# GETTING STARTED WITH CLOUD COMPUTING

Your instructor:

## Michael J Shannon

Certified Cloud Security Professional (CCSP),
AWS Certified Security – Specialty,
ITIL 4 Managing Professional

**Class will begin at 11:00 A.M. Eastern Standard Time (EST)**

# TODAY'S AGENDA

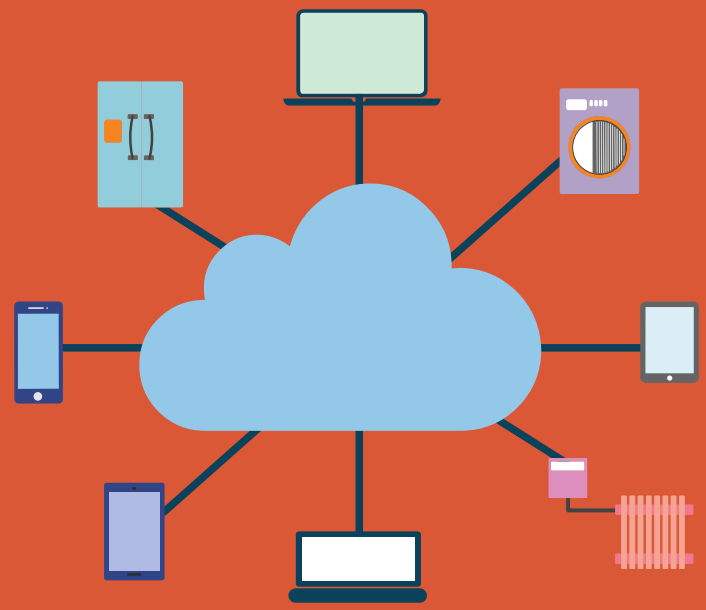
- Cloud Concepts, Architecture, and Design
- Cloud Networking and Content Delivery
- Cloud Compute Services
- Cloud Storage and Database Services
- Cloud Monitoring and Optimization
- Cloud IAM Security Basics

# CLOUD COMPUTING DEFINED


- **Cloud computing** – A network-accessible platform model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction
- Other references:
  - https://csrc.nist.gov/glossary
  - ISO/IEC 17788 “Cloud Computing – Overview and Vocabulary”

# The Value Proposition of Cloud

- The cloud computing value proposition:
  - On-demand self-service
  - Broad network access (CDN)
  - Resource pooling (multi-tenancy)
  - Rapid elasticity (auto-scaling)
  - Measured services (metering)
  - Pay-as-you-go (and grow)
  - Chargeback (showback) billing

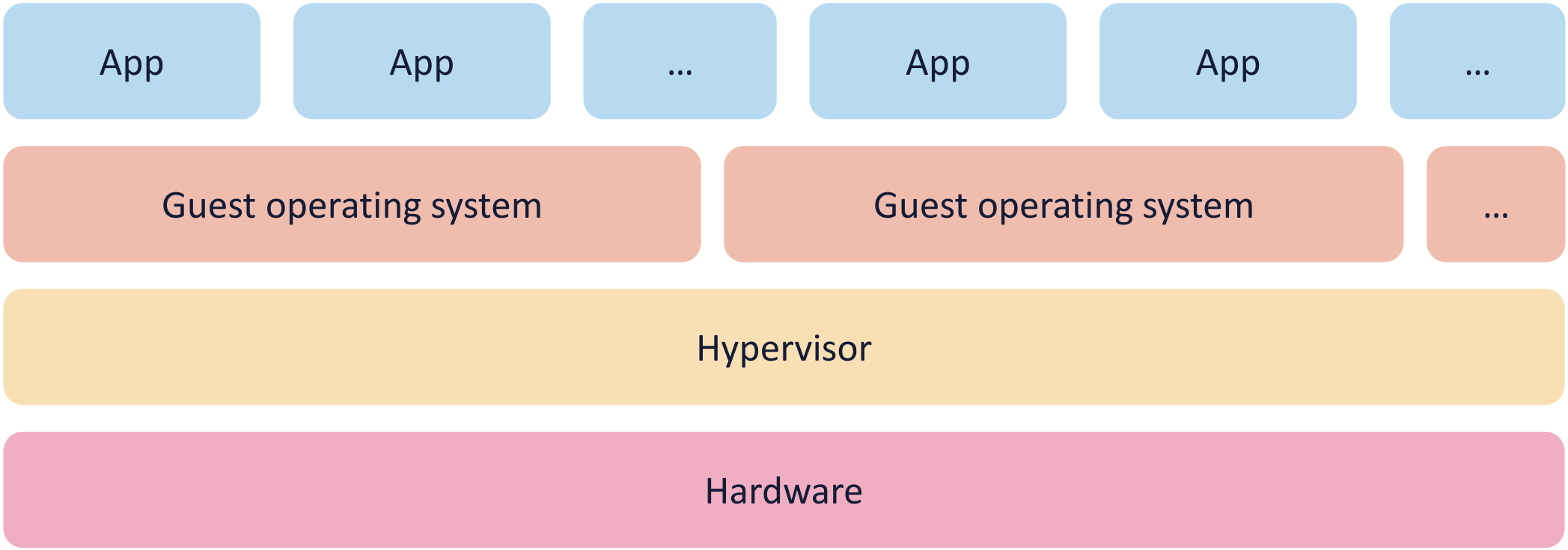


# CLOUD PROVIDER GLOBAL INFRASTRUCTURE

# VIRTUALIZATION DRIVES THE CLOUD COMPUTING DATA CENTER

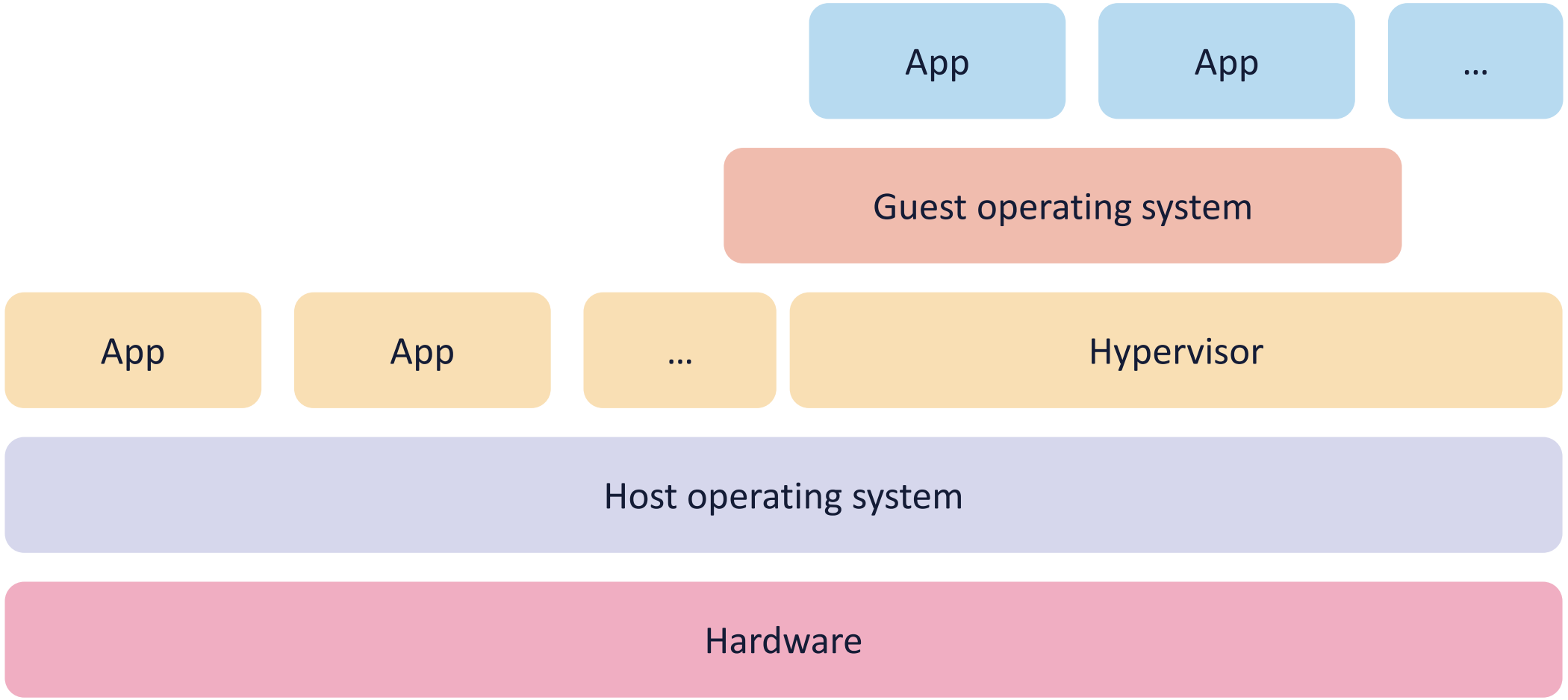

- Virtualization creates a layer of abstraction between the physical hardware and software and file-based simulated machines that run on top of it
- The hypervisor is software that creates, separates, manages, and **orchestrates** the virtual machines
  - A hypervisor is often called a Virtual Machine Manager (VMM)
  - Hypervisors can be proprietary or open-source
  - Hypervisors can be type 1 or type 2

# TYPE 1 HYPERVISORS (BARE METAL OR NATIVE)

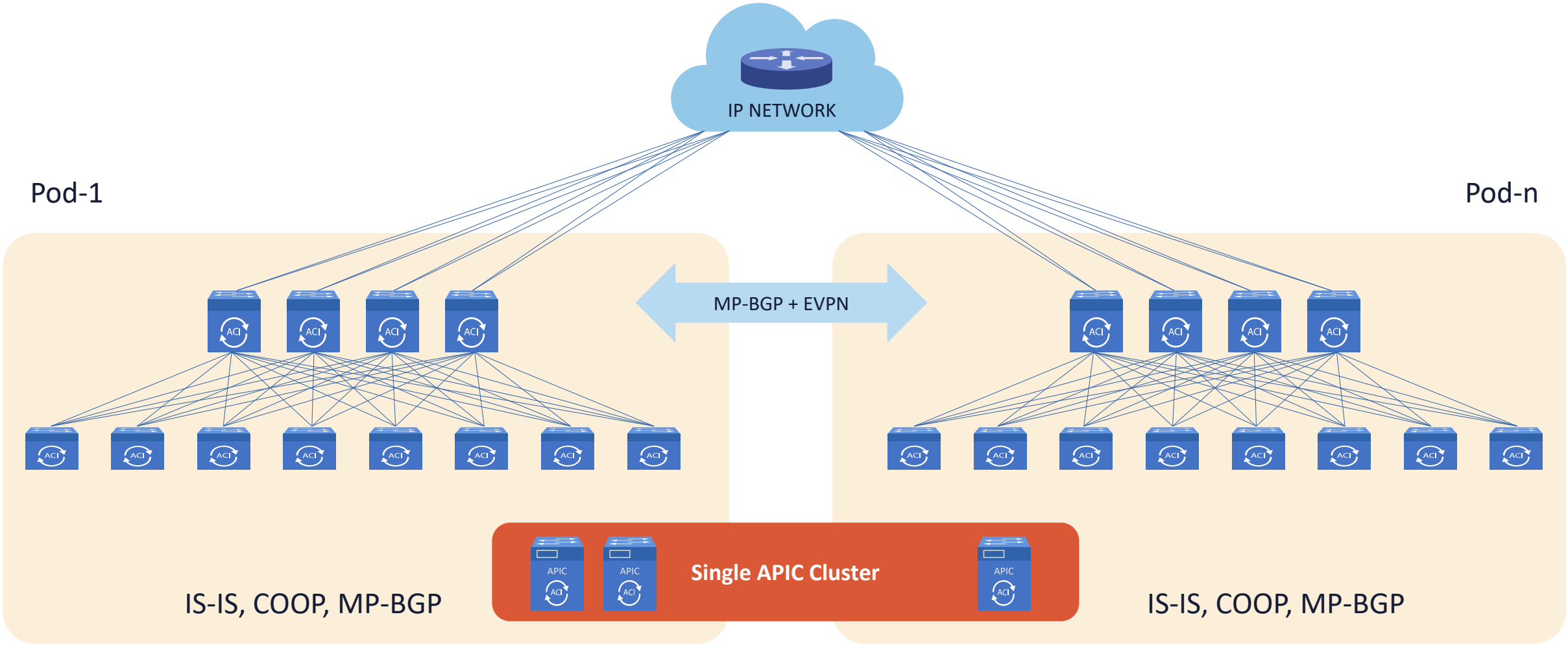



**Azure uses customized Hyper-V, AWS and GCP use KVM, IBM uses IBM z/VM**

# TYPE 2 HYPERVISORS



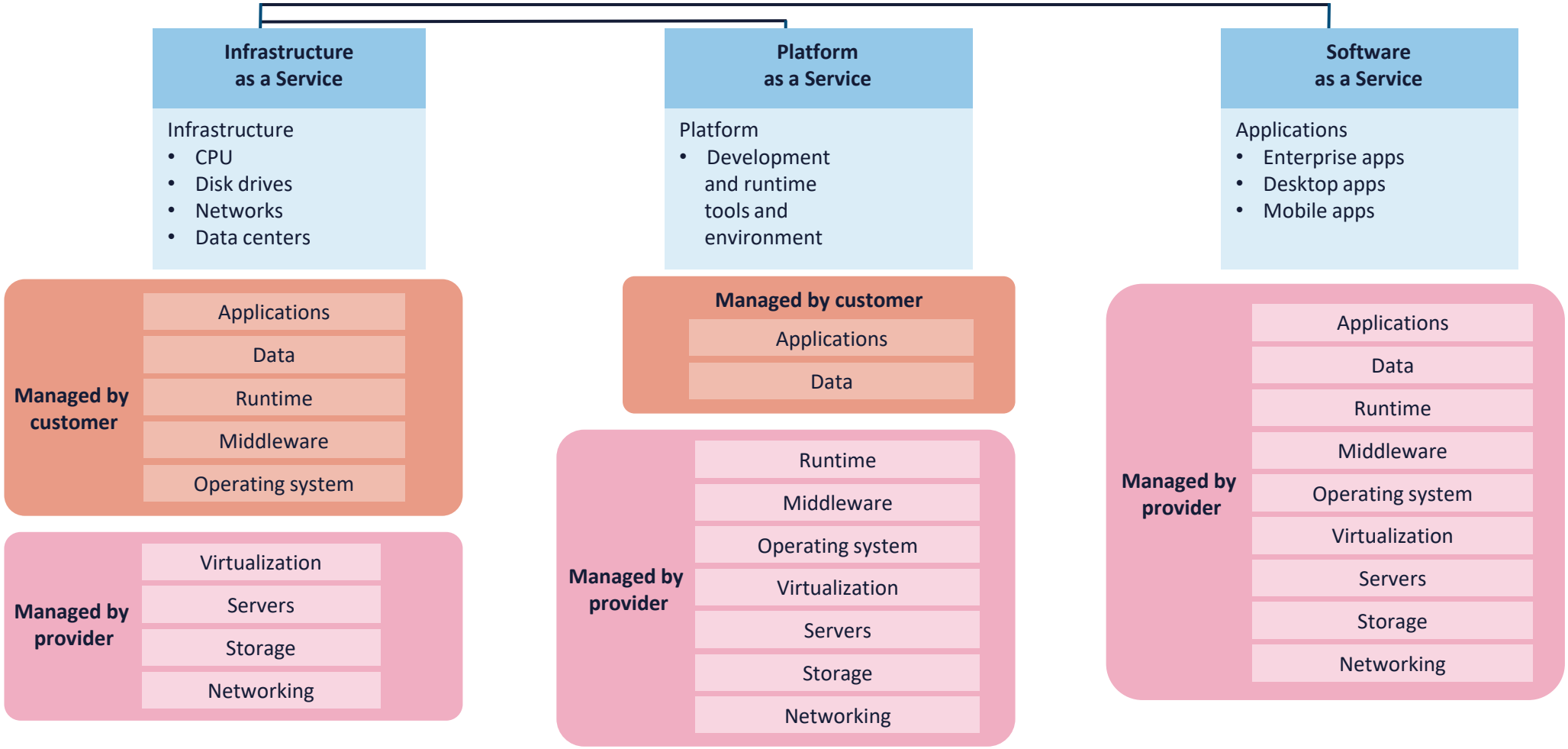
# DATACENTER VXLAN DESIGN



# CLOUD COMPUTING ROLES

- **Cloud Service Customer (user)** – the entity that is paying, leasing, renting, or trying cloud services. Also called the “consumer”.
- **Cloud Service Provider (CSP)** – the vendor that is providing the services from their data center, zones, regions, and edge computing locations. Examples: AWS, Rackspace, IBM Cloud, Microsoft Azure, and Google Cloud Platform
- **Cloud Service Partner** – an entity with various partnership agreements with the CSP such as telecoms, broadband providers, Software-as-a-Service providers, and security solution vendors. Example AWS (GuardDuty) and Rapid7
- **Cloud Service Broker** – an organization that buys hosting services from a CSP and then re-sells to their own consumers. Example: Direct Connect or ExpressRoute partners of AWS and Azure. Also “CASB”
- **Cloud Auditor** – Typically third-party regulators who are ensuring compliance with frameworks such as PCI-DSS

