AWS SERVICES :

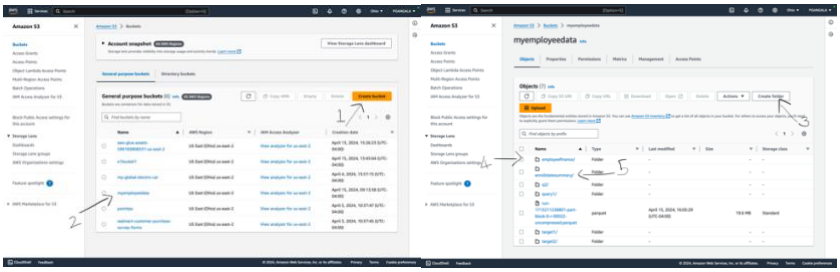
USED SERVICES: S3 BUCKET, IAM, AWS GLUE, AWS SLUE STUDIO, ATHENA

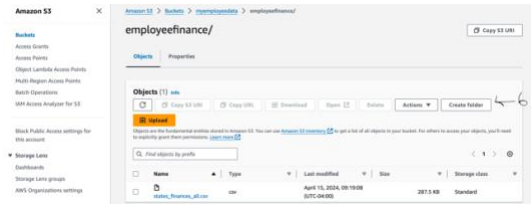
S3 BUCKET

step1:Create a new bucket in an S3 Bucket

step2:Create a folder and 2 subfolders in it

step3:Insert file (employee state and finance CSV files)

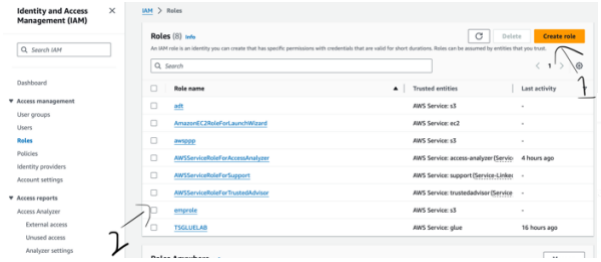


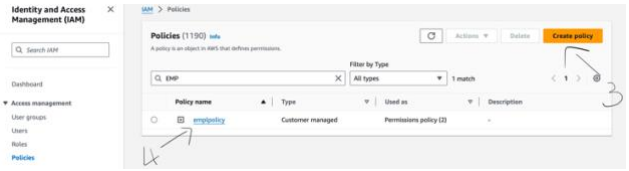


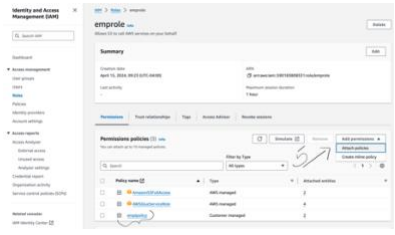
IAM

step1:Create a Role and a new policy

Step 2:Attach policy to the role assigned.

