**Converting CSV files to Parquet files using Glue**

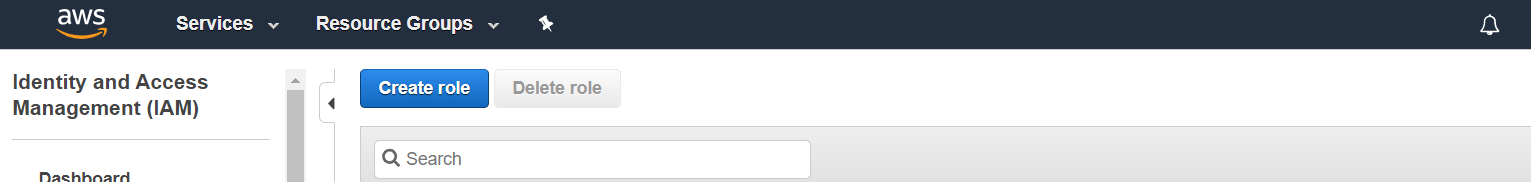
**Pre-requisites:**

* IAM Role
* S3 Buckets
* Glue Crawler
* Glue ETL Job

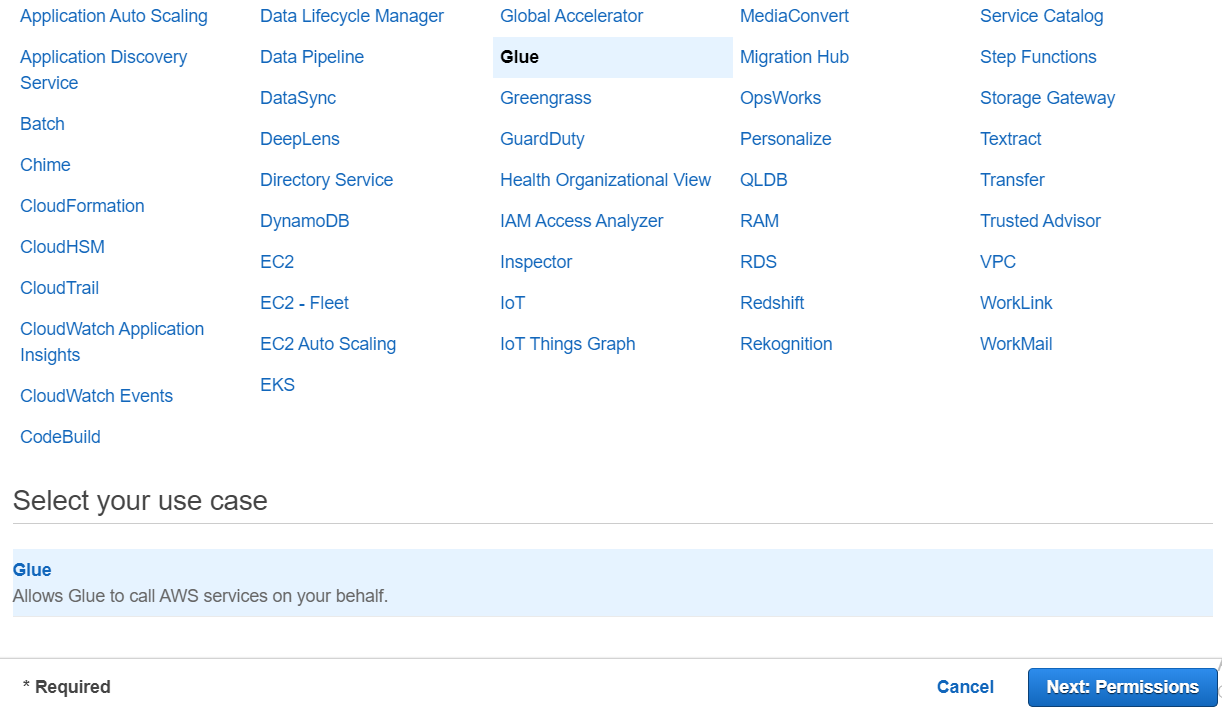
**Step:1 🡪 Creating IAM Role:**

Here we need to create Role for Glue to access Data from S3 Bucket

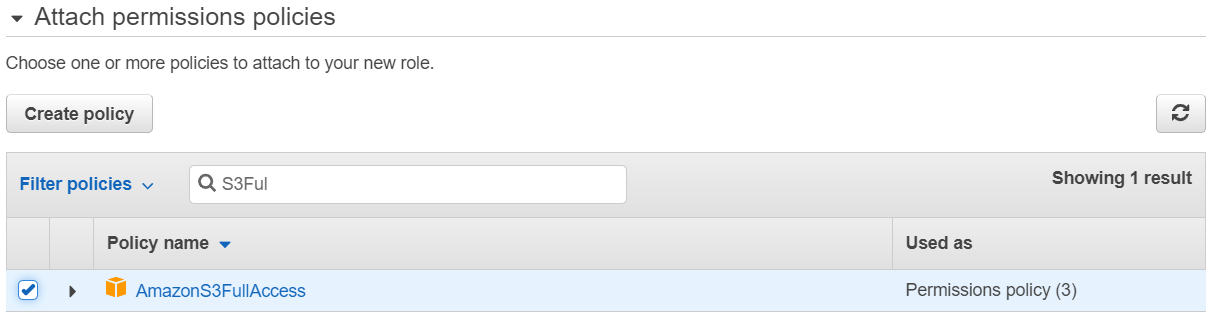
Goto AWS IAM Service and Click on Role

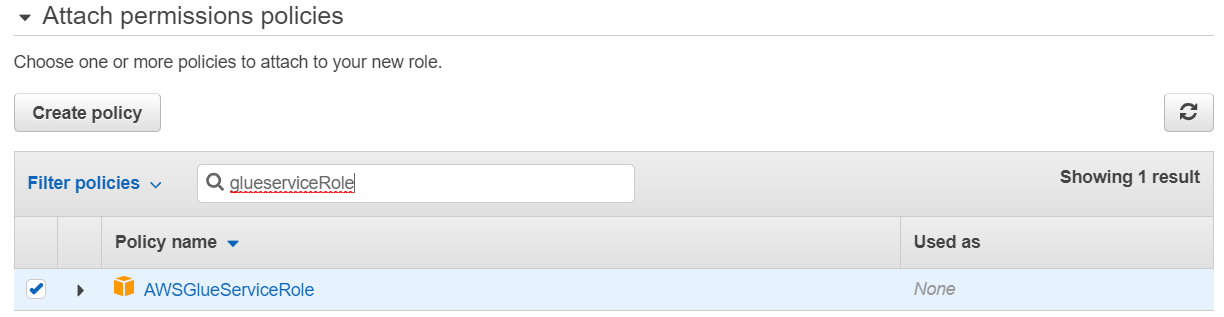


Click on **Create role**

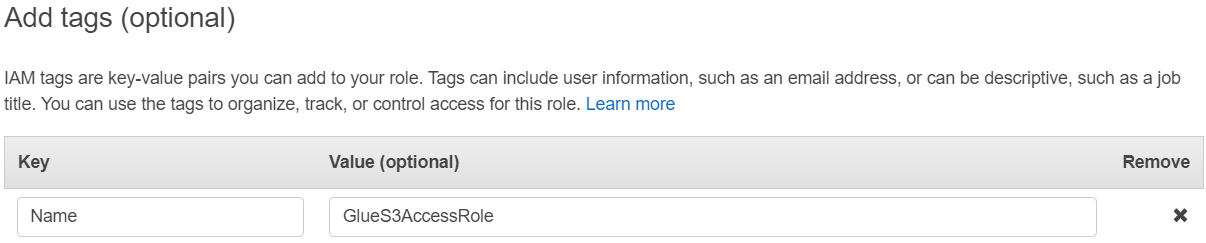


* Select **Glue** and Click on **Next**

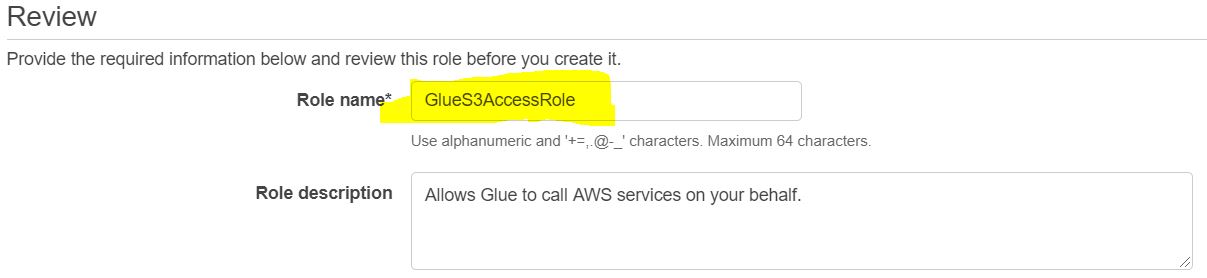