# CLOUD COMPUTING SERVICE TYPES

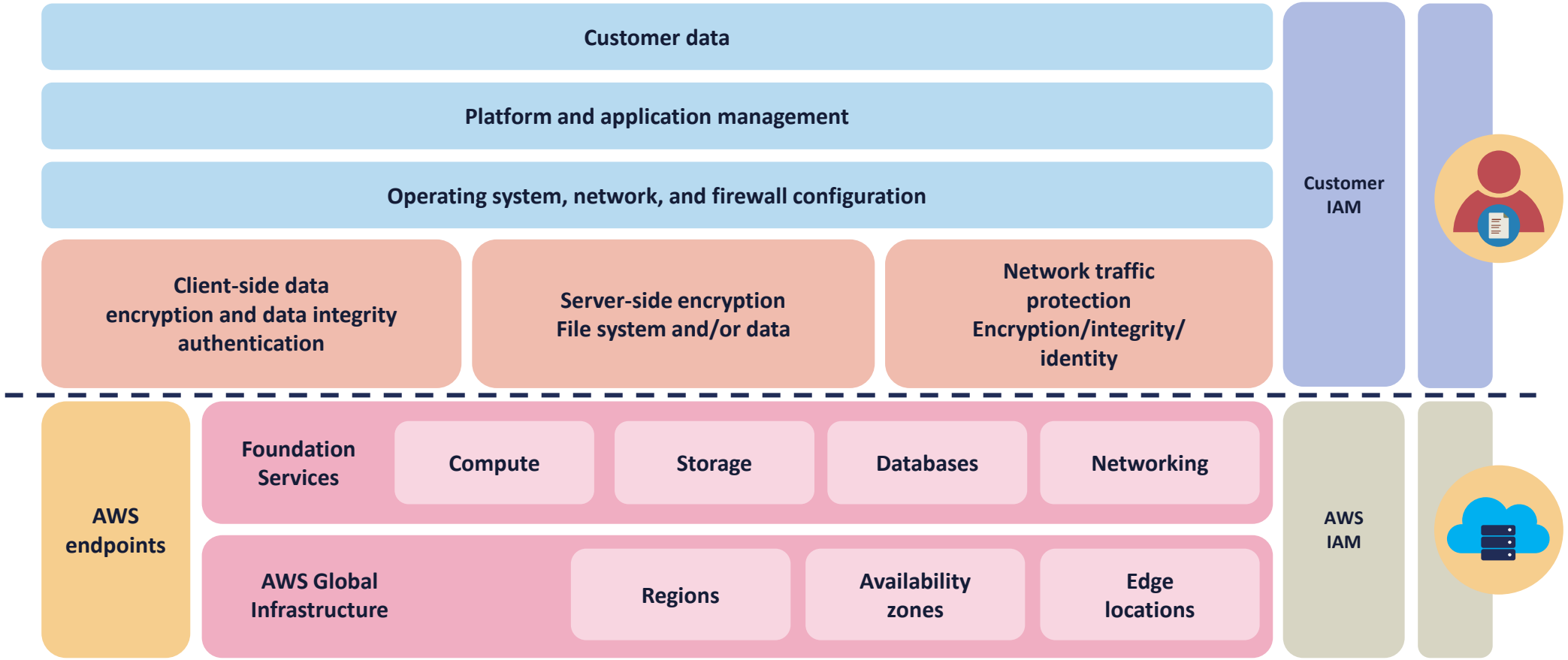


# IaaS ACCORDING TO NIST

“Infrastructure-as-a-service is where the capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer can deploy and run arbitrary software, which can include operating systems and applications.

The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls)."

# AWS INFRASTRUCTURE-AS-A-SERVICE (IaaS)



# PaaS ACCORDING TO NIST

“Platform-as-a-service is the when the provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider.

The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations"

# COMMON PLATFORM-AS-A-SERVICE OFFERINGS

Development and SDK platforms for Java, PHP, Python, etc.

Container services for Docker and Kubernetes

Managed and fully managed relational and document databases

Managed security and threat modeling services

SSO, machine learning, AI, IoT, blockchain, media services

# DEMO: AWS EC2 IaaS VS. PaaS DATABASES

# SaaS ACCORDING TO NIST

"The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser.

The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings."


# COMMON SaaS OFFERINGS

- Customer relationship management (CRM)
- Human resources and workplace tools
- Finance, sales, and marketing services
- Payroll services
- E-mail, collaboration, and cloud storage
- Help & service desk
- Virtual call center
- Business analytics

# DEMO:
# www.zoho.com

# PUBLIC CLOUD DEPLOYMENT

- A model where computing resources are owned and operated by a provider and shared across multiple tenants via the Internet or other public networks
- Enterprises often use public cloud for less-sensitive applications that have unpredictable spikes in usage or for storing data that does not require frequent access
- Public cloud makes computing resources available to anyone for purchase and multiple users usually share the use of a public cloud
- Many businesses use a public cloud to scale existing IT resources on demand without having to commit to growing their physical IT infrastructure

# PRIVATE CLOUD DEPLOYMENT

- A model that is dedicated to a single customer with no other sharing of cloud resources – not a multi-tenant environment
- The private cloud can be a dedicated part of the Cloud Service Provider in a sandbox environment for an additional cost
- The private cloud can be an on-premises solution using virtualization and other cloud service characteristics

# COMMUNITY CLOUD DEPLOYMENT

- A method for connecting infrastructure and applications between similar entities in a certain sector
- Can be for public or private use
- Often used to share information and research among parties with various types of agreements and cooperative relationships
- Common examples are:
  - Government agencies and departments
  - Healthcare provider networks
  - Gaming communities
  - Insurance holding companies
  - Financial services companies

# HYBRID CLOUD DEPLOYMENT

- Technically a combination of private, public, and/or community cloud deployments
- Can also be a method for connecting infrastructure and applications between cloud-based resources and other resources that are not placed in the cloud
- The most common type of hybrid deployment is between the provider’s Public cloud and a standing on-premises enterprise Private cloud
- Can be used to migrate, expand, or grow an organization's infrastructure into a cloud solution while linking internal systems to cloud resources
- Often used by organizations to "burst up" to the cloud during peak demand times or special situations

# CSP SERVICE CAPABILITIES (AWS)

**Archiving**

Reasonably priced solutions for data archiving up to petabyte scale

**Backup and restore**

Durable and cost-effective choices for backup and disaster recovery

**Blockchain**

Shared ledgers for trusted connections between multiple entities

**Business applications**

Simplified management and lower cost business applications

**Cloud migration**

Fluid and simplified migration of applications and data to AWS

**Containers**

Fully managed services for workloads with Docker and Kubernetes

**Content delivery**

Low latency, cached delivery of web sites, APIs, and video content

# CSP SERVICE CAPABILITIES (AWS)


**Database migrations**

Time and cost-effective migration to managed and fully managed databases

**Data Lakes and analytics**

Secure, scalable, and cost-effective data lake and analytics services

**DevOps**

Rapidly and consistently build and deliver solutions using DevOps practices

**E-Commerce**

Highly scalable and secure offerings for online sales and retail businesses

**High performance computing**

Superior networking and cloud sized clusters for multifaceted challenges

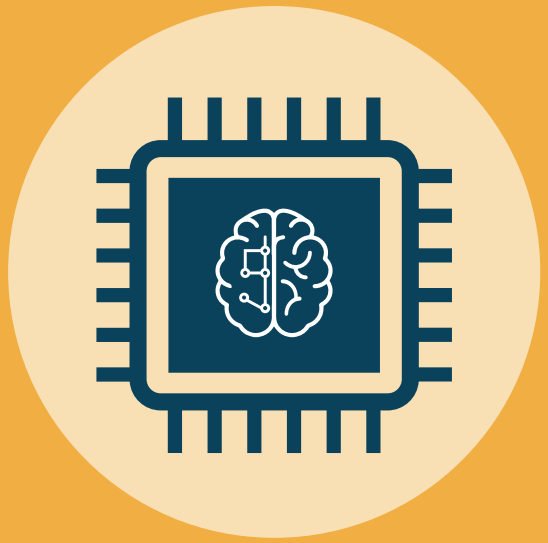
**Hybrid cloud architectures**

Extend the on-premise IT infrastructure to the AWS cloud

**Internet of things**

Scale to billions of devices and messages with cutting-edge solutions

# CSP SERVICE CAPABILITIES (AWS)



**Machine learning**

Leverage wide-ranging machine learning framework support

**Mobile services**

Solutions to help enable mobile application development at scale

**Modern application development**

Produce and advance applications through rapid innovation lifecycles

**Remote work and learning**

Modern solutions for remote teleworkers, students, and center agents

**Scientific computing**

Perform analysis, object storage, and distribution of enormous data sets

**Serverless Computing**

Build and run applications without needing underlying servers

**Web sites**

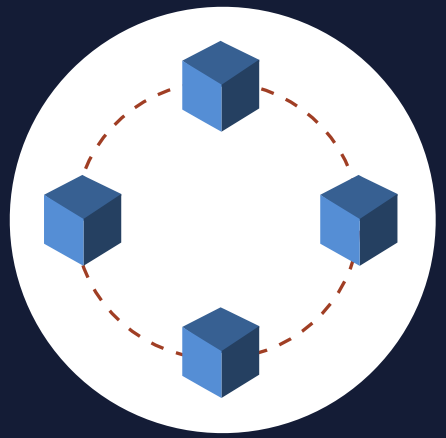
Use dependable, highly scalable, and affordable web application tools

# Refers to programs and machines acquiring, processing, correlating, and interpreting information

- The results of ML and AI are applied in a myriad number of ways without direct input from programmers or users
- There are a wide variety of cloud services that use machine learning algorithms and services:
  - Security automation
  - Language services
  - Intelligent contact centers and personalization
  - Intelligent search and document processing
  - Fraud detection
  - Media intelligence
  - Business forecasting

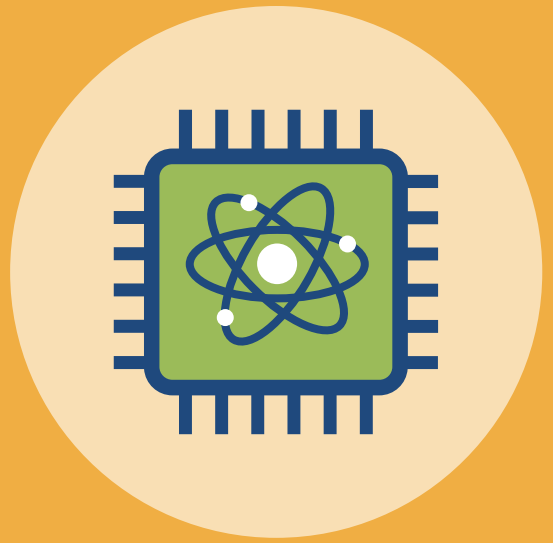
## MACHINE LEARNING

# BLOCKCHAIN



- A public ledger consisting of a digital "chain of blocks" storing information
- Data can be read or write but not modified – changes must be made to a subsequent block in the chain
- Transaction data such as date, time, and amount is verified with a consensus mechanism (PoW, PoS, etc.)
- The transaction participants identities are based on digital signatures
- Unique cryptographic hashes are used to distinguish the blocks from each other
- These things must occur for a block to be added
  - A transaction must take place
  - The transaction must be verified (consensus)
  - That transaction must be stored in a block and given a hash

# QUANTUM COMPUTING



- Personal computers use bits ( 1s or 0s) whereas quantum computers use qubits
- These are typically subatomic particles like electrons or photons
- Quantum computing derives its power from the fact that qubits can represent numerous possible combinations of 1 and 0 at the same time
- This ability to simultaneously be in multiple states is called superposition

# CONTAINERS

- A method for packaging and securely running an application within an application virtualization environment
- A container is a discrete modular and portable environment that includes the application binaries, software dependencies, and hardware requirements wrapped up into an independent, self-contained unit
- You can also use containers for processes and workflows in which there are important requirements for security, reliability, and scalability
- All cloud providers offer managed container development, automation and orchestration services
- Containers can be server-based or serverless (AWS Fargate)

# CLOUD SERVICE CONSIDERATIONS

*The CSP as the producer and we as the consumer co-create value to deliver data, applications, services, and solutions to the world*

- Agility (flexibility) and elasticity
- Reversibility
- Cost
- Security
- Interoperability and portability
- Availability vs. durability
- Resiliency

- Security and privacy
- Performance
- Governance and regulatory
- Auditability
- Maintenance and versioning
- Service levels and Service Level Agreements (SLA)

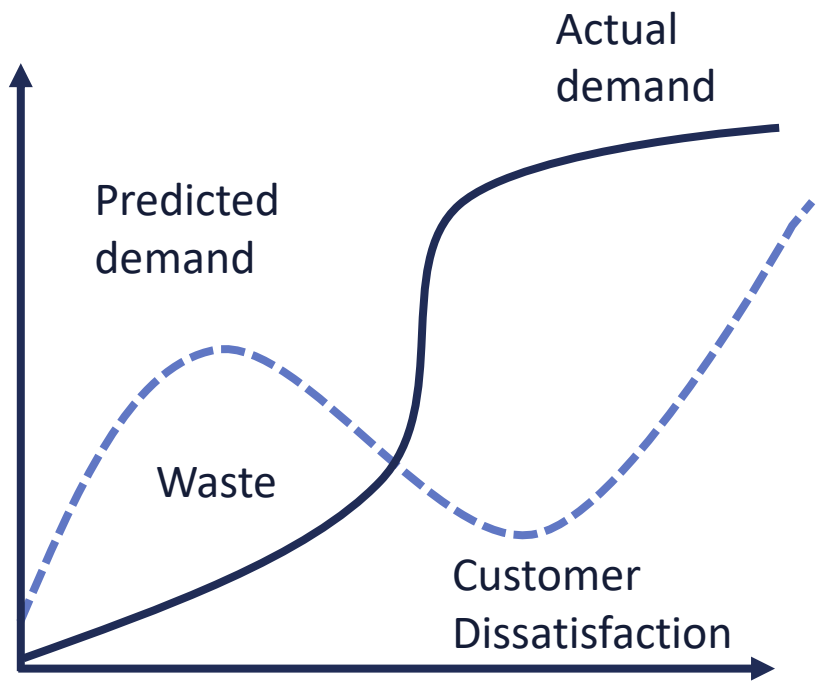
# AGILITY

- Leveraging for rapid deployment, testing, experimentation, & innovation
- Overcoming geographical limitations
- Getting content as close to the consumer as possible
- Reducing time and cost for testing and experimentation
- Allows consumers better innovation

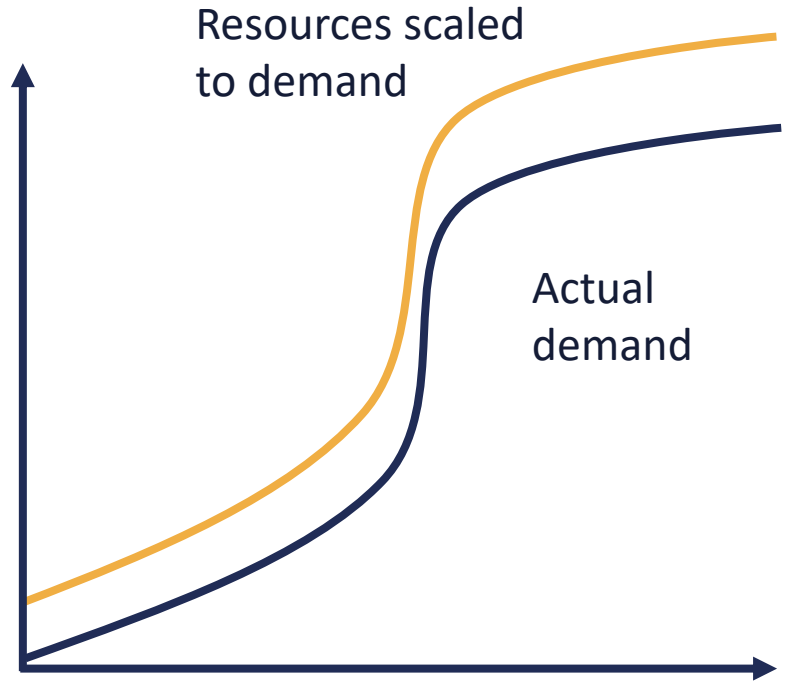
# ELASTICITY

- Elasticity is the ability to almost instantly provision and de-provision resources
- Challenges with predicting demand leads to higher costs
- Leveraging dynamic auto-scaling technologies
- “Elasticity of the cloud allows us to add thousands of virtual servers and petabytes of storage within minutes, making such an expansion possible. Leveraging multiple AWS cloud regions, spread all over the world, enables us to dynamically shift around and expand our global infrastructure capacity, creating a better and more enjoyable streaming experience for Netflix members wherever they are."
  - - Yury Izrailevsky, VP Cloud and Platform Engineering, Netflix (from Netflix Media Center)

