

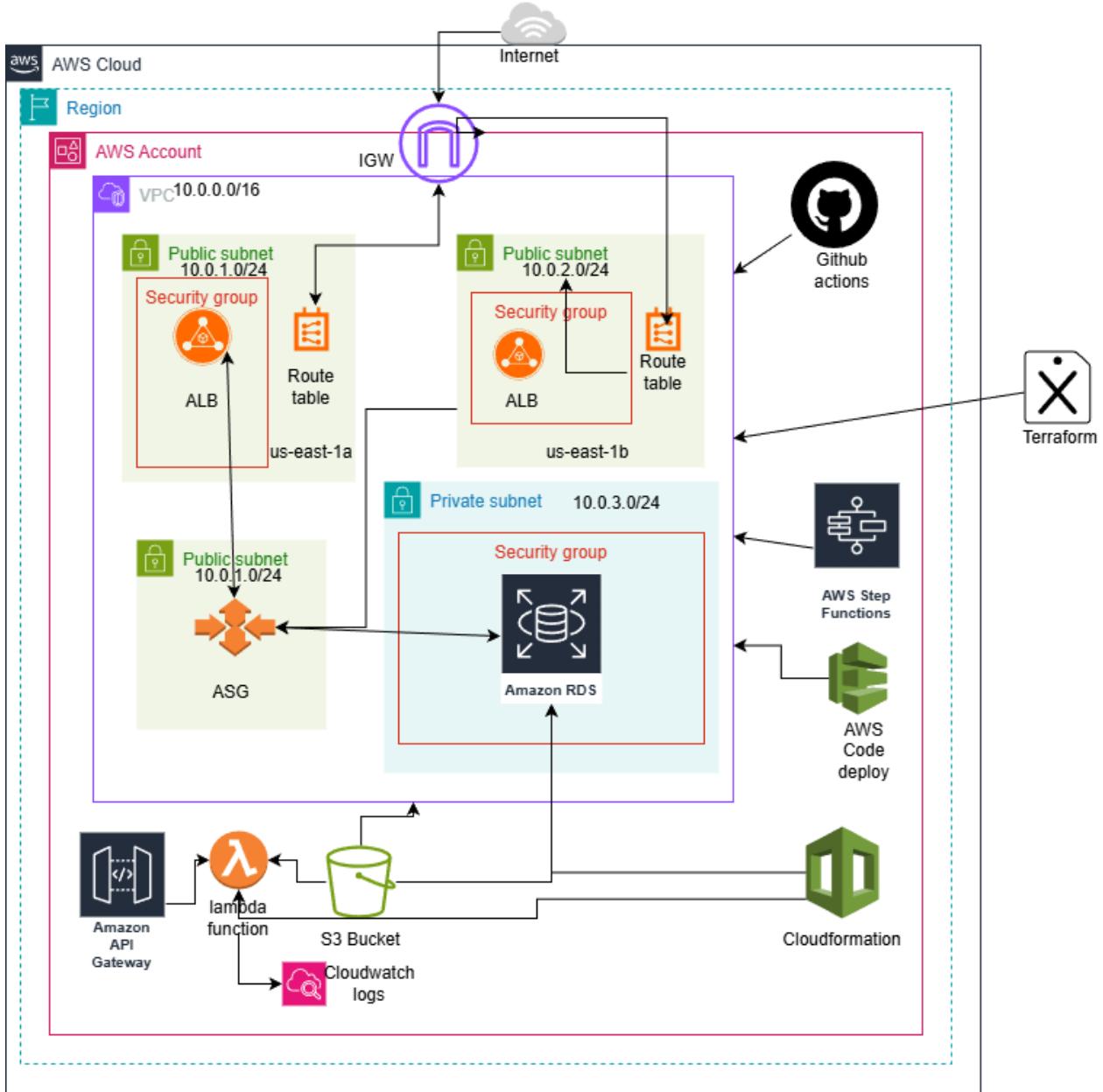
IS698 - Final Project

Deploying a Scalable AWS Architecture with Infrastructure as Code

Submitted By-
Sai Sumanth Reddy Kachi
Campus ID: DC93457

Project Requirements

1. Architecture Design



With this architecture, various AWS capabilities come together to form a system that is both reliable and easy to scale. It all starts with the construction of the VPC. In this instance, Terraform is employed to implement this layer so that consistency is achieved each time it is created. In the structure created by the construction of the VPC, subnets and internet gateways exist within the network design for enhanced organization.

