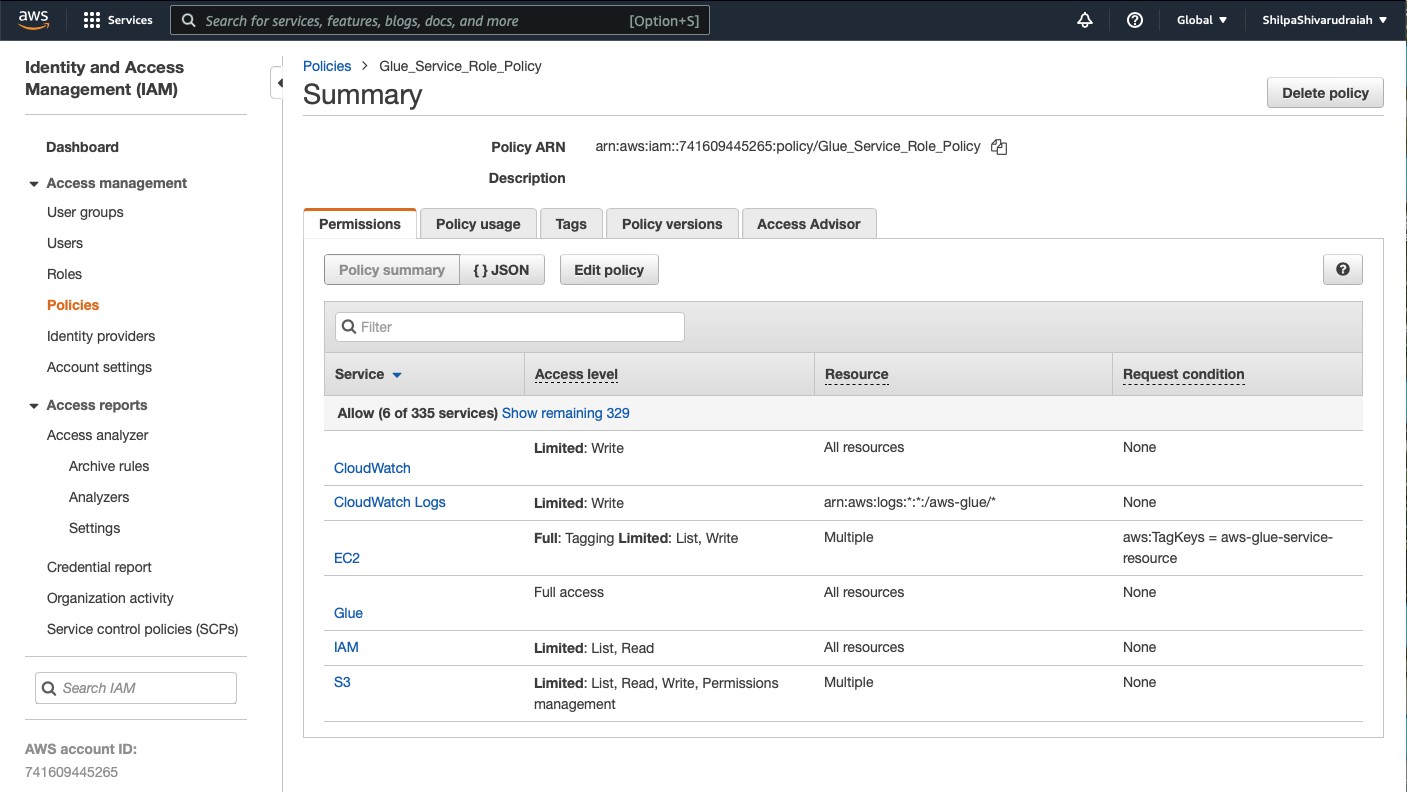
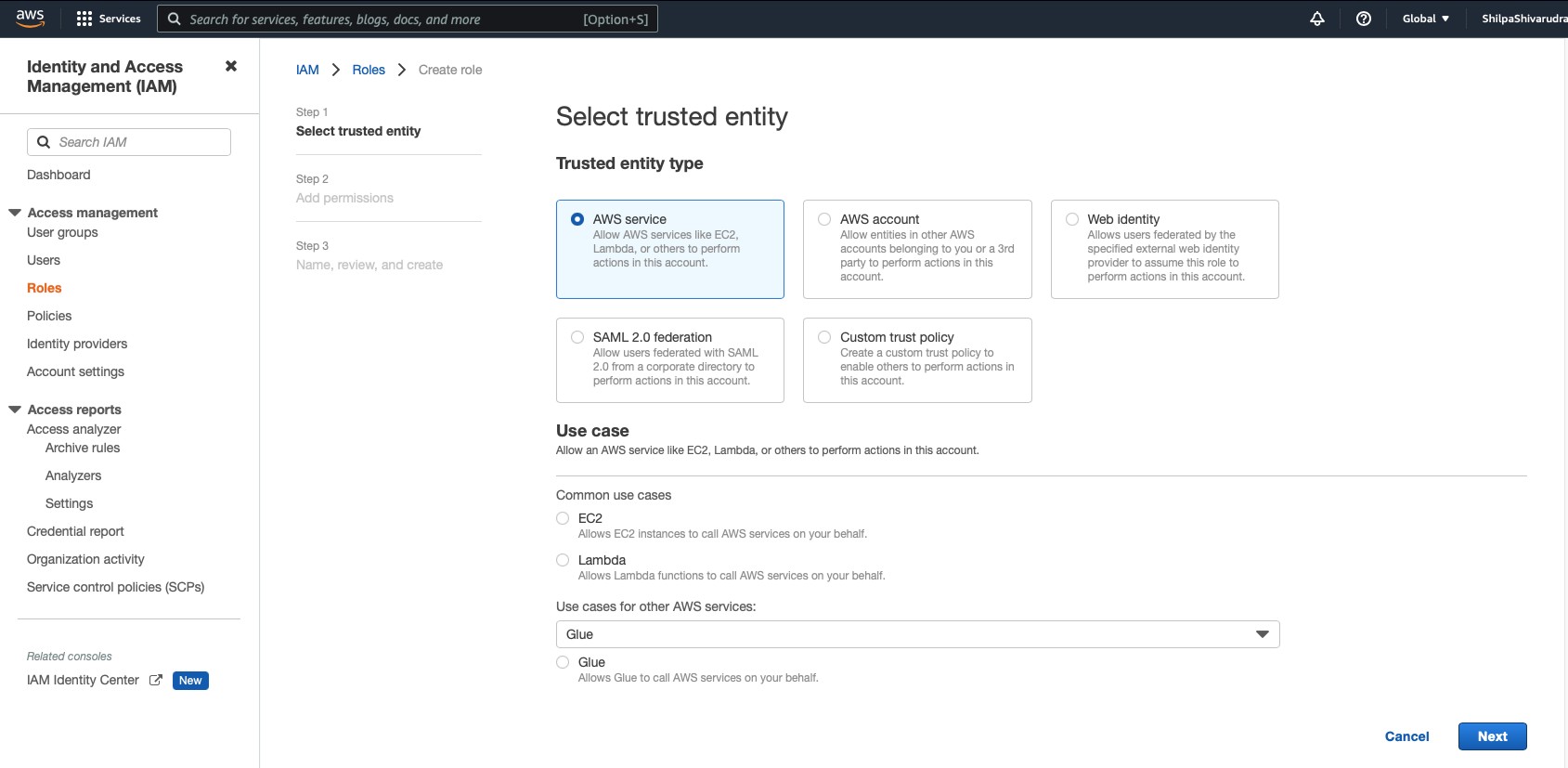
HW- 3

Step 1: Setting up IAM permission for AWS Glue

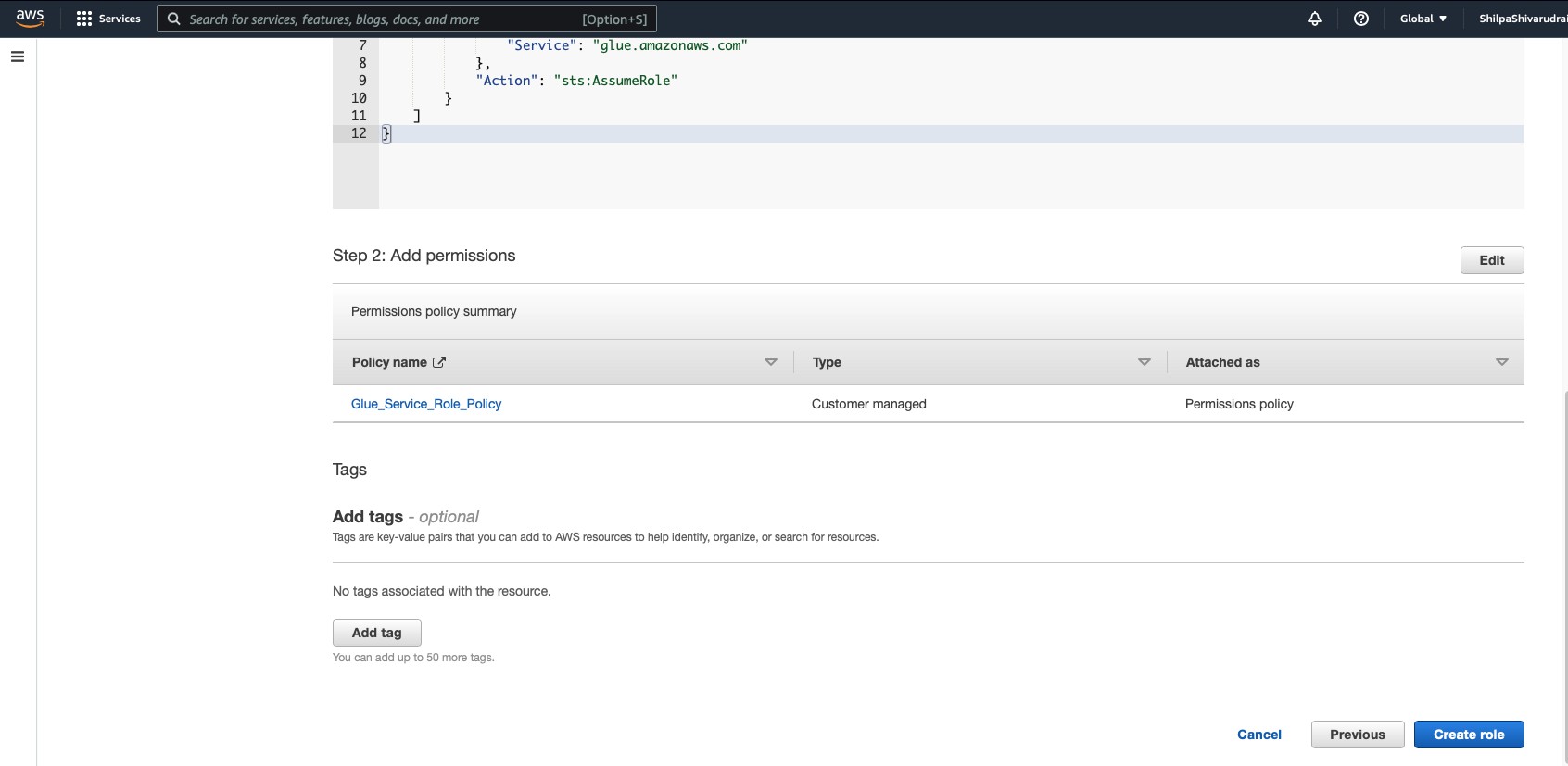
* 1. Logged in to the created AWS account.
  2. Creating a policy with given JSON statements and named the policy as Glue\_Service\_Role\_Policy.



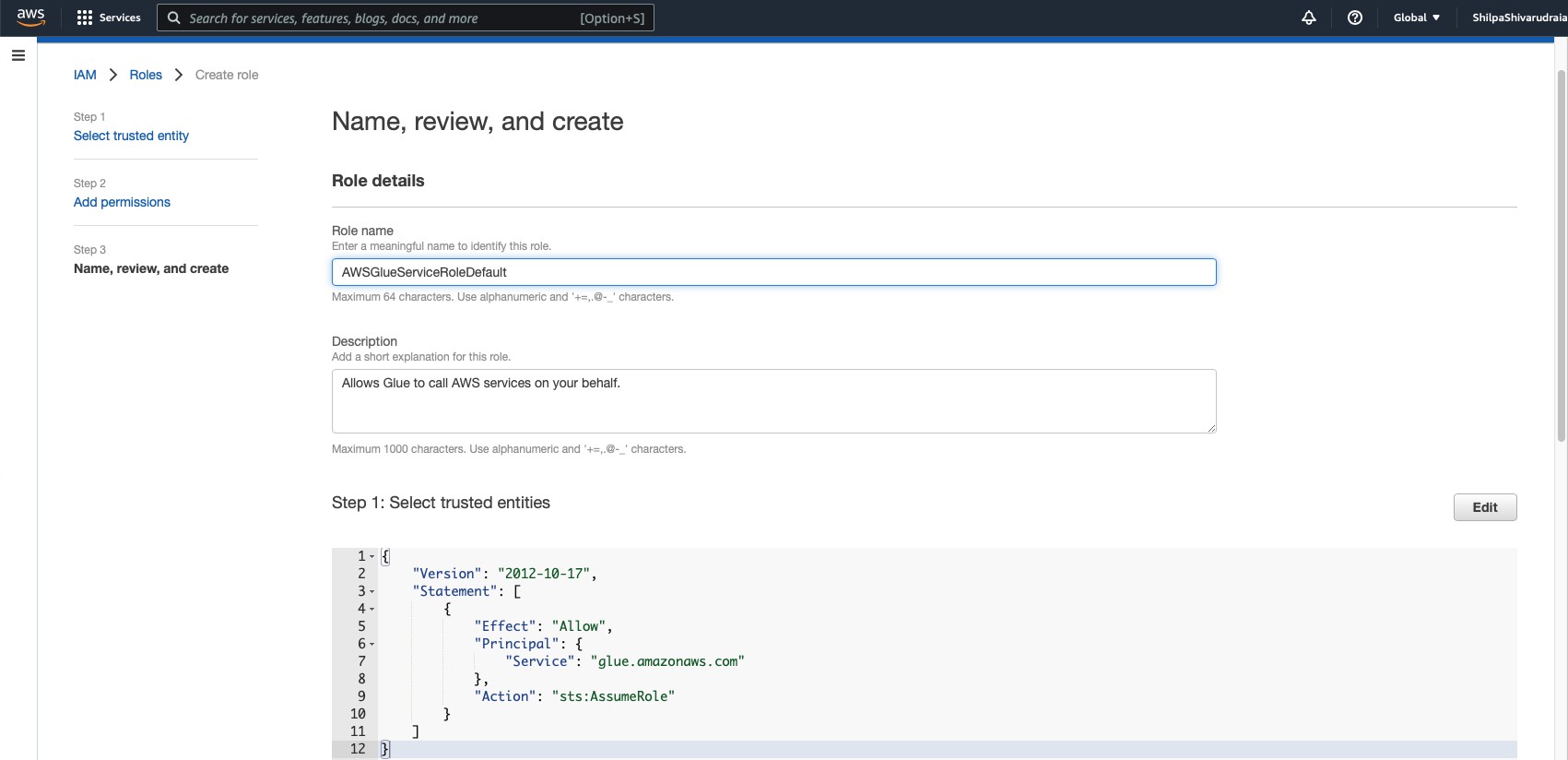
Step 2: To create an IAM role for AWS Glue by creating a role under new role and selecting Glue under AWS Service



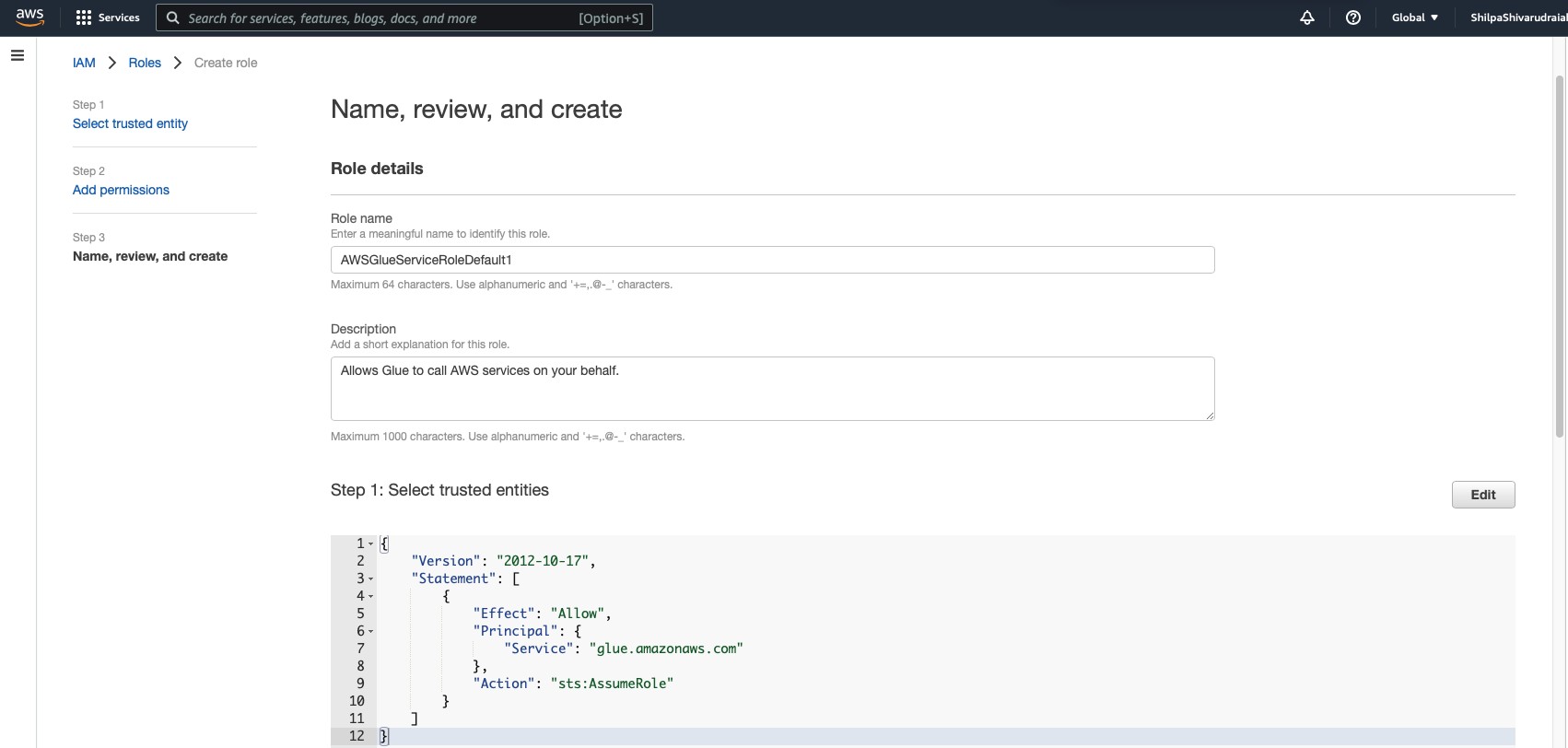
Step 3: On the Attach permissions policy page, select the policies Glue\_Service\_Role\_Policy for general AWS Glue permissions and the AWS managed policy AmazonS3FullAccess for access to Amazon S3 resources.



