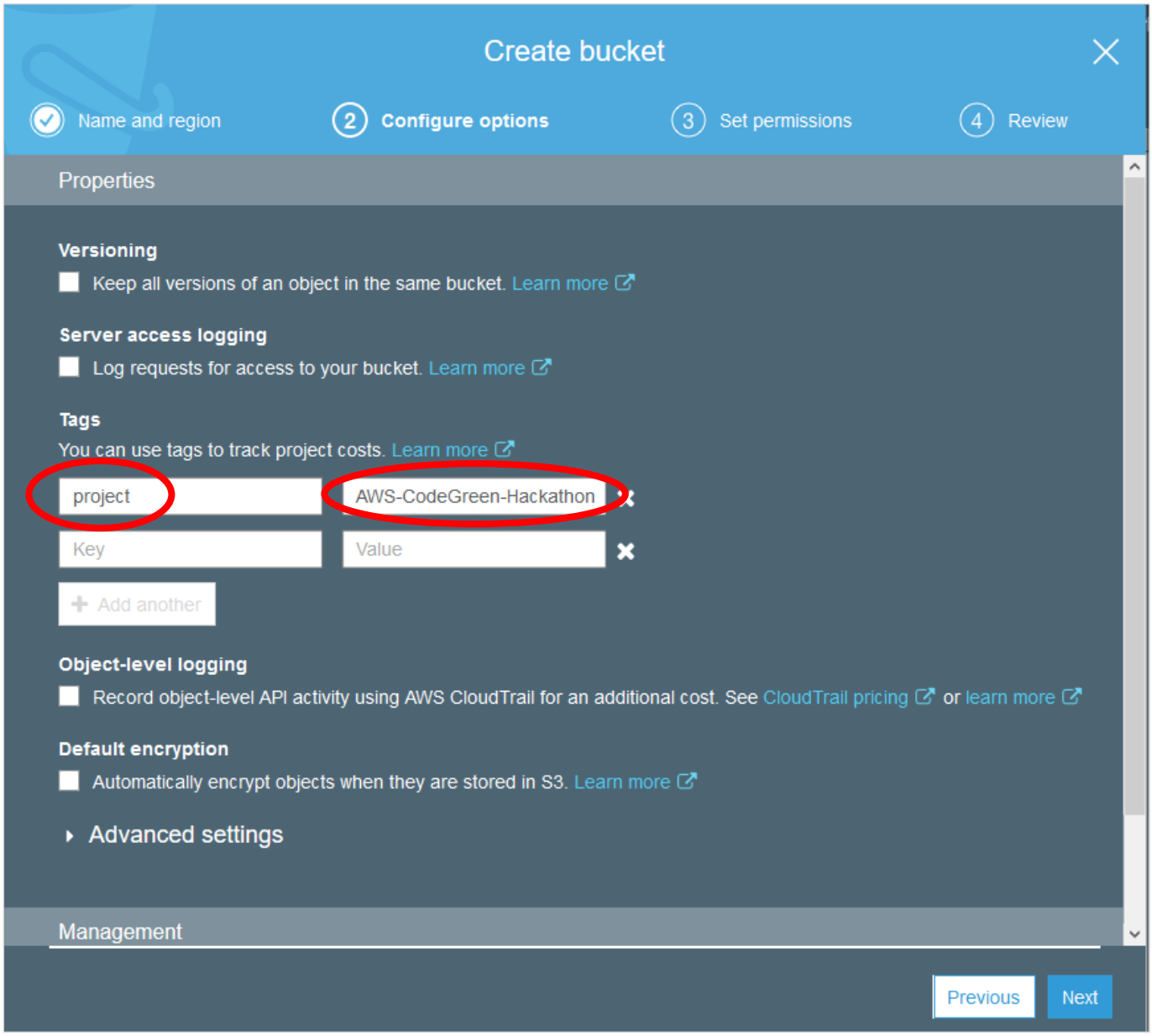
1) Log on to the AWS console and change your region to N. Virginia.



