**Secure Dataflow**

**Solution**

V1.0

Mahendra M

|  |  |  |
| --- | --- | --- |
| Revised by | Version | Date |
| Mahendra M | V1.0 | 7/4/2023 |
|  |  |  |

**Table of contents**

[Introduction 4](#_Toc231732121)

[Architecture Diagram 4](#_Toc38943365)

[Overview 5](#_Toc1341583883)

[Prerequisites 6](#_Toc1157036006)

[Steps to Implement solution using Terraform 6](#_Toc1139592920)

[Security 8](#_Toc1472353794)

[Monitoring and Alerting 8](#_Toc1621715496)

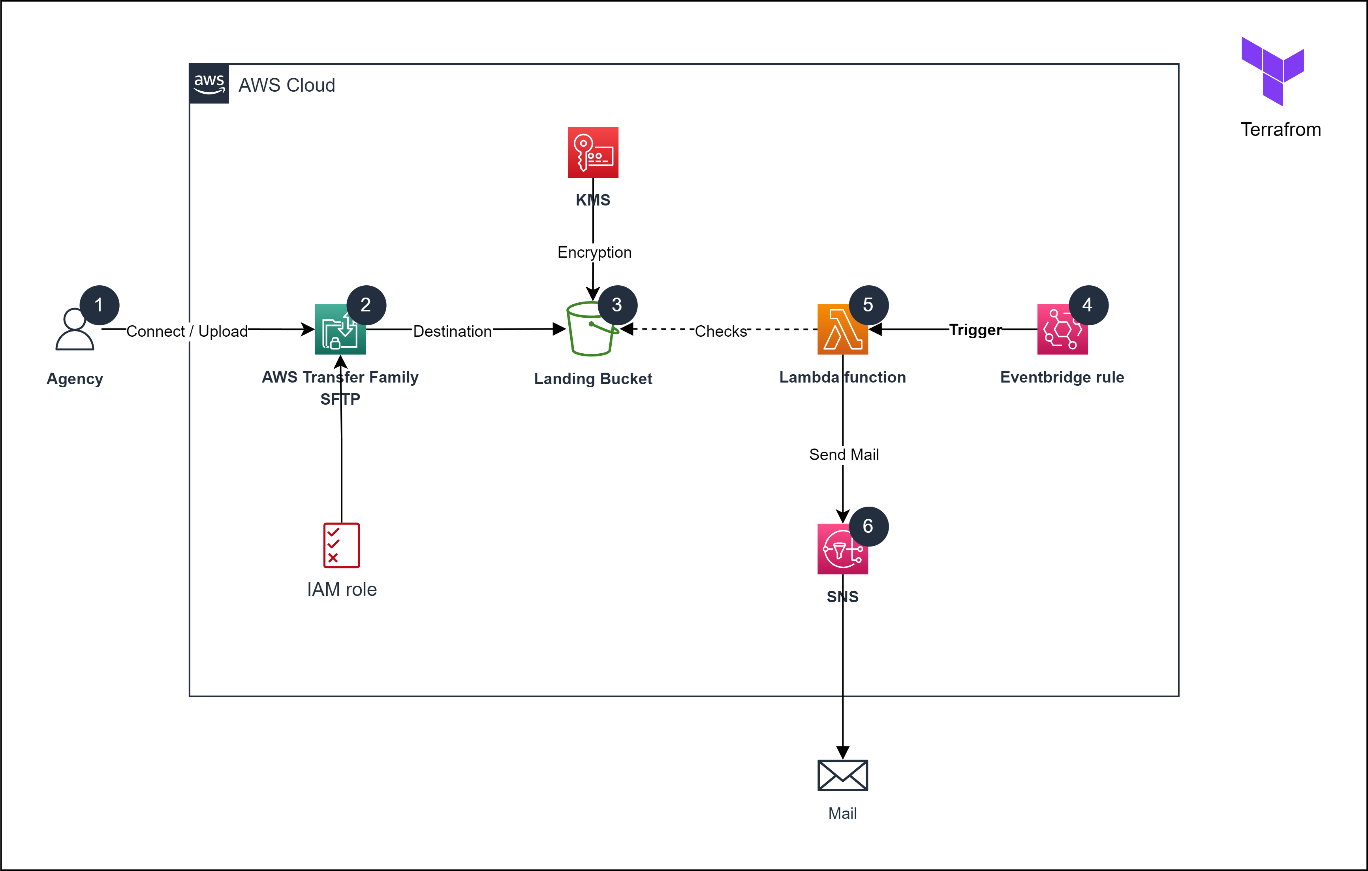
[Working 8](#_Toc1841026206)

