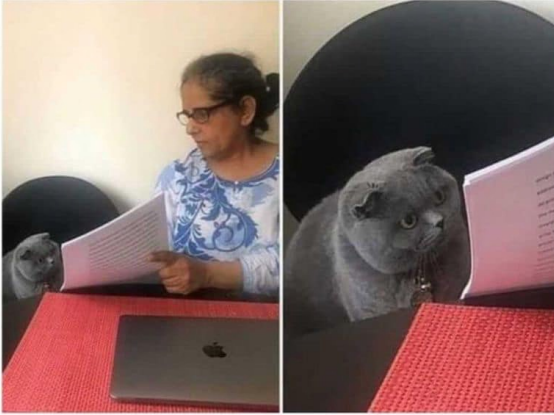# Deploy On Ecs

My manager showing me the code I wrote that took down the whole website

When you forget to turn off your AWS instance and receive your AWS bill

Agenda

Brief   Overview   of   Aws

To get our hands dirty → Deploy on EC2
                         → Deploy on ECS

Why Remote Computers
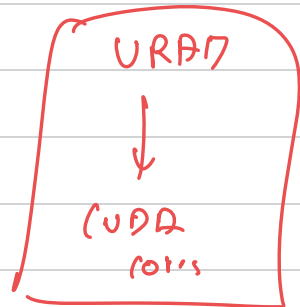
3.5L { GPU → RTX -5090
URAM-32-GB
14 B Param
mo

5 2014th

CPU → 9950x 3D
2016   16 - 6
              32 th

RAM → 128 GB
          DDR5
          70 k

Flipkart
2014

GPU →   A100 → 80 GB URAM-11J.
                GPT-4 — 25k

< 20%

CPU ——→ GPU | URAM
                          | ↓
                          | CUDA
                          | cores

1. AWS - Amazon
2. Azure - Microsoft
3. GCP - Google
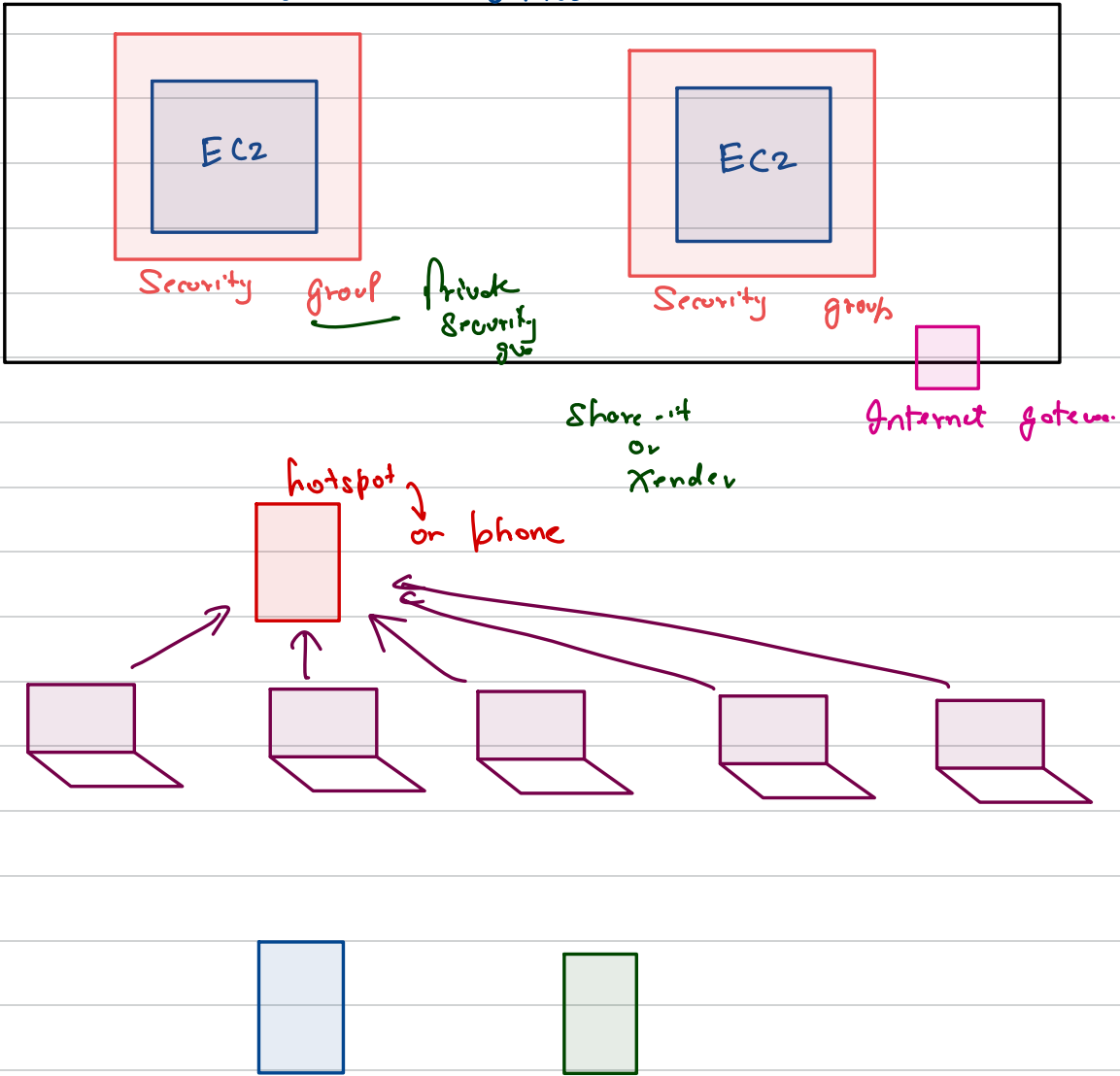4. Digital Ocean
5. Heroku
6. IBM
7. Oracle

