



# AWS Crawler Setup Documentation

## Introduction

In this document we will explore how to setup the *AWS crawler*.

## Prerequisites

Before we investigate how we setup this crawler, be sure you have the following at hand:

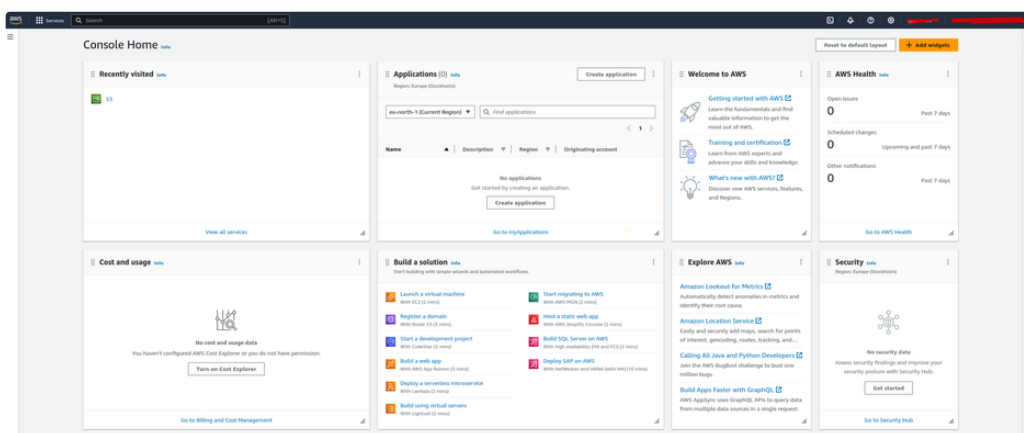
- **Bucket Name** - the name of the bucket we wish to *crawl*
- **Access Key** - unique *identifier* of the user trying to access the bucket (be sure you have permissions to access)
- **Secret Key** - key of the user
- **Region** - the region assigned to the bucket

Be sure that you have the *correct permissions* in your AWS account to access the desired buckets

## Retrieving Required Credentials

If you are unsure *where* to get the credentials described above, keep reading.

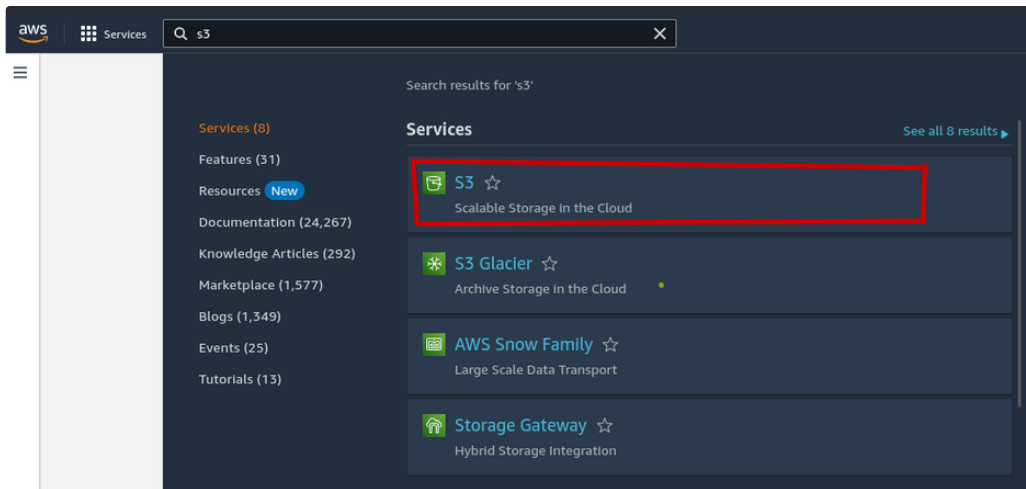
To begin, navigate to the *main dashboard* of your AWS account, it should look similar to this:



Console Home (Main AWS Page)