* Select **AmazonS3FullAccess** and **AWSGlueServiceRole** and Click on **Next**



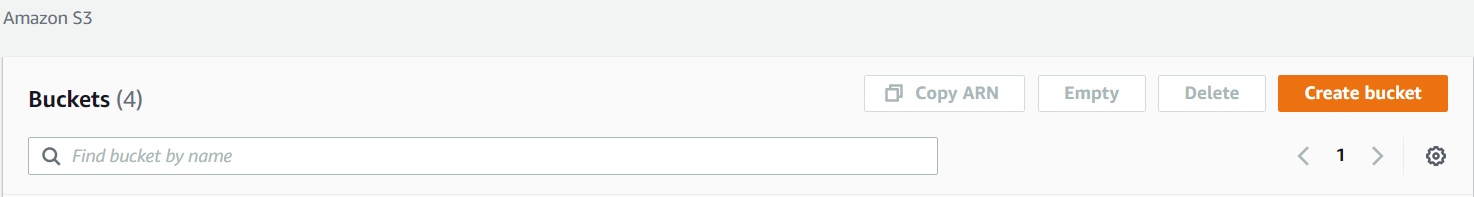
* Click on **Next**



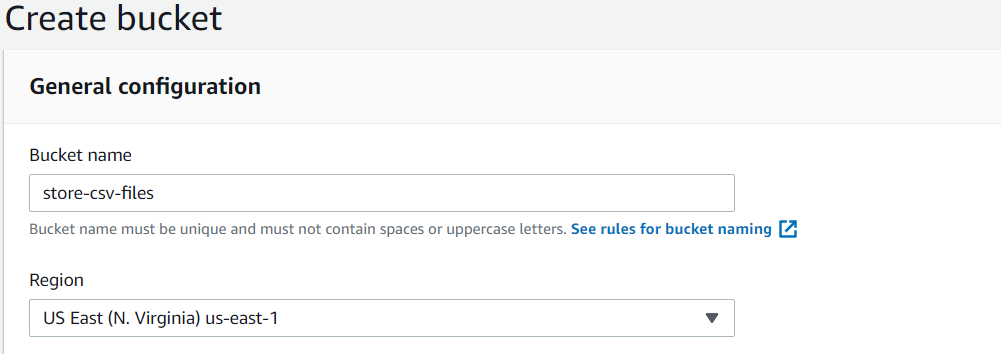
* Give a Name for Role and click on **Create Role**

**Step:2 🡪 Create S3 Buckets for storing CSV files and Storing Parquet files**

Goto AWS S3 Service:

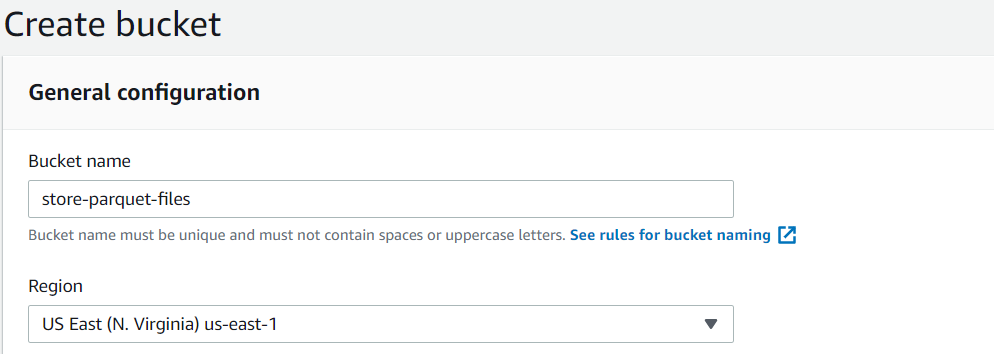


* Click on **Create bucket**

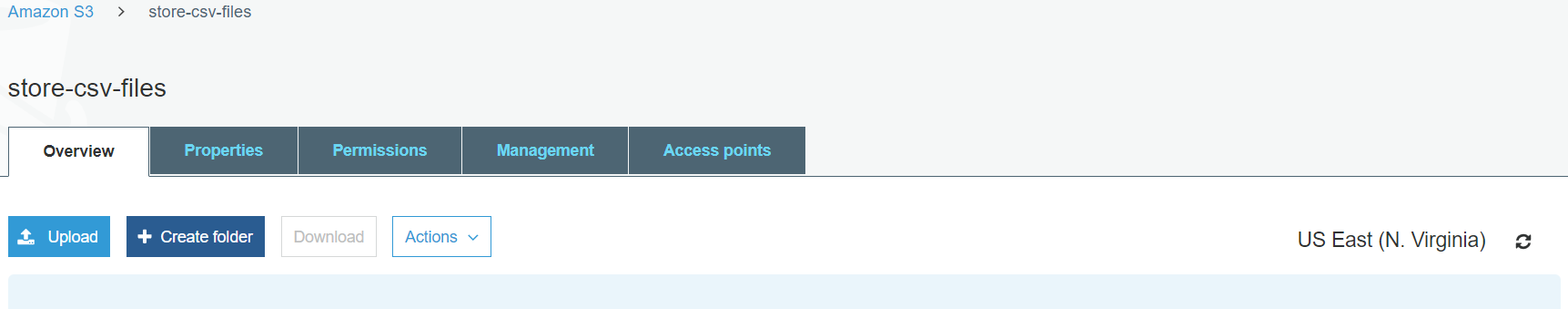


Give a name for S3 Bucket and Click on **Create bucket**

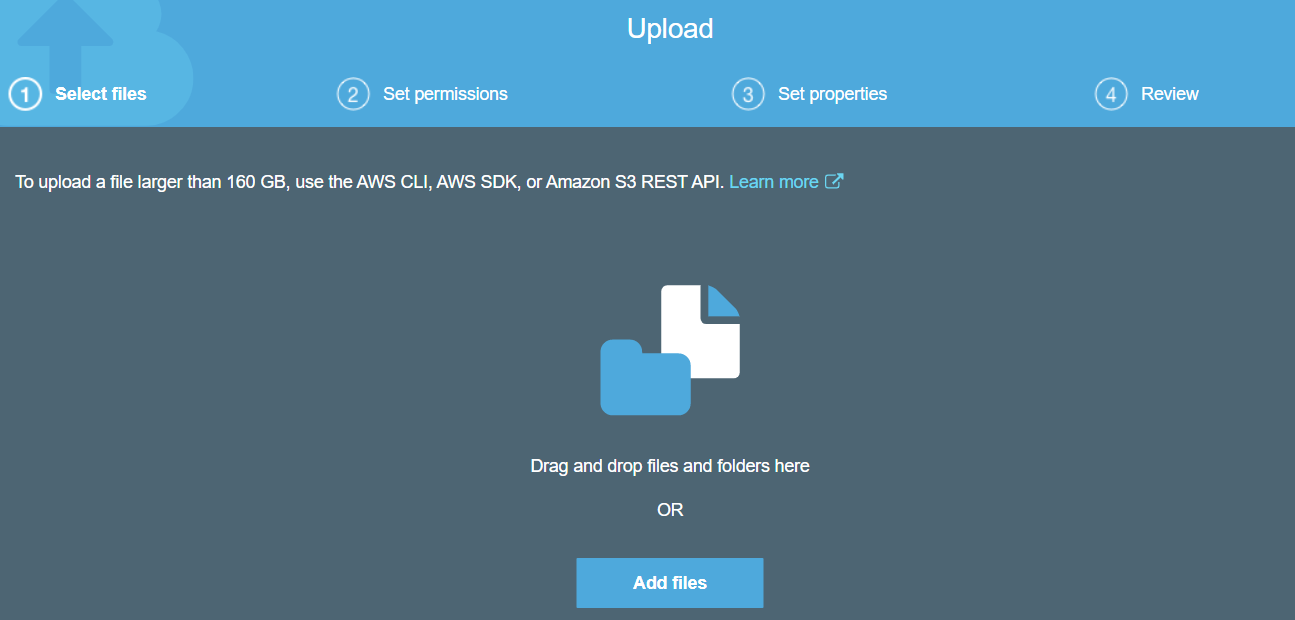
Create Another Bucket to store parquet files



