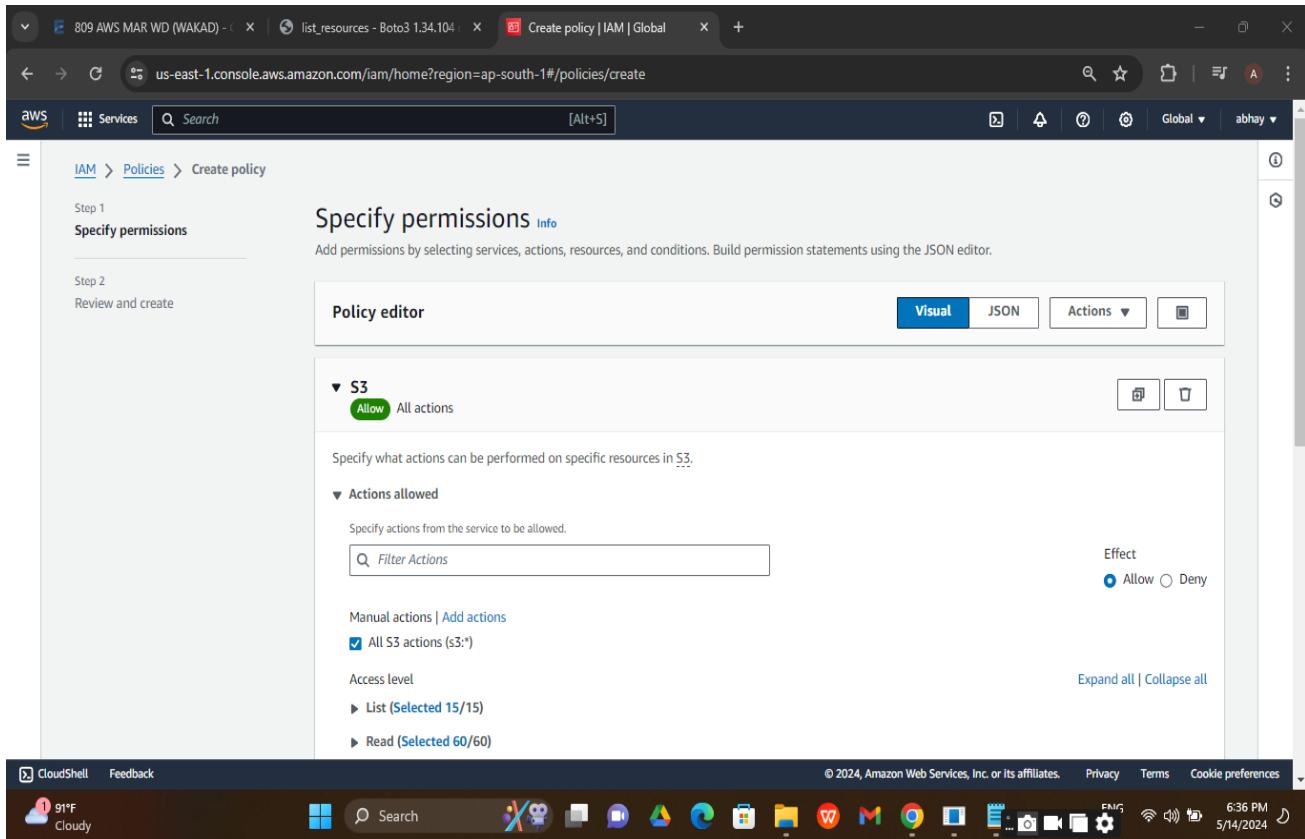


Lambda Automation Datapipeline

- Suppose I get a requirement from a client that he has a data and he wants to store that data in the dynamodb database of aws and all this should be done with single click execution(automated).
- We are creating datapipline using lambda service.
- First, we create 1 bucket in s3 and then upload this file in s3 bucket to using python boto3.
- We write the lambda function in this lambda function we write the program this program read this file and then write this data in dynamodb.
- lambda always run from trigger and this trigger means event of aws.
- Uploading file is an event so what i do i create that the movement some file will land in bucket that will invoke this lambda then lambda function code read the file from bucket and write the data in Database.
- So, we are creating pipeline.
- Why we select dynamodb database for storing data because this file data is json and dynamodb is pioneer database to hold json data.
- So first we go to IAM service and create policy.



