1

An Introduction to AWS (Mark Kerzner starting to verify that the format is good)

Overview

This chapter will introduce you to the Amazon Web Service (AWS) interface and will teach you how to store and retrieve data with Amazon Simple Storage (S3).

Next, you will apply your S3 knowledge by importing and exporting text data via the management console and the Command-Line Interface (CLI).

Lastly, you will learn how to locate and test Artificial Intelligence (AI) and Machine Learning (ML) services.

Introduction

What is AWS?

AWS is a secure cloud platform that delivers on-demand computing power, database storage, applications, and other IT resources, through a cloud services platform via the internet with a pay-as-you-go pricing model. As of 2018, AWS dominates the cloud infrastructure services worldwide market with ~30% of the market share, compared to the trailing Microsoft (~15%) and Google (~5%), as per Canalys. (https://www.canalys.com/static/press\_release/2018/270418-cloud-infrastructure-market-grows-47-q1-2018-despite-underuse.pdf).

What is Machine Learning?

Machine Learningis a subset of Artificial Intelligence in the field of computer science that often uses statistical techniques to provide computers with the ability to learn with data without being programmed. Machine Learning explores the building and construction of algorithms that can learn from and make predictions about, data. The algorithms transcend static instructions and make data-driven predictions and decisions with a model from sample inputs.

Machine Learning is essential to learn in today's world, since it is an integral part of all industries' competitive and operational data strategies. More specifically, ML allows insights from Natural Language Processing (NLP) to power chatbots; fraud detection ML insights are used in the financial industry, and ML applications allow for efficient online recommendation engines, such as friend suggestions on Facebook, Netflix displaying movies, you probably like, and more items to consider on Amazon.

What is Artificial Intelligence?

Artificial Intelligence is intelligence that's demonstrated by machines; more specifically, any device that perceives its environment and takes actions that increases its chance of successfully achieving its goals. Contemporary examples are understanding human speech, competing at the highest levels of strategic games (such as Chess and Go), and autonomous cars.

Artificial Intelligence is important because it adds intelligence to existing products. Products that are currently used will be further improved with AI capabilities, for example, Siri was added to the new generation of Apple products. Conversational chatbots can be combined with large amounts of data to improve technologies at home and in the office.

Overall, this chapter will provide you with the foundational knowledge of AWS to build sophisticated AI and ML applications in your projects. This will help you with the tools to identify free-tier AWS services. Thus, you will be able to use them exclusively, or integrate them to analyze data, build a conversational chatbot, store and process an incredible amount of data, and bring your ideas to fruition.

This chapter will introduce you to the AWS interface and help you learn how to store and retrieve data with Amazon Simple Storage Service (S3). You will apply your S3 knowledge by importing and exporting text data via the AWS Management Console and the CLI. You will also learn how to locate and test AI and ML services.

What is Amazon S3?

S3 is an online cloud object storage and retrieval service. Amazon S3 is a cloud object storage. Instead of data being associated with a server, S3 storage is server independent and can be accessed over the internet Data stored in S3 is managed as objects using an Application Programming Interface (API) that is accessible via the internet (HTTPS).

The benefits of using S3 are as follows:

* Amazon S3 runs on the largest global cloud infrastructure, to deliver 99.99% durability.
* It provides the widest range of options to transfer data.
* It allows you to run Big Data analytics without moving data into a separate analytics system.
* It supports security standards and compliance certificates.
* It offers a flexible set of storage management and administration capabilities.

Note

For more information refer: https://aws.amazon.com/s3/.

Why use S3?

The S3 is a place to store and retrieve your files. It is recommended for storing static content such as text files, images, audio files, video files, and so on. For example, S3 can be used as a static web server if the website consists exclusively of HTML and images. The website can be connected to an FTP client to serve the static files. In addition, S3 can also be used to store user generated image and text files.

However, the two most important applications of it are as follows:

* To store static data from web pages or mobile apps
* To implement Big Data analytics

It can easily be used in conjunction with additional AWS Machine Learning and infrastructure services. For example, text documents imported to Amazon S3 can be summarized by code running in an AWS Lambda function that is analyzed using AWS Comprehend. We will cover both in Chapter 2, Summarizing Text Document using NLP and Chapter 3, Perform Topic Modeling and Theme Extraction.

The Basics of Working on AWS with S3

The first step to accessing S3 is to create an AWS Free-tier account, which provides access to the AWS Management Console. The AWS Management Console is a web application that provides one method to access all of AWS's powerful storage and ML/AI services.

The second step is to understand the access level. AWS defines Identity and Access Management (IAM). The same email/password is used for accessing the IAM

AWS Free-Tier Account

AWS provides a free-tier (within their individual free usage stipulations) account, and one of the included storage services is Amazon Simple Storage (S3). Thus, you can maximize cost savings and reduce errors before making a large investment by testing services to optimize your ML and AI workflows.

Importing and Exporting Data into S3

AWS Import and Export is a service that you can use to transfer large amounts of data from physical storage devices into AWS. You mail your portable storage devices to AWS, and AWS Import/Export transfers data directly from your storage devices using Amazon's high-speed internal network. Your data load typically begins the next business day after your storage device arrives at AWS. After the data export or import completes, services return your storage device. For large datasets, AWS data transfer can be significantly faster than internet transfer and more cost-effective than upgrading your connectivity.

How S3 Differs from a Filesystem

S3 is used to store almost any type of file, thus, it can get confused with similarities to a traditional filesystem. However, S3 differs in a few ways from a traditional filesystem. Overall, the folders in a traditional file system are Buckets in S3; a file in a traditional filesystem is an object in S3. S3 uses objects, since you can store any data type (that is, more than files) in Buckets.

Another difference is how objects can be accessed. Objects stored in Buckets can be accessed from a web service endpoint (such as a web browser, for example, Chrome, Firefox, and so on), so each object requires a globally unique name. The name restrictions for objects are similar to the restrictions in selecting a URL when creating a new website. Obviously, you need to select a unique URL, according to the same logic that your house has a unique address.

For example, if you created a Bucket (with public permission settings) named myBucket and then uploaded a text file named pos\_sentiment\_\_leaves\_of\_grass.txt to the Bucket, the object would be accessible from a web browser via the corresponding subdomain.

Core S3 Concepts

The S3 hierarchy includes the following concepts:

Type of data storable: S3 is recommended for storing static content, such as text files, images, audio, video, and so on.

Objects: Objects are the most basic entities that are stored in S3. Every object contains data, metadata, and a key. Metadata is data about the data and provides basic information about the data stored in an object. Metadata is stored in a set of name-value pairs, used to describe the information associated with the object.

Keys: A key is the name assigned to an object that uniquely identifies it inside a bucket. All objects in a bucket have one key associated with them.

Bucket: Just like a folder, a Bucket is the container where you store objects. Buckets are created at root level, and do not have a filesystem hierarchy. More specifically, you can have multiple Buckets, but you cannot have sub-Buckets within a Bucket. Buckets are the containers for objects, and you can control (create, delete, and list objects in the Bucket) access to it, view access logs for it, and select the geographical region where Amazon S3 will store the Bucket.

Region: Region refers to the geographical region where Amazon S3 stores a Bucket, based on the user's preference. The region can be selected when creating a Bucket. The location should be based on where the data will be accessed the most. Overall, specific region selection has the biggest impact if S3 is used to store files for a website that's exclusively accessed in a specific geographic region. The object storage in a Bucket with different forms is as follows:

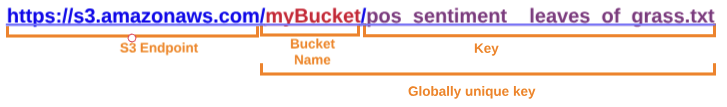


Figure 1.1: Object storage

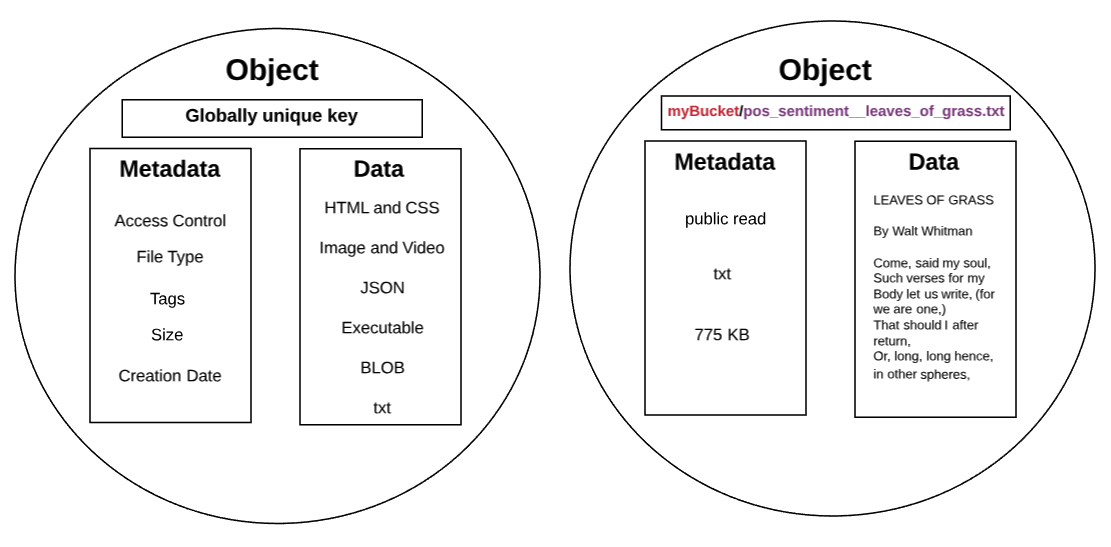


Figure 1.2: Object storage using a unique key and myBucket

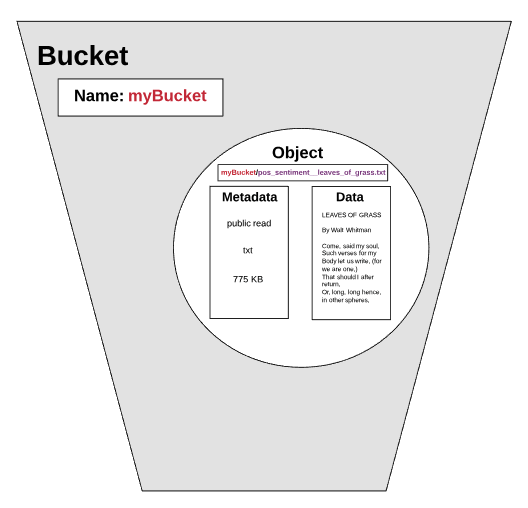


Figure 1.3: Object stored in myBucket

S3 Operations

The S3 Application Program Interface (API) is quite simple, and it includes the following operations for the respective entity:

* Bucket: Create, Delete, and List keys in a Bucket
* Object: Write, Read, and Delete

Data Replication

Amazon replicates data across the region in multiple servers located in Amazon's data centers. Data replication benefits include high availability and durability. More specifically, when you create a new object in S3, the data is saved in S3; however, the change needs to be replicated across the Amazon S3 regions. Overall, replication may take some time, and you might notice delays resulting from various replication mechanisms. Please consider the following when performing the indicated operations.

After deleting an object, replication can cause a lag time that allows the deleted data to display until the deletion is fully replicated. Creating an object and immediately trying to display it in the object list might be delayed as a result of a replication delay.

REST Interface

S3's native interface is a Representational State Transfer (REST) API. It is recommended to always use HTTPS requests to perform any S3 operations. The two higher-level interfaces that we will use to interact with S3 are the AWS Management Console and the AWS Command-Line Interface (CLI). Accessing objects with the API is quite simple, and includes the following operations for the respective entity:

* Bucket: Create, Delete, or List keys in a Bucket
* Object: Write, Read, or Delete

Exercise 1.01: Using the AWS Management Console to Create an S3 Bucket

In this exercise, we will import a text file into our S3 Bucket.

To import a file, you need to have access to the Amazon S3 console:

1. Press the Ctrl key while clicking https://console.aws.amazon.com/console/home to open the AWS Management Console in a new browser tab.
2. Click inside the search bar located under AWS services as shown:

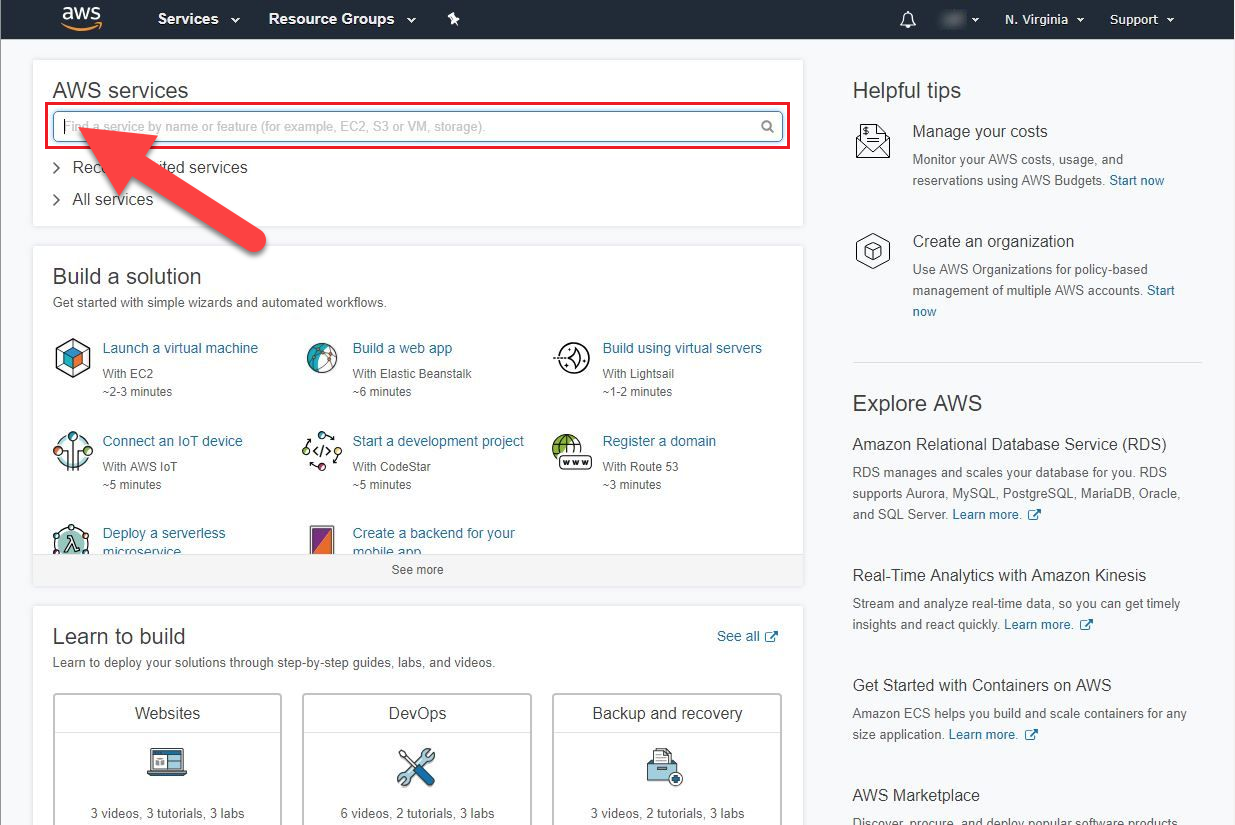


Figure 1.4: Searching AWS services

1. Type S3 into the search bar, and an auto-populated list will display. Then, click on the S3 Scalable Storage in the Cloud drop-down option:

