Data Sentinel 1.0.0 deployment using AWS Marketplace AMI

Introduction

Prescience Data Sentinel is a data quality assessment tool with over **~50 prebuilt data quality rules**. The tool allows enterprises to pick from these existing data quality rules or create their own new rules. It then provides a comprehensive analysis of data and automatically generates data quality scores and reports.

Overview

Data Sentinel supports ingestion of data from multiple data sources like **RDBMS, CSV, Parquet, JSON files** etc., this enabling an end-to-end view of the enterprise and business unit data. It securely ingests enterprise data without copying it and can be deployed as a standalone solution or be integrated with your existing data pipelines.

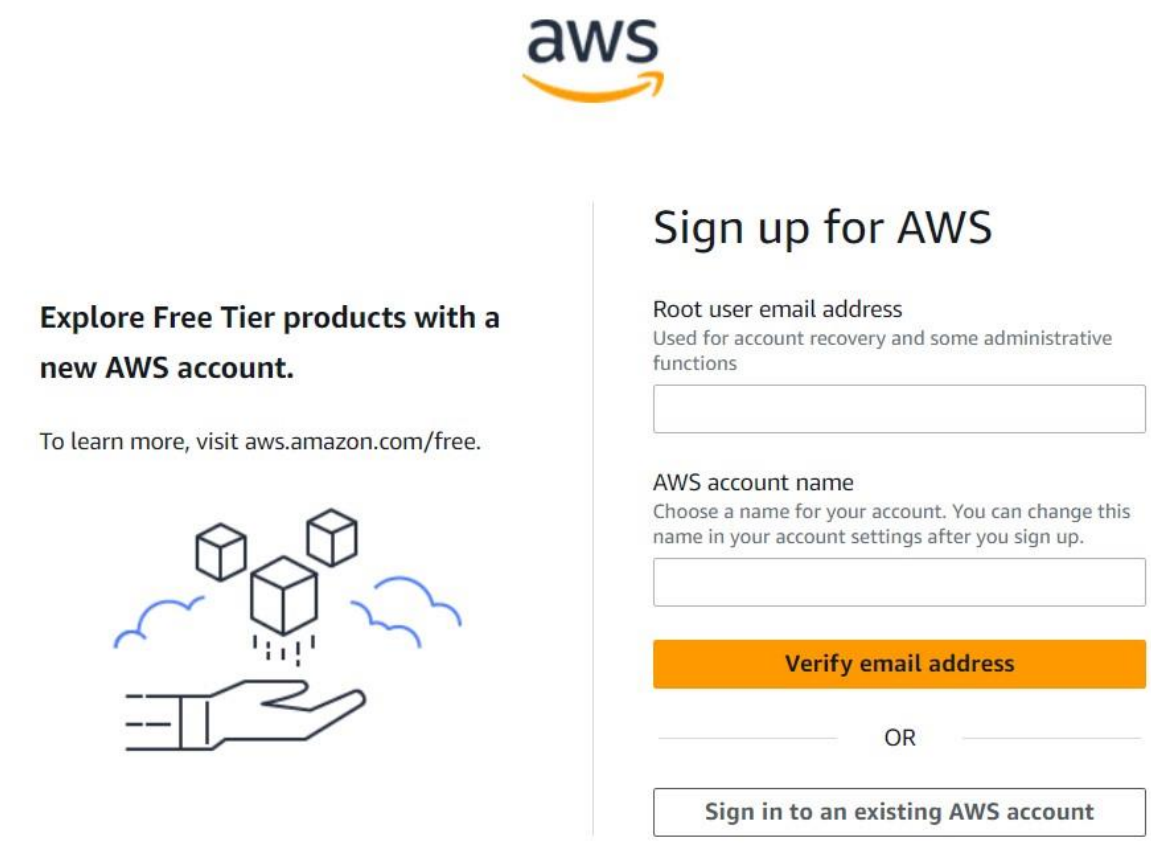
This trial version of the application can be used as many times as required, for a periods of **45 days from the registration date (first tenant creation date).**

Step 1: Register with Amazon Web Services (AWS)

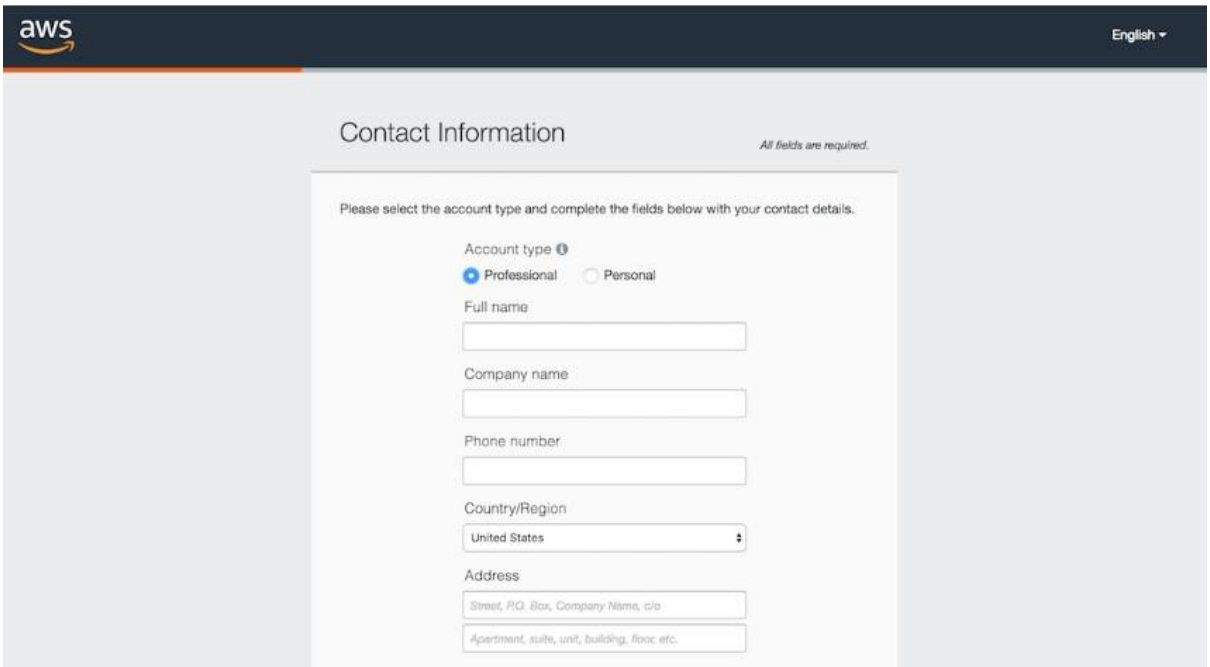
At the end of this step, you will have signed up for the **Amazon Web Services free tier**. If you already have an Amazon Web Services account, you may skip this step.

