**Create s3 buckets**

aws s3 mb s3://file-upload-source-bucket-123

aws s3 mb s3://processed-pdf-files-bucket-123

aws s3 mb s3://processed-image-files-bucket-123

**create iam role for lambda:**

aws iam create-role --role-name LambdaS3FileCategorizerRole --assume-role-policy-document '{

  "Version": "2012-10-17",

  "Statement": [

   {

    "Effect": "Allow",

    "Principal": {

     "Service": "lambda.amazonaws.com"

    },

    "Action": "sts:AssumeRole"

   }

  ]

}'

**Attach policies to role:**

aws iam attach-role-policy --role-name LambdaS3FileCategorizerRole --policy-arn arn:aws:iam::aws:policy/AmazonS3FullAccess

aws iam attach-role-policy --role-name LambdaS3FileCategorizerRole --policy-arn arn:aws:iam::aws:policy/service-role/AWSLambdaBasicExecutionRole

**s3 event notification:**

aws s3api put-bucket-notification-configuration \

  --bucket file-upload-source-bucket-123 \

```
--notification-configuration '{

  "LambdaFunctionConfigurations": [

   {

     "LambdaFunctionArn":
"arn:aws:lambda:YOUR_REGION:YOUR_ACCOUNT_ID:function:S3FileCategorizer",

     "Events": ["s3:ObjectCreated:*"]

   }

  ]

 }'
```

**Permission for s3 to invoke lambda:**

```
aws lambda add-permission \

  --function-name S3FileCategorizer \

  --principal s3.amazonaws.com \

  --statement-id s3-invoke \

  --action "lambda:InvokeFunction" \

  --source-arn arn:aws:s3:::file-upload-source-bucket-123 \

  --source-account YOUR_ACCOUNT_ID
```

**Crawler database:**

```
aws glue create-database --database-input '{

  "Name": "file_categorizer_db"

}'
```

**Create crawlers:**

**Pdf:**

```
aws glue create-crawler --name pdf_files_crawler \

  --role arn:aws:iam::YOUR_ACCOUNT_ID:role/service-role/AWSGlueServiceRole \

  --database-name file_categorizer_db \
```

```
  --targets '{

    "S3Targets": [

      {

        "Path": "s3://processed-pdf-files-bucket-123"

      }

    ]

  }' \

  --schedule "cron(0 12 * * ? *)"
```

**Image:**

```
aws glue create-crawler --name image_files_crawler \

  --role arn:aws:iam::YOUR_ACCOUNT_ID:role/service-role/AWSGlueServiceRole \

  --database-name file_categorizer_db \

  --targets '{

    "S3Targets": [

      {

        "Path": "s3://processed-image-files-bucket-123"

      }

    ]

  }' \

  --schedule "cron(0 12 * * ? *)"
```

**Start crawlers:**

```
aws glue start-crawler --name pdf_files_crawler
aws glue start-crawler --name image_files_crawler
```