CloudAcademy

Revision Cards

Solution Architect Associate Certification 2017

REVISION CARDS – HIGH AVAILABILITY

A region is a physical geographical location made up of groups of data centres Each region is designed to be independent and isolated from another.

Each region has two or more locations that are known as Availability Zones.

Availability zones are designed to be isolated and independent from failure of another AZ within a region. AZ’s have low latency network connectivity to other AZ’s within a region.

AZ’s enable customers to operate highly available, fault tolerant, scalable systems which wouldn’t be possible with a single data centre. Spreading services across AZ’s is vital, spreading across regions needs to be a considered choice as it involves using the public internet and may attract additional costs.

Amazon CloudWatch is a vital service for managing highly available environments. CloudWatch provides system wide visibility in to resource utilization and operational health. Can also monitor application resources using detailed monitoring / an installed agent.

Amazon Simple Queue Service allows you to decouple applications and services from one another improving durability.

VPC – private CIDR block of the AWS Cloud

Amazon DynamoDB is a fully managed NoSQL database service.

Amazon Route53 –simple, failover, latency based routing, geo-based routing

REVISION CARDS – AUTO SCALING

Must have elements of an Auto Scaling group are
Minimum size
Launch configuration
Health checks and desired capacity are optional

Must have elements of an Auto Scaling launch configuration are:
launch configuration name
AMI
Instance type

The default number of instances you can launch per region is 20

The default limit for launch configurations is 100 per region.

Run **AWS AUTO SCALING describe-account-limits**

ELBs check the health of an AWS resource, the ELB does not terminate an instance if it is unhealthy, that is done by the Auto Scale group

Auto Scaling designed to scale out based on an event like increased traffic and scale in when traffic drops off

CAN be used to control steady state workloads that need a consistent number of Amazon EC2 instances at all times with MIN and MAX values

Scheduled scaling is where you set a date / time combination to have AUTO SCALING increase an autoscale group based on predictive traffic patterns

An AUTO SCALING launch config cannot add an already running instance to an AUTO SCALING group

LAUNCH
Adds a new EC2 instance to the autoscale group, increasing its capacity

TERMINATE
Removes an EC2 instance from the group, decreasing its capacity

HEALTHCHECK
Checks the health of the instances. Auto Scaling marks an instance as unhealthy if Amazon EC2 or Elastic Load Balancing tells Auto Scaling that the instance is unhealthy. This process can override the health status of an instance that you set manually

REPLACEUNHEALTHY
Terminates instances that are marked as unhealthy and subsequently creates new instances to replace them

AZREBALANCE
Balances the number of EC2 instances in the group across the Availability Zones in the region

Auto Scaling launches new instances in an unaffected Availability Zone before terminating the unhealthy or unavailable instances

Auto Scaling launches an instance in the Availability Zone with the fewest instances unless told otherwise

If you suspend TERMINATE, your Auto Scaling group can grow up to 10% larger than its maximum size! Auto Scaling allows this growth during rebalancing activities

ALARMNOTIFICATION
Accepts notifications from CloudWatch alarms that are associated with the group

ADDTOLOADBALANCER

Adds instances to the attached load balancer or target group when they are launched

REVISION CARDS – ELASTIC LOAD BALANCER

Elastic load balancer takes requests and distribute traffic across your AWS resources

two types of ELB - the **CLASSIC ELB** and the **APPLICATION ELB**

Application Load Balancer operates at Layer 7 of the OSI model. Classic ELB operates at **Level 4** of the OSI model

At Layer 7, the application ELB can inspect application-level content, not just IP and port. So the application ELB lets you set more advanced rules and checks

ELBs support HTTP HTTPS TCP SSL

FEATURES OF BOTH CLASSIC AND APPLICATION LOAD BALANCERS
Native IPv6 support - both can support IPv4 and IPv6 but only Application Load Balancers support IPv6 in the VPC
Sticky sessions - a way to route requests from the same client to the same target
Health checks - routes traffic to healthy targets
Connection draining – graceful scale in / out - 300 SECONDS DEFAULT
HTTPS support - supports HTTPS termination between the clients and the load balancer