You will need an existing amazon account to log in and sign up. To create it, follow these steps:

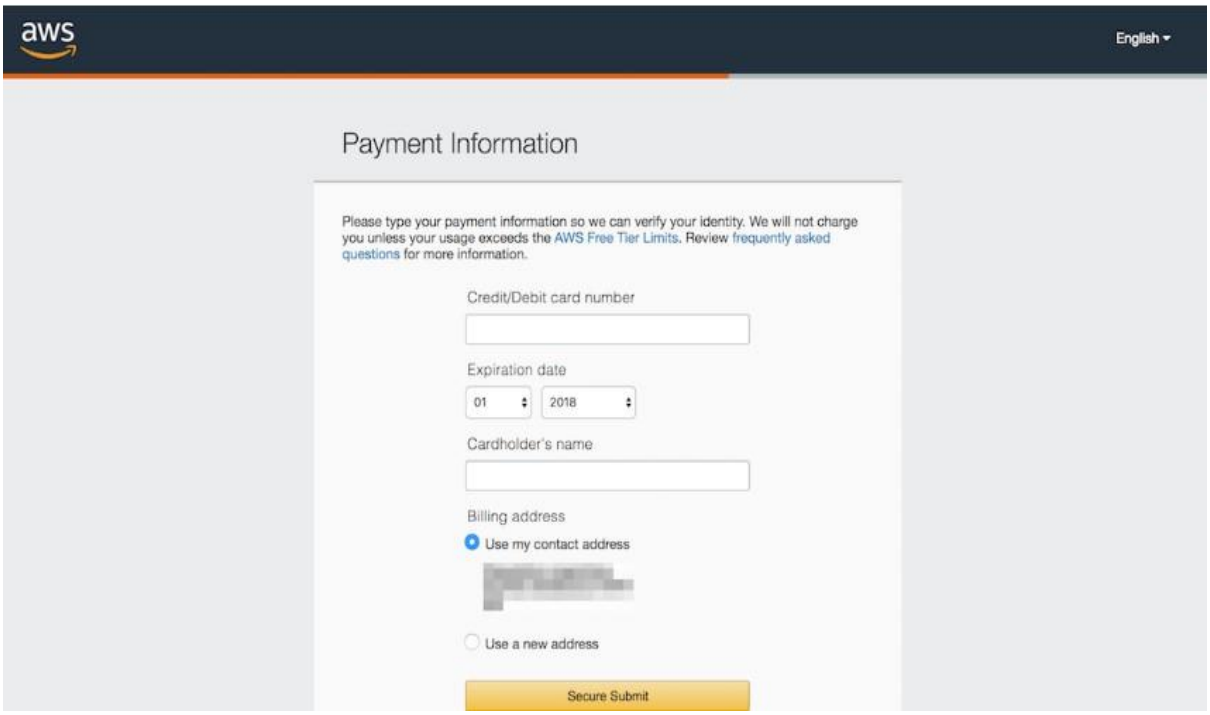
- Browse to http://aws.amazon.com and click the **Create an AWS account** button at the top of the page.
- In the resulting page, enter an email address, a password, and an AWS account name.
  Then, click **Continue** to start the registration process.



Once you've signed in to Amazon, sign up for AWS by selecting the account type and providing some basic contact information and your mobile phone number.



Once that's done, proceed to the next stage by entering your credit card information. Click the **Secure Submit** button to continue with the account creation.



If you are worried about how much you will be billed for services, relax. When you first sign up for AWS, you will get automatic access to the **AWS Free Tier"** which entitles you to 12 months of free usage up to certain limits. This includes 750 hours per month of free usage of **Amazon EC2** micro services, which are just right for development or low-traffic website hosting. SO long as your usage falls within the limits of the free tier, your credit card will never be billed. However, Amazon still needs your credit card information for security purposes, to avoid service misuse and to confirm your identify.

Important: You should fully understand the limits of the AWS free tier to avoid being unduly charged for service usage

- Amazon will now verify your identity, by making an automated call to your mobile phone number and prompting you to enter the PIN number displayed on the screen.
- Once your identity is verified, choose the **Basic support plan** (also free) and confirm your account.

Note: At this point, make sure that you have subscribed a plan, even if you decide to register for the free tier or **Basic support plan**.