**Introduction**

The organization is currently working on a data lake project and needs a solution to enable these agencies to securely upload data to their Amazon S3 buckets. However, these agencies are not very tech-savvy and have requested an SFTP connection for daily or weekly uploads.

In this document, we will outline a solution that utilizes AWS Transfer Family to create an SFTP server that enables the agencies to securely transfer their data to the S3 buckets. We will also provide step-by-step instructions for implementing this solution, as well as best practices to ensure secure and efficient file transfer.

# **Architecture Diagram**



## **Overview**

1. AWS Transfer Family is a managed service for file transfer to/from Amazon S3.
2. Organization provides agencies with username, hostname, and SSH key to log in.
3. Agencies can use FTP tools like FileZilla and WinSCP to connect to the server.
4. Once connected, agencies can securely transfer files to and from Amazon S3.
5. IAM role is assigned for each of the agencies while creating their access.
6. Role will restrict agencies from uploading any other than required file extension i.e., csv, excel and Json.
7. S3 bucket has been encrypted with AWS KMS which will restrict access to all other services except the agencies.
8. EventBridge triggers a Lambda function based on a specified schedule.
9. The Lambda function checks for missed file uploads by agencies for the current day.
10. If any missed file uploads are found, the Lambda function uses SNS to send an email to the SRE team, notifying them of the issue.