APPLICATION LOAD BALANCERS
Native IPv6 support within the VPC - IPv6 needs an AAAA NAME record
Container support - load balance containers across multiple ports on a single EC2 instance
Content-based Routing - can route a request to a service based on the content of the request
HTTP/2 support - Speeds up connections and page download times
Web sockets support - allows a server to exchange real-time messages with end-users
Delete protection - You can enable deletion protection - instances behind a deleted ELB will continue to run

An ELB SSL security policy definition requires -
SSL Protocols
SSL Ciphers
Server Order Preference
It does NOT require Client Order Preference

Need an X509 certificate for SSL
Does not support a client-side SSL
ELB SSL does NOT support TLS 1.3

Settings
idle timeout
cross-zone
connection draining
proxy protocol
sticky sessions
health checks

Listeners - Every listener is configured with a protocol and a port for a front-end connection and a protocol and a port for the back-end connection

CloudAcademy

REVISION CARDS – CLOUDWATCH

BASIC monitoring is free

CloudWatch keeps data for 2 WEEKS by default

Can have up to 5,000 CloudWatch ALARMS per account

To monitor provisioned read limit on DynamoDB Table

GET method to request **ProvisionedReadCapacityUnits** metric

Create a **threshold level** with an **alarm** set when a consecutive number of periods is crossed for that dynamoDB table

DETAILED monitoring attracts an additional cost

DETAILED monitoring enables you to aggregate metrics you define

CloudWatch does NOT aggregate data across REGIONS

CloudWatch CAN aggregate data across AVAILABILITY ZONES within a region

DETAILED monitoring enables you to set a TIME value

DETAILED monitoring enables you to report on Hypervisor visible data

EXCLUDES MEMORY
Reporting on memory utilisation is possible in CloudWatch however it requires you install a reporting agent on your EC2 instances to send custom data to CloudWatch

REVISION CARDS – ROUTE53

Domain names are registered with domain registrars that then register the domain name with InterNIC, a service of ICANN

Each domain name is registered in a central WhoIS database

Domains are defined by their top level domains (TLD) TLD's are controlled by IANA in a root zone database

Amazon Route53 enables you to register top level domains (TLDS)

Failover
traffic from your resources in a primary location to a standby location(TLDS)

Route53 organises your DNS records in to hosted zones

Zone records consist of any of the DNS supported domain extensions e.g.
ANAME - root record - amazon.com
CNAME - alias - www.amazon.com
MX - mail exchange -mx.amazon.com

Latency-Based
Used to route your traffic based on the lowest latency

Simple
Good for a single resource eg www.amazon.com

Geolocation
Used to route your traffic based on your end user's location

Weighted
route a percentage of your traffic to one resource

An AAAA record is used to route traffic to an IPv6 address
An ANAME record is used for an IPv4 address

A PTR record is used for reverse DNS

REVISION CARDS – DATA SECURITY

AWS secures the infrastructure, customers secure anything that goes in it

IAM is NOT an identity store

IAM is NOT an authorisation system for your application

Authentication is via user name and password via the console

IAM is a web service that enables customers to manage AWS users and AWS user permissions

IAM is NOT a way to manage permissions within your applications

Applications access service API's with an IAM user using a two part access key

MFA is multi factor authentication increases account security by adding a device specific one time password

ROOT user - associated with the account, cannot be restricted in any way

IAM users - persistent identities controlled via IAM

Roles - provide temporary access with different credentials

JSON doc includes
Effect
Service Name
Action
Resource

A policy can be associated with an IAM user in two ways:

User Policy - only exist in the context of the user
Managed policies - exist independently of users - created in the polices tab on the IAM page or via the CLI

Permission is denied by default

If two policies contradict each other the action is denied

To lock down an account
Add MFA
Implement a password policy

CloudAcademy

REVISION CARDS – DATA SECURITY

If your administrator leaves the company
Change the password and add MFA to the root account
Rotate keys and change the passwords for IAM user accounts
Delete the users personal account
Put an IP restriction on the root account