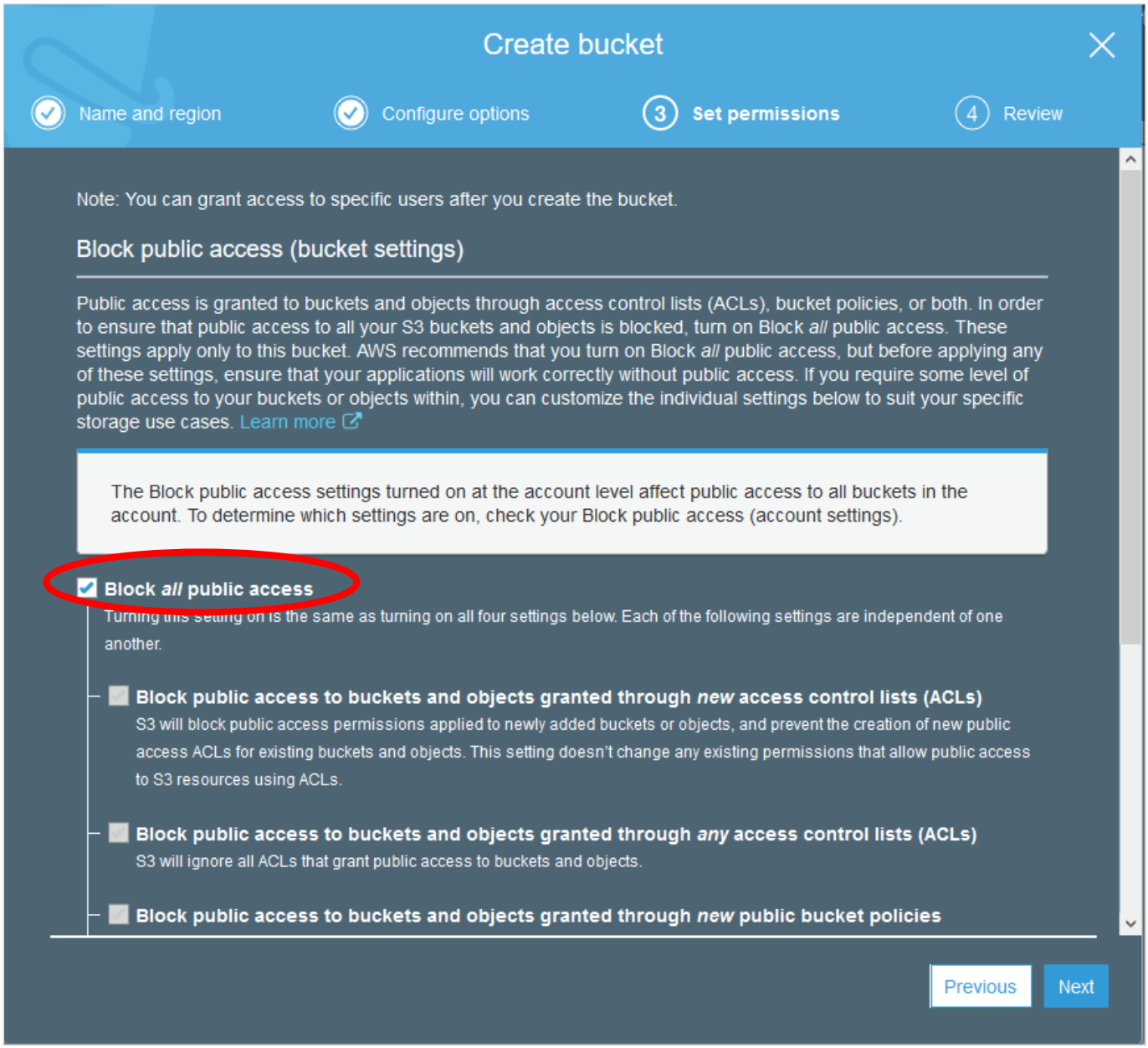
2) In the Find Services field, search for S3 and navigate to the S3 dashboard, then click on "Create bucket". All S3 buckets must have a globally unique name and must comply with DNS naming conventions, generally you should use lower-case letters and no underscores (more information). Select the region N. Virginia and a unique bucket name and click "Next".



3) Add a tag with “project” as the key and “AWS-CodeGreen-Hackathon” as the value, and click “Next”.



4) You don't need this bucket to be publicly available, so accept the default ("Block *all* public access"), and click "Next".