The screenshot shows the AWS IAM console with the URL us-east-1.console.aws.amazon.com/iam/home?region=ap-south-1#/policies/create. The page is titled "Create policy | IAM | Global". The main content area is titled "DynamoDB" with an "Allow All actions" button. It asks to "Specify what actions can be performed on specific resources in DynamoDB." Under "Actions allowed", there is a "Filter Actions" search bar and a list of actions grouped by access level:

- Manual actions | Add actions**:
 - All DynamoDB actions (dynamodb:*)
- Access level**:
 - ▶ List (Selected 6/6)
 - ▶ Read (Selected 27/27)
 - ▶ Write (Selected 32/32)
 - ▶ Permissions management (Selected 2/2)
 - ▶ Tagging (Selected 2/2)

On the right, there is an "Effect" section with "Allow" selected. At the bottom right of the main content area are "Expand all" and "Collapse all" buttons.

The bottom navigation bar includes CloudShell, Feedback, Search, and various system icons like battery, signal, and date/time (6:38 PM, 5/14/2024).

This screenshot shows the same AWS IAM policy creation interface, but the service selected is CloudWatch. The title "Create policy | IAM | Global" is visible at the top. The main content area is titled "CloudWatch" with an "Allow All actions" button. It asks to "Specify what actions can be performed on specific resources in CloudWatch." Under "Actions allowed", there is a "Filter Actions" search bar and a list of actions grouped by access level:

- Manual actions | Add actions**:
 - All CloudWatch actions (cloudwatch:*)
- Access level**:
 - ▶ List (Selected 6/6)
 - ▶ Read (Selected 20/20)
 - ▶ Write (Selected 25/25)
 - ▶ Tagging (Selected 2/2)

On the right, there is an "Effect" section with "Allow" selected. At the bottom right of the main content area are "Expand all" and "Collapse all" buttons.

The bottom navigation bar includes CloudShell, Feedback, Search, and various system icons like battery, signal, and date/time (6:38 PM, 5/14/2024).

The screenshot shows the AWS IAM 'Create policy' interface for CloudWatch Logs. The policy name is 'CloudWatchLogsPolicy'. The 'Actions allowed' section is expanded, showing a list of actions under 'All CloudWatch Logs actions (logs*)'. The 'Selected' column indicates 18 actions are selected: List (18/18), Read (18/18), Write (40/40), Permissions management (2/2), and Tagging (4/4). A warning message at the bottom states: 'Dependent permissions not selected. To grant permissions for the selected resource actions, including additional dependent actions might be required.' The 'Effect' dropdown is set to 'Allow'. The browser status bar at the bottom shows 'CloudShell Feedback' and the date '5/14/2024'.

The screenshot shows the 'Review and create' step of the policy creation wizard. The policy name is 'PolicyForDataPipeline'. The 'Policy details' section includes a 'Policy name' input field containing 'PolicyForDataPipeline' and a 'Description - optional' input field. The 'Permissions defined in this policy' section shows a search bar and an 'Edit' button. The browser status bar at the bottom shows 'CloudShell Feedback' and the date '5/14/2024'.

Screenshot of the AWS IAM Policy creation screen showing permissions defined in the policy.

Permissions defined in this policy Info

Permissions defined in this policy document specify which actions are allowed or denied. To define permissions for an IAM identity (user, user group, or role), attach a policy to it.

Search

Allow (3 of 411 services) Show remaining 408 services

Service	Access level	Resource	Request condition
CloudWatch Logs	Full access	All resources	None
DynamoDB	Full access	All resources	None
S3	Full access	All resources	None

Add tags - optional Info

Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.

No tags associated with the resource.

Add new tag
You can add up to 50 more tags.

Cancel Previous Create policy

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 6:42 PM 5/14/2024

Screenshot of the AWS IAM Policy creation screen showing the review and create step.

Step 1 [Specify permissions](#)

Step 2 [Review and create](#)

Review and create Info

Review the permissions, specify details, and tags.

Policy details

Policy name
Enter a meaningful name to identify this policy.
 PolicyForDataPipeline02
Maximum 128 characters. Use alphanumeric and '+-=_,@-_.' characters.

Description - optional
Add a short explanation for this policy.

Maximum 1,000 characters. Use alphanumeric and '+-=_,@-_.' characters.

Permissions defined in this policy Info

Permissions defined in this policy document specify which actions are allowed or denied. To define permissions for an IAM identity (user, user group, or role), attach a policy to it.

Search

Allow (3 of 411 services) Show remaining 408 services

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 6:43 PM 5/14/2024

The screenshot shows the AWS IAM Roles page. The left sidebar includes sections for Dashboard, Access management (User groups, Users, Roles, Policies, Identity providers, Account settings), Access reports (Access Analyzer, External access, Unused access, Analyzer settings), and Credential report. The main content area displays a table titled "Roles (38) Info" with columns for Role name, Trusted entities, and Last activity. The table lists various AWS service roles, such as AWSCodePipelineServiceRole and AWSCodePipelineServiceRoleForAmazonSSM, along with their respective last activity times.