EC2 instances cannot send spoofed or anonymous network traffic. You cannot run an instance in stealth or promiscuous mode

Port scans are not allowed under the AWS usage policy

AWS Cloudfront enables private content via Signed URLS, Signed cookies and Origin Access Identities

PEN testing is allowed…
but you need to ask for permission to run a test via AWS support

EC2 uses public key cryptography to encrypt and decrypt log in information

Linux instances have no password. You use a key pair to log in using SSH
For Windows you use a key pair to obtain the admin password and then log in using RDP

AWS KMS - a managed service that makes it easy to manage encryption keys

AWS CloudHSM is a dedicated key management appliance

AWS Directory service is a managed service that enables controlled information to information about your organisation

AD Connector is a proxy service which enables you to connect your on-premise Microsoft Active Directory to the AWS Cloud without the need for direct synchronisation and / or the complexity of hosting federation infrastructure

Security groups act as a virtual firewall

When you launch an instance you associate one or more security groups with the instance

You cannot access the underlying OS for Amazon RDS instances

Customer data is not exposed

Host operating systems should be protected using MFA

Encryption provided by - S3, EBS, Glacier, Storage Gateway, RDS, Redshift, Workspaces

REVISION CARDS – IMPLEMENTATION AND DEPLOYMENT

If a subnet has an Internet Gateway and a route to that internet gateway it is a public subnet

If a subnet doesn't have an IGW or a route then it is a private subnet

If a subnet only routes traffic to the Virtual Private Gateway (VPG) then it is a VPN only subnet

Route table are the rules for where traffic is allowed. A route table enables EC2 instances within different subnets to talk to each other

Public IP addresses are owned and assigned by AWS, they can be automatically assigned to instances launched within the VPC

An EIP is also a AWS owned public IP address but one that you can allocate to your account. By default we are allowed 5 IP addresses per region.EIP's are free if used, but we pay for any IP addresses that we request but do not use

Dedicated instances ensure your application will not run on hardware used by any other client

The minimum CIDR you can have in the VPC is /28
The maximum CIDR you can have in the VPC is /16

You need two public subnets and two private subnets for a HA design - 4 subnets

When you create a VPC all subnets can communicate with each other by default

Disabling source / destination checks on a NAT instance enables traffic to flow

An EIP remains associated with an instance when an instance is stopped

If you attach an EIP to an instance that is associated to a different subnet the instance will be dual-homed

With reserved instances you can change an instance type within the same instance family, and you can change the availability zone

Spot instances are a cost effective way to provide temporary compute resource

You cant change the CIDR block of a VPC once it has been created

NACLS are associated to a VPC subnet to control traffic flow

You can only have one IGW for each VPC

In an EC2 -classic network the EIP will be dissociated from the instance on a stop or start event

A stop / start of an EBS-backed EC2 instance always changes the underlying host computer

Reserved instances enable cost savings if you need to run instances full time

On demand instances enable flexibility to handle spikes in traffic

REVISION CARDS – IMPLEMENTATION AND DEPLOYMENT

Simple Workflow Service enables co-ordinated tasks across distributed components

Simple Workflow Service actors can be workers, workflow starters or deciders

Each workflow runs in a domain. You can have multiple domains per account. Domains can't interact with each other

Simple Notification Services is an asynchronous push notification service

Simple Notification Service enables a publisher to send notifications to individual or groups of subscribers

SNS can use HTTP, HTTPS, SMS, email, email-JSON, Amazon SQS or Lambda protocols

Key elements are
Publisher
Subscriber
Topic

Simple Queue Service visibility timeout - the period of time where SQS prevents other apps or services from accessing or receiving a message

Simple Queue Service default message visibility timeout is 30 seconds.
The maximum you can set is 12 hours

Simple Queue Service messages are retained for 14 days

SQS Long polling allows an application to poll the SQS queue with a wait factor which you can set to be between 1 and 20 seconds

CloudAcademy