

AWS ORDERS PIPELINE - BEGINNER (NOOB) STEP BY STEP GUIDE

1) LOGIN + REGION

- Login to AWS Academy.
- Top right region: select N. Virginia (us-east-1)

2) CREATE S3 BUCKET

- Go to S3
- Create bucket name: omsai-traya-bucket
- Create folders: raw , processed , analytics

3) CREATE VPC

- Go to VPC → Create VPC wizard
- Select VPC + 2 public + 2 private subnets
- Name: orders-pipeline-vpc-vpc

4) CREATE SECURITY GROUPS

A) lambda-rds-sg

- inbound: MYSQL 3306 from itself
- inbound: MYSQL 3306 from ec2-analytics-sg

B) ec2-analytics-sg

- inbound: SSH 22 from 0.0.0.0/0

5) CREATE RDS

- MySQL , db.t3.micro
- NO PUBLIC ACCESS
- VPC: orders-pipeline-vpc-vpc
- SG: lambda-rds-sg
- user: admin
- password: YOUR_PASSWORD
- endpoint: ordersdb.....amazonaws.com

6) CREATE LAMBDA

- python 3.12
- attach to private subnets
- SG = lambda-rds-sg
- environment vars set
- add pandas aws sdk layer
- upload pymysql layer

7) SET S3 TRIGGER

- S3 → bucket → raw/
- add trigger to lambda

8) CREATE SNS TOPIC

- subscribe your email
- then create VPC endpoint for SNS
- service: com.amazonaws.us-east-1.sns
- subnets: private
- SG: lambda-rds-sg

9) LAUNCH EC2

- amazon linux 2023 t3.micro
- public subnet
- SG ec2-analytics-sg
- connect via mobaxterm
- install pip + pymysql + boto3

10) COPY pipeline.py TO EC2

- run python3 pipeline.py
- It uploads analytics json

11) SET CRON

```
echo "0 */12 * * * /usr/bin/python3 /home/ec2-user/pipeline.py >> /home/ec2-user/pipeline.log  
2>&1" > mycron
```

```
crontab mycron
```

12) COST SAVING

- Stop EC2 when not using

- Stop RDS temporarily
- Release EIP