The screenshot shows the "Create role" wizard, Step 2: Trusted entity type. It asks for a name, review, and create. The "Trusted entity type" section contains five options: "AWS service" (selected), "AWS account", "Web identity", "SAML 2.0 federation", and "Custom trust policy". Below this, the "Use case" section allows selecting a service or use case, with "Lambda" selected. Under "Use case", "Lambda" is chosen, described as allowing Lambda functions to call AWS services on behalf of the user.

The screenshot shows the 'Add permissions' step of the IAM role creation wizard. It lists two customer-managed policies: 'policyfordatapipeline' and 'PolicyForDataPipeline02'. The second policy is selected. A note below says 'Set permissions boundary - optional'. Navigation buttons 'Cancel', 'Previous', and 'Next' are at the bottom.

Permissions policies (1/945) [Info](#)

Choose one or more policies to attach to your new role.

Filter by Type All types 2 matches

Policy name	Type	Description
policyfordatapipeline	Customer managed	-
PolicyForDataPipeline02	Customer managed	-

▶ Set permissions boundary - *optional*

Cancel Previous Next



The screenshot shows the 'Name, review, and create' step of the IAM role creation wizard. It displays 'Role details' with a role name 'RoleForDataPipeline02' and a description 'Allows Lambda functions to call AWS services on your behalf.'. Below is 'Step 1: Select trusted entities' with a trust policy editor containing the following JSON:

```
1 "Version": "2012-10-17",
2 "Statement": [
3     {
4         "Effect": "Allow",
5         "Principal": "*",
6         "Action": "sts:AssumeRole"
7     }
]
```

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Step 2: Add permissions

Permissions policy summary

Policy name	Type	Attached as
PolicyForDataPipeline02	Customer managed	Permissions policy

Step 3: Add tags

Add tags - optional Info
Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.

No tags associated with the resource.

Add new tag

You can add up to 50 more tags.

Cancel Previous Create role

- We are go to dynamodb and go to tables and create table.

The request to delete the "employee" table has been submitted successfully.

DynamoDB > Tables

Tables (1) Info

Name	Status	Partition key	Sort key	Indexes	Deletion protection	Read capacity mode	Write capacity mode
employee	⚠ Unknown						

⚠ Unavailable

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Screenshot of the AWS DynamoDB 'Create table' wizard.

The 'Table settings' step is displayed, showing the following configuration:

Auto scaling	On	Yes
Local secondary indexes	-	No
Global secondary indexes	-	Yes
Encryption key management	Owned by Amazon DynamoDB	Yes
Deletion protection	Off	Yes
Resource-based policy	Not active	Yes

Tags
Tags are pairs of keys and optional values, that you can assign to AWS resources. You can use tags to control access to your resources or track your AWS spending.

No tags are associated with the resource.

Add new tag
You can add 50 more tags.

Create table

Screenshot of the AWS DynamoDB 'List tables' page.

The 'Tables' section shows one table named 'employee'.

Name	Status	Partition key	Sort key	Indexes	Deletion protection	Read capacity mode	Write capacity mode
employee	Active	emp_id (\$)	-	0	Off	Provisioned (5)	Provisioned (5)