The public subnet handles the web-facing part. An Auto Scaling Group launches EC2 instances running the application, and then an Application Load Balancer is placed on top to manage traffic. These components function together to ensure that the application is still responsive even with increased traffic. Security groups are layered around each element to filter incoming and outgoing traffic appropriately.

In the private subnet environment, the database on Amazon RDS is maintained out of reach from the internet and is allowed limited connectivity from the application layer. This provides a significant safeguard for the data.

It is further supplemented by event-driven processing with the use of an S3 bucket and a Lambda function. Every time a file is uploaded, Lambda records the events on AWS CloudWatch. AWS CloudFormation facilitates the deployment process for core elements, while services including GitHub Actions and CodeDeploy simplify the process for updates and automation.

A. Infrastructure Deployment

1. Terraform: Networking and Infrastructure

Terraform was used to provision the entire network layer of the architecture. This included the VPC, subnets, route tables, internet gateway and security groups. Using Infrastructure-as-Code ensured that networking resources were consistent and repeatable across deployments.

A Terraform configuration was created to define:

- VPC (10.0.0.0/16)
- Public subnets for EC2 and autoscaling
- Private subnet for RDS
- Internet Gateway for outbound access
- Route tables for public routing
- Security groups for EC2 and RDS traffic control

```

E: > Sumanth UMBC > IS698-Project > terraform-698 > main.tf
 1 ######
 2 # Terraform and Provider
 3 #####
 4
 5 terraform {
 6   required_version = ">= 1.5.0"
 7
 8   required_providers {
 9     aws = {
10       source  = "hashicorp/aws"
11       version = "~> 5.0"
12     }
13   }
14 }
15
16 provider "aws" {
17   region = var.aws_region
18 }
19
20 #####
21 # VPC
22 #####
23
24 resource "aws_vpc" "main" {
25   cidr_block          = var.vpc_cidr
26   enable_dns_hostnames = true
27   enable_dns_support    = true
28
29   tags = {
30     Name = "${var.project_name}-vpc"
31   }
32 }
33
34 #####
35 # Internet Gateway
36 #####
37

```

[main.tf](#)

```

E: > Sumanth UMBC > IS698-Project > variables.tf
 1 variable "aws_region" {
 2   description = "AWS region to deploy into"
 3   type        = string
 4   default     = "us-east-1"
 5 }
 6
 7 variable "project_name" {
 8   description = "Base name for tagging resources"
 9   type        = string
10  default     = "cloud-arch-project"
11 }
12
13 variable "vpc_cidr" {
14   description = "CIDR block for the VPC"
15   type        = string
16   default     = "10.0.0.0/16"
17 }
18
19 variable "public_subnets" {
20   description = "CIDR blocks for public subnets"
21   type        = list(string)
22   default     = ["10.0.1.0/24", "10.0.2.0/24"]
23 }
24
25 variable "private_subnets" {
26   description = "CIDR blocks for private subnets"
27   type        = list(string)
28   default     = ["10.0.11.0/24", "10.0.12.0/24"]
29 }
30
31 variable "availability_zones" {
32   description = "AZs for subnets"
33   type        = list(string)
34   default     = ["us-east-1a", "us-east-1b"]
35 }
36