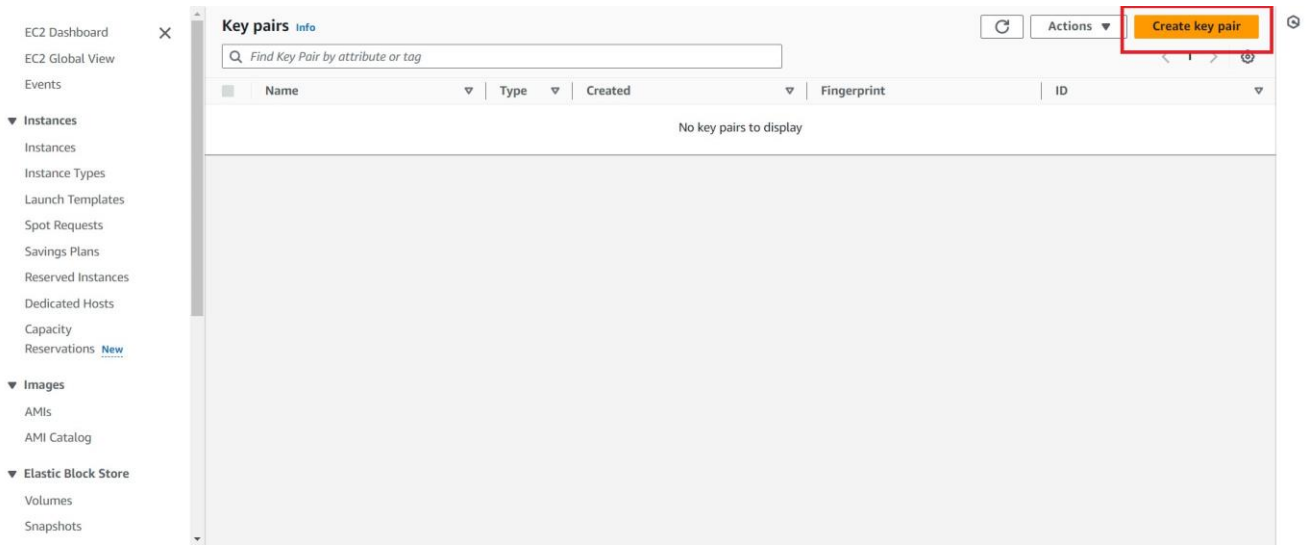
The AWS account registration machine will churn away for a minute or so, and you will then be redirected to a welcome page, which includes a link to the AWS management console. You should also receive an account confirmation email, which tells you that your account is good to go.

Step 2: Generate an AWS Key Pair

At the end of this step, you will have generated an SSH key pair to access your EC2 instances. If you already have an SSH key pair for the AWS region you are operating in, you can skip this step.

To generate an SSH key pair, which you will need to log in to your EC2 instances, follow the steps below:

- Log in to the **AWS Console**
- From the Amazon Web Services menu, Select the **EC2 Service.**
- Click on **key pairs** in the **EC2-Dashboard**

