EC2: (Elastic Compute Cloud)
• EC2 enables you to create virtual computers in the Cloud and you don't need to manage any hardware. It's a cost-e fficient service compared to an on-premises computational network.

• AWS has plenty of predefined EC2 images (i.e. Linux, Windows), and you can also use your own images. These i mages are called AMI (Amazon Machine Image *pretty unspectacular, isn't it*) . You can select how powerful the u nderlying hardware needs to be, by choosing storage type, CPU and if required GPU power.

• AWS offers different options for purchasing EC2:

o On-Demand Instances = you pay a fixed price per hour as long the instance is running.
Pros → High Availability, Pay for Use only, great for reliable short term computing tasks.
Cons → Can be expensive over a long period of time.
Use-Cases → Prototype Development, Extra capacity in case of unexpected computational demand.

o Reserved Instances = You sign a fixed term agreement to use instance for 1–3 years.
Pros → Cheaper than On-Demand Instances, High Availability, great for reliable long term computing tasks.
Cons → Can be expensive if the instance has a lot of unused idle time.
Use-Cases → Webserver, Relational databases, core infrastructure with high Up-Time requirements.

o Spot Instances = eBay-like pricing model, AWS offers unused EC2 resources to the highest payer. Usually, you c an save more than 50% compared to the On-Demand or Reserved Instances.
Pros → Cheapest EC2 option, great for short term computation which can be interrupted.
Cons → AWS can terminate the instance at any time, Depending on the market price you may have to wait for so me time, till your offer matches the market price and the instance can start booting.
Use-Cases → Extra computational capacity for unpredictable workloads that can be interrupted (for instance you can have a fleet of Reserved instances running the webservers, and in case of a sudden spike in demand you can req uest Spot instances to take care of the additional workload).

o Dedicated Hosting = You reserve the entire hardware server/rack for your instance (it's not a virtual machine any more).
Pros → Great for isolated computation (high security).
Cons → Most expensive EC2 service.
Use-Cases → Enterprise databases with high-security requirements.

• Spot-fleets can be used to combine all the services above to a cost-efficient strategy. You can have a bunch of Rese rved instances to take care of the core workload, and a combination of On-Demand and Spot instances to handle the workload spikes. Spot fleets are great to use in Auto Scaling Groups.
• Once you have defined your EC2 type and pricing model, let's have a deeper dive into the EC2 storage options.
EBS = Elastic Block Storage
EBS is a virtual file system drive attached to the instance. Important to remember that you can't attach the same EB S to multiple instances (it is not a network drive), but you can duplicate EBS by taking a snapshot and attaching the snapshot to another instance. EBS comes in multiple flavors:

o EBS(GP2) = General Purpose SSD
Pros → balanced work performance, suitable for most applications
Cons → I/O throughput (max. 16,000 IOPS) → remember this number!
Use-Cases: System boot drive, medium workload applications

o EBS(I01, I02) = Provisioned IOPS SSD
Pros → Excellent for low latency and high throughput requirements (up to 64,000 IOPS)
Cons → Expensive storage type
Use-Cases: Relational databases, High-performance computing (HPC)

o EBS(ST1) = Throughput Optimized HDD
Pros → Low-cost HDD volume designed for frequently accessed, throughput-intensive workloads
Cons → low IOPS
Use-Cases: Storage for streaming applications, Big Data Storage

o EBS(SC1) = Cold HDD
Pros → Cheapest EBS type
Cons → lower durability and availability
Use-Cases: Storage for files that require high throughput and are infrequently accessed.