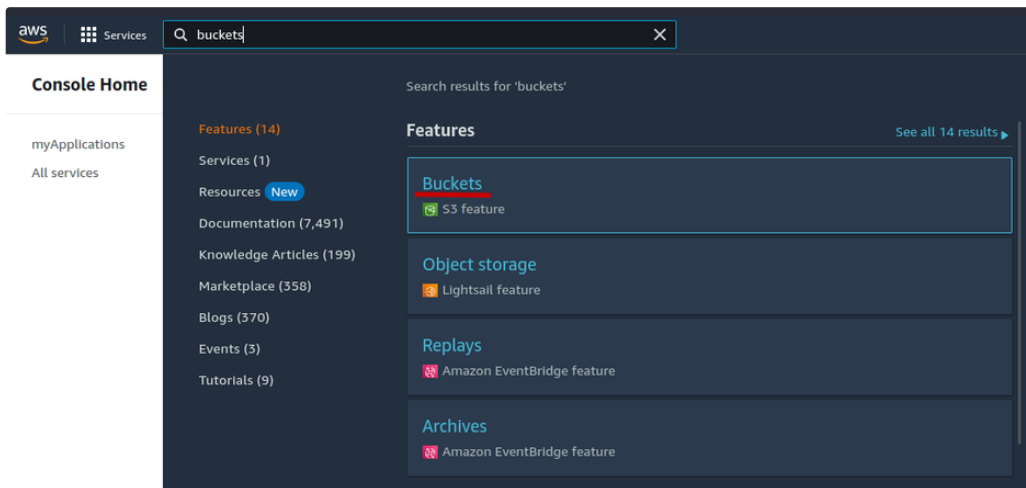
This should be the first page you are directed to when logging into your account.

To get to your desired *bucket*, type the *name* of your bucket in the search bar:



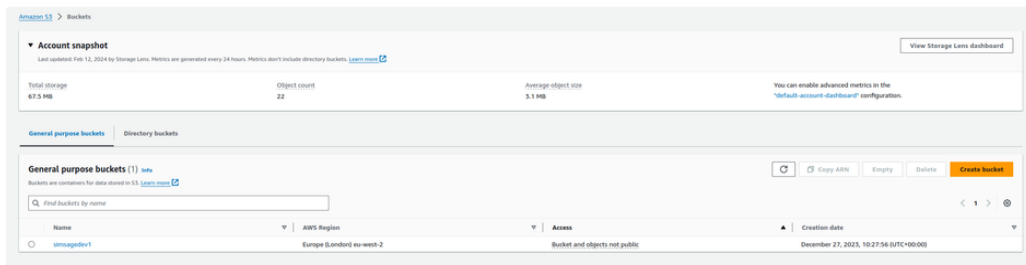
Buckets

Alternatively type buckets into the search bar:



## Bucket Name & Region

After selecting your bucket, you should be redirected to something similar to the following:



View a bucket

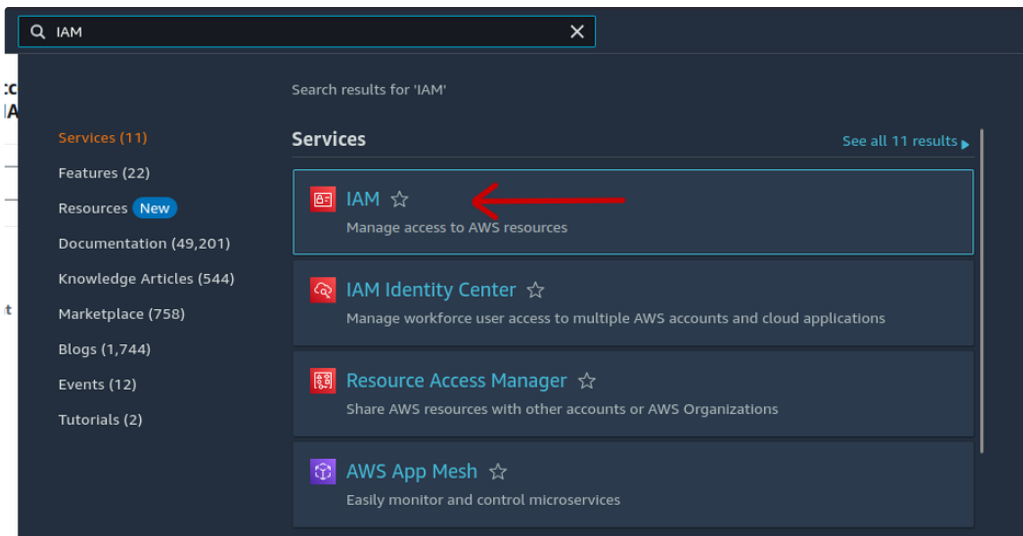
Note the columns 'Name' and 'AWS Region':

| Name               | AWS Region                       |
|--------------------|----------------------------------|
| <u>simsagedev1</u> | <u>Europe (London) eu-west-2</u> |