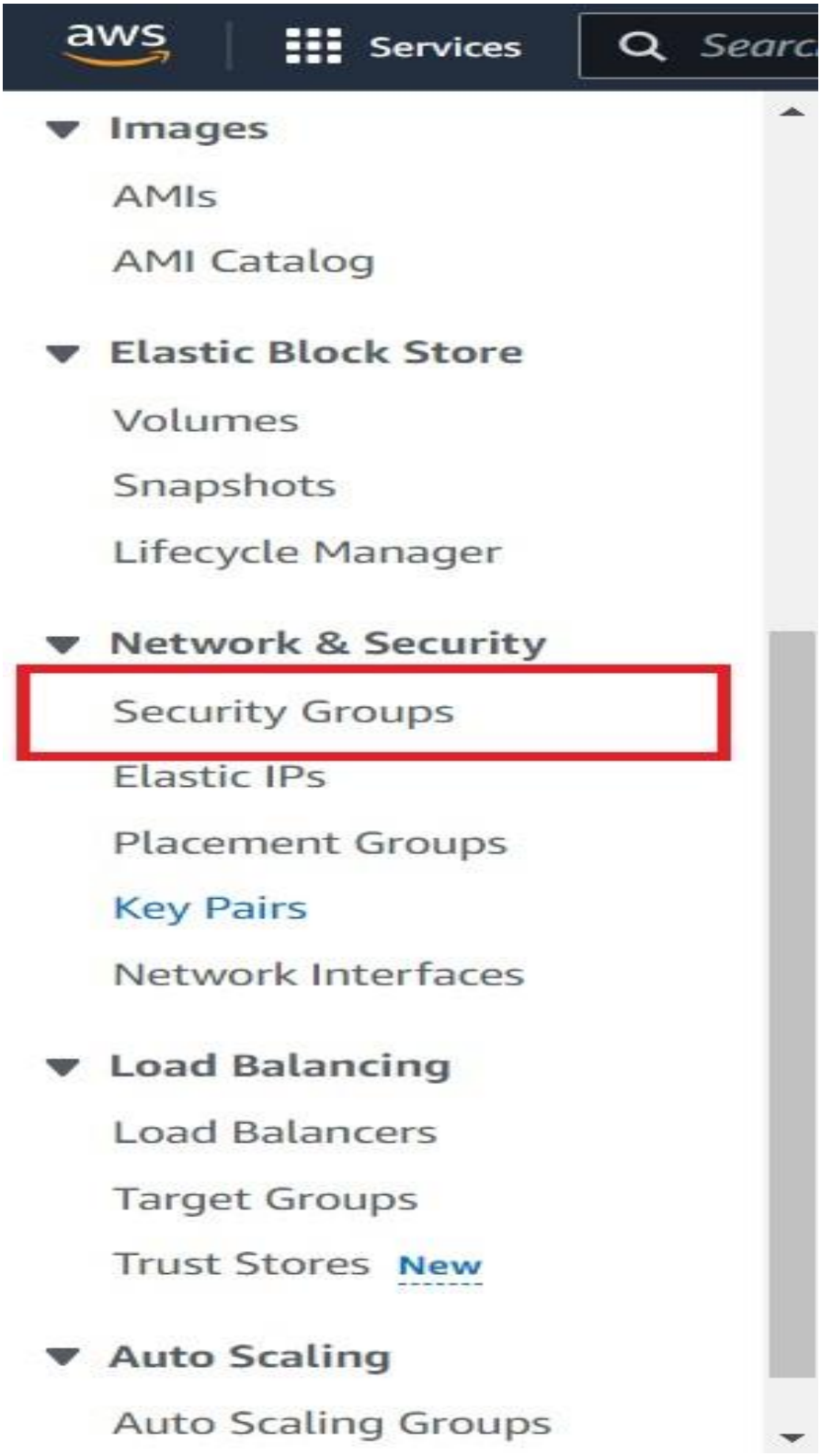


Step 3: Create an AWS Security Group

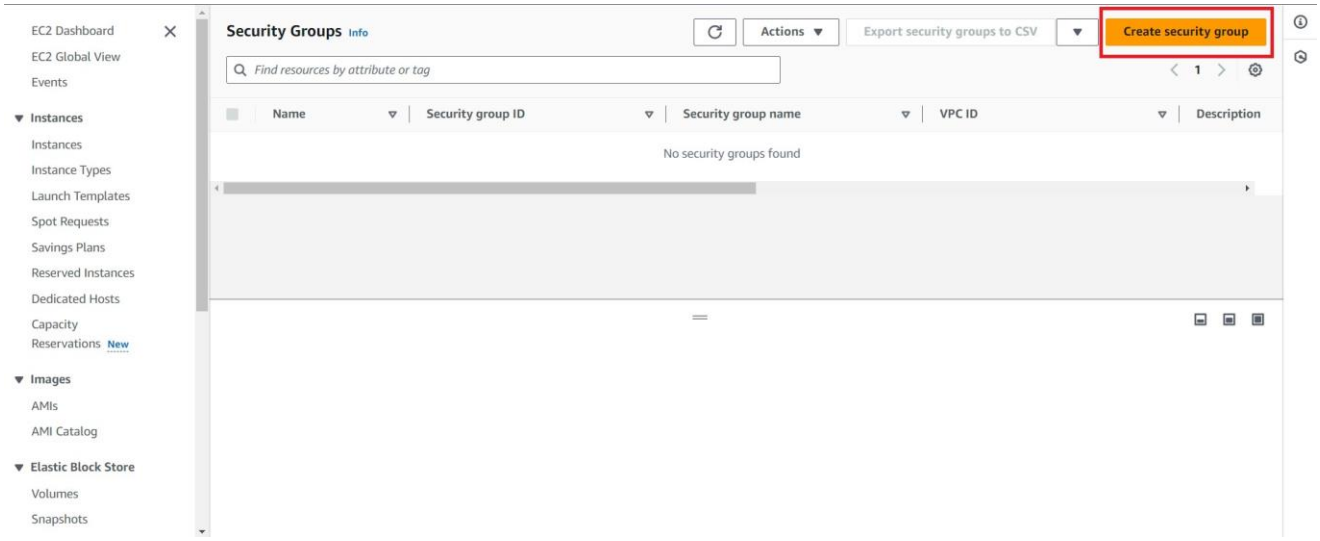
At the end of this step, you will have created an AWS security group for your cloud server.

By default, AWS cloud servers have their ports closed to secure them against external attacks. Since Data Sentinel is a Web application, it is necessary to open ports 80 for HTTP access, and port 22 for SSH access and port 8000 for backend of the application. To do this:

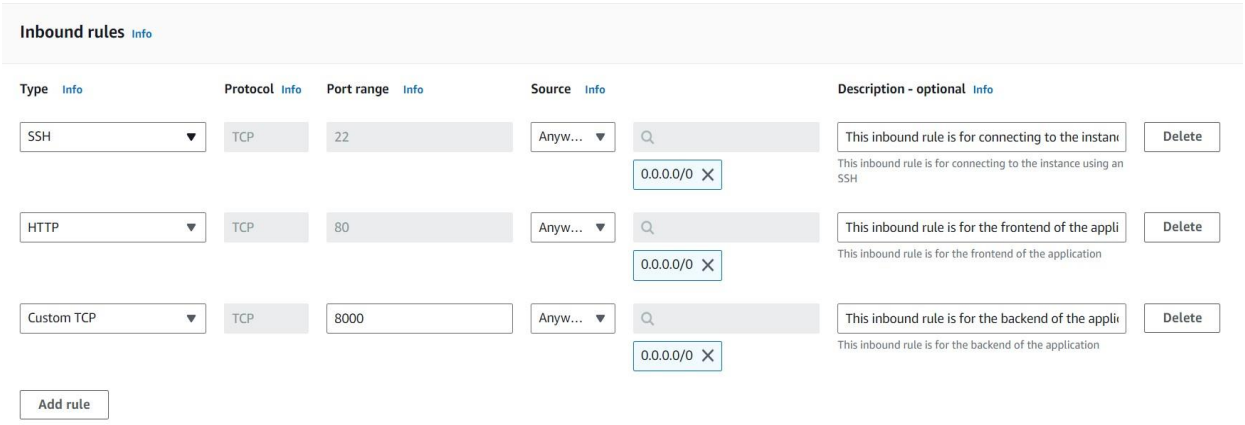
From the Amazon Web Services menu, select the EC2 service.