AWS   provides → Raw Infrastructure   EC2

Pre- Configured platforms → ECS

AWS Services

1. EC2 -> Elastic Compute Cloud (Blank Online Server)
2. S3 -> Google Drive (Cheaper Than Google Driver > 100 GB Simple Storage Service
3. ECS - Elastic Container Service
4. ECR - Elastic Container Registry (DockerHub equivalent for AWS)
5. SageMaker - Jupyter Notebook inside AWS
6. EMR - Distributed Computing - Apache Spark
7. IAM - Controls access across entire AWS system (Identity and Management)

UPC → Virtual Private Cloud

EC2

Security group Private Security gp

EC2

Security group

Share-it or Xender

Internet gateway

hotspot or phone

EC2

① Rent Server For 1 day
   Pay For entire day

ECS

② Pay as you go

Spain → Ronda →   Flat → 50K per mont
                    Hotel → 5K per day
                            ⌣
                            30 days

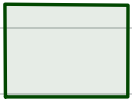2 day → Hotel   Flat   1.5 laks
        10K   50K

1 month   1.5 L   50K.

② Server → ↓ 1000)   ECS → autoscaling
              Crash

Autoscaling

① horizontal auto scaling →   Inc no of Serves

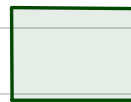② Vertical   ''   ''   →   Inc configuration
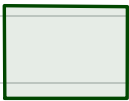   of server.

12 GB RAM
512 GB SSD
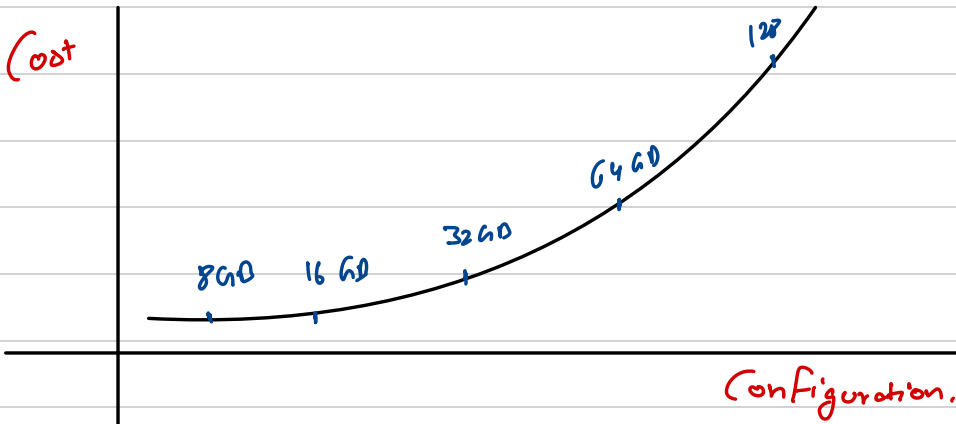
12 GB RAM
512 GB SSD

12 GB RAM
512 GB SSD

horizontal

12 GB RAM
512 GB SSD

128 GB
4 TB SSD

Vertical.