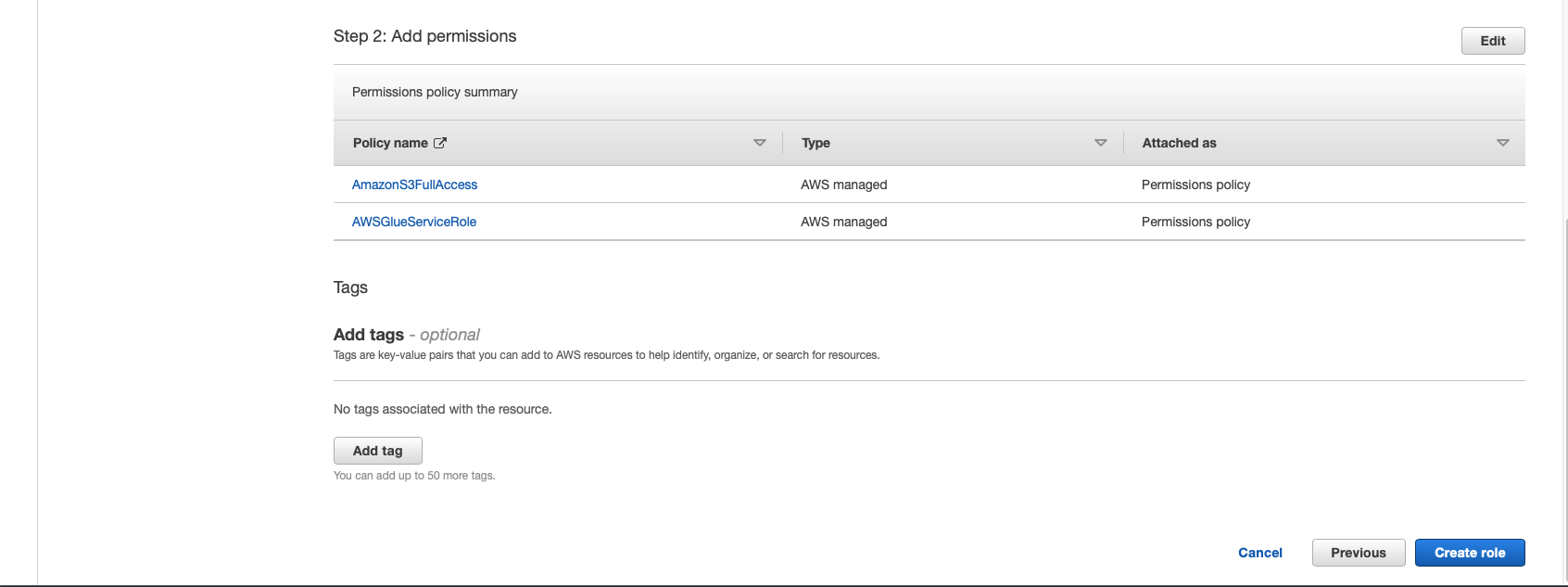
Step 4: Creating the name of the role as AWSGlueServiceRole. This role shall be passed from console users to the service.



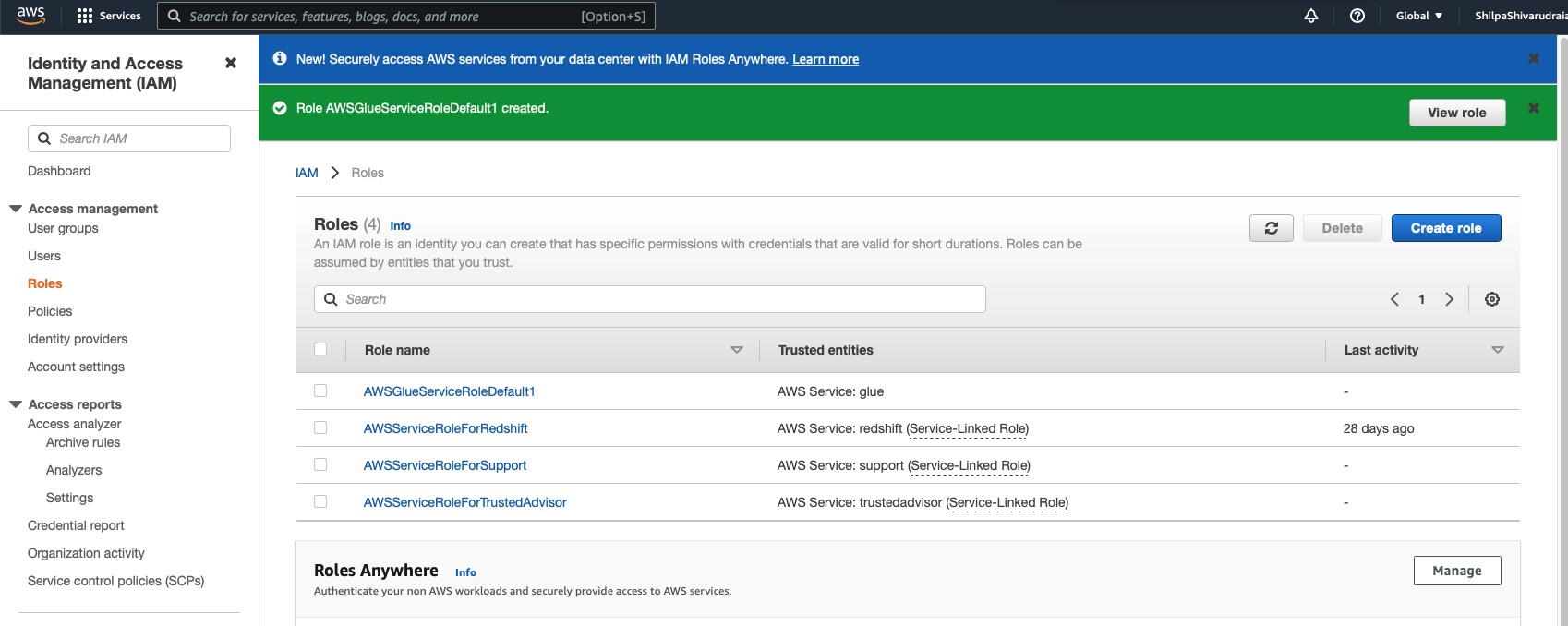
Step 5: Creating a new Role under the Role by providing a role name as AWSGLUESERVICEROLEDEFAULT1.



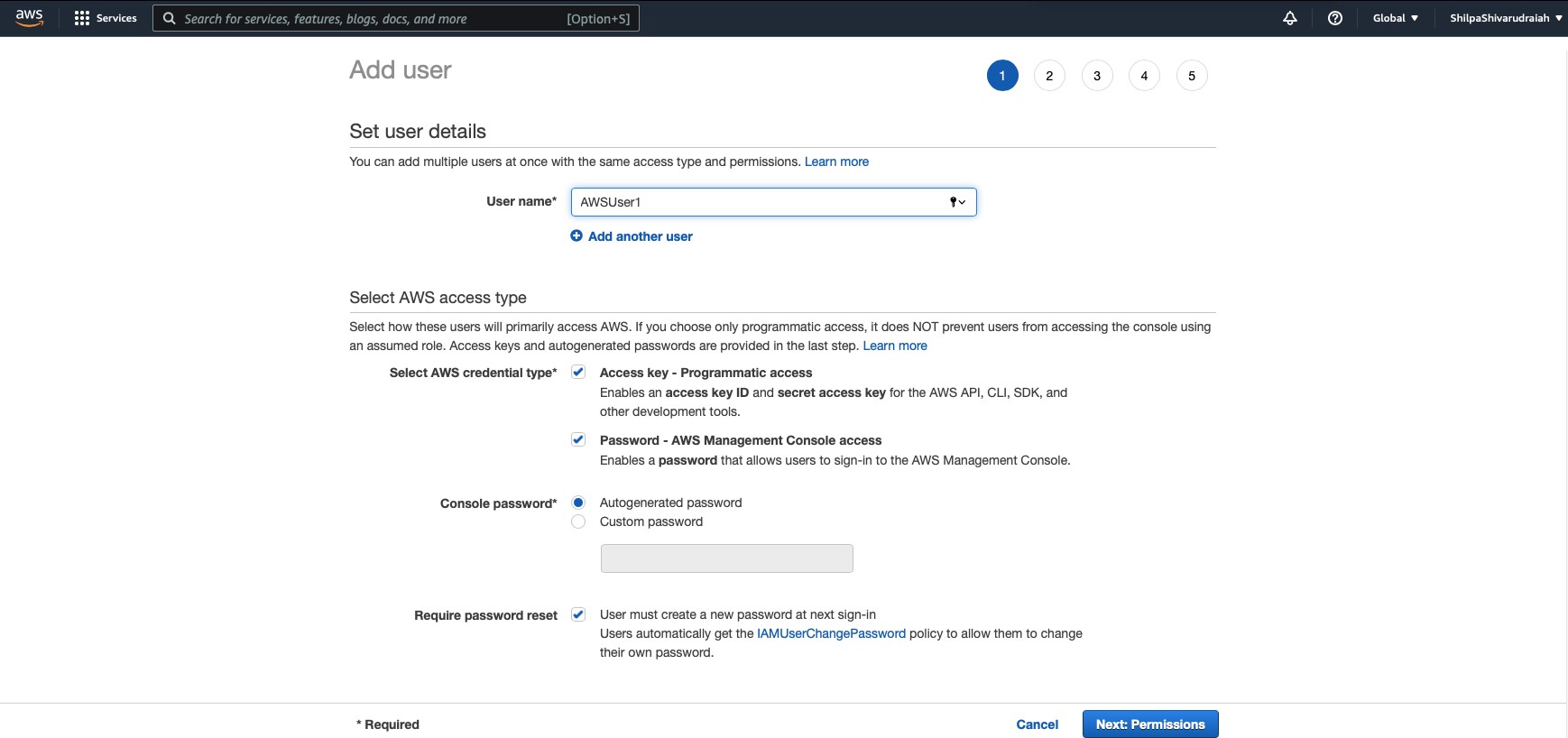
Step 6: Add permission on the Attach permissions policy page by choosing the policies that contain the required permissions. For example, the AWS managed policy AWSGlueServiceRole for general AWS Glue permissions and the AWS managed policy AmazonS3FullAccess for access to Amazon S3 resources.



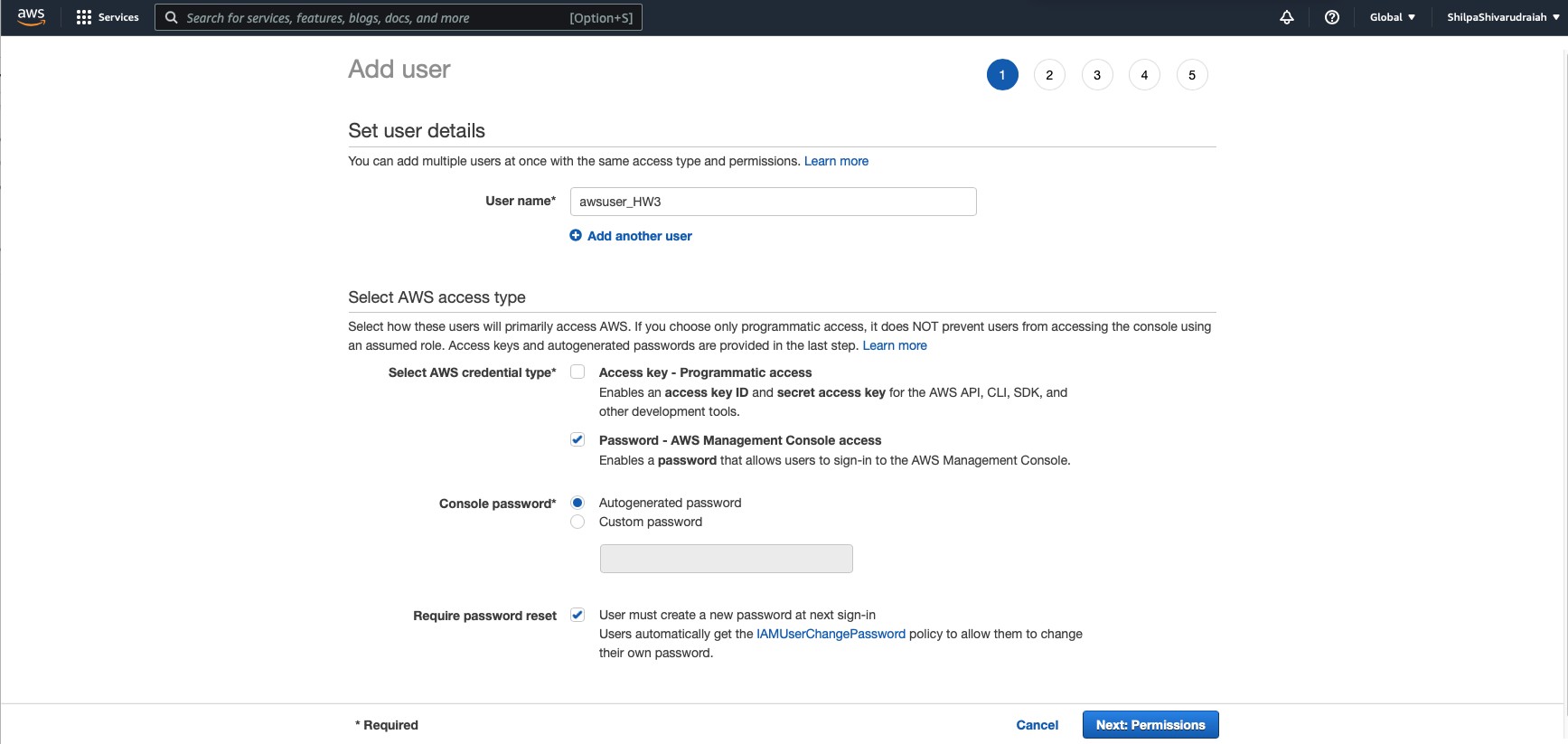
Step 7 : Creating the name of the role as AWSGlueServiceRoleDefault. This role shall be passed from console users to the service.



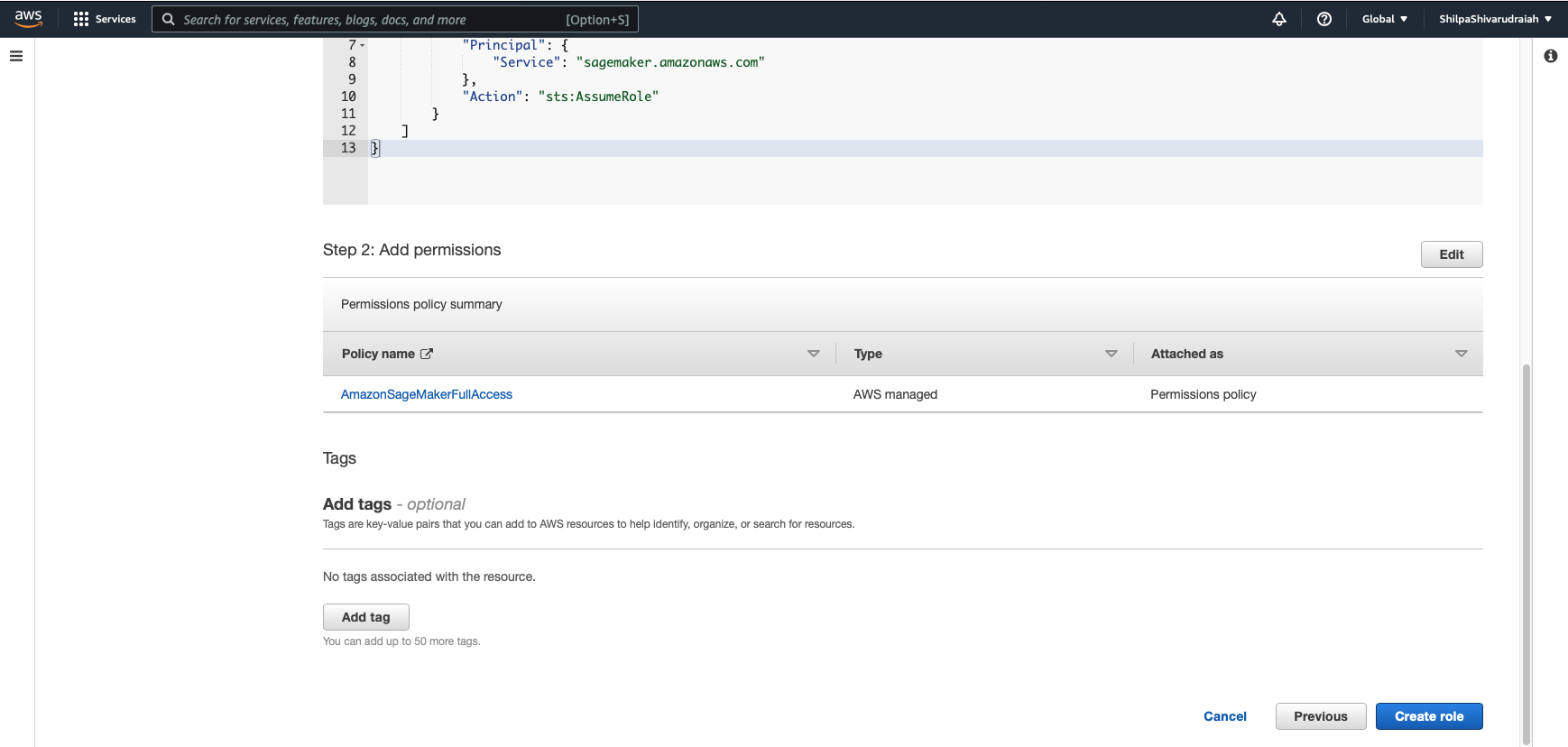
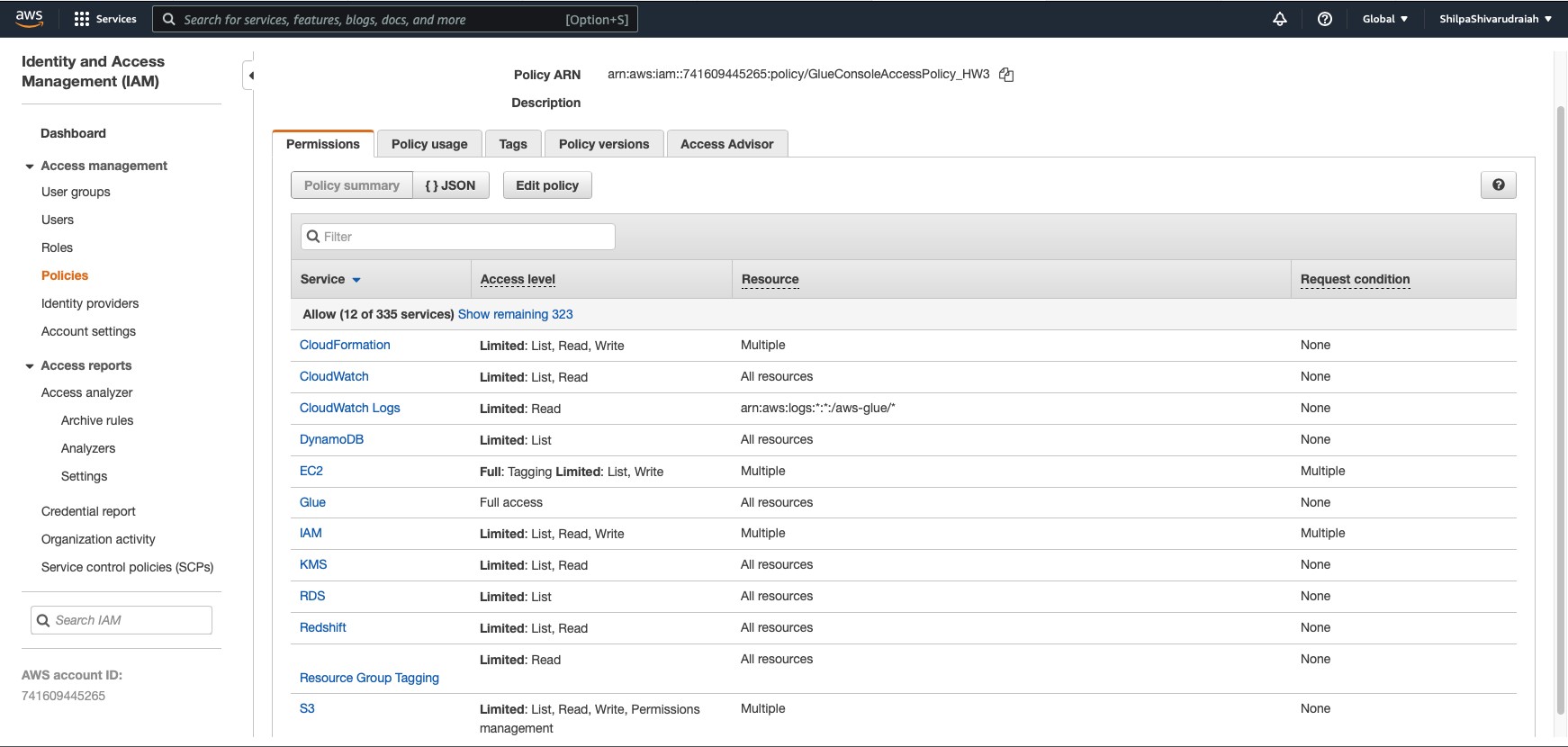
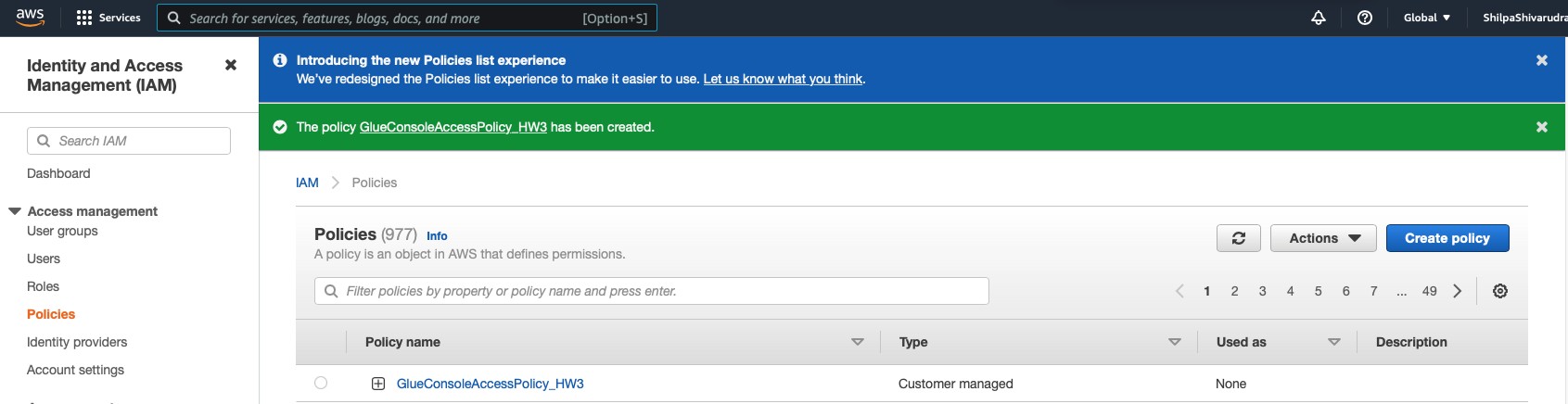
Step 8: Adding a user and selecting the necessary AWS credential type to embed a policy in.



