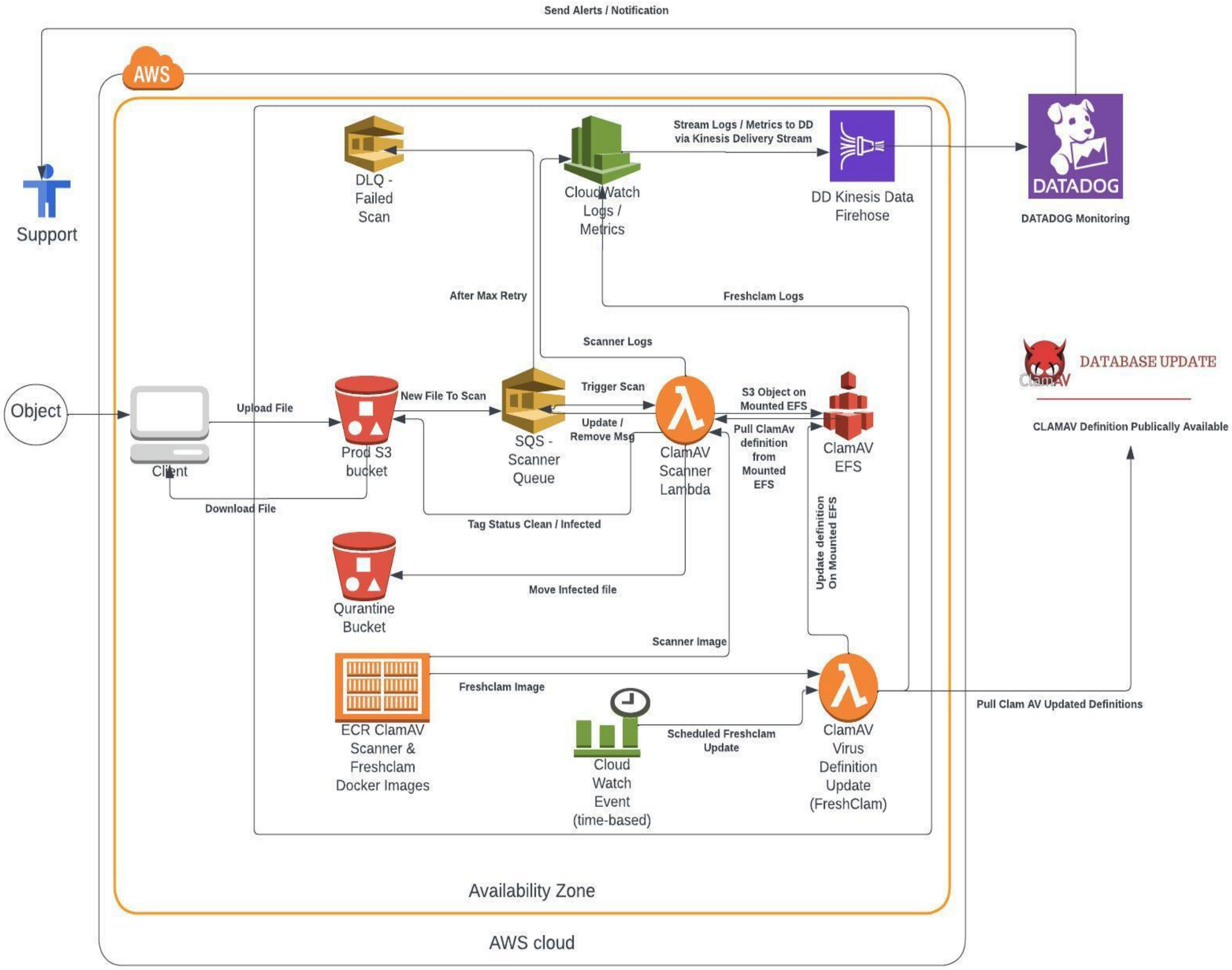
Antivirus for S3 Content

AWS Serverless solution to scan malware file when uploaded to S3 bucket. The solution is built upon Clam AV open-source antivirus feature. Project includes 2 Lambda (Java & Docker) and Terraform script to build serverless infra-structure.

