# Welcome to the CCSP Bootcamp

Your instructor:

## Michael J Shannon

CISSP #42221 / #524169,
CCSP, CSA-CCSK
CCNP-Security, PCNSE7,
AWS Certified Security – Specialty,
GIAC GSEC, OpenFAIR, and
ITIL 4 Managing Professional

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

# Certified Cloud Security Professional (CCSP) 2022 Exam

## CCSP Exam Details

- Length of exam : 4 hours
- Number of questions: 150
- Exam format: Multiple Choice
- Passing grade: 700 out of 1000
- Exam availability: English, Chinese, German, Japanese, Korean, Spanish
- Testing center: Pearson VUE Testing Center

# CCSP Examination Weights

| Domains | Weight |
| --- | --- |
| 1. Cloud Concepts, Architecture and Design | 17% |
| 2. Cloud Data Security | 20% |
| 3. Cloud Platform and Infrastructure Security | 17% |
| 4. Cloud Application Security | 17% |
| 5. Cloud Security Operations | 16% |
| 6. Legal, Risk and Compliance | 13% |
| **Total:** | **100%** |

# Domain 1

# Cloud Concepts, Architecture and Design


# Cloud Computing Definitions

- **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 definitions are covered throughout this course
- Study references:
  - https://csrc.nist.gov/glossary
  - ISO/IEC 17788 “Cloud Computing – Overview and Vocabulary”

# 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 Computing Roles

- **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

# Key Cloud Computing Characteristics

**On-demand self-service** - a cloud consumer provisions resources automatically when needed with little or no intervention from the provider

**Broad network access** – consistent availability to cloud resources using a myriad of components, with the help of load balancers, edge computing, and content delivery (distribution) networking (CDN)

**Multi-tenancy** - The ability to have multiple customers and applications running within the same

**Rapid elasticity** – allows customer to provision and de-provision cloud resources to meet immediate needs, often in minutes

# Key Cloud Computing Characteristics

**Scalability** – the ability to leverage the massive amounts of resources of the cloud provider to add additional compute, storage, database, and networking

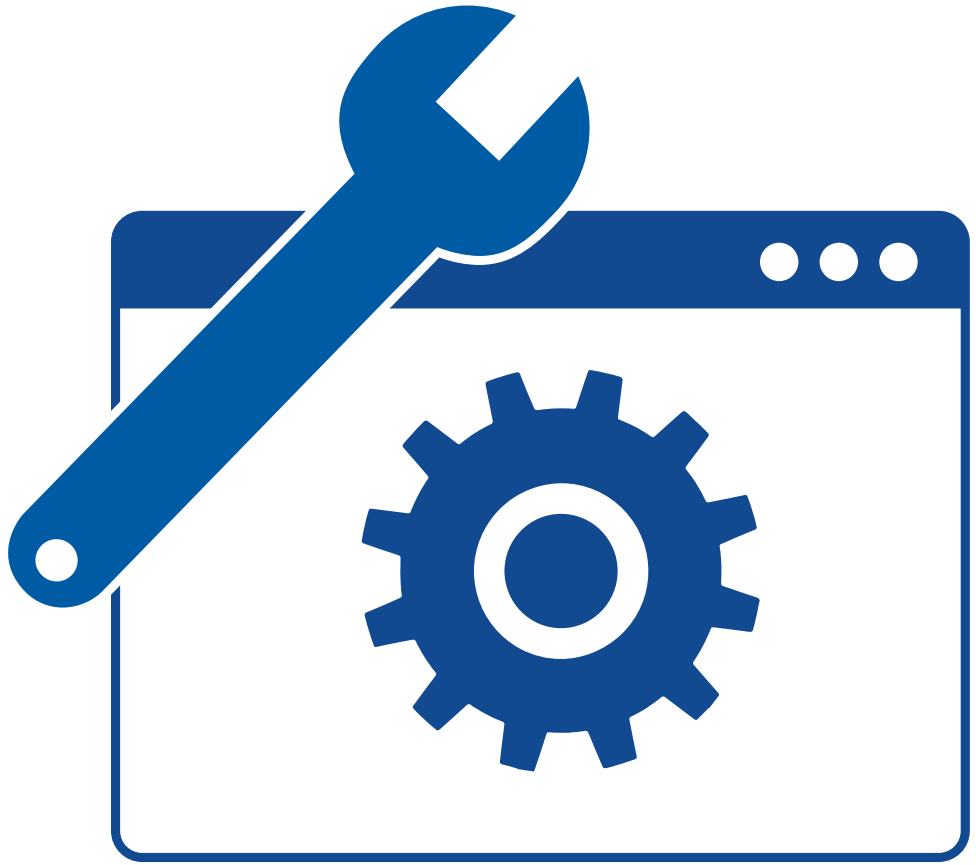
**Resource pooling** - the aggregation of resources allocated to cloud consumers by the cloud service provider

**Measured service** - the delivery and billing of cloud services in a metered fashion

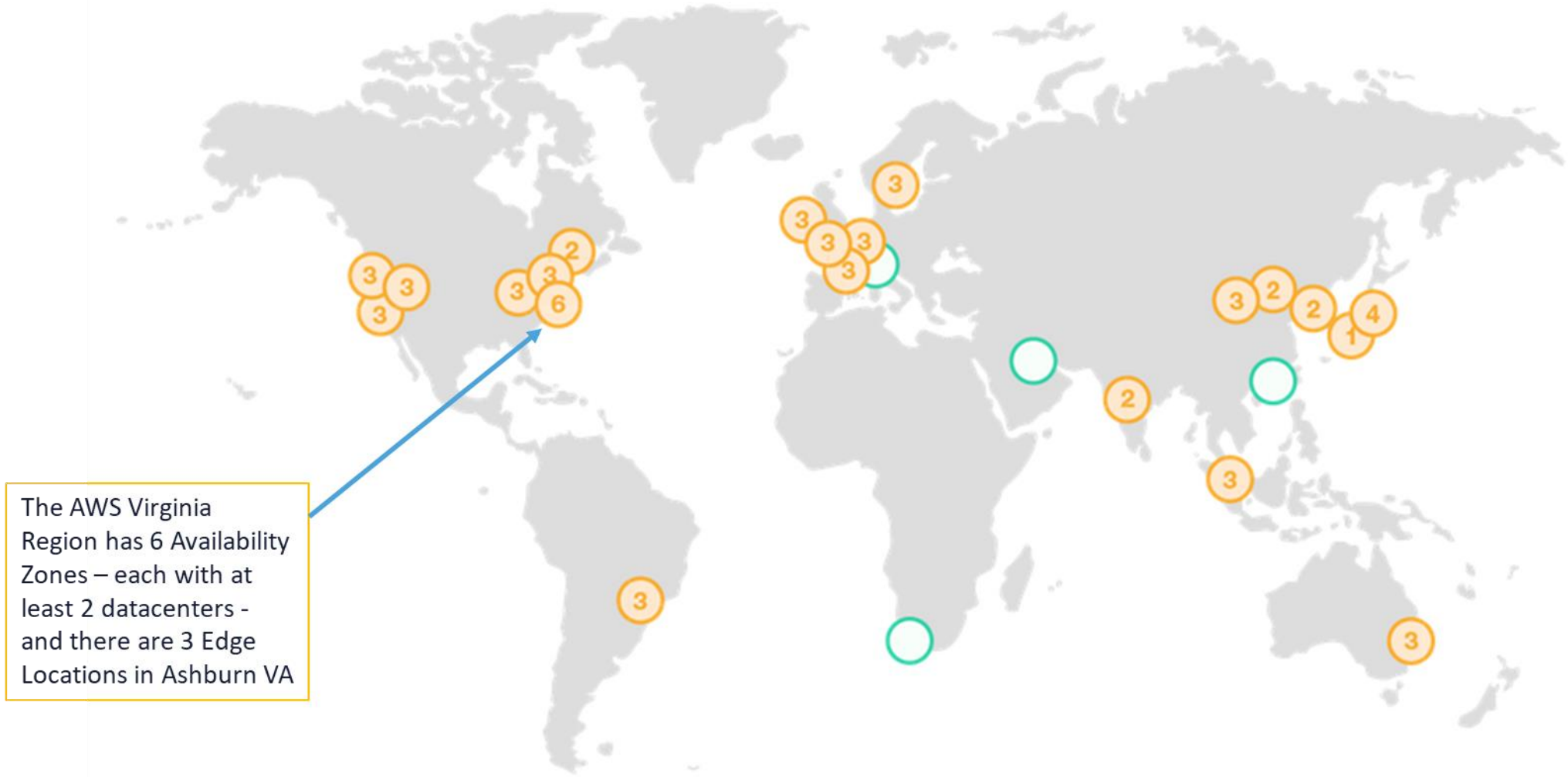
# Building Block Technologies

**Using “bleeding-edge” CSP tech**

- Virtualization (Type 1 Hypervisors)
- Storage and SAN
- Databases and RAID arrays
- Redis clusters
- Networking (VXLAN)
- SD-LAN and SD-WAN
- Automation and orchestration



# Amazon Web Services Global Infrastructure



# 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**

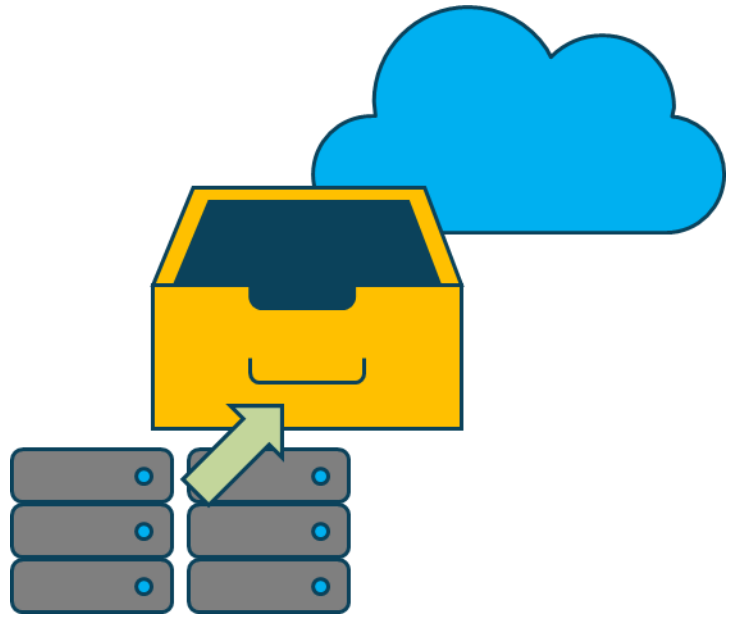
Fluid and simplified migration of applications and data to AWS

**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**

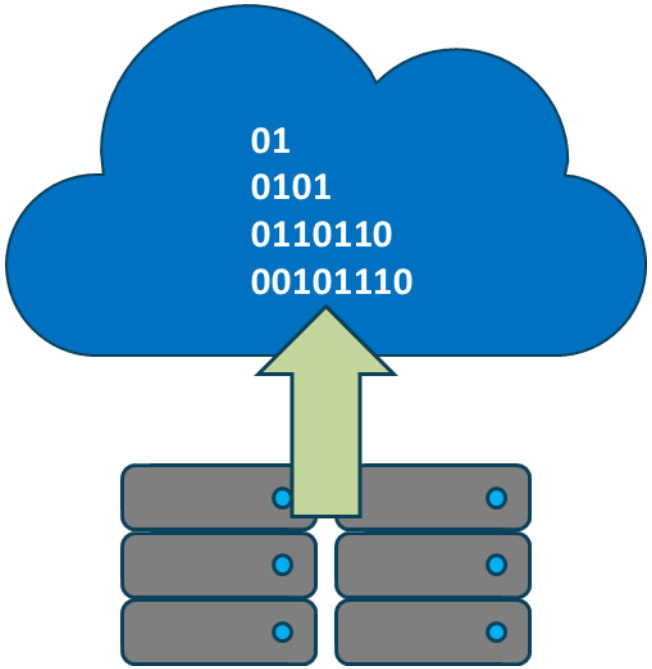
Superior networking and cloud sized clusters for multifaceted challenges

**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**

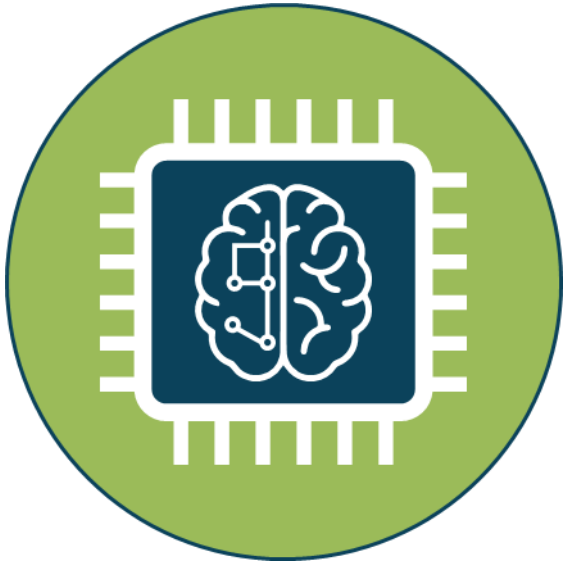
Perform analysis, object storage, and distribution of enormous data sets

**Serverless Computing**

Build and run applications without needing underlying servers

**Web sites**

Use dependable, highly scalable, and affordable web application tools

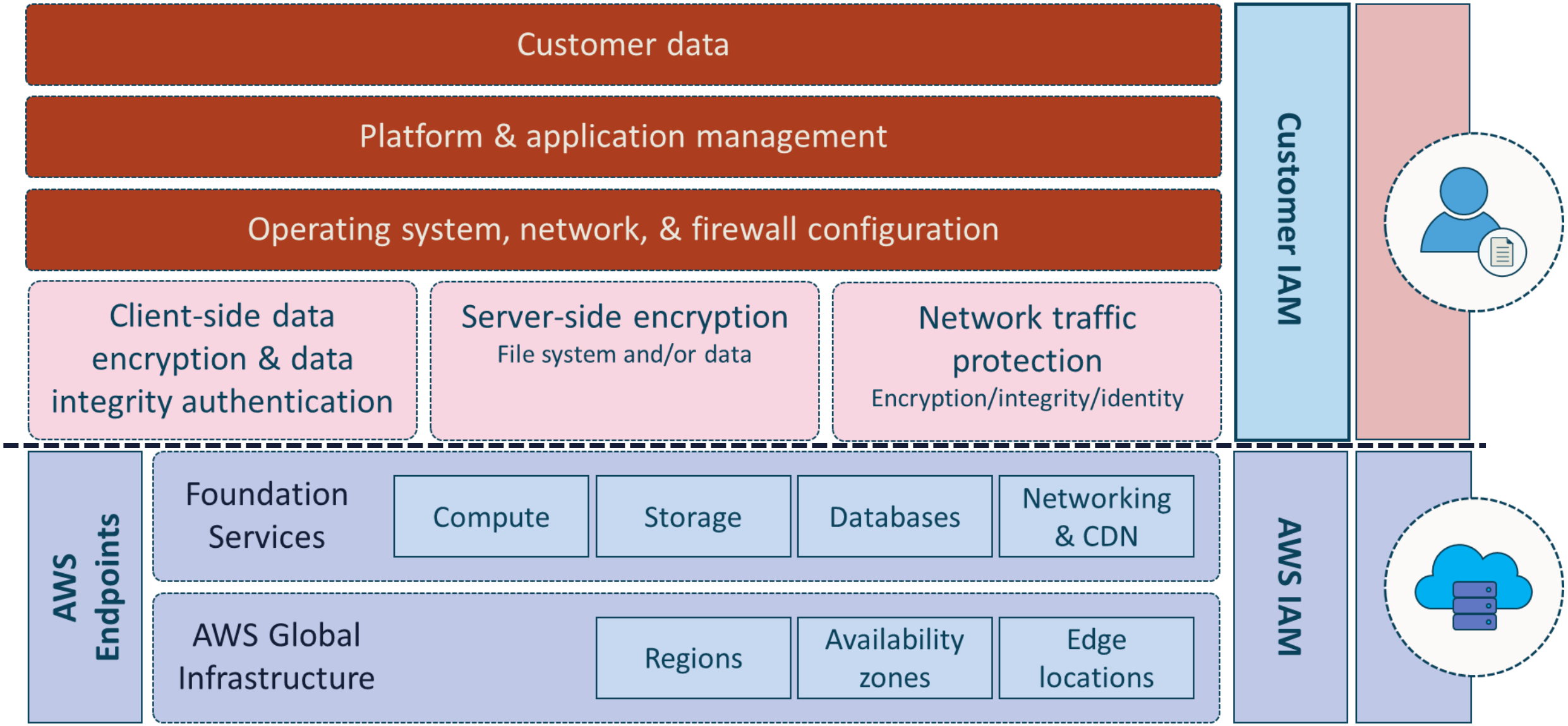


# IaaS According to NIST

“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 is able to 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)."