# ELASTICITY




Rigid On-premises Resources




Elastic Cloud-based Resources

# RESILIENCY

## Resiliency should be considered a de facto aspect of cloud computing

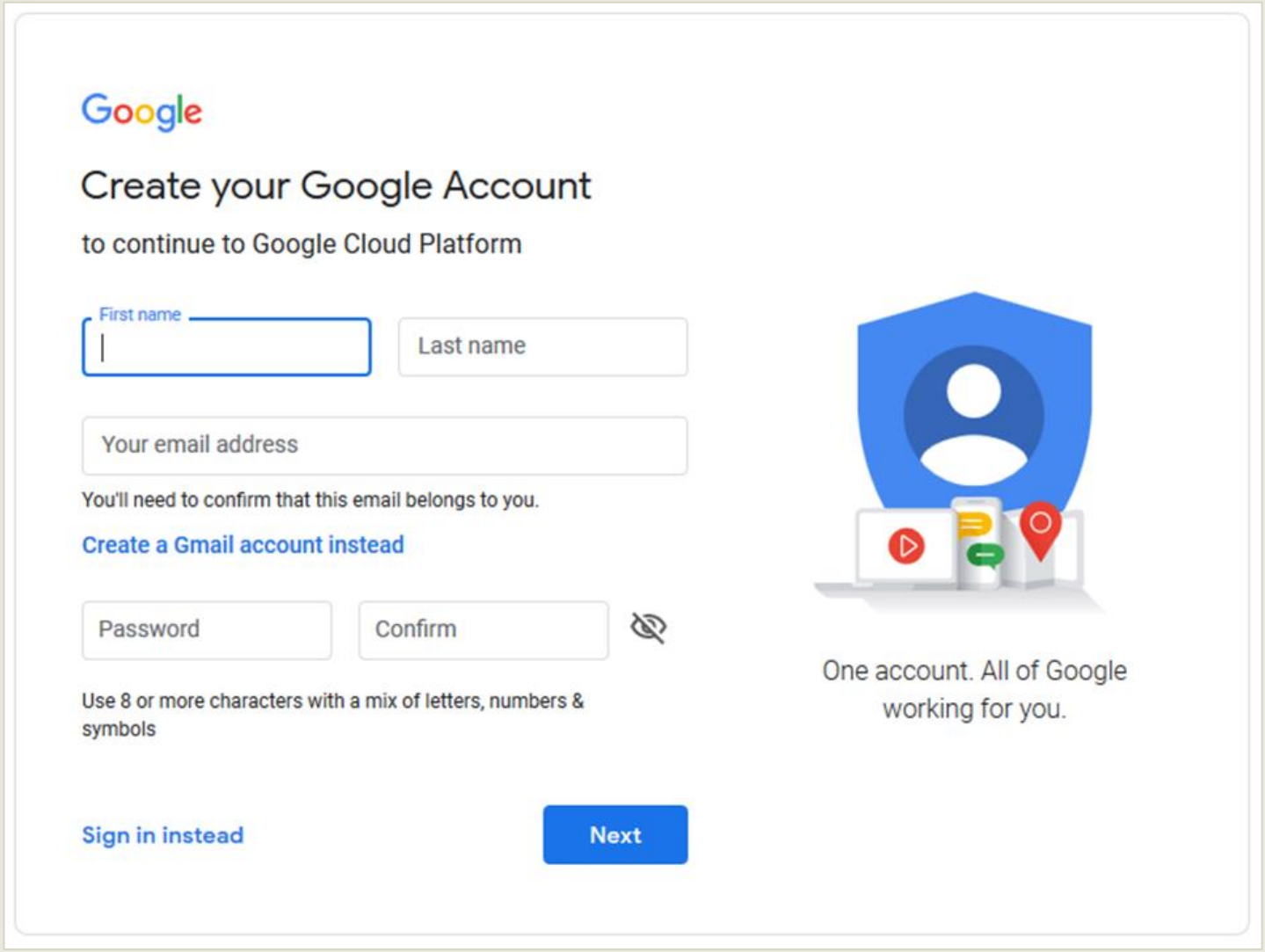
- Reliability is a measure of percentage uptime, considering the downtime due only to faults, whereas Availability is a measure of the percentage uptime, considering the downtime **due to faults and other causes such as planned maintenance**
  - For two different systems, it is possible for one system to be more reliable but less available than the other
- Reliability of a workload in the cloud hinges on a few factors, the primary of which is **resiliency**
- Resiliency is the ability of a workload to recover from infrastructure or service disruptions
- The customer should be able to dynamically obtain computing resources to meet demand and mitigate disruptions
  - Disruptions can be misconfigurations or transient network issues

# AVAILABILITY VS. DURABILITY

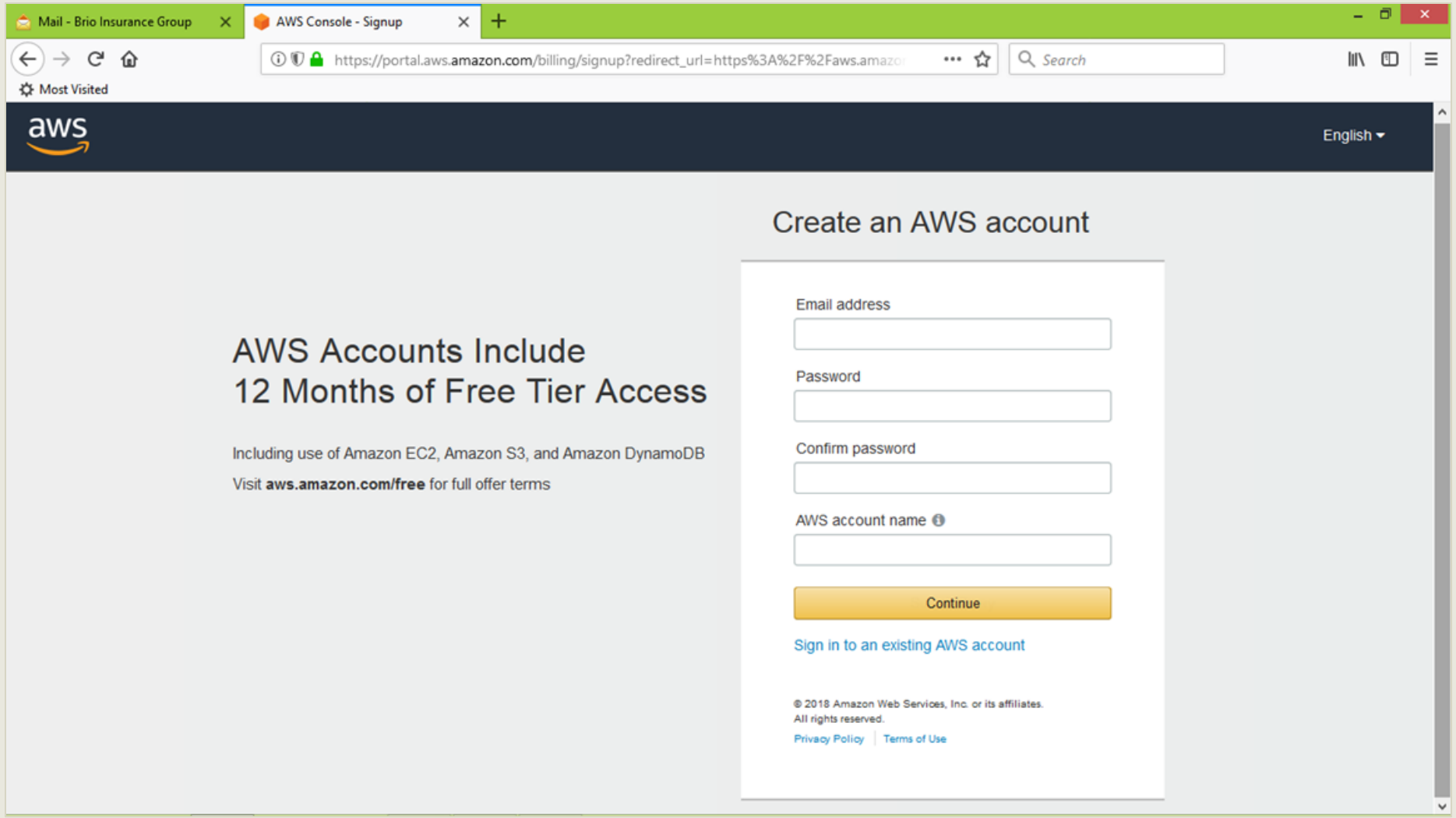
## Availability and durability are two very different aspects of data accessibility

- **Availability** refers to system uptime (i.e., the storage system is operational and can deliver data upon request)
  - Historically, this has been achieved through hardware redundancy so that if any component fails, access to data will remain
- **Durability**, on the other hand, refers to long-term data protection (i.e., the stored data does not suffer from bit rot, degradation or other corruption)
  - Rather than focusing on hardware redundancy, it is concerned with data redundancy so that data is never lost or compromised
- Example: CSP standard object storage provides high levels of data durability and availability by automatically and synchronously storing data across both multiple devices and facilities within a geographical region
  - AWS S3 and Google Cloud is designed for 99.999999999 percent (11 nines) durability per object and 99.99 percent availability over a one-year period
  - Azure claims 12 nines and even 16 nines durability for some services

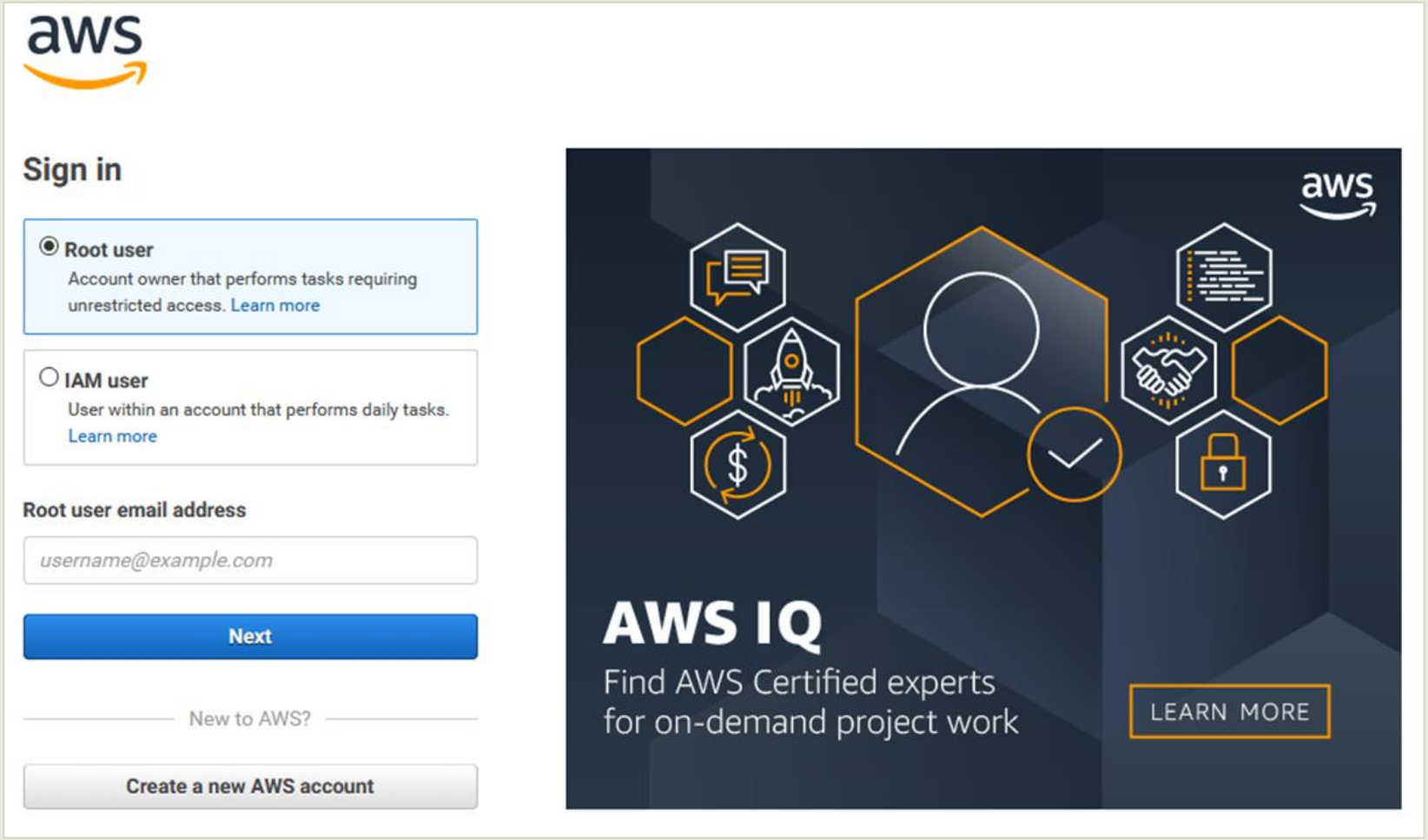
# OPENING YOUR FIRST ACCOUNT



# OPENING YOUR FIRST ACCOUNT



# OPENING YOUR FIRST ACCOUNT



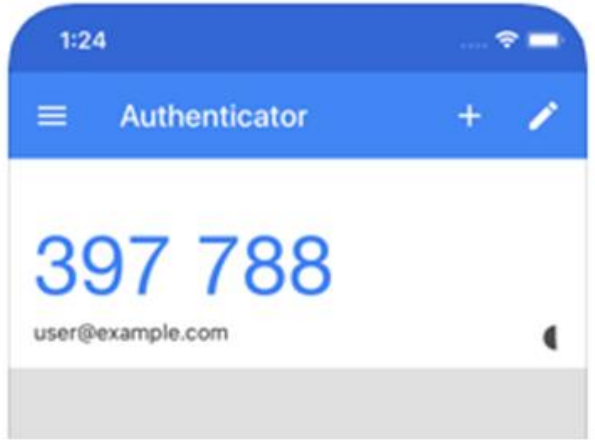
# ROOT AND BILLING ACCOUNTS

**These are standalone single-sign-on root credentials for the cloud accounts**

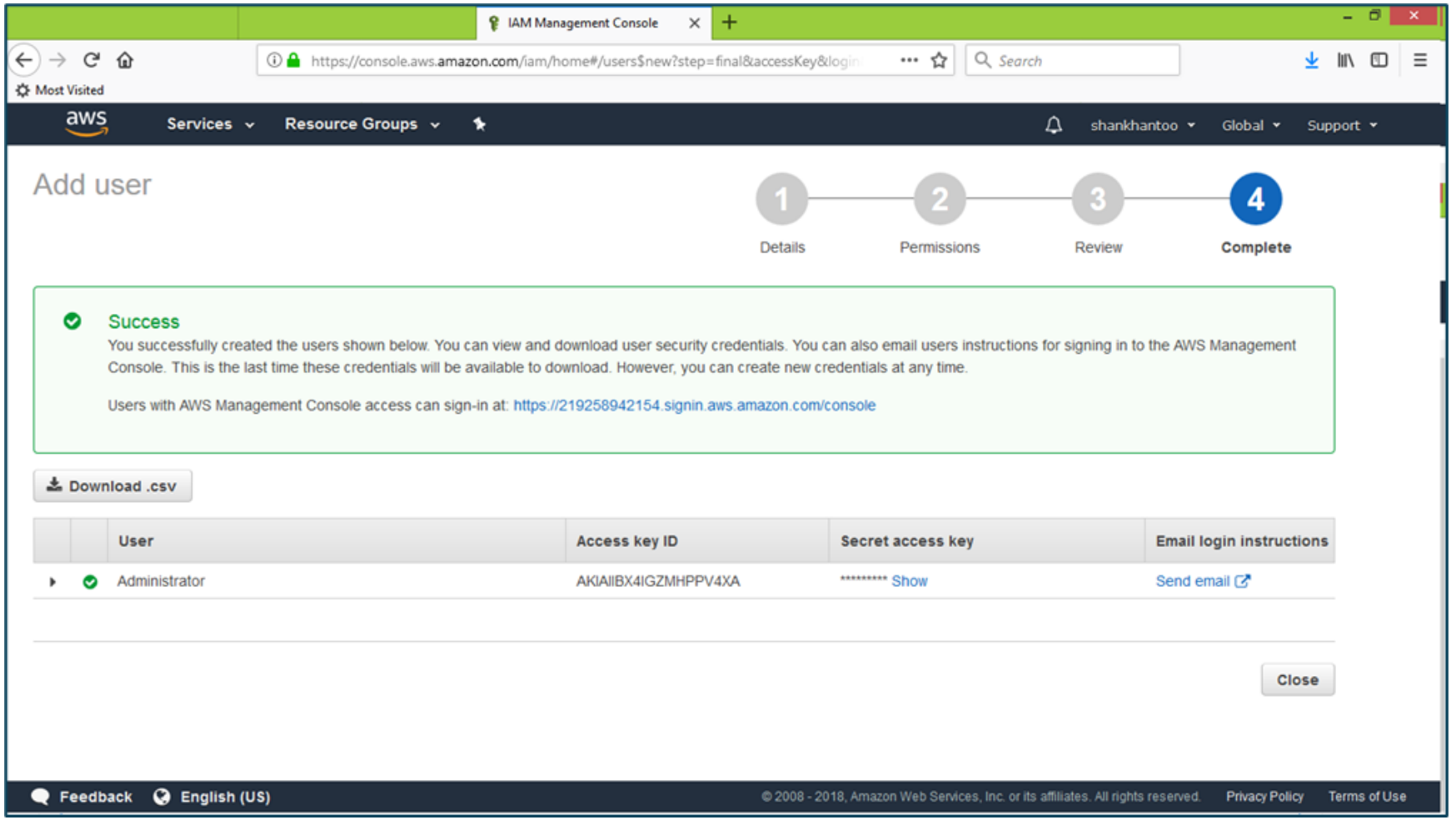
**There are certain tasks that can only be done by Root:**

- Change root user details (password)
- Change Support Plan
- Payment options and billing
- Close an AWS account
- Sign up for special services like GovCloud
- Create an X.509v3 signing certificate
- Transfer Route 53 domain to another account