- From the Amazon EC2 dashboard, select the **Security Groups** option in the **Network & Security** menu.
- Click the **Create Security Group** button.



- In the resulting dialog box, enter a name and description for the new security group.
- Click the **Add Rule** button and add rules for HTTP, HTTPS and SSH and access using the following guidelines:
    1. **Type**: Use the pre-defined types **HTTP**, **SSH** and a **Custom TCP**
    2. **Source**: Use **Anywhere** to allow access from anywhere or use Custom IP and specify an IP address range.



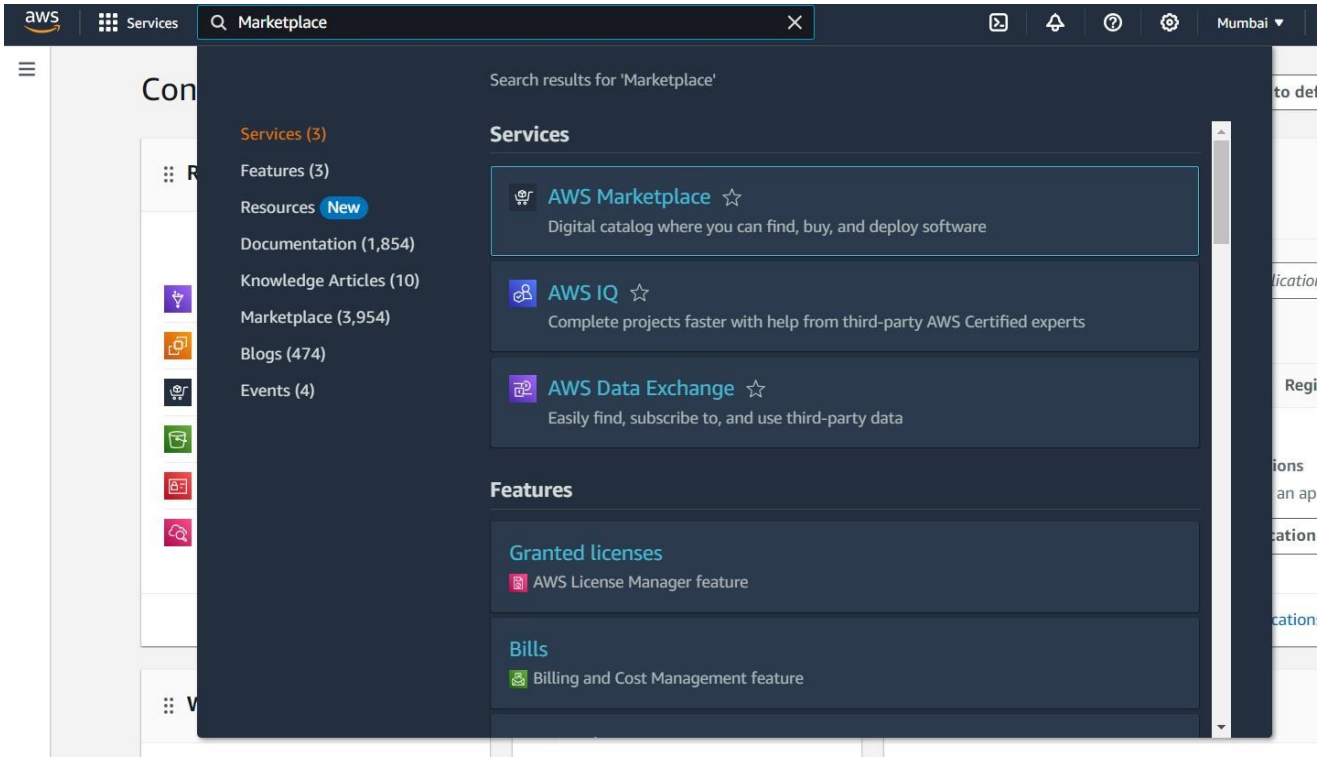
Click the **Create** button to your changes

Step 4: Deploy Data Sentinel on AWS Cloud Server

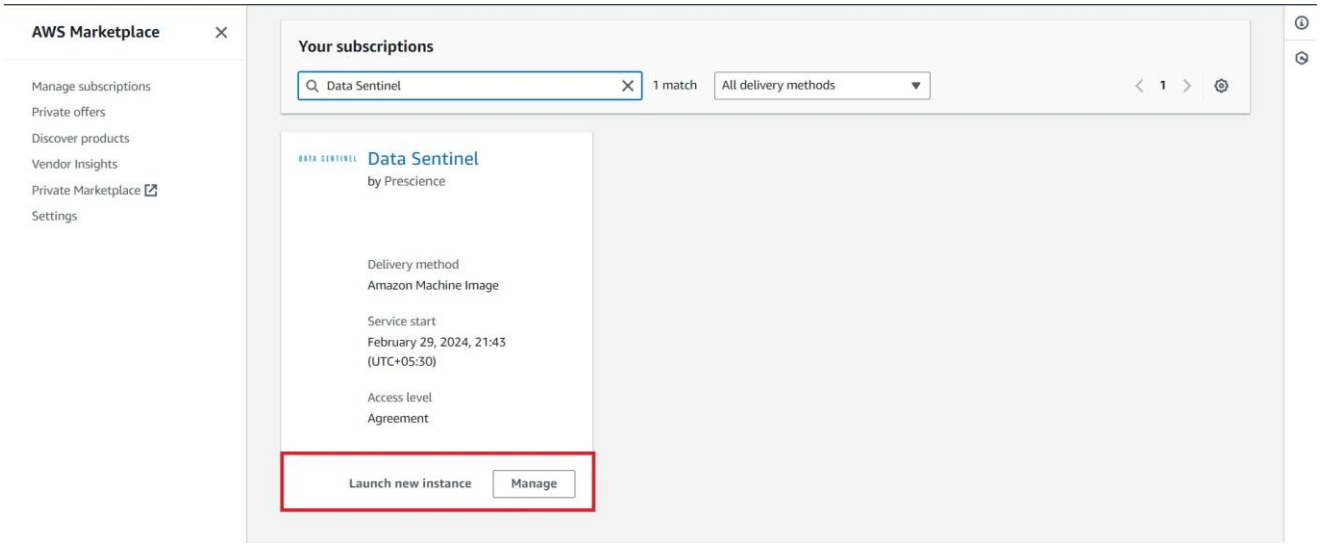
At the end of this step, your Data Sentinel application will be running on an AWS cloud server.