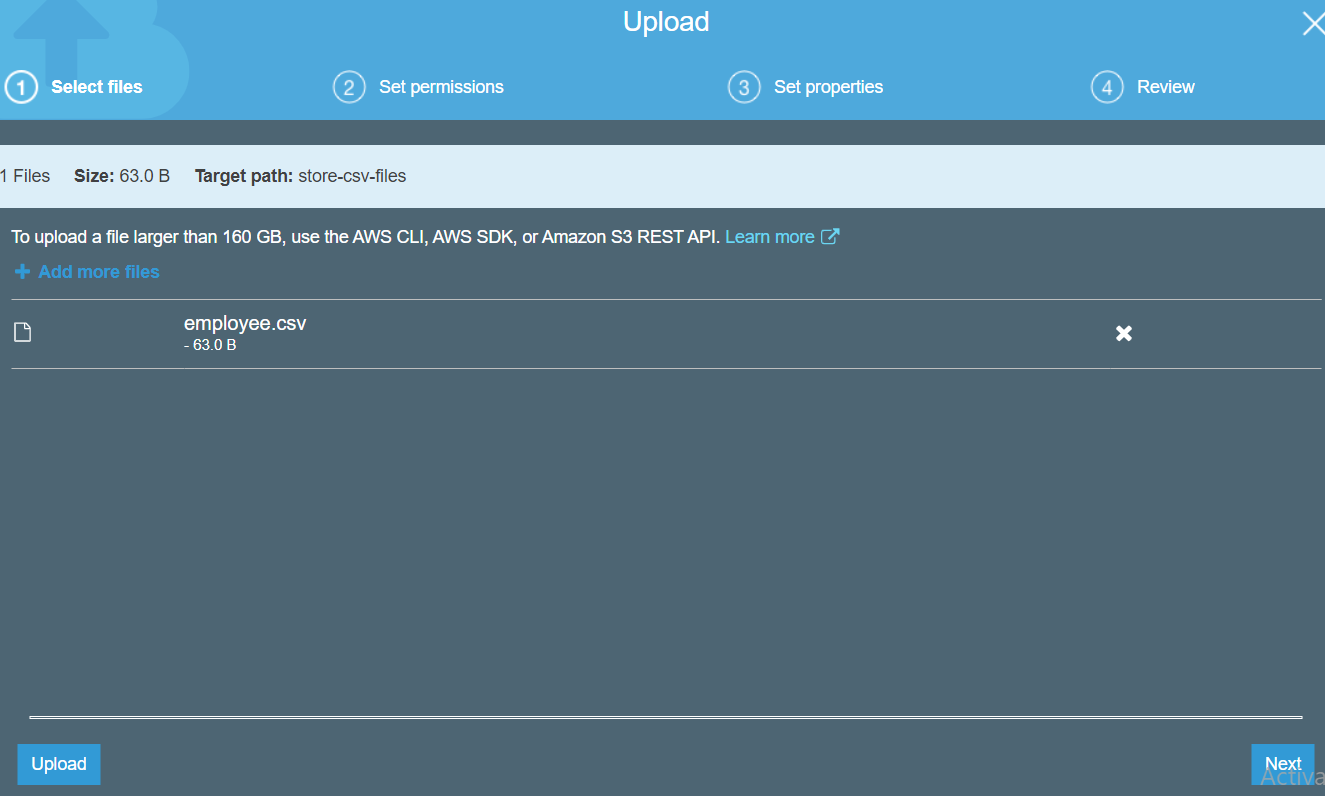
* Upload CSV file
* Click on store-csv-files bucket