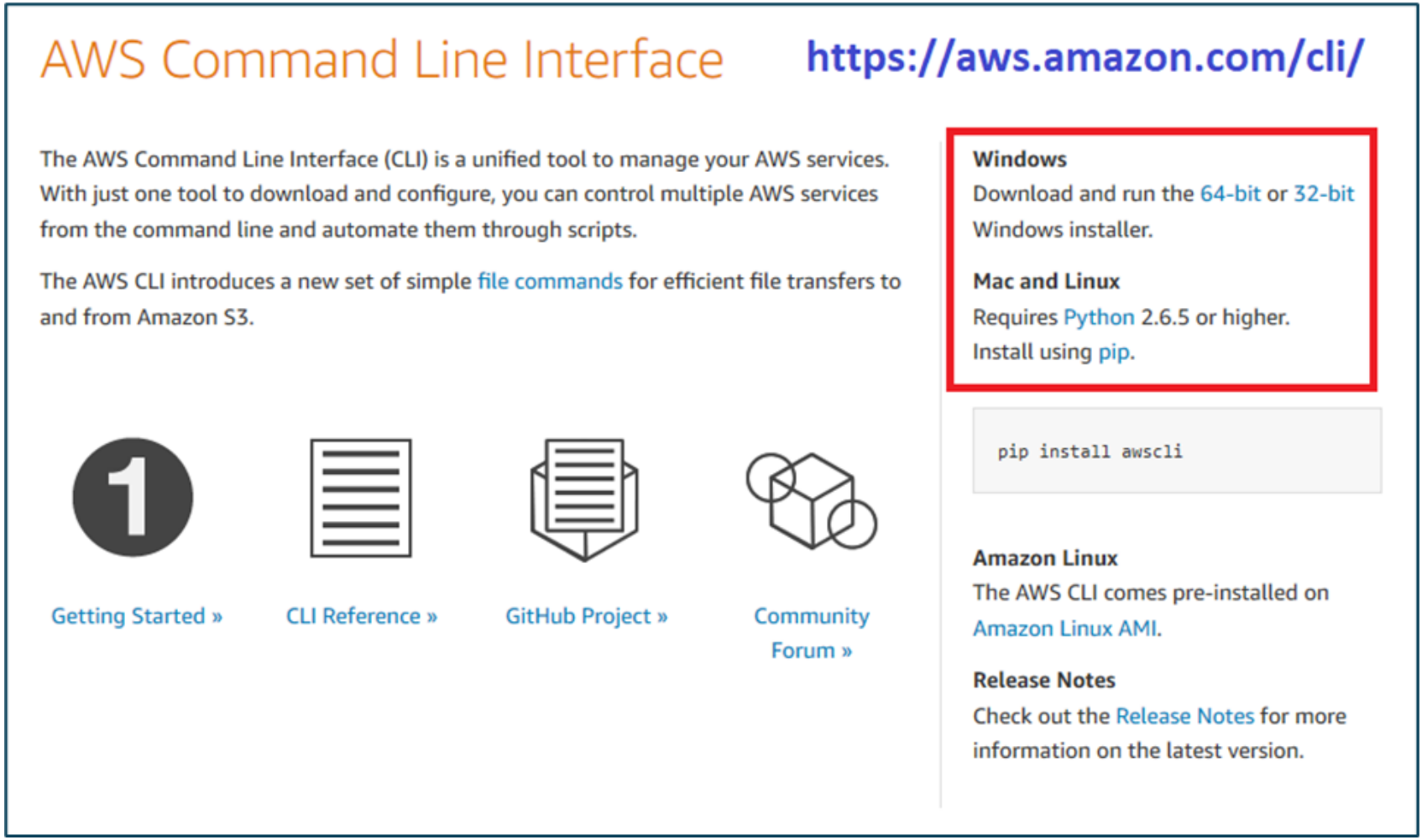
# MULTI-FACTOR AUTHENTICATION IS CRITICAL

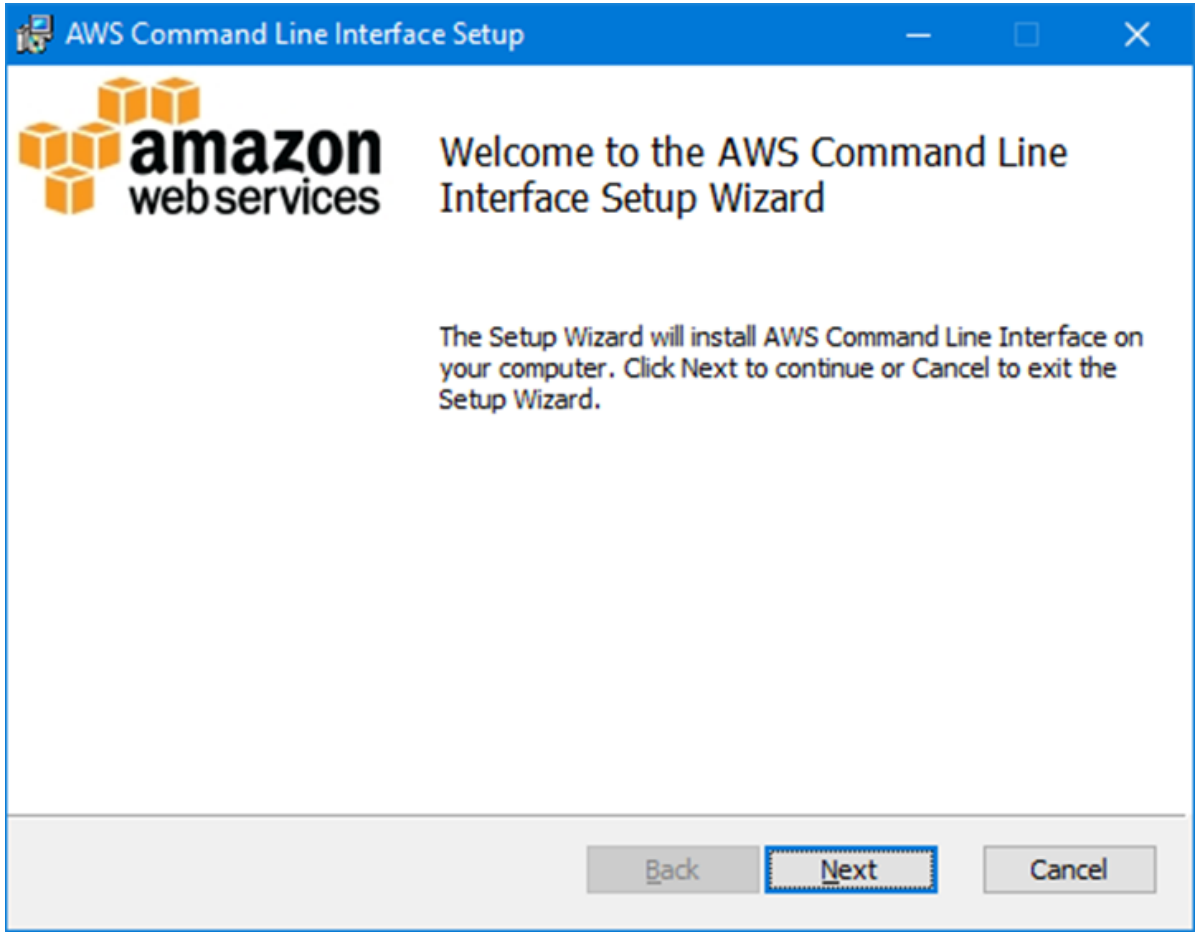
# CONFIGURING CLI ACCESS



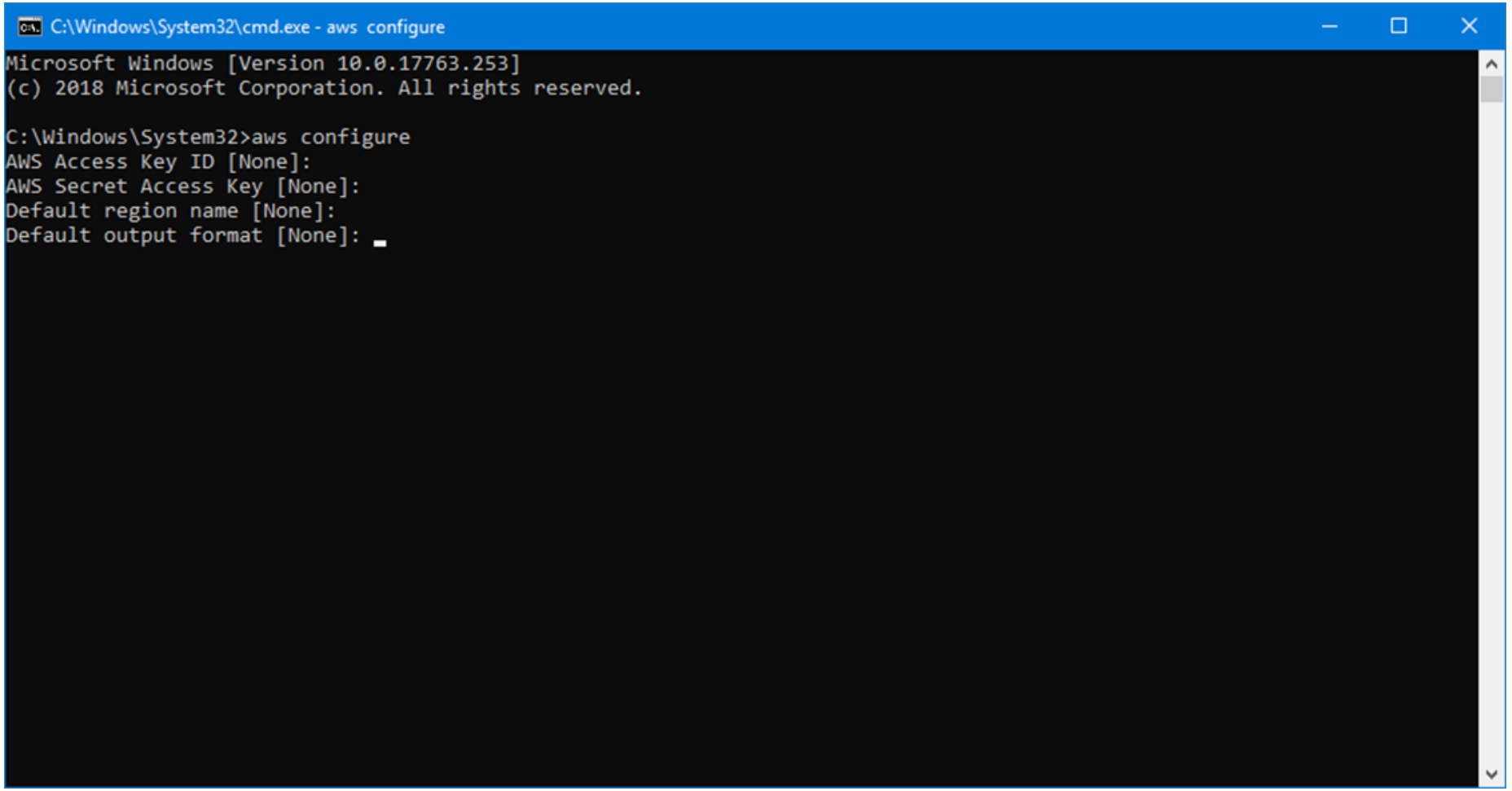
# CONFIGURING CLI ACCESS



# CONFIGURING CLI ACCESS



# CONFIGURING CLI ACCESS



# GOOGLE CLOUD SDK

## gcloud command-line tool

The gcloud CLI manages authentication, local configuration, developer workflow, and interactions with the Cloud Platform APIs.

## gsutil Tool

gsutil provides command line access to manage Cloud Storage buckets and objects.

## PowerShell cmdlets (Windows)

Google Cloud Tools for PowerShell is a collection of Windows PowerShell cmdlets for managing Google Cloud Platform resources within the Windows PowerShell environment.

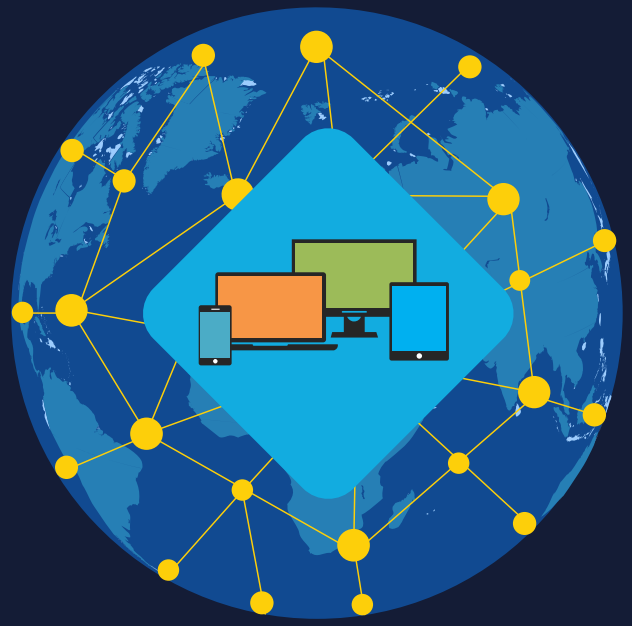
## bq Tool

bq allows you to run queries, manipulate datasets, tables, and entities in BigQuery through the command line.

## kubectl Tool

kubectl orchestrates the deployment and management of Kubernetes container clusters on the gcloud CLI.

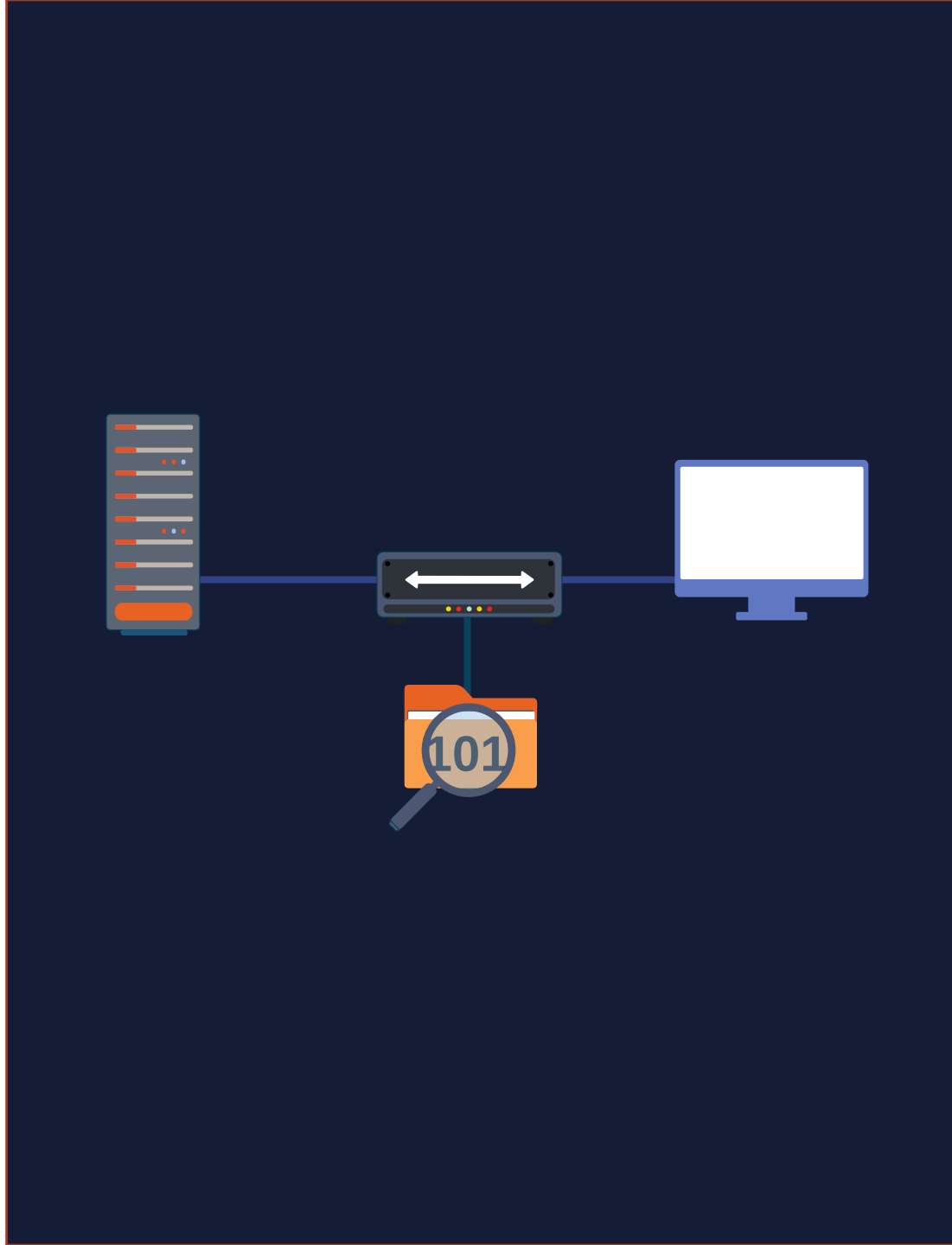
# CLOUD NETWORKING



- Cloud Service Providers offer network services called Virtual Networks (VNets) or Virtual Private Clouds (VPC)
- Consumers can use these to emulate LAN environments or as subnets to deploy other virtual resources
- Networking also involves several ways to connect to cloud resources from on-premises sites
- Content Delivery (Distribution) Networking is another edge computing aspect of cloud networking

# RFC 1918 ADDRESSING

- The Internet Assigned Numbers Authority (IANA) reserved these three blocks of the IP space for private internets address space:
  - 10.0.0.0 - 10.255.255.255  (10/8 prefix)
  - 172.16.0.0 - 172.31.255.255  (172.16/12 prefix)
  - 192.168.0.0 - 192.168.255.255 (192.168/16 prefix)
- Cloud providers use these addresses in their subnets and use special gateways and virtual load balancers to represent hosts using publicly routable addresses



# CLASSLESS INTER-DOMAIN ROUTING (CIDR)

- A compact method for representing IP addresses
- Sometimes called “Slash Notation”
- Eliminates the predefined partitioning of network and host numbers in the IP address
- The number of network (N) bits can be arbitrarily placed without regard to the legacy classes (A – E)
- For instance: the traditional Class C address 192.168.3.15 can be expressed as 192.168.3.15/24, 192.168.3.15/20, or 192.168.3.15/23

# IP VERSION 4 VS. VERSION 6

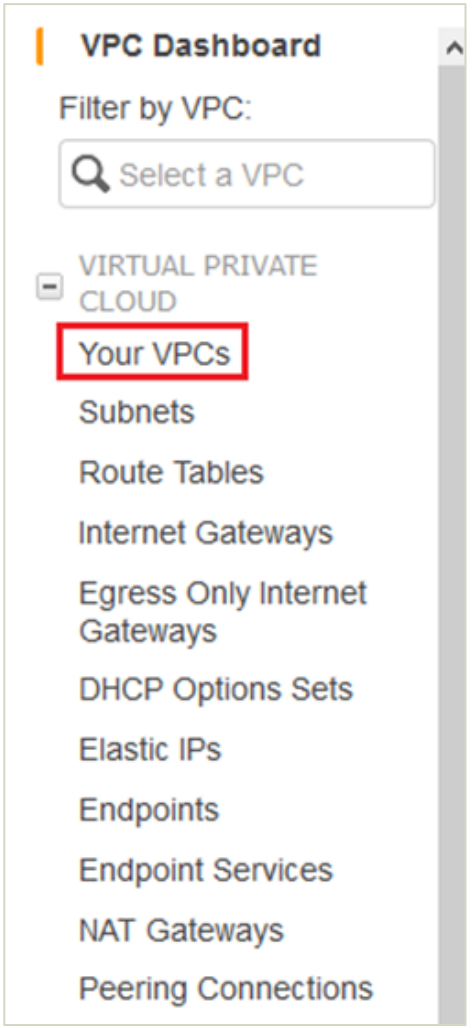
| Version 4 | Version 6 |
| --- | --- |
| $2^{32}$ address space | $2^{128}$ address space |
| Dotted decimal addressing | Hexadecimal notation |
| DHCP dynamic addressing | SLAAC and DHCPv6 |
| Header has 20 bytes and 13 fields | Header has 40 bytes and 8 fields |
| Variable header length | Fixed header length |
| Header option (obsolete) | Header extensions |
| Header checksum | No header checksum |

# IP VERSION 4 VS. VERSION 6

| Version 4 | Version 6 |
| --- | --- |
| Packet size: 576 bytes required fragmentation optional | Packet size: 1280 bytes required without fragmentation |
| Packet fragmentation: Routers and sending hosts | Packet fragmentation: Sending hosts only |
| IPv4 was never designed to be secure | Has native encryption and authentication |
| IPsec optional | IPsec mandatory |
| Non-equal geographical distribution (>50% USA) | No geographic limitations |

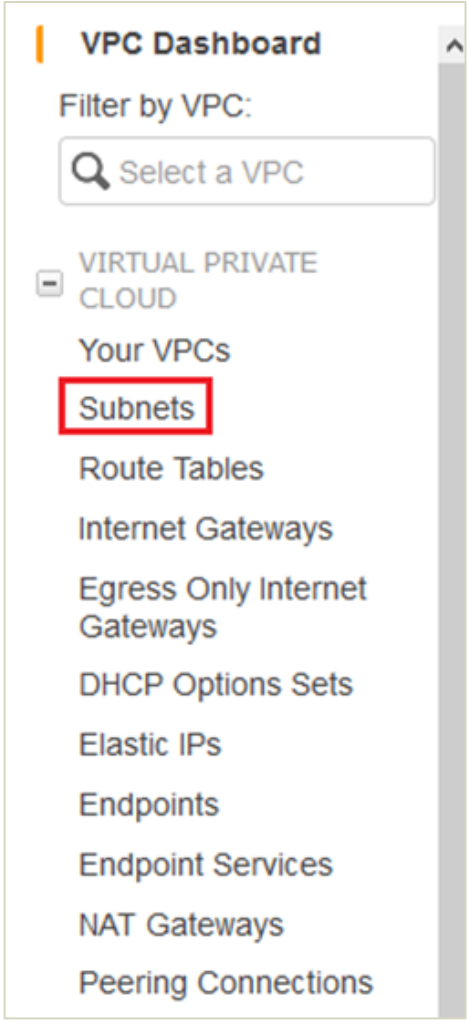
# VIRTUAL NETWORKS (VNets OR VPCs)

- A virtual private cloud (VPC) is a virtual network associated with an AWS account
- Amazon VPCs allow you to instantiate AWS resources into a virtual network that you have defined
- This virtual network is much like a classic network environment in your own data center, with the benefits of using the scalable cloud infrastructure
- Most providers offer a default virtual network and optional user-defined networks



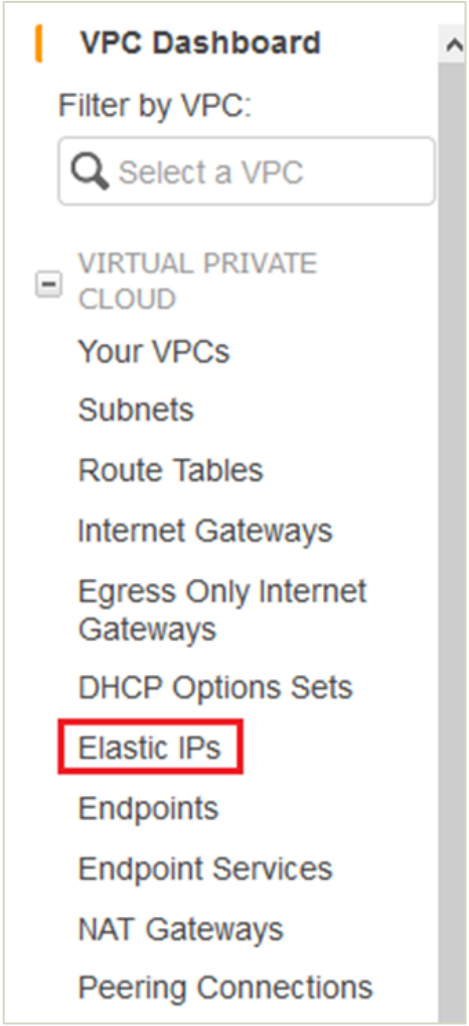
# SUBNETS

- A subnet is a range of IP addresses in your virtual network
  - At GCP and Azure the type of subnet is based on the configuration
  - At AWS a subnet is distinguished based on the entries in the Route Table assigned to it
- There are usually three types of subnets:
  - Private – there is no entry in the route table or internet gateway out of the subnet itself
  - Public – there is an entry in the table or a connection to an internet gateway
  - VPN-only – there is an entry or connection to a VPN gateway



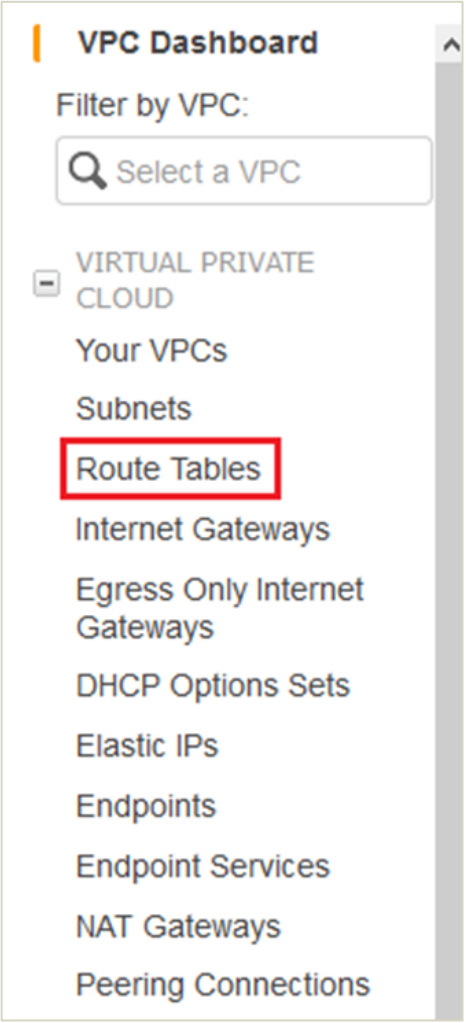
# ELASTIC IP ADDRESSES

- An Elastic IP addresses (EIP) address is a static, public IPv4 address used for dynamic cloud computing
- You can associate an EIP with any instance or network interface for any VPC in your account then mask an instance failure by quickly remapping the address to another instance in the VPC
- Associate the EIP with the network interface instead of directly with the instance so that you can move all the attributes of the vNIC from one instance to another in a single step

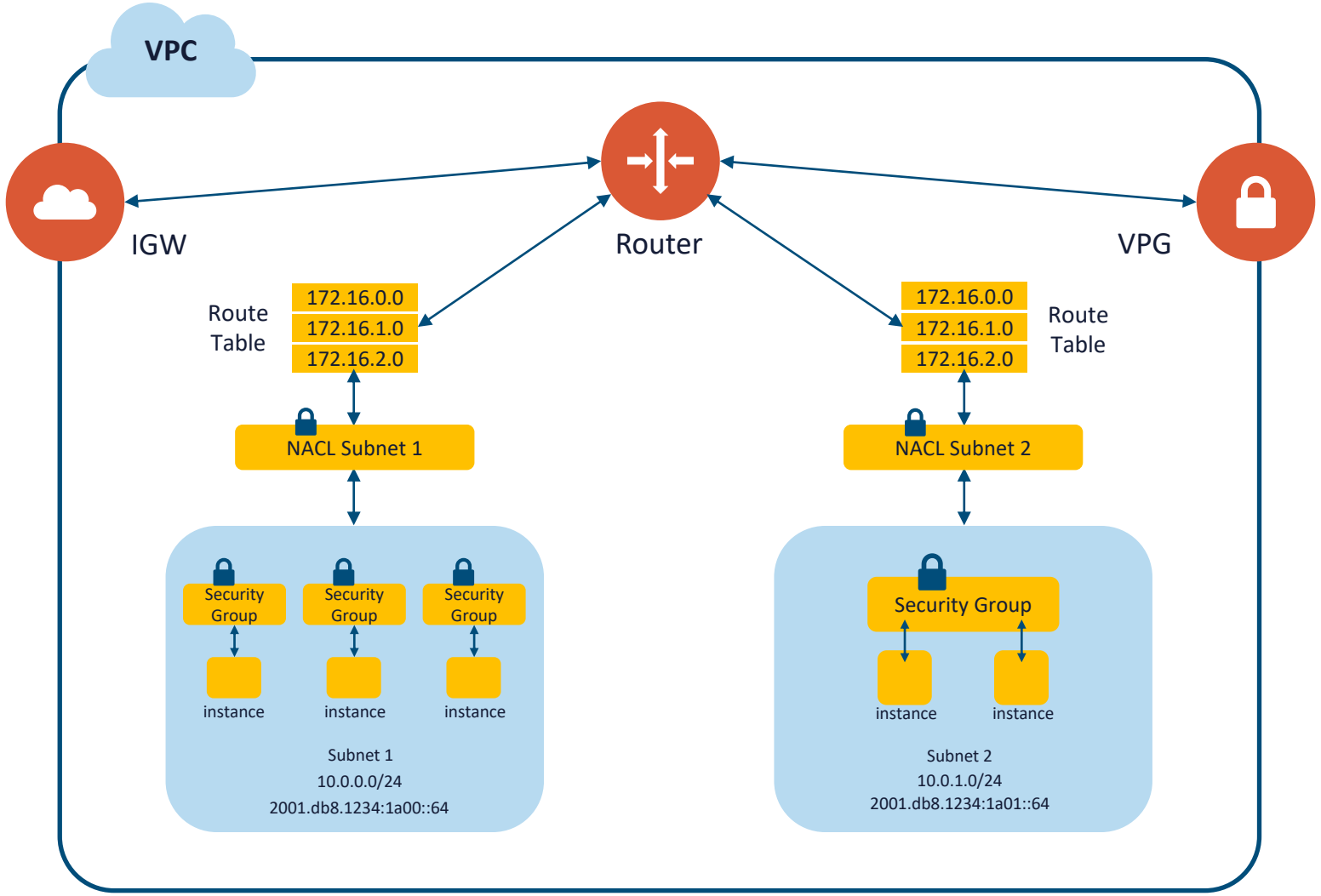


# ROUTE TABLES

- Route tables are virtual cloud components that contain a set of rules, called routes, that are used to determine where network traffic is directed within the virtual network and out to external components
- At AWS, the entries in your route tables determine the type of subnet you have deployed
- The default limit is 50 routes with 1000 route hard limit (new in 2019)

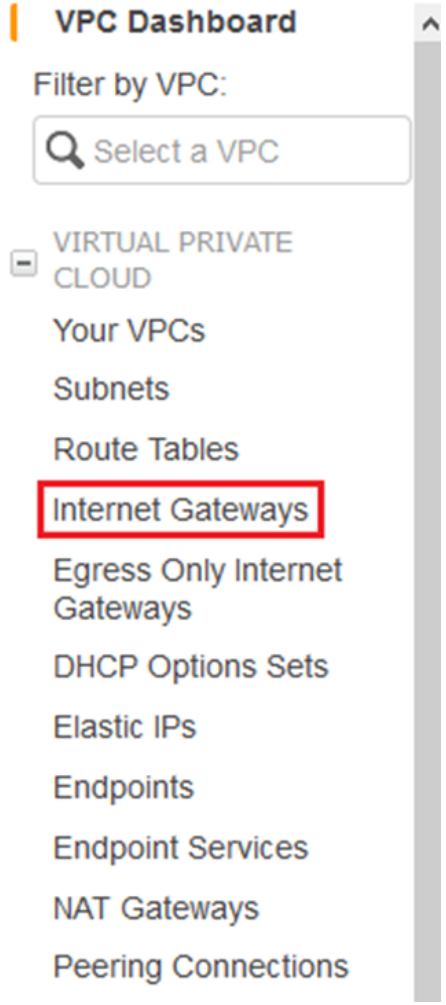


# NETWORKING BEGINS WITH DESIGN



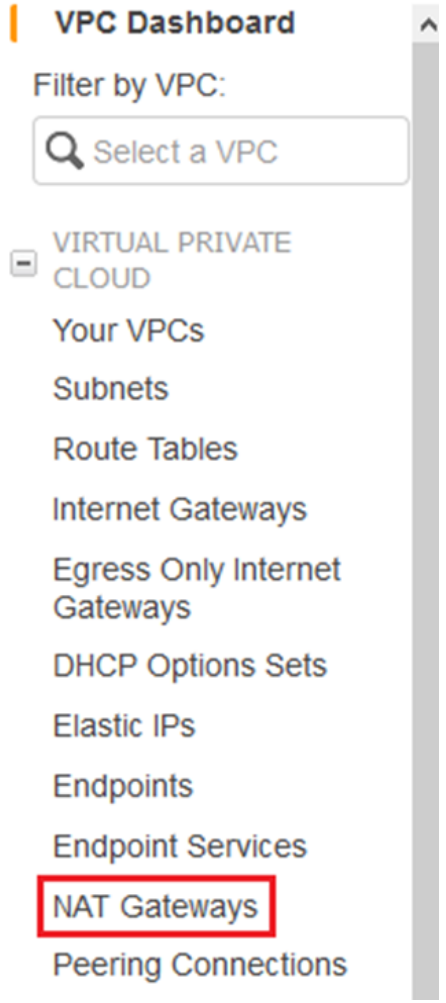
# INTERNET GATEWAYS

- An internet gateway (IGW) is a horizontally scaled, redundant, and highly available VPC component
- It facilitates communication between VPC instances and the internet
- It imposes no availability risks or bandwidth constraints on your network traffic
- An IGW serves two purposes:
  - To provide a target in your VPC route tables for traffic routable on the Internet
  - To perform network address translation (NAT) for instances that have been assigned public IPv4 addresses



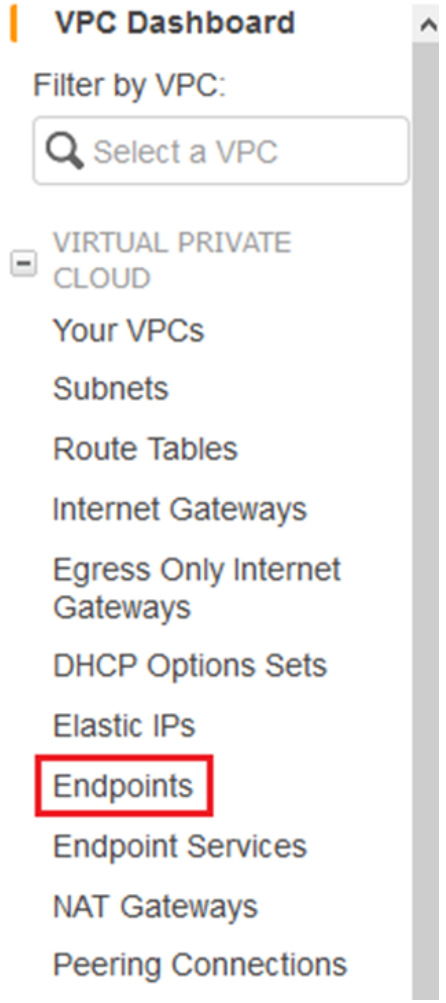
# NAT GATEWAYS

- NAT Gateways enable instances in a private subnet to connect to the internet or other AWS services, while preventing the internet from initiating a connection with those instances
- For example – Windows Update Services, upgrades, security fixes, etc.
- You are charged for creating and using a NAT gateway in your account with hourly usage and data processing rates applying
- The **Egress Only Internet Gateway** is used for IPv6 traffic to accomplish a similar service



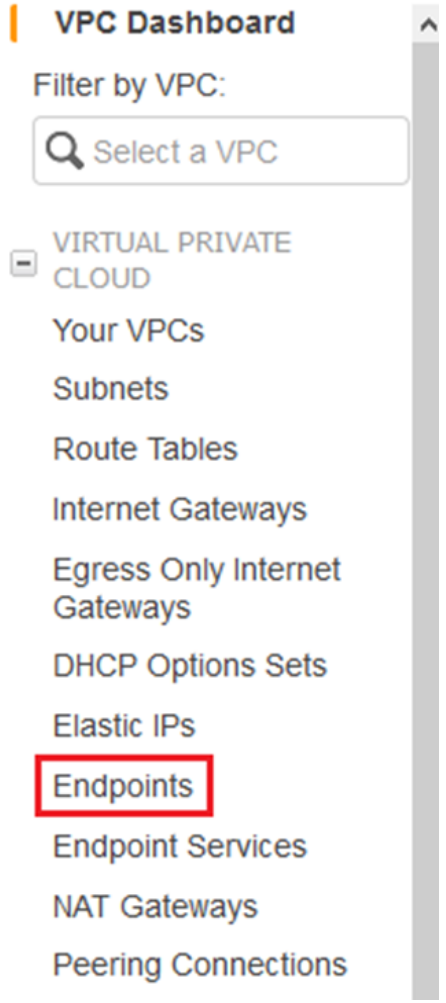
# ENDPOINTS

- Endpoints are virtual devices that allow you to privately connect your virtual network resources to supported services without needing an internet gateway, NAT device, VPN connection, or Direct Connect (ExpressRoute) connection
- Instances do not need public IP addresses to communicate with resources in the service since traffic between your VNet and the other service does not leave the provider network

