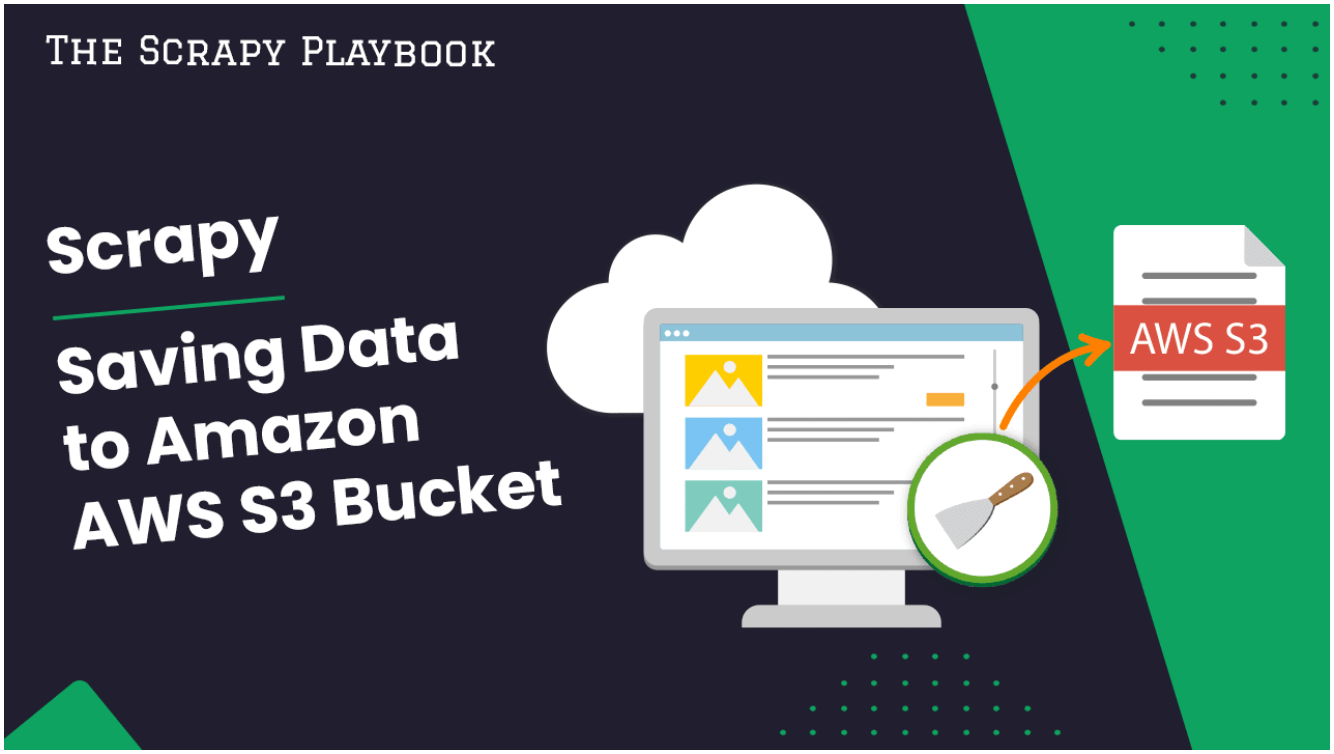


Saving Scraped Data To Amazon AWS S3 Bucket With Scrapy

Storing CSV and JSON files of scraped data on your local computer is find for small projects, however, a better option is to store it on a file storage system like in a Amazon AWS S3 bucket.

In this guide, we show you how to store your scraped data in CSV and JSON files and automatically sync it with AWS S3 bucket.

First, let's go over what are **Scrapy Feed Exporters**.

What Are Scrapy Feed Exporters?

The need to save scraped data to a file is a very common requirement for developers, so to make our lives easier the developers behind Scrapy have implemented Feed Exporters.

Feed Exporters are a ready made toolbox of methods we can use to easily save/export our scraped data into:

- JSON file format
- CVS file format
- XML file format
- Pythons pickle format

And save them to:

- The local machine Scrapy is running on
- A remote machine using FTP (file transfer protocall)
- Amazon S3 Storage
- Google Cloud Storage
- Standard output

In this guide, we will walk you through how you can save your data to CSV and JSON files and store those files in a AWS S3 bucket using Scrapy.