Step 9 :Select the name of the user to embed the policy. Also choose the permission tab and select the name of the user to embed the policy

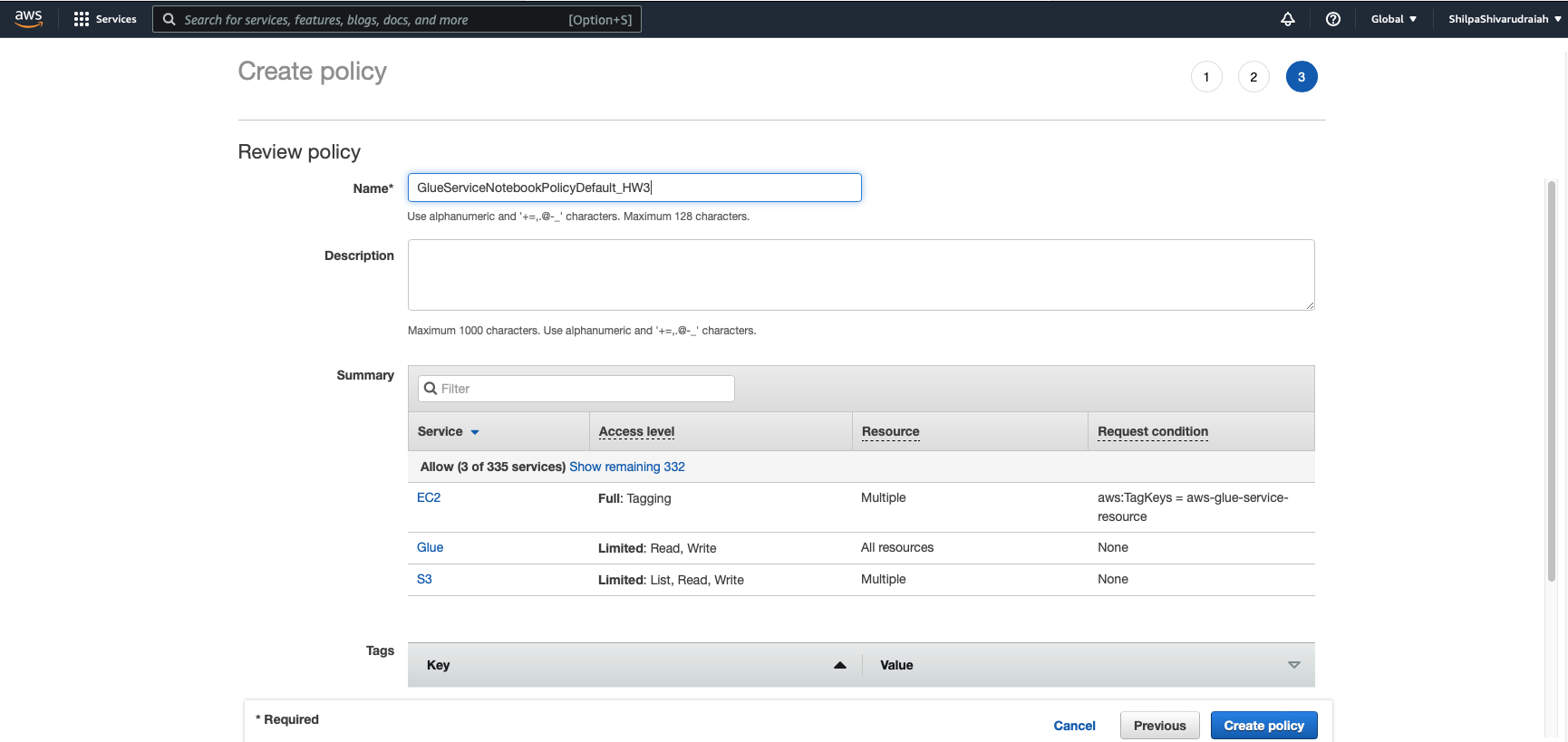
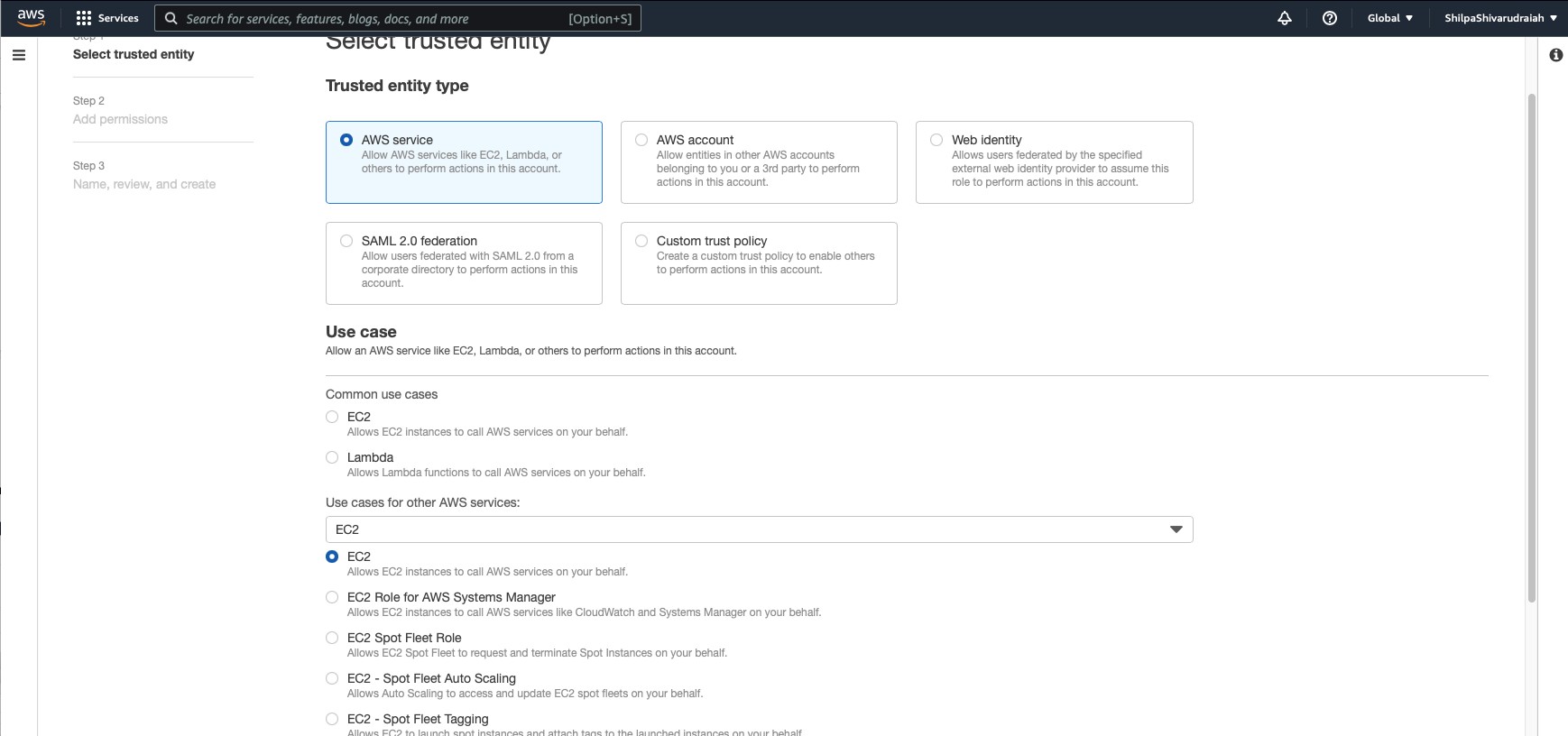


From the list of policies, select the GlueConsoleAccessPolicy\_HW3.



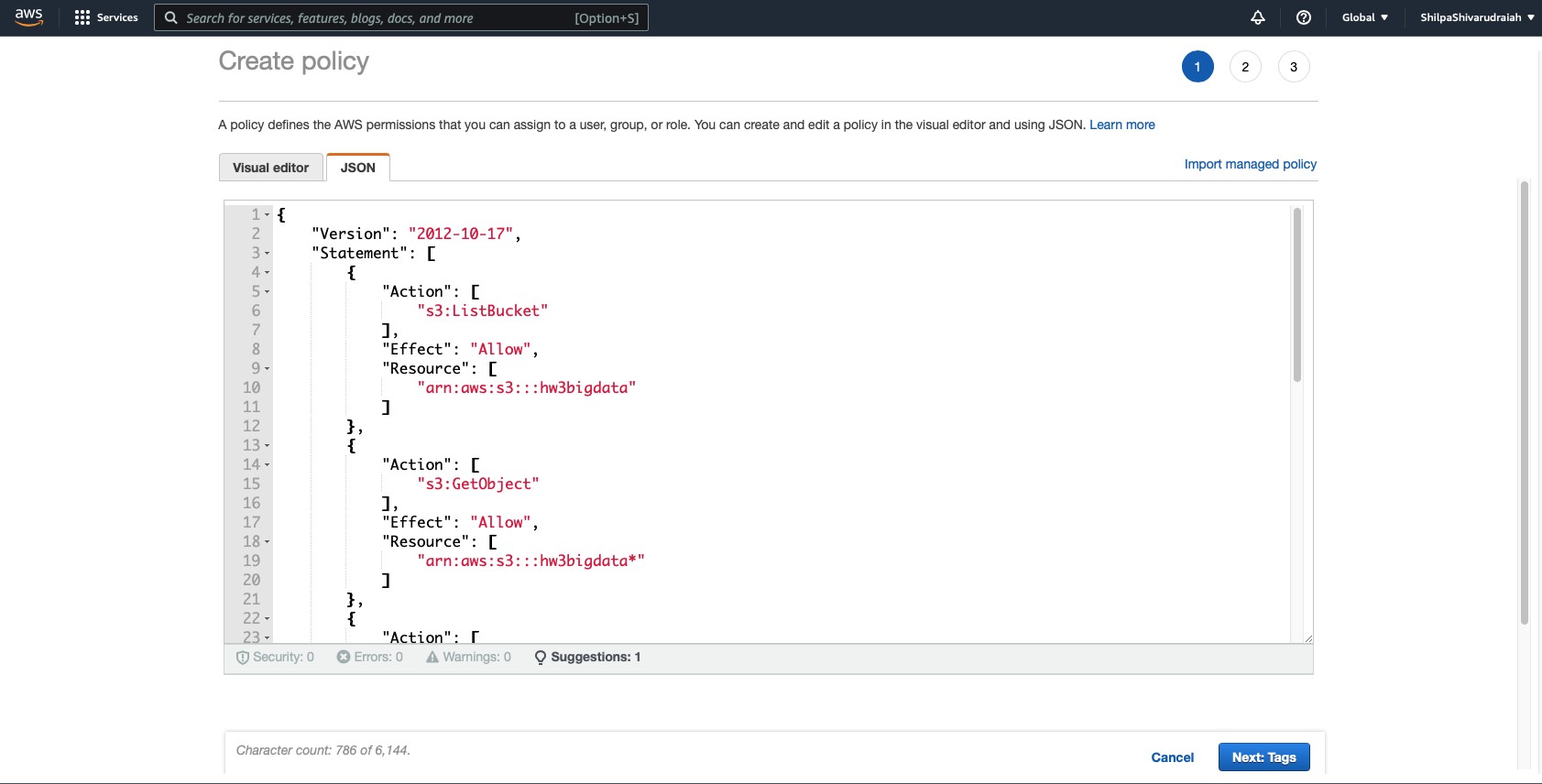
Step 10: Creating a policy with JSON statements given and named the policy as GlueServiceNotebookPolicyDefault.

* 1. In the left navigation pane, choose Roles and Choose Create role
  2. For role type, choose AWS Service, Select EC2, and choose the EC2 use case, then choose Next: Permissions.

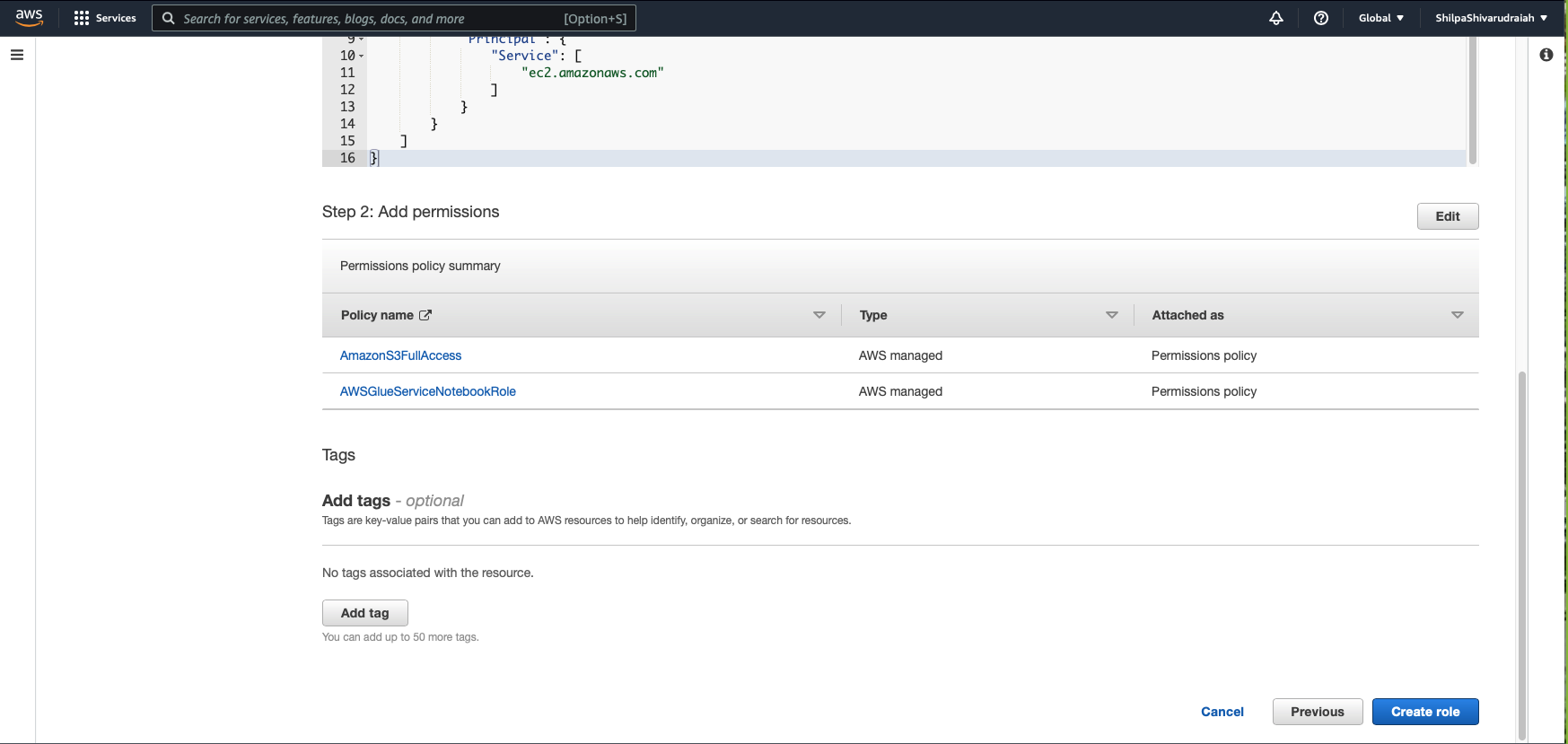


Step 11: To create an IAM role for notebooks by signing in to the AWS Management Console and open the IAM console at <https://console.aws.amazon.com/iam/>.