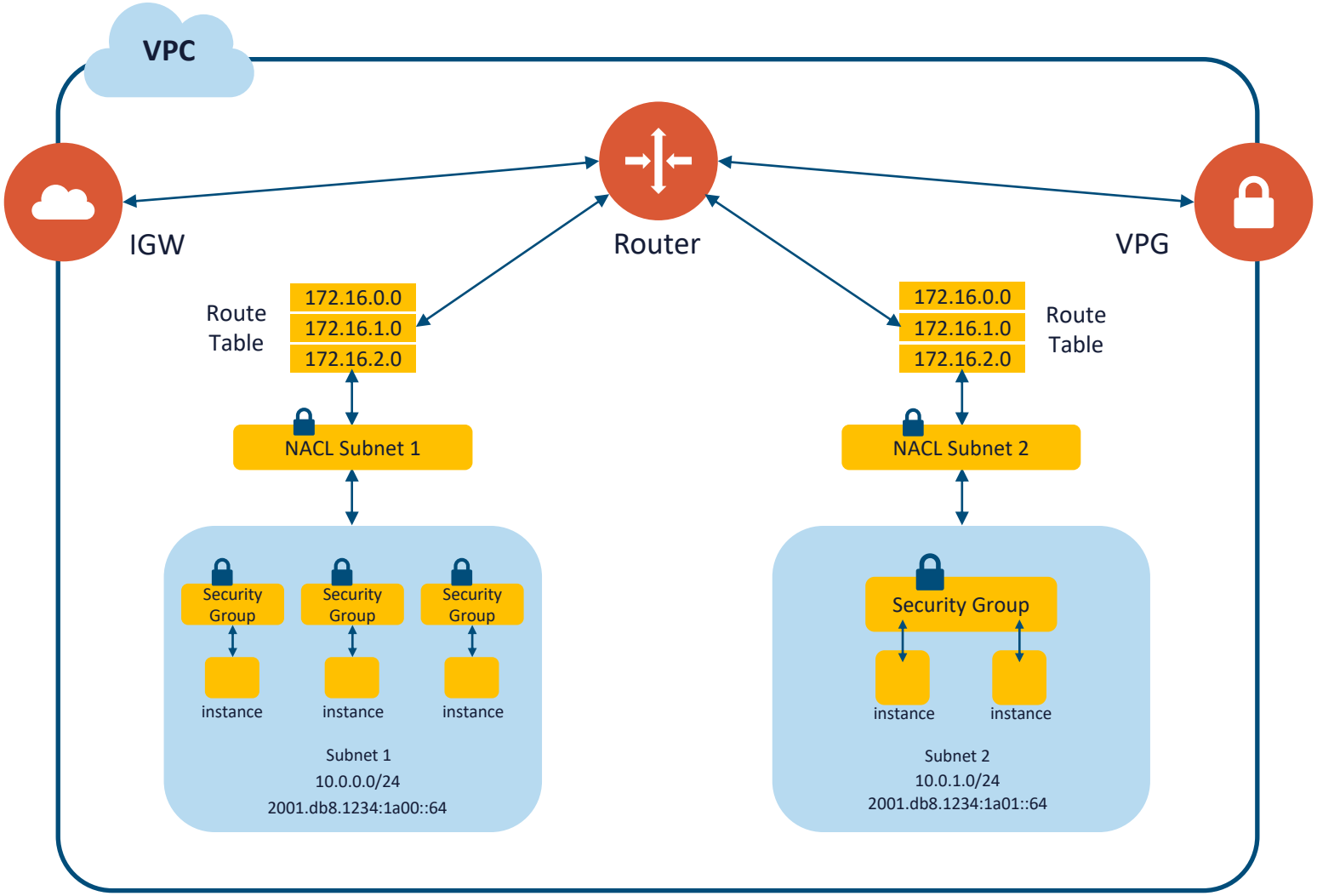


# ENDPOINTS (CONT.)

- Gateway endpoints - a gateway that you designate as a target for a route in your route table for traffic destined to a supported service
- Interface endpoints - An interface endpoint is an elastic network interface with a private IP address from the IP address range of your subnet that serves as an entry point for traffic destined to a supported service
- Endpoint services - Your own application in your VPC using PrivateLink. Other AWS principals can create a connection from their VPC to your endpoint service



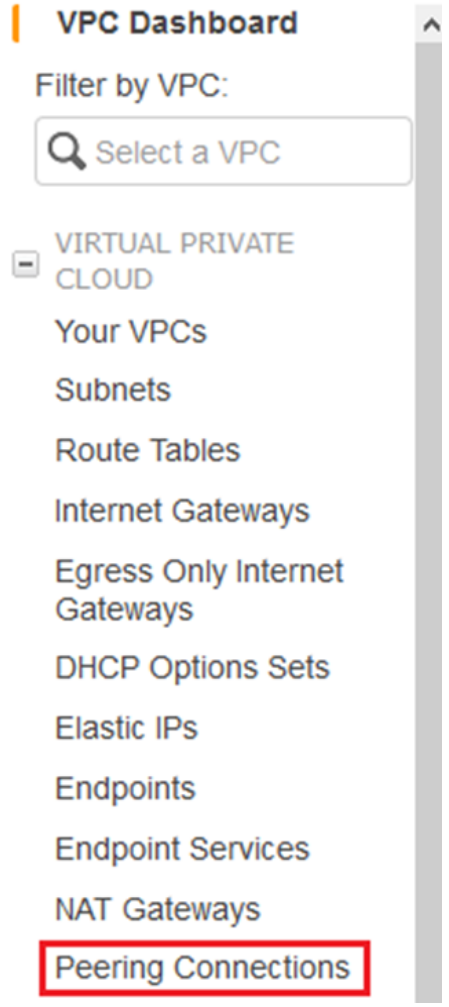
# USING ENDPOINTS



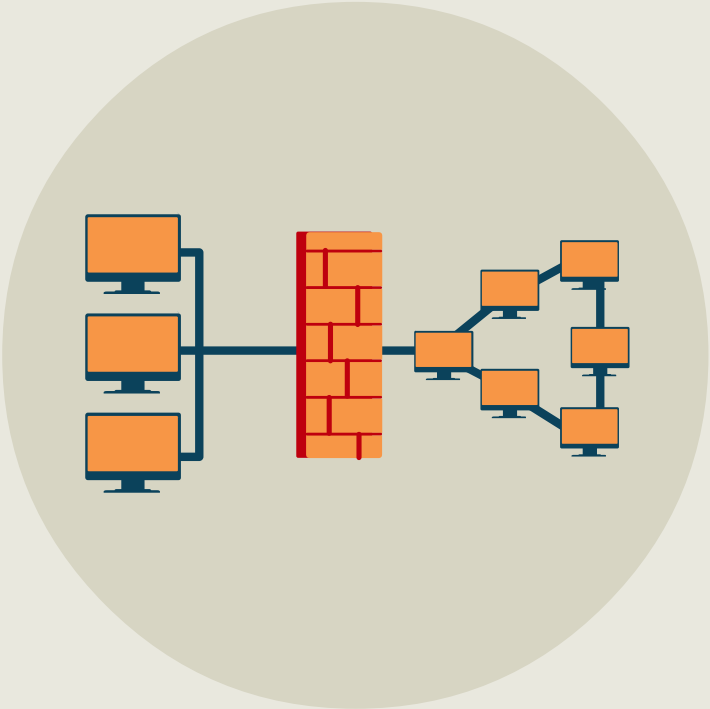

- S3 Storage
- DynamoDB
- Elastic Filesystem (EFS)
- CloudWatch and CloudTrail

# VPC PEERING

- A VPC peering connection is a networking connection between two VPCs that lets you route traffic between them using private IPv4 or IPv6 addresses
- Instances in either VPC can communicate with each other as if they are within the same network
- You can create a VPC peering connection between your own VPCs, or with a VPC in another AWS account
- Intra-region or inter-region peering connections

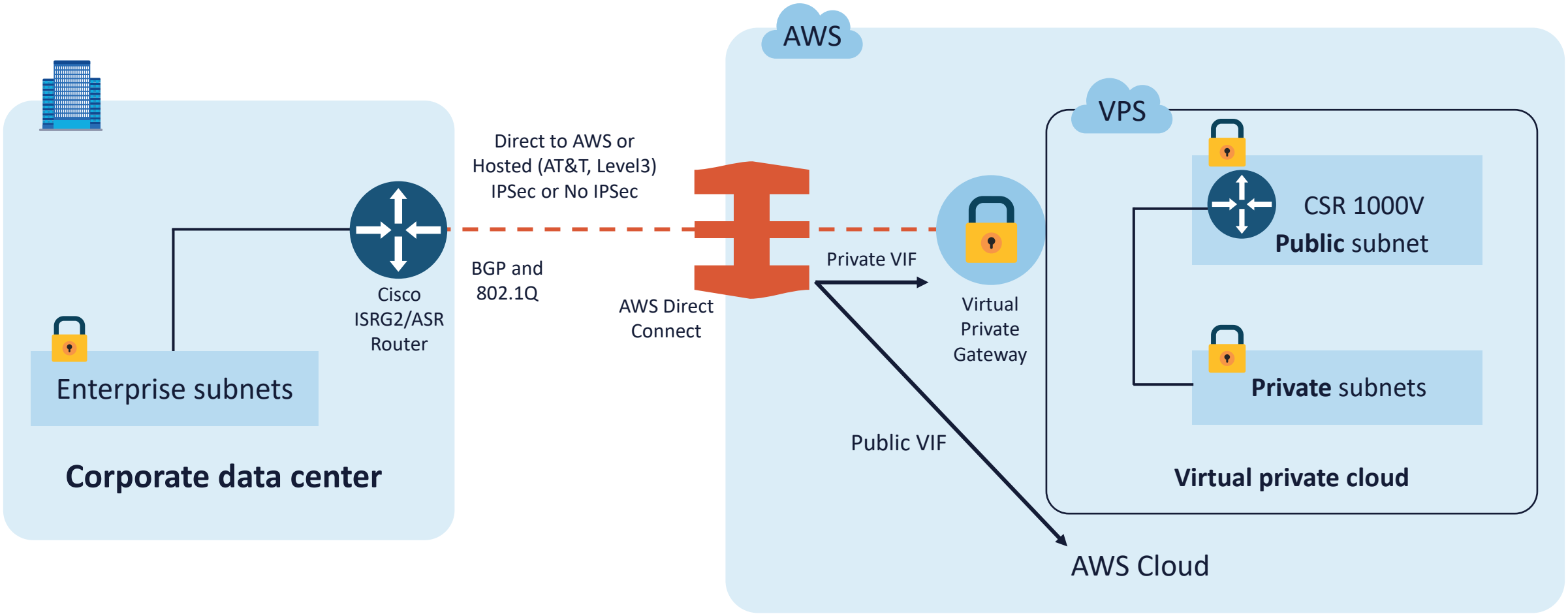


# CONNECTING TO THE CSP VIRTUAL NETWORK

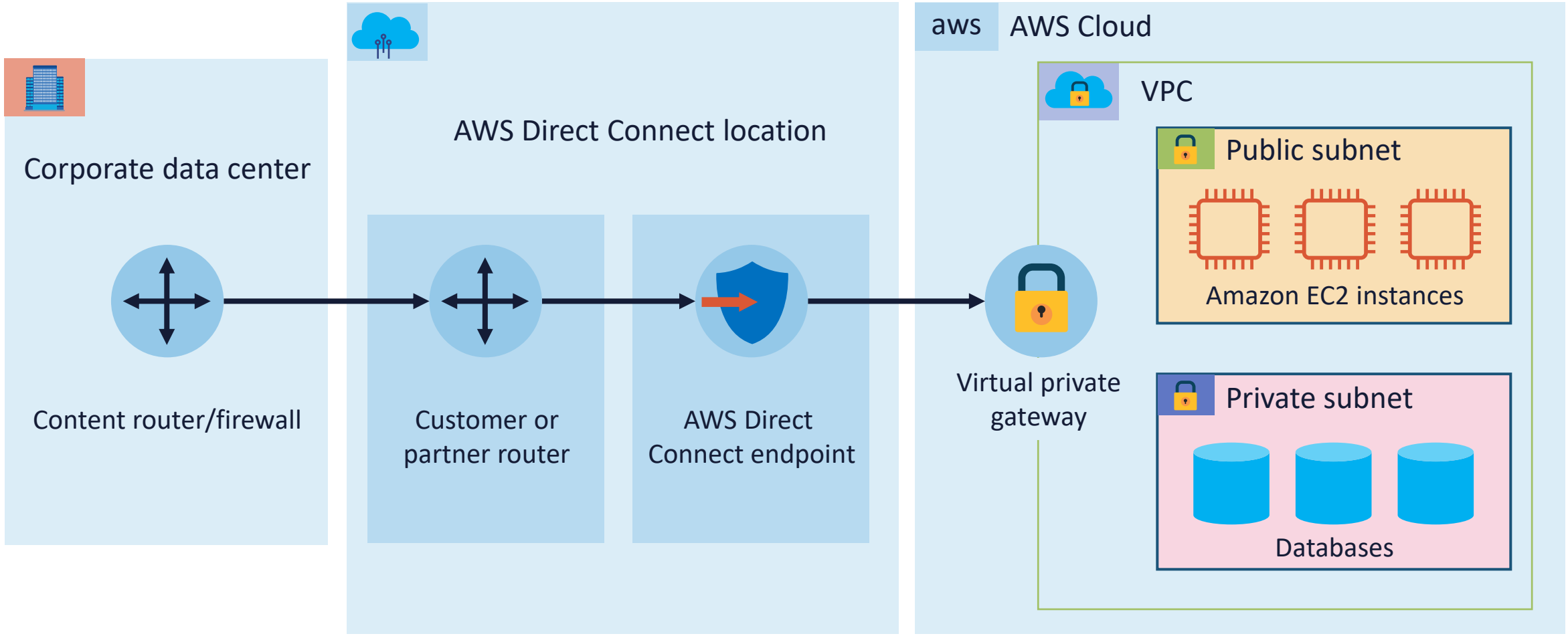


- **Azure** provides VNet Peering and Global VNet Peering, Azure VPN Gateway, Site-to-Site VPN, Point-to-Site VPN, and Express Route
- **GCP** has Google Global Networking, IPsec Cloud VPN, Cloud Interconnect (Direct Peering or Carrier Peering partner)
- **AWS** allows access from the internet via an internet gateway (IGW), VPNs (i.e., Site-to-Site VPG), Direct Connect, secure endpoints, and VPC peering (PrivateLink)

# DIRECT CONNECT SOLUTIONS



# DIRECT CONNECT SOLUTIONS



# DEMO: AWS SITE-TO-SITE VPN

# DEPLOYING CLOUD PROVIDER INSTANCES

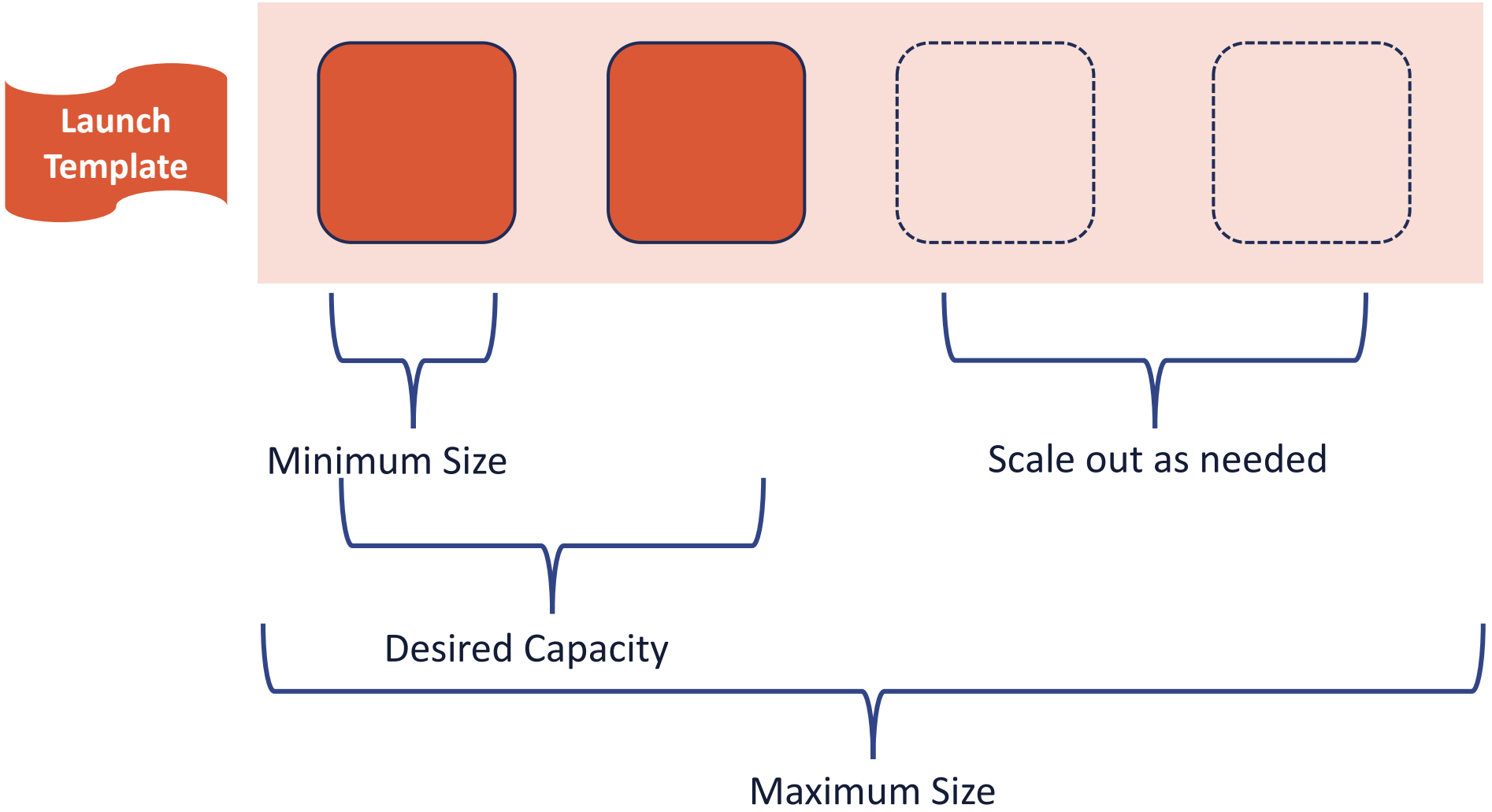
- All CSPs offer services that delivers resizable and secure compute capacity in the cloud
- This makes rapid web-scale cloud computing easier by using a simple web service interface (CLI/SDK as well)
- Provides control of your resources running on an established computing infrastructure

