**Convert Raw Data to Formatted Data**

**Idea:**

To design a pipeline which triggers the lambda function to format the raw data in a streaming manner.

**Pipeline:**

**Services Involved:**

1. AWS S3
2. Simple Notification Service (SNS) topic
3. Simple Queue Service (SQS)
4. AWS Lambda
5. IAM

**Steps Included:**

1. Create a SNS Topic
2. Create a SQS Queue
3. Subscribe the Queue to a SNS topic
4. Create a Lambda fn to convert JSON file in raw folder to CSV format and place in format folder
5. Add permissions to IAM role associated with Lambda fn for receiving SQS msg, putting files into S3
6. Add SQS queue as trigger source for Lambda fn
7. Create a Event notification for S3 bucket raw folder to send notification destined to SNS topic whenever new file inserted into the folder

**Procedure Followed:**

To understand how the message is getting travelled, Broken down the pipeline into 3 steps

 ***Step-1:*** *Only included Lambda, S3 in pipeline*



1. Created lambda function to format the data in the raw folder of S3 bucket
2. The code is as follows: [real\_time\_project/lambda\_fn\_json\_to\_csv.py at main · winnu3103/real\_time\_project (github.com)](https://github.com/winnu3103/real_time_project/blob/main/lambda_fn_json_to_csv.py)
3. Create Event notification for S3 bucket raw folder destined as lambda to get triggered
4. Add S3 as the trigger source to the lambda function
5. Give permissions to IAM role of lambda function for read and write to the s3

***Step-2:*** *Included Lambda, S3, SNS in pipeline*



1. Created lambda function to format the data in the raw folder of S3 bucket
2. The code is as follows: [real\_time\_project/lambda\_fn\_SNS.py at main · winnu3103/real\_time\_project (github.com)](https://github.com/winnu3103/real_time_project/blob/main/lambda_fn_SNS.py)
3. Create SNS topic
4. Create Event notification for S3 bucket raw folder destined as SNS topic
5. Add SNS as the trigger source to the lambda function
6. Give permissions to IAM role of lambda function for read and write to the s3, receive notification from SNS
7. SNS will send notification to lambda function to get it triggered whenever there is new data in the raw folder of S3 bucket.

***Step-3:*** *Included the entire pipeline i.e., Lambda, S3, SNS in pipeline*



1. Created lambda function to format the data in the raw folder of S3 bucket
2. The code is as follows: [real\_time\_project/lambda\_fn\_SQS.py at main · winnu3103/real\_time\_project (github.com)](https://github.com/winnu3103/real_time_project/blob/main/lambda_fn_SQS.py)
3. Create SNS topic, SQS Queue
4. In SNS, Create a Subscription to the SQS queue
5. Create Event notification for S3 bucket raw folder destined as SNS topic
6. Add SQS as the trigger source to the lambda function
7. Give permissions to IAM role of lambda function for read and write to the s3, receive notification from SQS
8. SNS will send notification to SQS and the queue will then send messages to lambda function to get it triggered whenever there is new data in the raw folder of S3 bucket.

**Testing:**

* Inserted a new data in the Raw folder of the s3 bucket
* Checked the delivery status logs of SNS topic to check whether it is receiving and sending the notification or not.
* Added loggers in the lambda function, and watched the logs in AWS CloudWatch to check whether it is receiving message and processing without errors or not.
* Finally checked the format folder in s3 bucket for CSV format file of the JSON file inserted into raw folder.
* If CSV file is found in format folder then the pipeline is working fine.