# 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

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

# Public Cloud Deployment

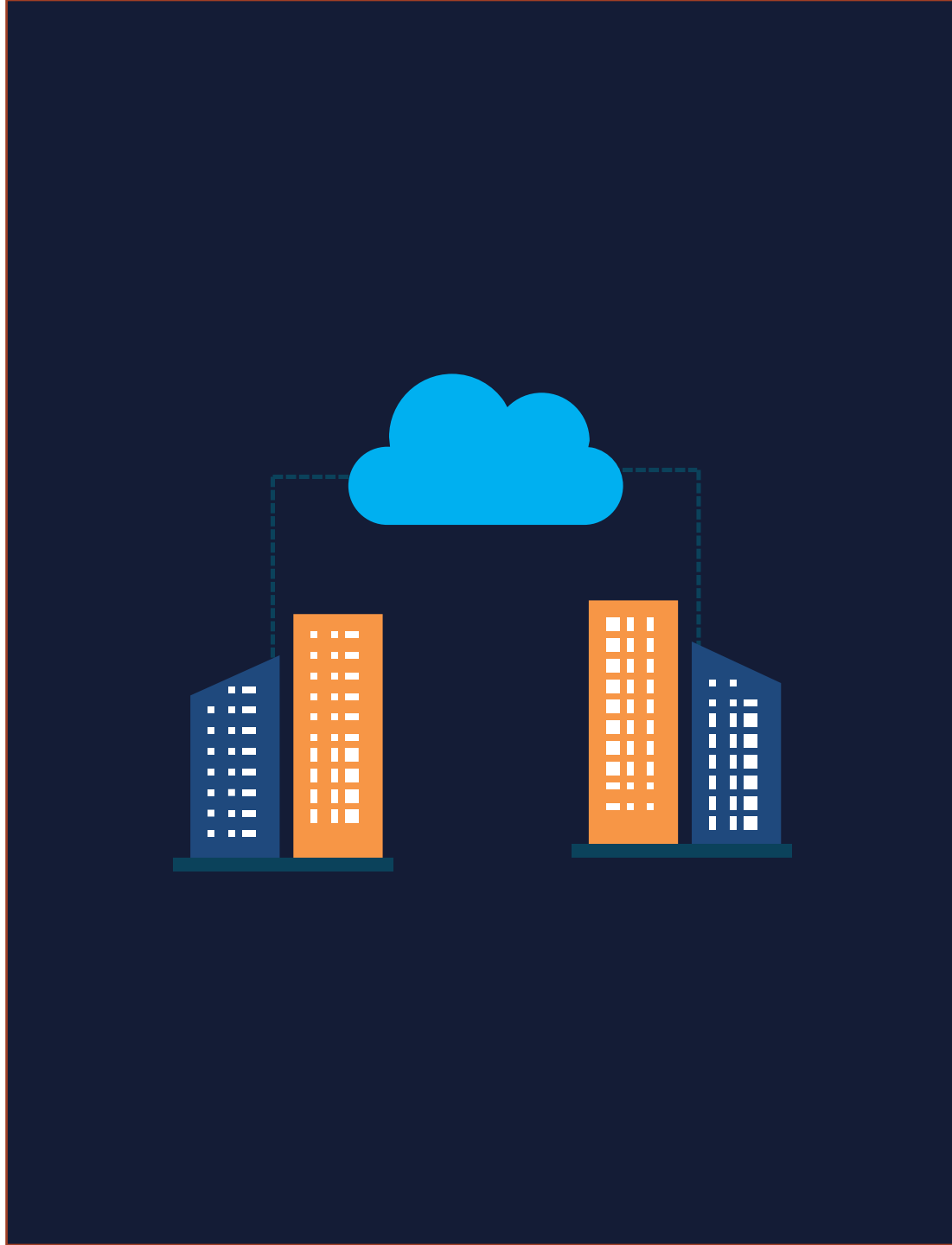
- A model where computing resources are owned and operated by a provider and shared across multiple tenants via the Internet or other public networks
- Enterprises often use public cloud for less-sensitive applications that have unpredictable spikes in usage or for storing data that does not require 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 (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

# Multi-cloud Deployment

- Multi-cloud is a cloud computing model where an enterprise leverages a combination of clouds (two or more public clouds, two or more private clouds, or a combination of public, private and edge clouds) to distribute applications and services and:
  - Accelerate app transformation and the delivery of new apps
  - Avoid vendor lock-in and ensure enterprise sovereignty: Total cloud spend, data sovereignty, vendor dependencies and lock-in are increasing concerns
  - Distribute applications and services to the edge in industries such as logistics, retail and manufacturing, the next generation of gains in automation, efficiency and improved customer experiences require applications to be distributed to the edge, closer to physical devices and users
  - Support the rise of the distributed workforce that secure and manage users and their devices in the new hybrid workforce challenge

# Cloud Shared Considerations

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

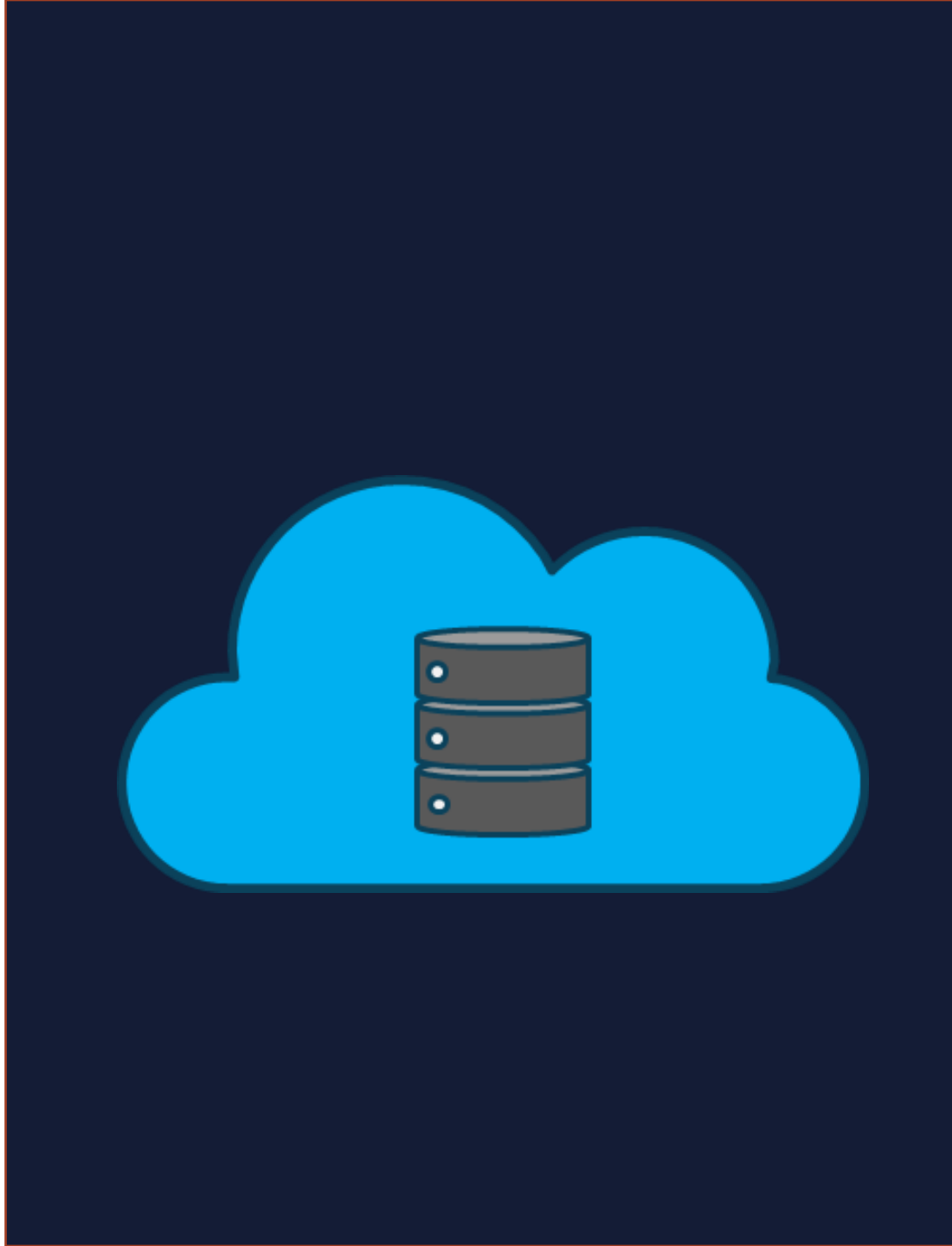
- Agility (flexibility) and elasticity
- Reversibility
- Cost
- Security
- Interoperability and portability
- Availability vs. durability
- Resiliency
- Security and privacy
- Performance
- Governance and regulatory
- Auditability
- Maintenance and versioning
- Service levels and Service Level Agreements (SLA)

# Agility

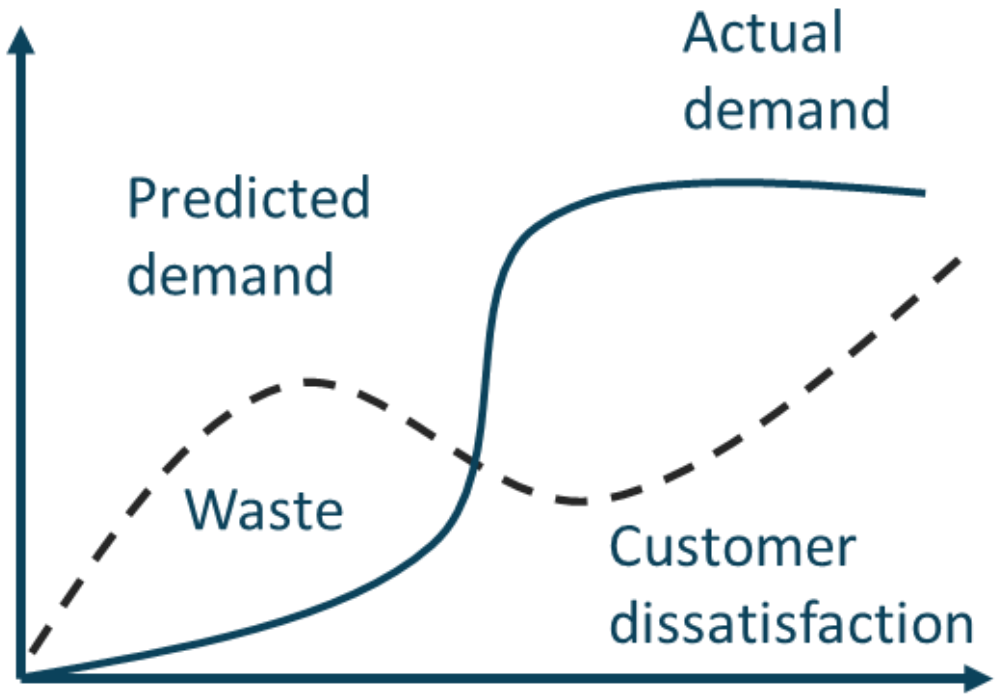
- Leveraging for rapid deployment, testing, experimentation, & innovation
- Overcoming geographical limitations
- Getting content as close to the consumer as possible
- Reducing time and cost for testing and experimentation
- Allows consumers better innovation

# Elasticity

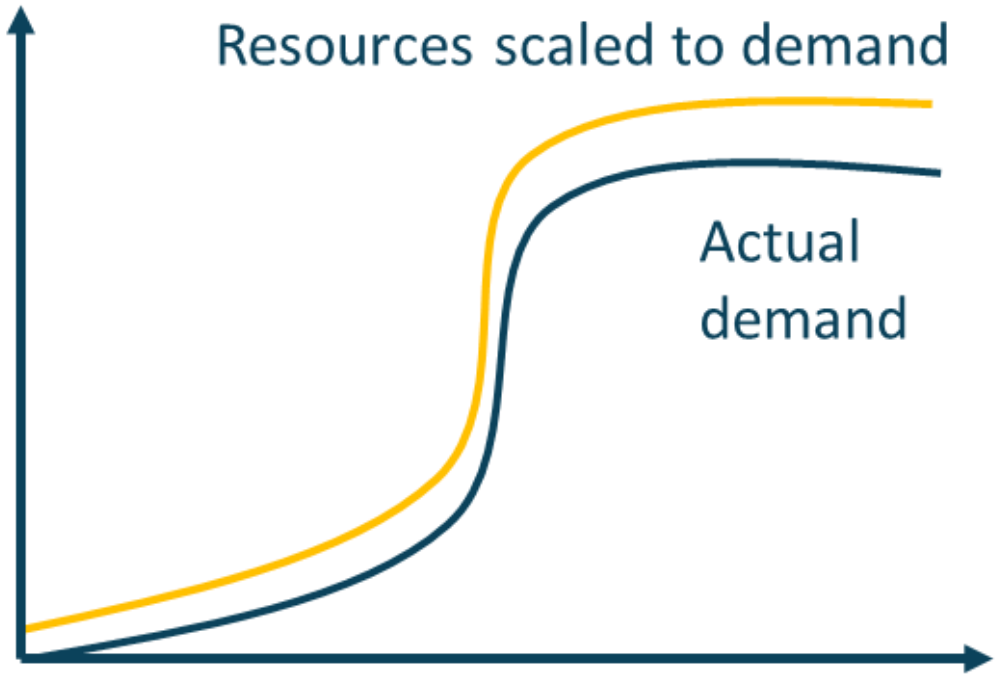
- Elasticity is the ability to almost instantly provision and de-provision resources by leveraging dynamic auto-scaling technologies
- Meets the challenges of predicting demand that leads to higher costs
- “Elasticity of the cloud allows us to add thousands of virtual servers and petabytes of storage 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