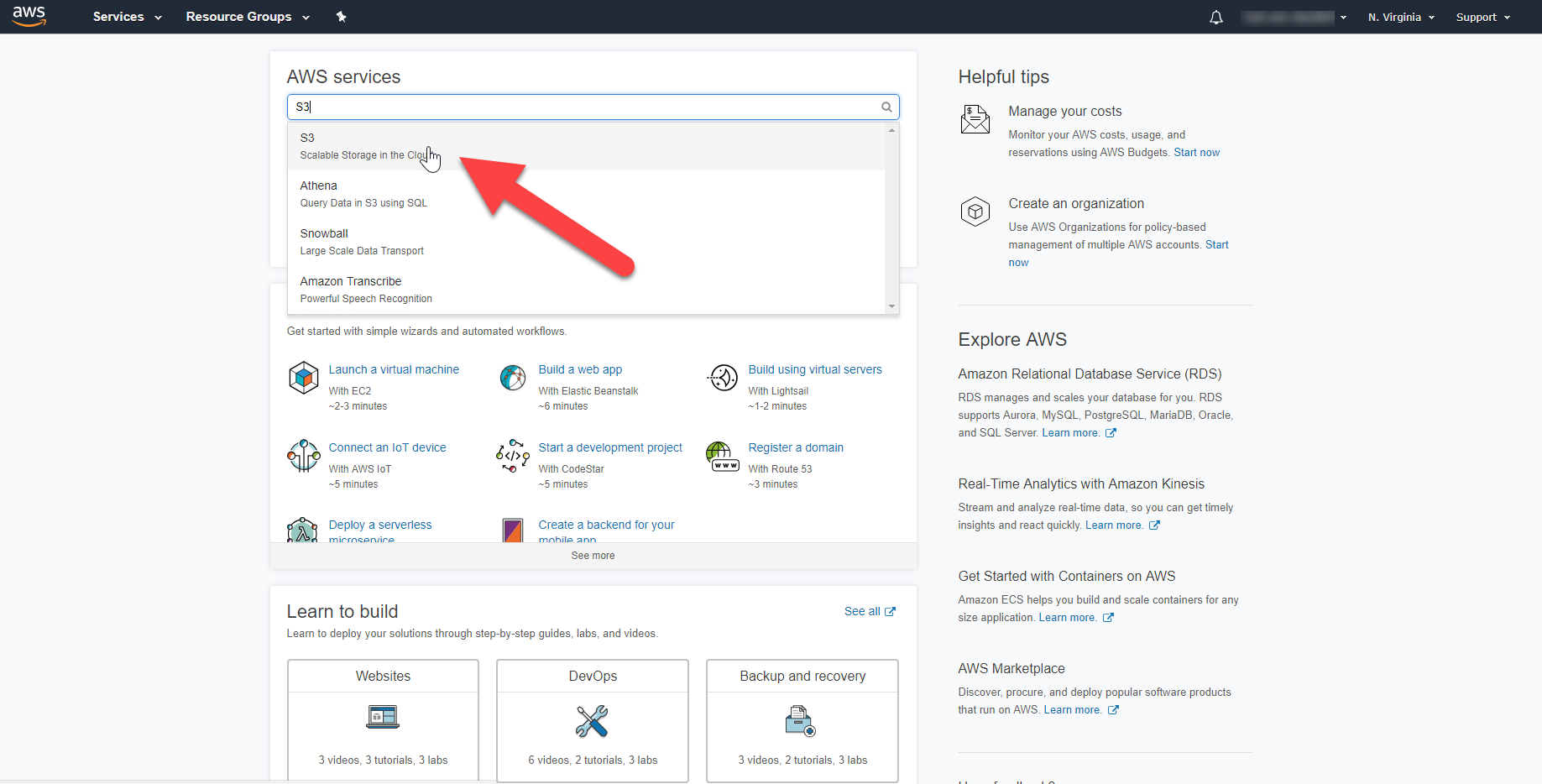


Figure 1.5: Selecting the S3 service

1. Now, we need to create an S3 Bucket. In the S3 dashboard, click the   
   Create Bucket button.
2. If this is the first time that you are that you are creating a bucket, your screen will look as follows:

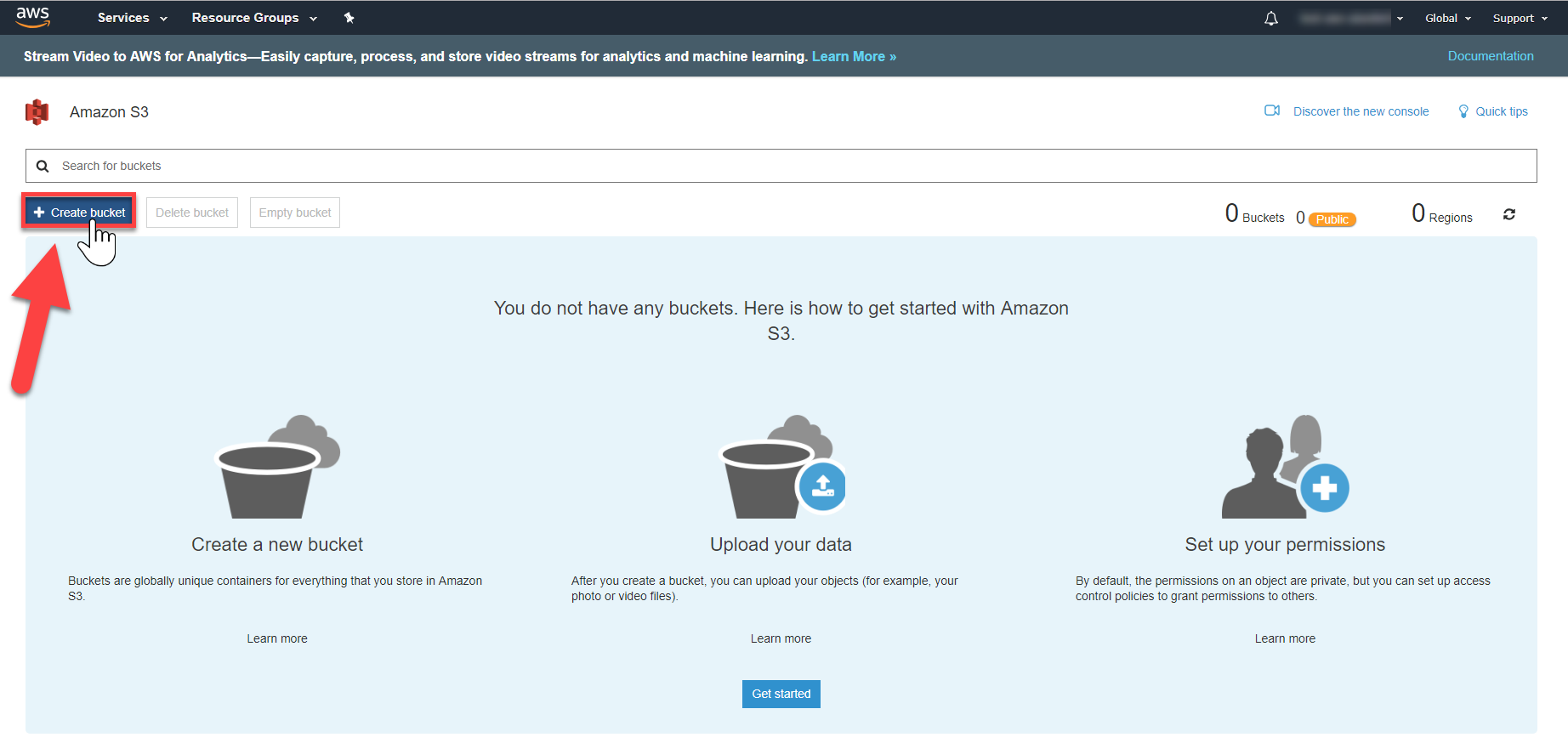


Figure 1.6: Creating the Bucket

If you have already created S3 Buckets, your dashboard will list all of the Buckets you have created.

1. Enter a unique Bucket name: Bucket names must be unique across all existing Bucket names in Amazon S3. That If you encounter a naming issue, please refer to https://docs.aws.amazon.com/AmazonS3/latest/dev/BucketRestrictions.html.
2. Region: If a default region is auto populated, then keep the default location. If it is not auto populated, select a region near your current location.
3. Click the Next button to continue for the creation of bucket:

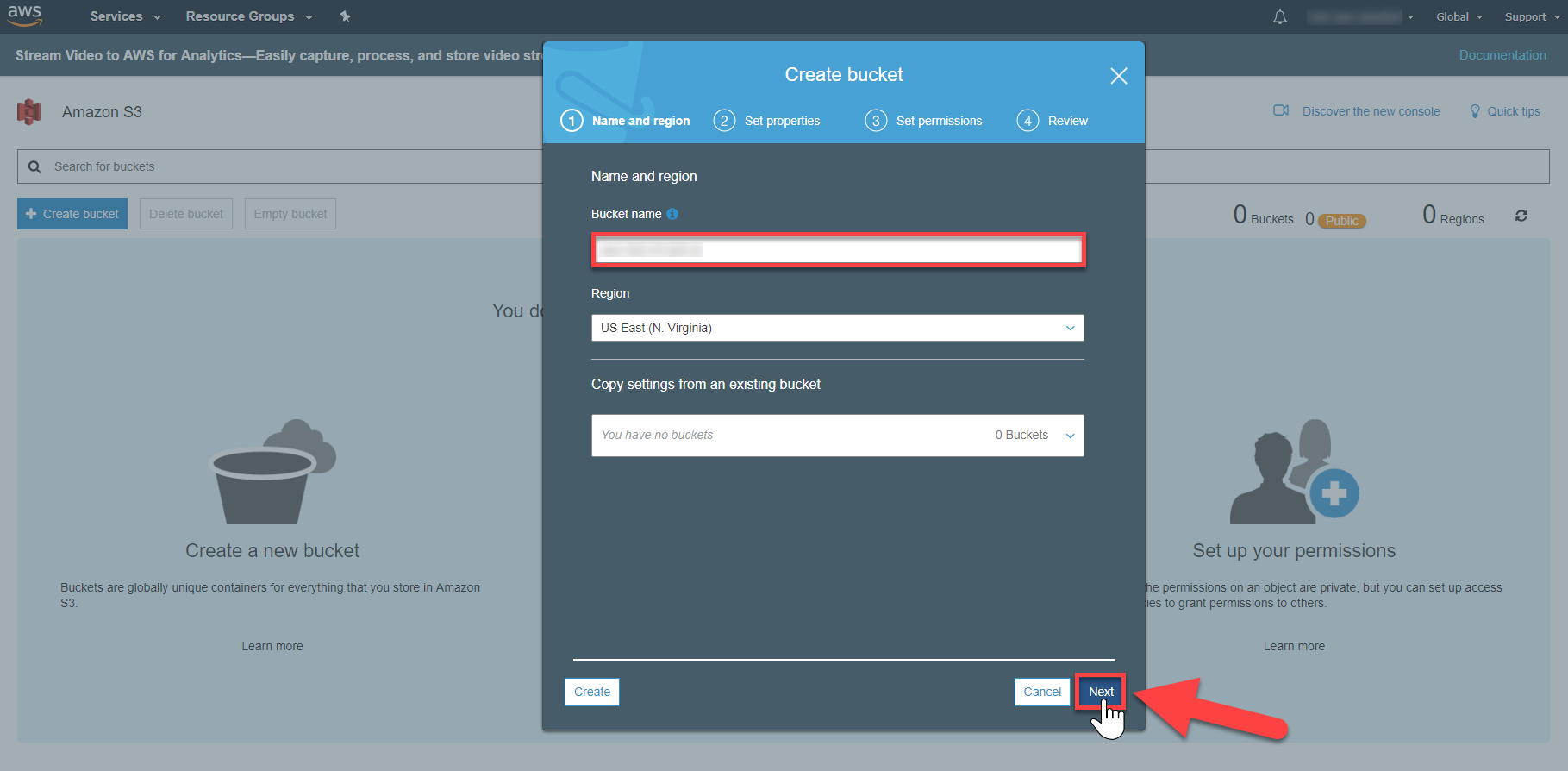


Figure 1.7: The Create Bucket window

1. An S3 Bucket provides the property options Versioning, Server Access Logging, Tags, Object-Level Logging, and Default Encryption. However, we will not enable them.
2. Your Bucket will be displayed in the bucket list, like so:

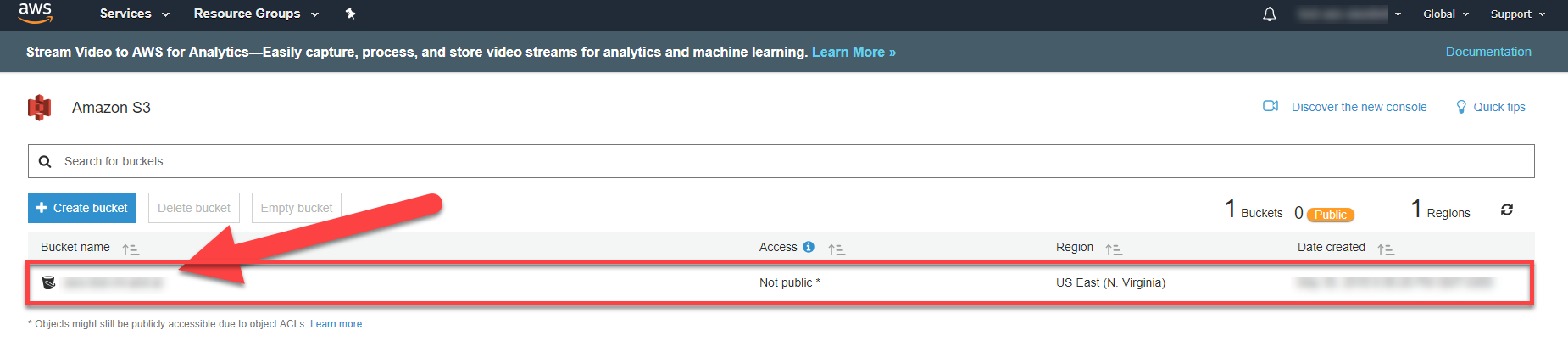


Figure 1.8: The Bucket has been created

Exercise 1.02: Importing and Exporting the File with your S3 Bucket

In this exercise, we will import and export the file with the S3 Bucket. The following are the steps for completion:

Importing a File:

1. You will import a file to your Amazon S3 Bucket.
2. Click the Bucket's name to navigate to the Bucket:

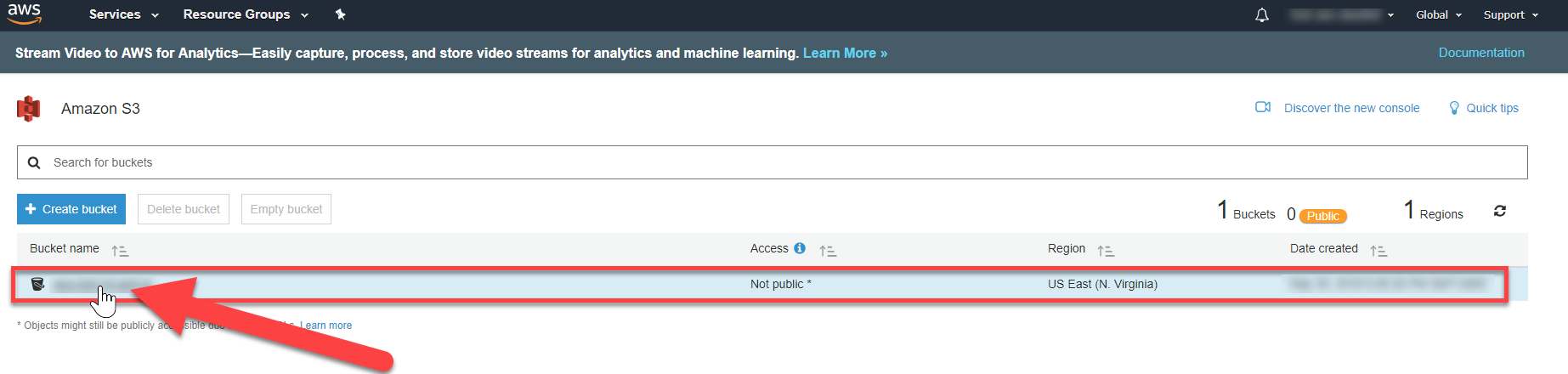


Figure 1.9: Navigate to Bucket

1. You are in the Bucket's home page. Select Upload:

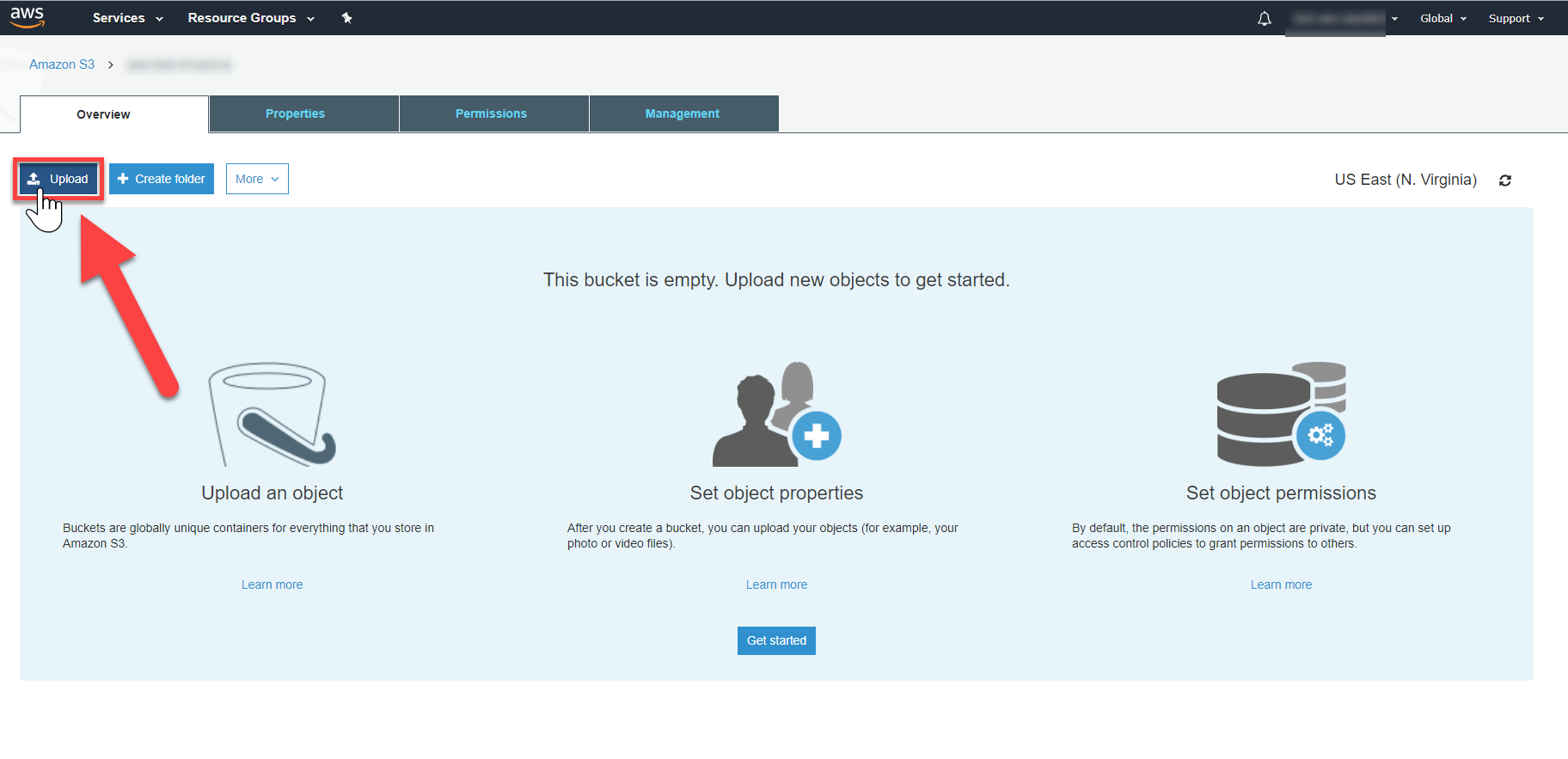


Figure 1.10: Uploading the file into the Bucket

1. To select a file to upload, click Add files. Navigate to the pos\_sentiment\_\_leaves\_of\_grass.txt location and select the sample file that you want to store:

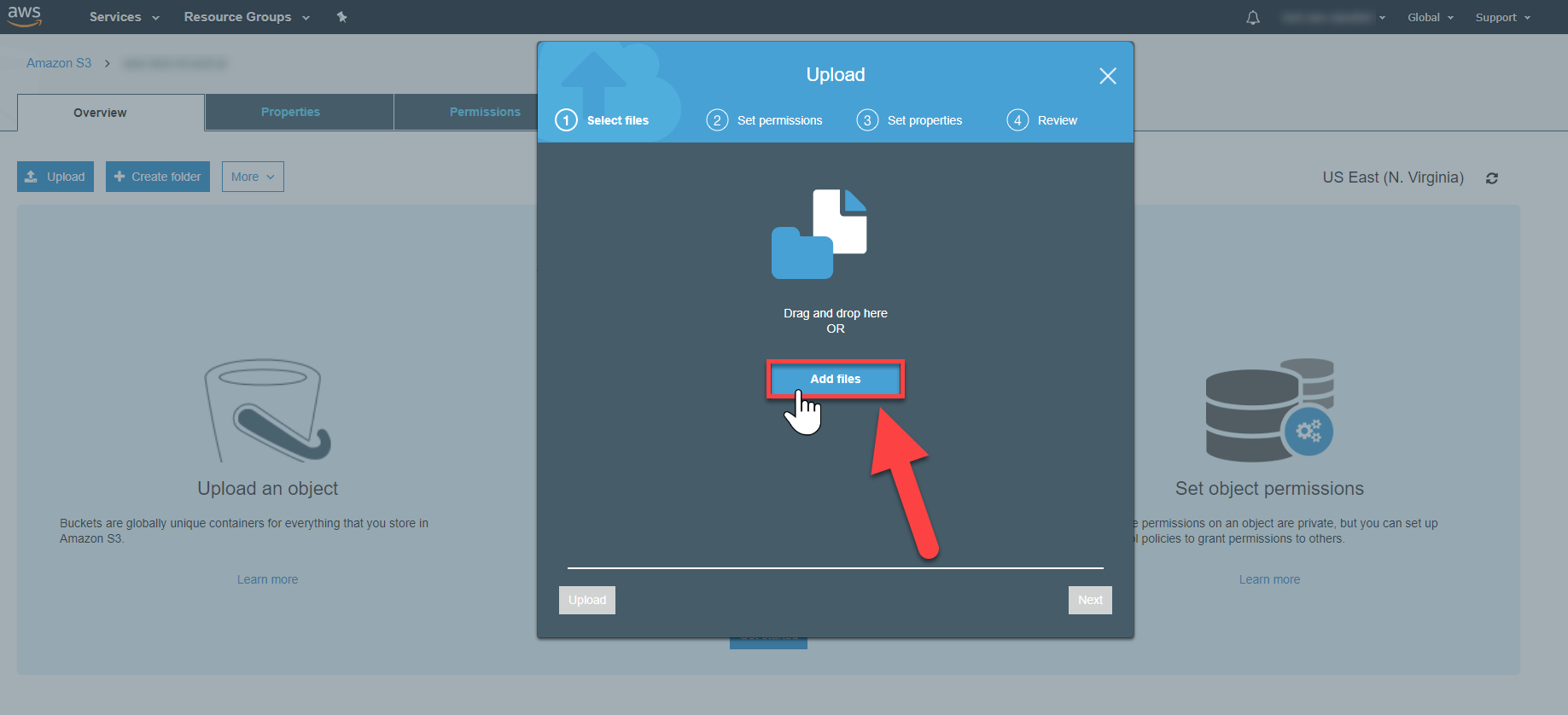


Figure 1.11: Adding a new file to the Bucket

1. After selecting a file to upload, select Next:

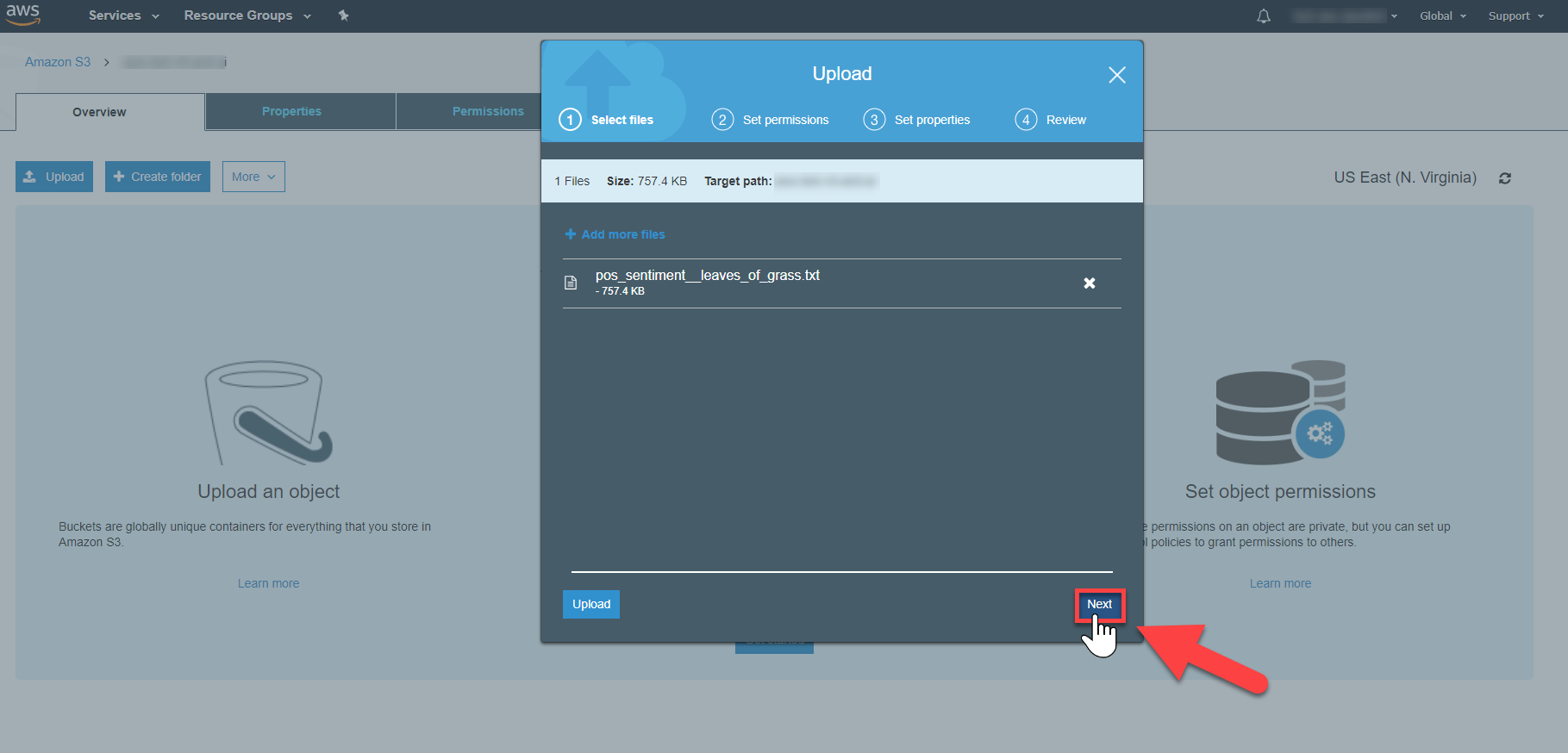


Figure 1.12: Select the file to upload to the Bucket

1. Click on the Next button and leave the default options selected:

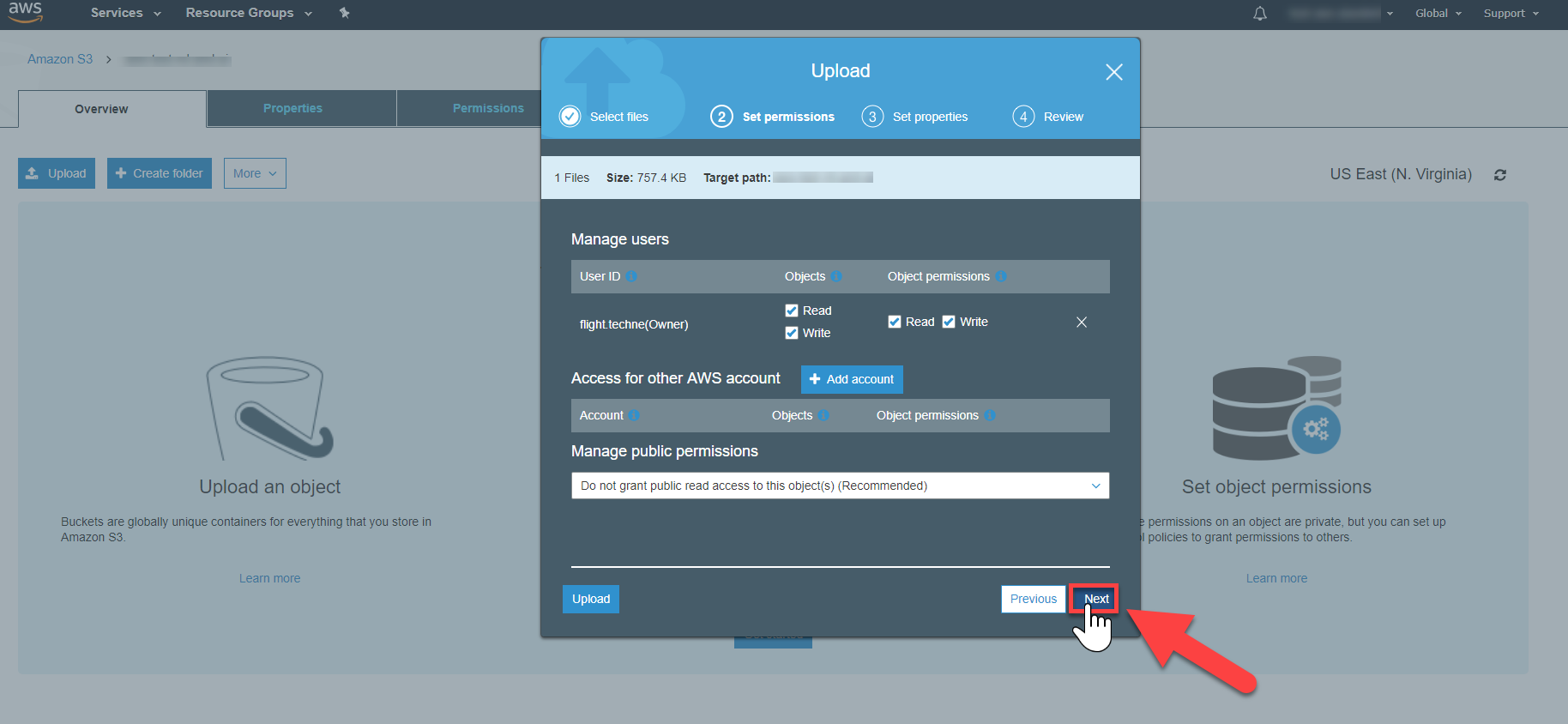


Figure 1.13: Default Options page while uploading the file

1. You have the ability to set property settings for your object, such as storage class, encryption, and metadata. However, leave the default values as-is, and then click on the Next button:

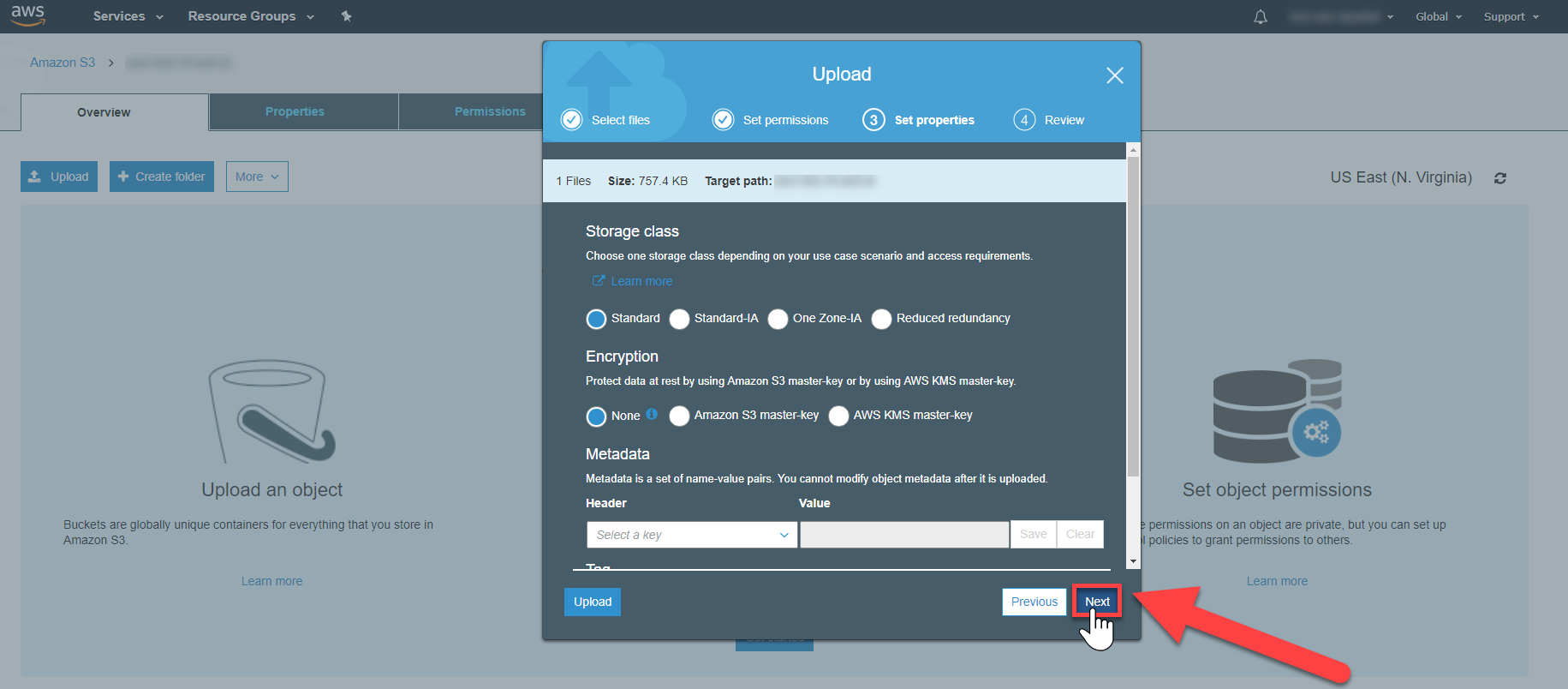


Figure 1.14: Setting properties

1. Click on the Upload button to upload the files:

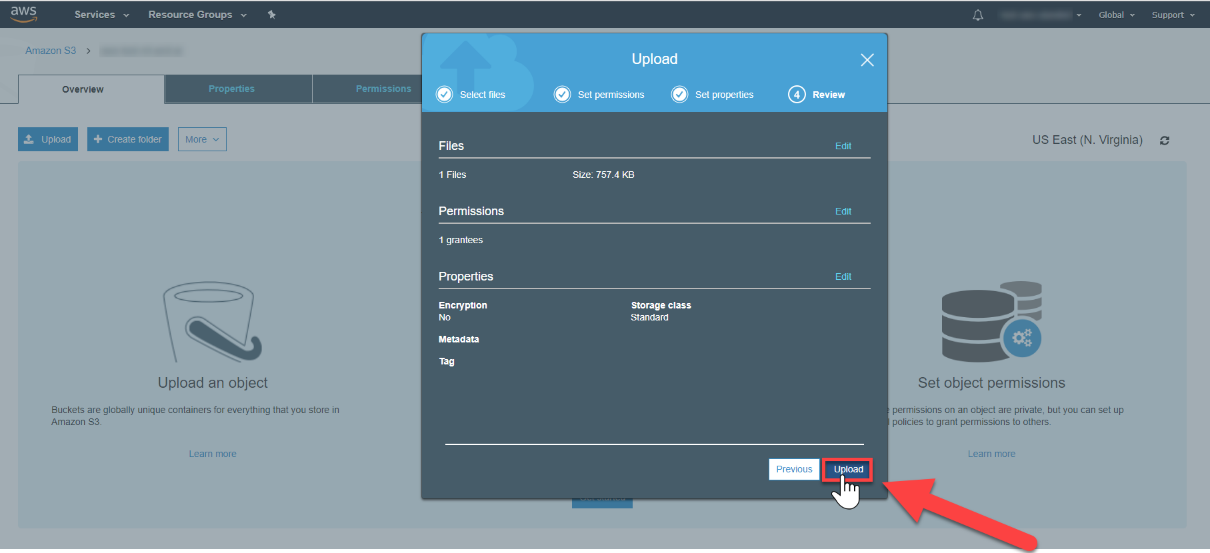


Figure 1.15: Uploading the files

1. You will be directed to your object in your bucket's home screen:

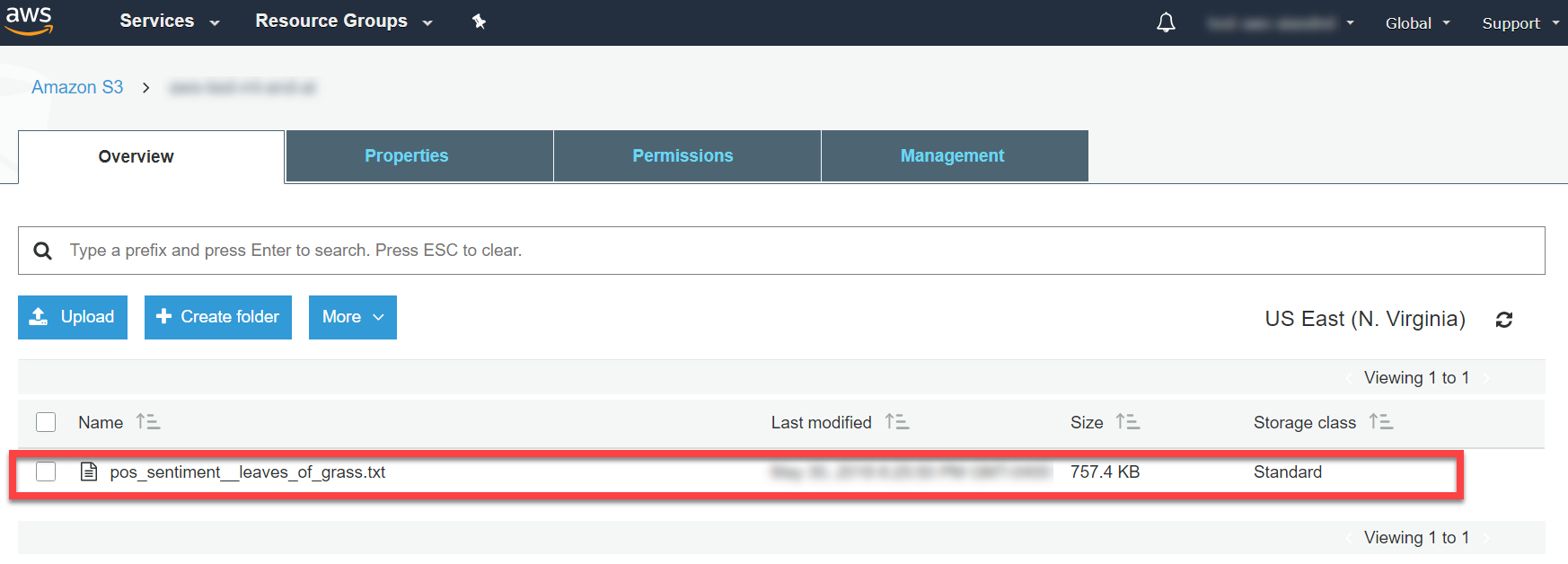


Figure 1.16: Files uploaded to the Bucket

Export File:

1. Select the checkbox next to the file to export (Red Marker #1 – see the following screenshot). This populates the file's information display screen. Click on Download (Red Marker #2 – see the following screenshot) to retrieve the text file:

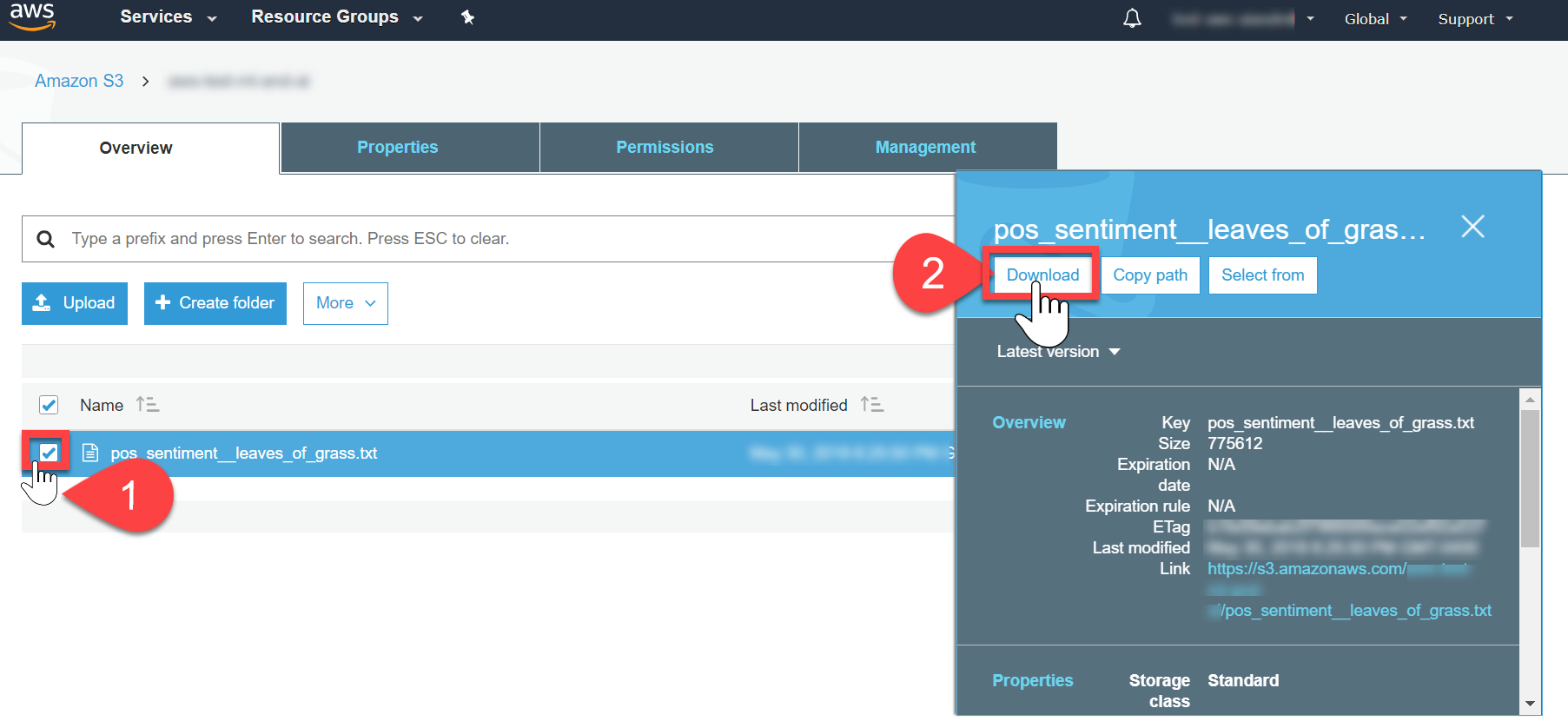


Figure 1.17: Exporting the file

1. The file will download, as shown in the lower left-hand corner of the screen:

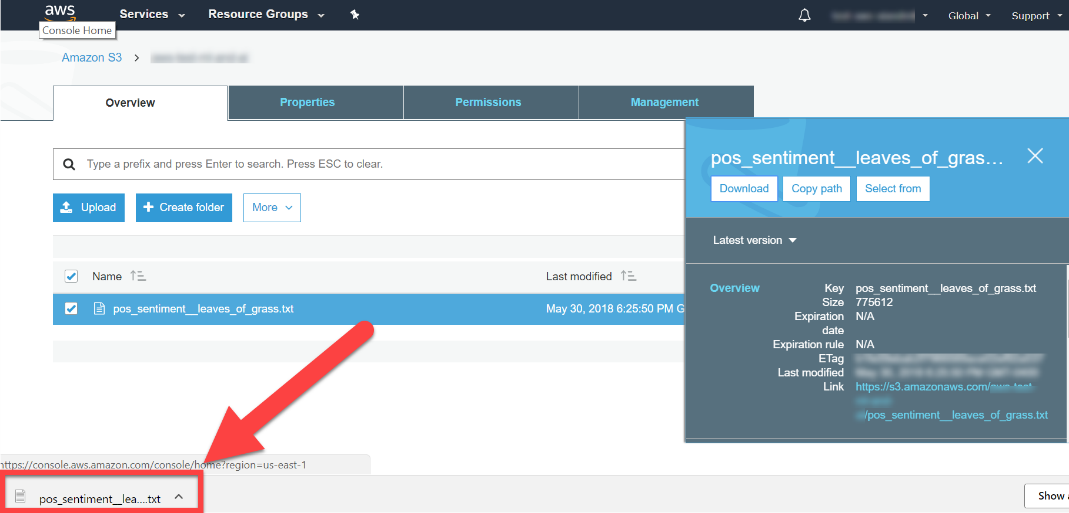


Figure 1.18: Downloading the file to export

AWS Command-Line Interface (CLI)

The CLI is an open source tool built on the AWS SDK for Python (Boto) to perform setups, determine if calls work as intended, verify status information, and so on. The CLI provides another access tool for all AWS services, including S3. Unlike the Management Console, the CLI can be automated via scripts.

To authenticate your AWS account to the CLI, you must create a configuration file to obtain your public key and secret key. Next, you will install, and then configure, the AWS CLI.

Exercise 1.03: Configuring the Command-Line Interface

In this exercise, we will configure the CLI with our respective AWS Access Key ID and AWS Secret Access Key. The following are the steps for completion:

1. Go to.: https://console.aws.amazon.com/console/home and then, click on Users:

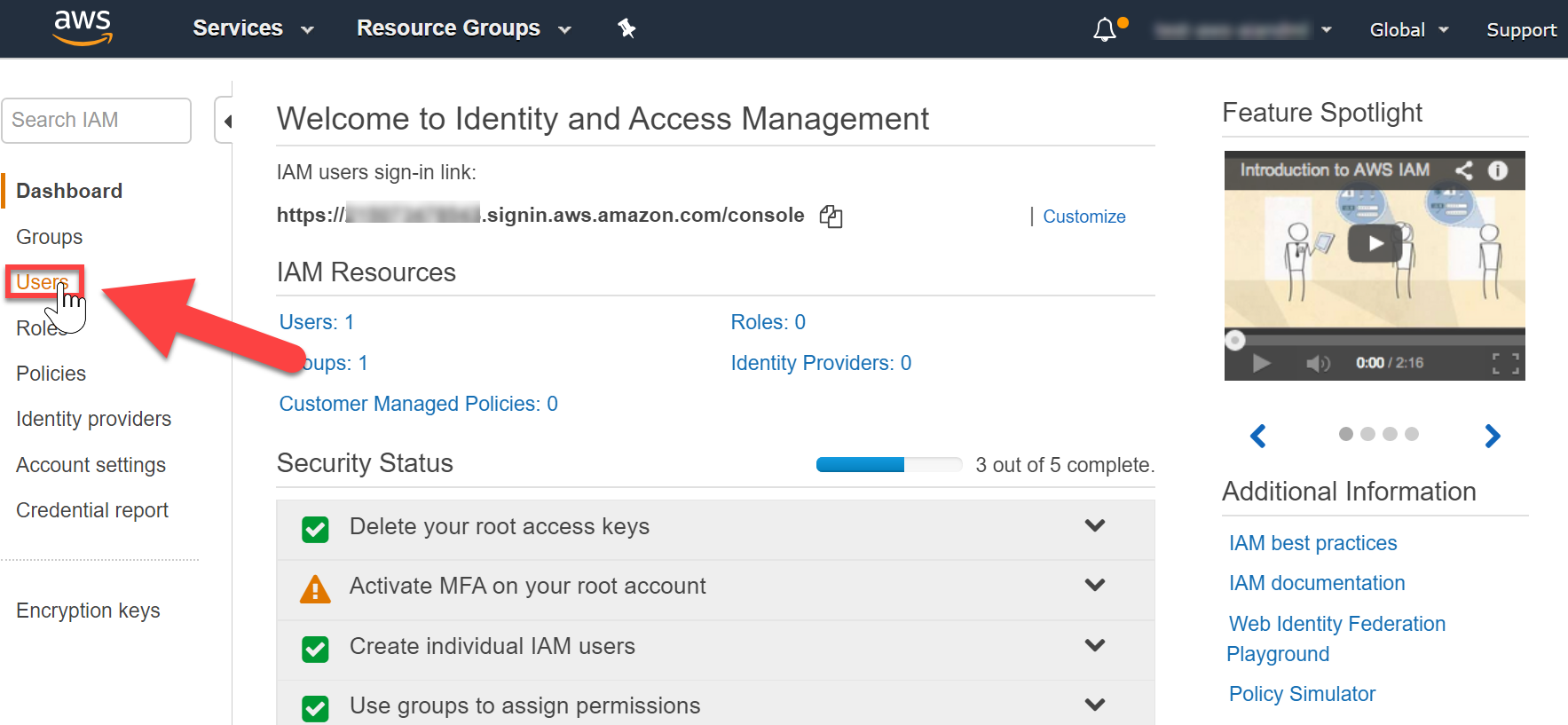


Figure 1.19: The Amazon Console home page with the Users option highlighted

1. In the upper-right corner of the signed-in AWS Management Console, click on My Security Credentials:

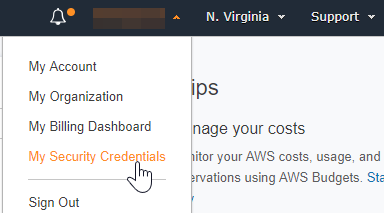


Figure 1.20: Selecting My Security Credentials

1. Next, click on Continue to Security Credentials:

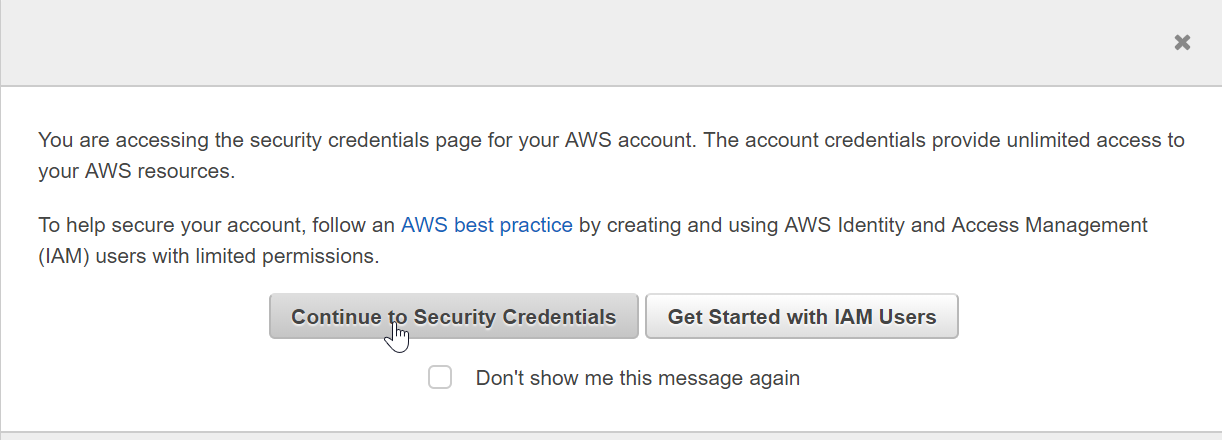


Figure 1.21: Security Credentials

1. Click on the Access keys (access key ID and secret access key) option:

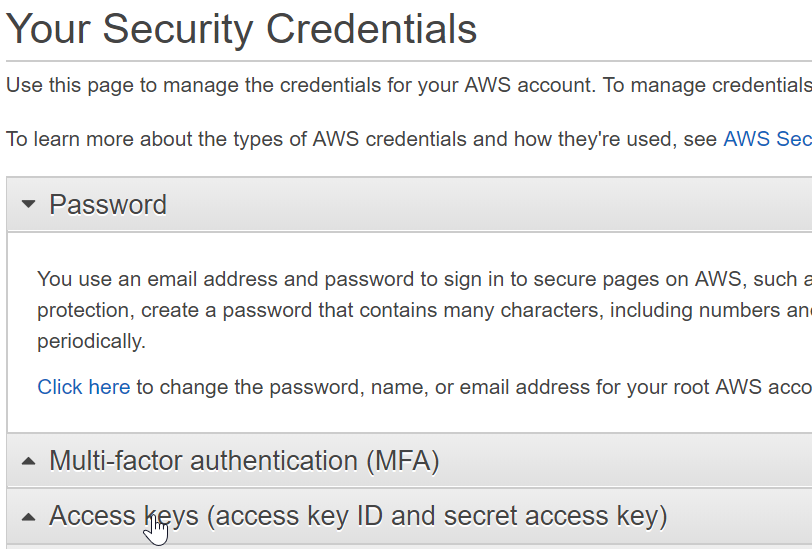


Figure 1.22: Access key generation

1. Then, click on Create New Access Key:

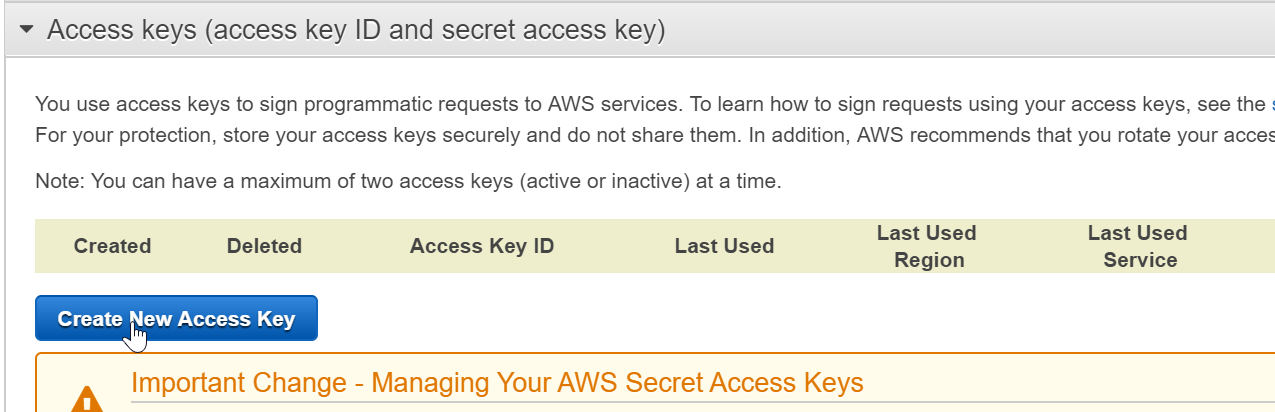


Figure 1.23: Creating a new access key

1. Click on Download Key File to download the key file:

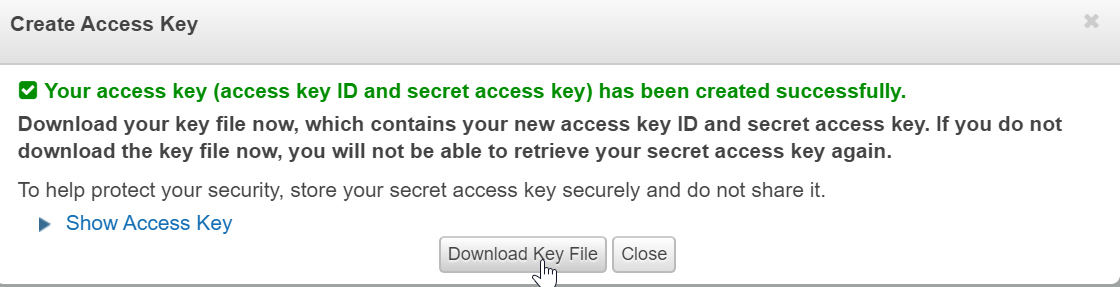


Figure 1.24: Downloading the key file

1. The rootkey.csv that contains the keys will be downloaded. Click it to view the the details:

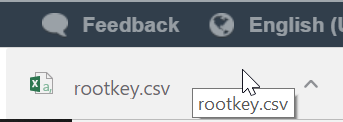


Figure 1.25: The downloaded key file

1. Store the keys in a safe location. Protect your AWS account, and never share, email, or store keys in a non-secure location. An AWS representative will never request your keys, so be vigilant when it comes to potential phishing scams.
2. Open the Command Prompt and type aws configure:
3. You will be prompted for four input variables, one by one type your respective information, then press Enter after each input:

AWS Access Key ID

AWS Secret Access Key

Default region

Default output format (json)

1. The name is obtained in your console (N. Virginia is displayed here, but yours is determined by your unique location):

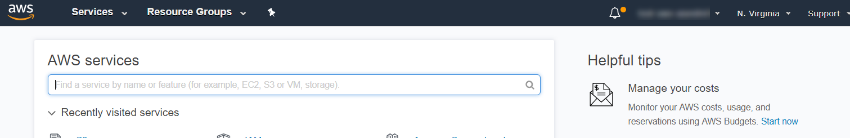


Figure 1.26: Location search

1. The code is obtained from the following Available Regions list:

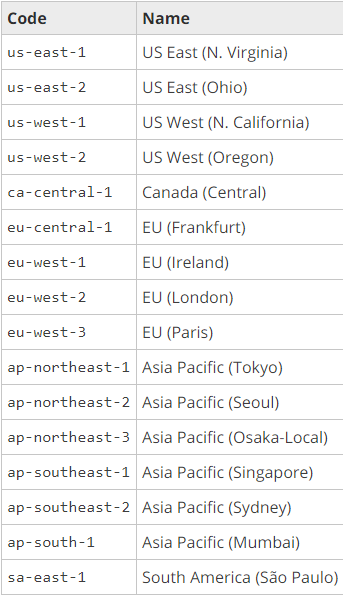


Figure 1.27: List of available regions

1. The Command Prompt 's final input variable will look as follows. Then, press Enter:

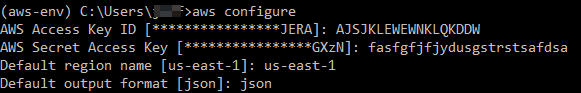


Figure 1.28: The last step in AWS CLI configuration in the Command Prompt

Command Line-Interface (CLI) Usage

When using a command, specify at least one path argument. The two path arguments are LocalPath and S3Uri:

LocalPath: This represents the path of a local file or directory, which can be written as an absolute or relative path.