* Click on **Upload**



* Click on **Add files**

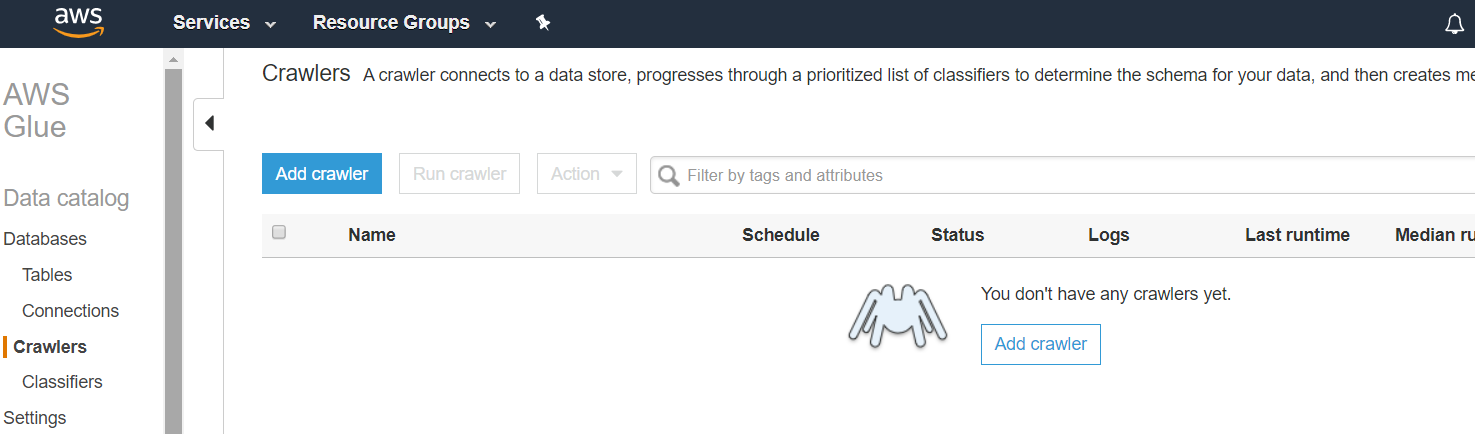
****

Select file and click on **Upload**

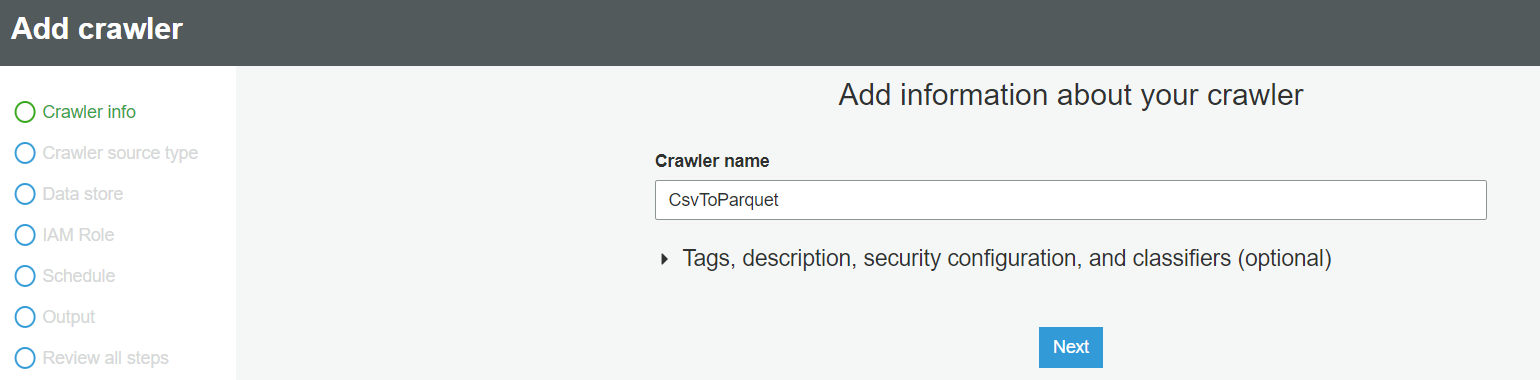
* Click on **Create bucket**

**Step:3 🡪 Create Glue Crawler**

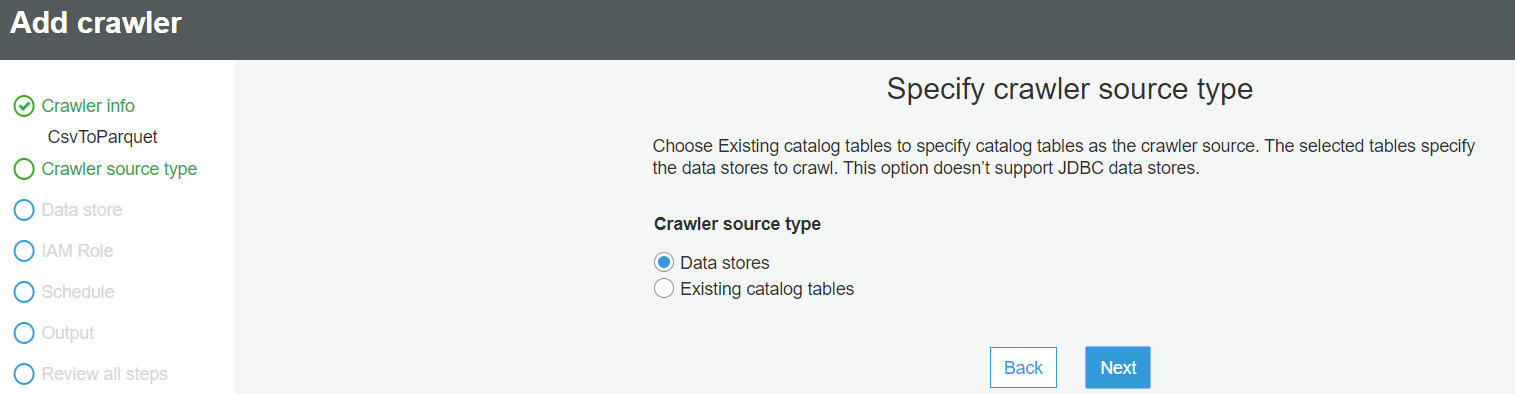
AWS Glue Service:



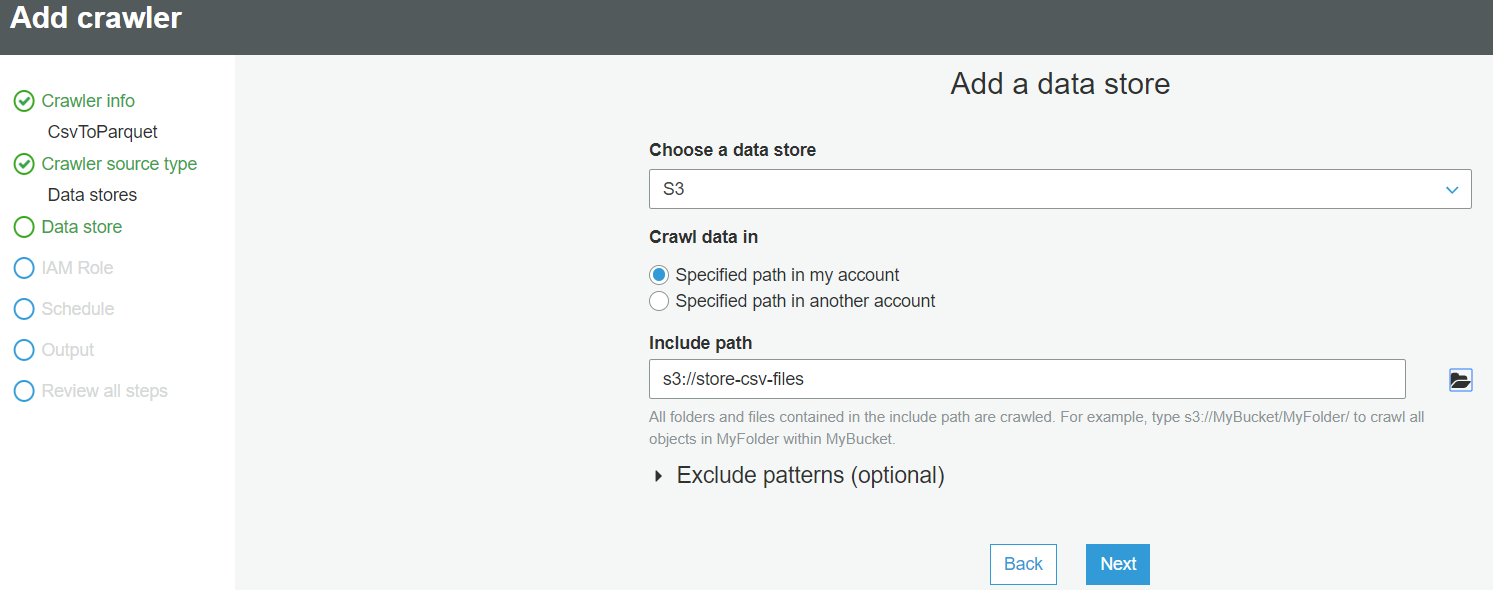
* Click on **Add crawler**



* Click on **Next**

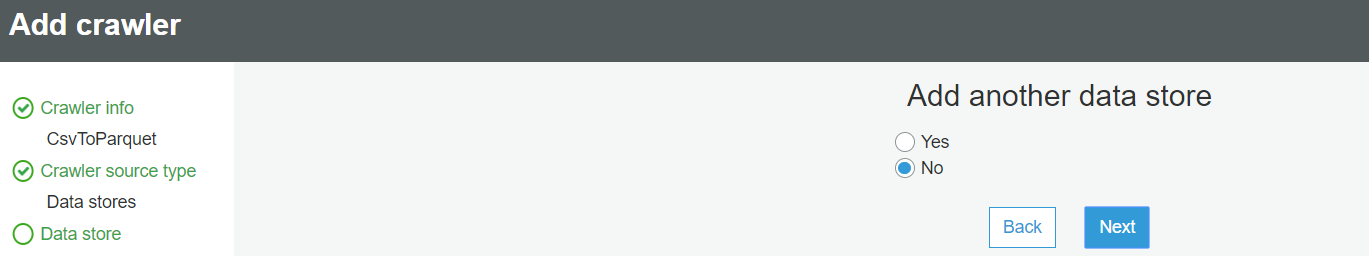


* Click on **Next**

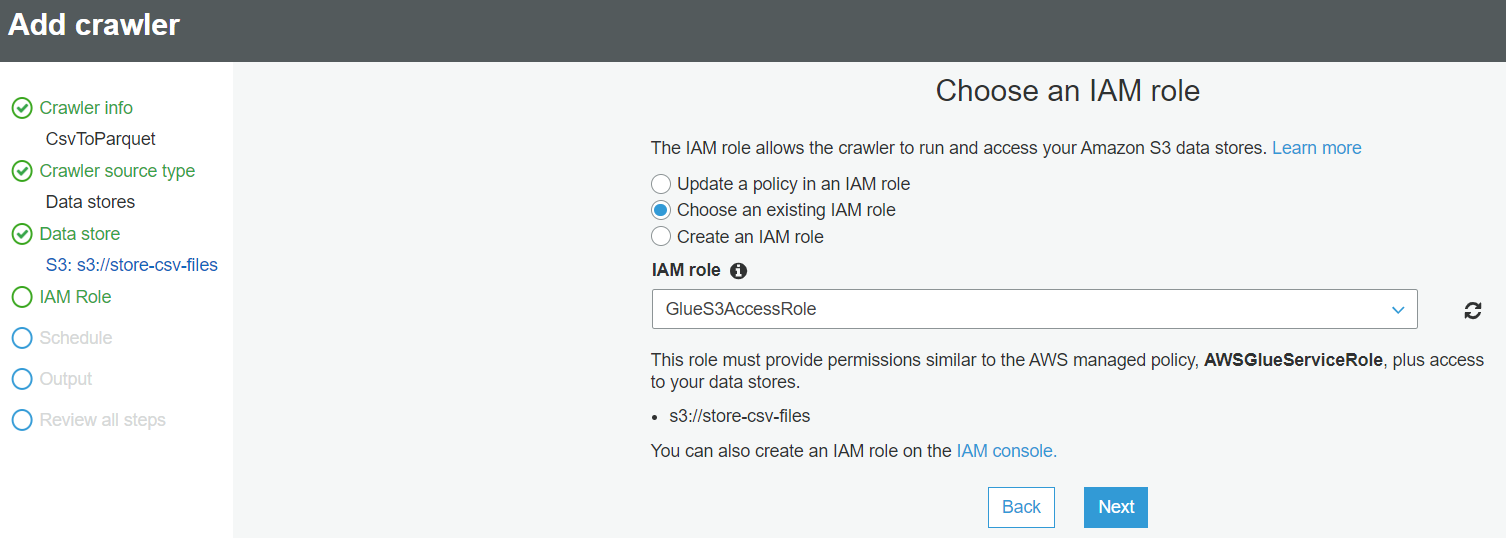


Select S3 bucket where we stored CSV files

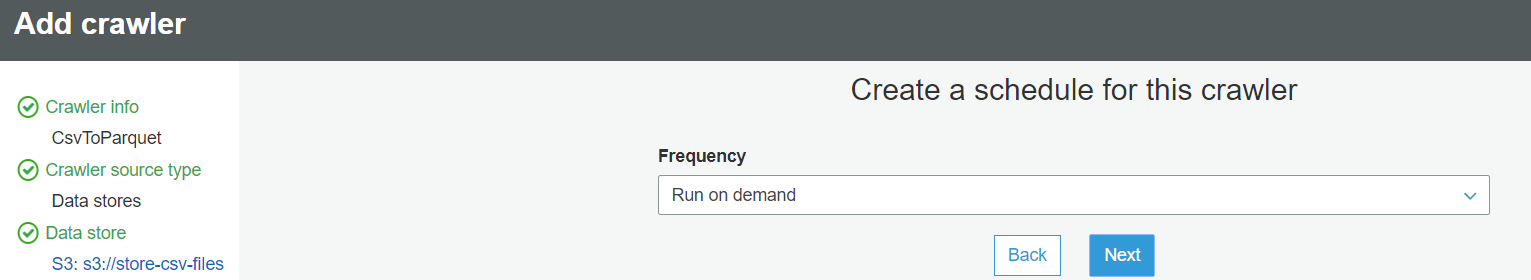
* Click on **Next**



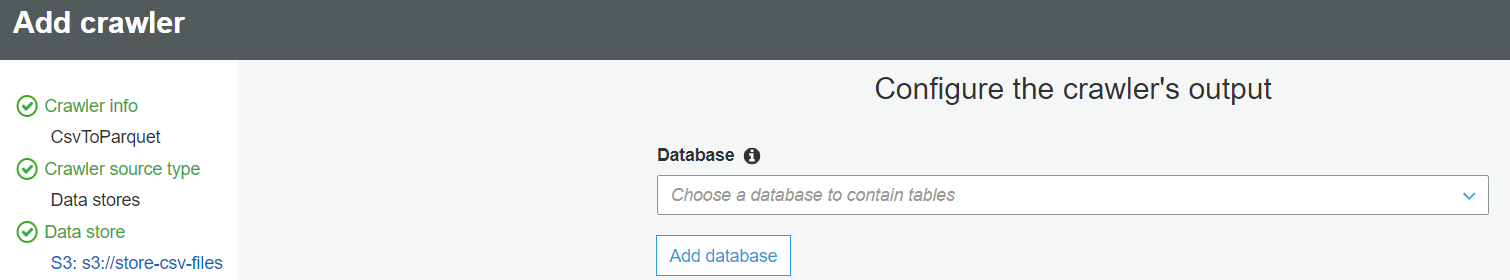
* Click on **Next**



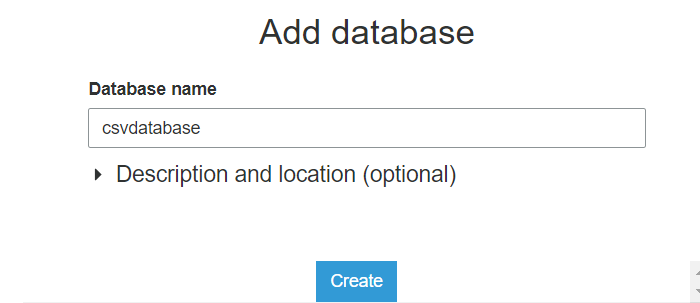
* Choose existing role which we create
* Click on **Next**