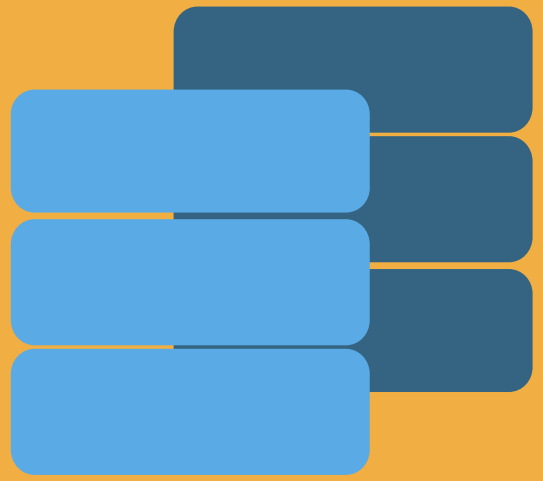



Rigid On-premises Resources




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** 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 multiple facilities within a selected 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

# Impact of Related Technologies

- Data science
- Machine learning
- Artificial intelligence (AI)
- Blockchain
- Internet of Things (IoT)
- Containers
- Quantum computing
- Edge computing
- Confidential computing
- DevSecOps

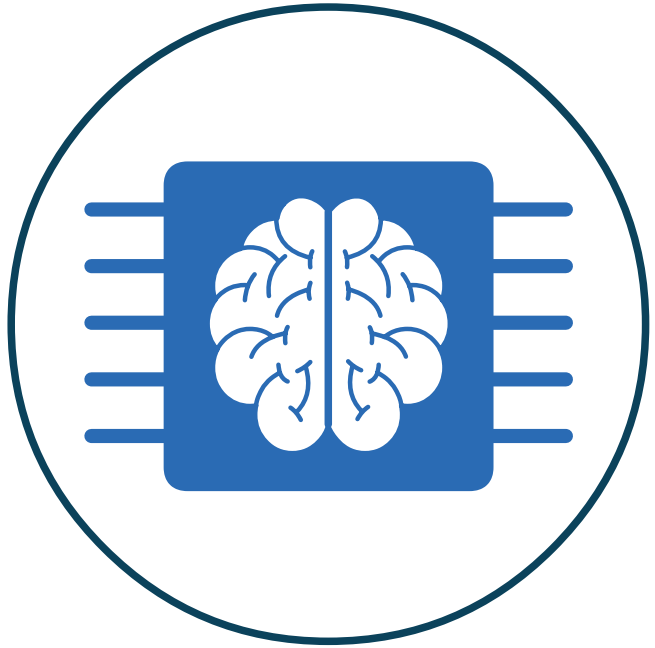
# Data Science

- Data discovery and ingestion
- Data lake and data warehouse
- Data preprocessing
- Data analysis and business intelligence
- Machine learning training and serving
- Accountable Artificial Intelligence (AI)
- Orchestration

# Machine Learning

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

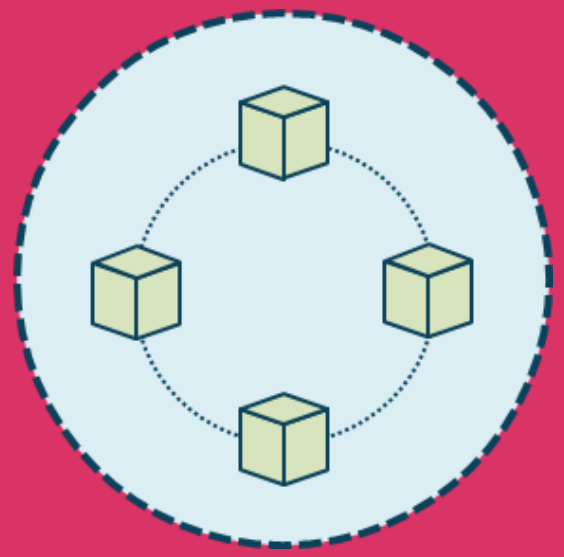
- The results of ML and AI are applied in a myriad number of ways without direct input from programmers or users
- There are a wide variety of cloud services that use machine learning algorithms and services:
  - Security automation
  - Language services
  - Intelligent contact centers and personalization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



# Artificial Intelligence (AI)

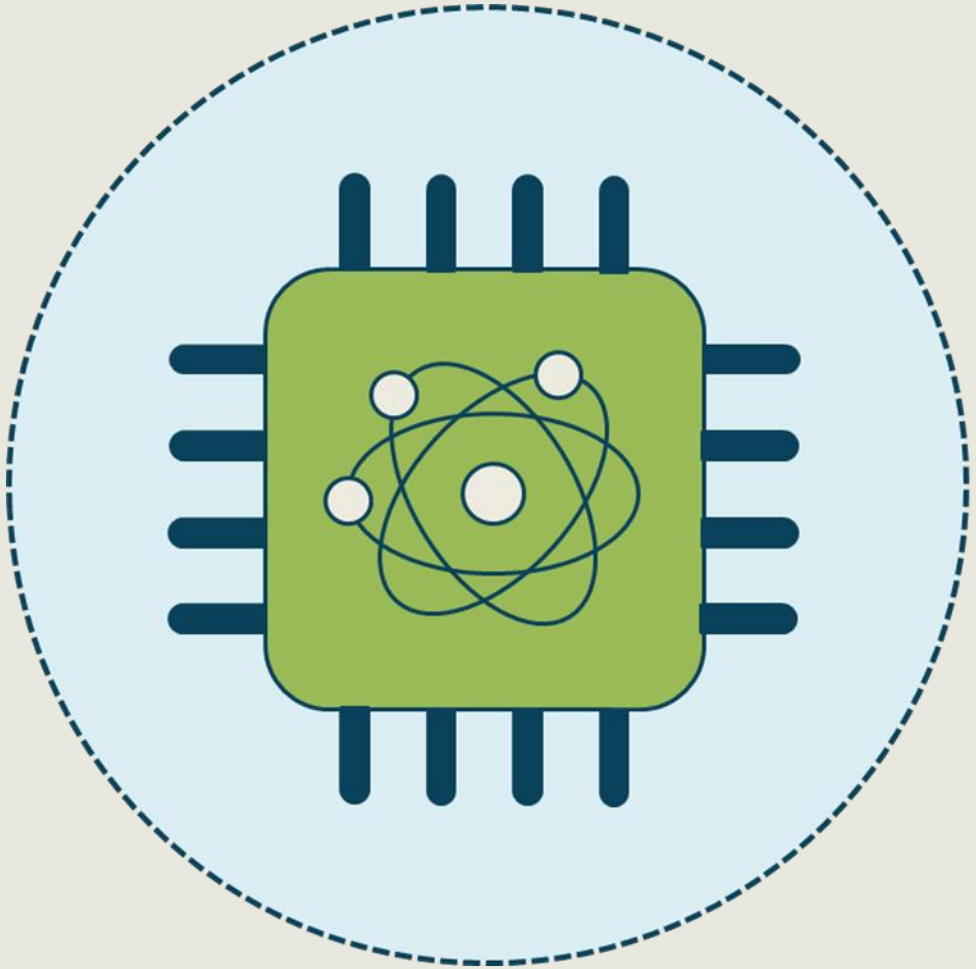
- The goal of Cloud AI is to create a virtual cloud space to emulate the human brain
- AI cloud delivers AI software-as-a-service offering enterprises access to AI tools and enabling them to harness AI capabilities
- AI can automate complex and repetitive tasks to boost productivity, as well as perform data analysis without any human intervention
- IT teams can also use AI to manage and monitor core workflows

# 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

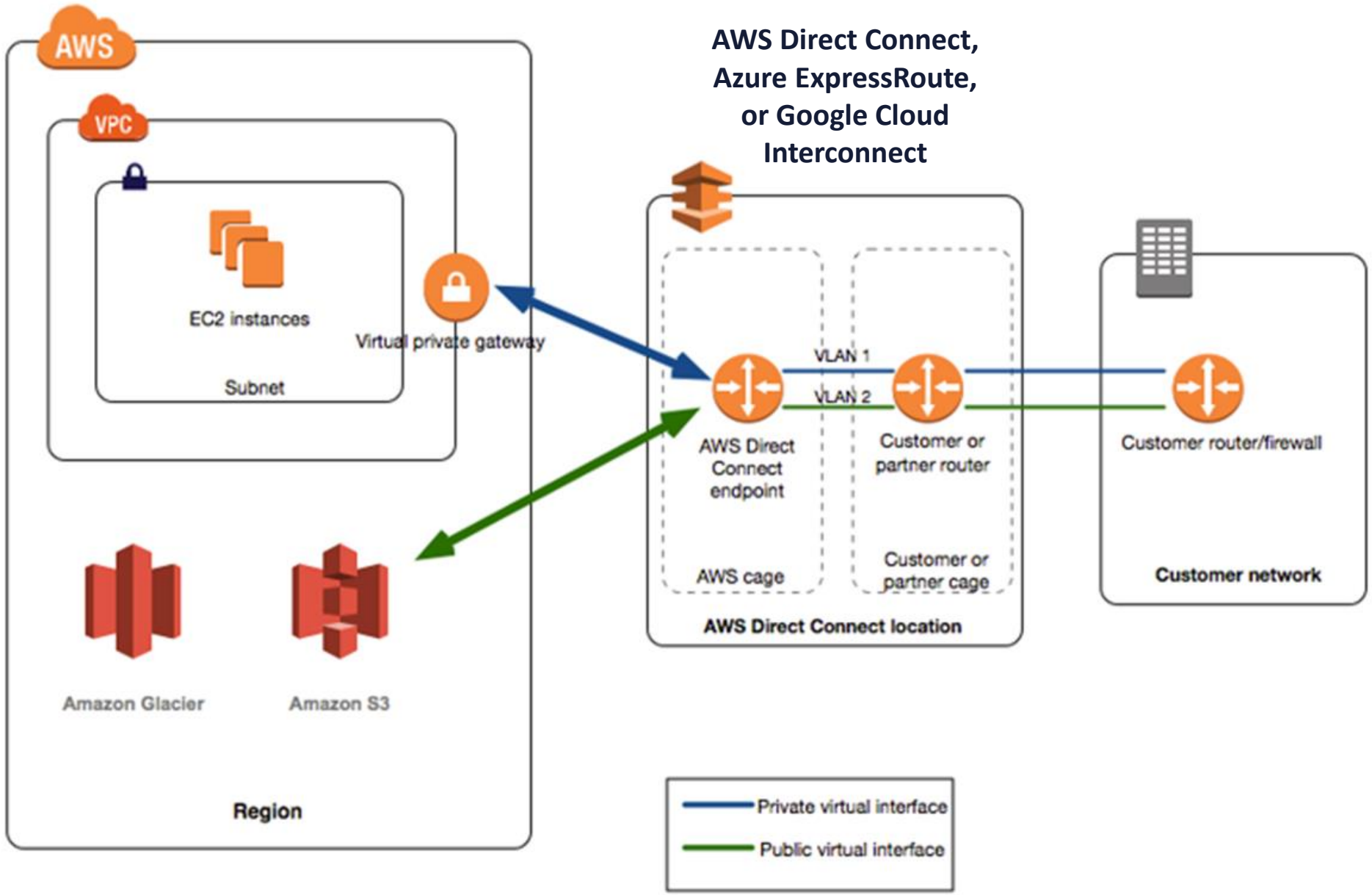


- Personal computers use bits (1s or 0s) whereas quantum computers use qubits
- These are typically subatomic particles like electrons or photons
- Quantum computing derives its power from the fact that qubits can represent numerous possible combinations of 1 and 0 at the same time
- This ability to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

# Edge Computing

- Edge computing is a distributed computing standard that brings compute services and data storage close to the site where it is needed to speed up response times and preserve bandwidth
- CDN solutions place cached versions of content (often in elastic Redis in-memory storage clusters) at metro area edge locations
- Security is a shared responsibility model between customers, ISPs, and CDN providers
- **Edge computing can also be characterized by using direct solutions like AWS Direct Connect, Azure ExpressRoute, and Google Interconnect**

# Edge Computing



# Confidential Computing

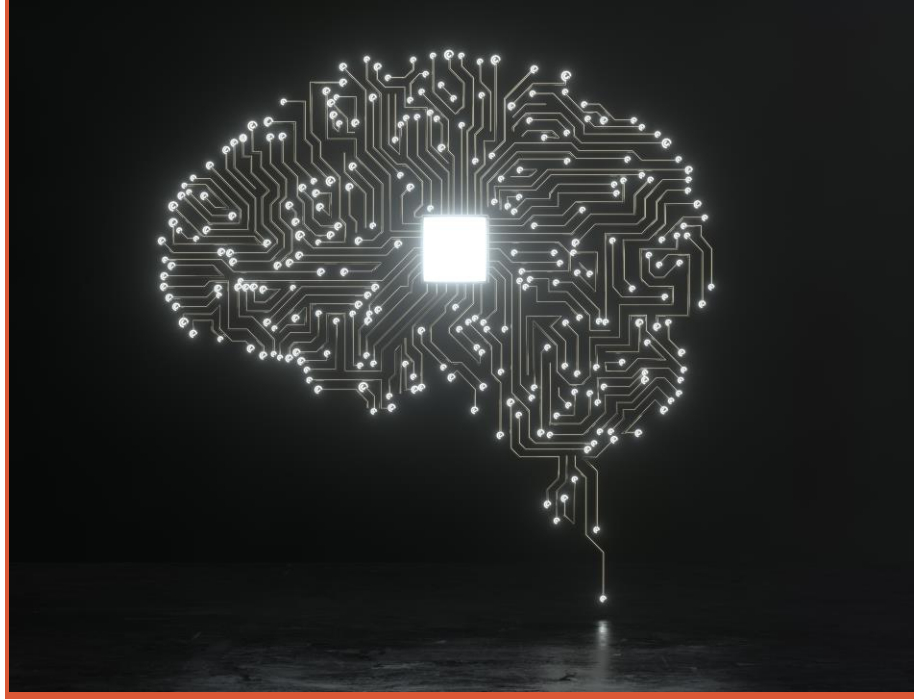
- Confidential Computing is a Private Cloud built within a Public Cloud Infrastructure
- Applications, data, and workloads within a Confidential Cloud deployment are protected by a blend of hardware-grade encryption, memory isolation, and other services that assure workload, data, and platform integrity
- Confidential Clouds are typically created on-demand at runtime and the workloads and data function completely masked from insiders, bad actors, and malicious processes
- All aspects of a workload are secure even in the event of a physical host breach

# 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)

# DevSecOps and Security

- DevOps is a clipped compound referring to a set of practices that accentuate the collaboration and communication of both software developers and IT professionals to improve and automate the software delivery process
- Can construct software quickly by linking development and operations considering application and infrastructure security from the start and automating some security gates to keep the DevOps workflow from slowing down
- Choosing the right tools to integrate security continuously is critical
- Example: AWS Dev = Elastic Beanstalk, Elastic Container or Kubernetes services; AWS Ops = CloudFront or API Gateway: AWS Sec = WAF, Shield and GuardDuty



# Internet of Things (IoT)

- IoT can be characterized as a long-term initiative to map everything on the global Internet to at least one IP version 6 address
- The explosion of Internet of Things (IoT) and Internet of Everything (IoE) presents a challenge for embedding computing vulnerability discovery
- Many systems are often powered by specialized chips or system-on-a-chip as well as some older unpatched version of Linux or Microsoft Windows