```

[variables.tf](#)

```

E: > Sumanth UMBC > IS698-Project > terraform-698 > outputs.tf
 1 output "vpc_id" {
 2   value      = aws_vpc.main.id
 3   description = "ID of the main VPC"
 4 }
 5
 6 output "public_subnet_ids" {
 7   value      = [for s in aws_subnet.public : s.id]
 8   description = "IDs of public subnets"
 9 }
10
11 output "private_subnet_ids" [
12   value      = [for s in aws_subnet.private : s.id]
13   description = "IDs of private subnets"
14 ]
15
16 output "alb_sg_id" {
17   value      = aws_security_group.alb_sg.id
18   description = "ALB security group ID"
19 }
20
21 output "web_sg_id" {
22   value      = aws_security_group.web_sg.id
23   description = "Web server security group ID"
24 }
25
26 output "rds_sg_id" {
27   value      = aws_security_group.rds_sg.id
28   description = "RDS security group ID"
29 }
30

```

outputs.tf

Terraform files are then executed using the following commands:

terraform init
terraform plan
terraform apply

After **terraform apply**, the AWS console confirms that all networking components were created successfully.

2. Cloudformation: EC2 and RDS with Static Web Page Deployment

CloudFormation was used to automate the creation of EC2 instances and an RDS MySQL database.

1. Two YAML templates were created:
 - **EC2 template** for launching the web server
 - **RDS template** for provisioning the MySQL database

```

E: > Sumanth UMBC > IS698-Project > cloudformation > ec2-alb.yaml
32   Resources:
33     WebServerSecurityGroup:
34       Type: AWS::EC2::SecurityGroup
35       Properties:
36         GroupDescription: "Web Server Security Group"
37         VpcId: !Ref VPC
38         SecurityGroupIngress:
39           - Type: "HTTP"
40             FromPort: 80
41             ToPort: 80
42             IpProtocol: "TCP"
43             CidrIp: 0.0.0.0/0
44         SecurityGroupEgress:
45           - Type: "HTTP"
46             FromPort: 80
47             ToPort: 80
48             IpProtocol: "TCP"
49             CidrIp: 0.0.0.0/0
50       Tags:
51         - Key: Name
52           Value: simple-ec2-public
53     WebServerInstance:
54       Type: AWS::EC2::Instance
55       Properties:
56         ImageId: !Ref AMI
57         InstanceType: !Ref InstanceType
58         KeyName: !Ref KeyName
59         SubnetId: !Ref PublicSubnetId
60         SecurityGroupIds:
61           - !Ref WebServerSecurityGroup
62         Tags:
63           - Key: Name
64             Value: simple-ec2-public
65         UserData:
66           Fn::Base64:
67             "#!/bin/bash
68             yum update -y
69             yum install -y httpd
70             systemctl enable httpd
71             systemctl start httpd
72
73             cat > /var/www/html/index.html << 'EOF'
74             <html>
75               <head>
76                 <title>AWS Simple EC2</title>
77               </head>
78               <body>
79                 <h1>Hello from a simple EC2 instance!</h1>
80                 <p>This instance was created using a basic CloudFormation template
81                 in a public subnet.</p>
82               </body>
83             </html>
84             EOF
85
86
87
88
89   Outputs:
```

Ec2-alb.yaml

```

E: > Sumanth UMBC > IS698-Project > cloudformation > rds-mysql.yaml
4   Parameters:
5
6     DBSubnetGroup:
7       Type: AWS::RDS::DBSubnetGroup
8       Properties:
9         DBSubnetGroupDescription: "Subnet group for RDS in private subnets"
10        SubnetIds:
11          - !Ref PrivateSubnet1Id
12          - !Ref PrivateSubnet2Id
13        # Name omitted so AWS generates a unique one
14
15     MyDBInstance:
16       Type: AWS::RDS::DBInstance
17       Properties:
18         # Let AWS pick a unique identifier automatically by omitting DBInstanceIdentifier
19         AllocatedStorage: 20
20         DBInstanceClass: db.t3.micro
21         Engine: mysql
22         DBName: !Ref DBName
23         MasterUsername: !Ref DBUsername
24         MasterUserPassword: !Ref DBPassword
25         VPCSecurityGroups:
26           - !Ref RdsSecurityGroupId
27         DBSubnetGroupName: !Ref DBSubnetGroup
28         MultiAZ: false
29         PubliclyAccessible: false
30         DeletionProtection: false
31         BackupRetentionPeriod: 0
32
33   Outputs:
34     DatabaseEndpoint:
35       Description: RDS database endpoint address
36       Value: !GetAtt MyDBInstance.Endpoint.Address
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
```

Rds-mysql.yaml

2. After creating these YAML files, stacks were launched through the AWS console by uploading the existing templates.
3. For the EC2 instance a static web page was deployed



Hello from a simple EC2 instance!