Cost

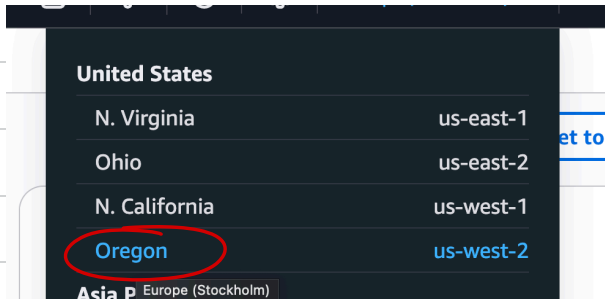8GB   16 GB   32GB   64 GB   128

Configuration.

Steps for Model Deployment (Platform Agnostic)

1. Containerize the model
2. Push the container to Registry
3. Define how to run it (CPU/memory needs, ports, environment variables, etc)
4. Setup Compute Platform, e.g. VM, an EC2.
5. Deploy and configure Network/Security.
6. Monitor

Same Steps in Context of AWS

1. Authenticate our computer with AWS, using IAM.
2. Containerize and push to ECR.
3. Create Task Definition
4. Create ECS Cluster
5. Deploy and configure security-group.

United States

N. Virginia	us-east-1
Ohio	us-east-2
N. California	us-west-1
Oregon	us-west-2

Asia P. Europe (Stockholm)

https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html

```
docker build --platform=linux/amd64 -t
flask_app .
```

Which AWS service is responsible for managing containers and enabling developers to run applications without configuring the environment?

0 users have participated

A	Amazon Elastic Container Registry (Amazon ECR)	0%
B	Amazon Elastic Compute Cloud (Amazon EC2)	0%
C	Amazon Elastic Container Service (ECS)	0%
D	Amazon Simple Storage Service (Amazon S3)	0%

End Quiz Now

What does the term "repository" signify in the context of Amazon ECR?

0 users have participated

	A	A server hosting web applications	0%
✓	B	A storage location for Docker or OCI images	0%
	C	A database for storing application data	0%
	D	A collection of virtual machines	0%

End Quiz Now

What purpose does Amazon Elastic Container Registry (Amazon ECR) primarily serve in the AWS ecosystem?

0 users have participated

	A	Managing serverless functions	0%
	B	Configuring virtual machines	0%
	C	Controlling access to S3 buckets	0%
✓	D	Storing and managing container images securely	0%

End Quiz Now

What is the primary purpose of AWS Fargate in Amazon ECS?

0 users have participated

	A	Managing S3 buckets	0%
	B	Securing virtual machines	0%
✓	C	Running containers without managing underlying infrastructure	0%
	D	Configuring IAM roles	0%

End Quiz Now

What is the primary function of a virtual private cloud (VPC) in AWS?

16 users have participated

A	Storing and managing virtual machine images	6%
B	Providing a dedicated virtual network for an AWS account	94%
C	Optimizing database queries	0%
D	Enabling high-performance computing	0%

End Quiz Now

SD
2

SK
1

3

🏆 **Sagar Deka**
5/5 ⚡ 466.59

🏆 **Santhosh K**
5/5 ⚡ 480.70

🏆 **kiran kumar**
5/5 ⚡ 442.00

4	AU	Ashwini Uday	5/5	⚡ 424.72
5		Rajaraman	4/5	⚡ 371.42
6	RS	Renu Singh	4/5	⚡ 365.10
7	CS	Chirantan Sharma	4/5	⚡ 354.16
8		Praveen MC	4/5	⚡ 353.00
9	L	Lokesh Sonawane	4/5	⚡ 350.17
10		Dinesh Patel	4/5	⚡ 348.67

✓ **Tasks launched**   ✕

○ arn:aws:ecs:us-west-2:261558455174:task/live-june13-ecs-cluster-a3/3b4ef980602543bcb2553c5525568561

3b4ef980602543bcb2553c5525568561

Last updated
13 June 2025 at 08:54 (UTC+5:30)   ↻   Stop

⊗ **Task is stopping**   ⊙ Diagnose with Amazon Q

ResourceInitializationError: unable to pull secrets or registry auth: The task cannot pull registry
auth from Amazon ECR: There is a connection issue between the task and Amazon ECR. Check your task
network configuration. RequestError: send request failed caused by: Post "https://api.ecr.us-west-
2.amazonaws.com/": dial tcp 34.223.27.76:443: i/o timeout

View troubleshooting guide 🔗

Configuration   Logs   Networking   Volumes (0)   Tags