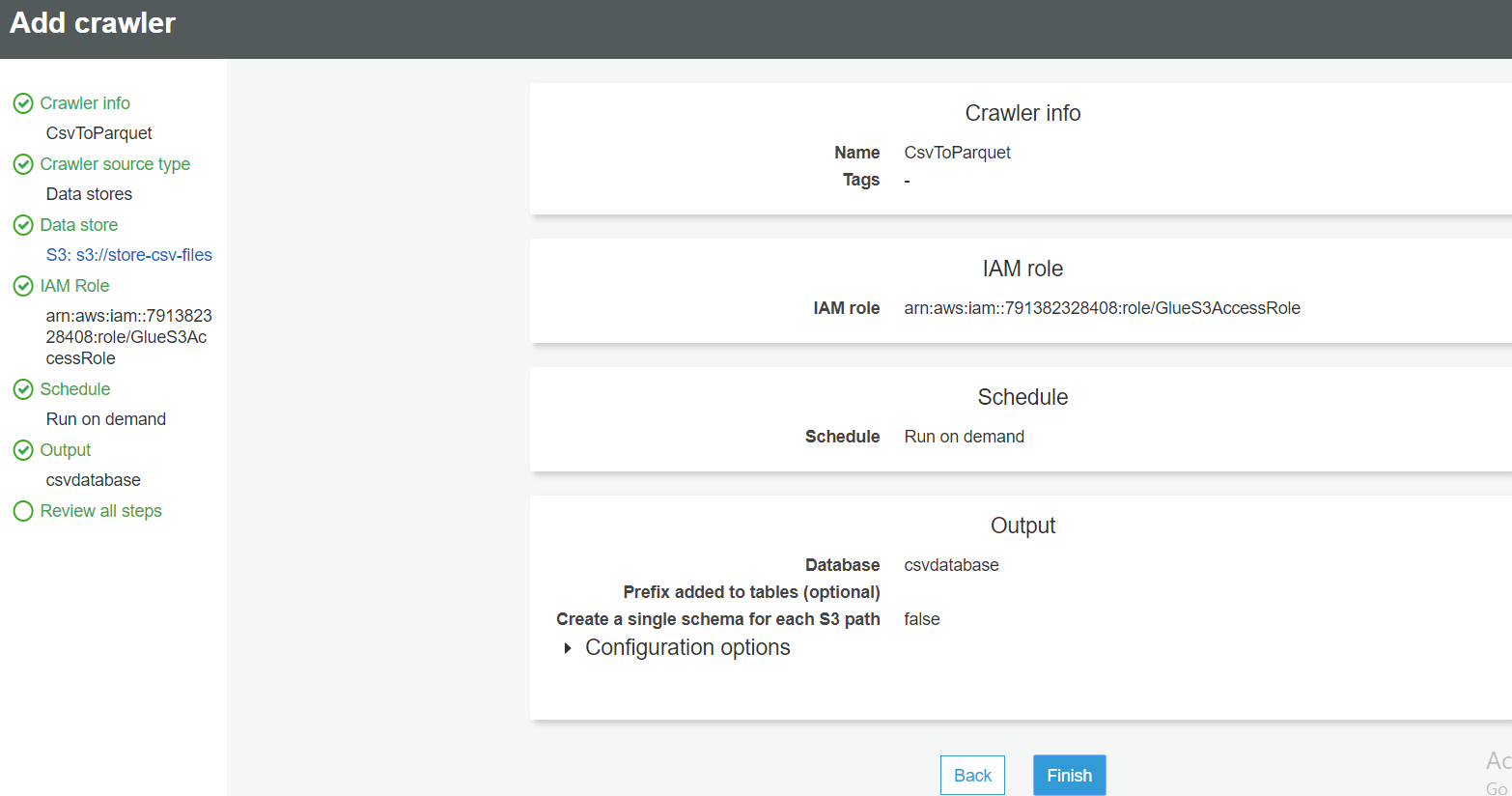
* Click on **Next**



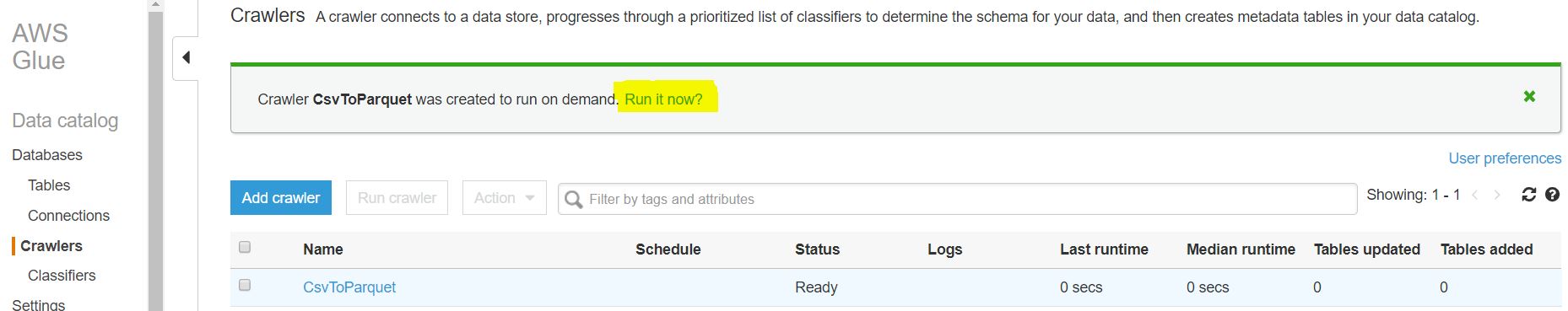
* Click on **Add database**(If not exit)



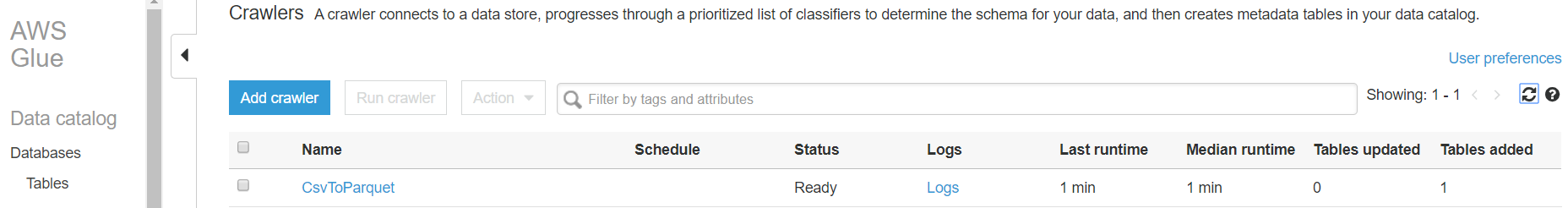
* Give some name and Click on **Create**
* Click on **Next**