This instance was created using a basic CloudFormation template in a public subnet.

4. For the RDS, a database was created with the specified characteristics.

3. Database Backend Configuration

The EC2 server was configured to use the RDS MySQL database as its backend.

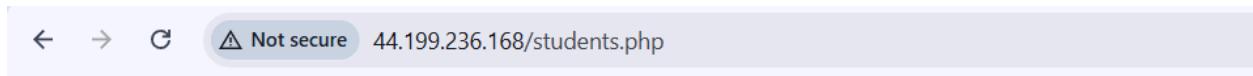
1. Connected to the EC2 instance via SSH using **putty.exe**
2. Installed Apache Web Servers with the following code:

```
sudo yum install httpd -y
sudo systemctl start httpd
sudo systemctl enable httpd
```
3. Uploaded the index.html file to **/var/www/html**
4. Created the database script file (db-check.php)

```
sudo nano /var/www/html/db-check.php
```

```
E:\> Sumanth UMBC > ls698-Project > db_check.php
 1  <?php
 2  error_reporting(E_ALL);
 3  ini_set('display_errors', 1);
 4
 5  $db_host = 'rds698-mydbinstance-qfjghsjrhrkc.cmtoeaqe4zjy.us-east-1.rds.amazonaws.com';
 6  $db_name = 'cloudprojectdb';
 7  $db_user = 'admin';
 8  $db_pass = '*****';
 9
10 $conn = new mysqli($db_host, $db_user, $db_pass, $db_name);
11
12 if ($conn->connect_error) {
13     echo "Database connection failed: " . $conn->connect_error;
14 } else {
15     echo "Database connection successful to database '$db_name' on '$db_host'." ;
16 }
17
18 $conn->close();
19 ?>
```

5. Inserted RDS endpoint, DB name, DB username and DB password to the script.
6. After checking the database connection, the Students data is added and a new students.php file is created.
7. After testing the script, the browser confirmed that EC2 and RDS are connected.



ID	Name	Major	Year
1	Ken Adams	Computer Science	2
2	Sumanth Reddy	Information Systems	3

4. Autoscaling Implementation

An Auto Scaling Group (ASG) was created to automatically add or remove EC2 instances based on CPU load. This allowed the architecture to scale dynamically.

1. A Launch Template is created using:
 - The same AMI
 - The same EC2 configuration
 - The same user data script used for the manual EC2 instance

```

E:\> Sumanth UMBC > IS698-Project > cloudformation > ! ec2-autoscaling.yaml
4 Parameters:
41
42 Resources:
43   WebLaunchTemplate:
44     Type: AWS::EC2::LaunchTemplate
45     Properties:
46       LaunchTemplateName: !Sub "${AWS::StackName}-web-lt"
47       LaunchTemplateData:
48         ImageId: ami-0fa3fe0fa7920f68e
49         InstanceType: t3.micro
50         KeyName: 698finalproject
51         SecurityGroupIds:
52           - !Ref WebSecurityGroupId
53         UserData:
54           Fn::Base64:
55             Fn::Sub:
56               #!/bin/bash
57               yum update -y
58               yum install -y httpd php php-mysqlnd
59               systemctl enable httpd
60               systemctl start httpd
61
62               # Static home page
63               cat > /var/www/html/index.html << 'EOF'
64               <html>
65                 <head>
66                   <title>AWS Simple EC2 (Auto Scaling)</title>
67                 </head>
68                 <body>
69                   <h1>Hello from an Auto Scaling EC2 instance!</h1>
70                   <p>This instance was launched by an Auto Scaling Group.</p>
71                   <p>To test the database connection, open <code>/db-check.php</code>.</p>
72                 </body>
73               </html>
74               EOF
75
76               # DB connection check page