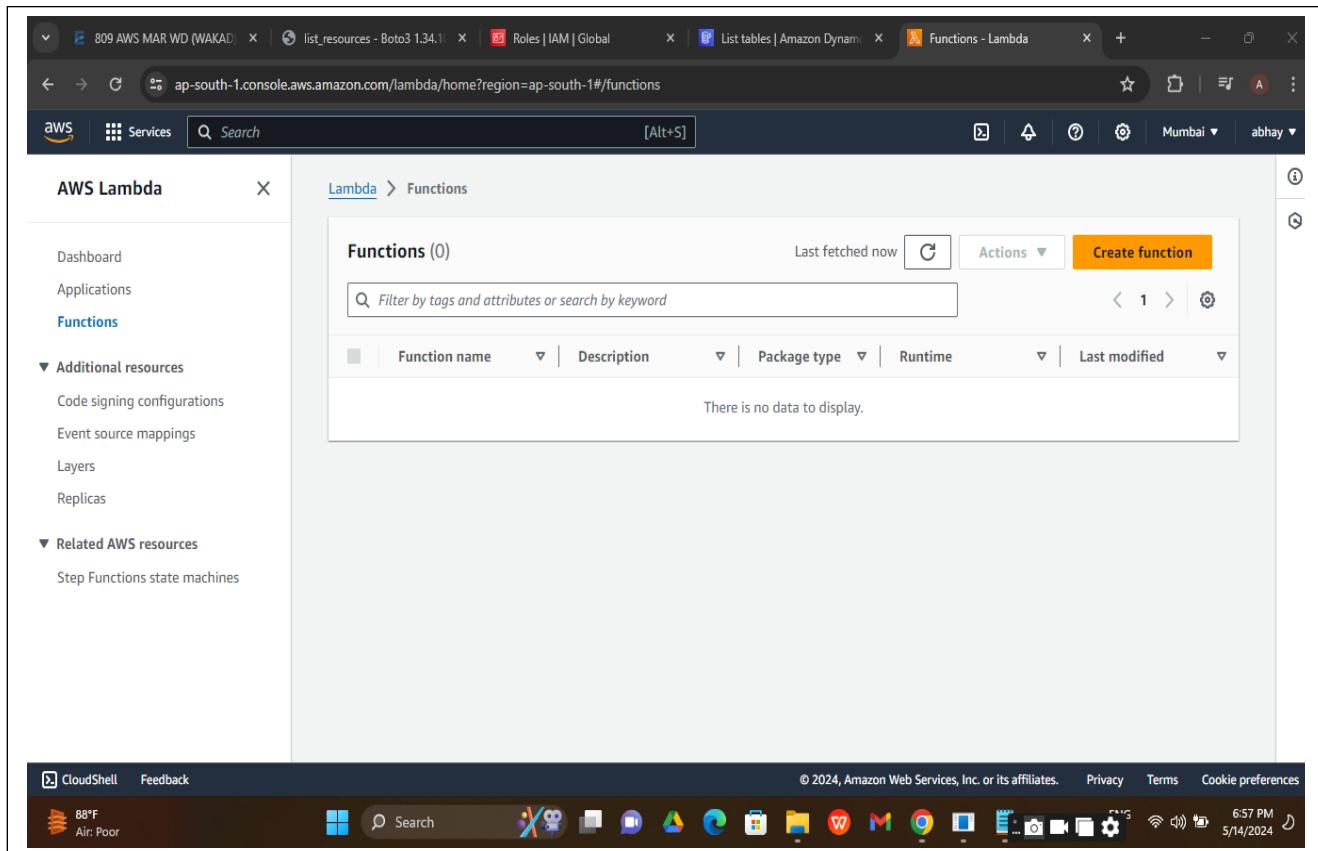
DynamoDB

- Dashboard
- Tables**
- Explore items
- PartiQL editor
- Backups
- Exports to S3
- Imports from S3
- Integrations [New](#)
- Reserved capacity
- Settings

- DAX**
- Clusters
- Subnet groups
- Parameter groups
- Events

CloudShell Feedback

- Go to lambda service and create function- Author from scratch.
- Add trigger s3
- Go to code and we just print the event and see what information his give - deploy the code
- Go to trigger and our function trigger is s3 so run the code and upload the file then go to CloudWatch and see logs groups aws/lambda/datapipline - see the information print by event.



The screenshot shows the 'Create function' wizard on the AWS Lambda console. The current step is 'Permissions'. It includes sections for 'Change default execution role' (with options to 'Create a new role with basic Lambda permissions', 'Use an existing role' (selected), and 'Create a new role from AWS policy templates'), 'Existing role' (set to 'RoleForDataPipeline02'), and 'Advanced settings'. At the bottom right is a 'Create function' button.

The screenshot shows the 'DataPipline809' Lambda function details page. The 'Function overview' section displays the function name, a diagram icon, and a 'Layers' section showing '(0)'. It also includes buttons for 'Throttle', 'Copy ARN', and 'Actions'. To the right, there are fields for 'Description', 'Last modified' (in 8 seconds), 'Function ARN' (arn:aws:lambda:ap-south-1:851725280943:function:DataPipline809), and 'Function URL' (Info). Below this are tabs for 'Code', 'Test', 'Monitor', 'Configuration', 'Aliases', and 'Versions'. The bottom navigation bar includes 'CloudShell', 'Feedback', and standard browser controls.

The screenshot shows the AWS Lambda console with a new trigger configuration. The trigger type is set to "All object create events". Under "Prefix - optional", the value "e.g. images/" is shown. Under "Suffix - optional", the value "e.g. .jpg" is shown. A note about recursive invocation states: "If your function writes objects to an S3 bucket, ensure that you are using different S3 buckets for input and output. Writing to the same bucket increases the risk of creating a recursive invocation, which can result in increased Lambda usage and increased costs." A checkbox is checked, acknowledging this note. A note at the bottom states: "Lambda will add the necessary permissions for AWS S3 to invoke your Lambda function from this trigger. Learn more about the Lambda permissions model." At the bottom right are "Cancel" and "Add" buttons.