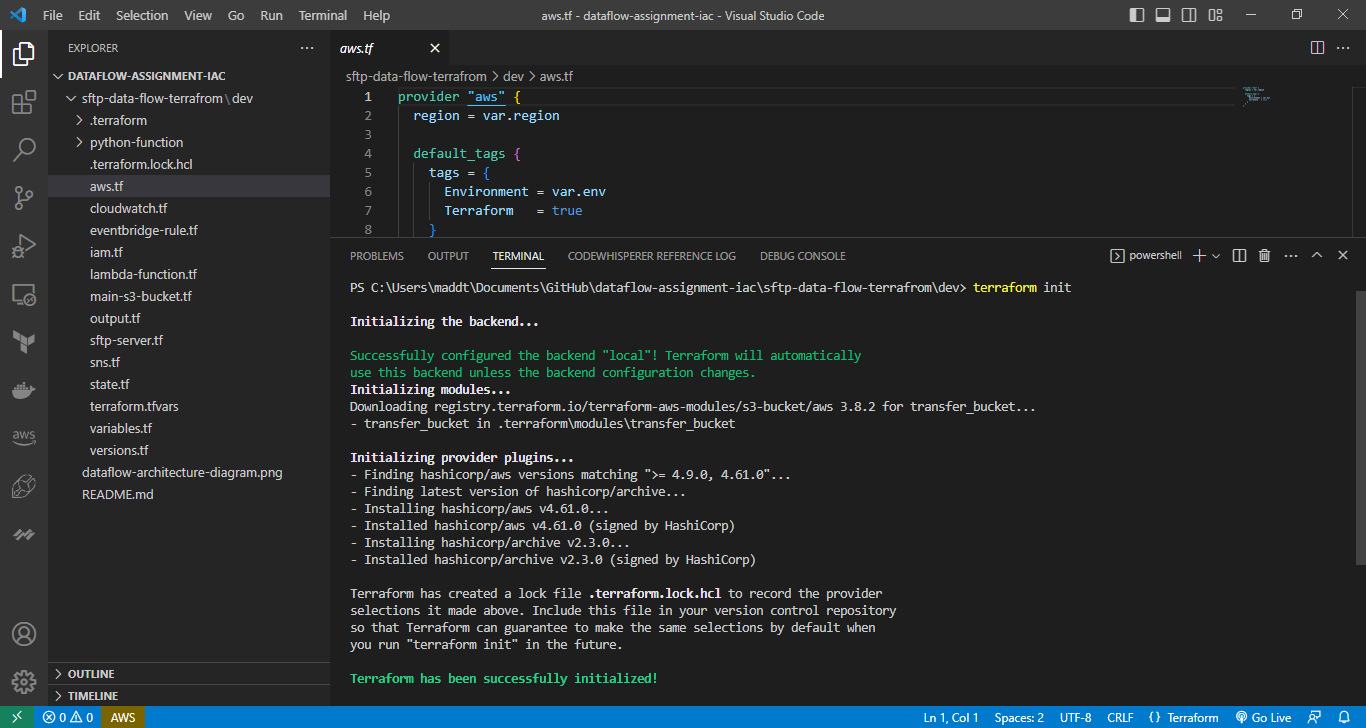
# **Prerequisites**

1. AWS Account
2. IAM User Credentials
3. Install [Terraform](https://developer.hashicorp.com/terraform/tutorials/aws-get-started/install-cli) in local machine
4. Terraform provisioned code for the AWS services

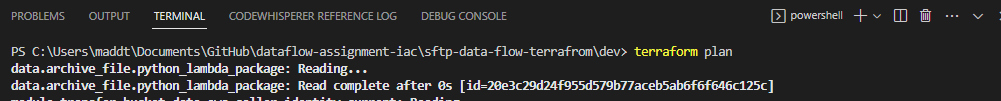
<https://github.com/creazer-I/dataflow-assignment-iac>

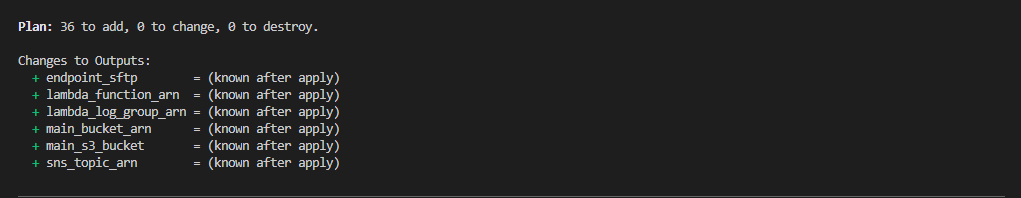
# **Steps to Implement solution using Terraform**

1. Clone the [repository](https://github.com/creazer-I/dataflow-assignment-iac/tree/main/sftp-data-flow-terrafrom/dev) in the local machine, the provided code is the terraform script for AWS services where the solution for the problem stated above is provided.
2. Configure AWS credentials: Terraform requires AWS access and secret keys to authenticate with your AWS account. You can either set these credentials as environment variables or use an AWS CLI profile.
3. Go to vaiables.tf and terraform.tfvars and make changes according to the need, e.g., bucket name, creating users in sftp server.
4. Go to the working directory and initialize it by terraform cli ‘terraform init’, so it would download necessary provider and modules.

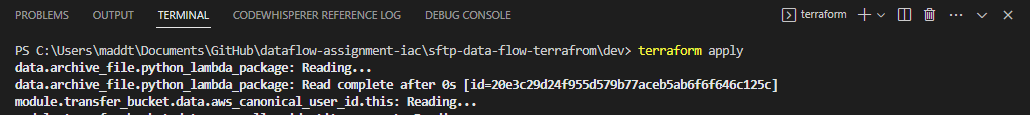


1. Plan the changes: Use the ‘terraform plan’ command to preview the changes that will be made to your infrastructure.





1. Apply the changes: If the plan looks good, use the ‘terraform apply’ command to apply the changes and create the resources in your AWS account.