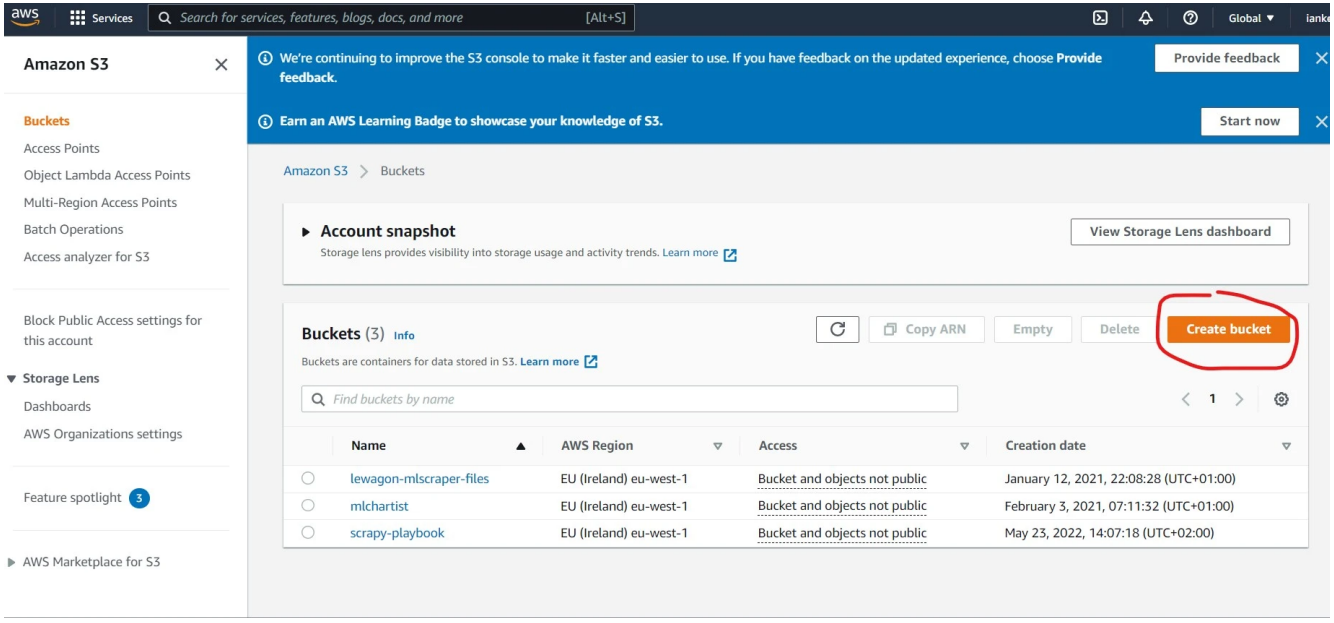
Getting Setup With Amazon AWS S3

The first step is we need to setup a AWS S3 bucket, get the appropriate access keys and install the `botocore` library if you haven't done so already.