The screenshot shows the AWS Lambda console with a success message: "The trigger 14may2024-1 was successfully added to function DataPipeline809. The function is now receiving events from the trigger." Below this, the "Function overview" section is displayed. It shows the function name "DataPipeline809" and its ARN: "arn:aws:lambda:ap-south-1:851725280943:function:DataPipeline809". The "Configuration" tab is selected. At the bottom, there are tabs for "Code", "Test", "Monitor", "Configuration", "Aliases", and "Versions".

The screenshot shows the AWS Lambda code editor interface. The main area displays a Python file named `lambda_function.py` with the following content:

```
1 import json
2
3 def lambda_handler(event, context):
4     print(str(event))
5     return "hello"
```

This screenshot is identical to the one above, showing the AWS Lambda code editor with the same Python function code in `lambda_function.py`.

PyCharm Project: practice

FileUpload.py content:

```
1 import boto3
2 s3 = boto3.client('s3')
3 s3.upload_file('F:/user/dell/Downloads/data.json', '14may2024-1', 'data.json')
```

Debug output:

```
C:\Users\Del\AppData\Local\Programs\Python\Python312\python.exe -X pycache_prefix=C:\Users\Del\AppData\Local\JetBrains\PyCharmCE2023.3\cpython-cache "F:\user\dell\PycharmProjects\practice\FileUpload.py"
Connected to pydev debugger (build 233.15026.15)
Process finished with exit code 0
```

AWS CloudWatch Log Groups

Log groups (1)

Log group	Log class	Anomaly d...	Dat...	Sen...	Re...
/aws/lambda/DataPipline809	Standard	Configure	-	-	Ne...

Screenshot of the AWS CloudWatch Log Groups page for the DataPipeline809 log group.

Log groups

- Log streams (1)
- 2024/05/14/[\$LATEST]9cf4066c9bc64802aaf7609fee7146ff

Log streams (1)

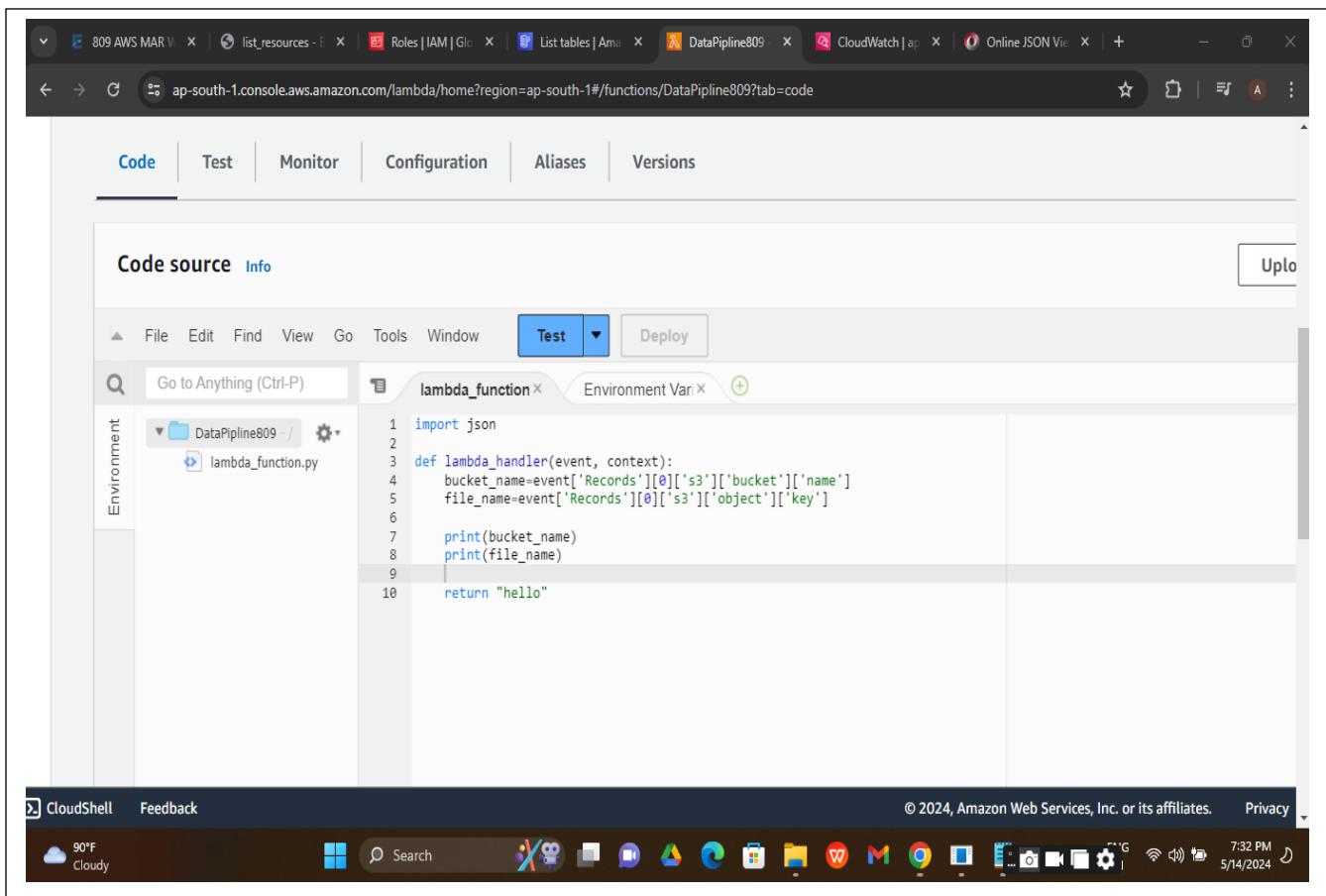
Log stream	Last event time
2024/05/14/[\$LATEST]9cf4066c9bc64802aaf7609fee7146ff	2024-05-14 13:46:35 (UTC)