The next step is to launch a cloud server with the Prescience Data Sentinel Amazon Machine Image (AMI) running on it. The AWS Console lets you do this in just a couple of clicks. Follow these steps:

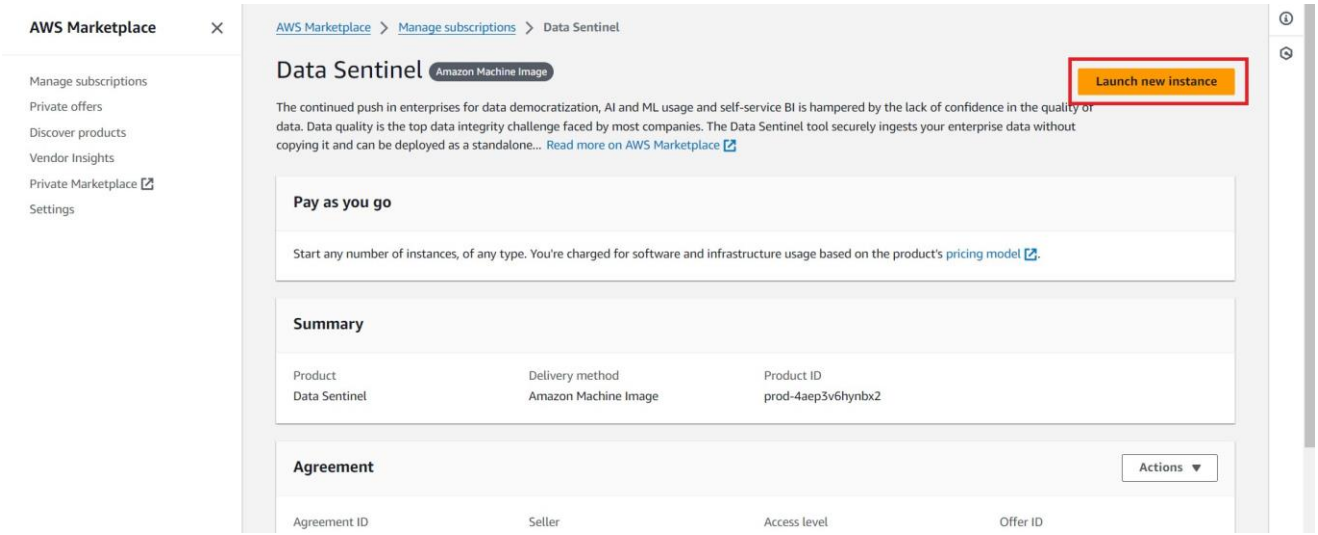
From the Amazon Web Services menu, search for AWS Marketplace



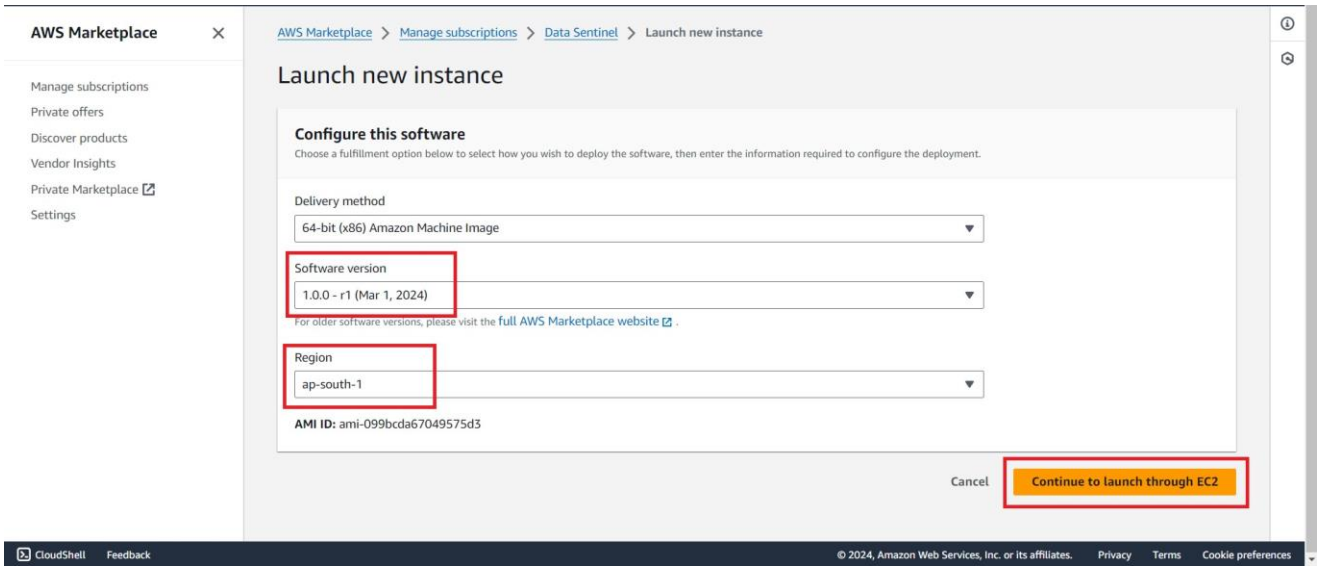
On the Marketplace page search **Data Sentinel**