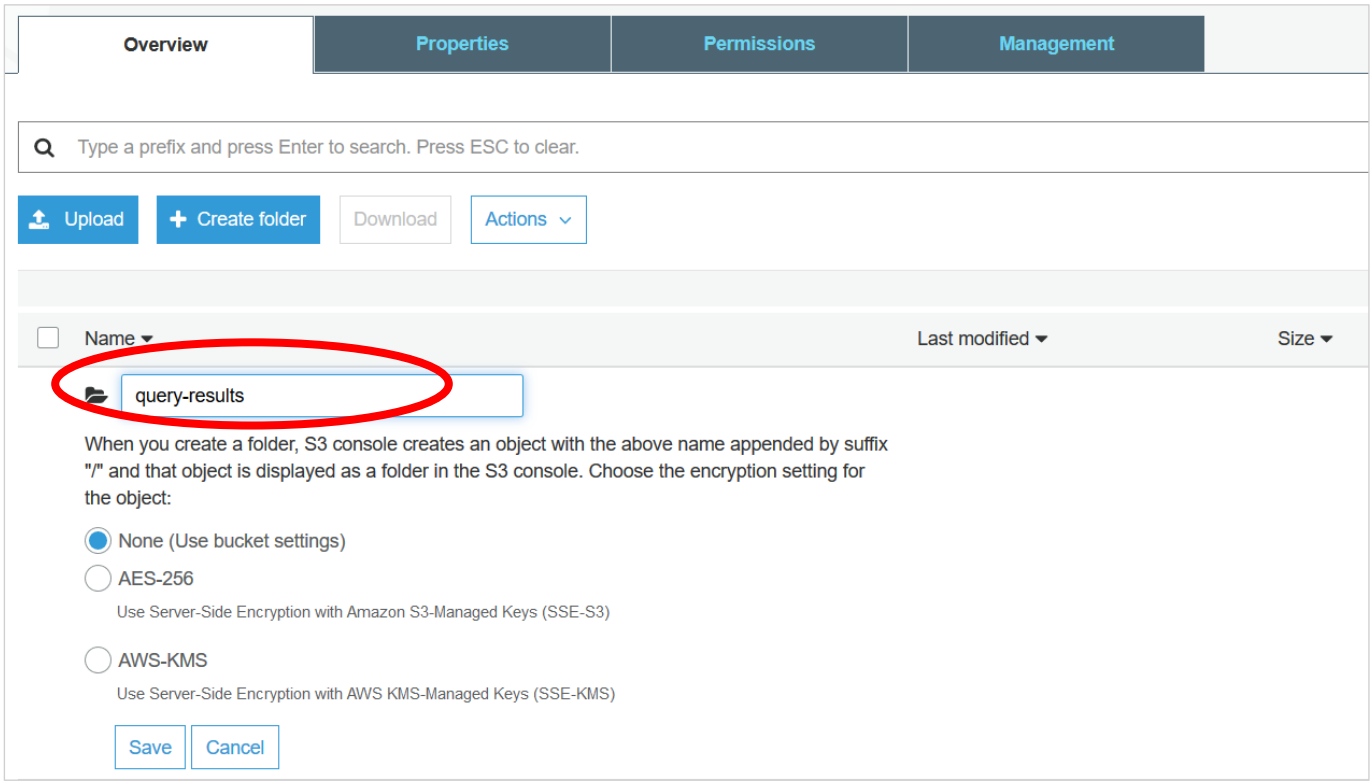


5) Click “Create bucket”.

6) Now you will see your bucket on the Amazon S3 dashboard, under Buckets.

7) Click on the bucket name you just created. It’s time to create a few folders!

8) Click on “Create folder,” then type “query-results” next to the folder icon. Accept the default encryption settings (“None”), and click “Save”.