S3Uri: This represents the location of an S3 object, prefix, or Bucket. The command form is s3://myBucketName/myKey. The path argument must begin with s3://, to indicate that the path argument refers to an S3 object.

The overall command structure is aws s3 <Command> [<Arg> …]. The following table shows the different commands with a description and an example:

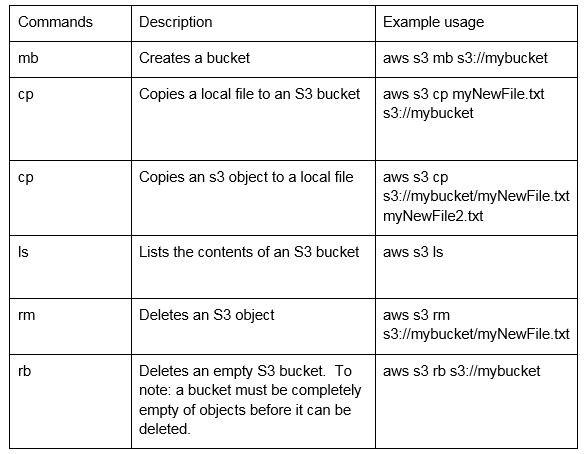


Figure 1.29: Command list

Recursion and Parameters

Importing files one at a time is time-consuming, especially if you have many files in a folder that need to be imported. A simple solution is to use a recursive procedure. A recursive procedure is one that has the ability to call itself and saves you, as the user, from entering the same import command for each file.

Performing a recursive CLI command requires passing a parameter to the API. This sounds complicated, but it is incredibly easy. First, a parameter is simply a name or option that is passed to a program to affect the operation of the receiving program. In our case, the parameter is recursive, and the entire command to perform the recursive command is as follows:

aws s3 cp s3://myBucket . --recursive

With the command, is all of the s3 objects in a respective Bucket are copied to a specified directory:

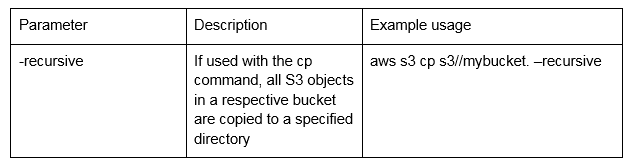


Figure 1.30: Parameter List

Activity 1.01: Importing and Exporting the Data into S3 with the CLI

In this activity, we will be using the CLI to create a Bucket in S3 and import a second text file. Suppose that you are creating a chatbot. You have identified text documents that contain content that will allow your chatbot to interact with customers more effectively. Before the text documents can be parsed, they need to be uploaded to an S3 Bucket. Once they are in S3, further analysis will be possible. To ensure that this has happened correctly, you will need to have installed Python, have an environment set up, and have a user authenticated with the CLI:

1. Configure the Command-Line Interface and verify that it is able to successfully connect to your AWS environment.
2. Create a new S3 Bucket.
3. Import a text file into the Bucket.
4. Export the file from the Bucket and verify the exported objects.

Note

To refer to the detailed steps, go to the Appendix A at the end of this book on Page no. 192

Solution:

1. Verify the configuration is correct by executing, "aws s3 ls" to output your bucket name (bucket name will be unique):



Figure 1.34: Command line

1. Execute "aws s3 ls <bucket-name>" to output the text file pos\_sentiment\_\_leaves\_of\_grass.txt" in the bucket.



Figure 1.35: Command line

1. Create a new S3 bucket with the following command:



Figure 1.36: Command line

Note

Your bucket name needs to be unique. Refer to the S3 ["Rules for Bucket Naming](https://docs.aws.amazon.com/AmazonS3/latest/dev/BucketRestrictions.html)" for specific details)

1. In the command prompt, navigate to the "neg\_sentiment\_\_dracula.txt" location Execute, "aws s3 cp neg\_sentiment\_\_dracula.txt" to import the text file to your S3 bucket:

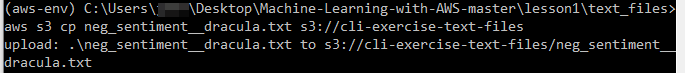


Figure 1.37: Command line

1. Navigate to the "peter\_pan.txt" file location with your command line. Import the file "peter\_pan.txt" to your S3 bucket (named "aws-test-ml-and-ai-two" in this example) with the following command:



Figure 1.38: Command line

1. Navigate to your Desktop in the command line. Create a new local folder named "s3\_exported\_files" with the command "mkdir s3\_exported\_files":



Figure 1.39: Command line

1. Next, recursively export both files ("neg\_sentiment\_\_dracula.txt" and "peter\_pan.txt") from the S3 bucket to your local directory with the "- -recursive" parameter. See below for the command's execution.



Figure 1.40: Command line

1. Verify the objects were exported successfully to your local folder with the, "dir" command (see below):

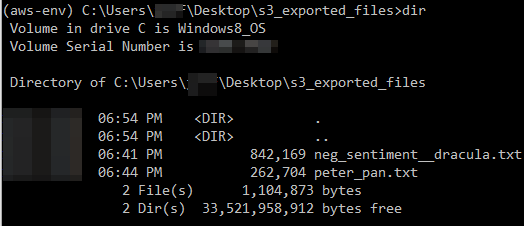


Figure 1.41: Output

Using the AWS Console to Identify Machine Learning Services

The AWS Console provides a web-based interface to navigate, discover, and utilize AWS services for AI and ML. In this topic, we will explore two ways to use the Console to search Machine Learning services. In addition, we will test an ML API with text data retrieved from a website.

Exercise 1.04: Navigating the AWS Management Console

In this exercise, we will navigate the AWS Management Console to locate Machine Learning services. Starting from the console https://console.aws.amazon.com/console/ and only using console search features, navigate to the Amazon Lex https://console.aws.amazon.com/lex/ service information page:

1. Click on https://console.aws.amazon.com/console/ to navigate to the AWS Console. Then, click on Services:

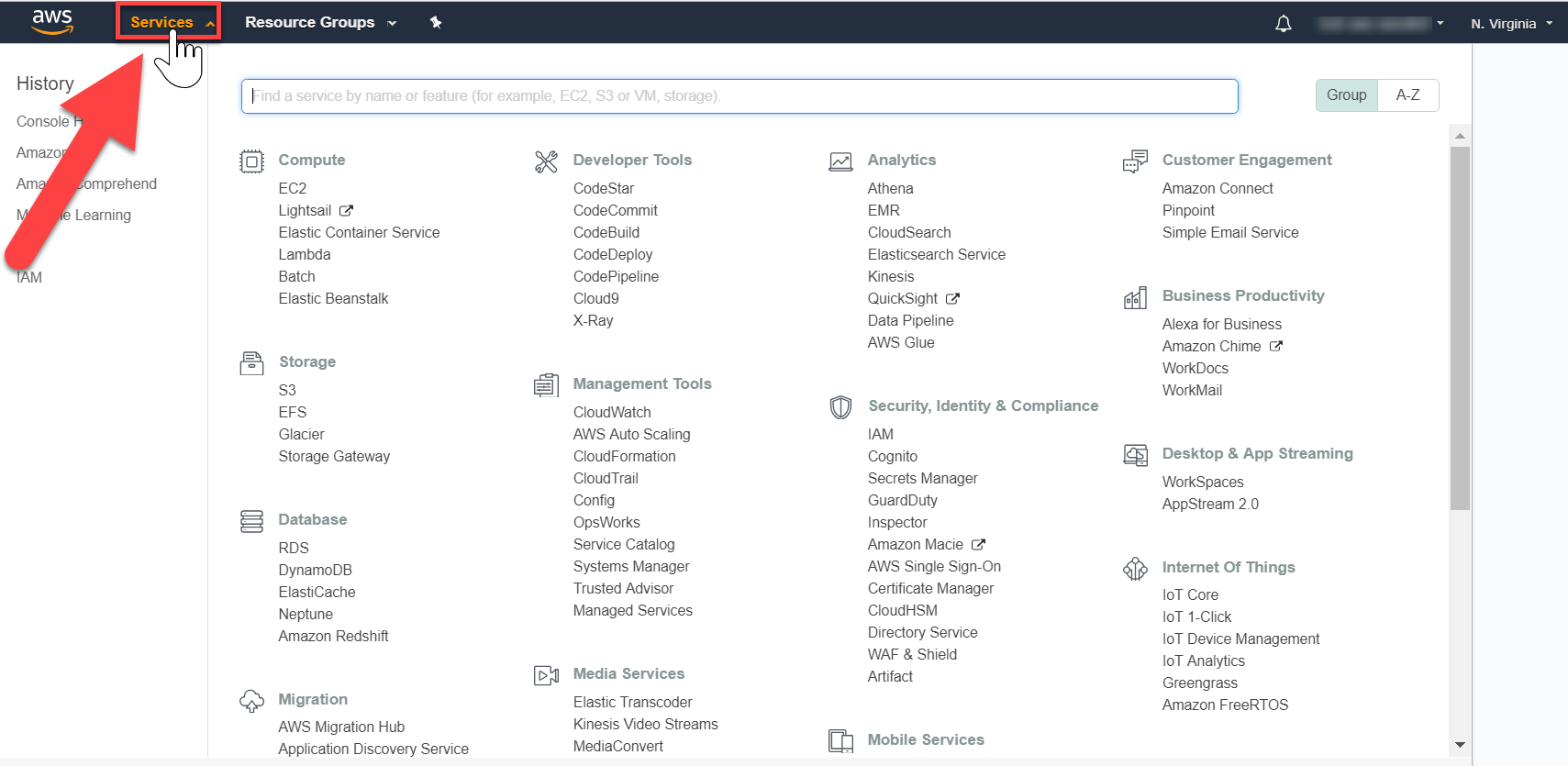


Figure 1.31: AWS Console

1. Scroll down the page to view all of the Machine Learning services. Then, click on Amazon Lex:

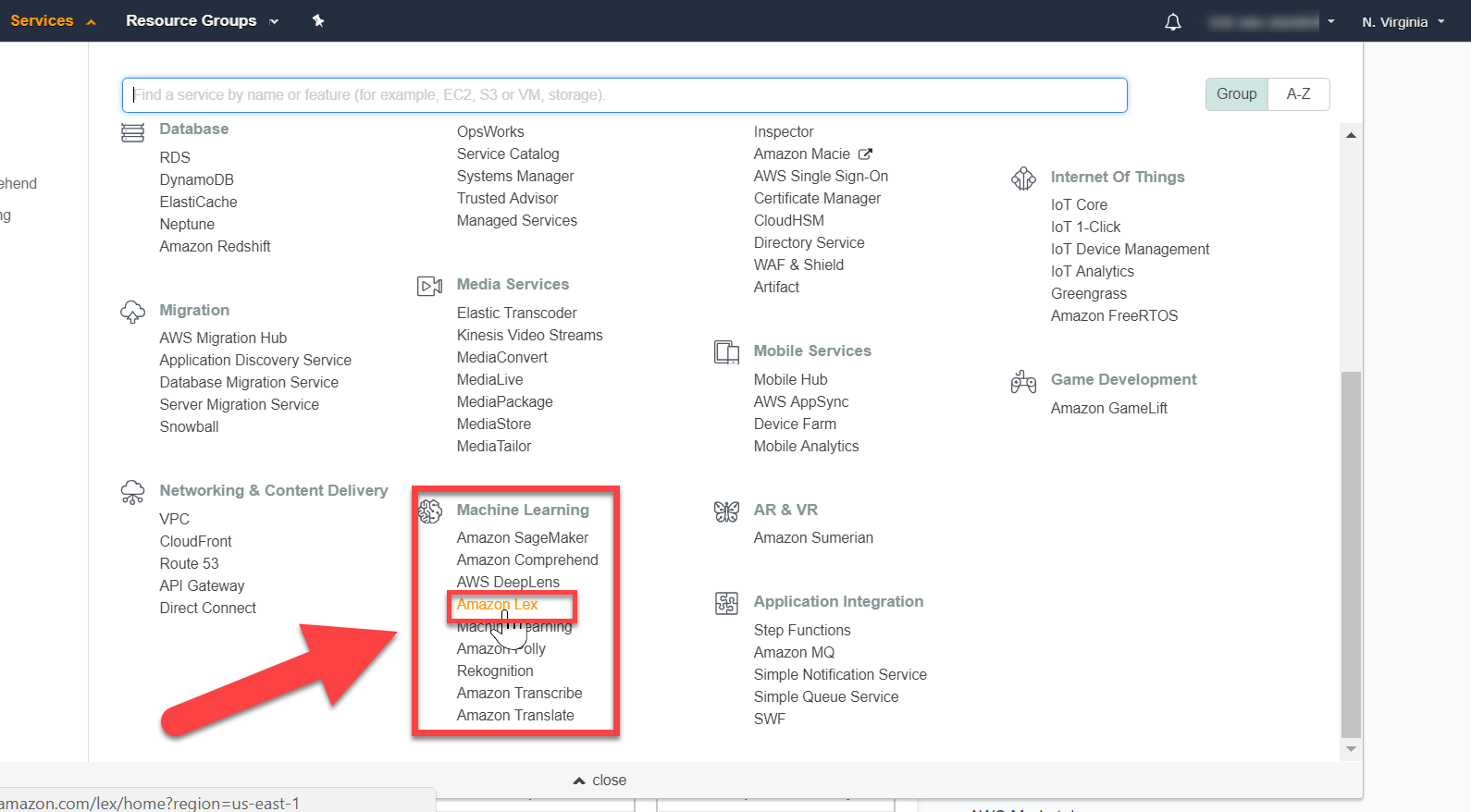


Figure 1.32: Options for Machine Learning

1. You will be redirected to the Amazon Lex home screen:

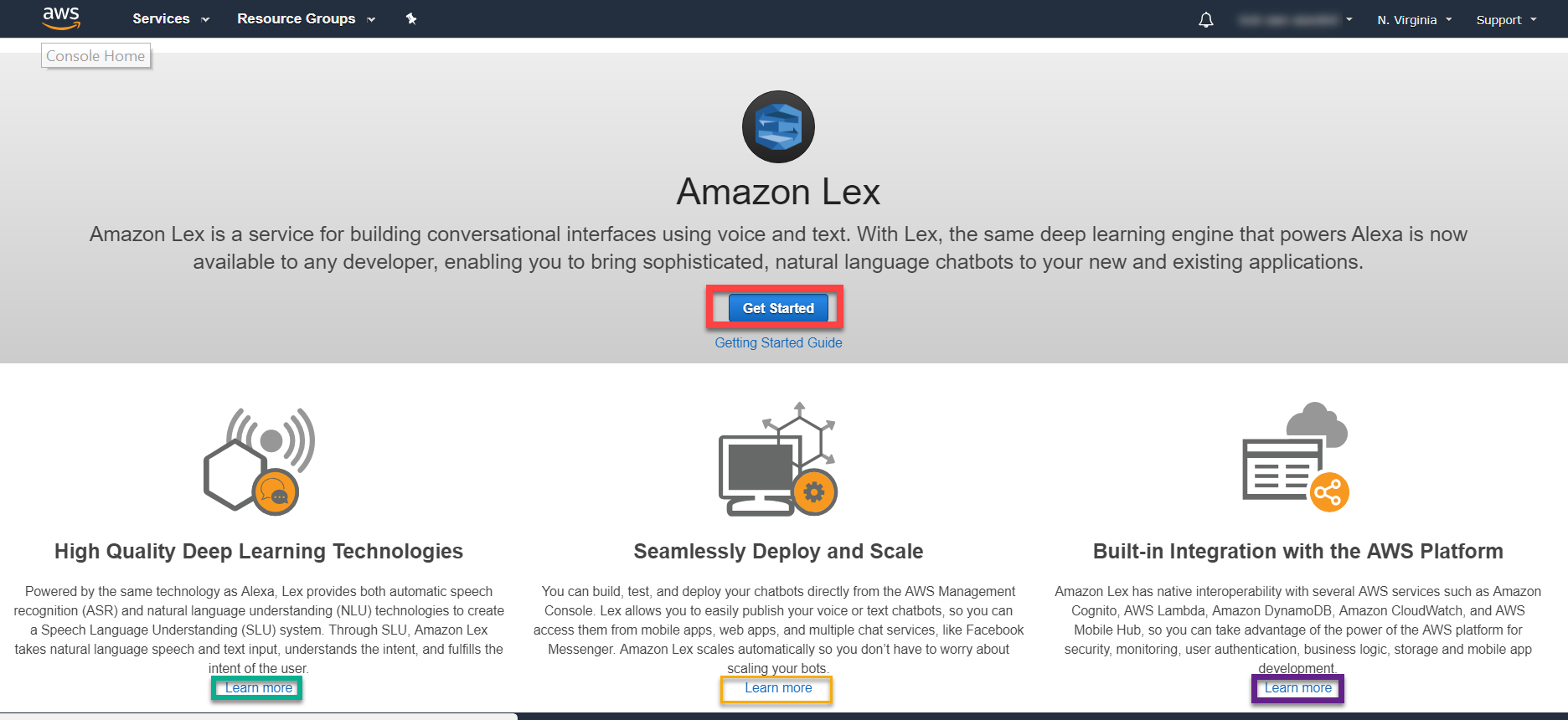


Figure 1.33: Amazon Lex home screen

Locating new AWS Services is an essential skill for discovering more tools to provide solutions for your data projects. Now, let's review another way to locate Machine Learning resources via the Search bar.

Activity 1.02: Testing the Amazon Comprehend's API Features

In this activity, we will display text analysis output by using a partial text file input in the API explorer. Exploring an API is a skill that saves development time by making sure that the output is in a desired format for your project. Thus, we will test Comprehend's text analysis features.

Suppose that you creating a chatbot. You have identified a business topic and the corresponding text documents, with content that will allow the chatbot to make your business successful. Your next step is to identify/verify an AWS service to parse the text document for sentiment, language, key phrases, and entities. Before investing time in writing a complete program, you want to test the AWS service's features via the AWS Management Console's interface. To ensure that this happens correctly, you will need to search the web for an article (written in English or Spanish) that contains the subject matter (sports, movies, current events, and so on) that you're interested in. The AWS Management Console is also accessible via the root user's account.

You are aware that exploring APIs is a skill that can save development time by ensuring that the output is in a desired format for your project. The following are the steps of completion:

1. Identify an AWS service, via the AWS Management Console, to accomplish your objectives.
2. Navigate to your web page of choice, which contains articles in English and Spanish.
3. Copy the text from the article written in English or Spanish, in order to identify the following features: sentiment, language, key phrases, and entities.
4. Obtain a score representing the articles: sentiment, language, key phrases, and entities.

Note

To refer to the detailed steps, go to the Appendix A at the end of this book on Page no. 194

Solution:

In this section, we will learn about display text analysis output using a partial text file input in the API explorer. We will be exploring API's is a skill to save development time by making sure the output is in a desired format for your project. Thus, we will test Comprehend's text analysis features.

We will consider an example suppose you are an entrepreneur creating a chatbot. You identified a business topic and corresponding text documents with content that will allow the chatbot to make your business successful. Your next step is to identify/verify an AWS service to parse the text document for sentiment, language, key phrases, and entities. Before investing time in writing a complete program, you want to test the AWS service's features via the AWS management console's interface. To ensure that it happen correctly, you will need to have search the web for an article (written in English or Spanish) that contains a subject matter (sports, movies, current events, etc.) that you are interested. And, also AWS Management Console accessible via the root user's account

You are aware exploring API's is a skill to save development time by making sure the output is in a desired format for your project.

1. Click in the Search bar of AWS Services to search the service name

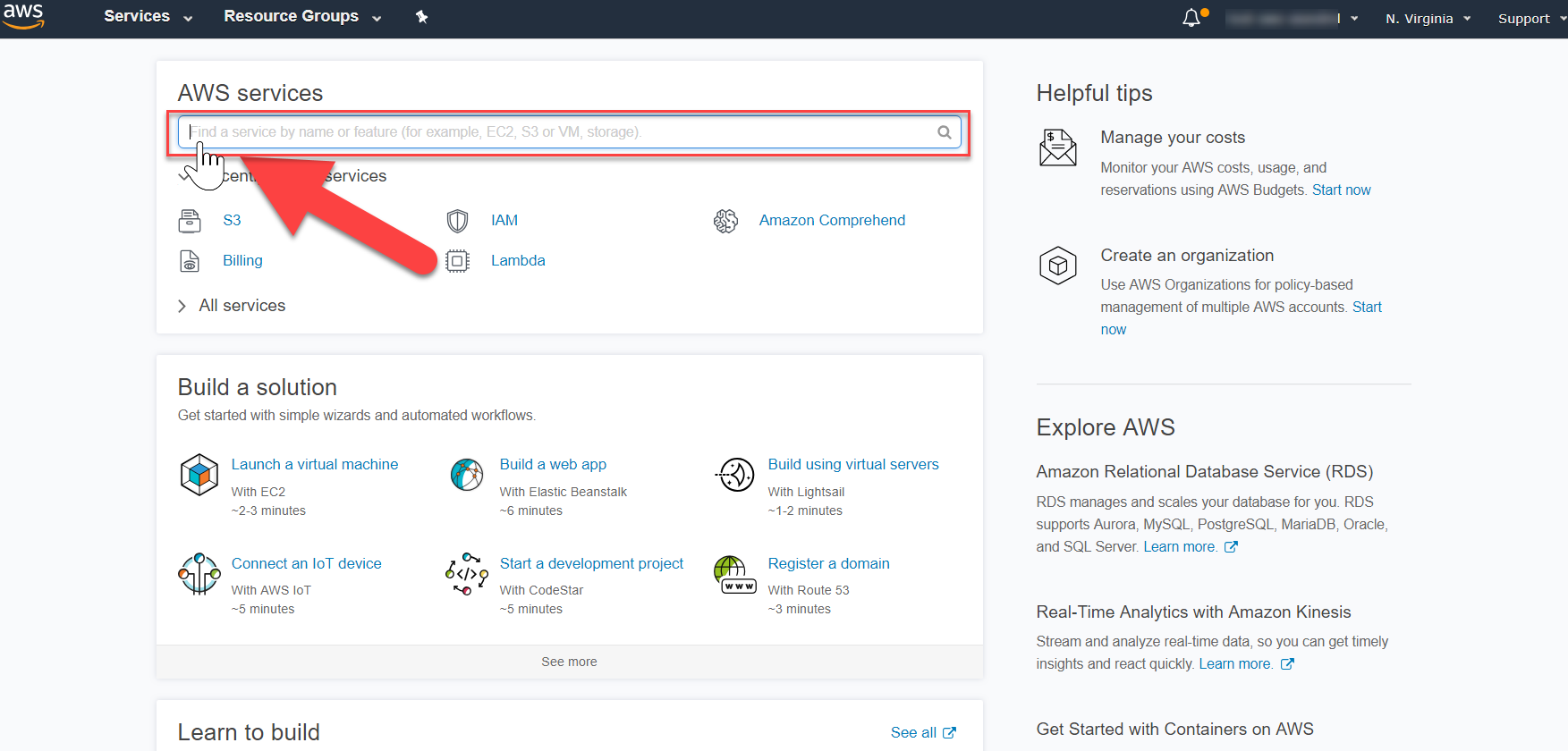


Figure 1.42: Searching of AWS Service

1. Search the Amazon Comprehend option and select the option you will be directed. Get started screen.

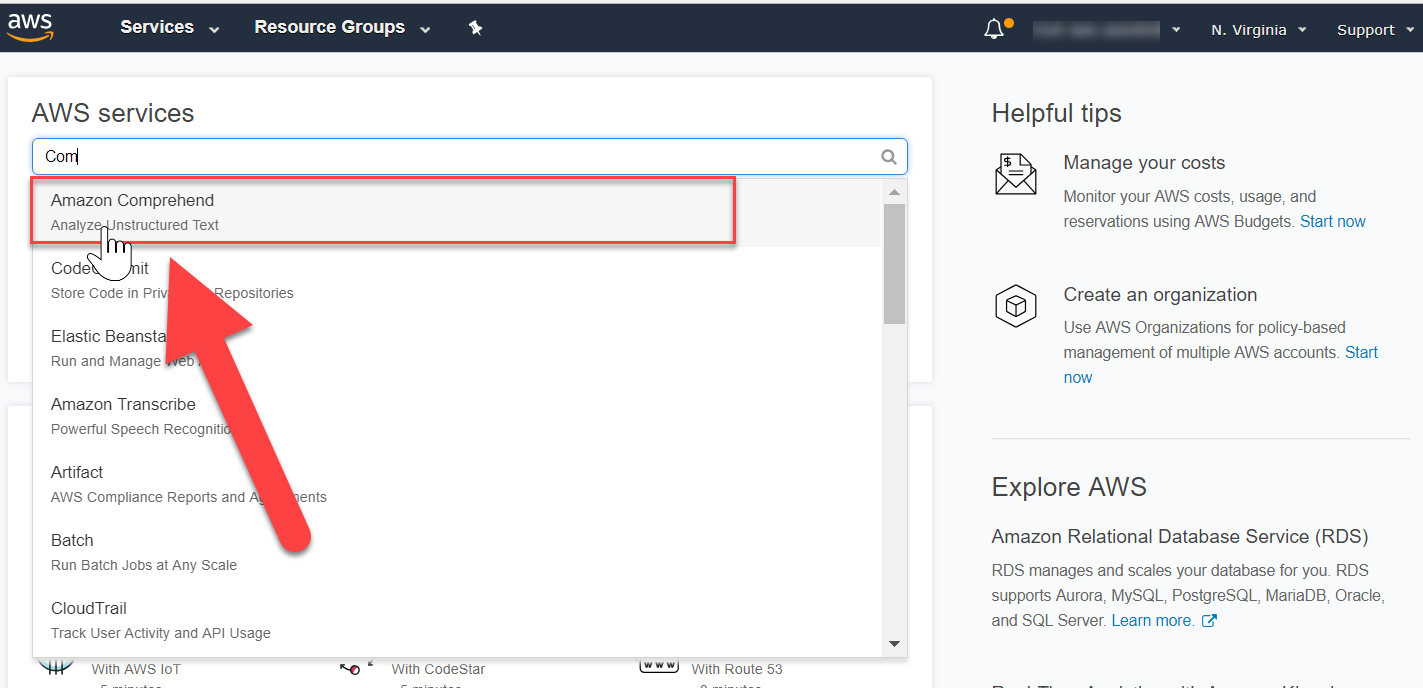


Figure 1.43: Selecting the Service

1. You will be directed to the API explorer. Navigation to Topic modeling and Documentation. The middle is a GUI to explore the API, and the right side provides real-time output for text input.