```

In 39 Col 22 Spaces: 2 LINES: 9 CRLF: 1 VAMI: 1 Signed out

2. The ASG is configured across two public subnets for high availability.
3. CPU-based scaling policies are applied.
4. To increase the load, CPU stress is generated using the SSH command:

```
stress --cpu 4 --timeout 120
```

5. When CPU usage crossed the threshold, a new EC2 instance was launched automatically.

B. AWS Lambda for Logging S3 Uploads

A Lambda function was created to automatically log S3 uploads. Whenever a file is uploaded to a S3 bucket, Lambda function is triggered and it writes the information about the upload to CloudWatch Logs.

1. Created an S3 bucket.
2. Wrote a Python Lambda function that reads S3 event metadata and prints it.
3. Linked S3 to Lambda using a bucket event trigger for operations.

```

E > Sumanth UMBC > IS698-Project > boto3-scripts > s3-upload-logger-project.py > ...
  1 import json
  2 import boto3
  3 import urllib.parse
  4 import logging
  5
  6 logger = logging.getLogger()
  7 logger.setLevel(logging.INFO)
  8
  9 s3 = boto3.client('s3')
 10
 11 def lambda_handler(event, context):
 12
 13     logger.info("Received event: %s, %s", json.dumps(event))
 14     for record in event.get("Records", []):
 15         try:
 16             s3_info = record.get("s3", {})
 17             bucket = s3_info.get("bucket", {}).get("name")
 18             key = s3_info.get("object", {}).get("key")
 19             size = s3_info.get("object", {}).get("size")
 20
 21             if key:
 22                 key = urllib.parse.unquote_plus(key)
 23             event_time = record.get("eventTime")
 24             event_name = record.get("eventName")
 25
 26             if size is None and bucket and key:
 27                 try:
 28                     head = s3.head_object(Bucket=bucket, Key=key)
 29                     size = head.get("ContentLength")
 30                 except Exception as e:
 31                     logger.warning("Could not head_object for %s/%s: %s", bucket, key, str(e))
 32
 33             logger.info("S3 Upload - bucket: %s, key: %s, size: %s, event: %s, time: %s",
 34                         bucket, key, size, event_name, event_time)
 35         except Exception as e:
 36             logger.exception("Error processing record: %s", str(e))
 37
 38     return {"status": "ok"}
 39
 40 ▾ 4

```

In 1, Col 1 | Spaces: 4 | UTF-8 | CRLF | () Python | 83 Signed out | 3.13.2

4. Uploaded test files to S3, and CloudWatch Logs confirmed the event was processed.

C. AWS Interaction via AWS CLI

AWS CLI commands were used to manage multiple AWS resources directly through terminal.

Commands:-

1. List EC2 Instances: `aws ec2 describe-instances`
2. List S3 buckets: `aws s3 ls`
3. Invoking Lambda Function: `aws lambda invoke`

D. Python Boto3 Interaction

To interact with AWS programmatically, python scripts were written.

