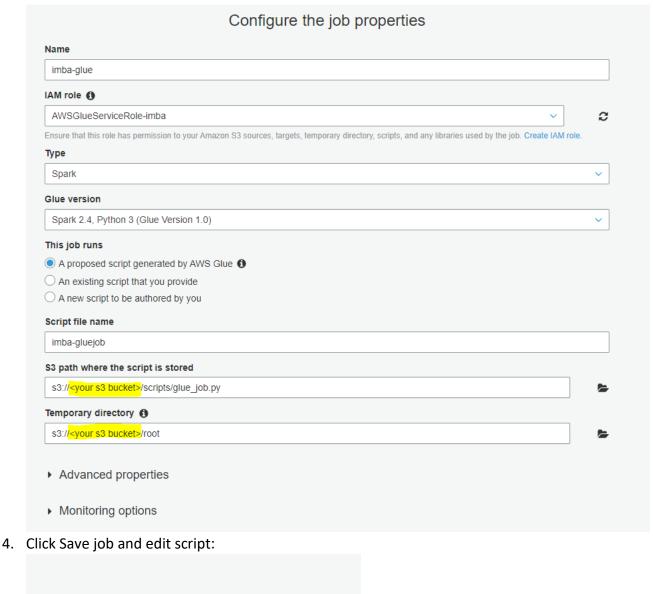
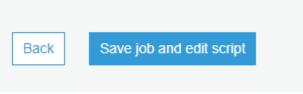
## **AWS Glue**

- 1. Open glue\_job.py, change <your s3 bucket> to your s3 bucket name, check to make sure all the s3 path is correct, save this file and upload it to s3://<your s3 bucket>/scripts/.
- 2. Open Glue console in AWS, click Jobs on the left pane:



3. Click Add job and you should fill in the details similar to below, name the job to "imbaglue", create a new IAM role or re-use an existing one (you just need to make sure AmazonS3FullAccess and AWSGlueServiceRole is attached). Make sure you select "An existing script that you provide" for "this job runs". Specify the s3 path where your script is stored: s3://<your s3 bucket>/scripts/glue\_job.py and Temporary directory: s3://<your s3 bucket>/root. Leave everything else as default and click next.





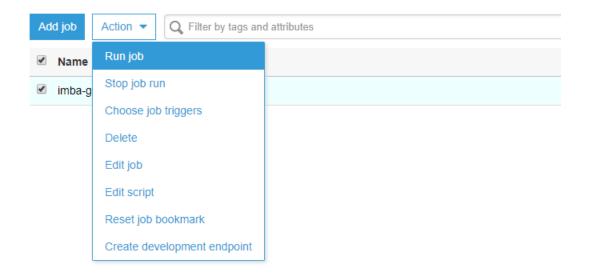
5. Have a look at the script and close it by clicking the top right X button:

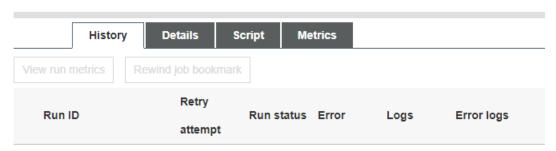


6. Select the job you created and click Run job from Action drop down menu:

## New in AWS Glue

Streaming ETL in AWS Glue (preview): Process streaming data and make it available for analysis in sec Reduced start times for AWS Glue Spark jobs (preview): Glue Spark jobs will start in under a minute. Le





- 8. Download this file to your local desktop and rename it as "data.csv"