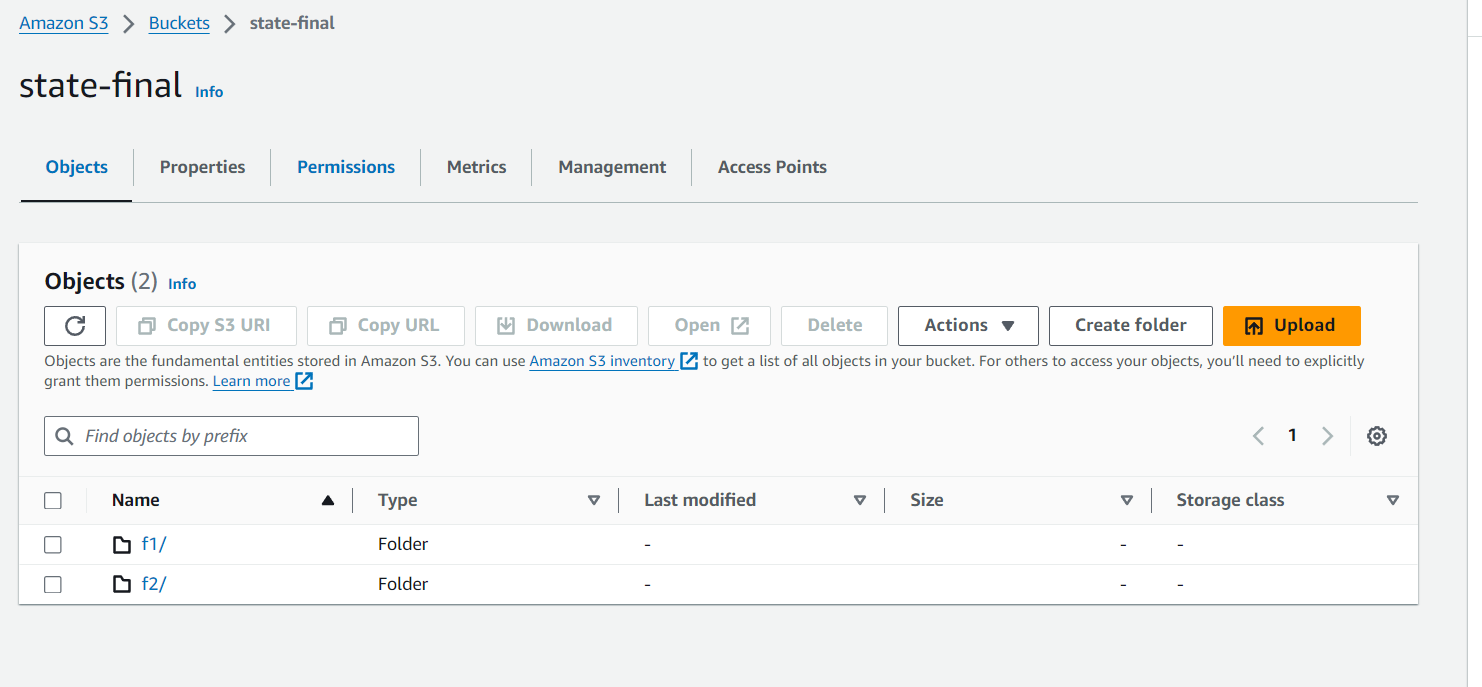
**Technical Documentation**

Creating S3 Buckets:

* Create one S3 bucket and inside it create a folder and after that create two files in it and insert the data files that we are having with us.



Creating an IAM Role:

* After getting the data into the buckets the next thing we should do is to define an IAM role and should attach policies to it.
* But instead of creating a new IAM Role I have used the old one which I have used for my previous task.

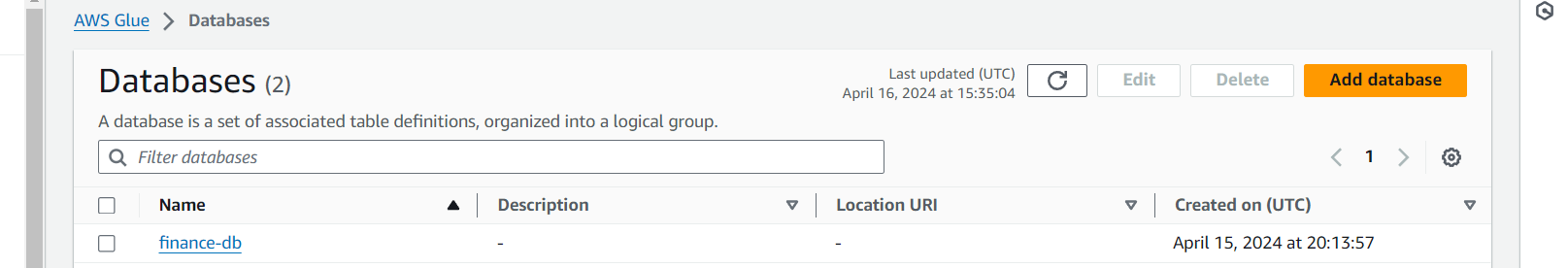
AWS Glue:

* In aws glue we will create a database and after that we will be creating crawlers.