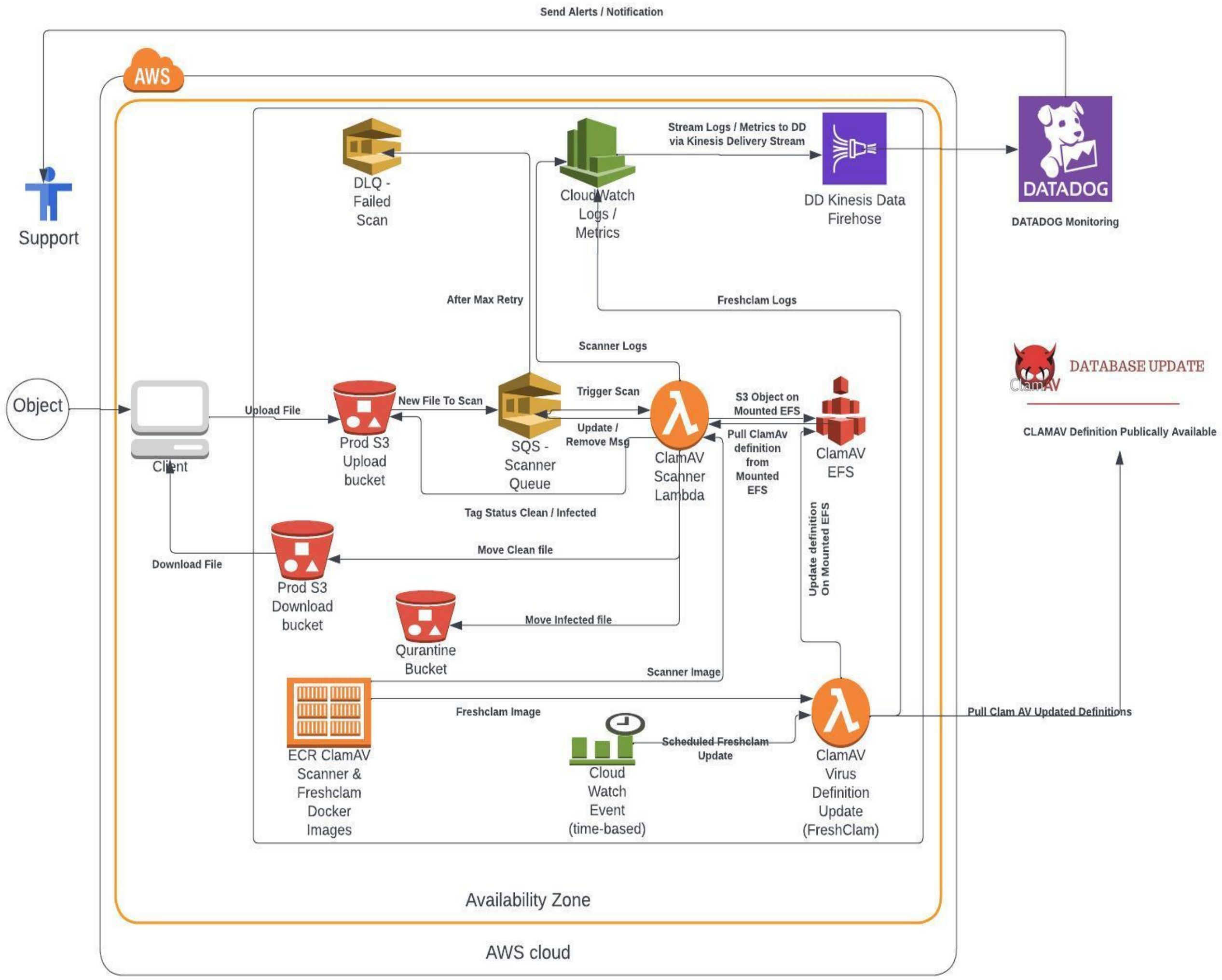
Design / Solution

Architecture Diagram

2 Bucket Solution


Send Alerts / Notification
AWS
Support
DLQ - Failed Scan
CloudWatch Logs / Metrics
Stream Logs / Metrics to DD via Kinesis Delivery Stream
DD Kinesis Data Firehose
DATADOG
DATADOG Monitoring
After Max Retry
Freshclam Logs
Scanner Logs
ClamAV
DATABASE UPDATE
CLAMAV Definition Publically Available
Object
Client
Upload File
Prod S3 bucket
New File To Scan
SQS - Scanner Queue
Trigger Scan
Update / Remove Msg
ClamAV Scanner Lambda
S3 Object on Mounted EFS
Pull ClamAV definition from Mounted EFS
ClamAV EFS
Download File
Tag Status Clean / Infected
Update definition On Mounted EFS
Qurantine Bucket
Move Infected file
Scanner Image
Freshclam Image
ECR ClamAV Scanner & Freshclam Docker Images
Cloud Watch Event (time-based)
Scheduled Freshclam Update
ClamAV Virus Definition Update (FreshClam)
Pull Clam AV Updated Definitions
Availability Zone
AWS cloud

3 Bucket Solution



Explanation

Fresh Clam Scenario (Update Virus Definition)

1. An event is configured in Amazon CloudWatch (time based) to trigger a Fresh Clam lambda.
2. Fresh Clam lambda is built up using docker image which is available on AWS ECR.
3. Fresh Clam Lambda synchronizes Virus definitions from Clam AV public server to predefined mounted EFS location.
4. Logs for the same are being updated to Cloud watch logs and the same has been stream (via Kinesis Firehose stream) to Datadog for monitoring.
5. Failure of Virus definition update can be monitored through Datadog.

Virus Scan Scenario

1. Client uploads the file to pre-defined AWS S3 bucket.
2. A message is added in scanner queue to process new file.
3. A Clam AV Scanner Lambda is triggered for scan new uploaded file.
    1. Clam AV Scanner lambda is built up using docker image which is available on AWS ECR.
    2. If Scanner lambda is not executing, a new execution context is being created and docker image is pulled from AWS ECR.
4. Clam AV Scanner lambda downloads the file from S3 bucket to mounted EFS /temp location.
5. Clam AV Scanner lambda refers the Virus definition which is being updated regularly from Clam AV Fresh Clam lambda. (As defined in previous section)
6. If scan is successful, then,
    1. Clam AV Scanner lambda will add TAG on S3 Object as CLEAN / INFECTED.
    2. A policy can be configured in S3 bucket to download files only which are not tagged as INFECTED.
    3. Message from Scanner queue can be removed.
    4. INFECTED files are moved to Quarantine bucket.
7. Else if scan is unsuccessful, then,
    1. Scanner queue can retry to scan the same file.
    2. If retry is successful, then same process as mentioned in step 7.
    3. If still unsuccessful, even after multiple attempts up to MAX_RETRY (configured in SQS), the message will be moved DLQ (Dead letter Queue).
    4. A DLQ metrics will be stream to Datadog via Kinesis Firehose stream.
8. Logs for scanning lambda is being updated to Cloud watch logs and the same has been stream (via Kinesis Firehose stream) to Datadog for monitoring.
9. Failure of scanning can be monitored through Datadog log / metrics.