* 1. In the left navigation pane, choose Roles.
  2. Choose Create role.
  3. For role type, choose AWS Service, find and choose EC2, and choose the EC2 use case, then choose Next: Permissions.



Step 12 : On the Attach permissions policy page, choose the policies that contain the required permissions; for example, AWSGlueServiceNotebookRole for general AWS Glue permissions and the AWS managed policy AmazonS3FullAccess for access to Amazon S3 resources. Then choose Next: Review.

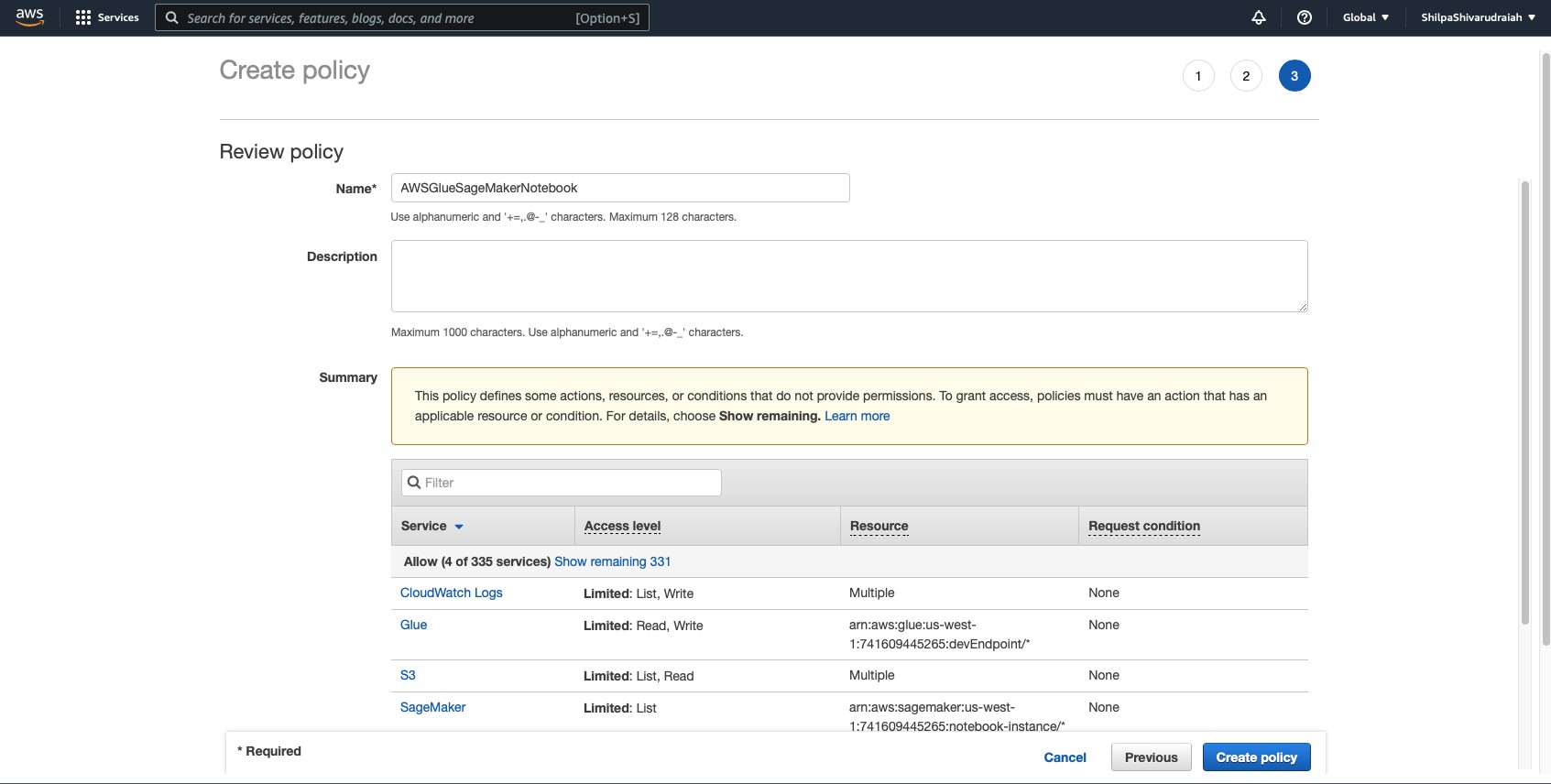


Step 13: Create an IAM policy for SageMaker notebooks.

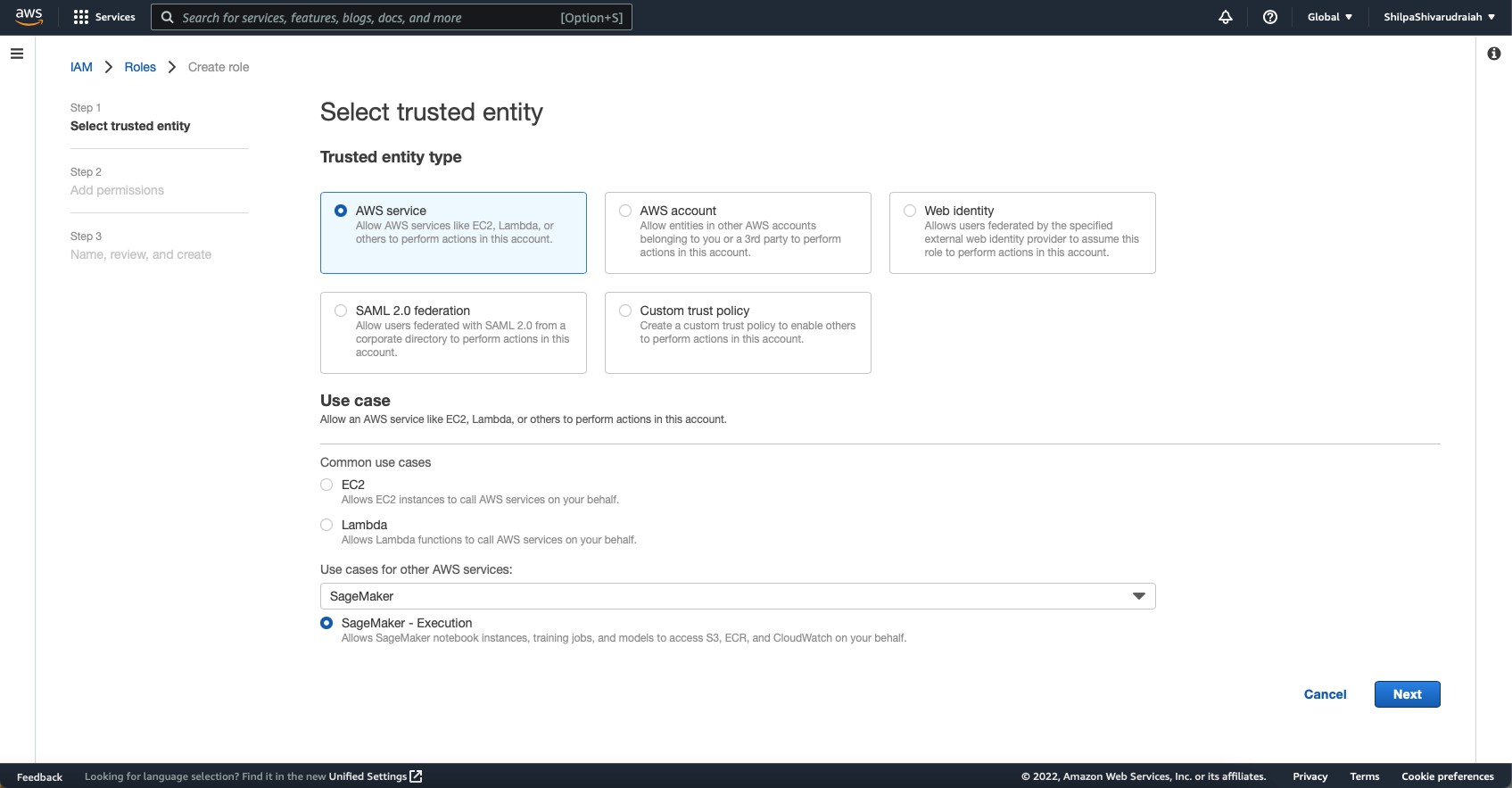
* 1. Sign in to the AWS Management Console and open the IAM console at <https://console.aws.amazon.com/iam/>.
  2. In the left navigation pane, choose Policies.

13.3. Choose Create Policy.

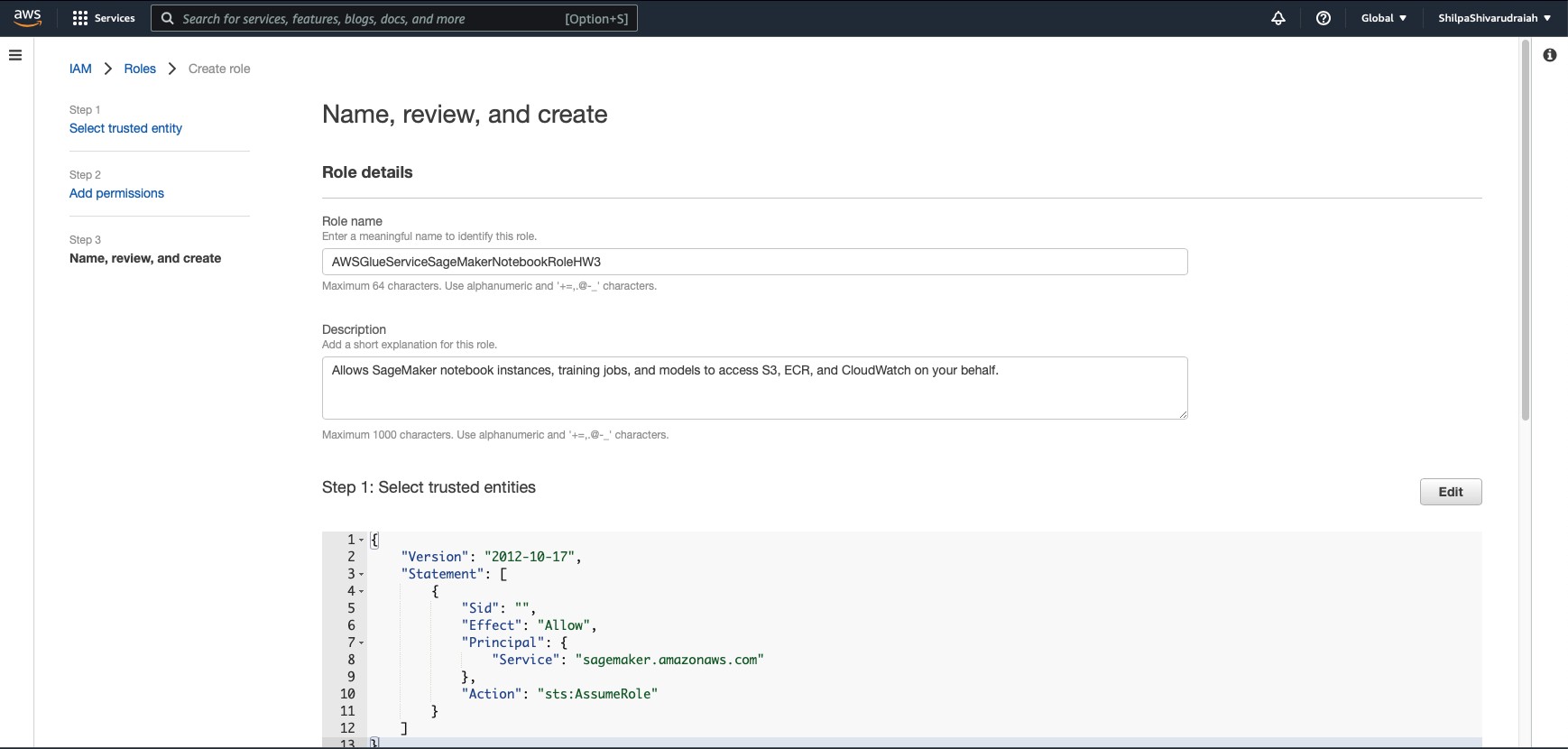
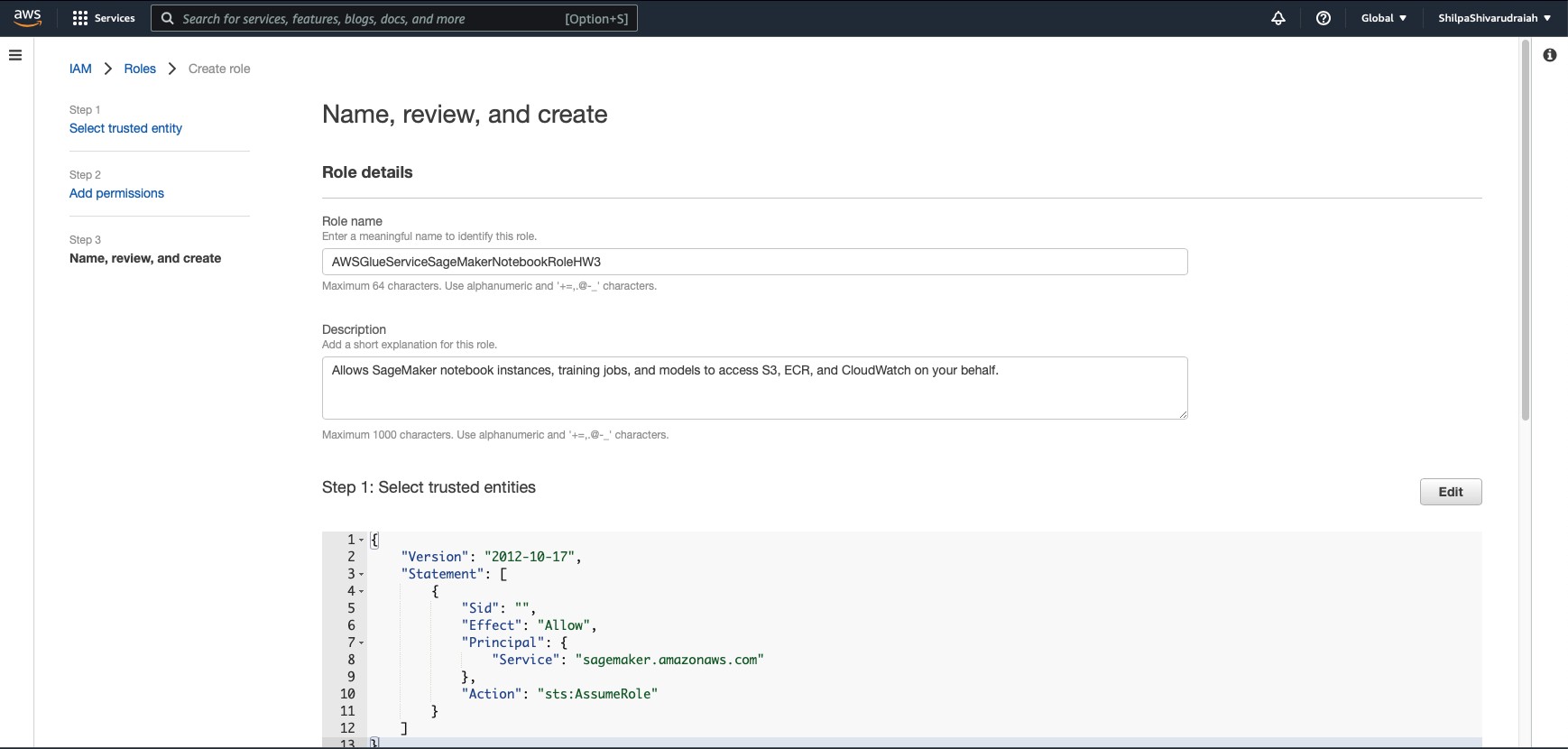
13.4. On the Create Policy page, navigate to a tab to edit the JSON. Creating a policy document with the following JSON statements with the bucket name h3bigdata, region as us-west-1,with the account id 741609445265 for the environment.



Step 14: To create an IAM role for SageMaker notebooks by selecting the SageMaker Execution use case and then click on permission



Step 15: For the role name, we choose AWSGlueServiceNotebookRole\_HW3 to allow the role to be passed from console users to SageMaker.



|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | |
| Identity and Access | X |  |  |
| Management (1AM)  Q *Sean:hf.AM*   * Access management User groups   Roles  k:lentltyprovlders Aec:ount settings   * Access reports Aec:essanalyzer Archive rules   Analyzers Settings  Credential report  OrganlzatOOactlvlty Servk;econtrolpollcles(SCPs)  1AM Identity Cent&( CZ' 8 |  | 1AM ) Roles ) AWSGlueServK:8NotebookRole\_HW3  AWSGlueServiceNotebookRole\_HW3  Al s EC2 lnstan s 10 call AWS services on your tlehalt  Summary  Creation date ARN  September 29, 2022, 13:33 {UTC-07:00) am:aws:lam::741609445265:role/AWSGlueServlceNotebookRol  e\_HW3  L.astactMty  Noo,  nust relatlonshlps Tags Access Advisor  Permissions policies (3)  You can attach up to 10 mWlaged policies.  Q *FIiier* pol/c/es *by propefty or poNcy name* and *press* enter,  Polley name CZ'  1±1 AWSGlueSageMakerNot&book Customer managed  1±) ii AmazonS3FullAccess AWSmanaged  1±1 ii AWSGlueServlceNotebookRole AWSmanaged | Instance profile ARN  l'tJ arn:aws:IBm::741609445265:lnstance-profiQIAWSGlueServlceN  otebookRole\_HW3  Add permissions •  < 1 > ®  Description  Polley for AWS Glue service role whldl allo... |



 S. :::krv1cco Q , 4 Ci) Cloblll • ShllpoShlvllnxlraollh •

1AM ) Roles ) Create role

0

Step1

Select trusted entity

Name, review, and create

Step2

AddpermissOOs

Role details

Step3

Name, review, andcreate

Enlll'o.meaningfuMITll!!OKlenOlythSrole

AWSGlueServiceSageMakerNotebookRoleHW3

MioomYm64 characters. UselllphanumMk; ilfld "+a,.9·

Description

AddashMex nationforthisrol11

Mows SageMaker notebook ins.tances, training )obs, and models to access S3, ECR, and CKll..ldWatdl on your behalf.