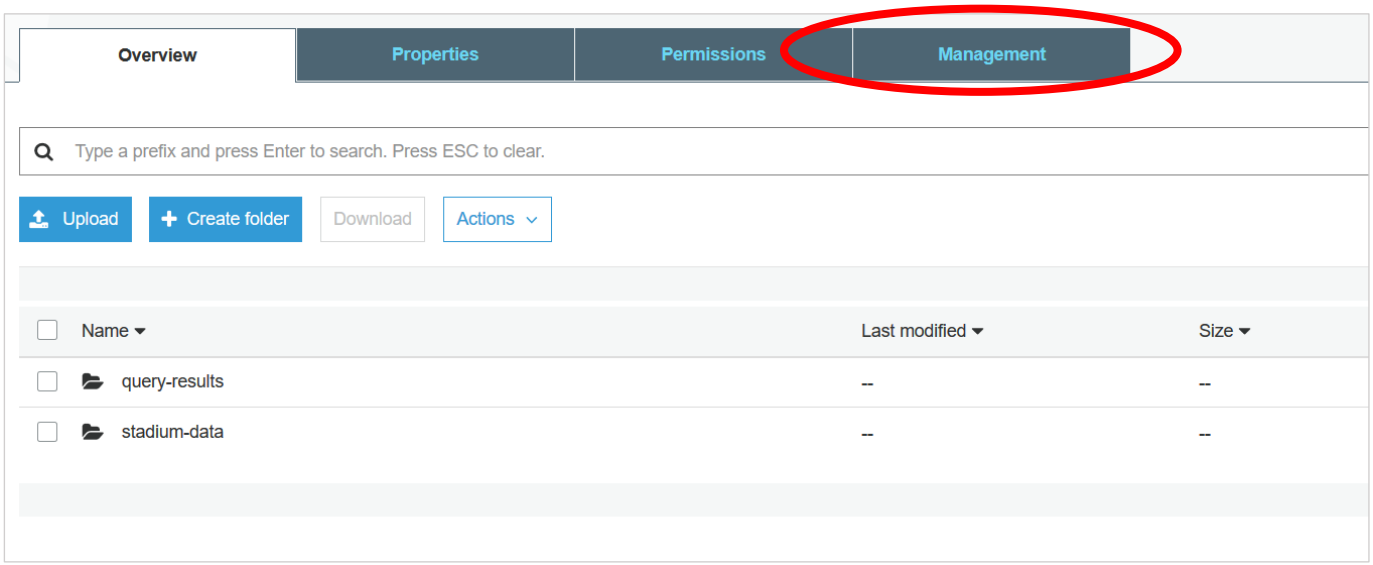


9) Repeat this process to create a folder named “stadium-data”.

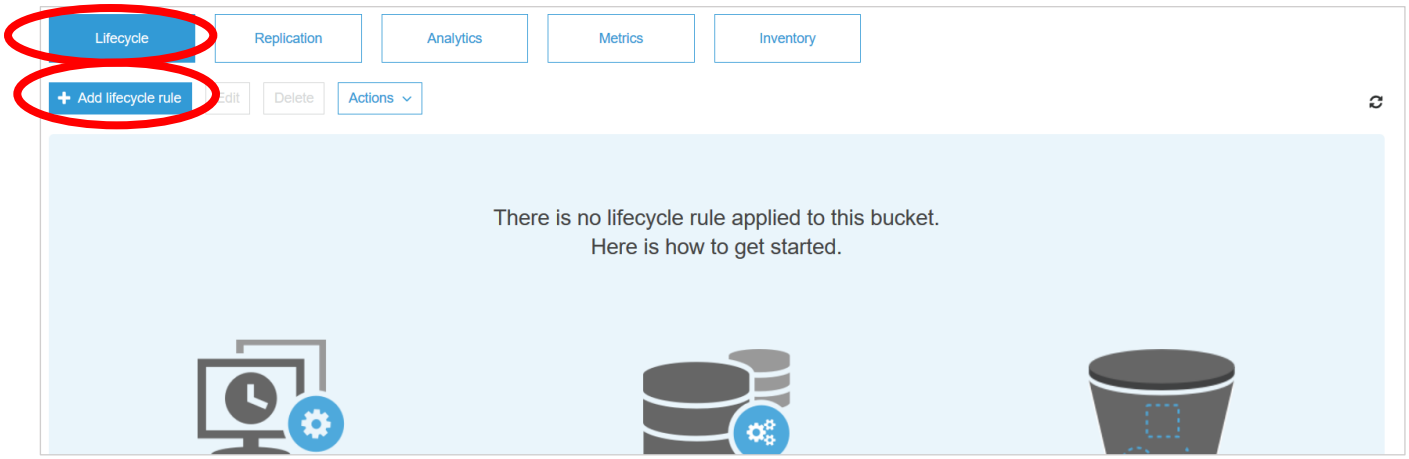
10) Using the AWS Console, upload the stadium_with_stations_global.csv file to the stadium-data folder: click on the stadium-data folder, then click on the “Upload” button. Navigate to where you saved the file, select it to be uploaded, and click “Next”. In the “Set permissions” section, accept the defaults and click “Next”, and likewise in “Set properties leave the default, “Standard”, and click “Next”. Finally, click “Upload”. You should see the object you just uploaded in the folder when it is complete.

11) Click on “Amazon S3” link above the Overview tab to go back to the S3 console. Then click on your bucket you created in this example.