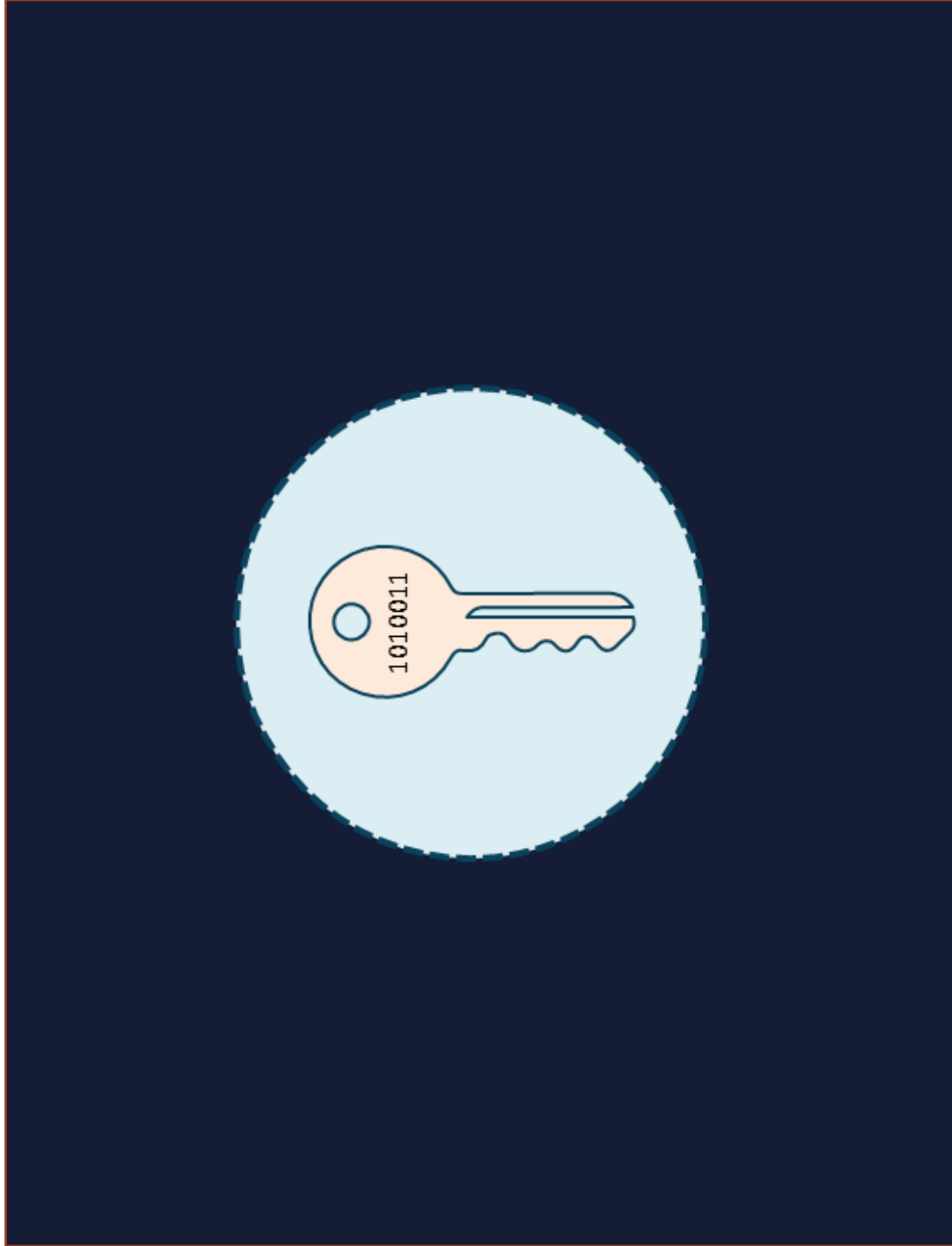
# Cryptographic Services

## Protecting data at rest and in transit

- **Confidentiality**
  - Hiding the data at rest, in transit, and in use from unauthorized entities
  - Often involves a system that converts plaintext data into ciphertext
- **Integrity**
  - Ensures the data has not been altered while at rest or in transit
  - Often involves attaching a cryptographic hash digest to the data
- **Authenticity**
  - Cryptographic controls that involve origin authentication and identity
- **Non-repudiation**
  - Ensures original sender cannot deny sending data or engaging in a digital transaction
  - Common to use digital certificates and digital signing

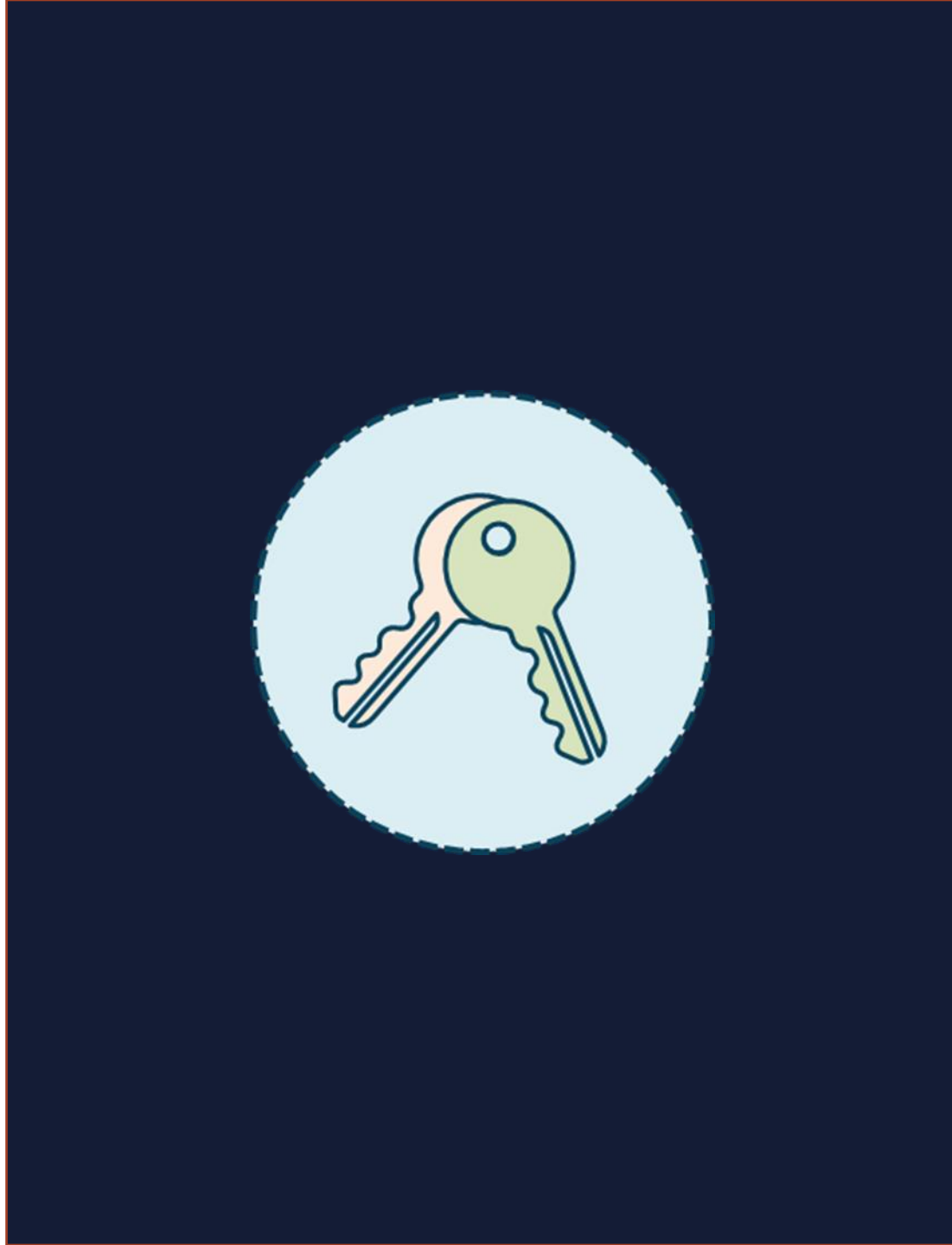
# Symmetric Key Cryptosystems

- Uses the same key to encrypt and decrypt
- Efficient, fast, computationally inexpensive, and handles high data rates of throughput
- Uses shorter key lengths (40 to 512 bits)
- Key management is more complex and does not scale well
- At a CSP, commonly used to protect storage and database data at rest as well as for TLS session keys for data in transit



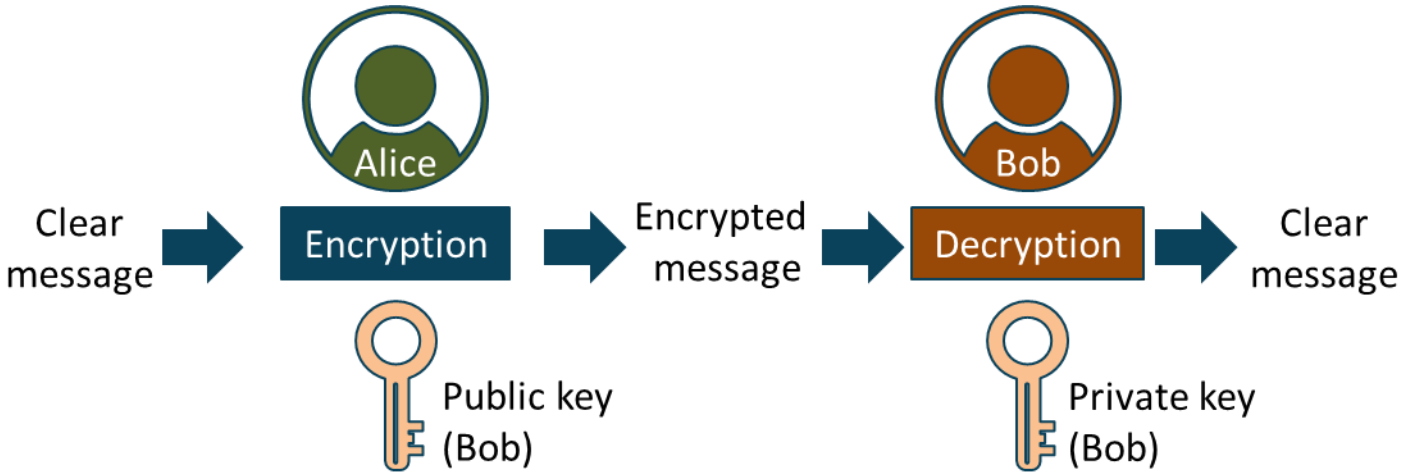
# Asymmetric Key Cryptosystems

- Uses a mathematically related pair of a public and private key - if one is used to encrypt - the other is used to decrypt
- Allows for efficient key management
- Highly scalable with a public key infrastructure
- Great for digital signatures and key exchange
- Employs longer key lengths than symmetric
- Slower and more computationally expensive
- At a CSP, KMS can be used to generate key pairs for:
  - Encrypting and decrypting
  - Digitally signing

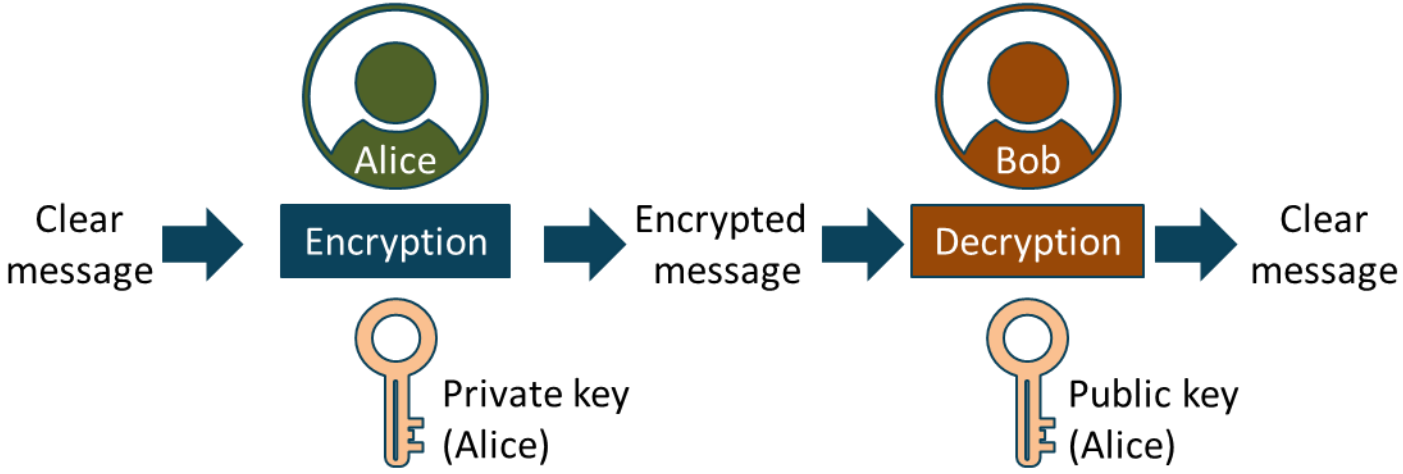


# Asymmetric Key Cryptosystems

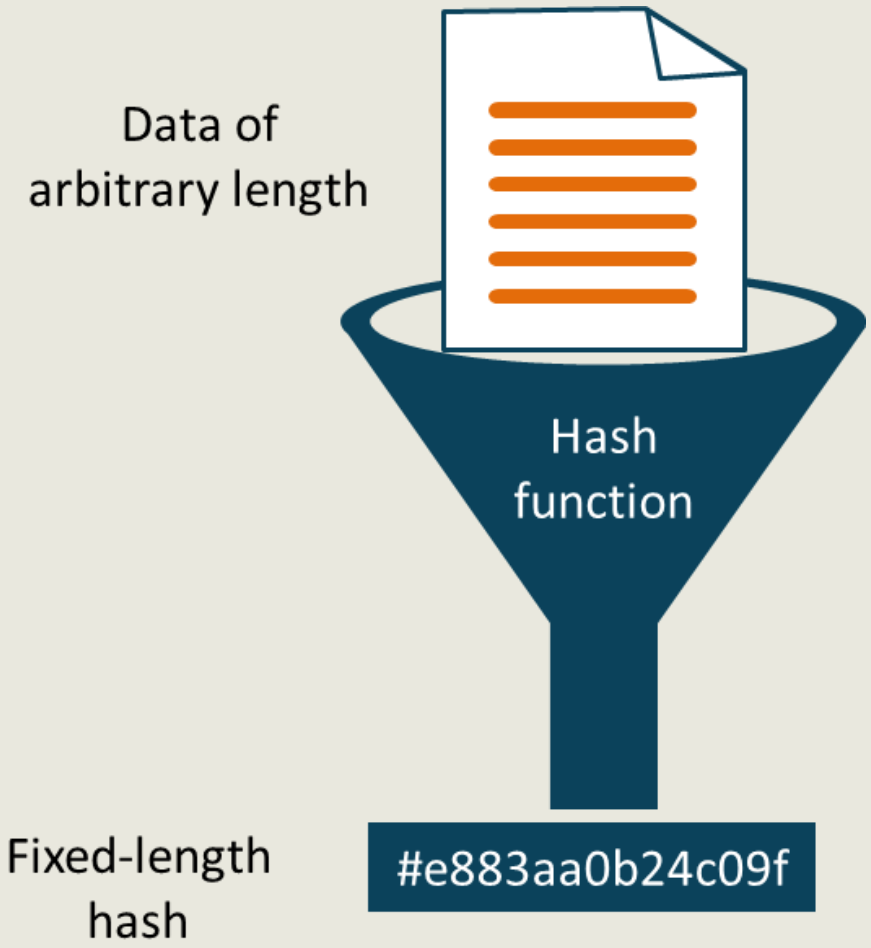
- Confidentiality
  - Encrypt with public key
  - Decrypt with private key