Mioomlllll1000characters.Use ar.umerlc.11d".;=,.0-

Step 1: Select trusted entitles

Edit

1·

2

3

4·

7

8

9

10

11

12

"Ve-rsion": "2012-10-17",

"Statement": [

{

"Sid":"",

"Effect": "Allow",

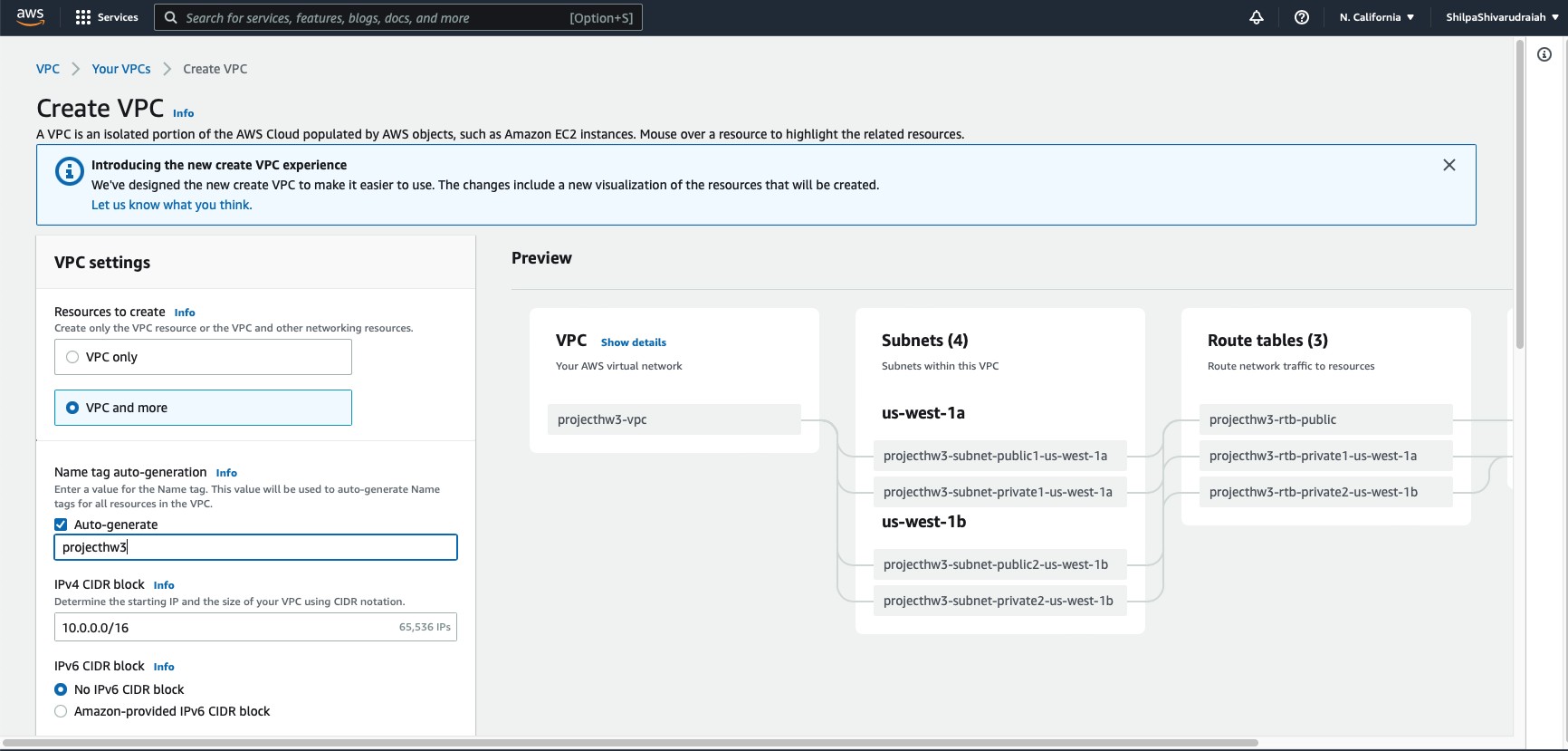
"Pri.ncipal": {

"Service": "sagemciker.anazonaws.com"

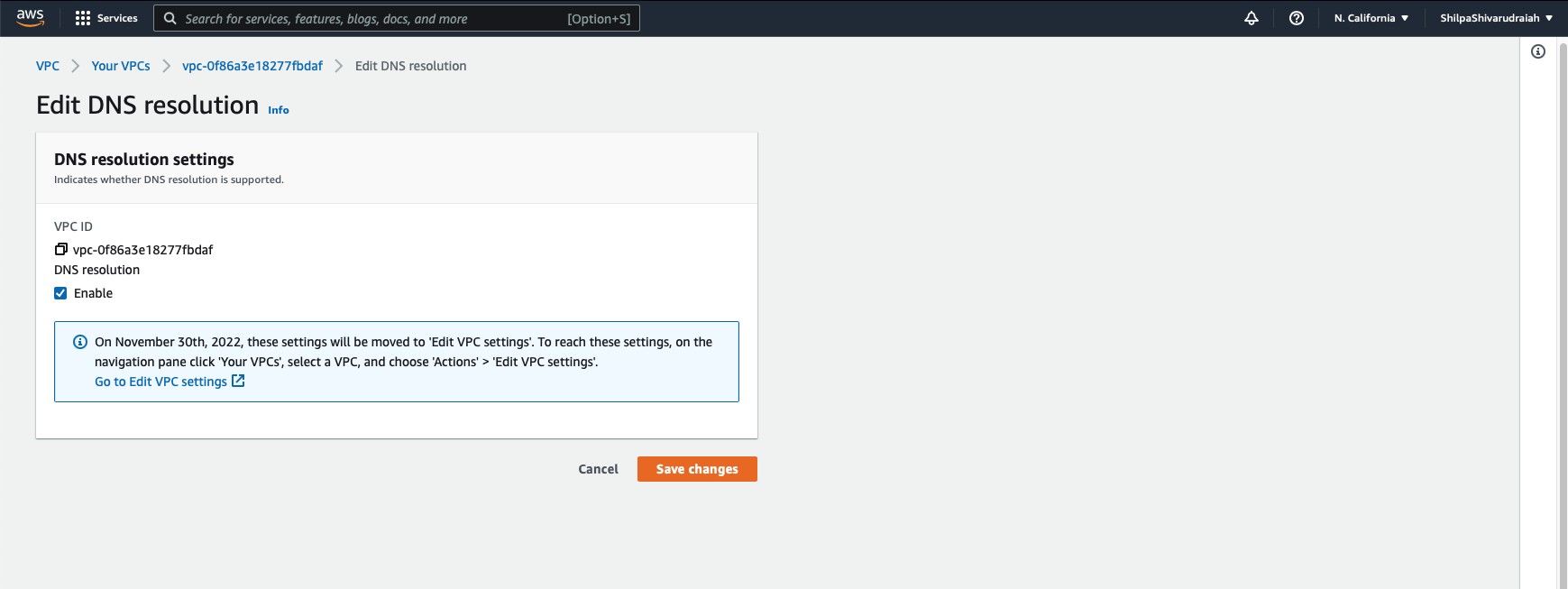
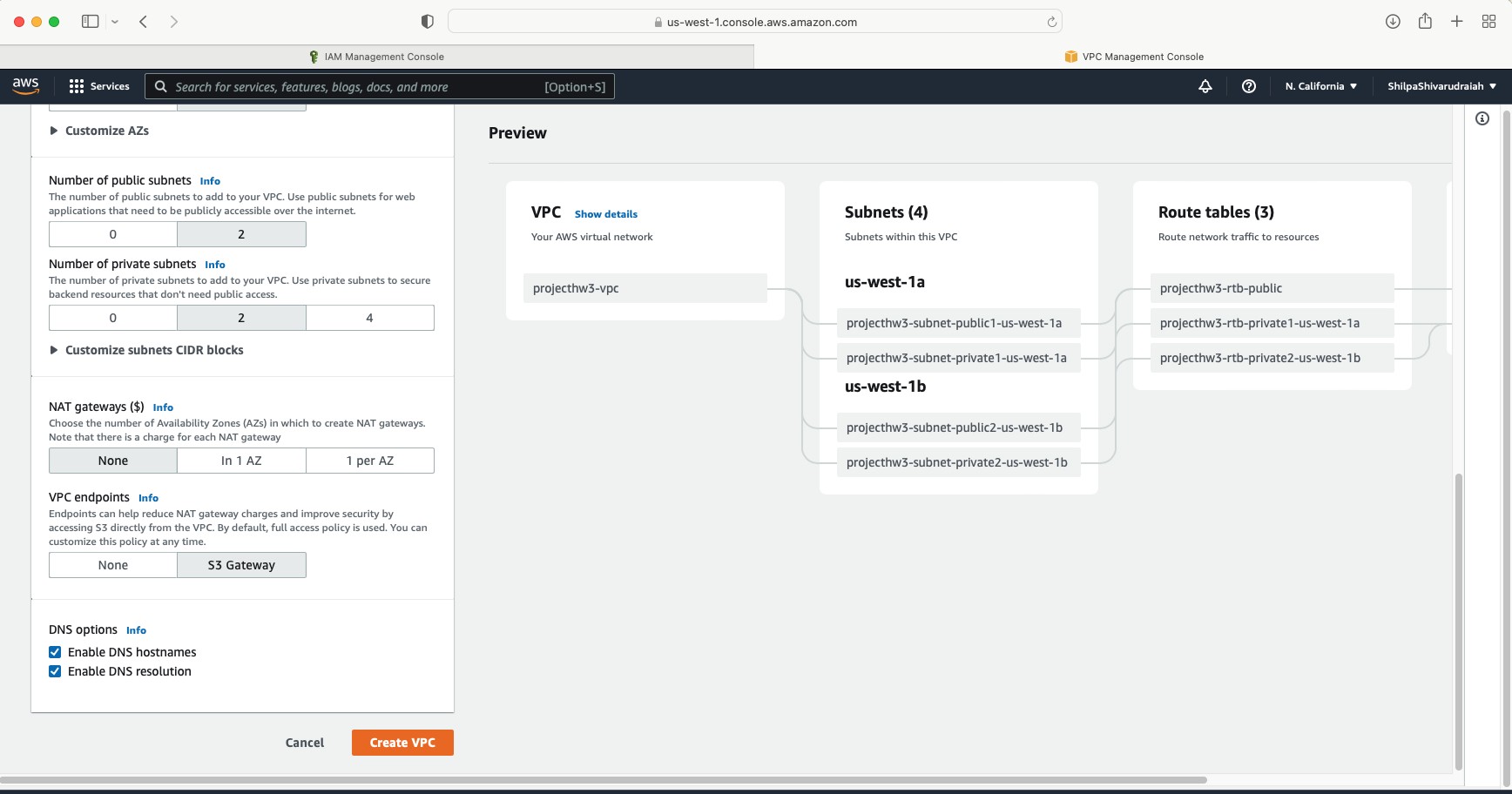
},

"Action": "sts:AssumeRole"

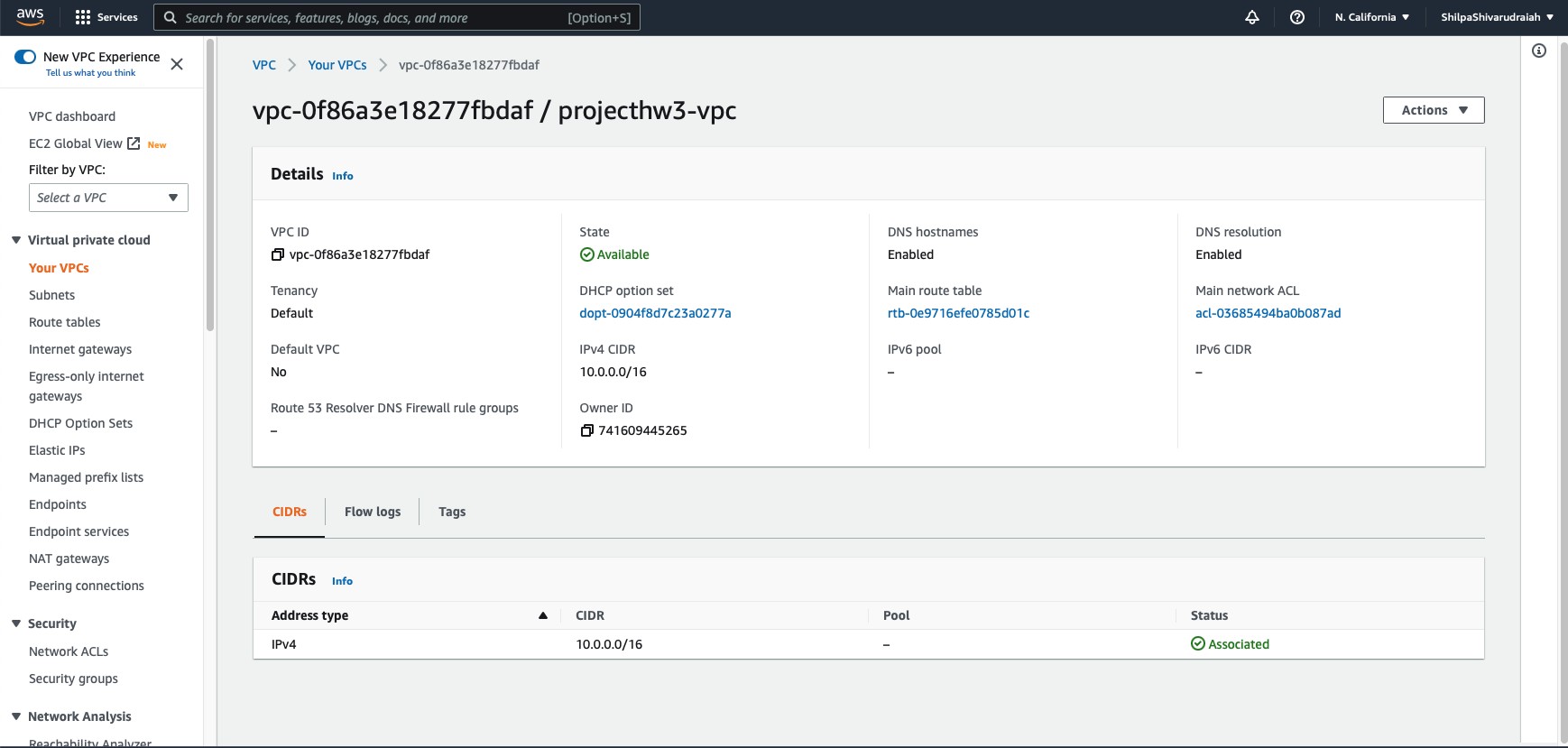
Step 16 : Setting up DNS in VPC by creating the VPC



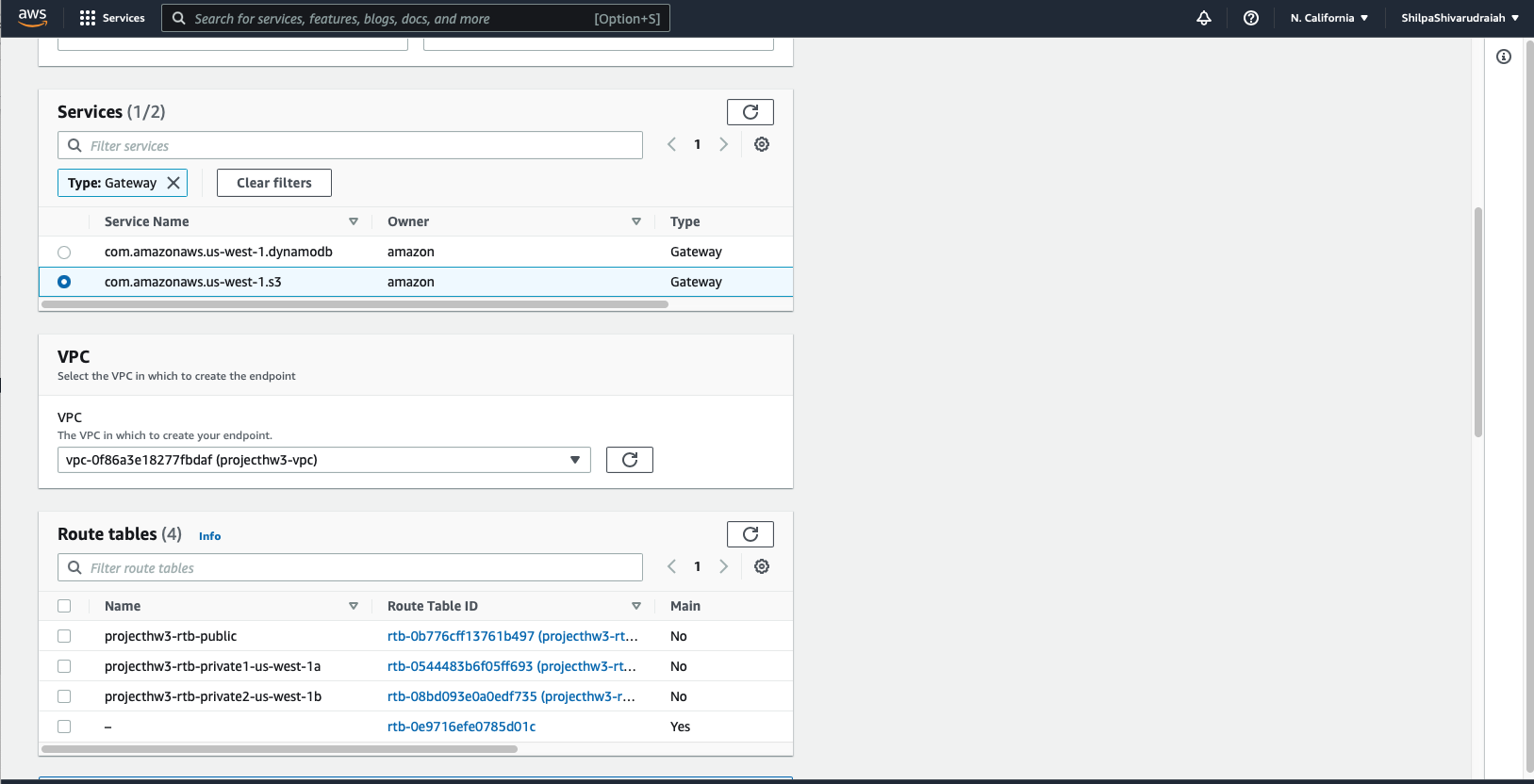
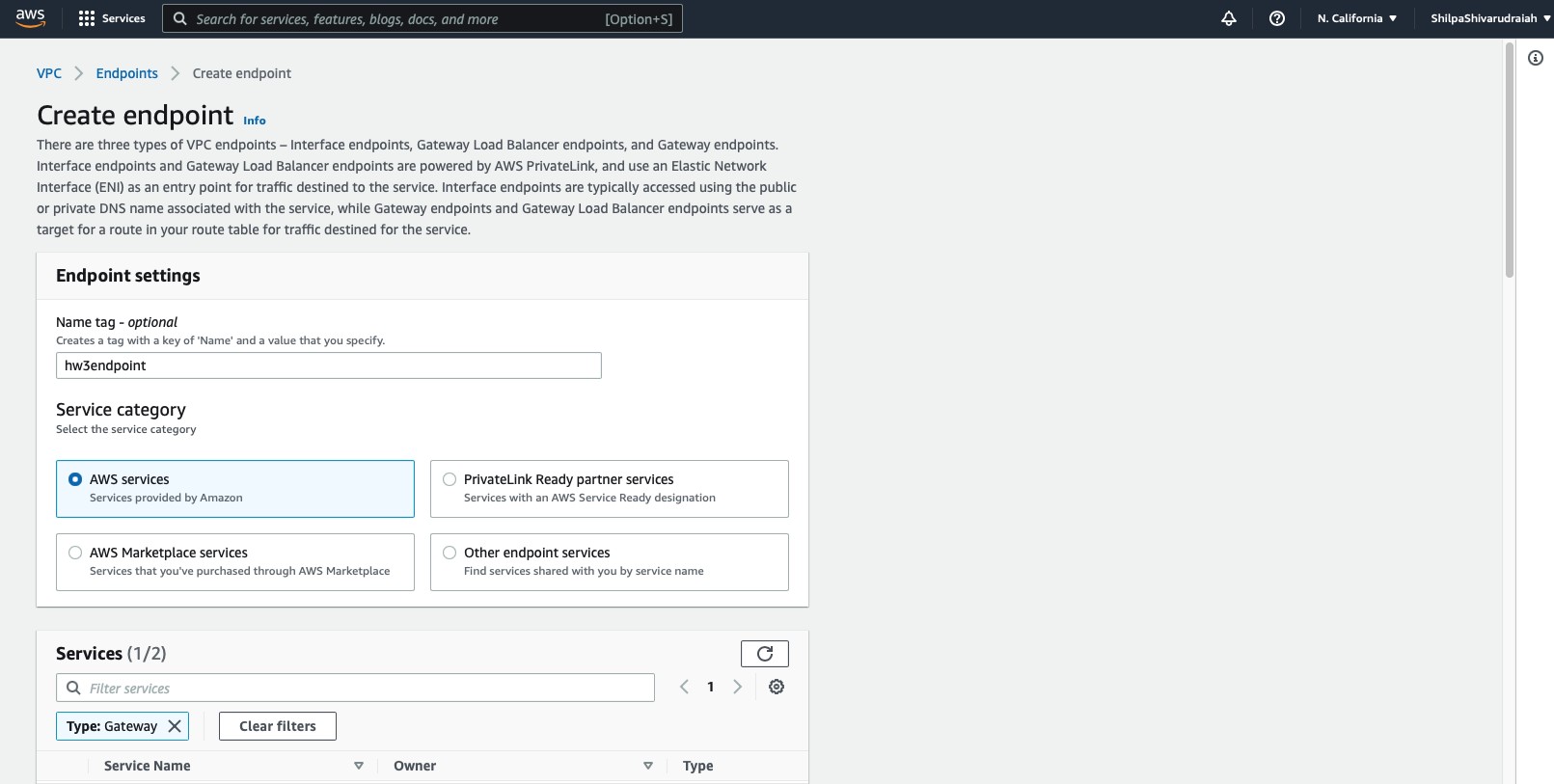
Step 17 : Ensure that DNS hostnames and DNS resolution are both enabled in your VPC