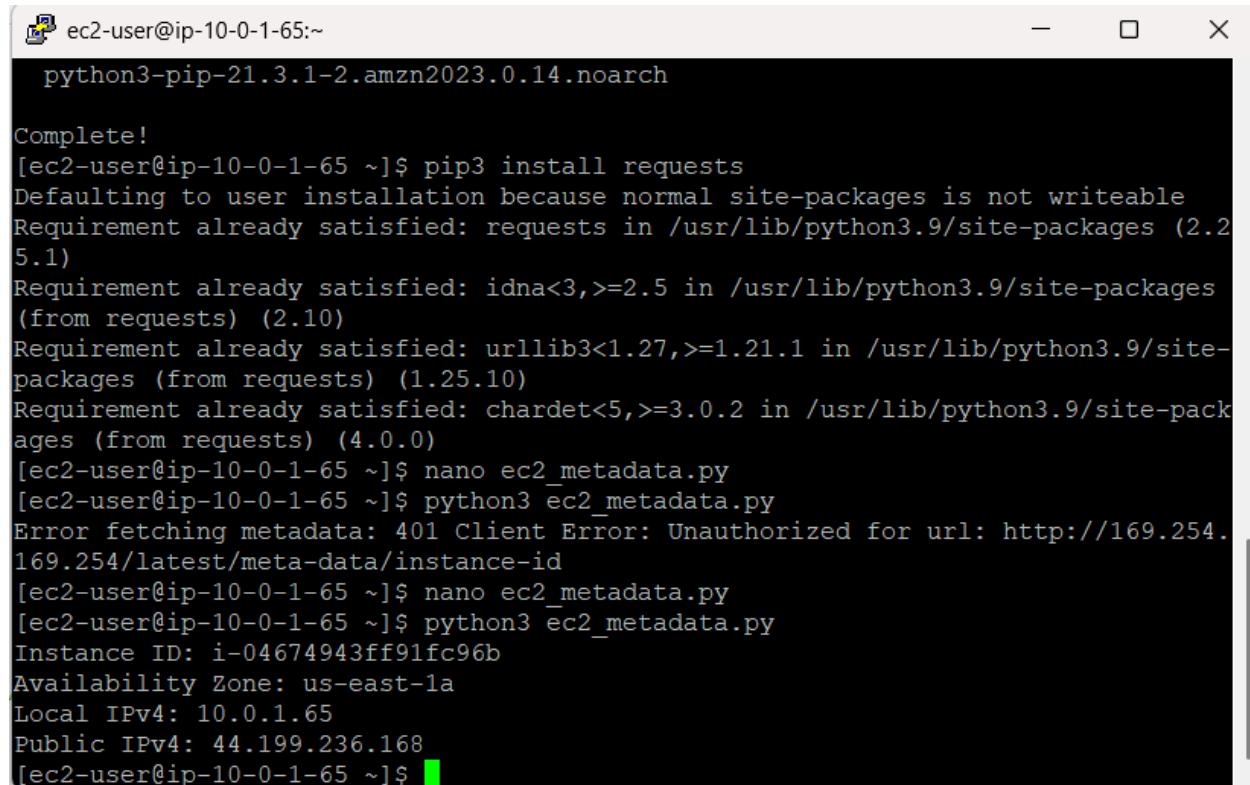
1. Creating an S3 bucket:

```
E: > Sumanth UMBC > IS698-Project > boto3-scripts > create_bucket.py > main
1 import boto3
2 import uuid
3
4 def main():
5     #region
6     region = "us-east-1"
7
8     #S3 client
9     s3 = boto3.client("s3", region_name=region)
10
11     #bucket name
12     bucket_name = f"boto3-demo-bucket-{uuid.uuid4().hex[:8]}"
13
14     print(f"Creating bucket: {bucket_name}")
15
16     s3.create_bucket(Bucket=bucket_name)
17
18     #simple file in memory
19     file_key = "hello.txt"
20     file_body = b"Hello from Boto3 S3 demo!"
21
22     print(f"Uploading {file_key} to {bucket_name}")
23     s3.put_object(Bucket=bucket_name, Key=file_key, Body=file_body)
24
25     print("Done. You can verify in S3 console.")
26
27 if __name__ == "__main__":
28     main()
```

2. Listing EC2 instances

```
E: > Sumanth UMBC > IS698-Project > boto3-scripts > list_ec2.py > ...
1 import boto3
2
3 def main():
4     region = "us-east-1"
5     ec2 = boto3.client("ec2", region_name=region)
6
7     response = ec2.describe_instances(
8         Filters=[{"Name": "instance-state-name", "Values": ["running"]}]
9     )
10
11     print("Running EC2 instances:")
12     for reservation in response["Reservations"]:
13         for inst in reservation["Instances"]:
14             instance_id = inst["InstanceId"]
15             state = inst["State"]["Name"]
16             public_ip = inst.get("PublicIpAddress")
17             name_tag = next(
18                 (t["Value"] for t in inst.get("Tags", []) if t["Key"] == "Name"),
19                 None,
20             )
21
22             print(f"- ID: {instance_id}, State: {state}, Name: {name_tag}, Public IP: {public_ip}")
23
24 if __name__ == "__main__":
25     main()
```