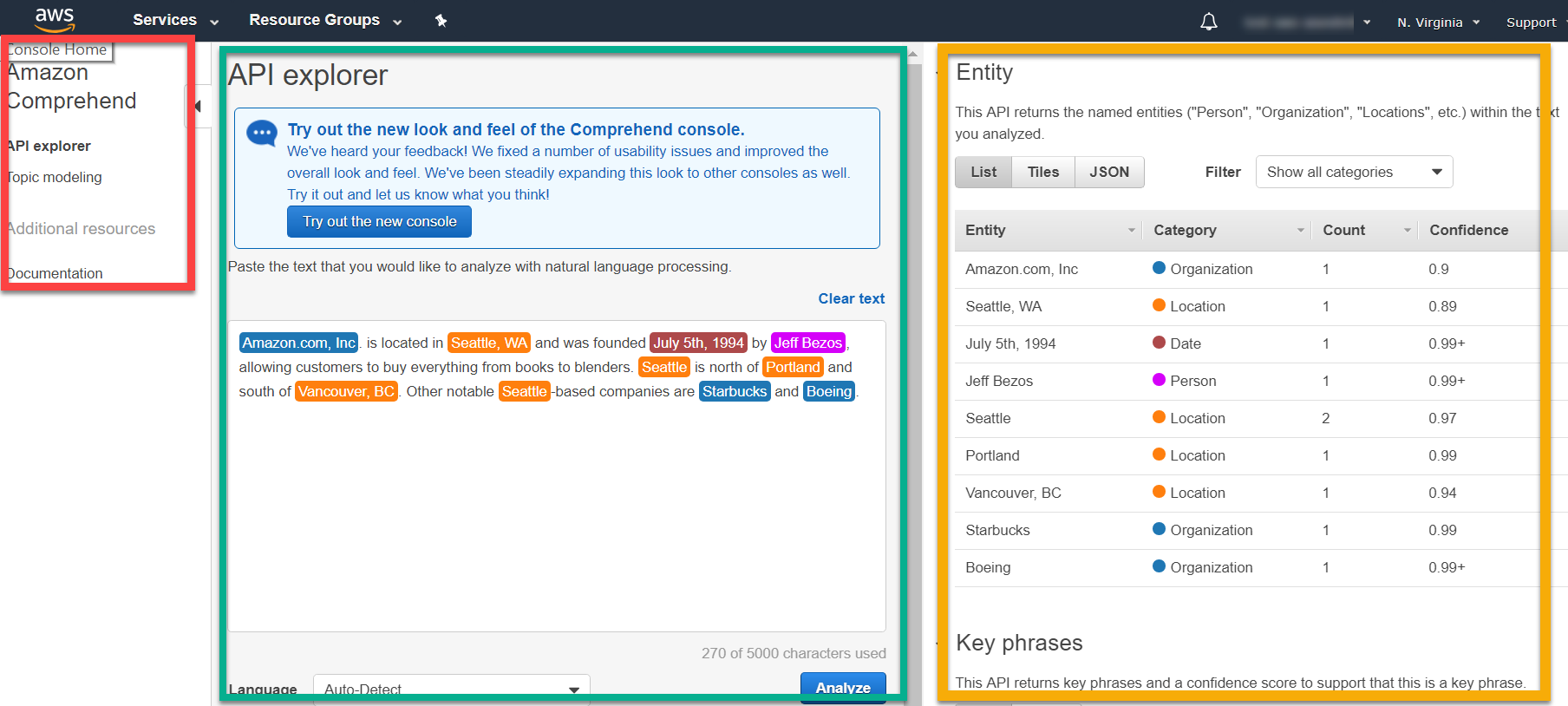


Figure 1.44: API Explorer

1. Click clear text to clear all default services. Navigate to open the following URL in a new tab <http://www.gutenberg.org/cache/epub/1322/pg1322.txt>

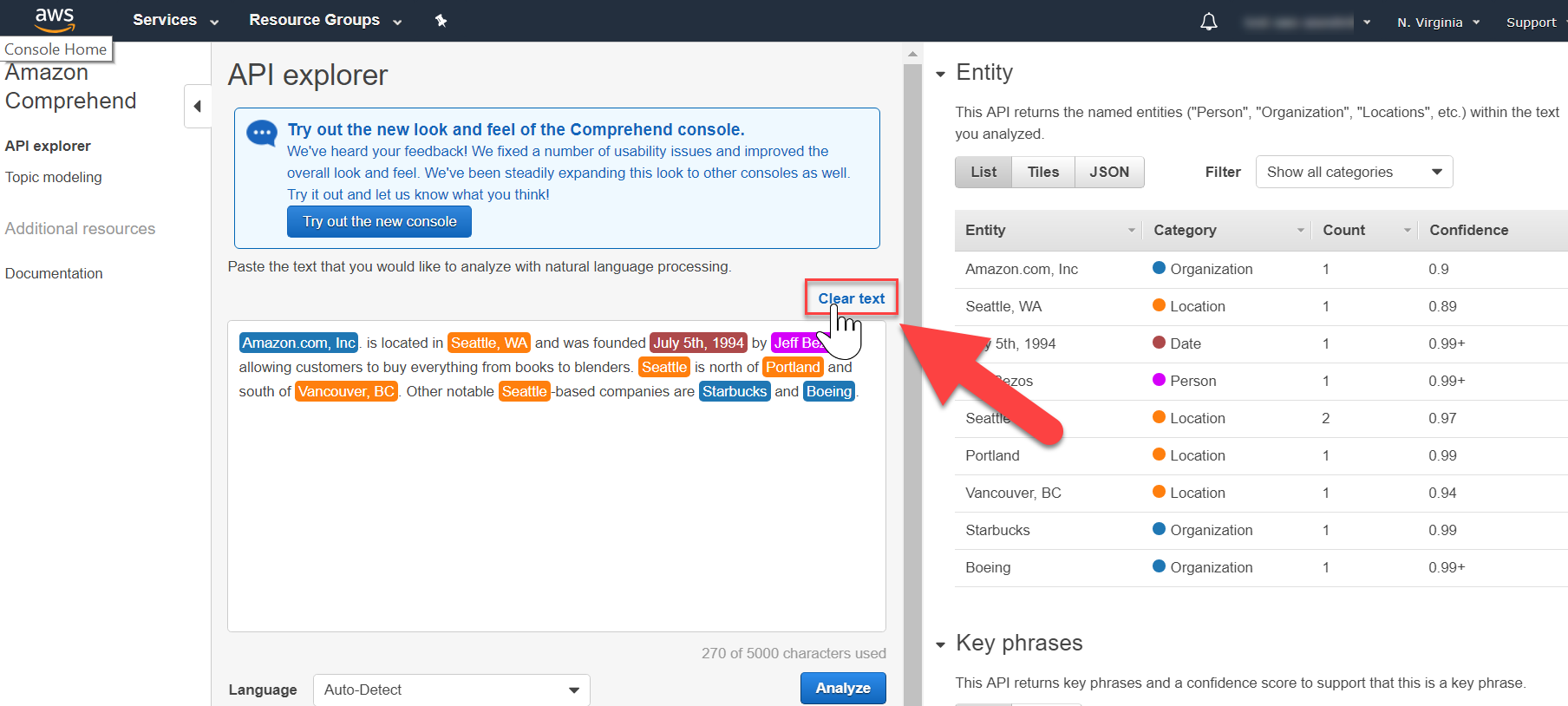


Figure 1.45: API Explorer Screen

1. Copy the first poem, and paste it in the Explorer and click Analyze to see the output:

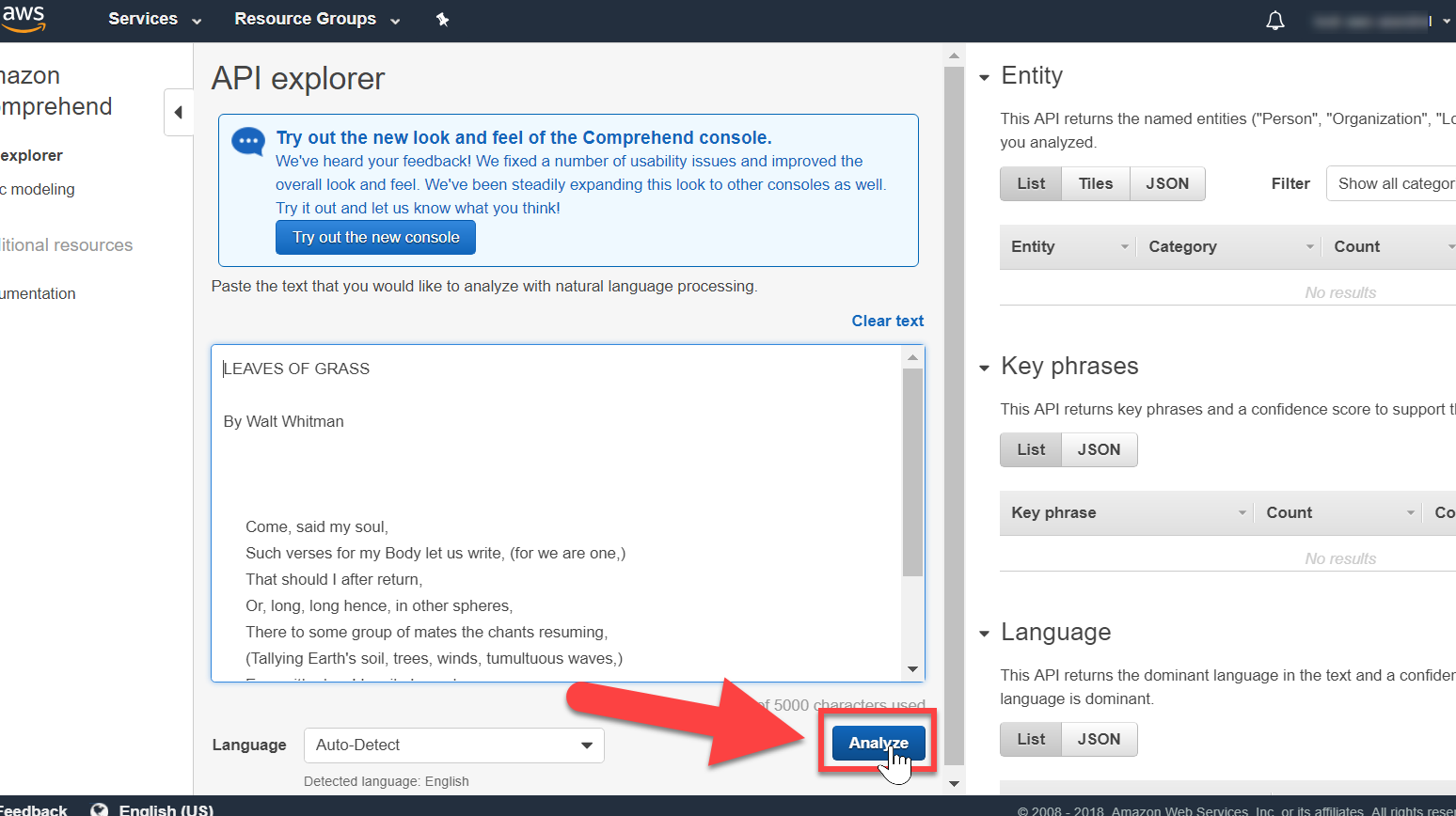


Figure 1.46: Analyzing the Output

1. Review the right side of the screen for Entity, Key phrases, Language, and scroll down to view Sentiment:

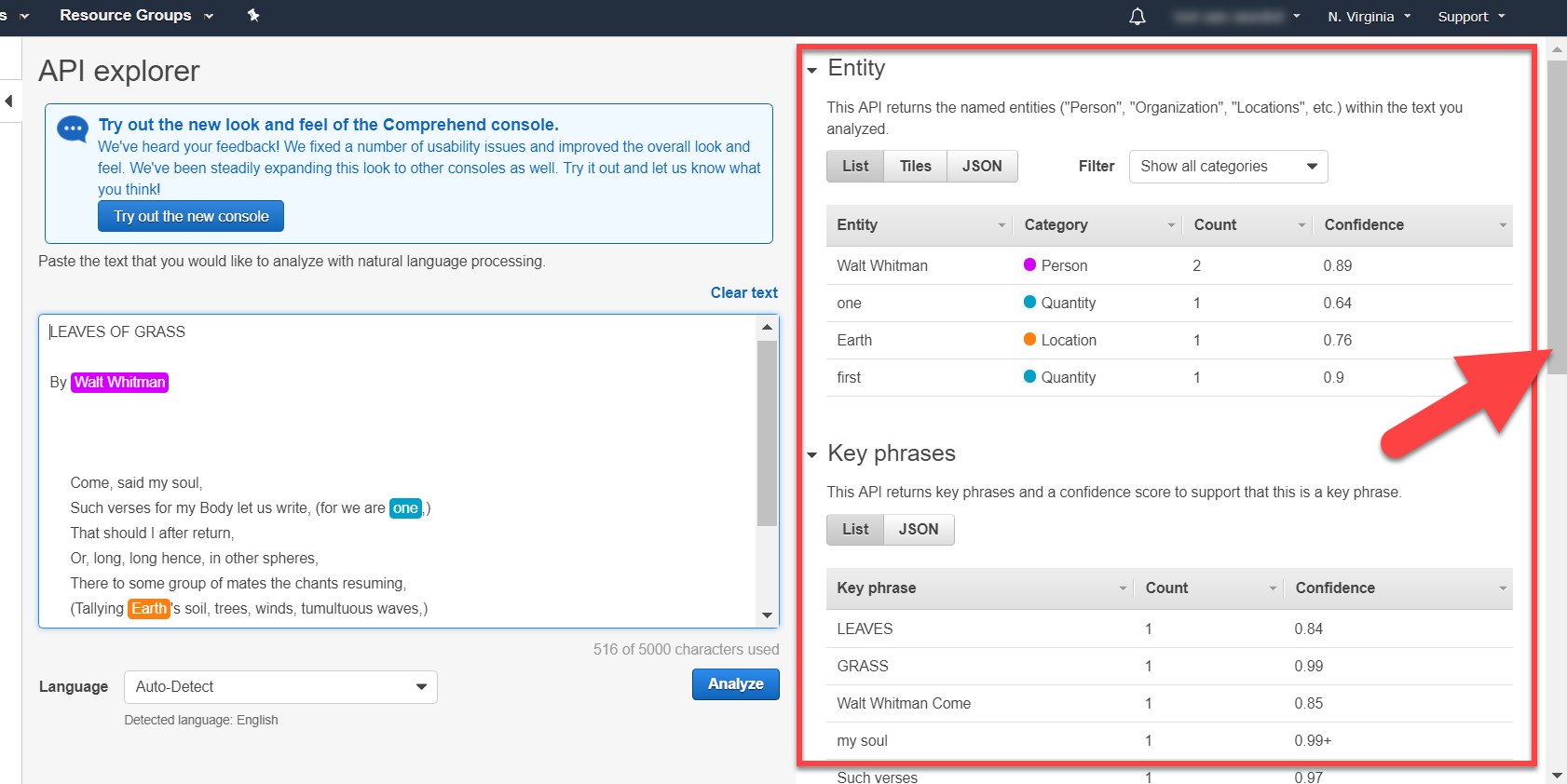


Figure 1.47: Result

You now know how to explore an API, and how to display sentiment output for the same.

Summary

At the beginning of this chapter, we explained what Amazon Web Services is, what Machine Learning is, and what Artificial Intelligence is. Following this, you learned what Amazon S3 is, and why Amazon S3 is used. You also explored the basic requirements for working with AWS using S3. With this, you used IAM (Identity and Access Management).

Next, you learned how to import and export data into S3. Following that, you explored the components of S3. At the same time, you learned about the REST Interface. In the last part of this chapter, we looked at the AWS command line and its usages. Finally, we explored the concepts of recursion and parameters, and how to use the AWS Console to identify Machine Learning services.

In the next chapter, you will learn how to summarize text documents by using Natural Language Processing (NLP). Researching new AWS services is essential for discovering additional solutions to solve any machine learning problems that you are working on.