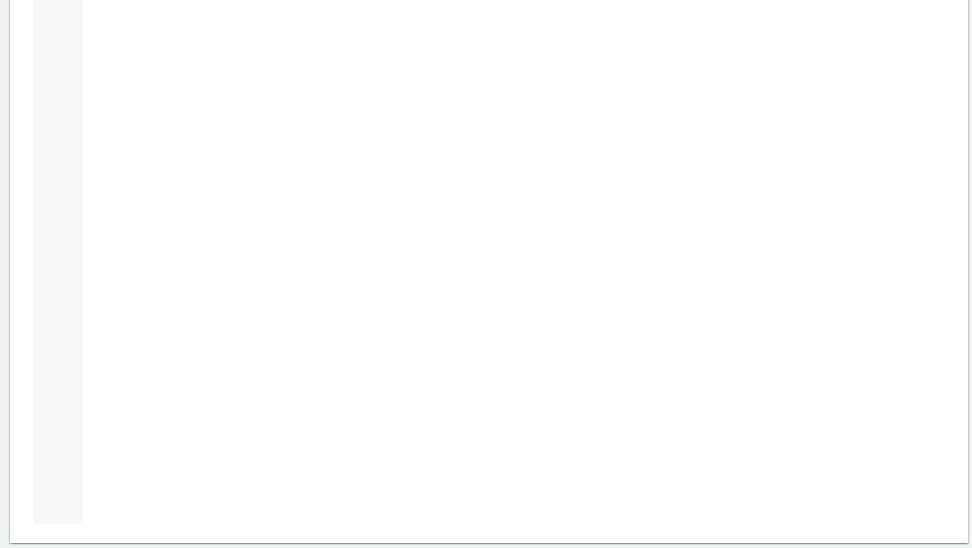


Step 18 : DNS is been enabled and created



Step 19: Setting up environment to access data stores by creating the Endpoints, and following the means to make an Amazon S3 VPC endpoint of type Gateway



a s :::SC!r<l(C!I Q ' ,, l ' l - *.Q.* ® N callfom"' • shllpaShlvarudra"'h.



©

Tags

Key

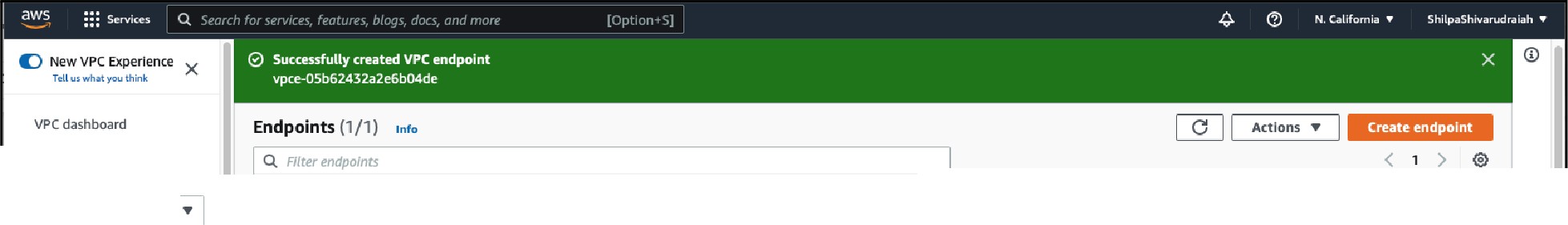
Value - *optional*

IQ\_N\_a\_m• x I I Q hw3endpo!m

I Add new tag

You<anadd49mor tags

cancel **FMfllJJ:[·@ffiH**



EC2 Global View - FllterbyVPC:

i se/ectoVPC

]

VPC endpoint ID: vpce-05b62432a2e5b04de X

Clear filters

!i

!i

Name

VPCendpolntlD

Endpoint type

* Virtual private cloud YourVPCs

Subnet5 Routl!'tables lnternetgatewa

Egress-only Internet gatl!'Ways

DHCP Option Set5 E(astlclPs

Managed prefix lists Endpoints

Endpoint services

NATgatl!'Ways Peering connections

hw3endpolnt

vpce-05b62432a2e6b04de

VPCID

vpc-Of86a3e18277fbdaf I proJecthw3-vpc

service name

com,amuonaws.us-west-1.s3

Gateway

Details

Endpoint ID

vpce-05b62432a2e6b04de

Status

€)All'allable

Creation time

Thursday, September 29, 2022, 15:50:00

**PDT**

Endpoint type

Gateway

VPCID

vpc-Of86a3e18277fbdaf (proJecthw3-vpc)

Status message

* security Networi:.ACls

SE'curltygroups

service name

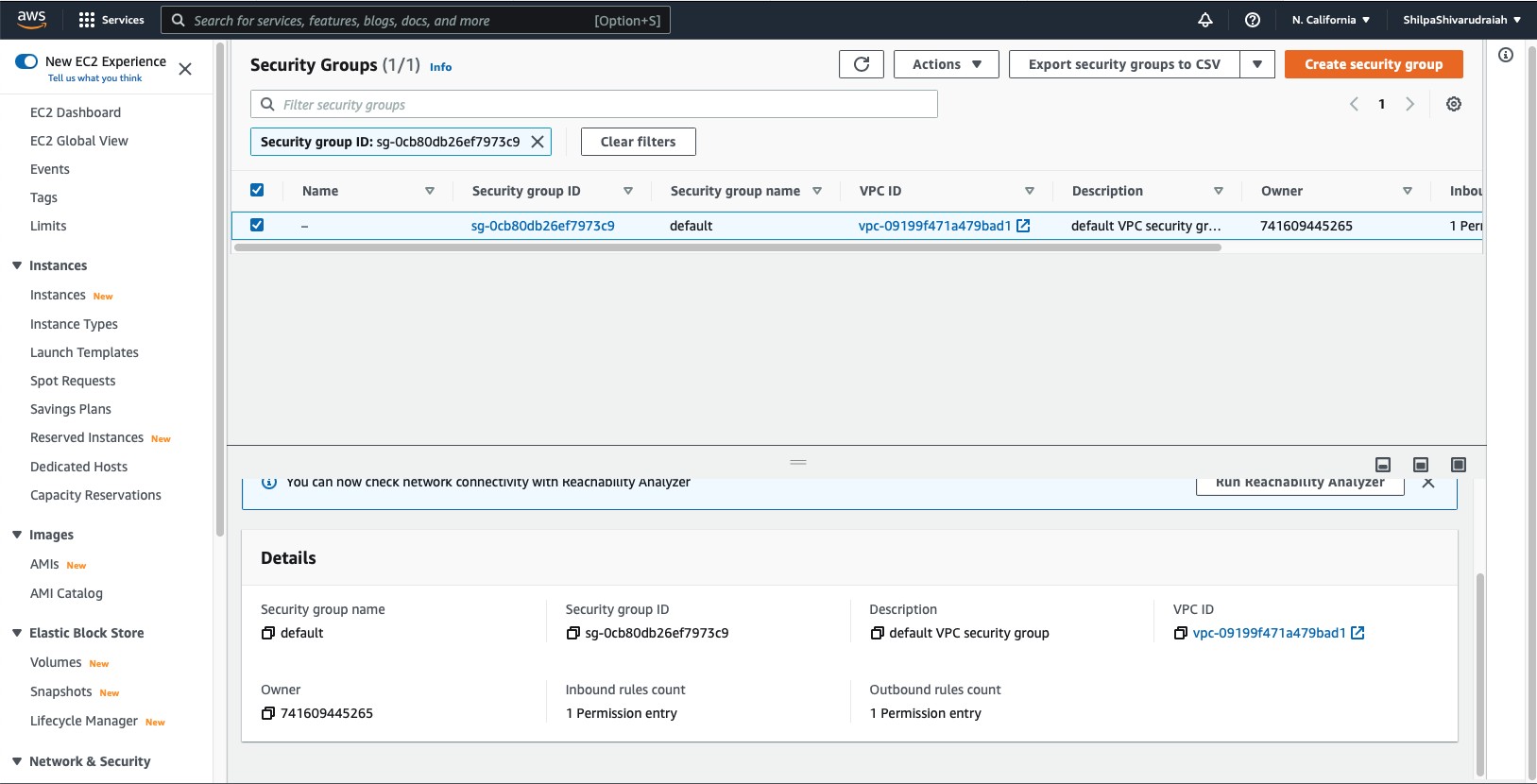
ol com.amazonaws.us-west-1.53

PrlVateDNSnamesenabled

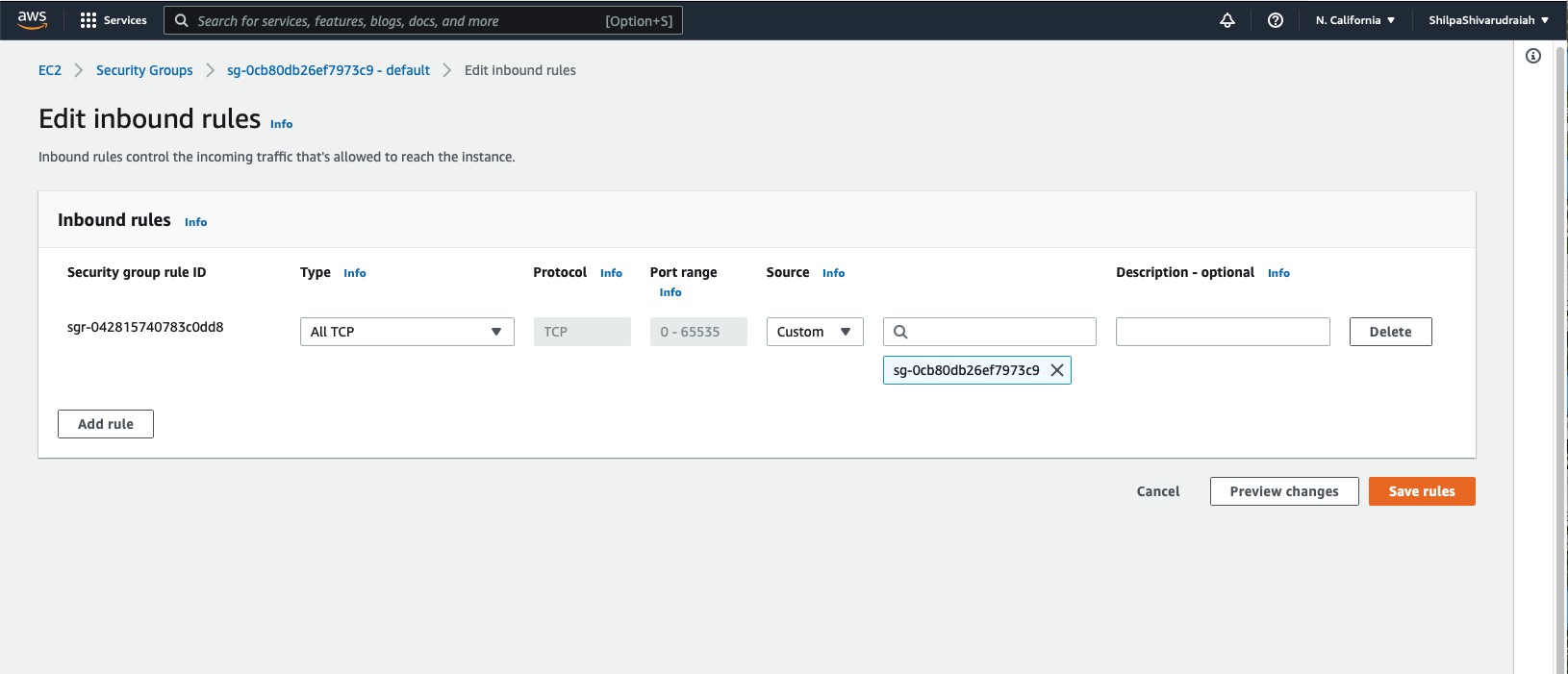
**No**

Step 20 : Setting Up a VPC to Connect to JDBC Data Stores.

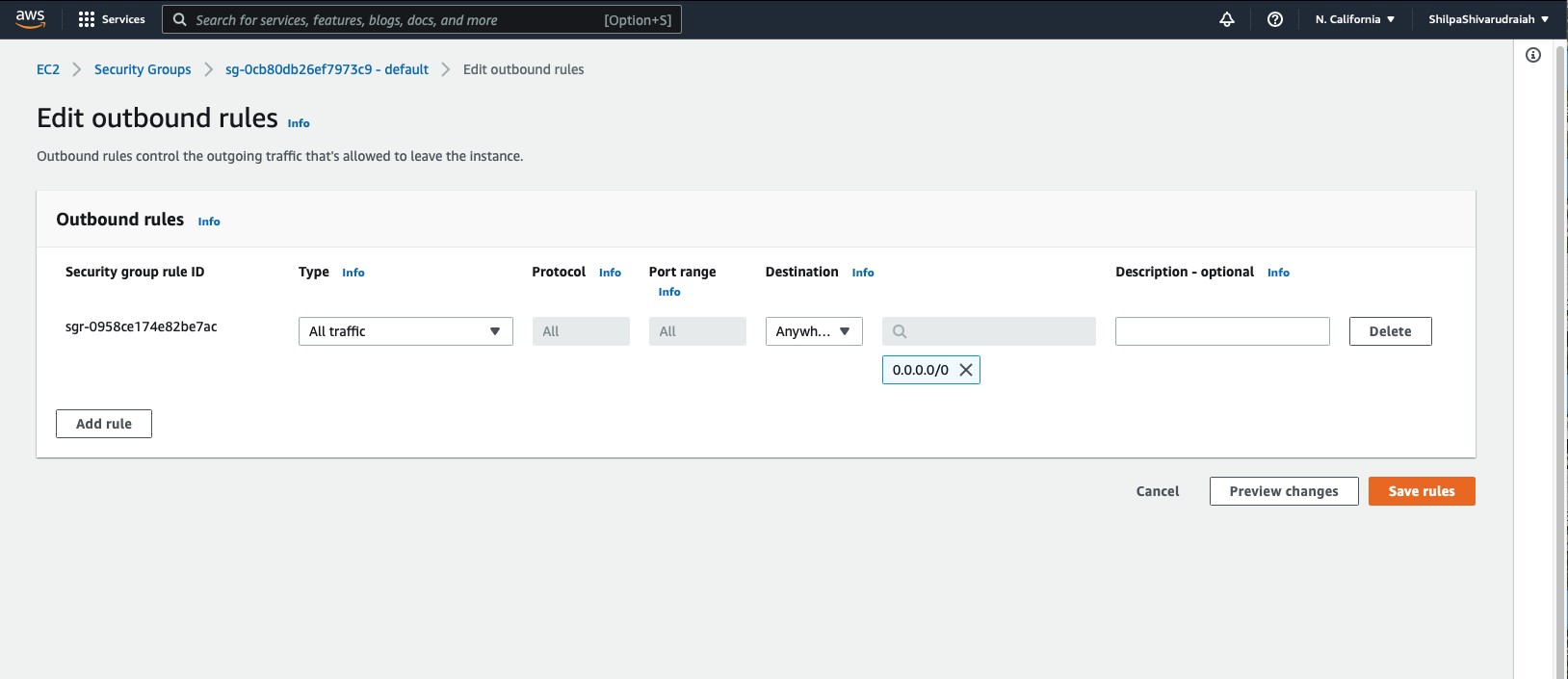
1. 1. Sign in to the AWS Management Console and open the Amazon Redshift comfort at <https://console.aws.amazon.com/redshift/>.
   1. In the left navigation pane, pick Clusters.
   2. Pick the cluster name that you need to access from AWS Glue.
   3. In the Cluster Properties segment, pick a security group in VPC security groups to permit AWS Glue to utilize. Record the name of the security group that you decided for future reference. Picking the security group opens the Amazon EC2 console Security Groups list.