- Origin authentication
  - Encrypt with private key
  - Decrypt with public key

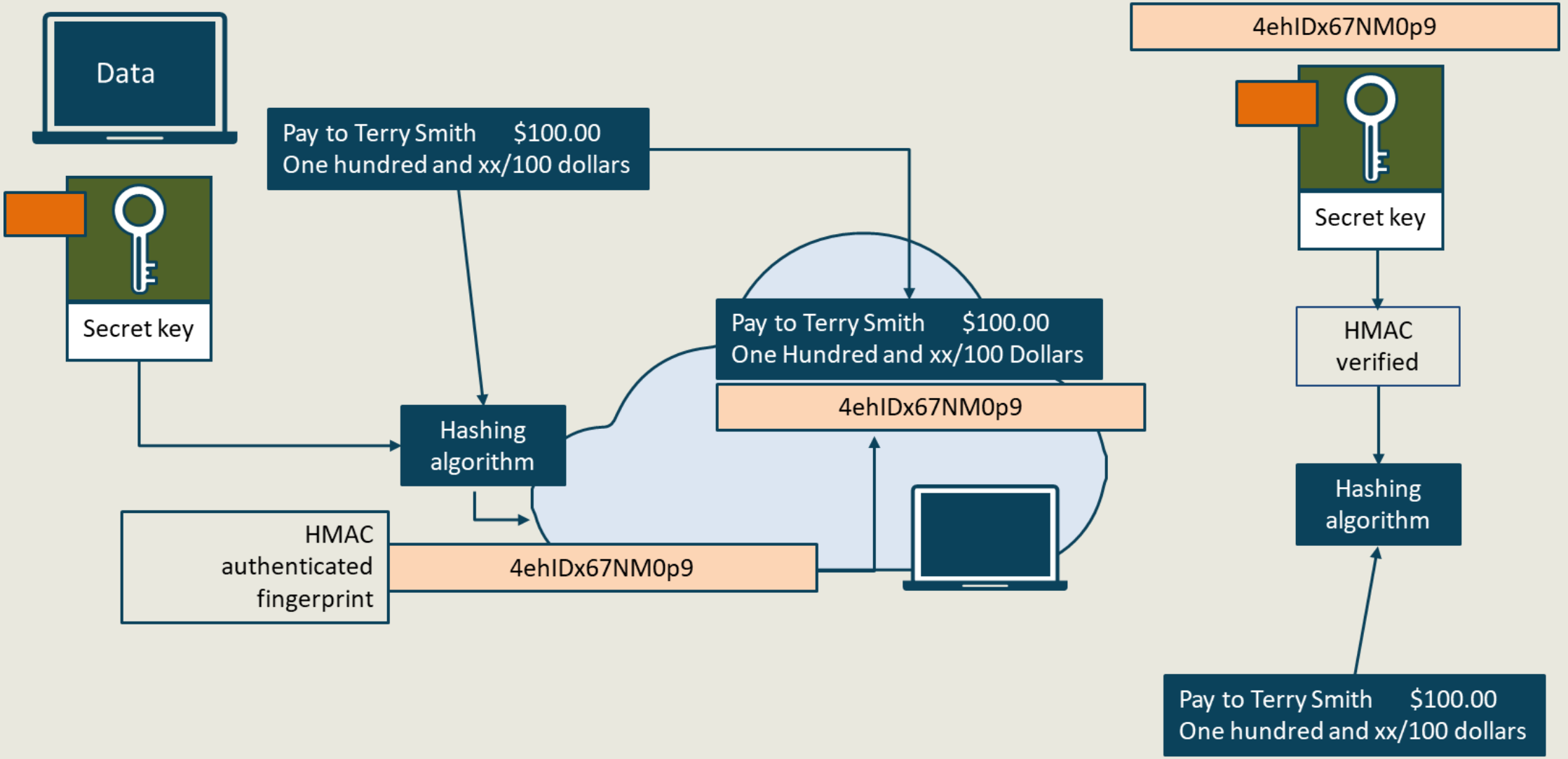


# Cryptographic Hashing



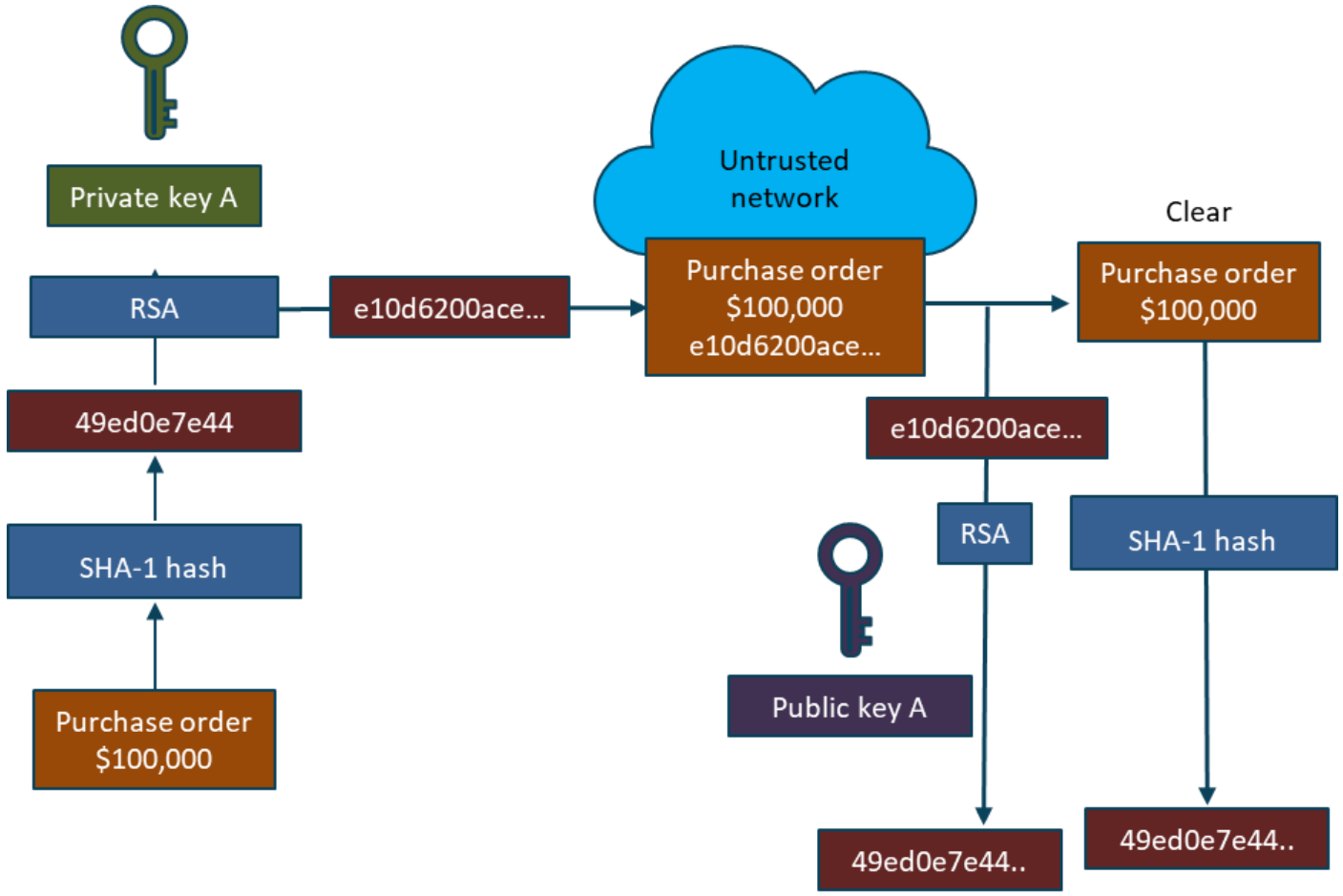

- Converts data of any input size to a fixed-length string called a hash value, message digest, or fingerprint
- An advanced version of a simple checksum
- A one-way mathematical function that produces a digest of 128 to 512 bit
- Birthday paradox and avalanche effect
- Used in authentication, data integrity, non-repudiation, fingerprinting, and password storage

# HMAC for Origin Authentication and Integrity



# Digital Signatures

- Scalable mechanism for providing authenticity, integrity, and non-repudiation
- Does not offer confidentiality
- Equivalent to a handwritten signature in many countries
- Random private/public key pair
- SHA1/2/3 hash algorithm
- Signing algorithms
  - RSA (Rivest, Shamir, Adelman)
  - DSA (digital signature algorithm)
  - ECDSA



# Key Management

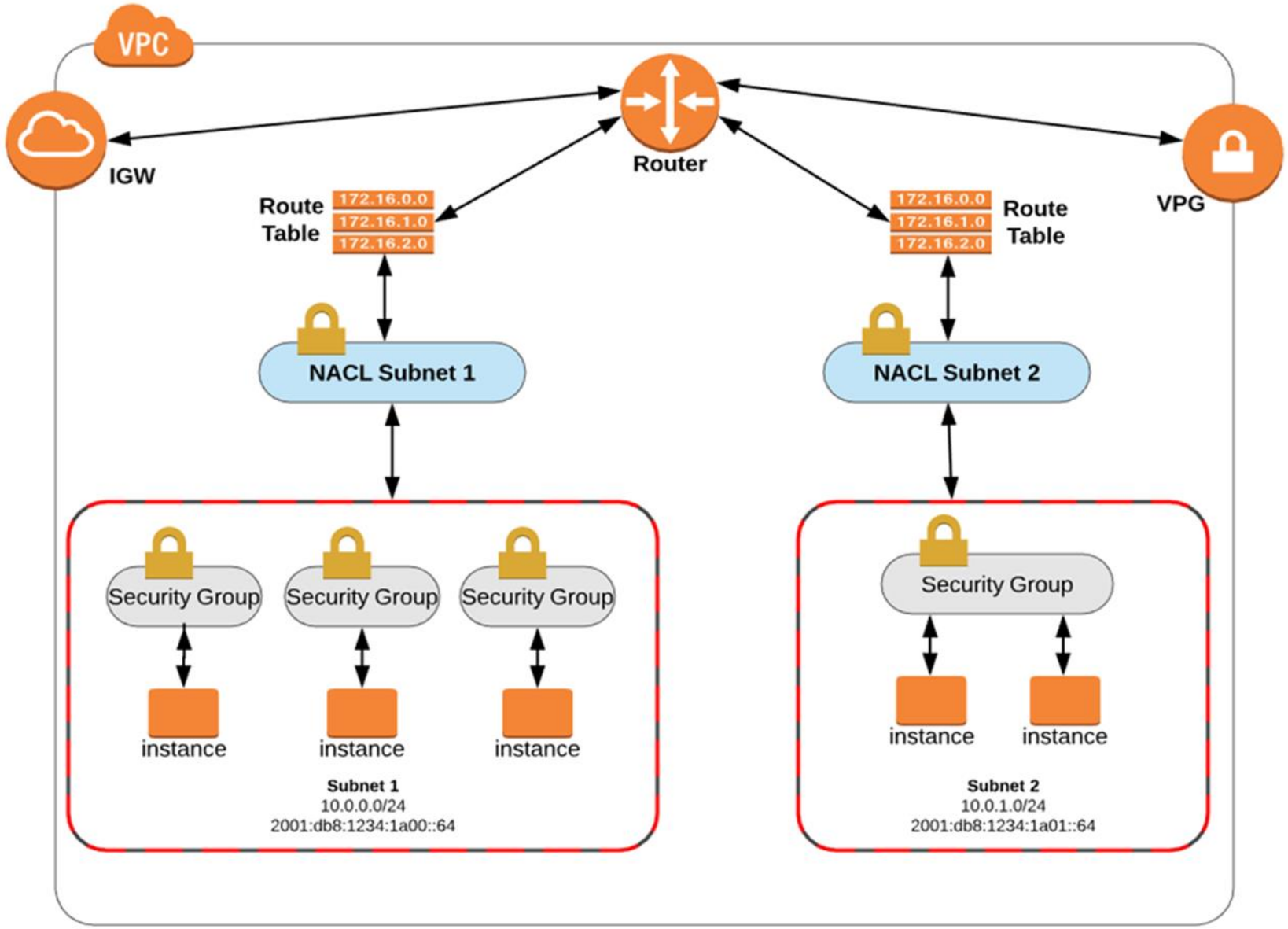
- **CSP key management is client-side, server-side, or KMS**
- Only authorized persons should be involved in the life cycle
- Long-term storage is often done with HSM or CloudHSM
- Removing keys from operation:
  - Destruction - removes an instance of a key in one of the permissible key forms at a specific location
  - Deletion – also removes an instance of a key, plus any data from which the key may be reconstructed, from its operational storage/use location
  - Termination - All instances and information of the key are completely removed from all locations, making it impossible to regenerate or reconstruct the key
  - **Crypto-shredding (cryptographic erasure) - is the singular pragmatic option for disposal of keys and data/media in the cloud**

# Access Control

## Combines Authentication, Authorization, and Accounting

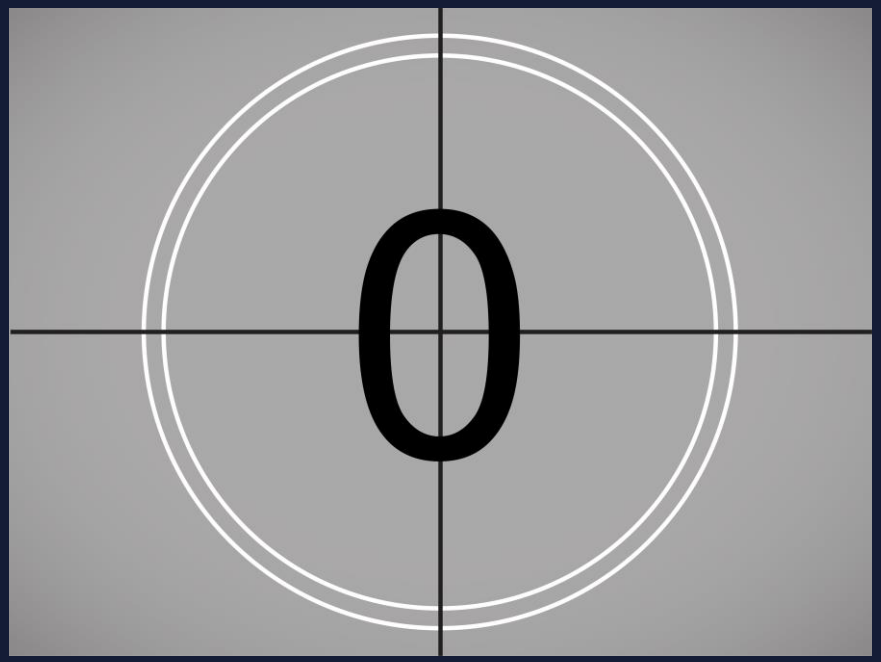
- Usernames (username or email) and passwords
- Optional MFA
  - OTP card, YubiKey, software using Google Authenticator
  - Biometric fingerprint or retinal scan
- Access key ID + access key for programmatic access
- Federated access and single-sign-on
  - SAML.2.0
  - Oauth + OIDC
  - Shibboleth