Screenshot of the AWS CloudWatch Log Events page for the DataPipeline809 log group.

Log events

Timestamp	Message
2024-05-14T13:46:35.398Z	INIT_START Runtime Version: python:3.8.v48 Runtime Version ARN: arn:aws:lambda:ap-south-1::runtime...
2024-05-14T13:46:35.518Z	START RequestId: 71abb4fb-5623-4ec7-8141-02f80319f654 Version: \$LATEST
2024-05-14T13:46:35.519Z	{'Records': [{}]} ('Records': [{}])
2024-05-14T13:46:35.523Z	END RequestId: 71abb4fb-5623-4ec7-8141-02f80319f654
2024-05-14T13:46:35.523Z	REPORT RequestId: 71abb4fb-5623-4ec7-8141-02f80319f654 Duration: 4.95 ms Billed Duration: 5 ms Mem...

- We are the code in lambda for read the file from bucket and write in dynamodb database and also, convert the format in which we want to write.
- We write bucket and file name and check whether it is resolved or not.
- Upload the bucket for trigger and go to logs groups and see the resolved.



The screenshot shows the AWS Lambda function editor interface. The top navigation bar includes tabs for Code, Test, Monitor, Configuration, Aliases, and Versions. The Code tab is selected. Below the tabs, there are buttons for File, Edit, Find, View, Go, Tools, Window, and a dropdown menu currently set to Test. A toolbar with Test and Deploy buttons is visible. The main area is titled "Code source" with an Info link. On the left, there's a sidebar for Environment variables and a CloudShell feedback button. The main code editor window displays the following Python script:

```

1 import json
2
3 def lambda_handler(event, context):
4     bucket_name=event['Records'][0]['s3']['bucket']['name']
5     file_name=event['Records'][0]['s3']['object']['key']
6
7     print(bucket_name)
8     print(file_name)
9
10    return "hello"

```

The AWS Lambda function editor interface includes a search bar, a file tree showing a folder named "DataPipeline809" containing "lambda_function.py", and a status bar at the bottom with weather information (90°F, Cloudy), system icons, and the date/time (7:32 PM, 5/14/2024).

The screenshot shows the PyCharm IDE interface. In the top right, there's a toolbar with icons for file operations like 'FileUpload' and 'Run'. The left sidebar shows a project named 'practice' containing files: 'create s3 bucket.py', 'FileUpload.py', 'praccreatebuckets3.py', and 'prafileupload.py'. The main editor window displays the 'FileUpload.py' code:

```
import boto3
s3 = boto3.client('s3')
s3.upload_file('F:/user/dell/Downloads/data.json', '14may2024-1', 'data.json')
```

