1. AWS Organization - Regions, Availability Zones, Edge Locations
2. Compute - EC2, Load Balancers, AMI
3. Storage - EBS, S3, EFS, Volumes, Snapshots
4. Networking - VPC, NAT, Bastion Hosts, Route53
5. Identity & Access Management - Users, Roles, Policies & Permissions

AWS Week 1 - Notes

Introduction to AWS

Amazon is the market leader and pioneer in cloud computing.
Also services name will be variable from one cloud provider to another. But the functionality will be similar across the cloud providers.

Note: Console layout may change so keep note of the functionality then finding the buttons will be easy

History:

Amazon - 1994 founded
Merchant.com - 2000 (Real experience of scaling, handling seasonal traffic)
EC2 - 2006 , 2007-08 becomes public

Problems:

Maintaining a server, backups, recovery, etc.

Solution:

EC2 plus added services to handle all the requirements of the cloud server.
EC2 can act as a database too.

Storage:

Storage devices are present as a separate service in cloud. All data cannot be stored on the same EC2 since EC2 can be replaced and respawned.
S3, Block storage. 4 types of storage in AWS.

Security:

Add a lock-in key (Virtual Private Cloud) - communicate only thru VPN.

Scalability:

S3 and database are auto-scaled by default in AWS.
EC2 can be scaled according to user requirement with the help of Elastic Load Balancer and automated scaling - “Autoscaling” with schedules or triggers from monitoring.

Route:

Route traffic with Route53. Uses custom domain names and acts as DNS and helps in redirecting the traffic to your domain to the actual cloud resource that you map.

- **Regions** are geographical locations
- **Availability Zones** – Individual Data Centres physically separated from one another but connected internally within a Region.
  Users pick up a certain availability zone as per the **user proximity** so that the network latency is minimum.
- **Launch**ing an instance is like starting a Virtual Machine in a specific region
- **Amazon Machine Image (AMI)** is a bundle which includes the OS and OPTIONALLY can also contain other apps like Ruby, Python, Java

Core Building Blocks

[CPU - RAM - Disk - OS - Network - App configuration] - Instance or VM

EC2 - Elastic Compute Cloud

Seven step workflow for EC2

1. AMI selection - either custom OVA(Open Virtualization Appliance) or use AWS marketplace or community edition or default AWS provided.
2. Instance Type - CPU intense /memory/ intense etc. T - general purpose
3. Configure - Network(VPC), subnets(AZ), spot instance(pricing), public IP, IAM role for the instance, shutdown behavior and deletion protection, Tenancy(shared/placement of servers), burstable-credits, extra NIC cards and userdata(for script running)

   Note: ‘Stop’ just removes the cpu, disk and storage stay. ‘Terminate’ removes everything. Instances are not charged when they are stopped or terminated.

4. Storage - adding a snapshot, delete protection, encryption
5. Tags - management ease and grouping the EC2 instances and helps in analyzing usage and billing
6. Security group - ports to be available and open. IP restrictions for accessing.

7. Review - Access Key with which secure login to EC2 is possible.

Public IPs can be given up in case the EC2 is not internet facing and traffic will redirect from Load balancer etc.

Types of instances

- Spot Instances are unpredictable. You bid and get them but as the price rises and crosses you limit; they will be reclaimed.
  It is therefore recommended to bid for a higher price, you will still be charged as per the current price.
- On demand instances have a fixed price.
- Reserved instances are significantly cheaper but the number of instances and the period for which they are taken remains unaltered.

AWS EC2 Advanced Features

CPU in EC2 can be defined as physical core and virtual core. 2 threads on a single physical core is taken as 2 virtual cores. Hence the core count and vCPUs count is different. Only opt in if a high level CPU intensive task is required. Else go for the default option where AWS decides the CPU according to the m, t, c, etc.

Capacity Reservation - Requesting AWS to hold some CPU capacity which you would use later which you planned with your capacity planning team for your organization. Costs extra.

EBS optimized instances - Faster transfer between the EC2 and the EBS storage - AWS has to place the physical hardwares nearby or add special optimization and hence costs more.

Elastic interface - GPU grade processing speed given to your EC2. Which is CPU intense and done on request and hence costs extra.

SSH in to an EC2 Instance

Protect the .pem file first. Restrict access.
Use the cmd from local(your PC) to connect to the EC2 that is launched.
ssh -i (location of the .pem file) ubuntu(centos/ec2-user)@(your public IP)

Launch more like this just picks up the launch configuration. Does not copy content from the already existing instance.

Load Balancing, Fault Tolerance & CloudWatch

Load balancer should be used as public facing unless it is configured for internal usage(micro-services)

Target groups health is checked via the health check pings from ELB to the EC2s in the group. If the interval is not calculated properly, when any one EC2 fails - ELB will show 502 bad gateway errors.

Cloud watch alarms can also be set up for monitoring the ELB. Notifications like mails, SNS or triggers to some other services can be created with reference to the parameter configured(like no. of EC2 to be live at all time).

Network Load Balancer functions at the fourth layer of the Open Systems Interconnection (OSI) model.
Gateway Load Balancer - manage your third-party virtual appliances.

Windows EC2 Instance, Instance Pricing

Get window password - converts pem file to windows password using which we can login to windows server via RDP access.

Reserved instances with discount cost - upto 75% compared to on-demand
Spot instances 90% discount
Dedicated instance cost more than the on-demand instances

Identity & Access Management

Policy is a JSON - which defines the access to a particular resource.
Hence policy allows and restricts access for a user/service to whomever the IAM is assigned to.
Policy generator can be used to create policy(JSON)

Set of policies makes up a ‘Role’ and is assigned to user or service(EC2)
A user is created inside the AWS account to whom we can assign the roles.