• EBS volumes are located in a single Availability Zone and if you need to provide High Availability and/or Disaster recovery, you will need to set up a backup process to create regular snapshots and save them in S3. EBS is a provisioned storage type and you pay for the storage capacity that you have defined.
• In case you need to your common network drive you will need to attach an EFS volume to your instances.
• Elastic Block Store (EBS) is a virtual hard disk. Snapshots are a point-in-time copy of that disk.
• Volumes exist on EBS. Snapshots exist on S3.
• Snapshots are incremental, only changes made since the last snapshot are moved to S3.
• Initial Snapshots of an EC2 instance will take longer to create than subsequent Snapshots
• If taking Snapshot of a root volume, the EC2 instance should be stopped before Snapshotting
• You can take Snapshots while the instance is still running.
• You can create AMIs from Volumes, or from Snapshots.
• EBS Volumes A durable, block-level storage device that you can attach to a single EC2 instance
• EBS Volumes can be modified on the fly eg. storage type or volume size.
• EBS Backed instances can be stopped and you will not lose any data.
• By default root volumes are deleted on termination.
• EBS Volumes can have termination protection (don't delete the volume on termination)
• Snapshots or restored encrypted volumes will also be encrypted.
• You cannot share a snapshot if it has been encrypted.
• Unencrypted snapshots can be shared with other AWS accounts or made public.

EFS = Elastic File System
EFS works only with Linux and is used as a common network drive that enables data sharing between the instances. For Windows instances, you should use “Fx for Windows” as a network file system. EFS is an automatically scalable storage service and you pay only for the storage space in use (+data transfer fees)

There are 2 EFS types:
o General Purpose Performance Mode → great for low latency, content management, and storage.
o Max I/O Performance Mode → good choice for big data and media processing that requires high IOPS, but it comes to higher latency as a trade-off.

• Elastic File System (EFS) supports the Network File System version 4 (NFSv4) protocol.
• You pay GB of storage per month
• Volumes can scale to petabyte size storage
• Volumes will shrink and grow to meet current data stored (elastic)
• Can support thousands of concurrent connections over NFS.
• Your data is stored across multiple AZs within a region
• Can mount multiple EC2 instances to a single EFS (as long as they are all in the same VPC)
• Creates Mount Points in all your VPC subnets so you can mount from anywhere within your VPC
• Provides Read After Write Consistency.

There is one more storage type you will need to know.

Instance Store is a temporarily high-performance storage with very high IOPS but the drawback is that is not durable storage solution
As soon instance stops, all data in the instance store get deleted. Instance Store is great for example for cache and session storage.

We've covered the fundamental knowledge about EC2 deployment and in the exam, you will get tested on scenarios that require high-availability, fault-tolerance, and quick up-times. There are a few keywords you need to remember:

• User Data = it defines the specific commands, which are getting executed during the instance boot process (for instance you could check for updates, every time instance has been booted)
• Meta data = it describes instance configuration and the meta data is accessible through SSH terminal by calling this address http://169.254.169.254/latest/meta-data/

Golden AMI = it is a custom instance image, that has all the necessary software and application installed, is used for quick recovery, and enables high availability. For high availability, you need to copy Golden AMI to different AZ or regions.

Placement Groups = EC2 instances can be grouped together in 3 different arrangments.
- o Cluster = All instances are located in the same hardware rack and have the highest network bandwidth. Great for machine learning and high-performance computing (HPS)
- o Spread = Instance are distributed across multiple AZ in order to maintain high availability
- o Partition = The instances are grouped in smaller groups and each group occupies its own hardware rack. It combines the benefits of Cluster placement with High Availability through a spreading approach.

In the exam, you will be tested on different scenarios for the most suitable EC2 configuration. Here are some general tips:
- o Use EBS(io1/io2) for application with high IOPS requirements (>16,000)
- o Use User Data to 'bootstrap' scripts during instance startup
- o Do NOT use Spot instances for critical workloads.
- o Strife for High-Availability by using EC2 instances in multiple Availability Zones and regions. You can have your production environment up in running one region, and the backup instances are either in Stop or Hibernate status in a different region. In case you production AZ becomes unavailable, you can switch to your backup instances and keep the production workload running from different region (Later I'll explain the concept of Elastic Load Balancing and Auto Scaling on AWS)
- o Use IAM roles to give permissions to EC2 instance
- o And last but not least:
  DO NOT STORE ANY CREDENTIALS INSIDE THE EC2 INSTANCE
  There are better ways to provide secure access by using IAM roles and AWS Secret Manager.