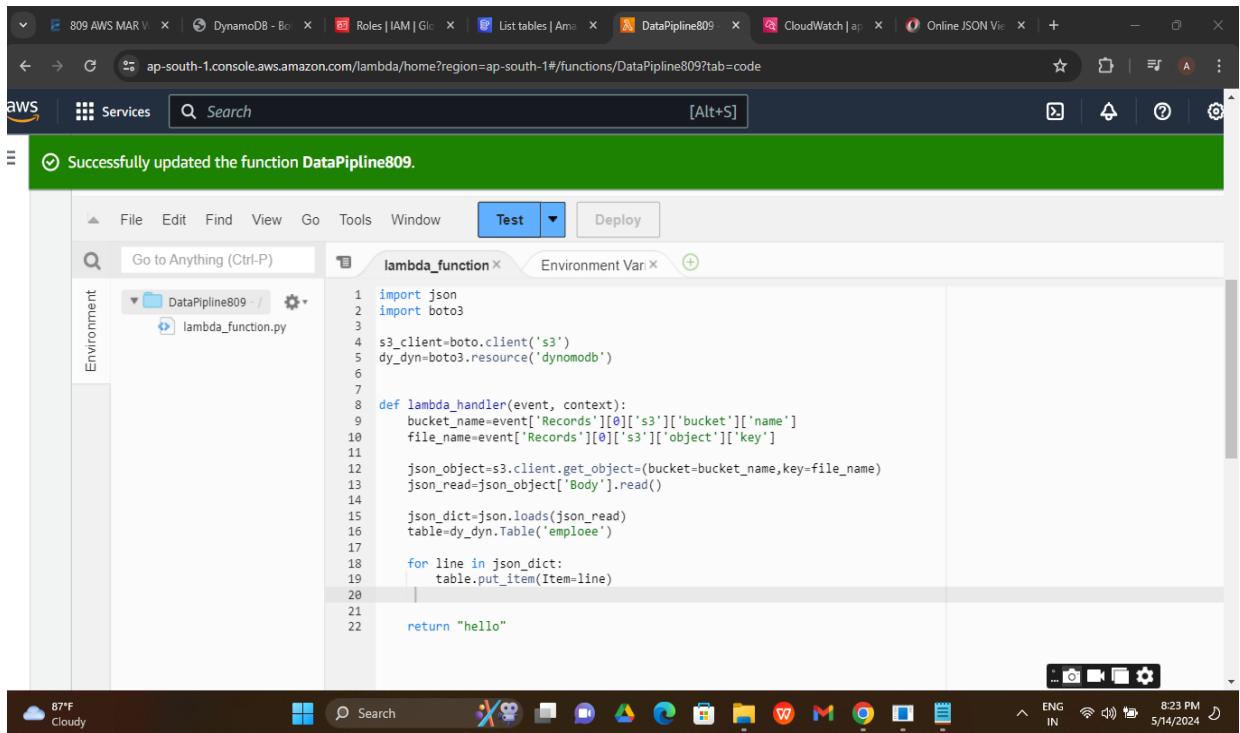
Below the editor is the 'Debug' tool window, which shows the output of the run: 'Connected to pydev debugger (build 233.15026.15)' and 'Process finished with exit code 0'. At the bottom, the status bar indicates 'Waiting for process detach' and shows system information like '3:68 CRLF UTF-8 4 spaces Python 3.12'. The taskbar at the bottom includes icons for various applications like File Explorer, Edge, and Google Chrome.

The screenshot shows the AWS CloudWatch Log Events interface. The left sidebar has a navigation tree with 'CloudWatch' selected, followed by 'Logs', 'Log groups', 'Logs', 'Log groups', and finally 'aws/lambda/DataPipeline809'. The main content area shows a table of log events:

Timestamp	Message
2024-05-14T14:03:33.433Z	INIT_START Runtime Version: python:3.8.v48 Runtime Version ARN: arn:aws:lambda:ap-south-1::runtime:83...
2024-05-14T14:03:33.558Z	START RequestId: 5851321b-2ad3-47b7-a5df-8878f0ef2629 Version: \$LATEST
2024-05-14T14:03:33.559Z	14may2024-1
2024-05-14T14:03:33.559Z	data.json
2024-05-14T14:03:33.560Z	END RequestId: 5851321b-2ad3-47b7-a5df-8878f0ef2629
2024-05-14T14:03:33.560Z	REPORT RequestId: 5851321b-2ad3-47b7-a5df-8878f0ef2629 Duration: 2.17 ms Billed Duration: 3 ms Memory...

At the bottom, there are links for 'CloudShell', 'Feedback', and the AWS footer with copyright information.

- Write code for lambda automation datapipeline
- Upload file as trigger
- Go to dynamodb database and check data will store.



