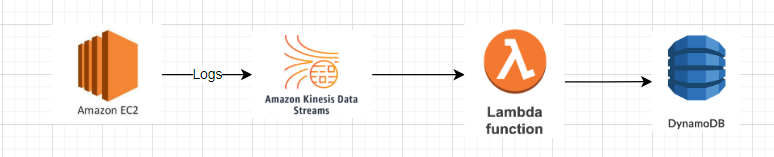
**AIM:** Display real time processing process through AWS Kinesis Data stream and use AWS Lambda function as a consumer to store data into Dynamo DB table.

**Block Diagram:**



**Services Used:** AWS EC2, AWS DynamoDB, AWS Kinesis Data Stream, AWS IAM, AWS Lambda Function.

**Source Data File:** CSV Format.

**Steps:**

1. Create kinesis Data Stream.
   1. Source: Direct Put (Will be using Kinesis Client Application in order to publish data in the delivery stream) from Amazon Kinesis Agent.
   2. Configurations for Data stream:
      1. No. of shards: 1.
2. Set up T2/Micro EC2 Instance to publish the logs into the Firehose Stream.
   1. Choose Amazon Linux AMI as the EC2 instance type.
   2. Generate Key pair to log into the EC2 instance.
   3. Connect to the instance through putty/Terminal.
3. Install Amazon Kinesis Agent in the instance launched.
   1. sudo yum install -y aws-kinesis-agent.
   2. Load the data file and the logs generator script file in the EC2 instance.
   3. Change permissions for the log generator file.
      1. chmod a+x LogGenerator.py. #LogGenerator.py: Log generator script.
   4. Create a log directory in which we will put the log files generated from LogGenerator.py.
      1. sudo mkdir /var/log/<Folder Name>. (Same is being monitored by the data stream and it will pick up the files to stream from the same folder).
4. Configure Kinesis Agent file for the delivery stream.
   1. sudo nano agent.json (in etc/aws/).
   2. Set up the variables accordingly.
   3. Use CSVTOJSON data processing option in order to convert the log file from CSV to json which is easy to consume for the lambda function.
5. Create a role for your EC2 instance to access data stream without creating a user and hence making sure the pipeline is secure.
6. Start the Amazon Kinesis Agent:
   1. sudo service aws-kinesis-agent start.
7. Create a dynamo DB table:
   1. Specify your partition and sort keys according to the data.
8. Create a role for lambda in order it to get access to AWS Kinesis data streams and dynamo DB table.
9. Create a Lambda function (Using python for this POC):
   1. Select the role created.
   2. Add a trigger of AWS Kinesis data stream to the lambda function in order to feed the logs into the function.
      1. Consumer: None (as custom consumer is used to process the log files).
      2. Batch size: 100 (Largest number of records that will be read from your stream once).
10. Generate logs using the script LogGenerator.py:
    1. sudo LogGenerator.py 100 #100 is the number of lines to be published.
    2. Check the log for Kinesis agent:
       1. tail -f /var/log/aws-kinesis-agent/aws-kinesis-agent.log