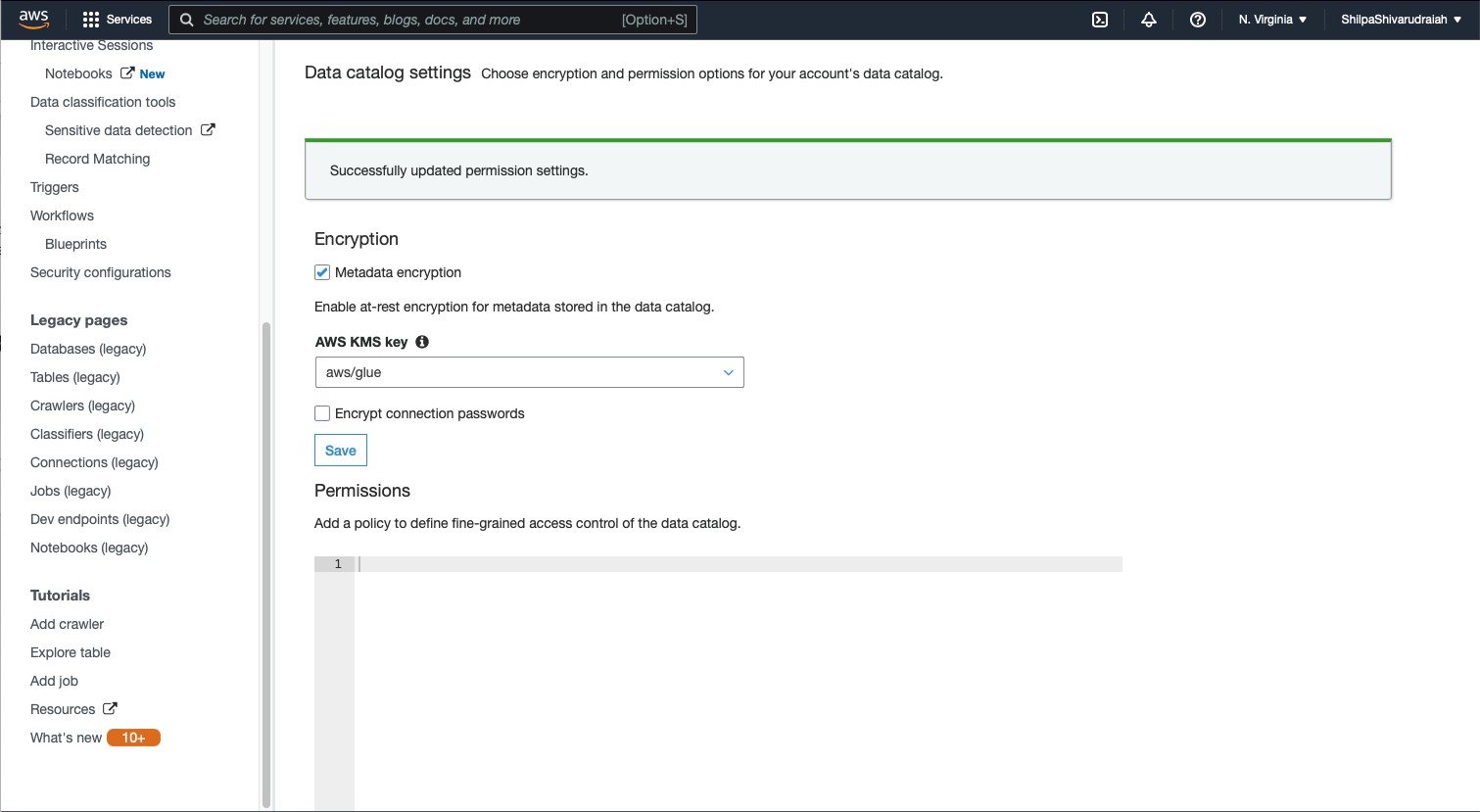
Step 21: Pick the security group to adjust and explore to the Inbound tab.



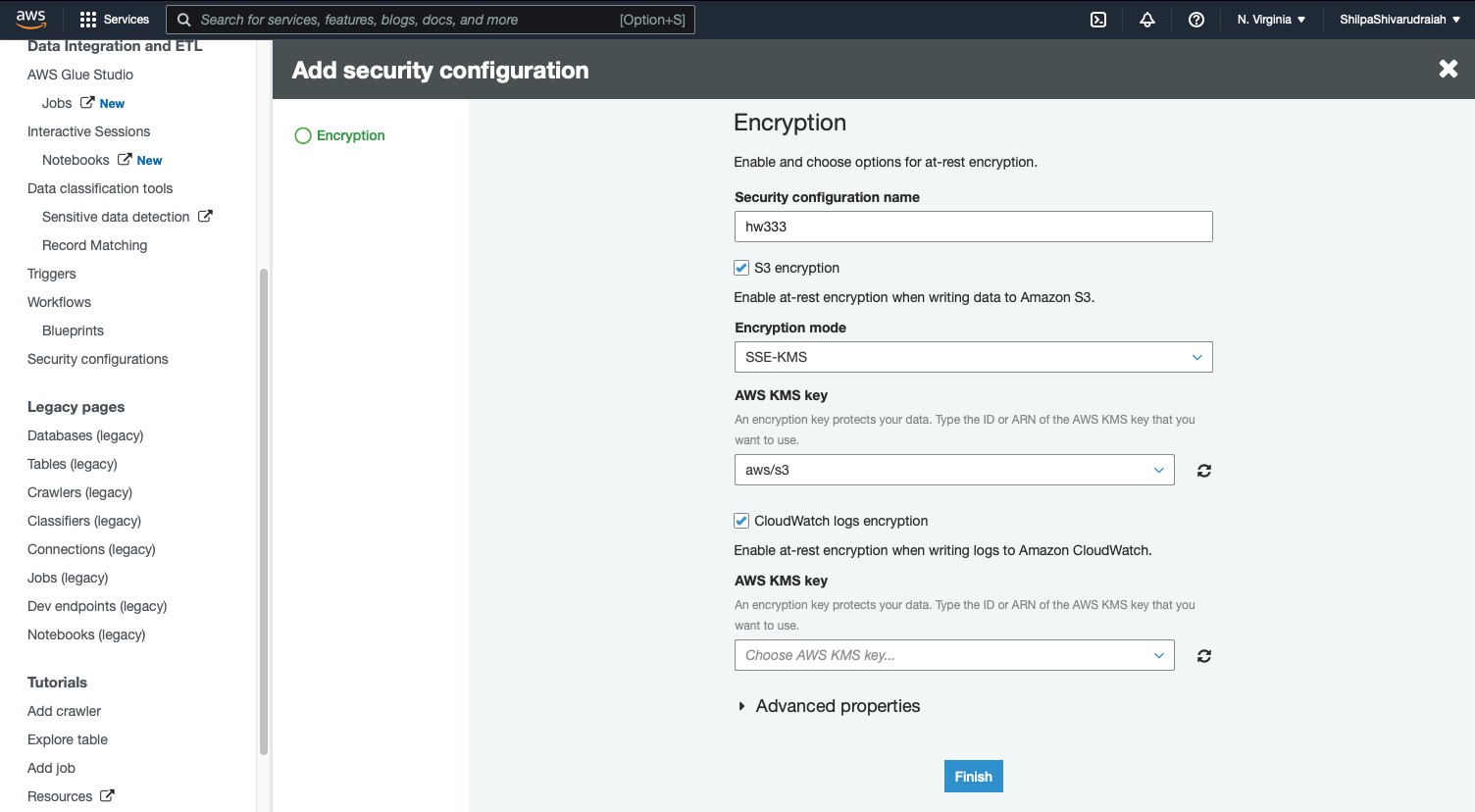
Step 22:Add a self-referring to run to permit AWS Glue parts to convey. In particular, add or affirm that there is a standard of Type All TCP, Protocol is TCP, Port Range incorporates all ports, and whose Source is a similar security bunch name as the Group ID.Add a standard for outbound traffic moreover. Either open outbound traffic to all ports or then again make a self-referring to lead where Type All TCP, Protocol is TCP, Port Range incorporates all ports, and whose Destination is a similar security bunch name as the Group ID. In the event that utilizing an Amazon S3 VPC endpoint, likewise add a HTTPS rule for Amazon S3 access. The s3-prefix-list-id is needed in the security bunch rule to permit traffic from the VPC to the Amazon S3 VPC endpoint.



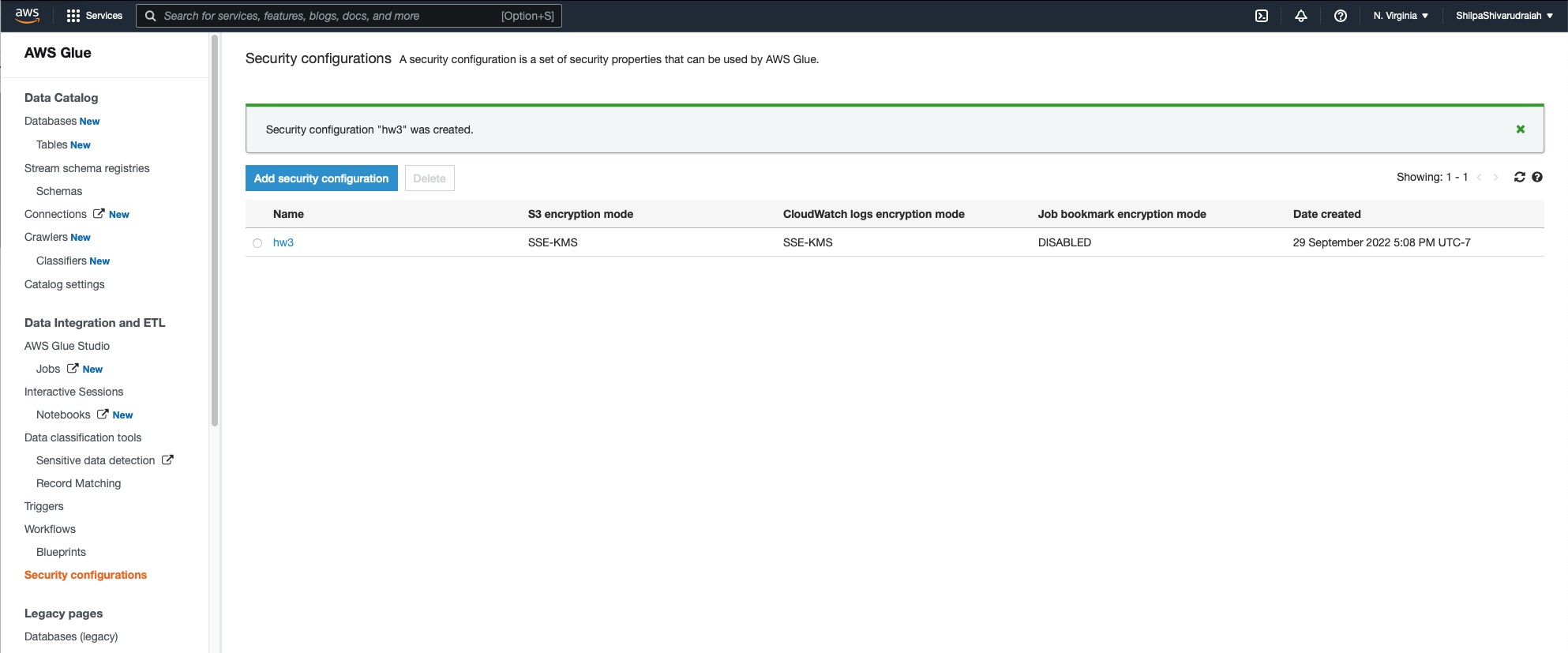
Step 23: Setting up encryption in AWS Glue. On the AWS Console select settings in the navigation pane. On the Data catalog settings page, encrypt your Data Catalog by selecting Metadata encryption. This option encrypts all the objects in the Data Catalog with the AWS KMS key that you choose.For AWS KMS key, choose aws/glue. You can also choose a customer master key (CMK) that you created.



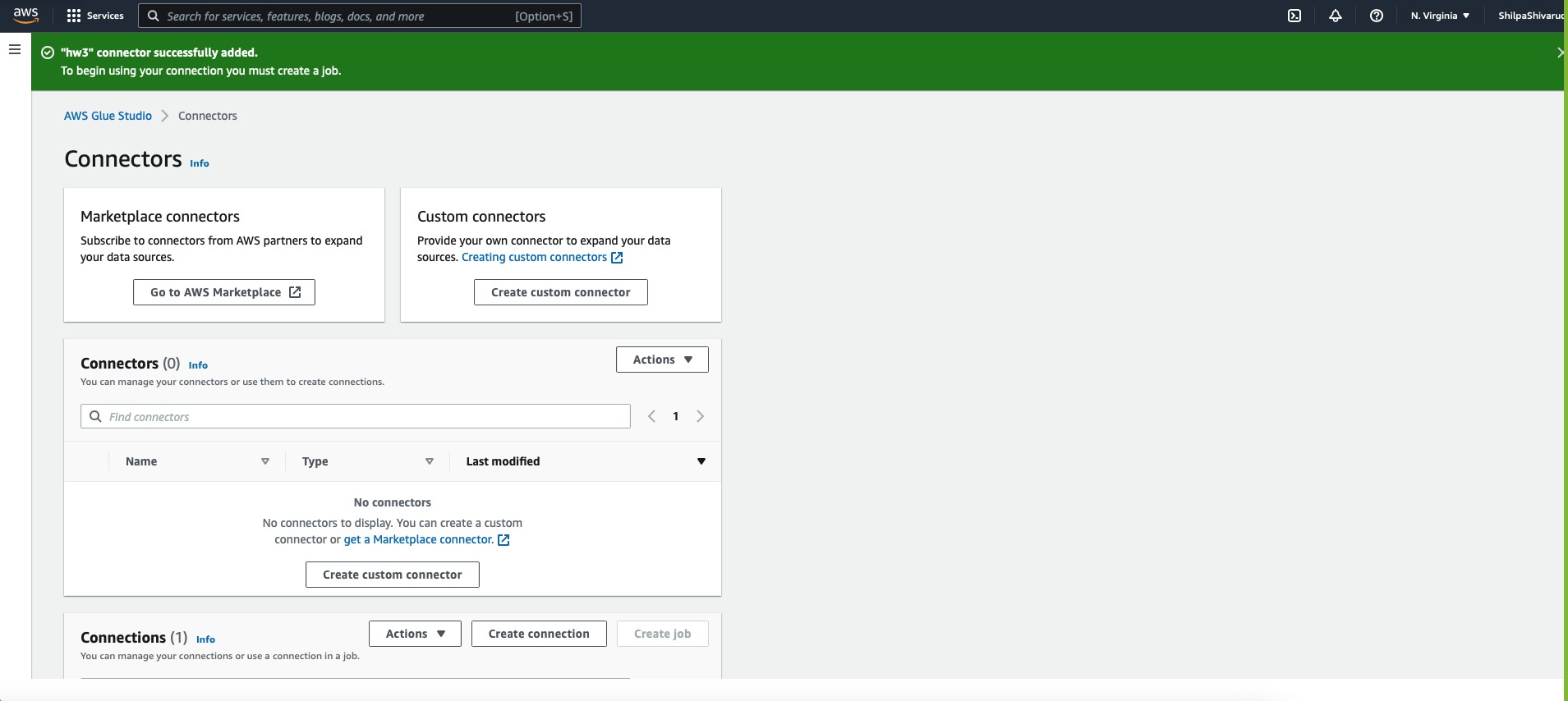
Step 24: In the navigation pane, choose Security configurations. A security configuration is a set of security properties that can be used to configure AWS Glue processes. Then choose Add security configuration name ‘hw333’.



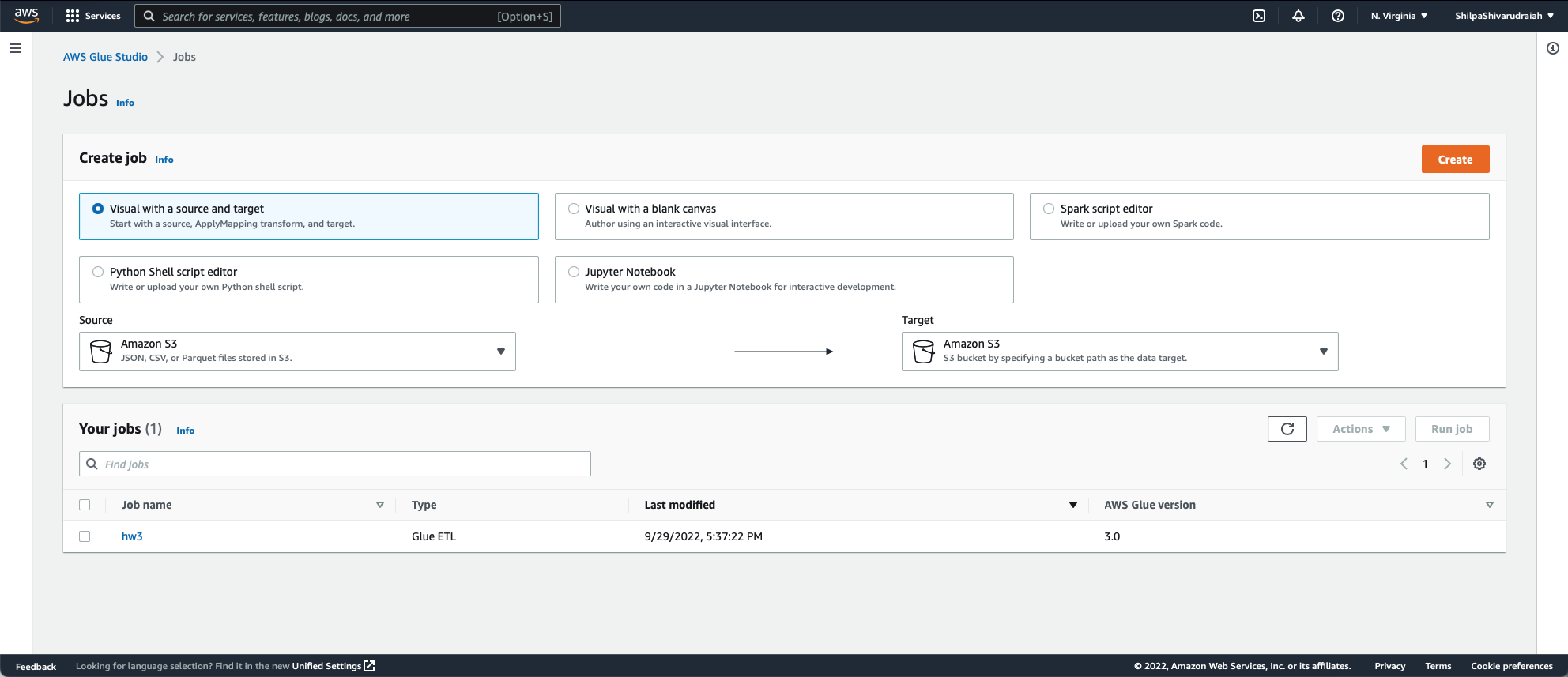
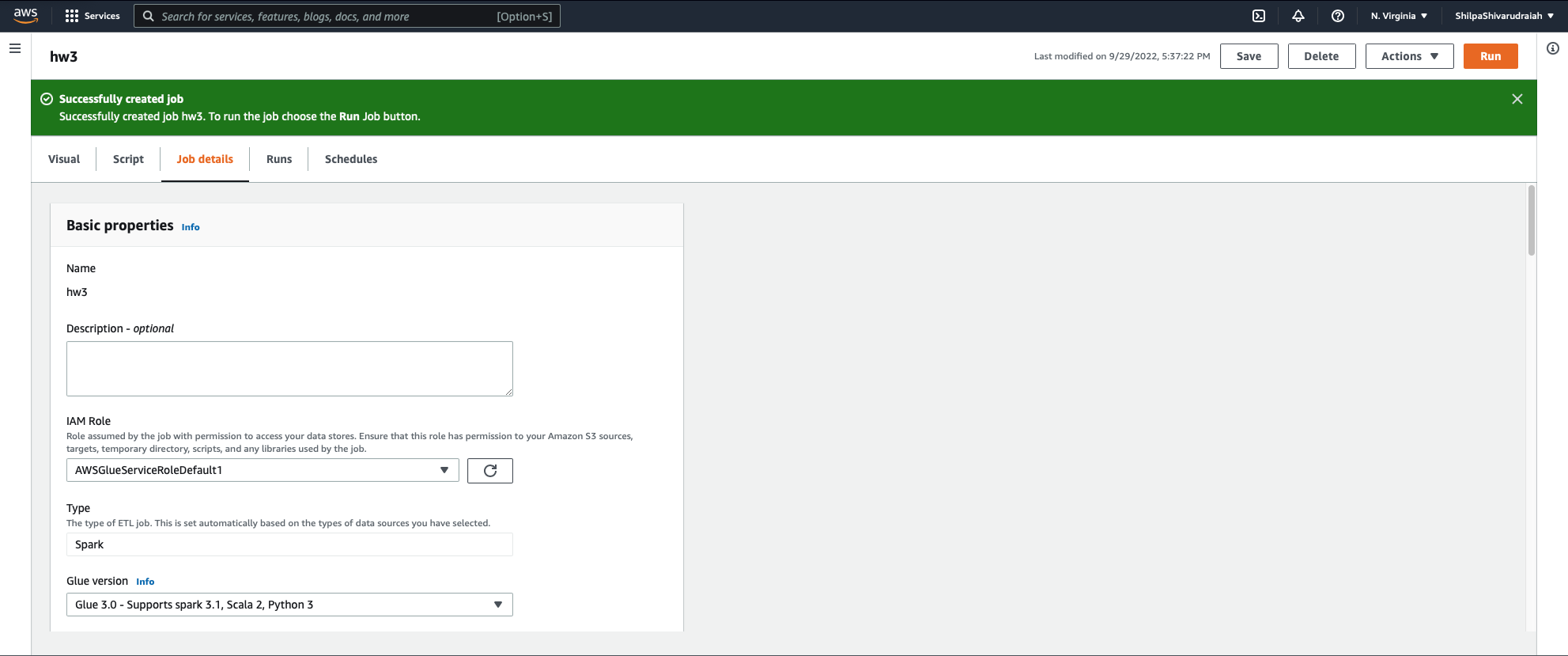
Below is the screenshot of KMS Key.



