**Lesson 4 Demo 9**

**Create a Glue Data Catalog**

**Objectives:** To set up the prerequisites and create a Data Catalog

**Tools require:** AWS workspace

**Prerequisites:** AWS account

**Steps to be followed:**

1. Create a VPC endpoint
2. Create a Data Glue Catalog

**Step** **1:** **Create a VPC endpoint**

1. Navigate to the AWS portal home screen, and searchfor and click on **VPC**
2. Here, click on **Endpoints**

Graphical user interface, text, application, website

Description automatically generated

1. In the **Endpoints** settings page, click on **Create endpoint**
2. Click on **Endpoint settings,** enter the name

Graphical user interface, text, application

Description automatically generated

1. A screenshot of a computer

   Description automatically generated Under **Services,** enter **s3** and choose **s3-global.accesspoint**
2. Under **VPC**,and choose the default option
3. In **Subnets,** click on the box next to **Availability Zone,** and then click on **Subnet ID**

Graphical user interface, application

Description automatically generated

1. Select **Group ID** under **Security groups**,and then click on **Create endpoint**

**Step** **2:** **Create a Data Glue Catalog**

1. Navigate to the AWS portal home screen, and search and click on **AWS Glue**

**Text

Description automatically generated**

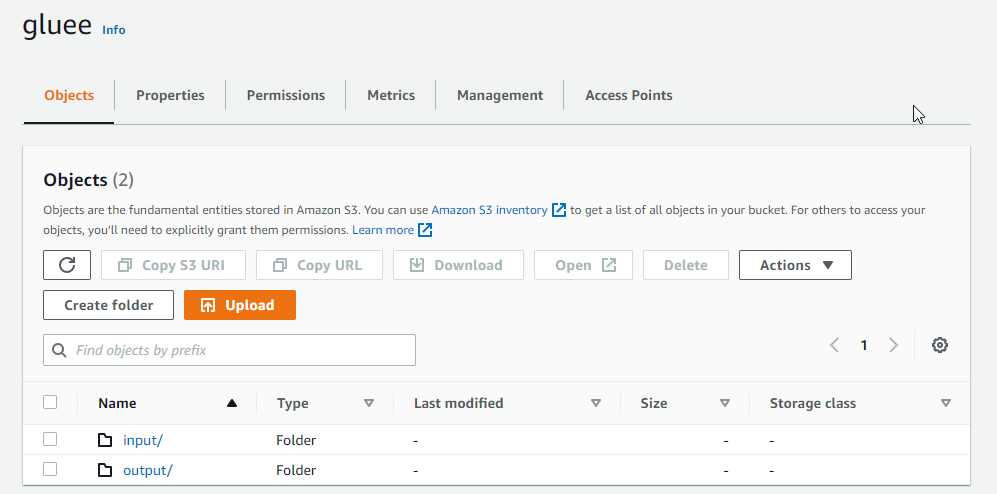
1. Click on the **Crawlers** option under **Data Catalog**

Graphical user interface, text, application, email

Description automatically generated

1. Click on **Create crawler** as per the **new console**

**Note:** Please make sure an S3 bucket with 2 folders called input and output are created where in add a particular file to the input folder as shown in the below screenshots



Graphical user interface, text, application, email

Description automatically generated

1. Enter the **Crawler details** as shown in the below screenshot:

Graphical user interface, text, application, email

Description automatically generated

1. Select **Data stores** and **Crawl all folders,** and click on **Next**

Graphical user interface, text, application, email

Description automatically generated

1. Click on **Add connection**

Graphical user interface, text

Description automatically generated

1. Enter name, choose the default VPC under **VPC**

Graphical user interface, application

Description automatically generated

1. Choose a subnet

Graphical user interface, text, application

Description automatically generated

1. Select the box next to **Group ID**, and click on **Add**

2.10 Select the option created in the previous steps under **Connection**

A screenshot of a computer

Description automatically generated

2.10 Select the folder by adding the path and click on **Add an S3 data source**

Graphical user interface, text, application, email

Description automatically generated

1. Under **Data source configuration**, click on **Next**

Graphical user interface, text, application, email

Description automatically generated

1. Create a **Database**, and under **Frequency** select on demand and click on **Next**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated with medium confidence

1. Select **Create an IAM role**, enter a name under **AWSGlueServiceRole**,and click on **Next**

Graphical user interface, text, application, email, Teams

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Once the **IAM role** is created click on Next and click on **Create crawler**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. The **crawler** is created successfully. Click on the Run option to start the crawler

Graphical user interface, text, application, email

Description automatically generated

1. The **crawler** is running successfully, and the duration can visible as shown in the below screenshot:

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated