**EC2 (Elastic Compute Cloud)**

**Technical**:

-Resizable compute capacity in the cloud

-99.99% availability

-Each region has at least 3 availability zones

-275 typs of instances

-Can hibernate EC2 instance (uses EBS and billed at EBS rates). Persists RAM. Applications that take awhile to bootstrap or persist state into memory benefit from this.

-Customers requiring massive floating point power will benefit from GPU instances

-Customers requiring high graphics capabilities will benefit from GPU Graphic instances

-Scalable computing capacity in the cloud.

-No need to invest in hardware upfront so develop and deploy applications faster.

-Scale up or down reducing need to forecast traffic

-AMI (Amazon Machine Image) preconfigured templates that package up what you need for a server including OS and additional software.

-Instance Types configurations of CPU, memory, storage, and networking capacity

-Secure login via Key Pairs AWS stores the public, you keep the private

-Instance Store Volumes temporary data that's deleted when you stop, hibernate, or terminate an instance

-Elastic Block Store (EBS) volumes for persisting data

-Regions and availability zones for both instances and EBS volumes

-Security Groups firewalls that allows you to control protocols, ports, and IP ranges that can reach your instance

-Static IPV4 addresses for dynamic cloud computing known as Elastic IP addresses

-Manage with AWS Console in web or via CLI

-Can use AWS Cloud Formation, use templates to launch resources

-Supports Query API GET and POST

**Payment:**

-On-Demand

Pay by the hour, no long term commitments or up front costs

-Savings Plan

Reduced cost by committing to a consistent usage amount, paid per hour in 1 or 3 years terms

**Security**:

-Lockdown security model prevents administrative access

-AWS Nitro offloads EC2 hardware functions to dedicated areas reducing attack surface

-Supports 89 security standards including CI-DSS, HIPAA/HITECH, FedRAMP, GDPR, FIPS

140-2, and NIST 800-171

**Billing:**

-On-Demand:

-Pay for compute capacity by the hour or second.

-Applications with short term or spikey or unpredictable workloads that can’t be interupted

-Spot Instances:

-Request spare capacity at up to 90% off

-Applications that have flexible start and stop times

-Apps that are only feasible at very low compute prices

-Users with urgent needs for large amounts of additional capacity

Savings Plans:

-Lower pricing model in exchange for a committment of consistent amount of usage

Reserved Instances:

-Discount on price by paying up front, reserves a specific availability zone instance

Dedicated Hosts:

-A Physical server dedicated for your use. Can help reduce costs by allowing you to use your existing server-bound software licenses, including Windows Server, SQL Server, SUSE Linux Enterprise server