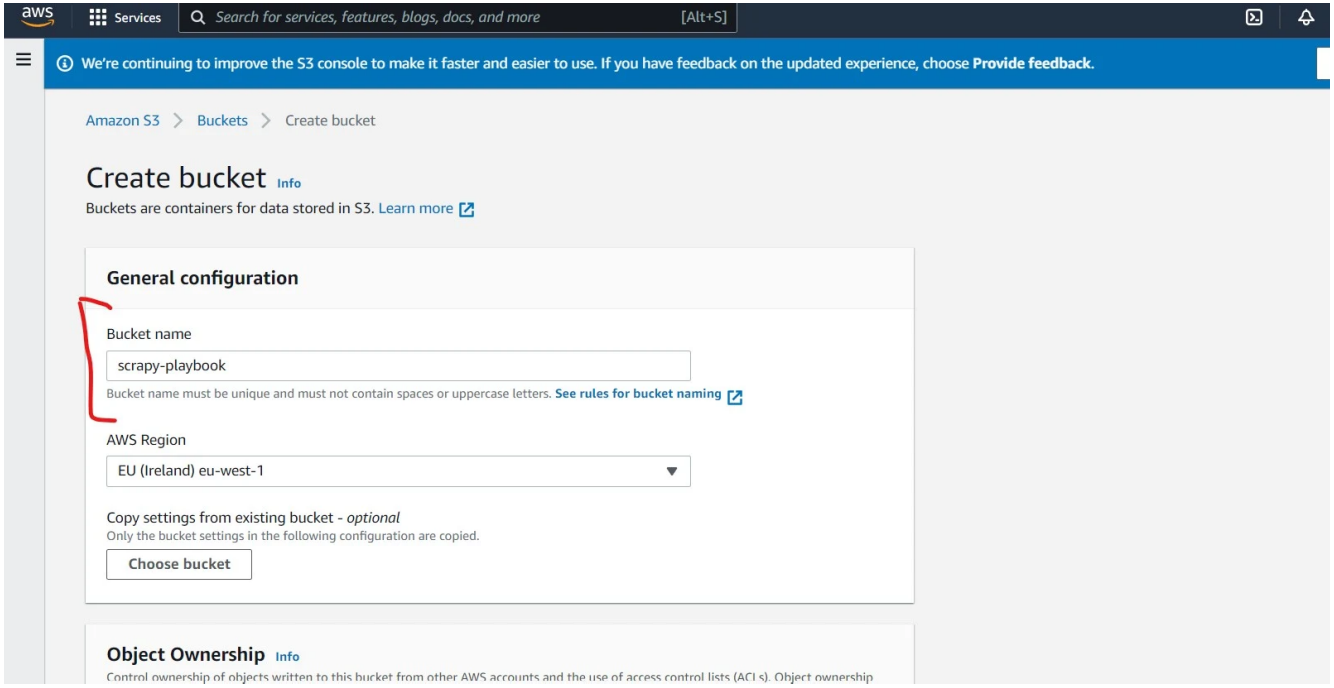
1. Create AWS S3 Bucket

First we need to signup for a AWS account here. Amazon gives you 5GB of free S3 storage for 12 months on their free tier.

Once logged in navigate to S3 by entering **"S3"** into the search bar, and once there click on the **"Create Bucket"** button.



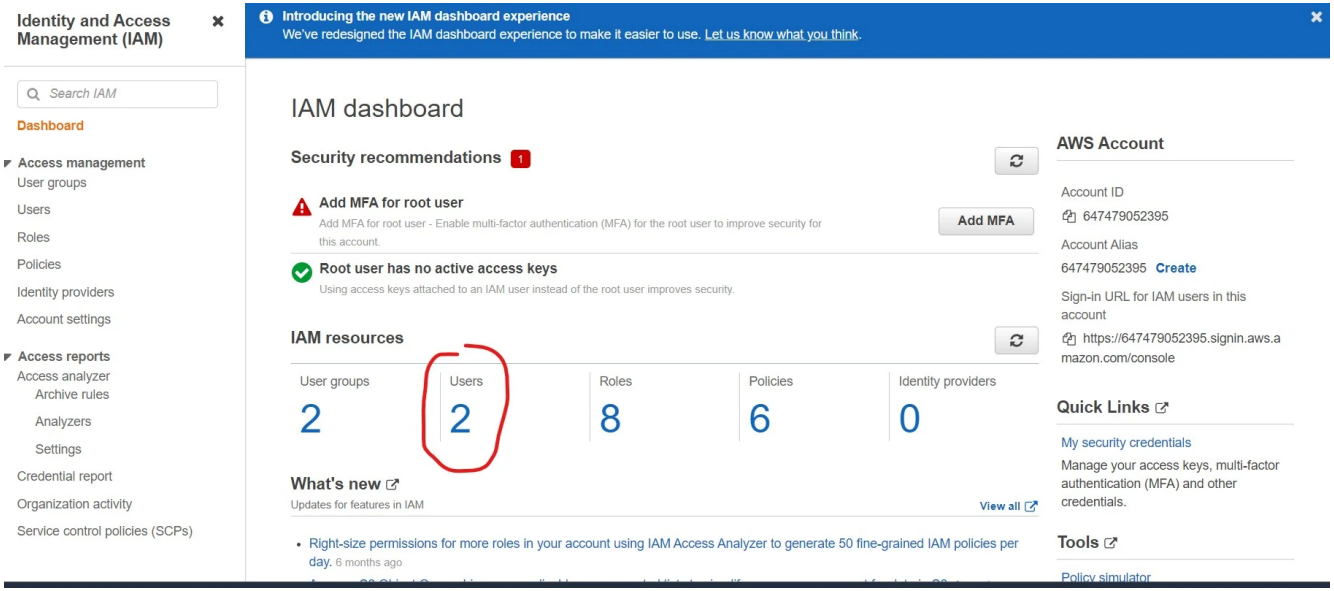
Then give your bucket a name, and you can leave all the default settings as they are.