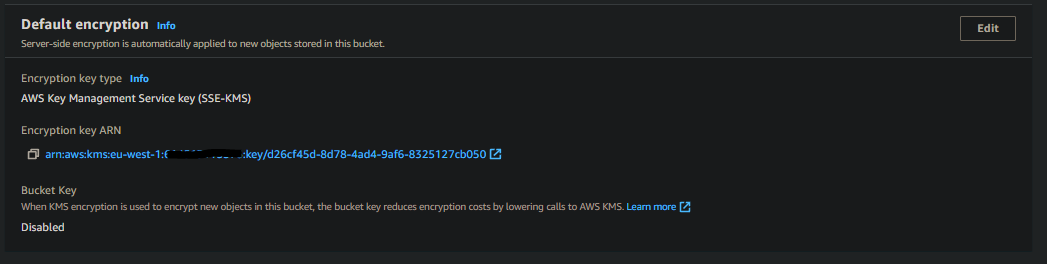




1. Verify the resources: Once the deployment is complete, verify that the resources are created as expected in your AWS account. You can use the AWS Console.

# **Security**

1. S3 bucket have been encrypted with AWS KMS service which does not allow other services to access the bucket except the agencies.



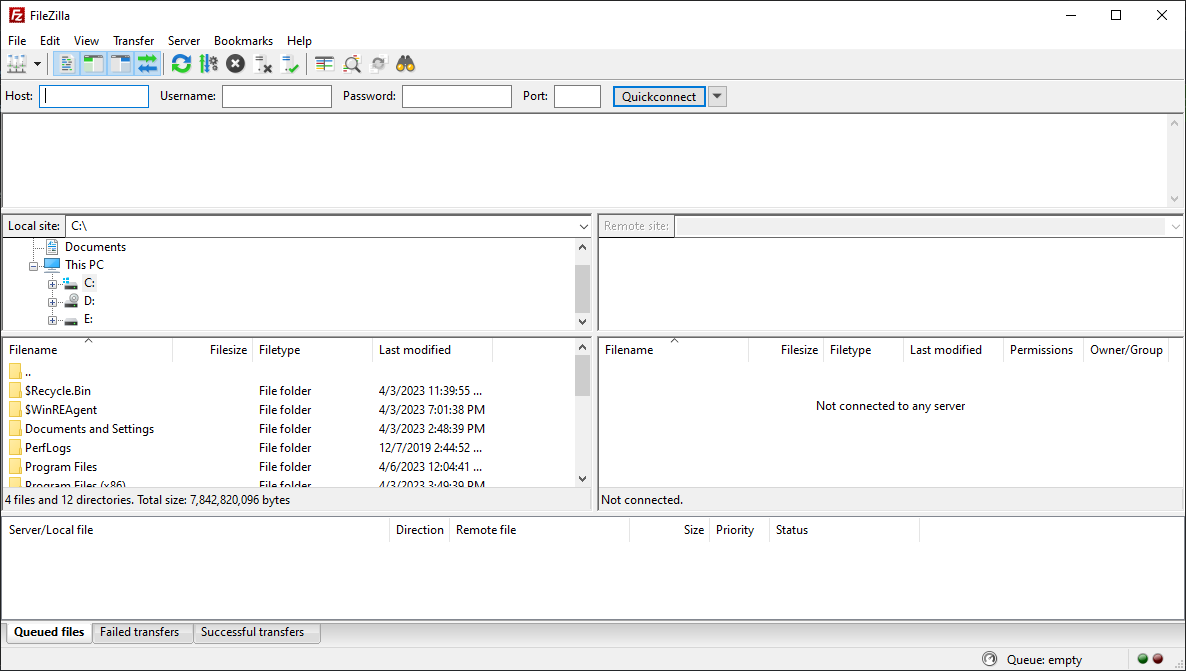
1. Each Agencies have been restricted to only to their directories so they can’t upload files in others directory. The IAM has been provided with least privileges.