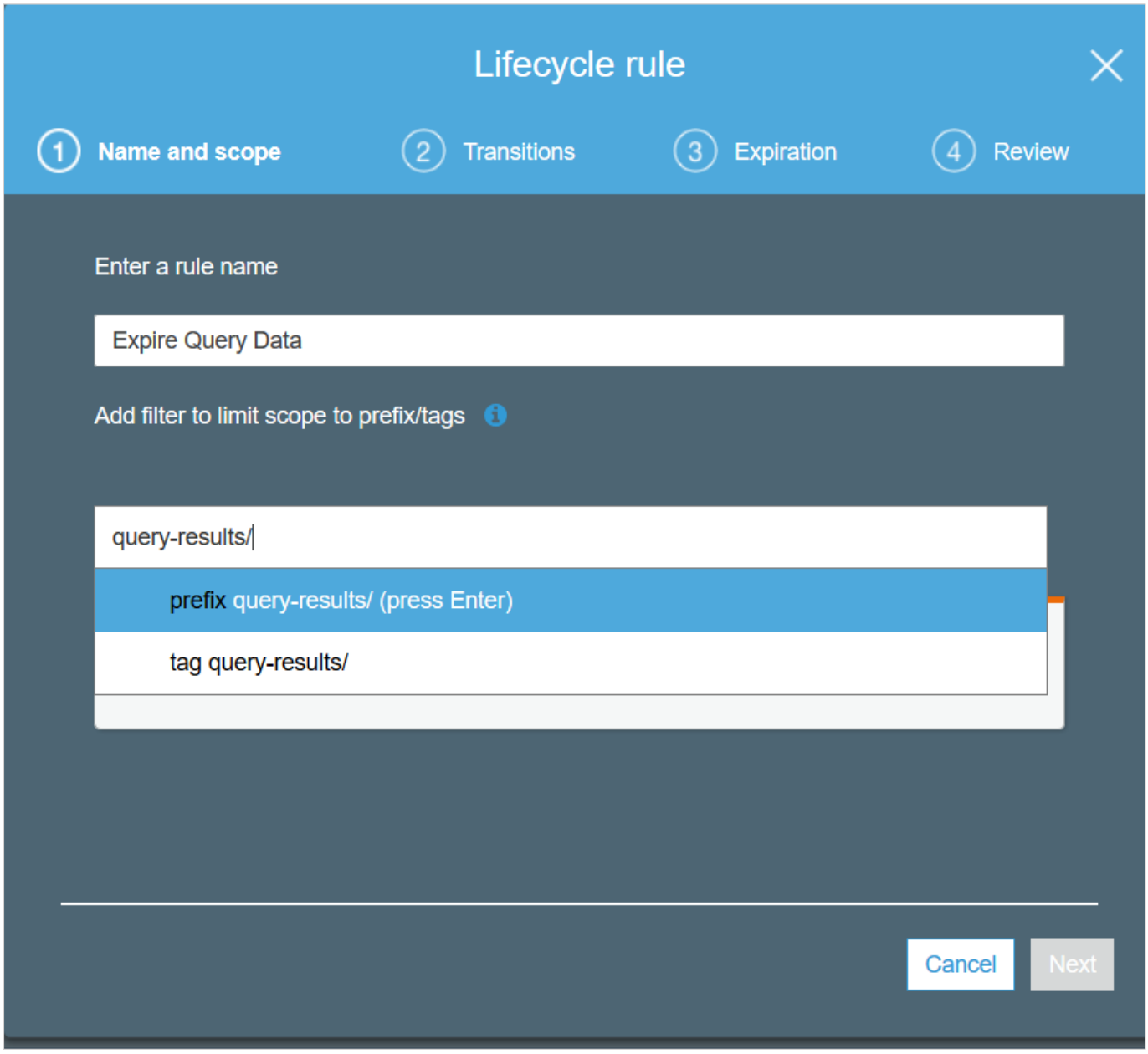
12) Since the query-results folder will be used to temporarily store results from our Athena queries, you will want to create a lifecycle policy to delete these files when they are no longer necessary. Start by clicking on the “Management” tab.



13) With the Lifecycle tab highlighted, click the “Add lifecycle rule” button.



14) In the rule name field type “Expire Athena query results”. In the prefix/tags field type the folder you created for the Athena query results followed by a slash (“query-results/”), select prefix, then click “Next”.



15) On the “Transitions” pane leave the two boxes for “Current version” and “Previous version” unchecked (we are not moving files to other S3 storage classes), and click “Next”.

16) On the “Expiration” pane select the “Current version” checkbox, change the expiration value to 1 day, and click “Next”.

17) On the “Review” pane click “Save”.

You've now completed section 1 of the workshop and can move on to the next session, "Connecting Athena to the NOAA data repository."

In this session, you created the S3 repository that Athena will use to store query results, specified a folder to hold the mappings file of the cities to sensor locations, and created a lifecycle policy to delete items in your query-result/ folder.