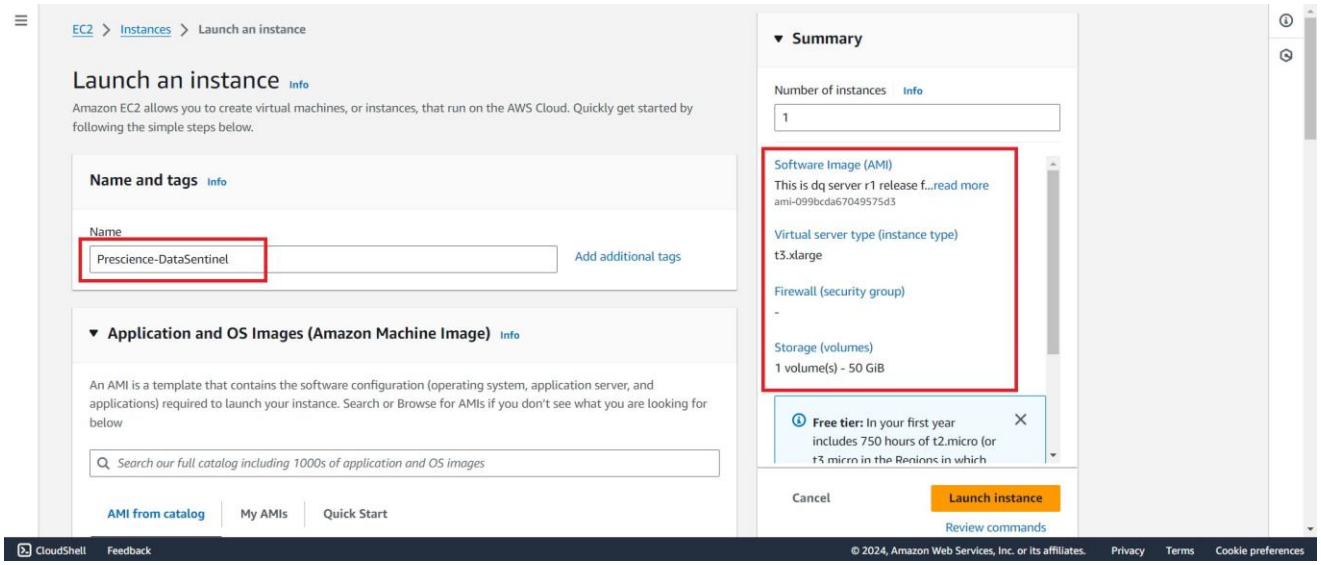
On the landing page click on **Launch new instance** button.



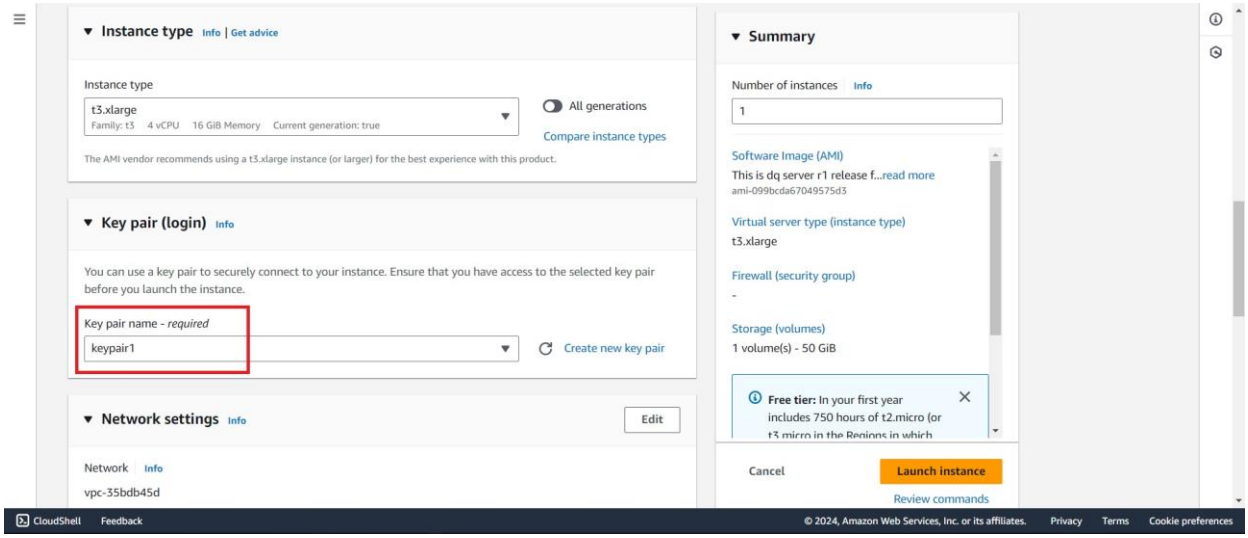
Select the latest version of Data Sentinel Released and select the desired region for setting up the EC2 instance.