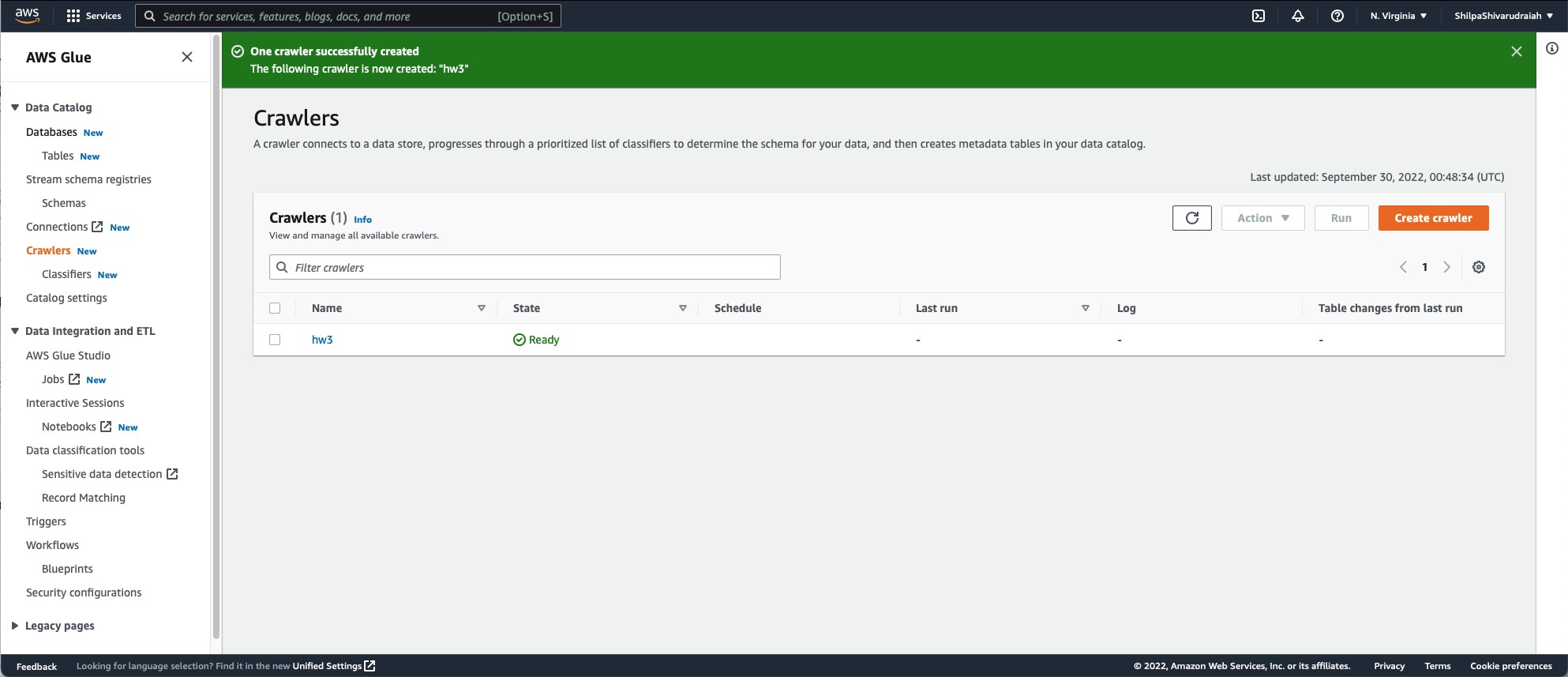
Step 25: Add connection to create a connection to the Java Database Connectivity (JDBC) data store ( this is the target of our ETL job)



Step 26: In the navigation pane, pick Jobs.Choose Add job to create a job that transforms data.In the job definition, choose the security configuration that you created.On the AWS Glue console, run your job on demand. Verify that any Amazon S3 data written by the job, the CloudWatch Logs written by the job, and the job bookmarks are all encrypted. Type is Spark for ETL.

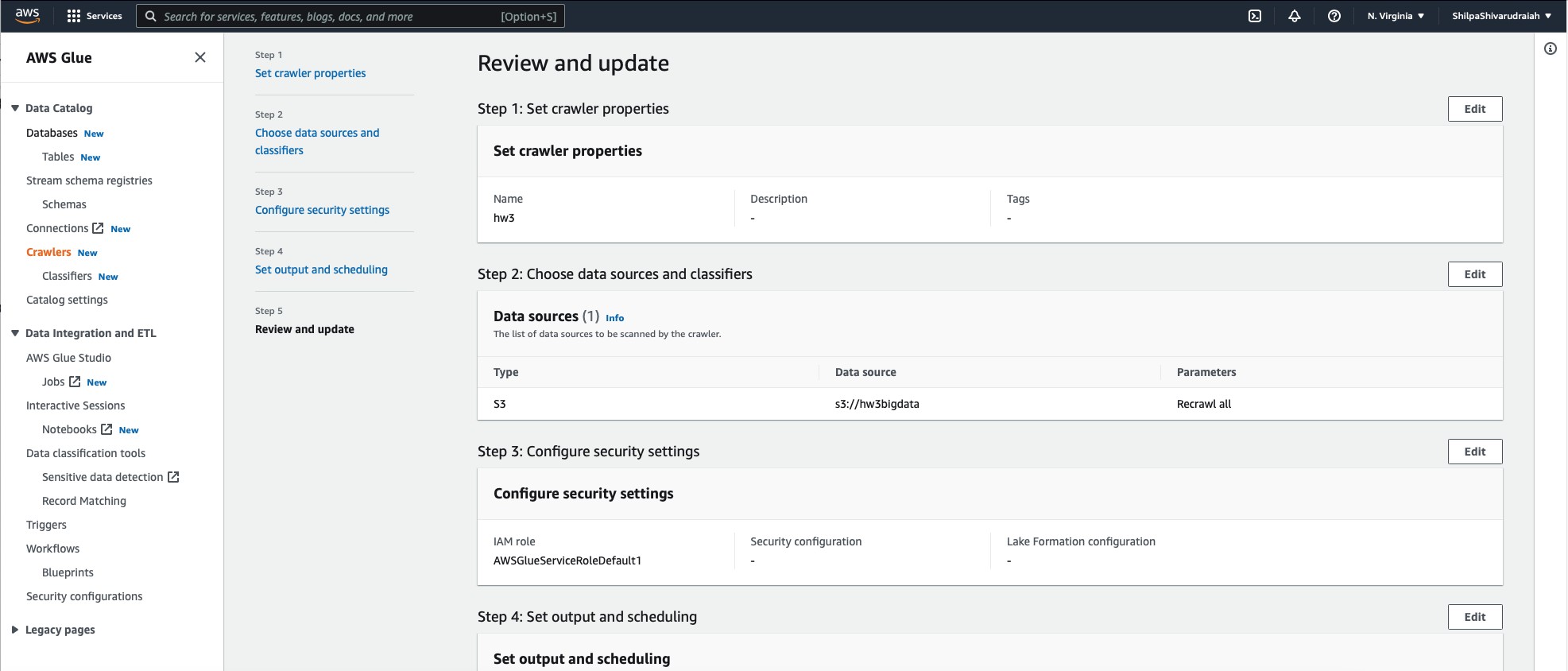


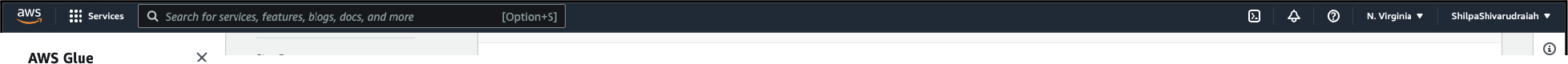
Step 27: Add crawler wizard from the list of tables or the list of crawlers



Step 28: On the AWS Glue service console, on the left-side menu, choose Crawlers.

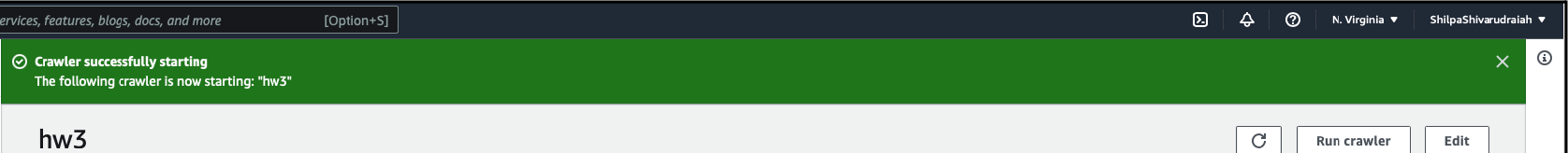
* 1. On the Crawlers page, choose Add crawler. This starts a series of pages that prompt you for the crawler details.For the crawler source type, choose Data stores and choose Next.
  2. Now let's point the crawler to your data. On the Add a data store page, choose the Amazon S3 data store.
  3. The crawler needs permissions to access the data store and create objects in the AWS Glue Data Catalog. To configure these permissions, choose Create an IAM role.
  4. The IAM role name starts with AWSGlueServiceRoleDefault1 and in the field, you enter the last part of the role name
  5. Verify the choices you made in the Add crawler wizard.After you have reviewed the information, choose Finish to create the crawler.







|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Conflguresecur1tyse111ngs | Descr1pllon  Step 2: Choose data sources and classifiers Datasources{1) info  Tho <t of <ht• sour«< to be «ann«!by tho rrawlor.  s3://hw3blgdala  Step 3:Configuresecuritysettings Configure security settings  IAMrole Secur1tyconflgura11on  AWSGluesuvlceRoleDefaul!l  Step 4: Set output and scheduling Set output and scheduling  Database Tablepreflx-opr/onor  hw3updated | "'' |  |
| * Da1acatalog |  |  |  |
|  | SelOUIPUlindschedullng |  |  |
| s1reamschemareg!s1r1es |  |  |  |
|  | Revlewandupda1e |  |  |
| Catalogsetllngs |  |  |  |
| * Da1a1n1egra1lonandETL |  |  |  |
| AWSGlueSludlo |  |  |  |
| Jobs Now |  |  |  |
|  |  | Lake Formation configuration |  |
| Nolebooks[!l Now |  |  |  |
| Sensltlvedatadetectlon[!l |  |  |  |
| Record Matching |  |  |  |
| Triggers |  |  |  |
| WorHlows |  |  |  |
| Blueprl IS  Security configurations |  | Maxlmumtablethreshold-oprlonor | Schedule On demand |
| * Legacy pages |  |  |  |
|  | | | | |



AWSGlue

X

* Data catalog

Pagelastupdated:SeptemtJer30,2022,00:S6:48(UTC)

Streamschemaregktr1es

Crawler properties

11\Mrole AWSGlueServlceRoleDefaultl [!l

Database hw3updated

Catalog settings

Descr1p!lon

Secur11yconflgura11on

Lake Formation configuration

Table prefix

* DalalntegratlonandETL

Jobs[!lNow

Notebooks[!l Now

* Advanced settings

Sensltlvedatadetectlon[!l Rec rdMatchlng

Triggers Workflows

8lueprlnt,;

Securltyconflguratl ns

* Legacypages

CreateslngleschemaforeachS3palh False

Inherit schema from table False

Schemaupdateslnthedatastor@

Updattthe1abledfflnltlonlntheda1aca1alo9

ObJectdelMlonlnthednastore Mark1hetableasdeprera1edlnthedatacatalog.

Repeatcrawl.sofS3datastores Crawlallfold@rsagalnwl!heverysubsequentcrawl.

Classifiers Tags

re«!b ,k

Un,A«! Sott"'9•1:!l

c, 2022 Amazon Wob Sem<<'> In< or t< ffil.,,,e, >rr,acy Term, Cook!o profercr><«



AWSGlue

hw3updated

Pagelastupdatec:September30,20Z2,01:01:Sl(UTC)

* Dan catalog

Database properties

Streamschemareg!stries

Description

hw3updatad

CrNtedon(UTC)

Saplamber30,2022,00:47:47

UIS!upd1m,d:Sep1embft"30,Z0Z2,01:0l:SZ(UTC)

Cat.!logsettlngs

* DanlnlegratlonandETL AWSGlueStudlo

J,bsl!:Ntw

Tables(1)

Q *F//tl!f'tables*

Notebooks Now

□

[ 0 hw3blgd111a

CQJ Delete I Addtilbluuslngcr•wler 1rm

< 1 > ®

Deprecaied

hw3updated

s3://hw3blgdalll/

Sensltlvedatadetectlon Record Matching

Triggers

Wor'<flows

Blueprints Securltyconflgurallon,

* Legacypages

reedbll<k

Un,Aed5ol'."'9•1!:

C,2022 ,.,,,.,onWobse,.,..1<<'> lo< ort> ffil"''"' Privacy Term, Cook!oprofercr><«

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
| AWSGlue   * Data catalog   Streamschemareg!stries    Catalog settings   * Da1a1n1egra1lonandETL AWSGlueStudlo   Jobsl!:Now  Notebooks Now  Sensltlvedatadetectlonl!: Rec:ordMnchlng  Tnggers  Workftows  Blueprints Securltyconftguratlons   * Legacypages | X | **nw**  Pagelastupdated:September30,20Z2,01:04:34(UTC)  Crawler properties  IAMrole AWSGlueServlceRoleDefaultl[!'.'.:  Description Sec:urityconflgura11on  I ► Advanced settings  Classifiers Tags  Crawlerruns(1)  *IO.F//tl!f'doto*  Stantlme(llTC) • Endtlme(llTC)  I O September30,2022,01:00:21 September30,Z022,0l:Ol:26 | I | Datatlase hw3updated  Lake Formation configuration  I I[) *Fllterbyodoteandtimerange*  €)completed | [QJ  I | L\_CU L"'"""""J L""J  Table prefix  Stoprun View CloudWatch logs I!: View run details  < 1 > ®  v Table changes  0.03744400C000000005 - ta\_bl\_e c\_hange, o\_p\_art\_ltlon\_ch\_a\_ng\_e\_ I |
|  | | | | | | |

I a!!\_S m Sc"'""' a, [) .0. (2) N Vlf'9",. • Shllp•Slwa, h •



CloudWatch X CloudWatch) Loggroups) /aWi-glue/crawlers) hw3

Log events

You can use the filter bar below to search for and much terms, phra'>!'S, or values In your log events. Learn mar. about filter patterns

□ -astext @J

* + - Logs

**ii:i!l!: d&ii\M**

Io.\_·,\_,,.\_,\_,,\_,\_,\_,... .,\_,.\_, ,,\_,\_,.\_,\_'"-'-"-"\_c x II clear lh 1Zh Custom I[) I @

Loggroups Logslnslgh:s

X-Raytraces

* Appllcatlonmonltorlng
* Insights

Tlmestamp

2a22-09-29Tl8:01:23.9S0·07:00

2a22-09-29Tl8:01:26.039·07:00

2a22-09-29Tl8:01:37.1!46·07:00

2a22-09-29Tl8:01:37.!160-07:00

2a22-09-29Tl8:01:37.!61-07:00

2a22-09-29Tl8:03:40.2S9·07:00

Message

[81l94721>4-421>4-4198-b39f·l>4d2866S63M] 8ENCHMAR : Running Start Crawl for Crawler hw3

[H:"4fl1>4·4l1>4·41Y3·b.:l f-04dl3bt»b.:lec] HNUfoW(\ : llasnflcanoo c.,,.lete, IW'ttlng results to aatacase nw:ul)(lated

[81l94721>4-421>4·4198·b39f·l>4d2866S63ec] INFO: Crawterconfigu,.,,dwithSchemaOlar, Policy {"UpdoteBehavior":"UPDATE\_IN\_DATIJIASE"," l•tdehavior":"DEPRECATI\_IN\_DATABASE"}. [81l94721>4-421>4-4198-b39f-1>4d2866S63•c]INFO:Createdtablehli.3bigdataindatabasehw3"pdoted

[81l94721>4-421>4-4198-b39f-1>4d2866S63M] 8ENCHMAR : Finished wri.tirig to Catalog

[81l94721>4-421>4-4198-b39f-1>4d2866S63M] INFO: RJn Suomory For TABLE:

[!ll94721>4-421>4-4198-b39f-b4d2!66563M] INFO: A : 1

[!ll94721>4-421>4-4198-b39f-b4d2!66563 c] 8ENCHMAR( : Crawt r has finish<d running and is in stat READY



Settings Gettlngs1artea

reed .,,k Un,Aed 5ott"'9•C!'.'.l c, 2)22 Am.,onWob ...,,.,..",.. In< or t> ffih.,,e, Priv•<) Term, Cook!oprofercr><«