1. Set S3 bucket
2. Create bucket: [raw-data-sc171](https://us-east-1.console.aws.amazon.com/s3/buckets/raw-data-sc171?region=us-east-1&bucketType=general)
3. Upload [airbnb\_ratings\_new.csv](https://us-east-1.console.aws.amazon.com/s3/object/raw-data-sc171?region=us-east-1&bucketType=general&prefix=airbnb_ratings_new.csv) to [raw-data-sc171](https://us-east-1.console.aws.amazon.com/s3/buckets/raw-data-sc171?region=us-east-1&bucketType=general)
4. Create bucket: [myresult-sc171](https://us-east-1.console.aws.amazon.com/s3/buckets/myresult-sc171?region=us-east-1&bucketType=general)
5. Create bucket: [my-code-sc171](https://us-east-1.console.aws.amazon.com/s3/buckets/my-code-sc171?region=us-east-1&bucketType=general)
6. Upload: [ml\_code.py](https://us-east-1.console.aws.amazon.com/s3/object/my-code-sc171?region=us-east-1&bucketType=general&prefix=ml_code.py) to [my-code-sc171](https://us-east-1.console.aws.amazon.com/s3/buckets/my-code-sc171?region=us-east-1&bucketType=general)
7. Create bucket: [ml-model-sc171](https://us-east-1.console.aws.amazon.com/s3/buckets/ml-model-sc171?region=us-east-1&bucketType=general)
8. Glue Job
9. go to ETL job
10. create Script Editor
11. Select upload script
12. upload etl\_job.py
13. Job detail
14. named etl\_job
15. Labrole
16. save
17. Set Glue Clawer
18. Create glue clawer
19. Named data\_crawler
20. Data source configuration, Select no yet
21. Add Data sources: s3://raw-data-sc171/processed
22. IAM role: LabRole
23. Add database, named data\_db
24. Target database: data\_db
25. Create crawler
26. Create lambdas
27. [1\_ingest\_data](https://us-east-1.console.aws.amazon.com/lambda/home?region=us-east-1#/functions/1_ingest_data)
28. [2\_get\_output](https://us-east-1.console.aws.amazon.com/lambda/home?region=us-east-1#/functions/2_get_output)
29. [3\_ml\_process](https://us-east-1.console.aws.amazon.com/lambda/home?region=us-east-1#/functions/3_ml_process)
30. [4\_refresh\_asg](https://us-east-1.console.aws.amazon.com/lambda/home?region=us-east-1#/functions/4_refresh_asg)

Reminder: Need to change runtime and memory size

1. Set Auto Scaling Group

(1). [Launch Templates](https://us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#LaunchTemplates:)

a. template name: ml\_web

b. QuickStart: Amazon Linux

c. Instance type: t2.micro

1. Key pair (login): optional
2. Network Setting: Create security group -> Security group name: ml\_sg , give port 22, 8000, 80, 3306. anywhere access
3. Advanczed detail -> IAM instance profile: LabInstanceProfile
4. User data -> copy userdata.sh

(2). Target Group

1. Choose a target type -> Instance
2. Target group name: ml-tg-sc171
3. Protocol : Port -> 8000
4. Protocol version -> HTTP1
5. Health checks: /
6. Advanced health check settings: Traffic port, Healthy threshold:5, Unhealthy threshold:10, Timeout:10, Interval:300, Success codes: 200-400
7. Create target group

(3). Load Balancers

1. Create load balancer
2. Select Application Load Balancer
3. Load balancer name: alm-ml-sc171
4. Scheme: Internet-facing
5. Security groups: ml\_sg
6. Listeners and routing: **Protocol -> HTTP, Port->80, Default action->** ml-tg-sc171

(4). Auto Scaling groups

a. Named asg\_ml

b. Launch template -> ml\_web, version->default

5. SNS

1. Create topic
2. Type: Standard
3. Name: query\_result
4. create topic
5. create subscription
6. Protocol -> email
7. Endpoint -> [jimmysc171@outlook.com](mailto:jimmysc171@outlook.com)
8. create subscription
9. change arn in lambda

6. RDS

open public access

make sure subnet is public

security group has inbound: 3306 port

**(vpc:igw, security group:3306, public access:yes)**

launch template

security group

sns

set etl

s3

Step functions

{

"bucket": "raw-data-sc171",

"prefix": "airbnb\_ratings\_new.csv"

}

|  |
| --- |