* Click on **Finish**

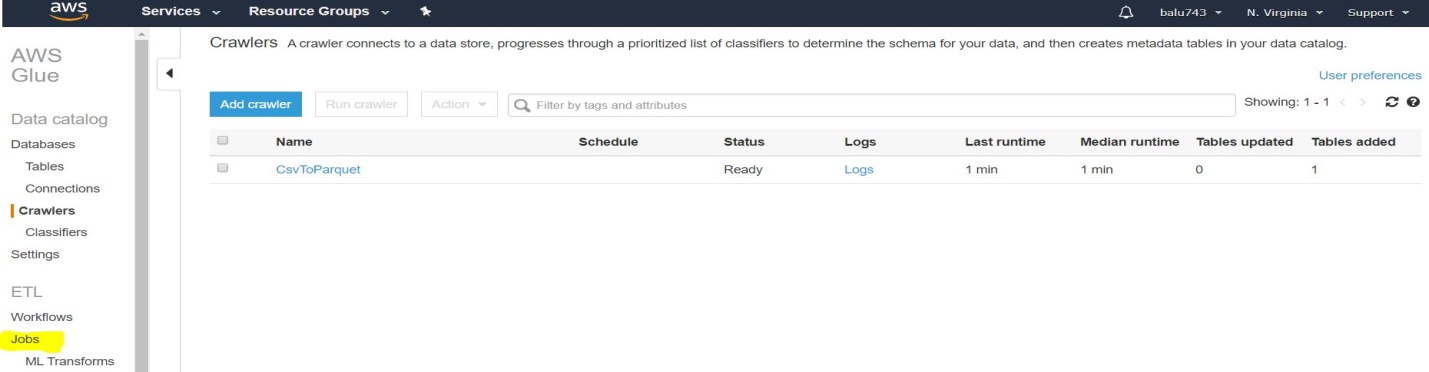


* Click on **Run it now**

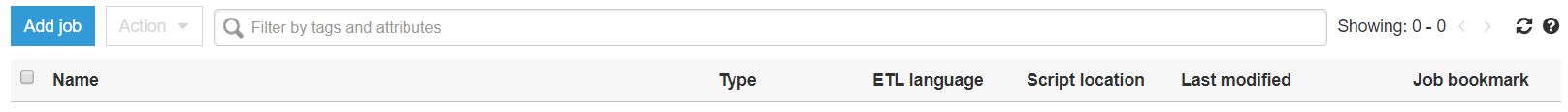


**Step:4 🡪 Create Glue ETL Job**

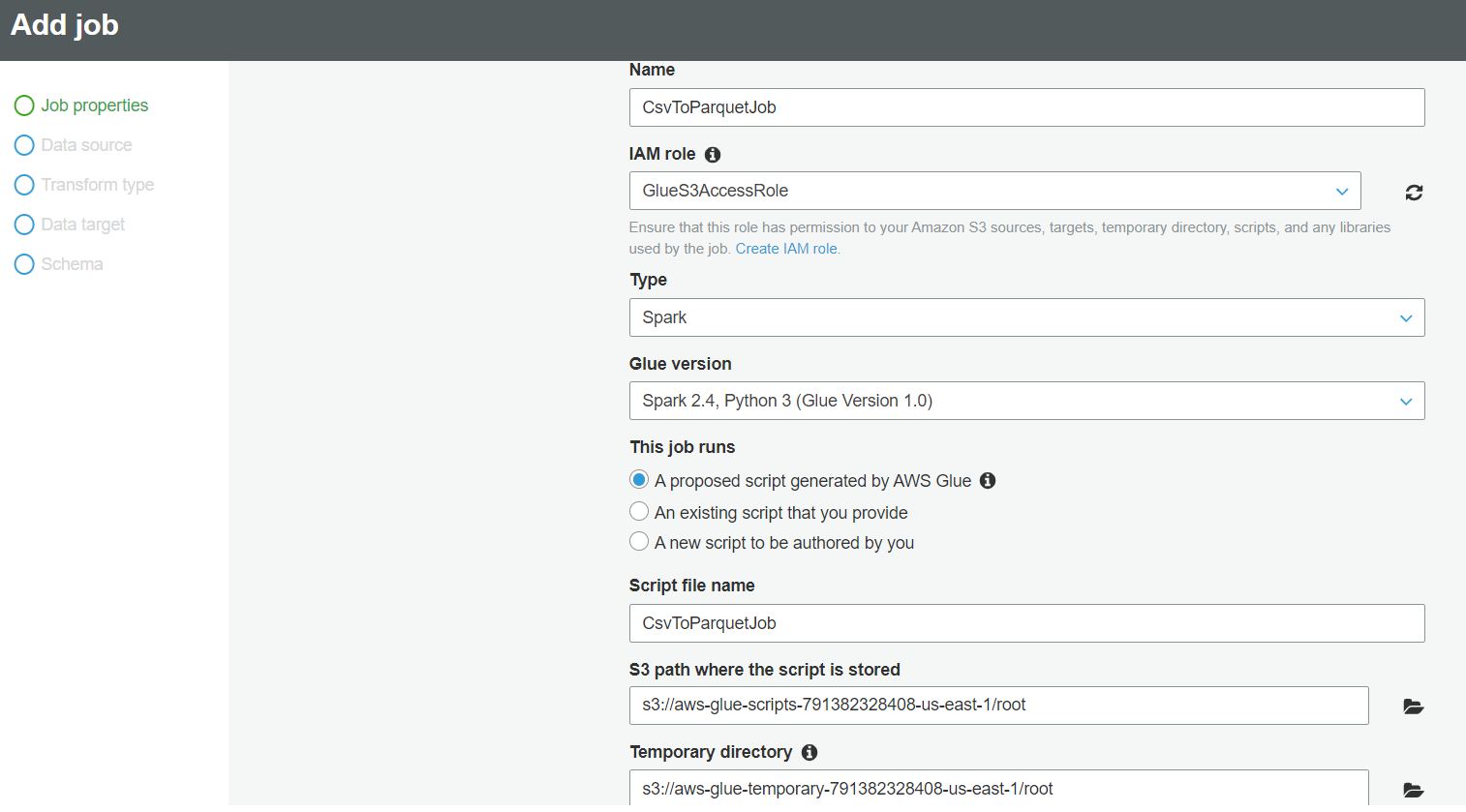
Goto AWS Glue Services



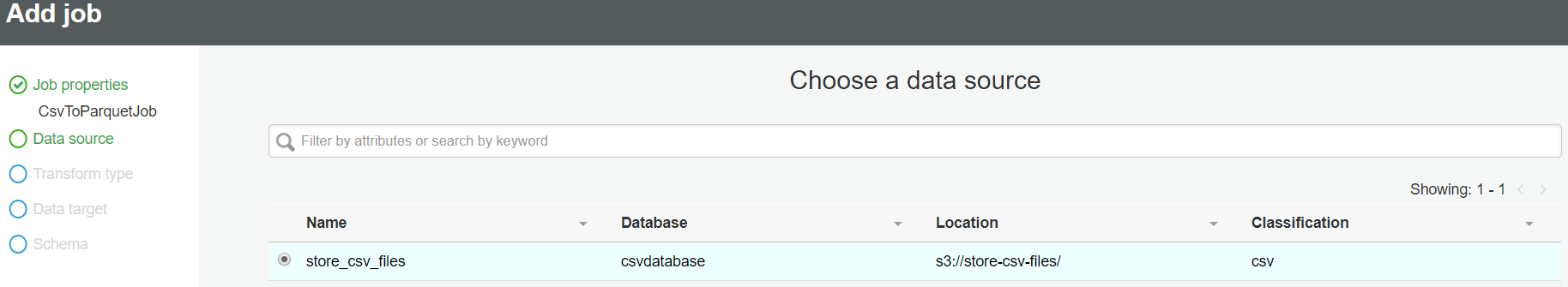
* Click on **Jobs**



* Click on **Add job**

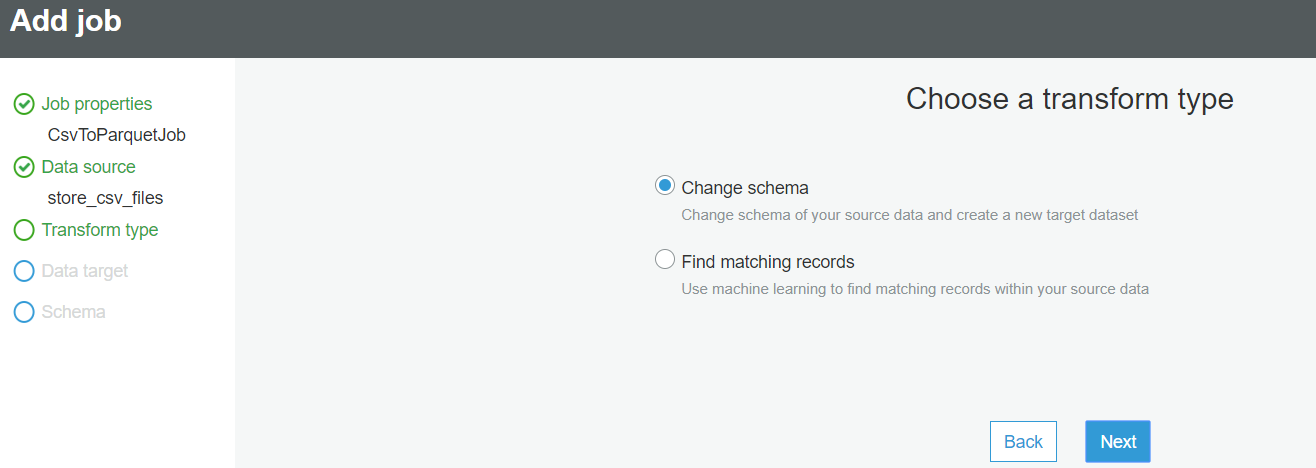


* Click on **Next**

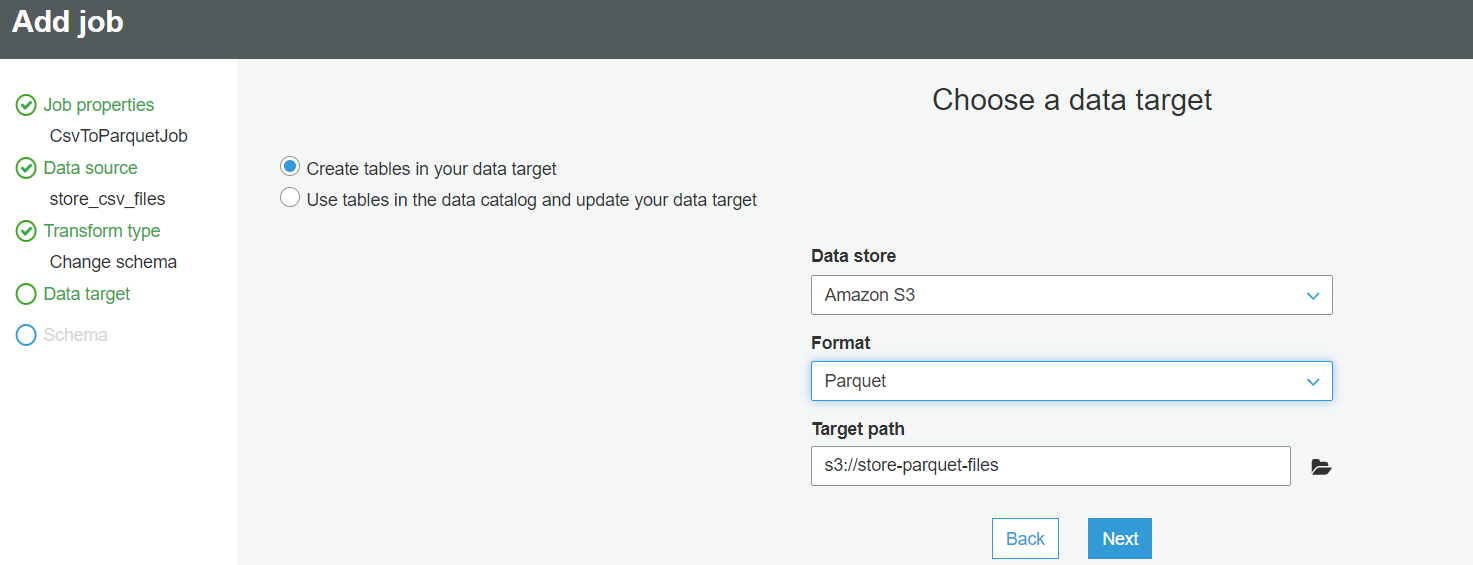


Select data source which we created database in crawlers