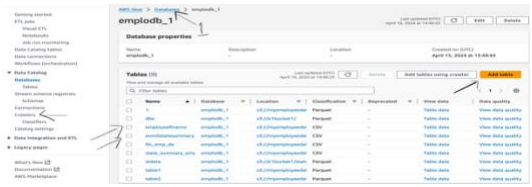






AWS GLUE

step1:Create a database following up with a table and crawler



AWS GLUE STUDIO

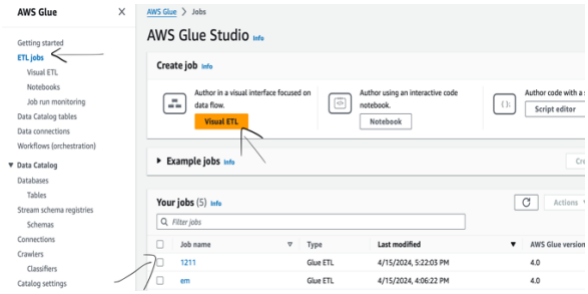
step1: ETL Visual

step2:Start node diagrams with

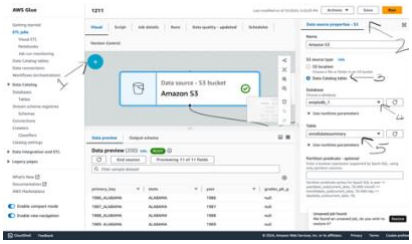
step3 :Amazon S3 – Node 1

step4 :Amazon S3 – Node 2

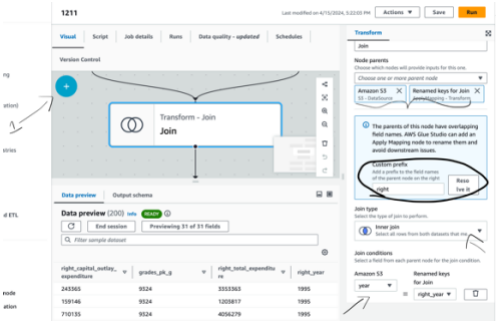
1.



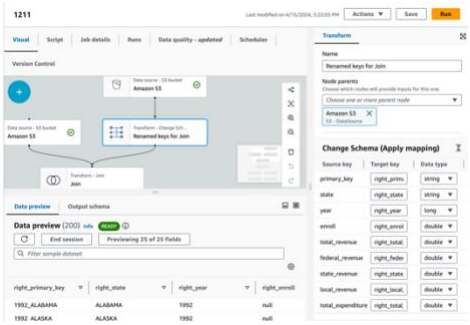
2.



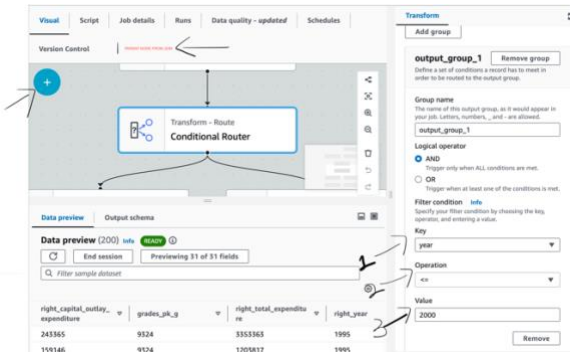
3.



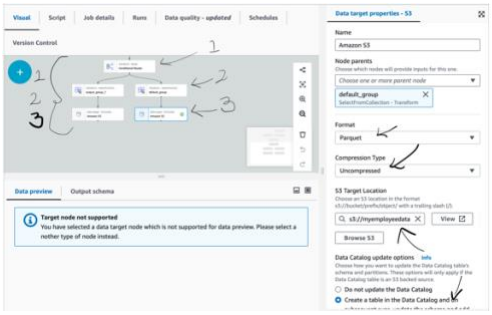
4.



5.



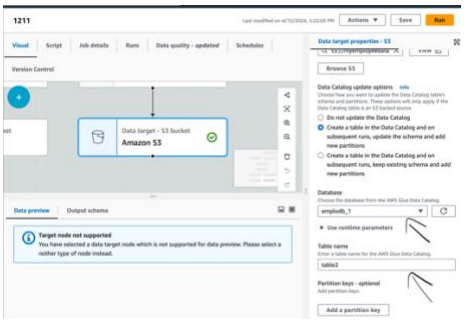
6.



under transforms:

step5 :JOIN NODE (year =right year)

step6 :Conditional Node {AND – Year(key)- <=2000(value)}



Under Targets: (Assign IAM)

