

Instrucciones modificación JSON

1. Crear un job con las siguientes opciones:

The screenshot shows the AWS Glue Studio 'Create job' interface. On the left is a sidebar with navigation links: 'Getting started', 'ETL jobs' (with sub-links for 'Visual ETL', 'Notebooks', 'Job run monitoring'), 'Data Catalog tables', 'Data connections', 'Workflows (orchestration)', 'Data Catalog', 'Data Integration and ETL', and 'Legacy pages'. The main panel is titled 'AWS Glue Studio' and 'Create job'. It features six radio button options: 'Visual with a source and target' (selected), 'Visual with a blank canvas', 'Spark script editor', 'Python Shell script editor', 'Jupyter Notebook', and 'Ray script editor'. Below these options, there are 'Source' and 'Target' dropdown menus. The 'Source' is set to 'Amazon S3' and the 'Target' is set to 'Amazon Redshift'. A red box highlights the 'Create' button in the top right corner. Another red box highlights the 'Visual with a source and target' option. A third red box highlights the 'Target' dropdown menu.

2. En la pestaña Job details, copiar y guardar (en un lugar donde los puedan consultar después) los path que se ven a continuación:

The screenshot shows the 'Job details' tab of the AWS Glue Studio interface. The tab is titled 'Untitled job'. Below the tab are several sections: 'Visual', 'Script', 'Job details' (active), 'Runs', 'Data quality', 'Schedules', and 'Version Control'. The 'Job details' section contains several configuration options: 'Script path' (s3://aws-glue-assets-697517630681-us-east-1/scripts/), 'Job metrics' (checked), 'Continuous logging' (checked), 'Spark UI' (checked), 'Spark UI logs path' (s3://aws-glue-assets-697517630681-us-east-1/sparkHistoryLogs/), 'Maximum concurrency' (1), and 'Temporary path' (s3://aws-glue-assets-697517630681-us-east-1/temporary/). Each path has a 'View' button and a 'Browse S3' button. Red boxes highlight the 'Script path', 'Spark UI logs path', and 'Temporary path' fields.

3. Abrir el JSON en el editor de su preferencia (se recomienda VSCode) y modificar los path que encuentren en el .json cambiando el número por el número de los paths que guardaron en el paso anterior:

Por ejemplo:

```
16 "scriptLocation": "s3://aws-glue-assets-697517630681-us-east-1/scripts/",
17 "language": "python-3",
18 "jobParameters": [],
19 "tags": [],
20 "jobMode": "VISUAL_MODE",
21 "createdOn": "2023-04-27T03:31:27.197Z",
22 "developerMode": false,
23 "connectionsList": [
24   | "uniandes-bi-lab3-connection"
25 ],
26 "temporaryDirectory": "s3://aws-glue-assets-697517630681-us-east-1/temporary/",
27 "logging": true,
28 "glueHiveMetastore": true,
29 "etlAutoTuning": false,
30 "metrics": true,
31 "spark": true,
32 "pythonPath": "s3://aws-glue-studio-transforms-510798373988-prod-us-east-1/gs
33 "bookmark": "job-bookmark-disable",
34 "sparkPath": "s3://aws-glue-assets-697517630681-us-east-1/sparkHistoryLogs/",
```

4. Modificar los path relacionados con el cluster S3 y la conexión que crearon al seguir los pasos del video.

Por ejemplo:

```
187  | "paths": [
188  |   | "s3://uniandes-bi-lab3-source/OrderLines.csv"
189  | ],
190  | "connectionsList": [
191  |   | "uniandes-bi-lab3-connection"
192  | ],
```

5. Hacer upload del .json modificado, guardar y seguir los pasos del video.