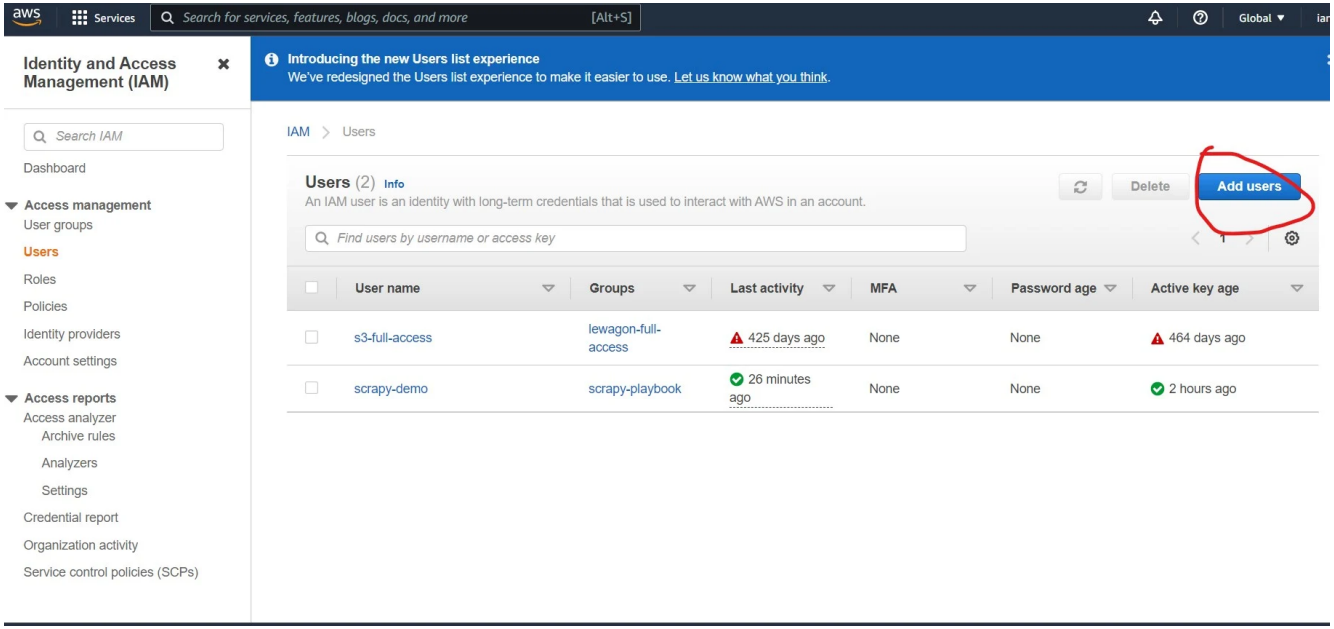
2. Generate S3 Access Credentials

Next, we need to enter **IAM** into the search bar and navigate to the **IAM service**.

Once there, click on **Users**.