A screenshot of a computer

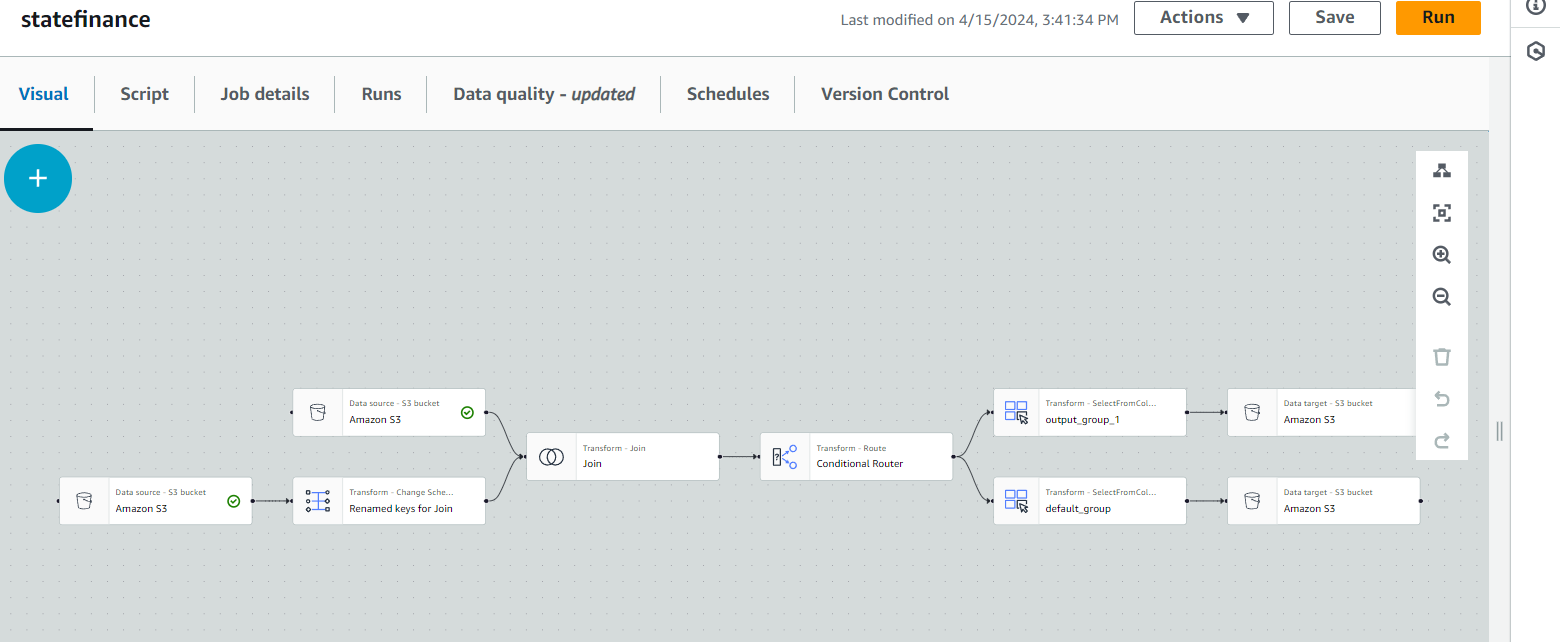
Description automatically generated

* We should create 2 crawlers and after creating them run the crawlers so that we will be getting data into the database. We will be adding the database while we are creating the crawlers.



ETL Visual:

* Below are the steps for creating ETL visual.
* Firstly, we are creating two S3 Buckets from source nodes and after that we will be adding join from the transformations.
* In join we will be combining the data and we will use inner join to do that while doing this we will be seeing overlapping issues and in order to solve it we will be using right join to resolve it.
* Aws glue will be using mapping and will rename the columns by adding right in front of all the column names.
* After that we will be adding conditional router and if we do that, we will have output group 1 and default group.



* Usually in conditional router we will be using the AND condition in order to keep files with years more than 2000 in one file and all the rest in another file.

A screenshot of a computer

Description automatically generated

* After that we will have two target files and in S3 we will create a bucket with 2 separate output files in it.
* We will be adding S3 bucket as target and node and we will be adding the URL’S of output files that we have created previously.

A screenshot of a computer

Description automatically generated

* Now we will be saving the job and running it.

A screenshot of a computer

Description automatically generated

AWS Athena:

* After running the ETL job we will be running the queries in aws athena in order to check whether the data has been sorted according to the condition that we have defines in the ETL flow.

A screenshot of a computer

Description automatically generated

* We will be creating a folder and two separate files in it and we will be giving this URL’s for this tables and we will be saving all of these results into them.

A screenshot of a computer

Description automatically generated