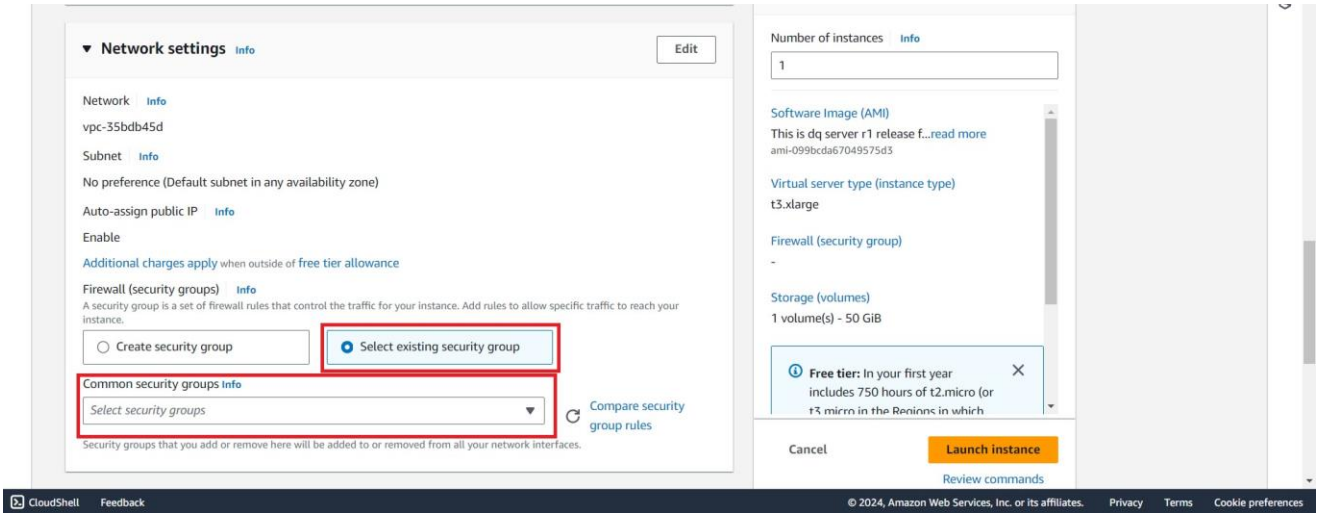
Give a name to EC2 instance of your choice.



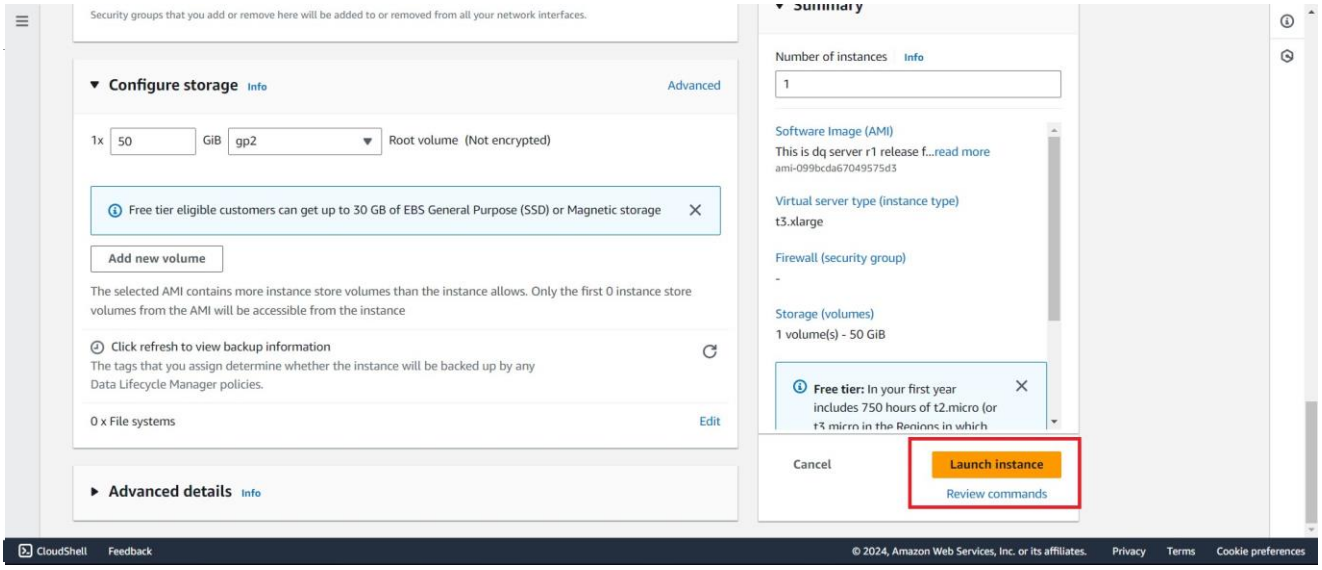
For the key pair (login) from the dropdown list select the previously created pair for this purpose.



Inside the **Network settings** select the **security group** created in **step 3.**



After reviewing and performing a final check on the settings click on **Launch Instance**



The process usually takes a few minutes, and you can use the EC2 Dashboard to check the status of the server. Once the server has deployed, you will be able to obtain its public IP address or public DNS name from the EC2 Dashboard to access the application. Application can be accessed using http://{PUBLIC_IP} or http://{PUBLIC_DNS_NAME} URLs.