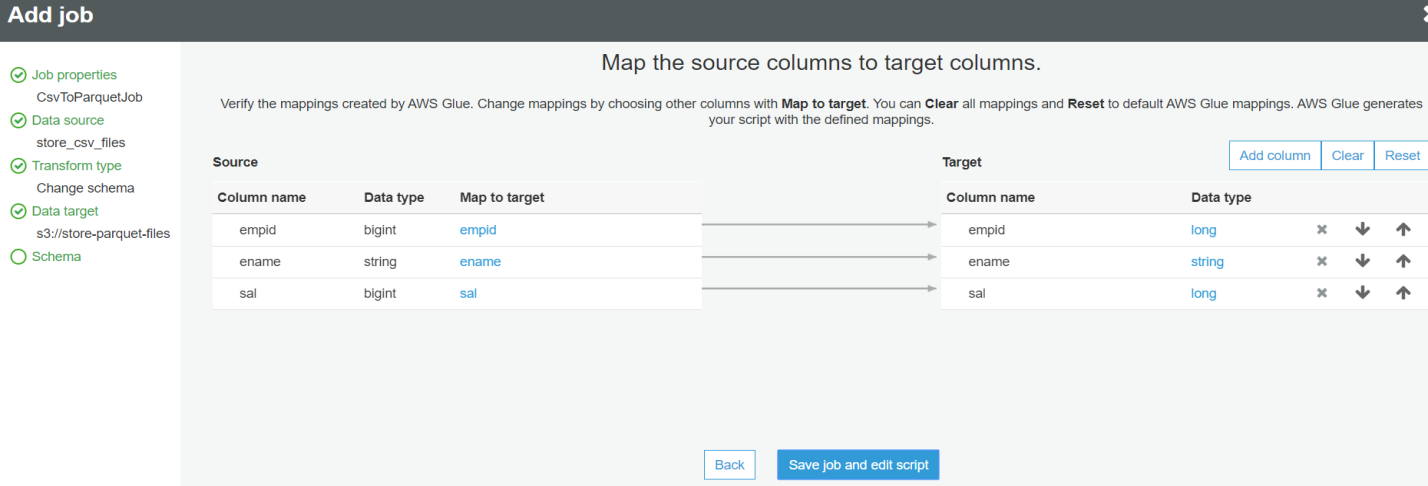
* Click on **Next**



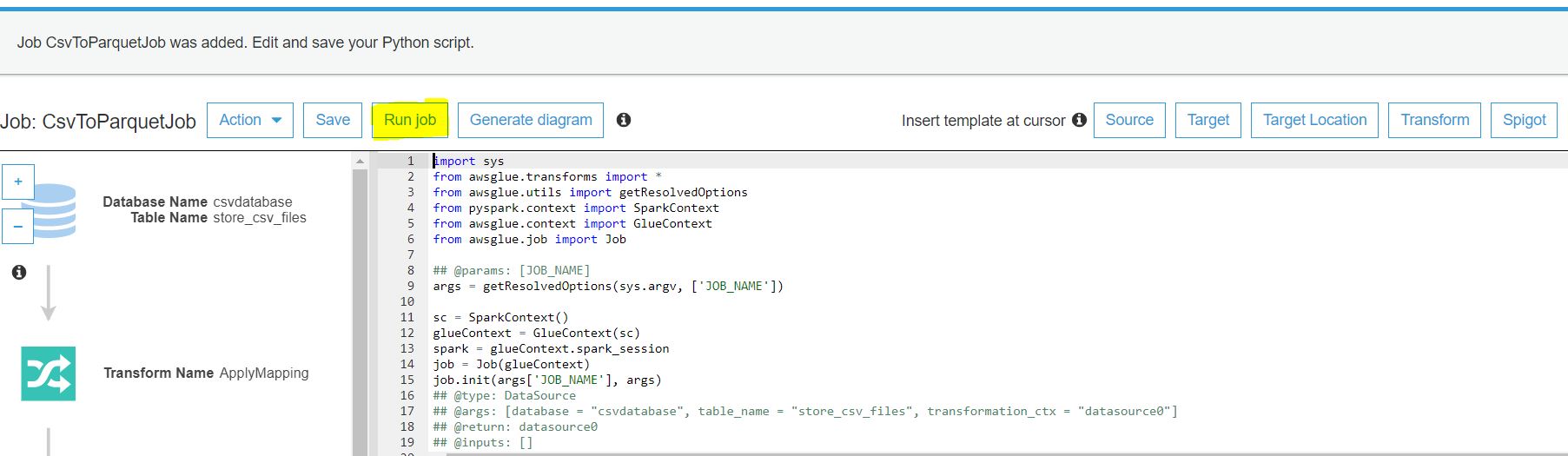
* Click on **Next**



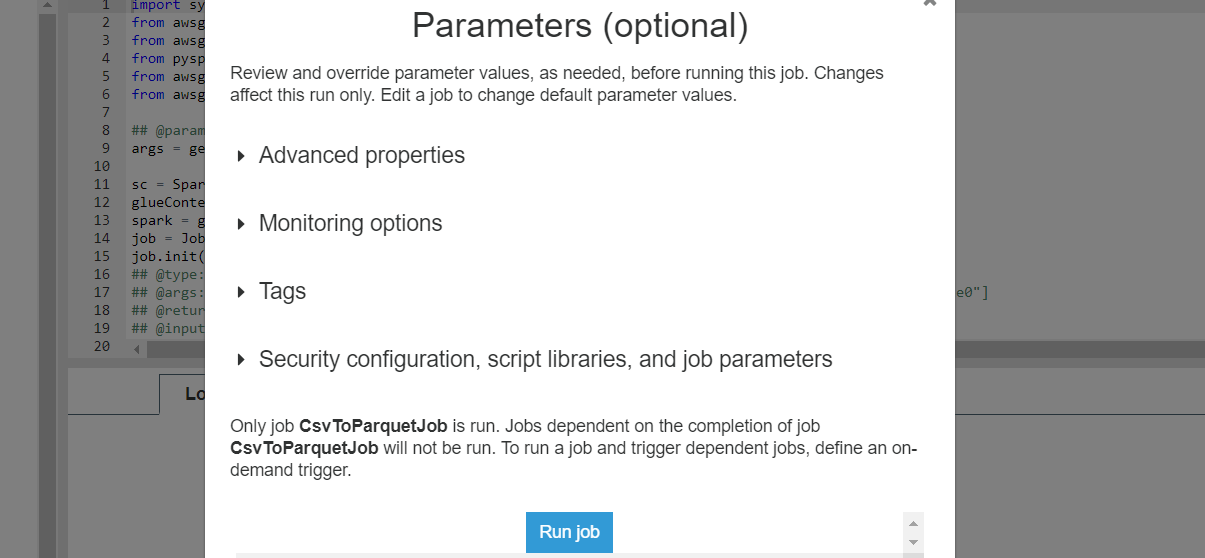
* Choose a data target as above
* Click on **Next**

****

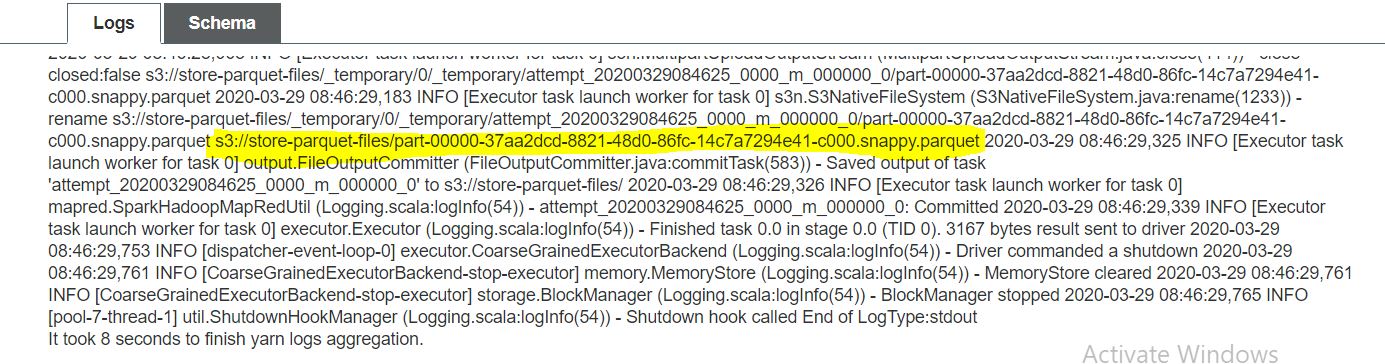
* Click on **Save job and edit script**



* Click on **Run Job**



* Click on **Run job**
* Wait until we get logs



* We get logs like as above
* Goto S3 bucket and check whether parquet file saved in target folder or not