# Customer Network Security

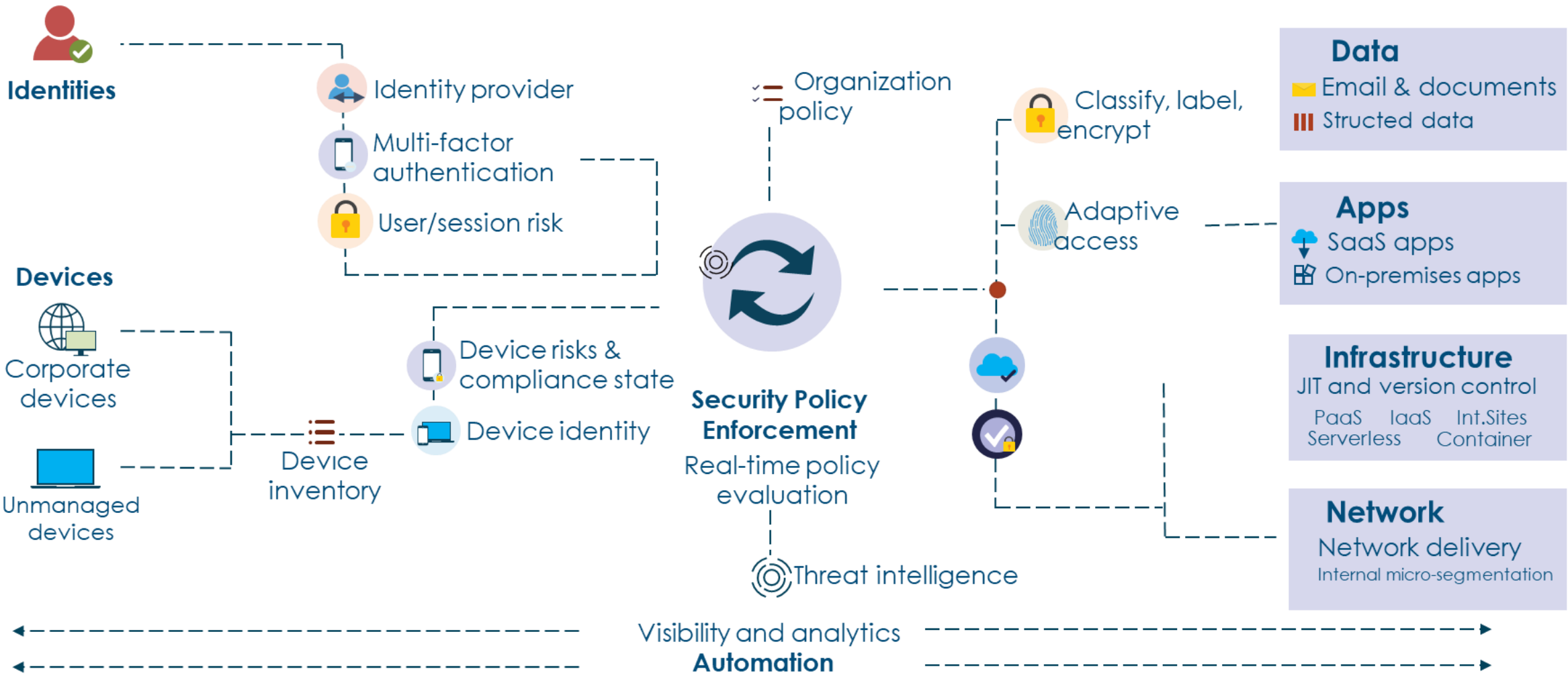


# Zero Trust

- Is an evolving paradigm moving focus to users, assets, and resources and away from a hardened corporate edge
- Uses zero trust principles to design industrial and enterprise infrastructure and workflows
- Assumes no implicit trust given to subjects based merely on their physical or network location – abandons “Trust but Verify”
- Performs authentication and authorization as distinct tasks before a session is established
- Performs authentication and authorization as distinct tasks before a session is established



# Zero Trust Enterprise

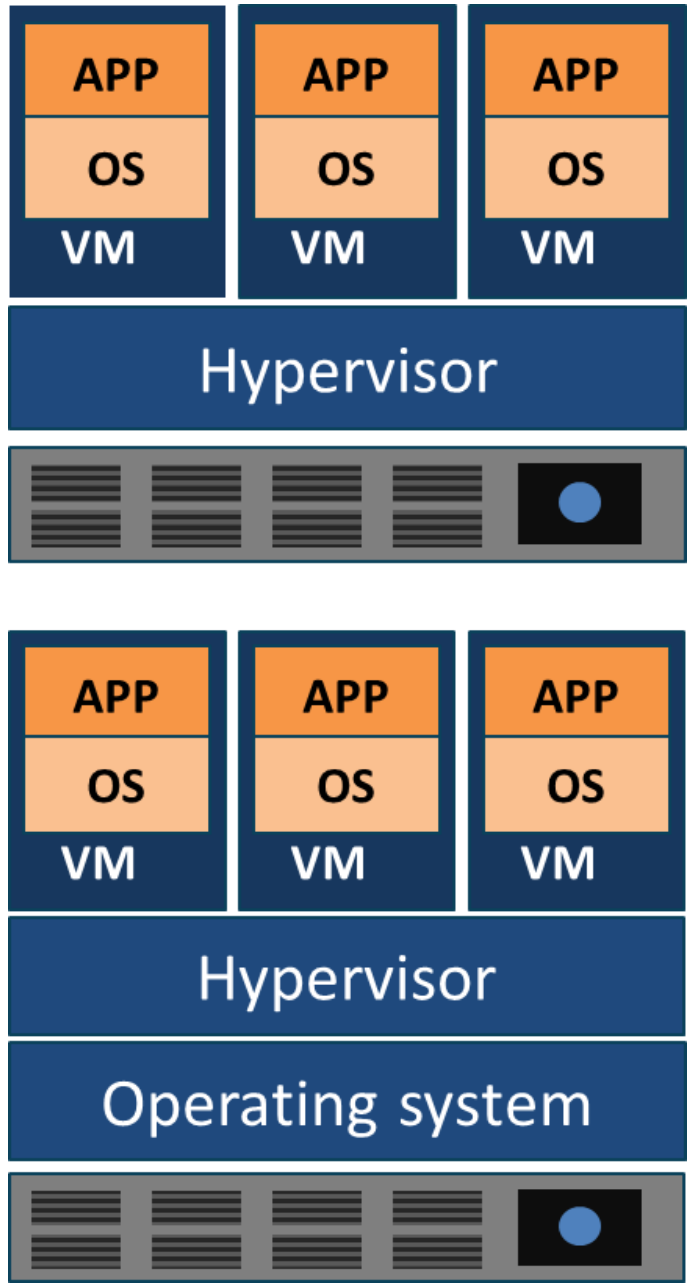


# Virtualization (VMs)

- Virtualization is the process of running a virtual instance of a computer system in a layer abstracted from the underlying hardware server
- It most often refers to running multiple operating systems on a computer system simultaneously
- To the applications running on top of the virtualized machine, it can seem as if they are on their own dedicated operating system with libraries, DLLs, and associated programs

# Hypervisors

- Software that runs virtual machines
- Controls interaction between the VMs and the hardware
- **Type I - bare metal or native**
  - Runs directly on the underlying hardware
  - XenServer, KVM, Hyper-V, ESXi
- **Type II - hosted**
  - Runs on the OS installed on the hardware
  - Oracle VirtualBox 6, VMWare Player/Workstation



# Virtualization Considerations

- **Ephemeral Computing**
  - Ephemeral computing is the process of creating a virtual computing environment on an ad-hoc temporary basis and then disposing of the environment when necessary or the resources are no longer in demand
  - The consumer only pays for what is used
  - Examples would be functions-as-a-service with AWS Lambda and Azure Functions

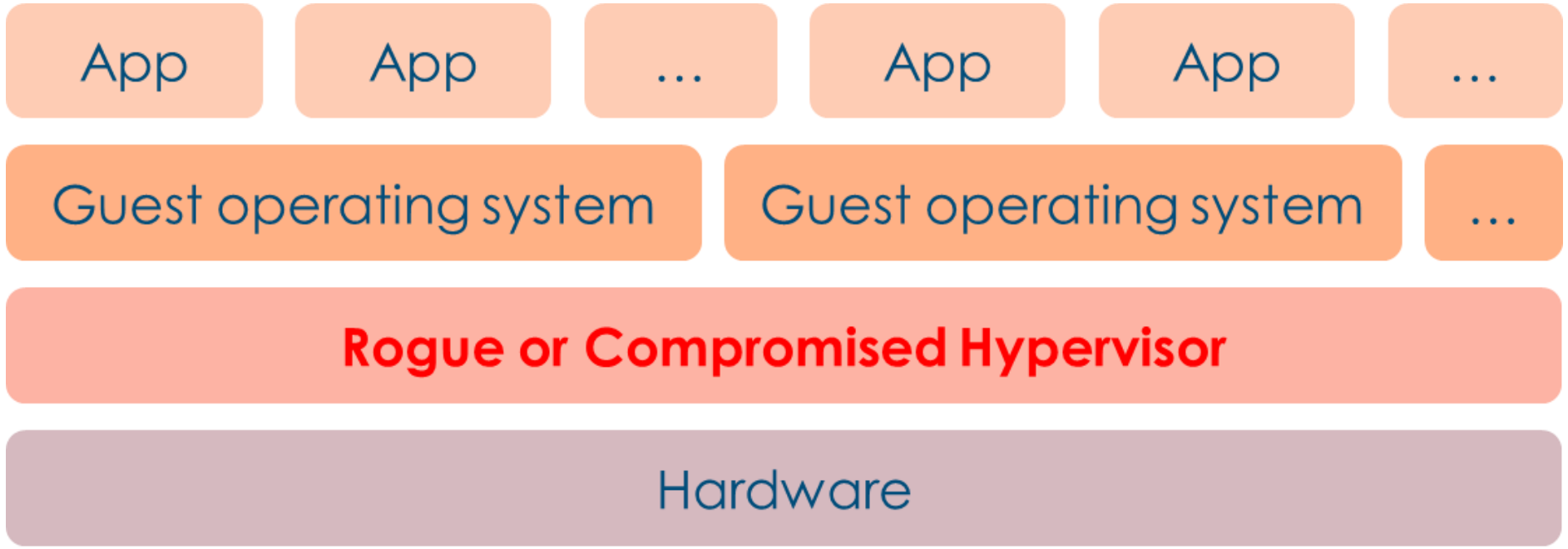
- **Serverless Technology**
  - Functions are a form of serverless technology
  - These are technologies for running code, managing data, and integrating applications, all without managing Windows, Linux, and MacOS servers
  - Serverless technologies feature automatic scaling, built-in high availability, and a pay-for-use billing model to increase agility and optimize costs

# Virtualization Vulnerabilities

- **VM sprawl**
  - When the number of VMs overtakes the administrator's ability to manage them and the available resources
- VM sprawl avoidance
  - Enforce a strict process for deploying VMs
  - Have a library of standard VM images
  - Archive or recycle under-utilized VMs
  - Use a Virtual Machine lifecycle management tool or a cloud service provider-managed service
- **VM escape**
  - A serious threat where a process running in the guest VM interacts directly with the host OS
- VM escape protection
  - Patch VMs and VM software regularly
  - Only install what you need on the host and the VMs
  - Install verified and trusted applications only
  - Strong access control policies and passwords

# Virtualization Vulnerabilities

- **Hyperjacking**
  - A hyperjacking attack is an attempt by an attacker to take control of the hypervisor, using a rootkit installed on a virtual machine



# Common Cloud Threats

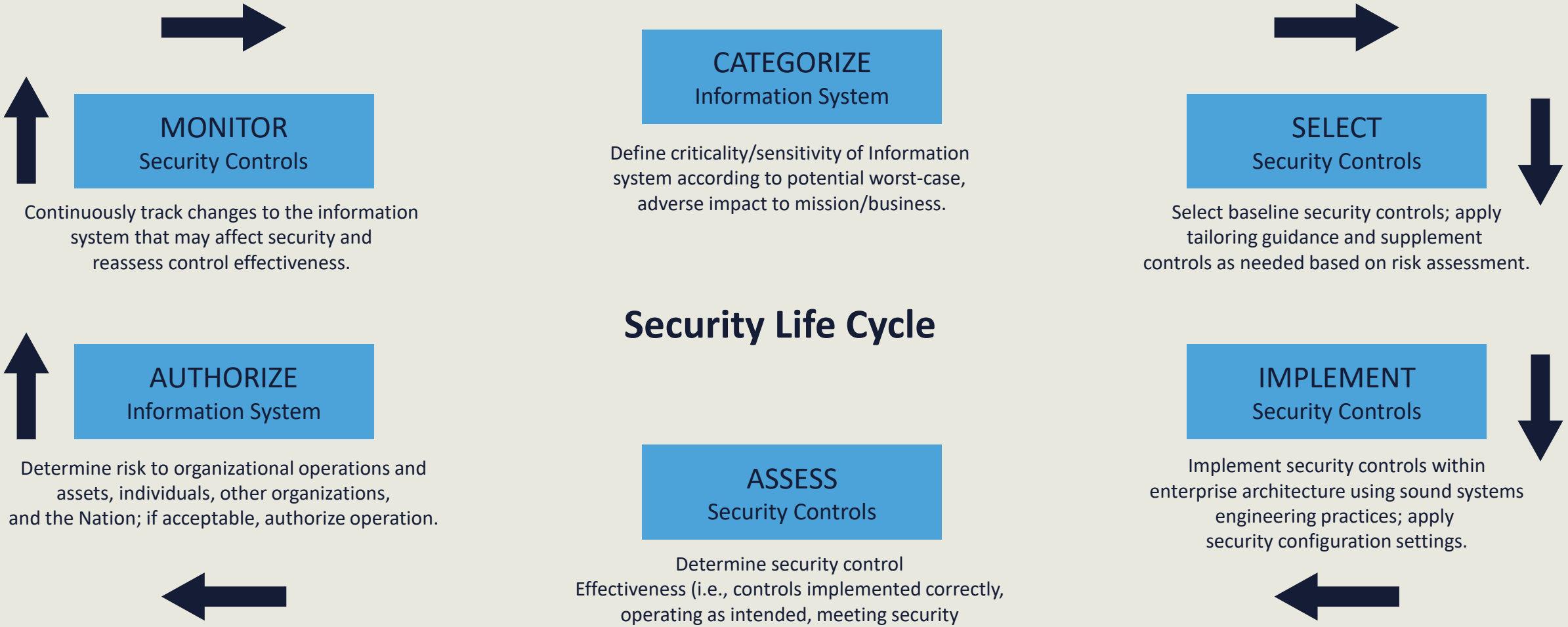
**“The Treacherous Twelve: Cloud Computing Top Threats” from the Cloud security Alliance (CSA)**

1. Data breaches
2. Insufficient identity, credential and access management
3. Insecure interfaces and APIs
4. System vulnerabilities
5. Account hijacking and cryptojacking
6. Malicious insiders
7. Advanced persistent threats
8. Data loss
9. Insufficient due diligence
10. Abuse and nefarious use of cloud services
11. Denial of service (DDoS and botnets)
12. Shared technology vulnerabilities

# Security Hygiene

- Cyber hygiene refers to practices and steps such as baselining and patching cloud resources to maintain system health and improve online security
- These practices are often part of a routine to ensure the safety of identity and other data that could be stolen or corrupted
- Much like physical hygiene, cyber hygiene is frequently and often automatically performed to mitigate against data and system deterioration and common internal and external threats
- Initiatives such as Center for Internet Security Top 18 are excellent foundations for security hygiene

# NIST SECURITY LIFECYCLE (SP 800-37)



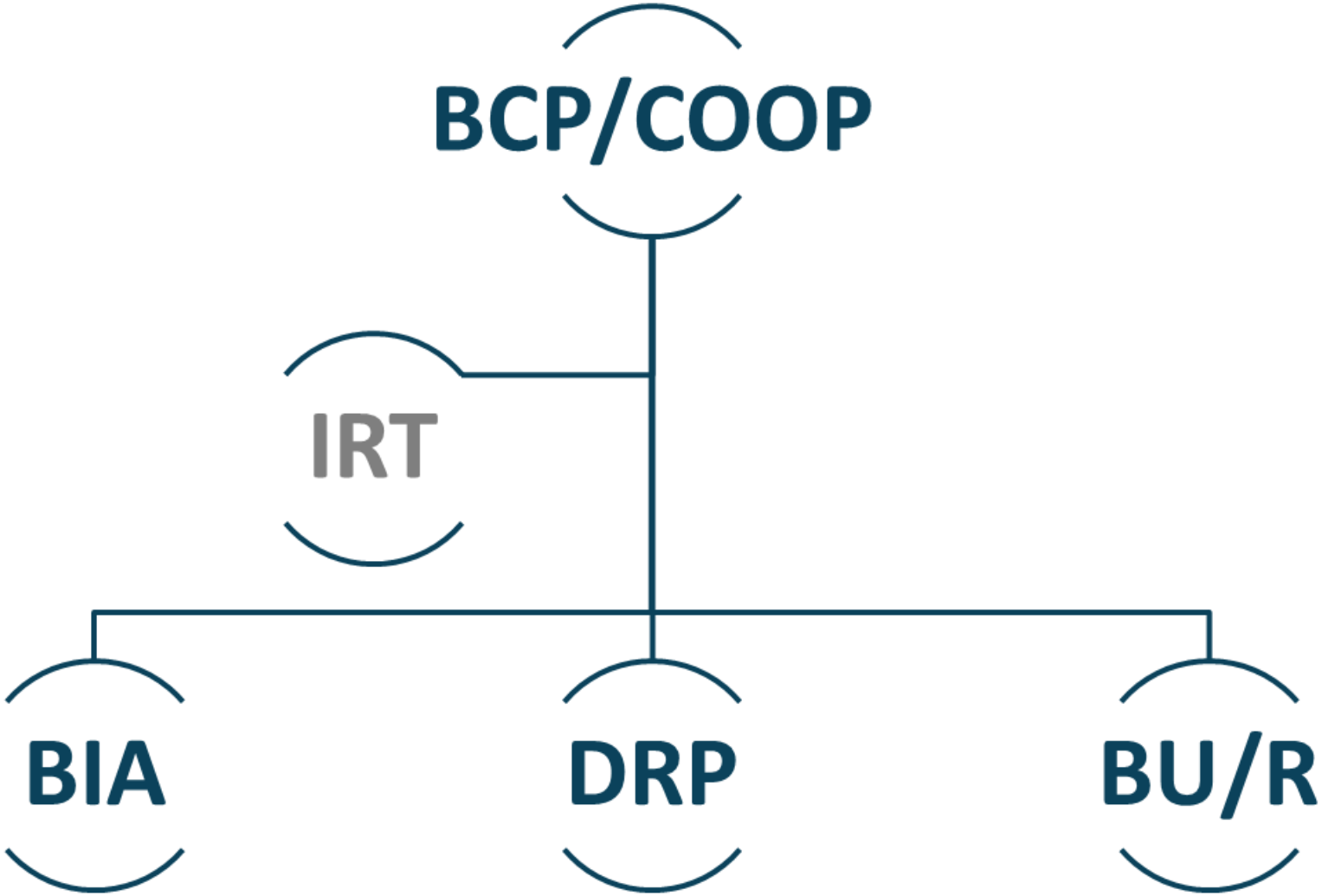
# Cloud based Disaster Recovery (DR) and Business Continuity (BC) planning

## There are two aspects to consider with BCP and DR:

1. The continuity of operations and disaster recovery responsibilities of the cloud provider and

2. The customer using the CSP for their own business continuity, site resiliency, and disaster recovery

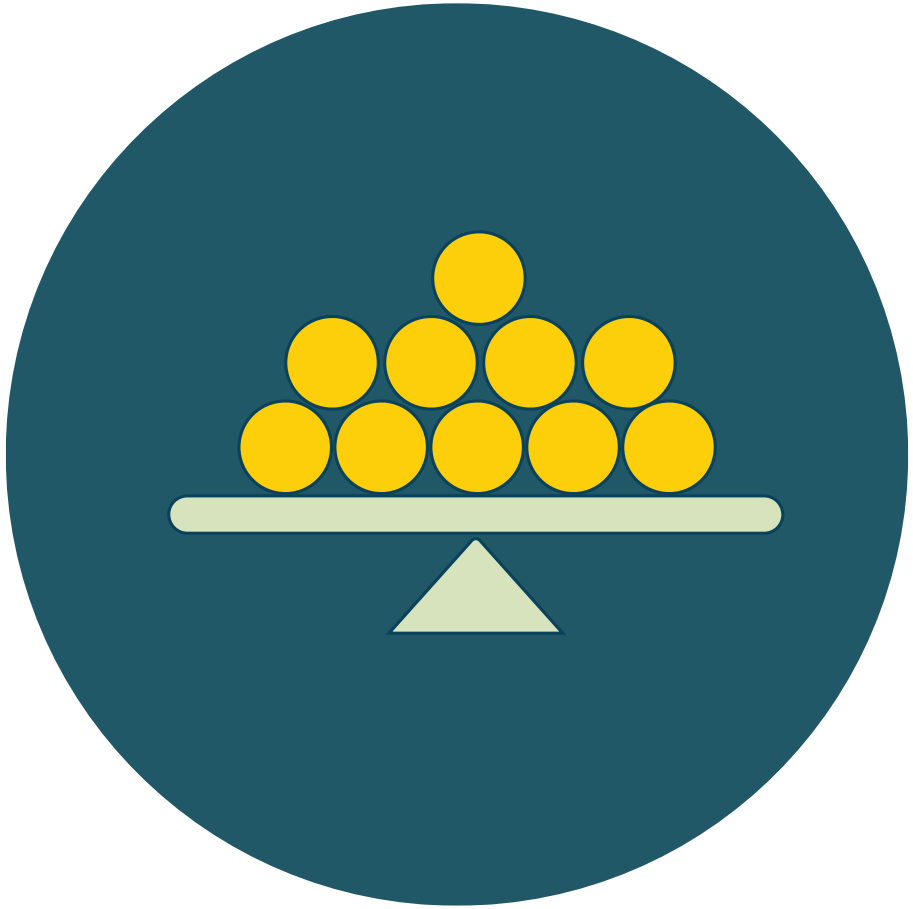
# Business Continuity Planning (BCP/COOP)



# Cost Benefit Analysis

**This should be conducted early in the security management lifecycle**

- After identifying all possible controls and evaluating their feasibility and effectiveness, enterprises should allocate resources and implement cost-effective controls based on cost-benefit analysis
- The CBA can be qualitative or quantitative based on risk management methodologies
- The goal is to demonstrate that the costs of implementing the controls can be justified by the reduction in the level of risk
  - For example, the organization may not want to spend $10,000 on a control to reduce a $2000 risk



# Cloud Return-on-Investment (ROI)

- Cloud return on investment (ROI) is a cloud economics metric of the bearing a cloud investment has on the enterprise
- For most organizations, ROI is a critical success factor for any cloud-based initiative
- It indicates that a particular business decision (database migration, DevOps using containers, content distribution networking, edge computing, etc.) resulted in a positive effect on the organization's bottom line

- Critical factors when calculating ROI according to VMware:
  - Productivity
  - Leverage
  - Pay as you go and need
  - Provisioning time
  - Reduction of capital spending
  - Access to new markets and regions
  - Cloud risk management

# Functional Security Requirements

- Portability (cloud-based) - is the ability to move applications, containers, code, and associated data from one CSP to another or between legacy on-prem environments and the cloud
- Interoperability – when the customer application does not function properly when changes are made to an environment; more common in a PaaS scenario since the O/S is managed and updated by the provider
- **Vendor lock-in** – a scenario where the consumer is unable to retain, migrate, or transfer to another provider based on technical or non-technical restrictions
- **Vendor lock-out** – A situation where the customer is not able to recover or access their own data due to a vendor going out of business, going bankrupt, or legal holds are implemented

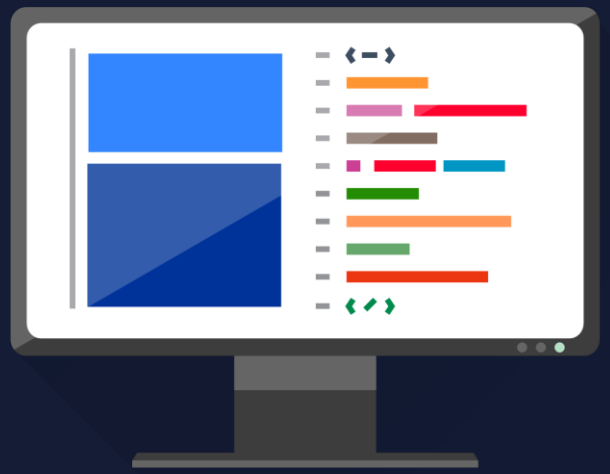
# Security Considerations for IaaS

- **The cloud customer has the highest degree of responsibility for security:**
  - Network design and firewalls
  - Operating system and applications
  - Data encryption (client-side vs. server-side)
  - Data protection
- CSP will secure the facility and datacenter – everything at Layers 1 and 2 of the OSI model including the hypervisors
- Customer has limited visibility and monitoring of the provider site and datacenter which makes audits challenging – must rely on third-parties
  - Must be negotiated in early negotiations
- Some activities may be prohibited by regulations such as Cloud-based HSM instead of on-prem HSM

# Security Considerations for PaaS

- The customer loses additional control when using platform-as-a-service since the provider installs, manages, upgrades, updates the operating systems, database systems, and even some applications
- There is more gray area with managed and “fully-managed” services (example: AWS Aurora vs. RDS MySQL)
- SLA’s and policies must be more structured and customized, especially when regulatory compliance is involved
- **Exam tip: Customer must have adequate monitoring and metering when using PaaS**

# Security Considerations for SaaS



- Practically all control is assumed by the service provider
- The consumer will only create, deliver, and modify data delivered to the cloud system
- Often, the applications are being rendered via XML or web-based tools on the client system or VDI environment
- Customer has limited administrative rights, elevated privileges, permissions, and authorization
- Consumer may still have contractual responsibilities for safeguarding data and federated access therefore a Cloud Access Security Broker (CASB) is often employed to assist SaaS solutions
- **Exam Tip: SaaS is the most likely service type to experience vendor lock-in**

# Cloud Design Patterns

- SANS security principles
- AWS Well-Architected Framework
- Cloud Security Alliance (CSA) Enterprise Architecture

# 1.5 Evaluate Cloud Service Providers

- International Organization for Standardization/International Electrotechnical Commission **(ISO/IEC) 27017** offers a set of standards for provisioning controls for cloud services and cloud service customer information and privacy
- Payment Card Industry Data Security Standard **(PCI DSS)** is for entities that process major credit or bank card transactions; some cloud providers will be “PCI-compliant”

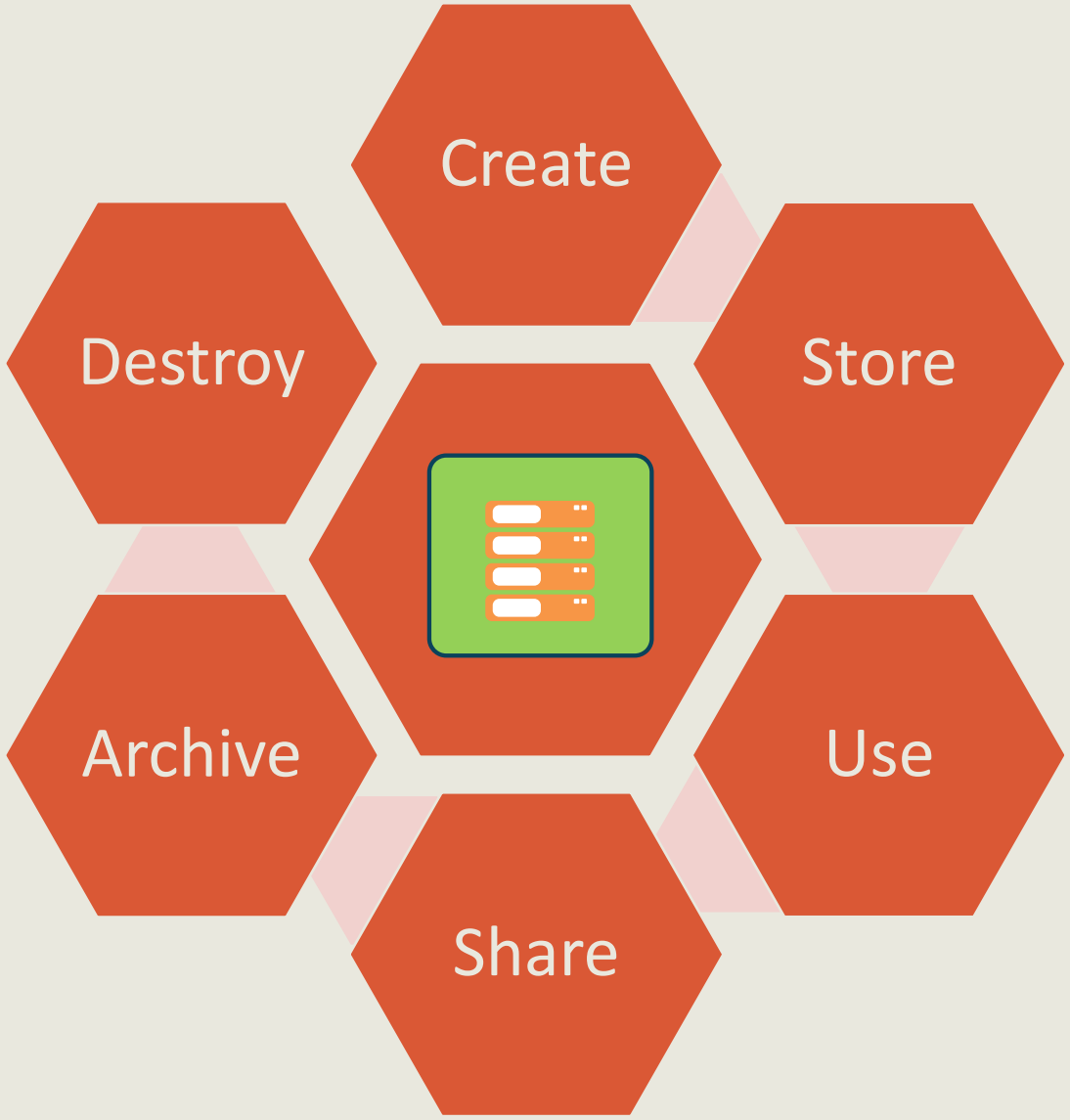
# 1.5 Evaluate Cloud Service Providers

- Common Criteria **(CC)** offers assurance and certification that the  descriptions, implementation and evaluation of a cloud provider service is performed in a rigorous and repeatable manner that adequately maps to its target use environment
- Federal Information Processing Standard **(FIPS 140-2)** is a standard that will is used by Federal organizations when cryptographic-based security systems are used to provide confidentiality, integrity, and non-repudiation for sensitive or valuable data