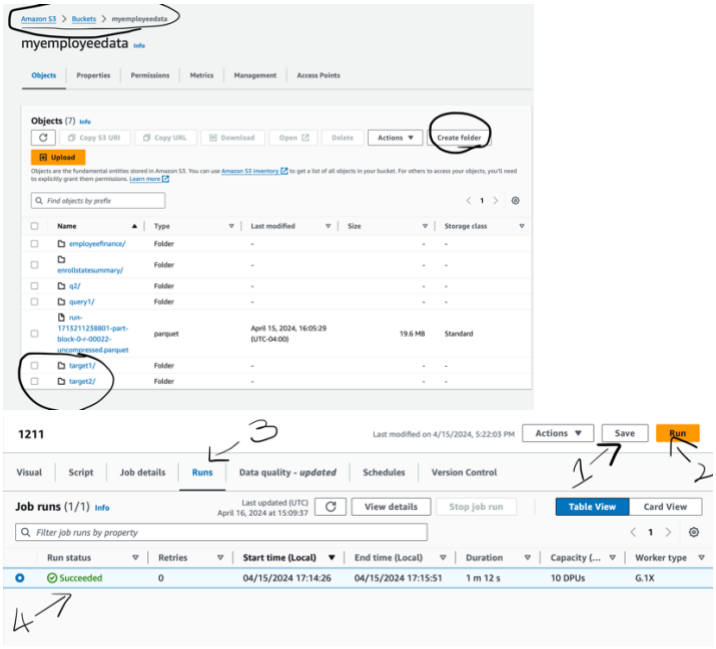
step7 :S3 Node : Folders(f1, f2)

step8 :Browse the S3 target path

step9 :Uncompress

step10:Data Catalog

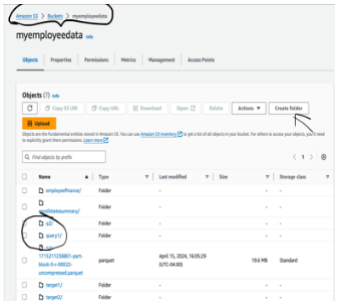
step11 : Create table



step12:Do for both targets: Save & Run
Running Status: Succeeded

S3 BUCKET:

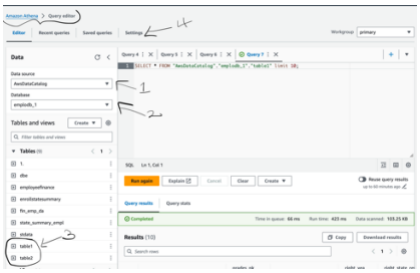
step1:Back to the Bucket path
*step2 :*create new Query1 & Query2 folders

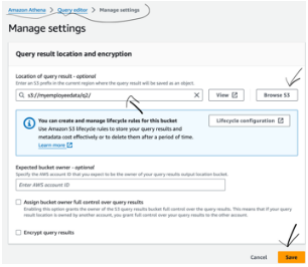


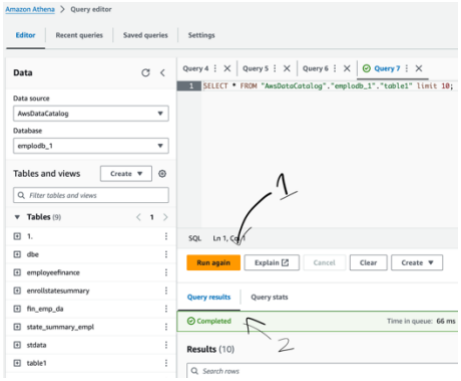
step3:And a table
step4:Table data for files
*step5:*Connecting to Athena

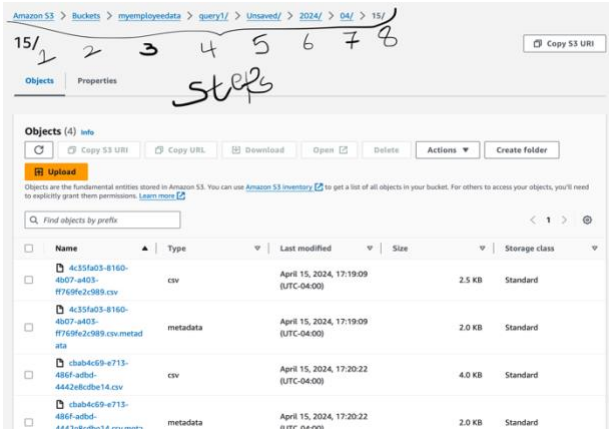
ATHENA:

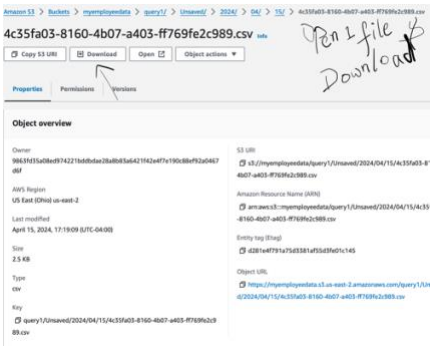
*step1 :*settings | Manage | Browse S3 path(q1) | Save | Editor | Run
step2 :Check the table in the S3 bucket: under the Q1 folder & download the table to view the data
*step3:*Same for Query2 Run step: 1 in Athena and view

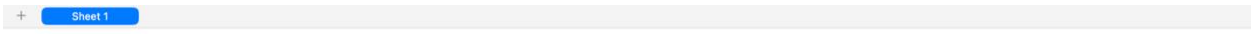












4c35fa03-8160-4b07-a403-ff769fe2c989

capital_outlay_expenditure	grades_pk_g	right_total_expenditure	right_year	right_state_revenue	grades_12_g	right_grades_8_g	right_instruction_expenditure	grades_all_g	right_grades_12_g	state	right_enroll	grades_8_g	right_federal_revenue	year	right_grades_4_g	right_support_services_expenditure	right_state	right_grades_bg_g	primary_key	grades_bg_g	grades_4_g	right_grades_all_g	grades_1_8_g	right_primary_key	right_local_re
			1988		45571.0	55262.0		724751.0	45571.0	ALABAMA		55262.0		1988	57531.0		ALABAMA	54265.0	1988_ALABAMA	54265.0	57531.0	724751.0		1988_ALABAMA	
			1988		45571.0	7164.0		724751.0	6851.0	ALABAMA		55262.0		1988	8517.0		ALASKA	8637.0	1988_ALABAMA	54265.0	57531.0	106481.0		1988_ALASKA	
			1988		45571.0	39317.0		724751.0	38962.0	ALABAMA		55262.0		1988	46518.0		ARIZONA	48132.0	1988_ALABAMA	54265.0	57531.0	574890.0		1988_ARIZONA	
			1988		45571.0	32481.0		724751.0	33074.0	ALABAMA		55262.0		1988	33777.0		ARKANSAS	34323.0	1988_ALABAMA	54265.0	57531.0	436387.0		1988_ARKANSAS	
			1988		45571.0	318856.0		724751.0	257450.0	ALABAMA		55262.0		1988	352076.0		CALIFORNIA	403237.0	1988_ALABAMA	54265.0	57531.0	4818120.0		1988_CALIFORNIA	
			1988		45571.0	39524.0		724751.0	37538.0	ALABAMA		55262.0		1988	43381.0		COLORADO	46599.0	1988_ALABAMA	54265.0	57531.0	560081.0		1988_COLORADO	
			1988		45571.0	31713.0		724751.0	32164.0	ALABAMA		55262.0		1988	34074.0		CONNECTICUT	38688.0	1988_ALABAMA	54265.0	57531.0	460607.0		1988_CONNECTICUT	
			1988		45571.0	6857.0		724751.0	6857.0	ALABAMA		55262.0		1988	7982.0		DELAWARE	7976.0	1988_ALABAMA	54265.0	57531.0	99578.0		1988_DELAWARE	
			1988		45571.0	5342.0		724751.0	4345.0	ALABAMA		55262.0		1988	6042.0		DISTRICT_OF_COLUMBIA	6092.0	1988_ALABAMA	54265.0	57531.0	84782.0		1988_DISTRICT_OF_COLUMBIA	
			1988		45571.0	127642.0		724751.0	99658.0	ALABAMA		55262.0		1988	132484.0		FLORIDA	153773.0	1988_ALABAMA	54265.0	57531.0	1720630.0		1988_FLORIDA	