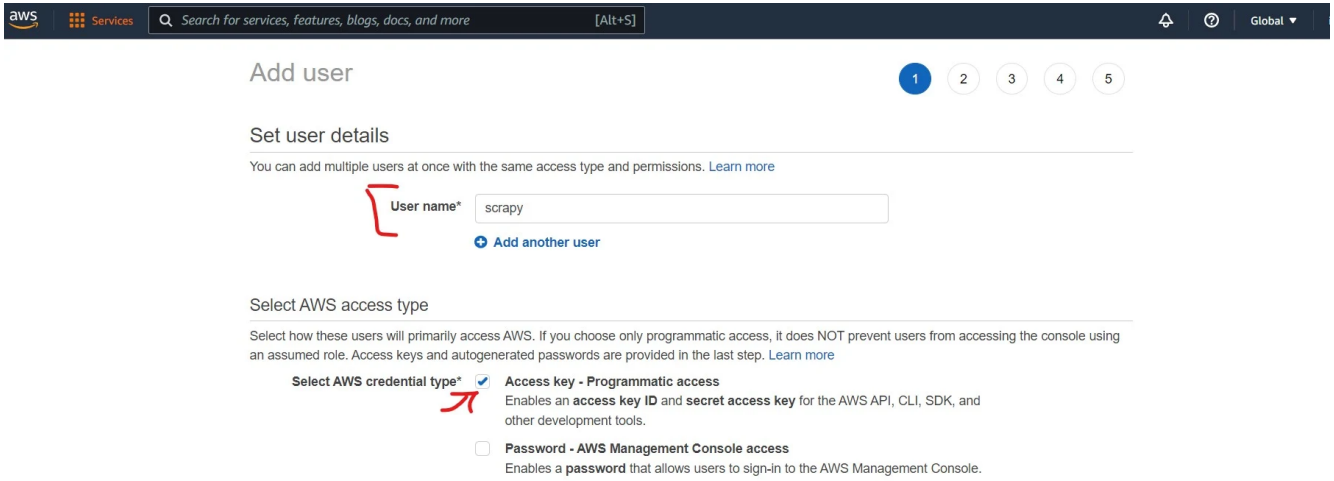


Then on **Add Users**.

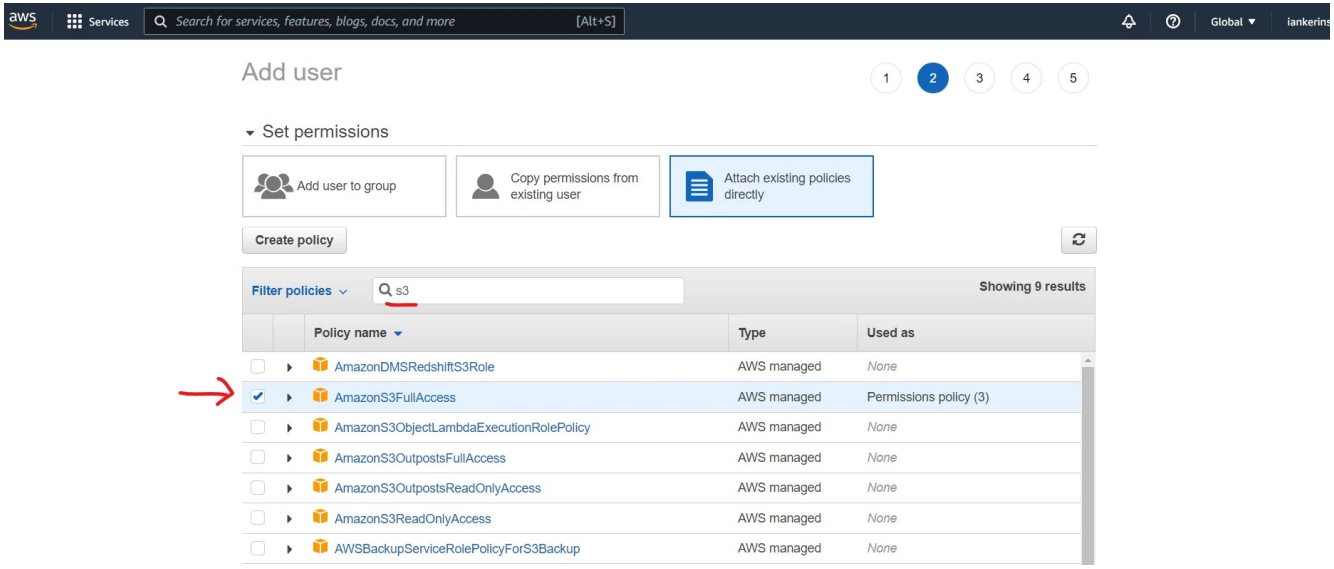


Here give your user a user name, and select **Access key - Programmatic access** in the **Select AWS access type**

section.

