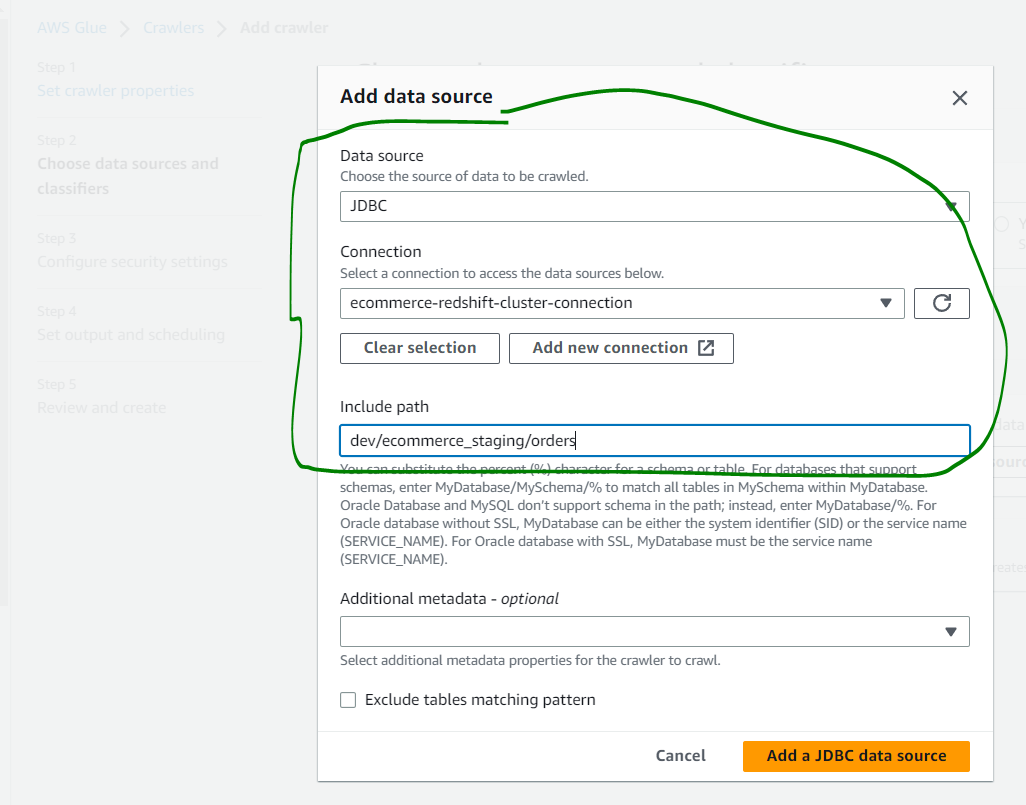
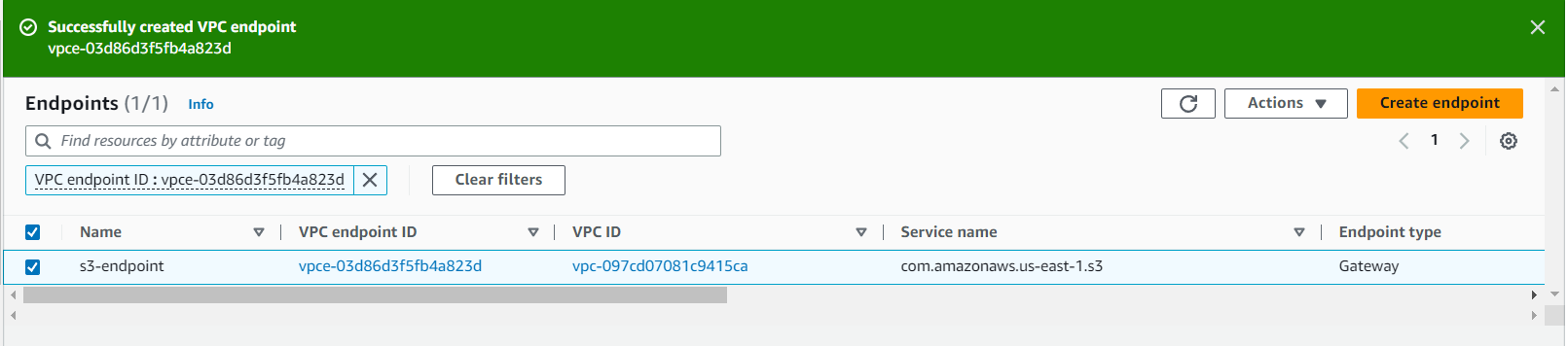
Create a Glue Crawler on the Redshift ecommerce\_staging.orders table using the Connection created in the previous step. The Glue Data Catalog table that will be created by the Crawler will be used in the glue Spark ETL job that will be created next, to move data from S3 staging area into Redshift staging area.



The Glue Crawler was failing with the below error:

VPC S3 endpoint validation failed for SubnetId: subnet-006d417e0769d0d1f. VPC: vpc-097cd07081c9415ca. Reason: Could not find S3 endpoint or NAT gateway for subnetId: subnet-006d417e0769d0d1f in Vpc vpc-097cd07081c9415ca (Service: AWSGlueJobExecutor; Status Code: 400; Error Code: InvalidInputException; Request ID: 3351b4df-6422-4bc3-a84a-de9c54cd7359; Proxy: null)

Created an S3 endpoint in the VPC/subnet of the Redshift cluster as shown below, to resolve this:



Verify that a Glue Data Catalog table is created corresponding to the Redshift staging table once the Crawler has completed successfully.