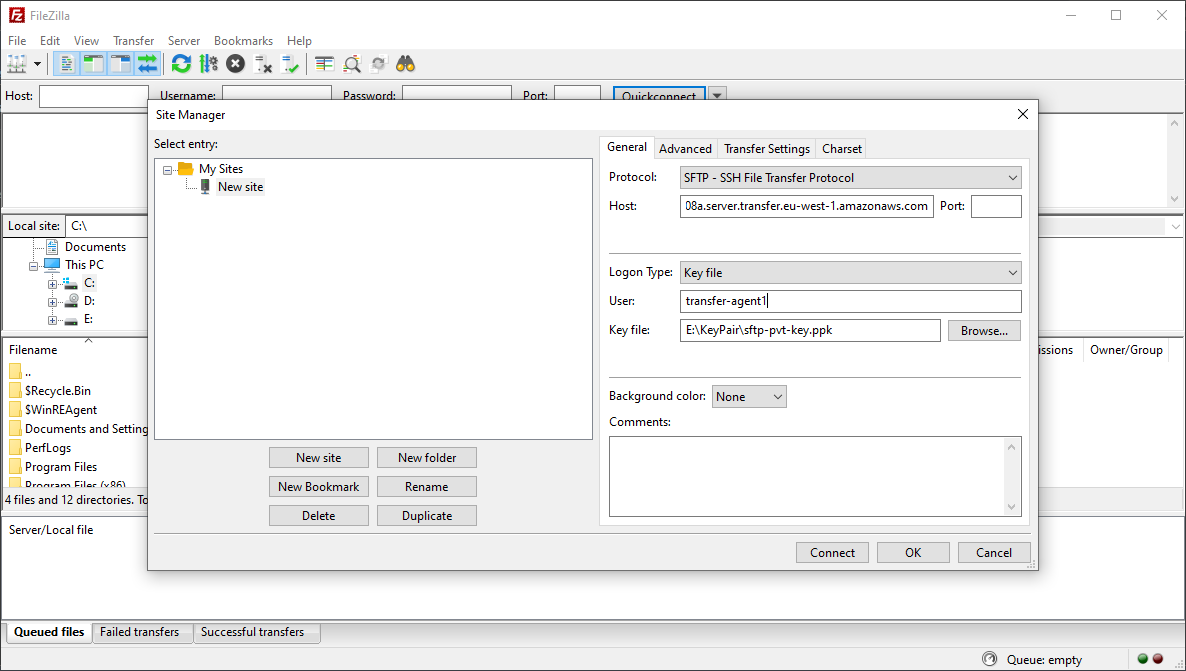
# **Monitoring and Alerting**

1. CloudWatch logs have been enabled for AWS Transfer family to Monitor
2. Eventbridge rule is an event service which will trigger the Lambda function every 5 pm Ireland time which then will investigate the bucket to check for the any data files uploaded for current day if any agencies fail to upload for the day, then lambda function will send an alert using AWS SNS to SRE team every day