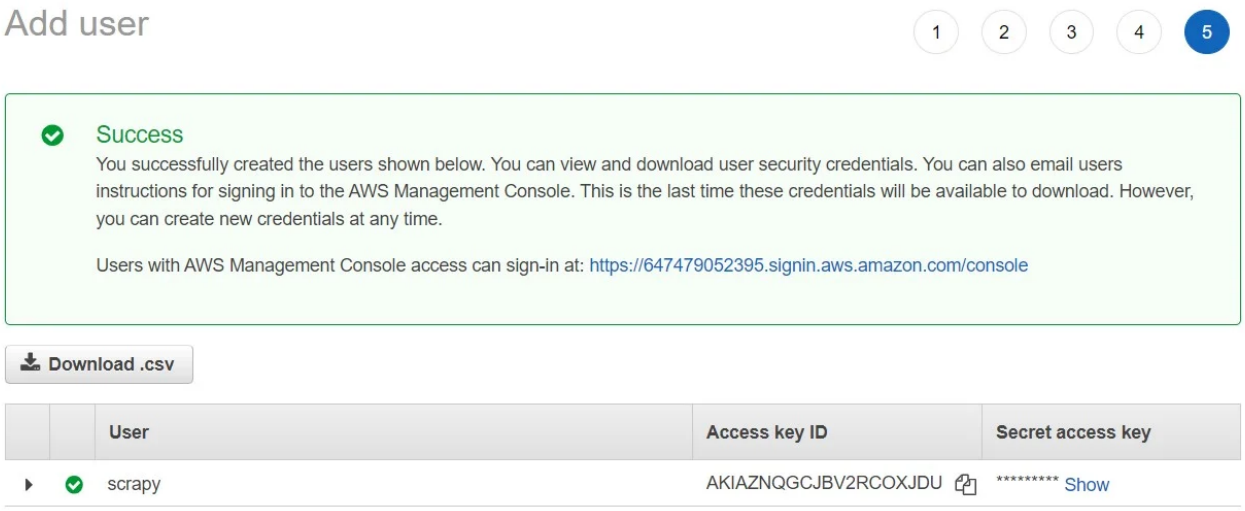


Then click on **Attach existing policies directly**, and search for **"S3"** in the search bar and select **AmazonS3FullAccess**.



Finally, skip the **Tags** section and click **Create User**. Here you will be shown the following screen showing your

username, **Access key ID**, and **secret access key**. Copy these down.



---

3. Install botocore

Finally, the last step to getting setup is to install `botocore` on our machine which is as simple as:

```
pip install botocore
```

Saving CSV Files To AWS S3 Bucket

Configuring Scrapy to save our S3 bucket is very simple. We just need to update the `settings.py` with the following:

```
FEEDS = {
```

```
    "s3://scrapy-playbook/%(name)s/%(name)s_%(time)s.csv": {
    "format": "csv",
    }
}

AWS_ACCESS_KEY_ID = 'YOUR_AWS_ACCESS_KEY_ID'
AWS_SECRET_ACCESS_KEY = 'YOUR_AWS_SECRET_ACCESS_KEY'
```

Here, I'm saving my CSV files into the `scrapy-playbook` S3 bucket, and inside that bucket it will create/use the folder using the name of my spider and call the file **SpiderName_TimeCreated.csv**.

Now, everytime I run my spider the CSV files will be saved to my S3 bucket.

Saving JSON Files To AWS S3 Bucket

Configuring Scrapy to save our CSV files to our S3 bucket is very simple. We just need to update the `settings.py` with the following:

```
FEEDS = {
    "s3://scrapy-playbook/%(name)s/%(name)s_%(time)s.jsonl": {
    "format": "jsonlines",
    }
}

AWS_ACCESS_KEY_ID = 'YOUR_AWS_ACCESS_KEY_ID'
AWS_SECRET_ACCESS_KEY = 'YOUR_AWS_SECRET_ACCESS_KEY'
```

Here, I'm saving my JSON files into the `scrapy-playbook` S3 bucket, and inside that bucket it will create/use the folder using the name of my spider and call the file **SpiderName_TimeCreated.jsonl**.

Now, everytime I run my spider the JSON files will be saved to my S3 bucket.

More Scrapy Tutorials

For more information on saving your scraped data to different formats then be sure to check these guides:

- Saving Data to CSV
- Saving Data to JSON
- Saving Data to SQLite Database
- Saving Data to MySQL Database
- Saving Data to Postgres Database

If you would like to learn more about Scrapy in general, then be sure to check out The Scrapy Playbook.