The screenshot shows the AWS Lambda console interface. At the top, there are several tabs: '809 AWS MAR...', 'DynamoDB - Bo...', 'Roles | IAM | G...', 'List tables | Ana...', 'DataPipeline809', 'CloudWatch | ap...', 'Online JSON Vie...', and a '+' button. Below the tabs, the URL is 'ap-south-1.console.aws.amazon.com/lambda/home?region=ap-south-1#functions/DataPipeline809?tab=code'. The main area has a green success message: 'Successfully updated the function DataPipeline809.' The interface includes a navigation bar with 'File', 'Edit', 'Find', 'View', 'Go', 'Tools', 'Window', 'Test' (which is currently selected), and 'Deploy' buttons. On the left, there's a sidebar with 'Environment' and a search bar 'Go to Anything (Ctrl-P)'. The central workspace shows a file tree with 'DataPipeline809' and 'lambda_function.py'. The code editor displays the following Python script:

```

1 import json
2 import boto3
3
4 s3_client=boto3.client('s3')
5 dy_dyn=boto3.resource('dynamodb')
6
7
8 def lambda_handler(event, context):
9     bucket_name=event['Records'][0]['s3']['bucket']['name']
10    file_name=event['Records'][0]['s3']['object']['key']
11
12    json_object=s3_client.get_object(Bucket=bucket_name,Key=file_name)
13    json_read=json_object['Body'].read()
14
15    json_dict=json.loads(json_read)
16    table=dy_dyn.Table('employee')
17
18    for line in json_dict:
19        table.put_item(Item=line)
20
21
22    return "hello"

```

The screenshot shows the PyCharm IDE interface. The top navigation bar includes 'FileUpload' and other standard options. The left sidebar displays a project structure for 'practice' containing files like 'create s3 bucket.py', 'FileUpload.py', 'praccretebuckets3.py', and 'prfileupload.py'. The main editor window contains two files: 'create s3 bucket.py' and 'FileUpload.py'. The code in 'FileUpload.py' is:

```
1 import boto3
2 s3 = boto3.client('s3')
3 s3.upload_file('F:/user/dell/Downloads/data.json', '14may2024-1', 'data.json')
```

The bottom panel shows the 'Debug' tool window with the 'FileUpload' configuration selected. The 'Console' tab displays the output of the run command: 'Connected to pydev debugger (build 233.15026.15)' and 'Process finished with exit code 0'. The status bar at the bottom right shows the date and time as '5/14/2024 8:23 PM'.

The screenshot shows the AWS CloudWatch Logs interface in a browser. The left sidebar lists services: CloudWatch, Services, and a search bar. Under 'Logs', 'Log groups' is selected, showing 'Log streams (3)'. The log streams listed are:

Log stream	Last event time
2024/05/14/[\$LATEST]a5d385e483e742dd8d0fc9c877729cbc	2024-05-14 14:53:49 (UTC)
2024/05/14/[\$LATEST]441ca85a99334682befea919a77c1c2b	2024-05-14 14:03:33 (UTC)
2024/05/14/[\$LATEST]9cf4066c9bc64802aaaf7609fee7146ff	2024-05-14 13:46:35 (UTC)

The status bar at the bottom right shows the date and time as '5/14/2024 8:24 PM'.

The screenshot shows the PyCharm IDE interface. On the left, the Project tool window displays a 'practice' project containing several Python files: 'create s3 bucket.py', 'FileUpload.py', 'praccretebuckets3.py', and 'prafileupload.py'. The right side shows the code editor with 'FileUpload.py' open, containing the following code:

```
1 import boto3
2 s3 = boto3.client('s3')
3 s3.upload_file('F:/user/dell/Downloads/data.json', '14may2024-1', 'data.json')
```

Below the code editor is the Debug tool window, which shows the command run: 'C:\Users\...\\Python312\python.exe -X pycache_prefix=C:\Users\...\\PyCharmCE2023.3\cpython-cache *F:'. It also displays the message 'Process finished with exit code 0'. At the bottom, the status bar shows the path 'practice > FileUpload.py' and the message 'Waiting for process detach'.

The screenshot shows the AWS DynamoDB item explorer interface. The left sidebar lists navigation options: Dashboard, Tables, Explore items (selected), PartiQL editor, Backups, Exports to S3, Imports from S3, Integrations (New), Reserved capacity, and Settings. Below this is a DAX section with Clusters, Subnet groups, Parameter groups, and Events. The main area displays a table titled 'Items returned (4)' with the following data:

	emp_id (String)	Age	Cars	Location
22	27	Swift	USA	
24	29		India	
23	16		London	
21	31			

A green notification bar at the top right says 'Completed. Read capacity units consumed: 0.5'. The bottom status bar shows the date '5/14/2024' and time '8:45 PM'.