3. Retrieving EC2 metadata

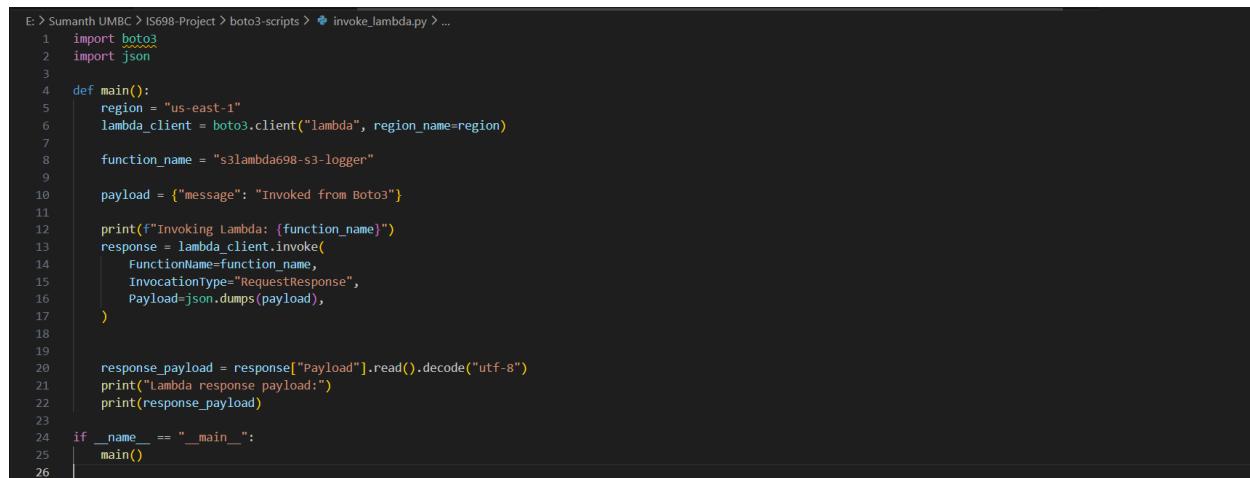


A terminal window titled "ec2-user@ip-10-0-1-65:~". The window shows the output of a pip3 install requests command, followed by the execution of ec2_metadata.py which prints EC2 instance metadata.

```
python3-pip-21.3.1-2.amzn2023.0.14.noarch

Complete!
[ec2-user@ip-10-0-1-65 ~]$ pip3 install requests
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: requests in /usr/lib/python3.9/site-packages (2.25.1)
Requirement already satisfied: idna<3,>=2.5 in /usr/lib/python3.9/site-packages
(from requests) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/lib/python3.9/site-packages
(from requests) (1.25.10)
Requirement already satisfied: chardet<5,>=3.0.2 in /usr/lib/python3.9/site-packages
(from requests) (4.0.0)
[ec2-user@ip-10-0-1-65 ~]$ nano ec2_metadata.py
[ec2-user@ip-10-0-1-65 ~]$ python3 ec2_metadata.py
Error fetching metadata: 401 Client Error: Unauthorized for url: http://169.254.169.254/latest/meta-data/instance-id
[ec2-user@ip-10-0-1-65 ~]$ nano ec2_metadata.py
[ec2-user@ip-10-0-1-65 ~]$ python3 ec2_metadata.py
Instance ID: i-04674943ff91fc96b
Availability Zone: us-east-1a
Local IPv4: 10.0.1.65
Public IPv4: 44.199.236.168
[ec2-user@ip-10-0-1-65 ~]$
```