# CLOUD COMPUTING INSTANCES

- Allows you to select flexible configurations (virtual machine images)
- Increase or decrease capacity within minutes
- Create templates known as “golden images”
- Securely integrated with most cloud services
- Highly available, reliable, and durable

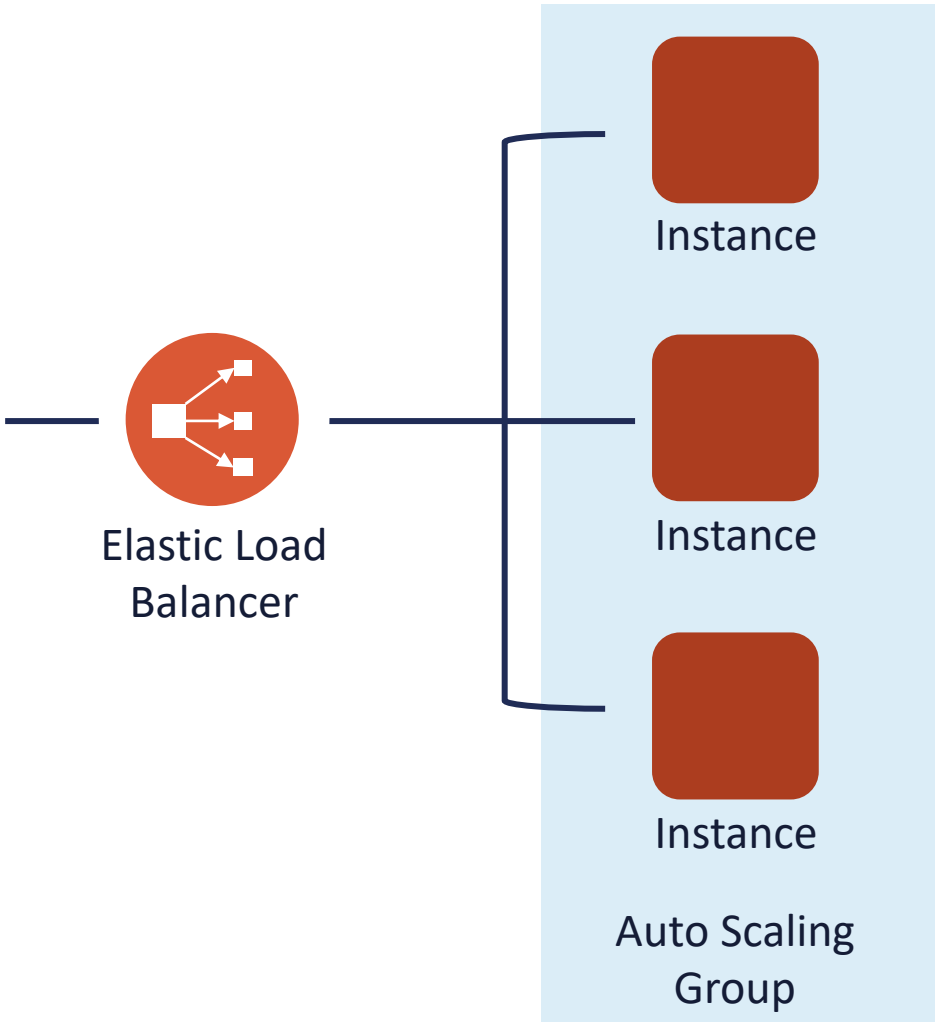
# DEMO: AWS ELASTIC COMPUTE CLOUD (EC2)

# AUTO SCALING GROUPS



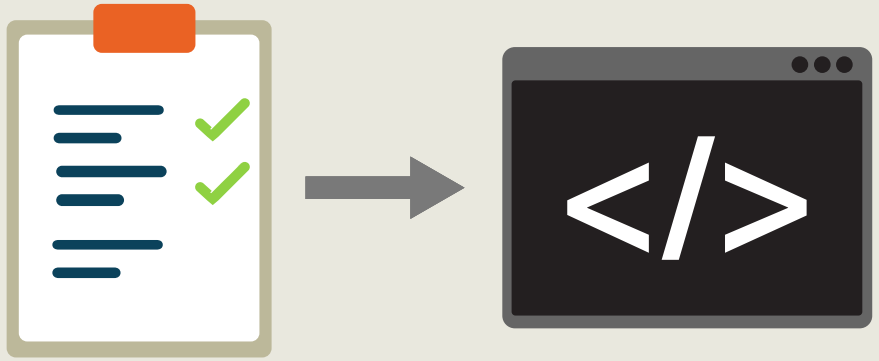
# ELASTIC LOAD BALANCING

- Elastic Load Balancing (ELB) automatically dispenses incoming traffic across several targets including O/S instances, containers, Lambda functions, and even IP addresses
- They can be public-facing or internal
- Application Load Balancers are for load balancing HTTP and HTTPS traffic for delivering modern application architectures
- Network Load Balancers are for TCP, UDP, and TLS traffic routing traffic to VPCs optimized for high-speed, low-latency traffic



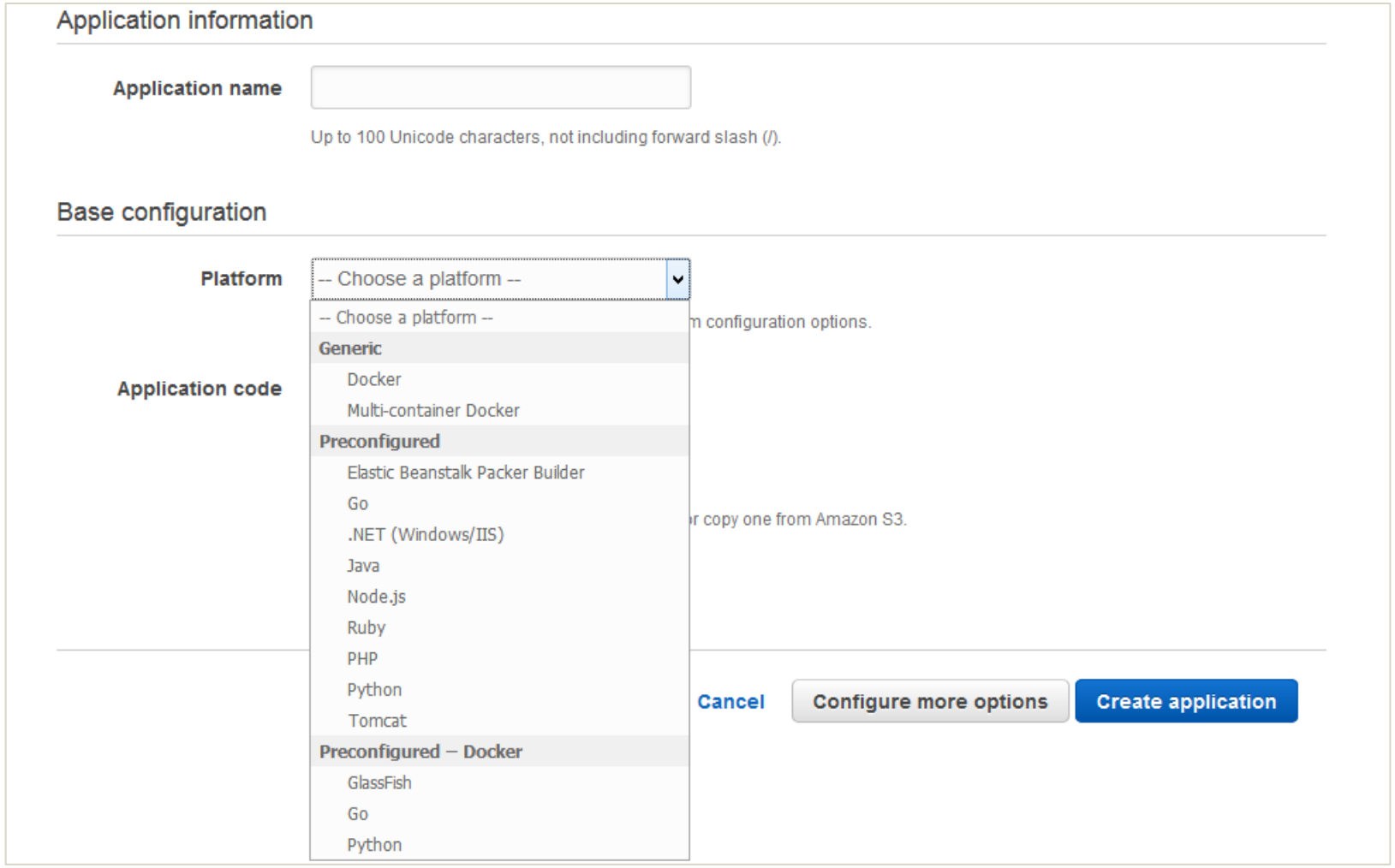
# ELASTIC BEANSTALK

**The traditional Dev platform**




- AWS Elastic Beanstalk is an easy-to-use service for deploying, monitoring and scaling web applications and services developed on several different platforms and applications
  - Choose your platform (Generic Docker, Preconfigured, Preconfigured Docker)
  - Upload an application or use a sample code from AWS
  - Run it

# ELASTIC BEANSTALK



# Functions-as-a-Service

- Run code in the cloud without deploying or managing servers (AWS Lambda, Azure Functions, GC Functions)
- Pay only for the compute time you consume and there is no charge when your code is not running
- You can run code for virtually any type of application or backend service—all with zero administration
- Functions can be triggered in several ways
  - Manual, alerts, alarms, URLs, API calls, events, etc.

# COMPARING CSP COMPUTE SERVICES

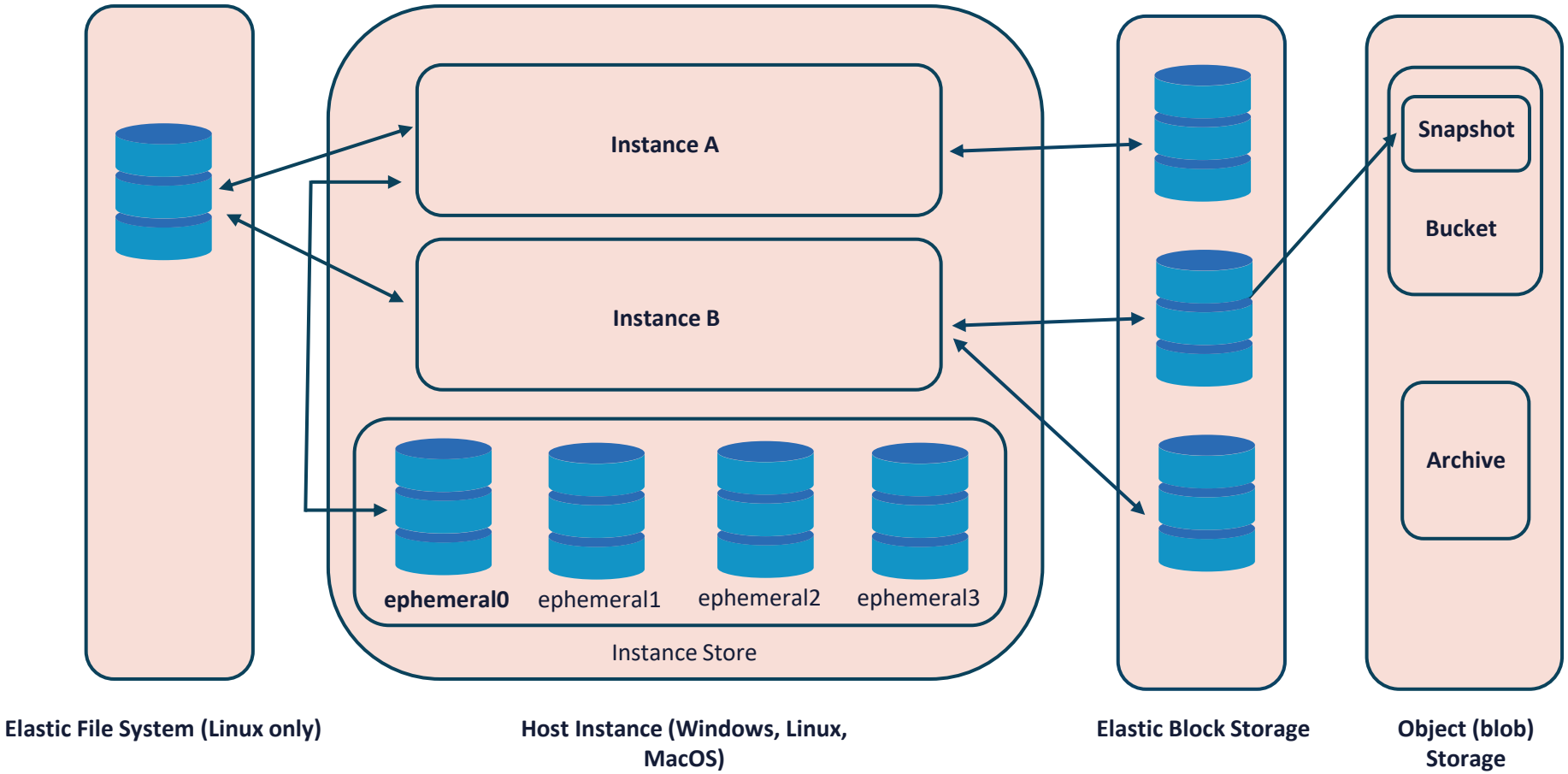
| Amazon Web Services | Google Cloud Platform | Microsoft Azure |
|---|---|---|
| Amazon EC2 | Google Compute Engine | Azure Virtual Machine |
| AWS Elastic Beanstalk | Google App Engine | Azure App Services |
| Amazon S3 | Google Cloud Storage | Azure Blob Storage |
| Amazon EC2 Containers | Kubernetes Engine | Azure container Service |
| Amazon Dynamo DB | Google Cloud Bigtable | Azure Cosmos DB |
| Amazon Redshift | Google BigQuery | Azure SQL Data WH |
| Amazon Lambda | Google Cloud Functions | Azure Functions |
| AWS Direct Connect | Google Cloud Interconnect | Azure ExpressRoute |
| AWS CloudWatch | Stackdriver Monitoring | Application Insights |

# CSP BLOCK STORAGE

## Network-based virtual HDD and SSD

- Files are split up and stored in fixed-sized blocks (volumes) on SSD or HDD arrays
- Uses cases are hosting databases, supporting random read/write operations, and keeping system files of the running virtual machines
- Suitable for apps that need high IOPS, database, and transactional data
- Capacity increased by adding more nodes
- Most offer Ephemeral (directly attached to hypervisor) and Elastic volumes (accessed over datacenter SAN)

# VOLUME (BLOCK STORAGE) (HDD AND SDD)



# DEMO: AWS ELASTIC BLOCK STORAGE (EC2)

# AMAZON ELASTIC FILE SYSTEM (AMAZON EFS)

- Amazon Elastic File System (Amazon EFS) provides a simple, scalable, elastic file system for Linux-based workloads for use with AWS Cloud services and on-premises resources
- It is built to scale on demand to petabytes without disrupting applications, growing and shrinking automatically as you add and remove files
- EFS is a fully managed service that requires no changes to your existing applications and tools, providing access through a standard file system interface for seamless integration

# CSP OBJECT STORAGE

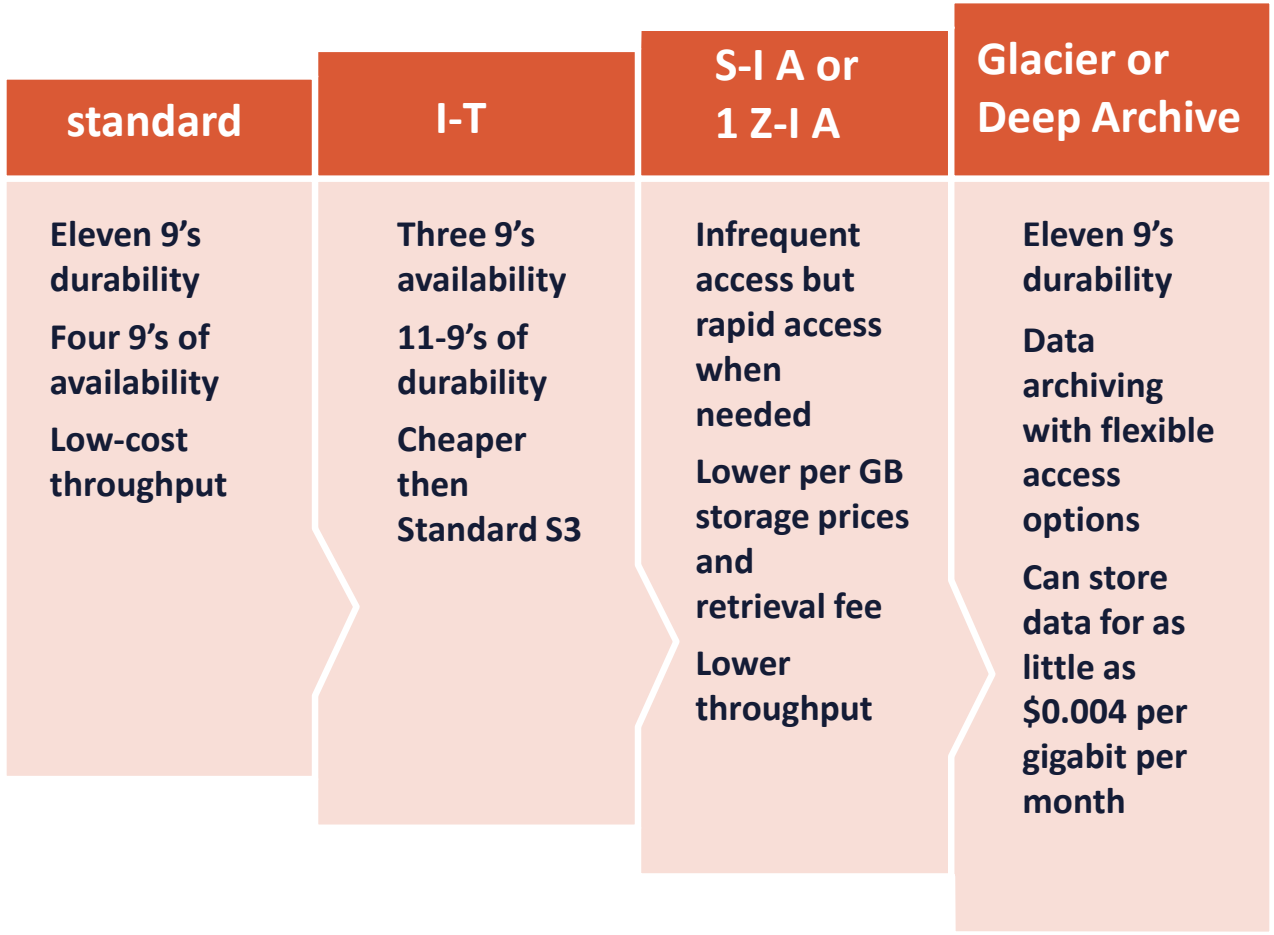
## Also referred to as blob storage

- **Amazon S3 / Google Cloud Storage / Azure Blob Storage**
- Data is stored as discrete objects – virtually unlimited file storage with extensible metadata capabilities
- Data is not placed in a hierarchy of directories and instead resides in a flat address space
- Objects are accessed using HTTP/S URLs or RESTful APIs
- Applications identify discrete data objects by their unique address
- Often the backend repository for CDN and other web applications