Name and AWS Region Credentials

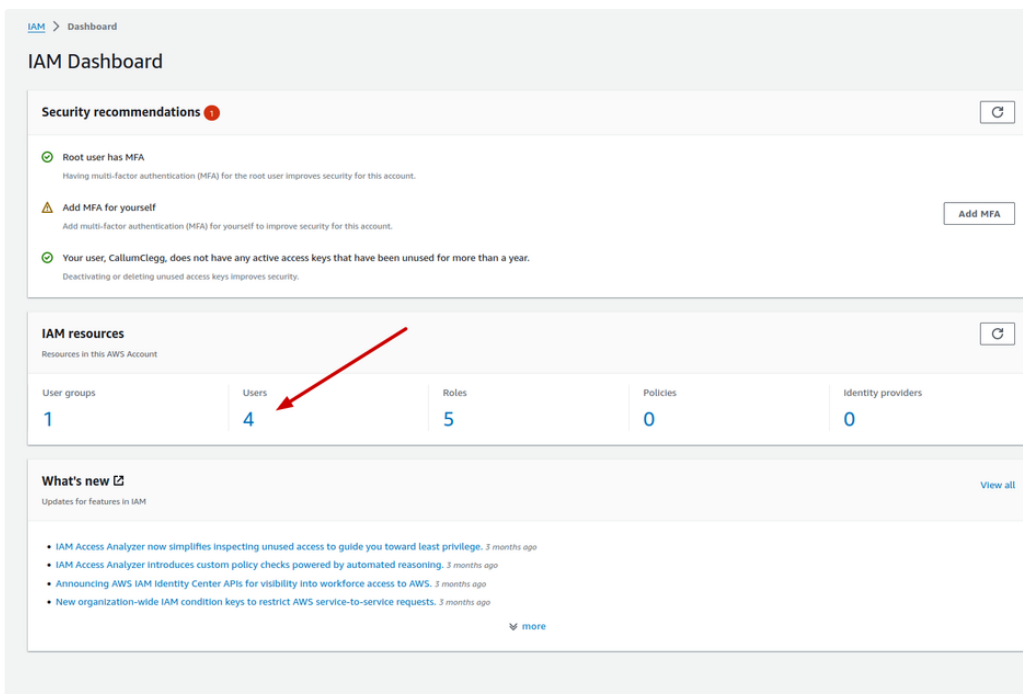
## Access Key & Secret Key

Type 'IAM' into the AWS search bar, and navigate to this page:



Select 'IAM'

You should be presented with the following page below, select 'Users':



Go to 'Users'

Next, select your account in the table shown:

[IAM](#) > [Users](#)

▼ Ready to streamline human access to AWS and cloud apps?

Identity Center is enabled. We recommend managing workforce users' access to

[Learn more](#) | [Watch how it works](#)

**Users (4)** [Info](#)

An IAM user is an identity with long-term credentials that is used to interact with AWS in an account.

| <input type="checkbox"/> | User name ▲                                     | Path ▼ | Group: ▼ | Last |
|--------------------------|---|--------|----------|------|
| <input type="checkbox"/> | <a href="#">CallumClegg</a>                     | /      | 1        | ...  |
| <input type="checkbox"/> | <a href="#">max@simsage.co.uk</a>               | /      | 1        | ...  |
| <input type="checkbox"/> | <a href="#">MaxGibbshotmail</a>                 | /      | 1        | ...  |
| <input type="checkbox"/> | <a href="#">rbandaru@ondevicesolutions.c...</a> | /      | 1        | ...  |

Select your account

We should now be presented with the option to create a new access key, if one is not yet setup:

[IAM](#) > [Users](#) > [CallumClegg](#)

**CallumClegg** [Info](#) [Delete](#)

**Summary**

**ARN:** `arn:aws:iam::190671109365:user/CallumClegg`

**Created:** February 13, 2024, 13:59 (UTC)

**Console access:** [Enabled without MFA](#)

**Last console sign-in:** Today

**Access key 1:** [Create access key](#)

Select 'Create access key', or not the key shown

Follow the AWS guidance if a access key needs to be created. We *recommend* you select 'Application running outside AWS' when creating a access key

⚠ Once created, be sure to note the *secret* - you will not be able to view this again

## Setting up the Crawler

Here is a example of what the input of some credentials *could* look like:

Edit Source: AWS TEST

General

AWS Crawler

Metadata

ACLs

Processors

Schedule

Bucket Name \*

simagedev1

Access Key (leave blank to keep previous)

\*\*\*\*\*

AWS

Setup Guide

Region \*

eu-west-2

Secret Key (leave blank to keep previous)

\*\*\*\*\*

Close


Test

Reset Delta

Save

Example AWS Crawler Setup

## Additional Material

 [Buckets overview - Amazon Simple Storage Service](#)