4. Invoking lambda

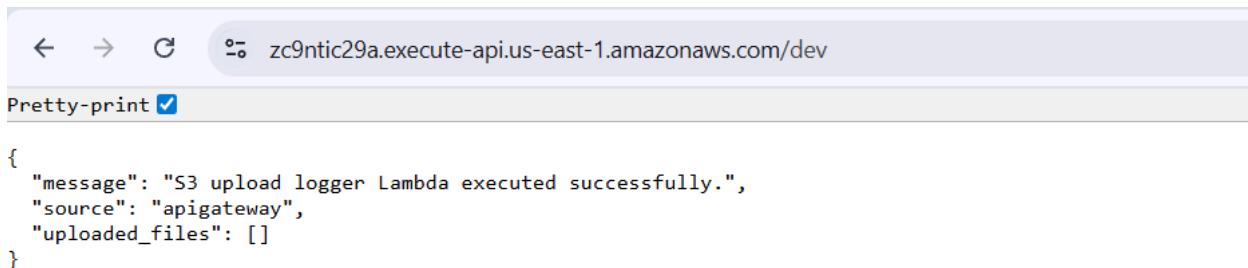


A terminal window titled "E: > Sumanth UMBC > IS698-Project > boto3-scripts > invoke_lambda.py > ...". The window shows the execution of invoke_lambda.py, which uses Boto3 to invoke a Lambda function and print the response payload.

```
1 import boto3
2 import json
3
4 def main():
5     region = "us-east-1"
6     lambda_client = boto3.client("lambda", region_name=region)
7
8     function_name = "s3lambda698-s3-logger"
9
10    payload = {"message": "Invoked from Boto3"}
11
12    print(f"Invoking Lambda: {function_name}")
13    response = lambda_client.invoke(
14        FunctionName=function_name,
15        InvocationType="RequestResponse",
16        Payload=json.dumps(payload),
17    )
18
19
20    response_payload = response["Payload"].read().decode("utf-8")
21    print("Lambda response payload:")
22    print(response_payload)
23
24 if __name__ == "__main__":
25     main()
```

E. Bonus Challenge

Deploy **API Gateway** to invoke Lambda via HTTP requests.



A screenshot of a web browser window. The address bar shows the URL: `zc9ntic29a.execute-api.us-east-1.amazonaws.com/dev`. Below the address bar, there is a "Pretty-print" checkbox which is checked. The main content area displays a JSON object:

```
{  
  "message": "S3 upload logger Lambda executed successfully.",  
  "source": "apigateway",  
  "uploaded_files": []  
}
```

A REST API was added using API Gateway to trigger the existing Lambda function through an HTTP request. The Lambda code was updated to recognize API Gateway events and return a simple JSON message. After creating a new resource and GET method in API Gateway, the method was linked to the Lambda function using proxy integration and deployed to a stage. The generated URL allowed the Lambda function to be invoked from a browser, confirming that the function works for both S3 uploads and manual API calls.

F. Challenges Encountered and Solutions Applied

During the course of this project, various difficulties were encountered while working with Terraform, CloudFormation, EC2, RDS, and Lambda functions. The first difficulty was encountered while creating the RDS instance. The stack was failing consistently due to improper DB parameters and the security group configuration. This problem was fixed by correcting the DB name, checking the credentials, and allowing MySQL access from the EC2 security group on the RDS security group. After this was done, the stack was created successfully.

Another issue at this stage was ensuring the connection between the EC2 instance and RDS was established. Despite updating the PHP script, the issue persisted. It was necessary to allow outbound traffic on port 3306.

Test cases for auto-scaling also included a wait for newly created EC2 instances. Initially, there was assumed to be a problem with the auto-scaling configuration, though the actual issue was the time for image initialization. Once realizing that image deployment does take about 2-3 minutes, the auto-scaling function was tested successfully.

At first, the script for EC2 metadata failed because the instance was set to use IMDSv2. This requires the use of a session token. Incorporating this token logic into the script solved the issue with authentication.

G. Future improvements and Recommendations

In the future, this architecture could be improved in a few practical ways:

- Enabling RDS Multi-AZ and automated backups would protect the database from Availability Zone failures and make recovery from issues much easier. Performance Insights could also help tune queries if the workload grows.
- CloudWatch dashboards and alarms can be added for CPU usage, scaling events, RDS performance, and Lambda errors. Connecting these alarms to SNS notifications would make it easier to react quickly when something goes wrong.