# AWS SIMPLE STORAGE SERVICE (S3)

- AWS S3 is object-based storage that is constructed to store and get unlimited volumes of data from anywhere on the Internet
- It provides a highly-available, extremely durable, and enormously scalable data storage infrastructure at very low cost

# S3 STORAGE PLANS (TIERS)



# DEMO: S3 SERVICES

# STORAGE GATEWAY

- AWS Storage Gateway is a hybrid storage service that enables your on-premises applications to seamlessly use AWS cloud storage
- You can use the service for backup and archiving, disaster recovery, cloud data processing, storage tiering, and migration
- Can be appliance-based or in a hypervisor
- Often used in conjunction with Direct Connect 10 Gbps

# The Snow Family

- **Snowcone** is a portable, rugged, and secure data box used to collect, process, and transfer up to 8 terabytes of  data to AWS, either offline by shipping the device, or online with an AWS DataSync solution
- **Snowball** is a data migration and edge computing device that comes in two device options: Compute Optimized and Storage Optimized
  - Snowball Edge Storage Optimized devices offer 40 vCPUs of compute capacity combined with 80 terabytes of usable block or Amazon S3-compatible object storage
- **Snowmobile** moves up to 100 PB of data in a 45-foot-long rugged shipping container and is ideal for multi-petabyte or Exabyte-scale digital media migrations and data center shutdowns

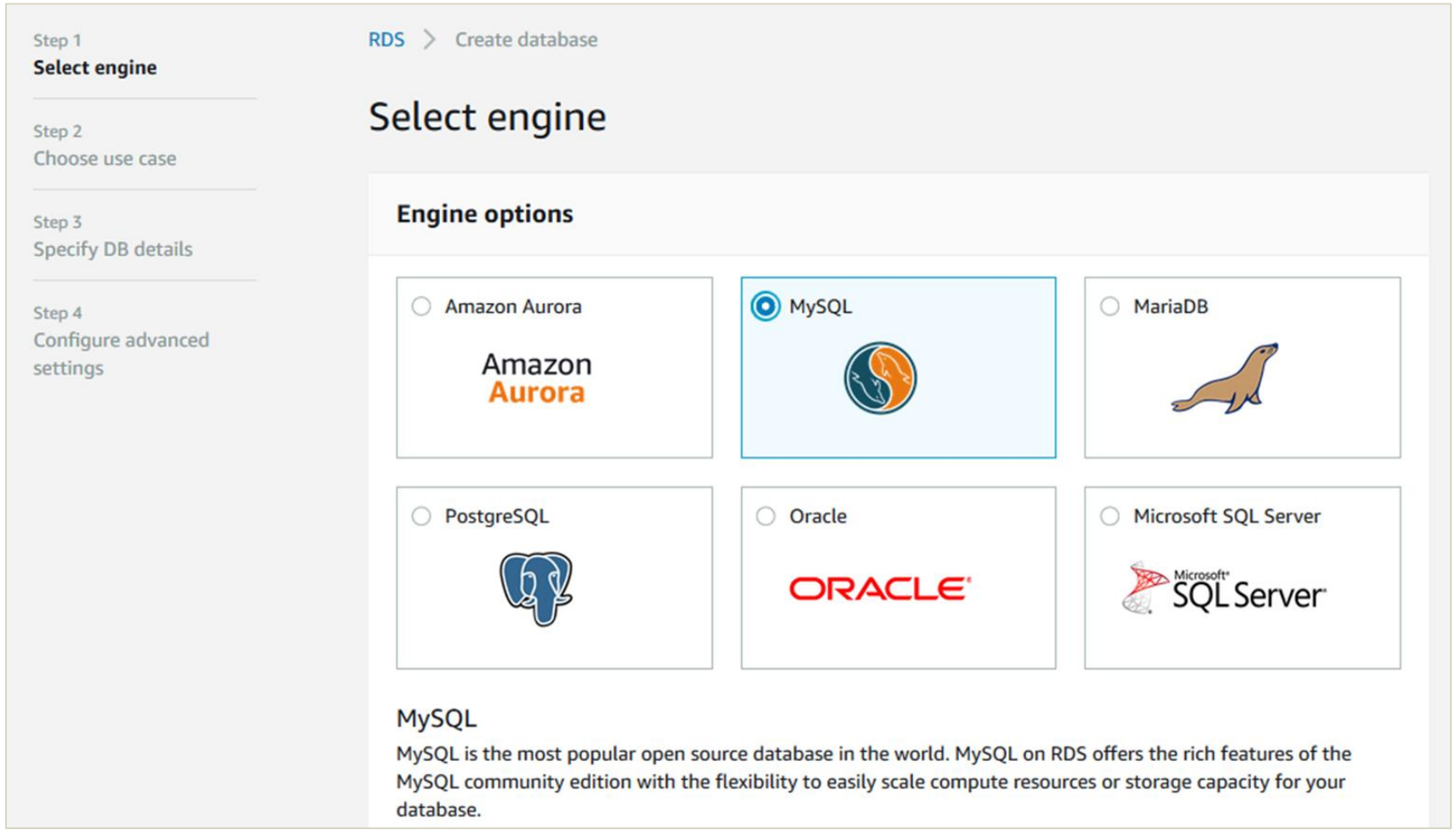

# SURVEY OF AWS DATABASE SERVICES

| Database type | Use cases | AWS service |
|---|---|---|
| **Relational** | Traditional applications, ERP, CRM, e-commerce | Amazon Aurora, Amazon RDS, Amazon Redshift |
| **Key-value** | High-traffic web apps, e-commerce systems, gaming applications | Amazon DynamoDB |
| **In-memory** | Caching, session management, gaming leaderboards, geospatial applications | Amazon ElastiCache for Memcached<br>Amazon ElastiCache for Redis |
| **Document** | Content management, catalogs, user profiles | Amazon DocumentDB |
| **Wide column** | High scale industrial apps for equipment maintenance, fleet management, and route optimization | Amazon Managed Apache Cassandra Service |
| **Graph** | Fraud detection, social networking, recommendation engines | Amazon Neptune |
| **Time series** | IoT applications, DevOps, industrial telemetry | Amazon Timestream |
| **Ledger** | Systems of record, supply chain, registrations, banking transactions | Amazon QLDB |


# AWS RELATIONAL DATABASE SERVICES (RDS)

- Amazon Relational Database Service (RDS) is a managed service for setting up, operating, and scaling a cloud-based relational database
- RDS is available on several database instance types that are optimized for memory, performance or I/O
- Can choose from Amazon Aurora, PostgreSQL, MySQL, MariaDB, **Oracle Database**, and SQL Server
- **Use the AWS Database Migration Service and/or Storage Gateway to migrate or replicate your existing databases to Amazon RDS**

# AWS RELATIONAL DATABASE SERVICES (RDS)



# DynamoDB

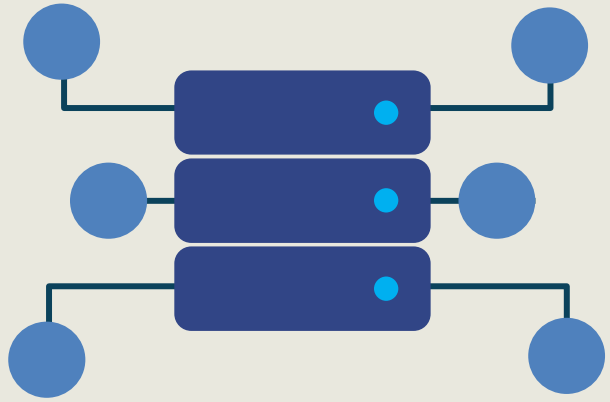
- Amazon DynamoDB is a key-value and document database (NoSQL) that provides single-digit millisecond performance at any scale
- It is a fully managed, multi-region, multi-master database with built-in security, backup and restore, and in-memory caching for internet-scale applications
- It can handle more than 10 trillion requests per day and support peaks of more than 20 million requests per second
- Over 100,000 AWS clients use DynamoDB as their key-value and document database

# AMAZON ELASTICACHE

- Amazon ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud
- The service improves the performance of web applications by empowering one to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases
- Amazon ElastiCache supports Redis and Memcached open-source in-memory caching engines

# AMAZON REDSHIFT

## Data Warehousing and Data Lakes



- Amazon Redshift clusters provide a fast, scalable data warehouse for cost-effective analysis of data across data warehouses and data lakes
- Uses machine learning, massively parallel query execution, and columnar storage on high-performance disks
- High security is provided using a 4-key nested encryption model

# COMPARING CLOUD DATABASE SERVICES

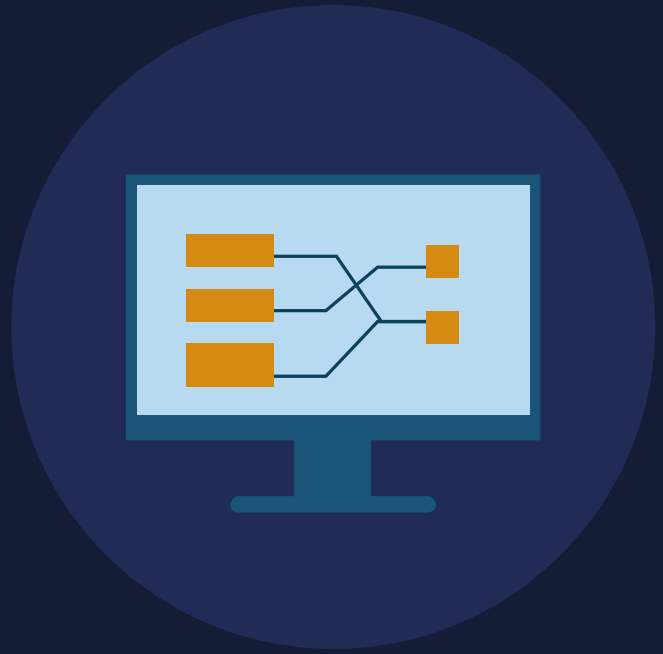
| Amazon Web Services | Google Cloud Platform | Microsoft Azure |
|---|---|---|
| Amazon EC2 | Google Compute Engine | Azure Virtual Machine |
| AWS Elastic Beanstalk | Google App Engine | Azure App Services |
| Amazon S3 | Google Cloud Storage | Azure Blob Storage |
| Amazon EC2 Containers | Kubernetes Engine | Azure Containers Services |
| Amazon Dynamo DB | Google Cloud Bigtable | Azure Cosmos DB |
| Amazon Redshift | Google BigQuery | Azure SQL Data WH |
| Amazon RDS | Google Cloud Spanner/SQL | Azure Database |
| AWS Direct Connect | Google Cloud Interconnect | Azure ExpressRoute |
| AWS CloudWatch | Stackdriver Monitoring | Application Insights |

# CLOUDWATCH

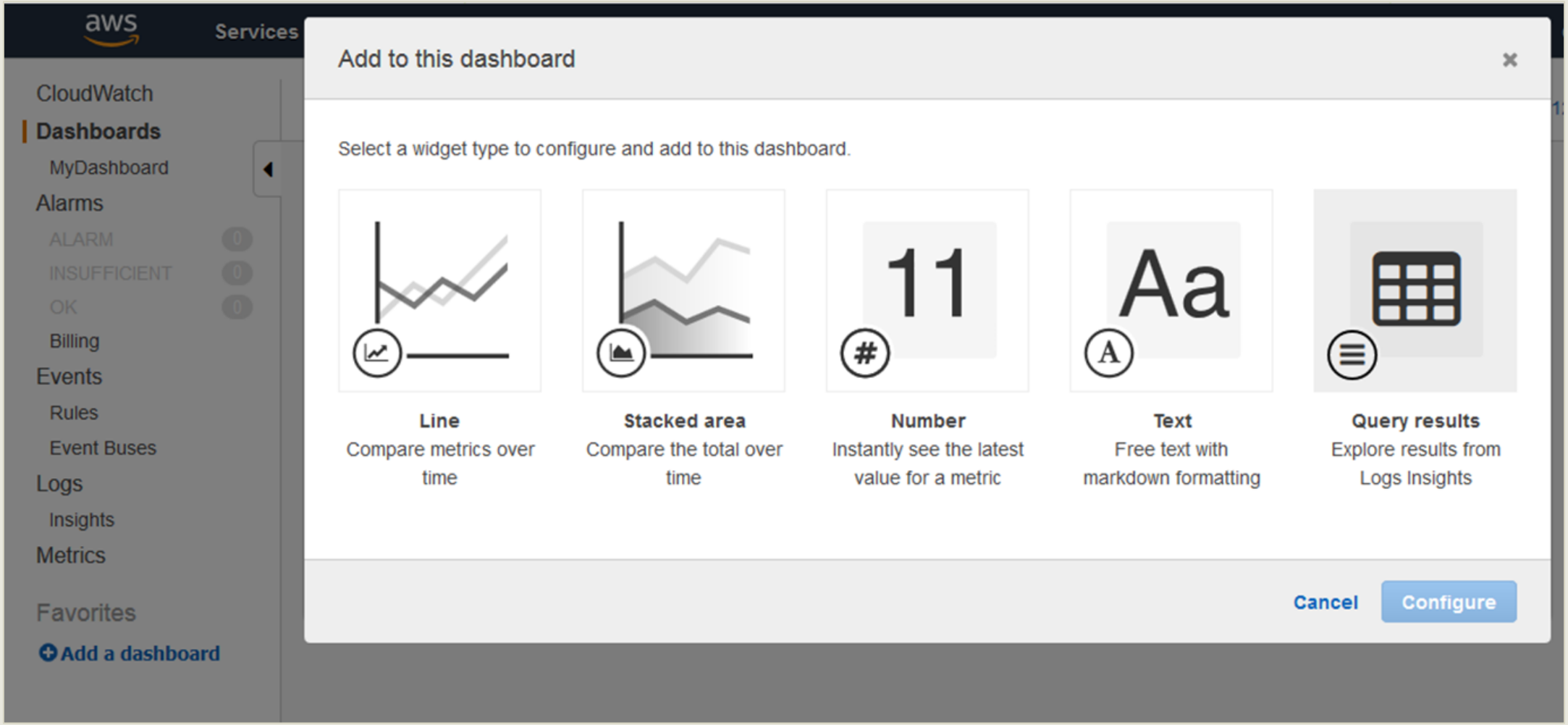
- Amazon CloudWatch is used for management and governance
- It is a monitoring and management service designed for developers, system operators, site reliability engineers (SRE), and managers
- CloudWatch offers data, meaningful metrics, and actionable insights to:
  - Monitor applications
  - Recognize and respond to system-wide performance changes
  - Optimize resource utilization
  - Gain a unified view of operational health

# CLOUDWATCH USE CASES

- Monitor critical metrics and logs, visualize application and infrastructure stacks, generate alarms, and correlate metrics and logs to recognize and resolve the root cause of performance issues
- Monitor applications and trigger automated CloudWatch Alarms and Lambda workflows to enhance the customer experience
- Explore, analyze, and visualize logs instantly to optimize resources, leverage CloudWatch Alarms to automate capacity, and do resource planning for auto scaling



# CLOUDWATCH DASHBOARDS



# CLOUDTRAIL

- With CloudTrail, customers can log, continuously monitor, and retain account activity related to all API calls across the AWS infrastructure
- Within CloudTrail, CloudTrail Insights can be enabled where CloudTrail can automatically detect unusual API activities in AWS accounts
- **Example**: CloudTrail Insights could detect that a higher number of Amazon EC2 instances than usual have recently launched in an account or abnormal account activity has occurred then review the full event details to determine which actions need to be taken next

# CLOUDTRAIL USE CASES

- **Exam: CloudTrail is one of the most common tools for getting insights into security events at AWS**
- Detect that a higher number of Amazon EC2 instances than usual have recently launched
- Identify which users and accounts called AWS, the source IP address from which the calls were made, and when the calls occurred
- Create a workflow to add a specific policy to an Amazon S3 bucket when CloudTrail logs an API call that makes that bucket public
- Connect your VPC to CloudTrail by defining an interface VPC endpoint for CloudTrail

# DEMO:
# AWS IAM