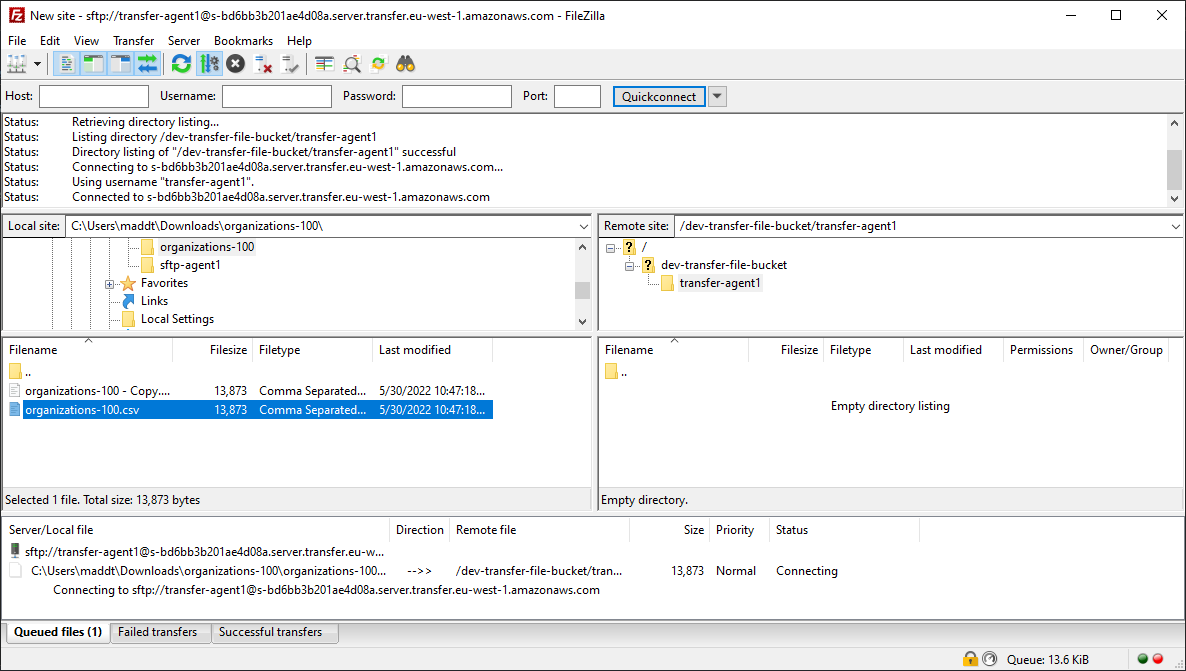
# **Working**

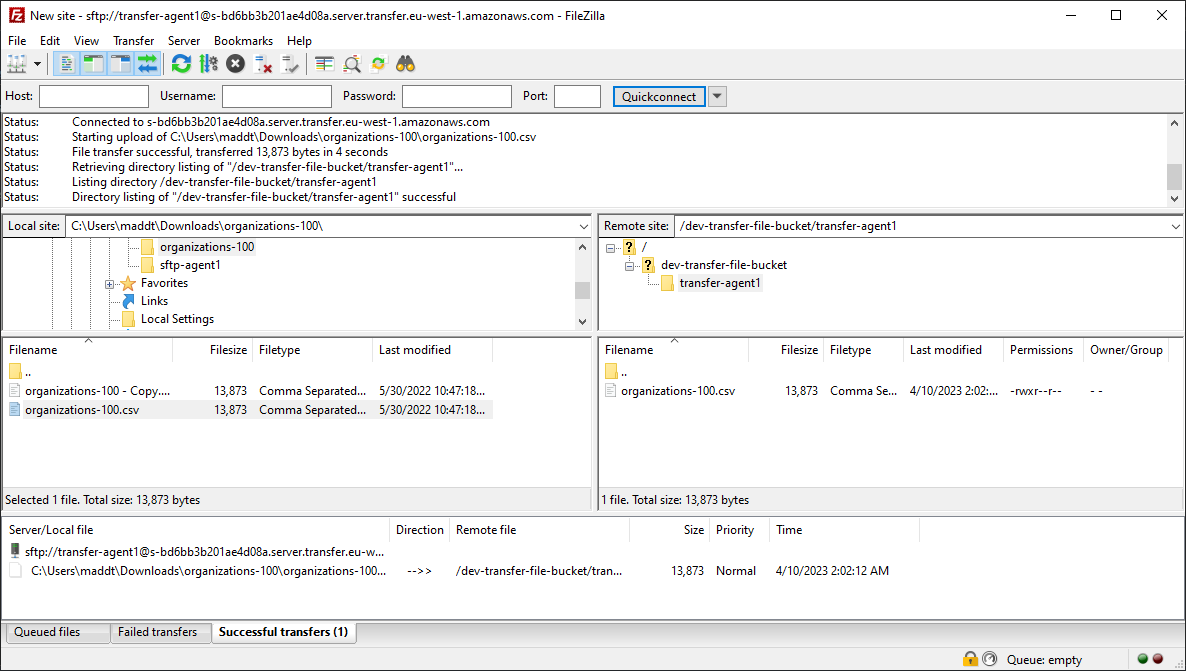
1. Client logs in via ftp tool and uses the credentials provided by the organization





1. After connecting, the agencies will be directed to their directory, and they can start uploading their files into their directory.





1. If any other than csv, excel and Json files are uploaded the files will be rejected.

