MODULE 6: Compute Services Overview.

1. Compute service overview

a. EC2
- Provides VM
- Instance based service
- IAAS

b. AWS LAMBDA:
- Low cost
- Serverless computing
- Function based

c. AMAZON ECS, EKS, FARGATE, ECR
- Container based computing and instance based computing

d. AMAZON BEANSTALK
- PAAS
- Web application used

How to choose optimal service?
- It will depend on the use case.
- Consider 3 aspects :
    1. What is application design
    2. Usage pattern
    3. Configurations service you want to manage
- If we choose the wrong compute service, then work efficiency will be low.
- It is good practice to understand this before picking

1. Amazon EC2

EC2:
- Web server
- application server
- Mail server
- Game server
- Media server
- Database server
- Catalogue server
- File server
- Computing server
- Proxy server
- Provides VM
- You can launch as many instances of any size into any AZ anywhere in the world.

- You can control traffic to and from the instances.

AWSCompute services

1. Ec2
2. EC2 autoscaling
3. Elastic Container Registry
4. Amazon Elastic Container Service
5. VMWare Cloud on WAs
6. WAs Elastic Beanstalk
7. Lamba
8. Elastic Kubernetes Services
9. Lightsail
10. AWS Batch
11. aWS fargate
12. AWS outposts
13. AWS Serverless Application repository

Launch EC2:

- AMI
- Instance Type
- Network setting
- IAM Role
- User Data
- Storage Option
- Tags
- Security Groups
- Key Pairs
- Option to launch Instance with AWS CLI

Cloudwatch is used to monitor the instances.

- Basic : every 5 min, free of charge
- Detailed : every 1 min, fix charges

2. EC2 Cost Optimisation

6 types of pricing models:

- On demand instances: pay by the hour, no long term commitments, eligible for AWS tier.
- Reserved Instances: full partial or no upfront payment for instance you reserve, discount on hourly charge for that instance, one year term or three year term.
- Dedicated host: a physical server with ec2 instance, capacity fully dedicated to your use.

- Dedicated instances: tsnaces that run in a vPC or hardware that is dedicated to a single customer.
- Scheduled reserved instance: 1 year term, purchase a capacity reservation that is always available on a recurring schedule you specify.
- Spot instances: it runs as long as they are available, they can be interrupted by AWS with a 2 min notification.

Note: per second billing available for on-demand, reserved, spot instances that run Amazon Linux or Ubuntu.

Benefits of pricing models:

- On-demand: low cost and flexibility
- Spot: large scale and dynamic workload
- Reserved: predictability ensures compute capacity is available when needed.
- Dedicated host: save money on licensing cost, help meet compliance and regulatory requirements.

Use cases of pricing models:

- On-demand: spiky workloads
- Spot : time insensitive workload
- Reserved instance: steady state workload
- Dedicated host: highly sensitive workloads, BYOL

4 pillars of Cost optimisation:

- Right size
- Increase elasticity
- Optimal pricing model
- Optimise storage choice

3. Containers Services

Containers: a method of operating system virtualisation

Benis:

- Repeatable
- Self contained environment
- Software runs the same in different environment
- Faster to launch and stop or terminal than VM

Docker:

- Software platform that enables you to build Test and deploy application weekly
- You run containers on docker

Elastic Container Service:

- High scalable
- Fast container management service

Kubernetes:

- Open source software for container orchestration
- Automates container provisioning, networking, load distribution, scaling

EKS:

- Enables you to run kubernetes on AWS
- Supports linux and windows containers
- It is used to manage clusters to manage clusters of EC2 instances, run containers that are orchestrated by kubernetes on those instances.

ECR

- It is fully managed docker container registry that make it easy for developer to store, manage, and deploy docker container images

4. Introduction to AWS Lambda

Serverless compute service build to run and build applications.

Benefits :

- Supports multiples programming languages
- Completely automated administration
- Built in fault tolerance
- Supports of orchestration of multiple functions
- Pay per use pricing

AWS Lambda Event Source:

- S3
- DynamoDB
- SNS: simple notification service
- SQS: simple queue service
- API Gateway
- Application Load Balancer

Configure LAMBDA function:

- Function code
- Dependencies
- Execution role

AWS Lambda Quotas:

- Soft limits per region: concurrent execution = 1000 & function and layer storage = 75GB
- Hard limits for individual function: max function memory allocation 3008 Mb & and function time out 15 min & deployment packet size = 250 Mb unzipped.

5. AWS Elastic Beanstalk:

Easiest way to run web application

It automatically handles few things:

- Infrastructure provisioning and configuration
- Deployment
- Load balancing
- Automate scaling
- Health monitoring
- Analysis and debugging
- Logging
- Free of charges

Support web application written in :

- Java
- PHP
- .net
- Python
- node js
- go
- Ruby
- docker

Benefits:

- Fast and simple to start using
- Developer productivity
- Difficult to outgrow
- Complete resource control