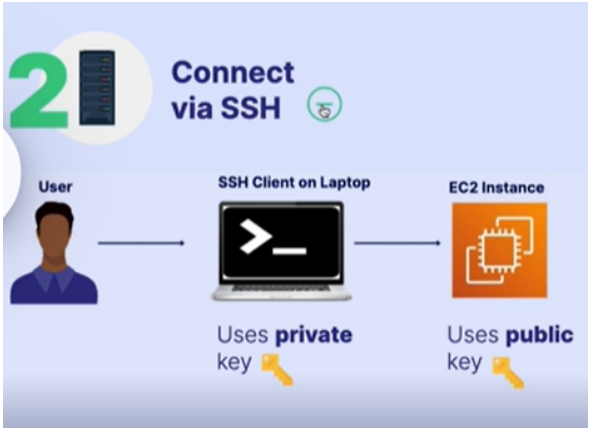
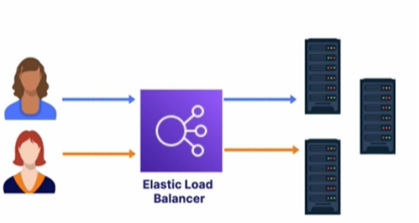
Exploring compute services : EC2

Questions	Notes
elastic compute cloud EC2	• allwos you to rent and manage virtual servers in the cloud • servers are the physical compute hardware running in a data center • ES2 instances are the virtual servers running on these physical servers • instances are not considered serverless • you're able to provision an EC2 instance at the click of a button • you can use preconfigured template called an amazon machine image(AMI) launch your instnace • you can deploy your applications directly to EC2 instnaces • you receive 750 compute hours per month on the free tire plan
methodes to access an EC2 Instance	• aws management console:you're able configure and manage your instance via a web browser • EC2 instance connect (EIC): allows you to use IAM policies to control ssh acess to your instnaces , removing the need to manage ssh keys • secure shell ssh: ssh allows you to establish a secure connection to your instance from your local laptop • aws systems manager: systems manager allows you to manage ec2 instnaces via a web browser or the aws CLI • the most secure way to connect to linux EC2 instances is via Secure shell(ssh): 1. **generate a pair key:**a key pair , which consists of a private key and a public key , proves your identity when connecting to an EC2 instance 2. **connect via ssh:** 
EC2 instances types	• on-demand • spot • reserved instances • dedicated hosts • savings plan

On-demand	• a fixed price in which you are billed down to the second based on the instance type.there is no contract, and you pay only for what you use. • **use On-demand instances when:** 1. *you care about low cost without any upfront payment or long-term commitment* 2. *your applications have unpredictable workloads that can't be interrupted* 3. *your applications are under development* 4. *your workloads will not run longer than a year* • **fun facts:** 1. you can reserve capacity using on-demand capacity ReservThe EC2 capaciy is held for you whether or not you run the instance.
Spot instances	• spot instances let you take advantage of unused EC2 capacity.your request is fulfilled only if capacity is available. • **use spot instances when :** ◦ you are not concerned about the start or stop time of your application ◦ your workloads can be interrupted ◦ your application is only feasible at very low compute prices ◦ the cheapest option • **Fun facts:** ◦ you cansave up to 90% off on-demand prices ◦ you pay the spot price that's in effect at the beginning of each hour.
Reserved instances	• RIs allow you to commit to a specific instance type in a particular region for 1 or 3 years • **Use Reserved Instances when:** ◦ your application has steady state usage , and you can commit to 1 or 3 years ◦ you can pay money Upfront in order to receive a discount on On-demand prices ◦ your application requires a capacity reservation • **Fun facts:** ◦ you can save up to 75% off On-demand prices ◦ you are required to sign a contract. ◦ you can reserve capacity in an AZ for any duration ◦ you can pay all Upfront , Partial upFront , or No Upfront . All upfront for the max term earns the highest discount ◦ provides convertible types at 54% discount

a Dedicated instance runs on the host	

savings plans	• Savings plan allows you to commit to compute usage (measured per hour) for 1 or 3 years
	• **use savings plans when:**
	  ◦ you want to lower your bill across multiple compute services
	  ◦ you want the flexibility to change compute services, instance types, operating system, or regions
	• **Fun facts:**
	  ◦ you can save up to 72% off on-demand instances prices
	  ◦ you are not making any commitement to any dedicates hosts, just compute usage
	  ◦ savings can be shared across various compute services like EC2 , Fargate and lambda
	  ◦ this does not provide a capacity reservation.

Features	• EC2 instances ofer load balancing and auto scaling • Elastic load balancing automatically distributes your incoming application traffic across multiple EC2 instances  • types: classic load balancers, application load balancers , Gateway load balancers, Network load balancers • EC2 auto scaling adds or replaces EC2 instances automatically across AZs, based on need and changing demand • horizontal scaling or scaling out: Auto scaling reduces the impact of system failures and improves the availabilty of your applications

Résumé

•