# Domain 2

# Cloud Data Security

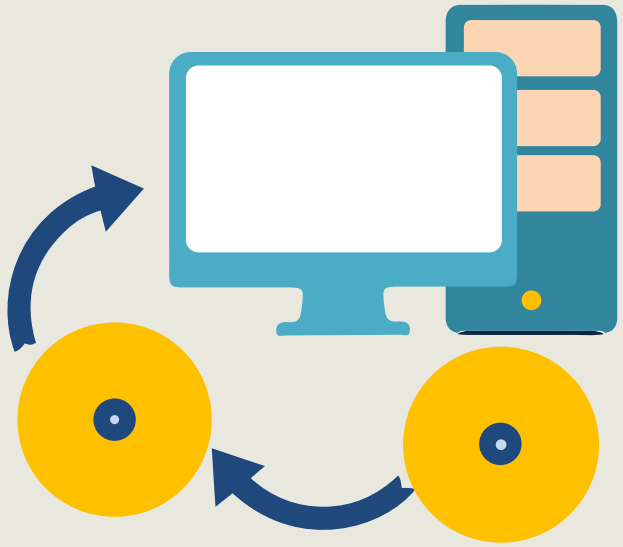
# Cloud Secure Data Lifecycle



# Cloud Data Life Cycle Phase 1: Create

- Data is either generated from scratch, inputted, or modified into another format either locally or in the cloud
- If created locally it may need to be sent over IPsec or TLS VPN (S2S or P2S) or the customer can perform client-side encryption and send over a clear channel
- The data owner is identified in the create phase
- Other key activities of phase one include:
  - Data discovery
  - Data categorization
  - Data classification
  - Data mapping
  - Data labeling (tagging)

# Cloud Data Life Cycle Phase 2: Store



- After the Create phase, the data is put into a volume (block)/object storage system or into one of several types of database systems
- This phase relates to transactional, near-term usage data as **opposed to long-term cold data storage**
- It also includes files and spreadsheets, typically done at during or at the end of the Create phase
- **Activities of this phase can also occur simultaneously when the data is generated in phase one**
- Protection of data at rest and data in transit will often occur in this phase unless default encryption is implemented in the Create phase

# Cloud Data Life Cycle Phase 3: Use

- Data is utilized by people, applications, and tools as well as being changed from the original state
- Raw data becomes information
- If data is used remotely then protection mechanisms must be in place (VPN, secure endpoints, digitally signed API calls)
- The systems that “use” the data must be secured as well; for example, endpoint detection and response (EDR) or host-based IPS agents (Palo Alto Traps)
- Technologies like VPN, Identity Rights Management (IRM), and Data Loss Prevention (DLP) engines may be introduced
- Assistance can come form a Managed Security Service Provider (MSSP) or Cloud Access Security Broker (CASB)

# Cloud Data Life Cycle Phase 4: Share

- Data is visible, analyzed, and apportioned among users, systems, and applications
- Global collaboration and sharing of data introduces obvious risks and lack of control
- Most of the control used in the previous phases will be implemented here in phase four (such as IRM and DLP services)
- Stringent Identity and Access Management (IAM) and/or Identity Management (IdM) should be used to enforce the least privilege principle in line with access control model (DAC, RBAC, MAC, ABAC, etc.)
- It may be beneficial to implement egress DLP on the email message transfer agents (MTA) to and from the cloud provider and partners using the same CSP

# Cloud Data Life Cycle Phase 5: Archive

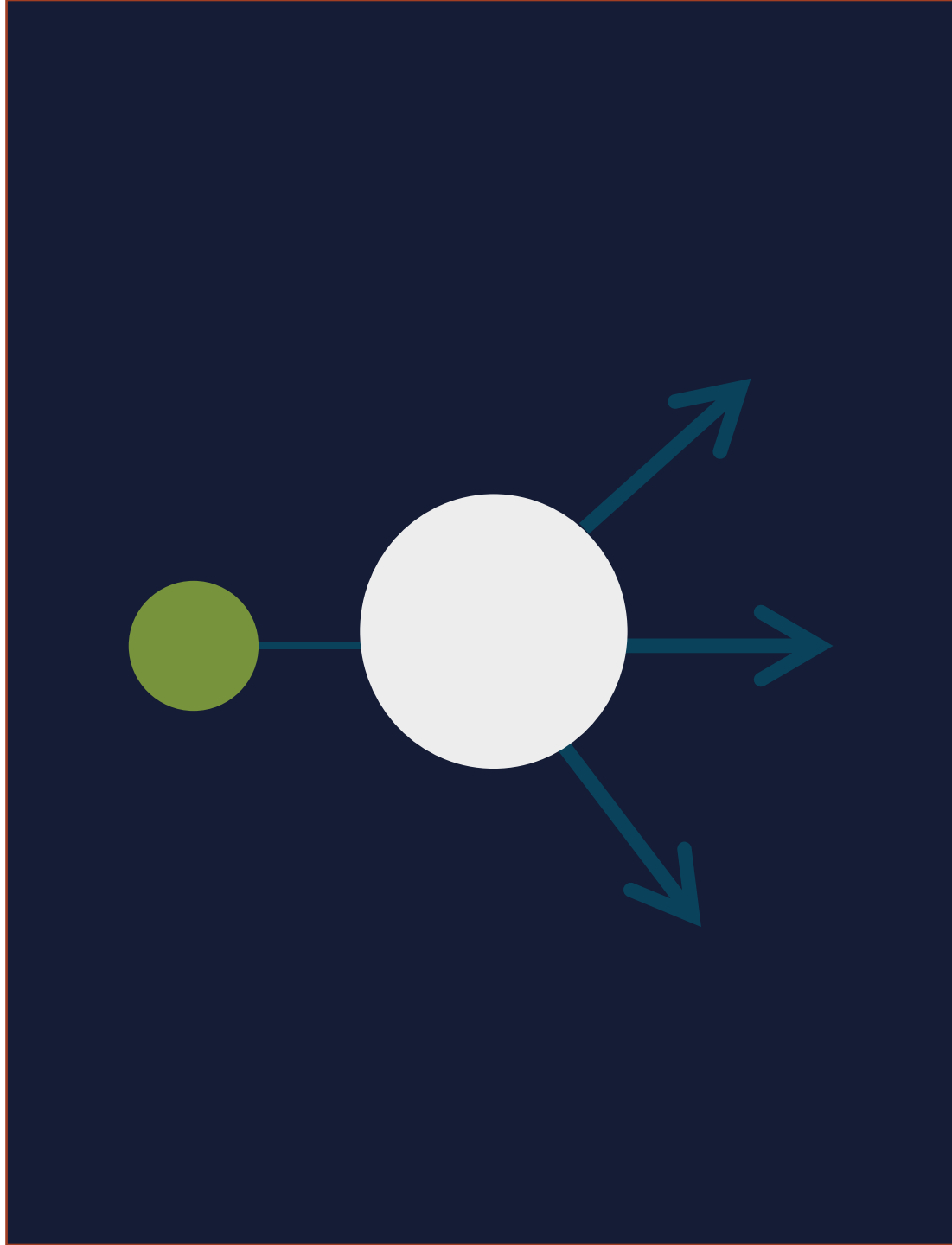
- Data is stored for long-term and removed from active usage
- It can be sanitized based on policy
- Stringent cryptography will be introduced for data at rest – as in AES-GCM-256 AEAD solutions
- Archiving is often automated and based on governance or regulations for example AWS S3 Intelligent Tiering or Storage Gateway management over a Direct Connect link
- Factors in choosing long-term storage:
  - Location
  - Media format
  - Staffing
  - Operating procedures

# Cloud Data Life Cycle Phase 6: Destroy

- Data is no longer accessible or usable based on lifetime, utility, policy, governance, and/or regulations
- **Although data can be disposed of using a variety of methods, when storing data at a CSP, cryptoshredding (cryptographic erasure) is the only practical and comprehensive solution**
- The provider will have their own established methods for disposal of data and media, often using military grade programs or physical destruction

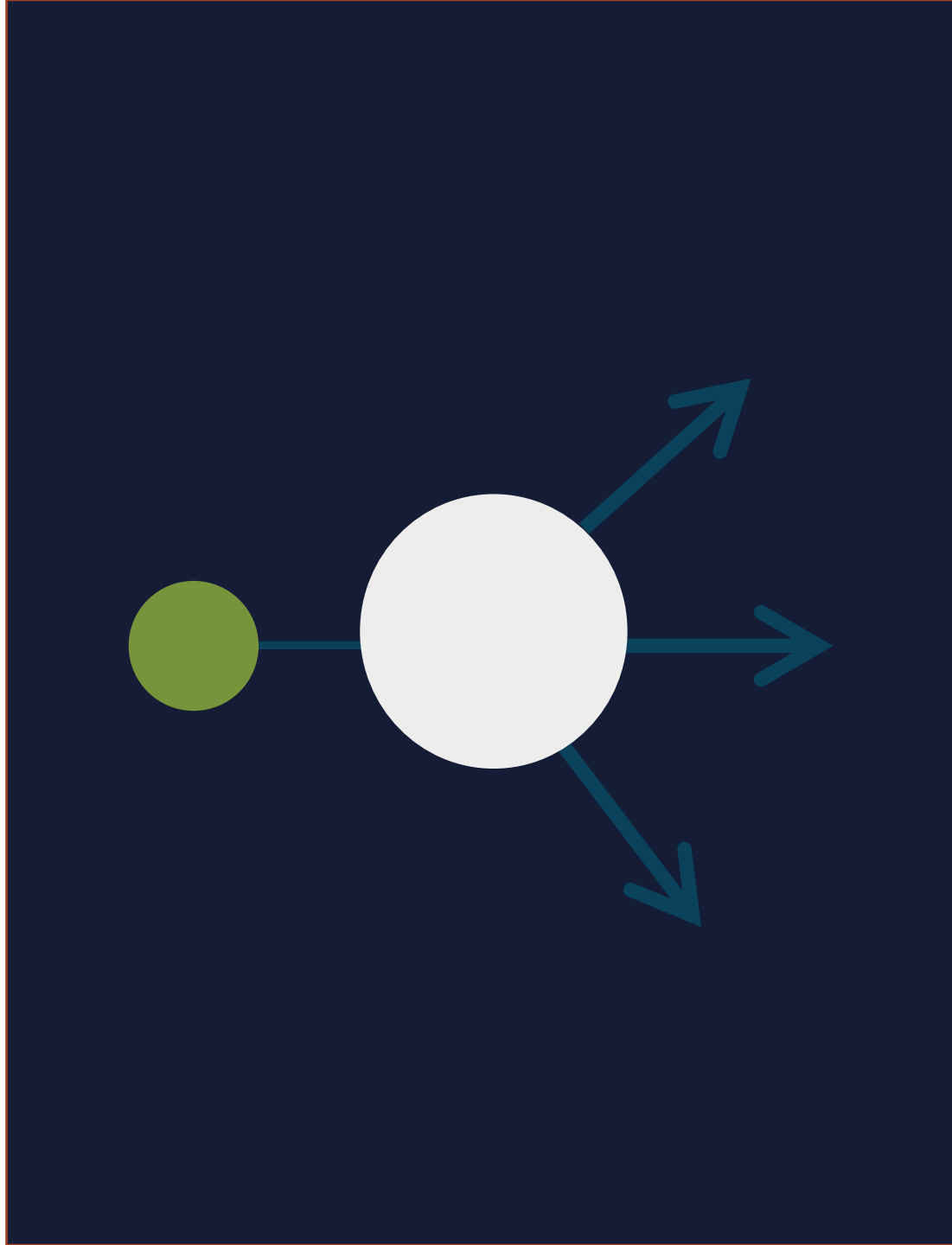
# Data Dispersion

- Data dispersion is a method that is often used to increase data security, but without the use of encryption
- Dispersion is like legacy RAID in that data is spread across different storage areas and **even different cloud providers (MULTI-CLOUD)** in disparate geographic locations
- However, if data is spread across multiple cloud providers, an outage at one could make the dataset unavailable to users, regardless of location

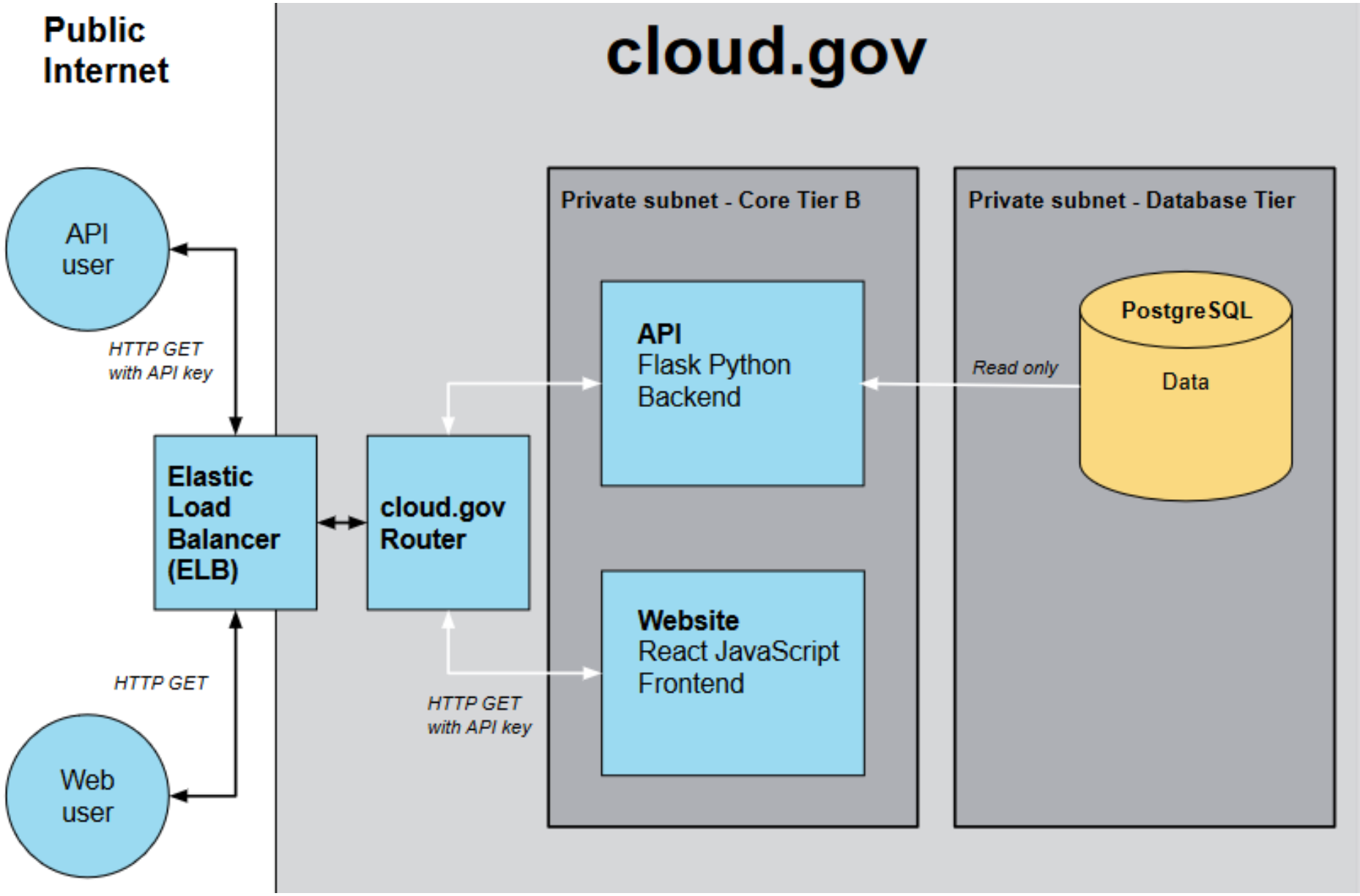


# Data Dispersion

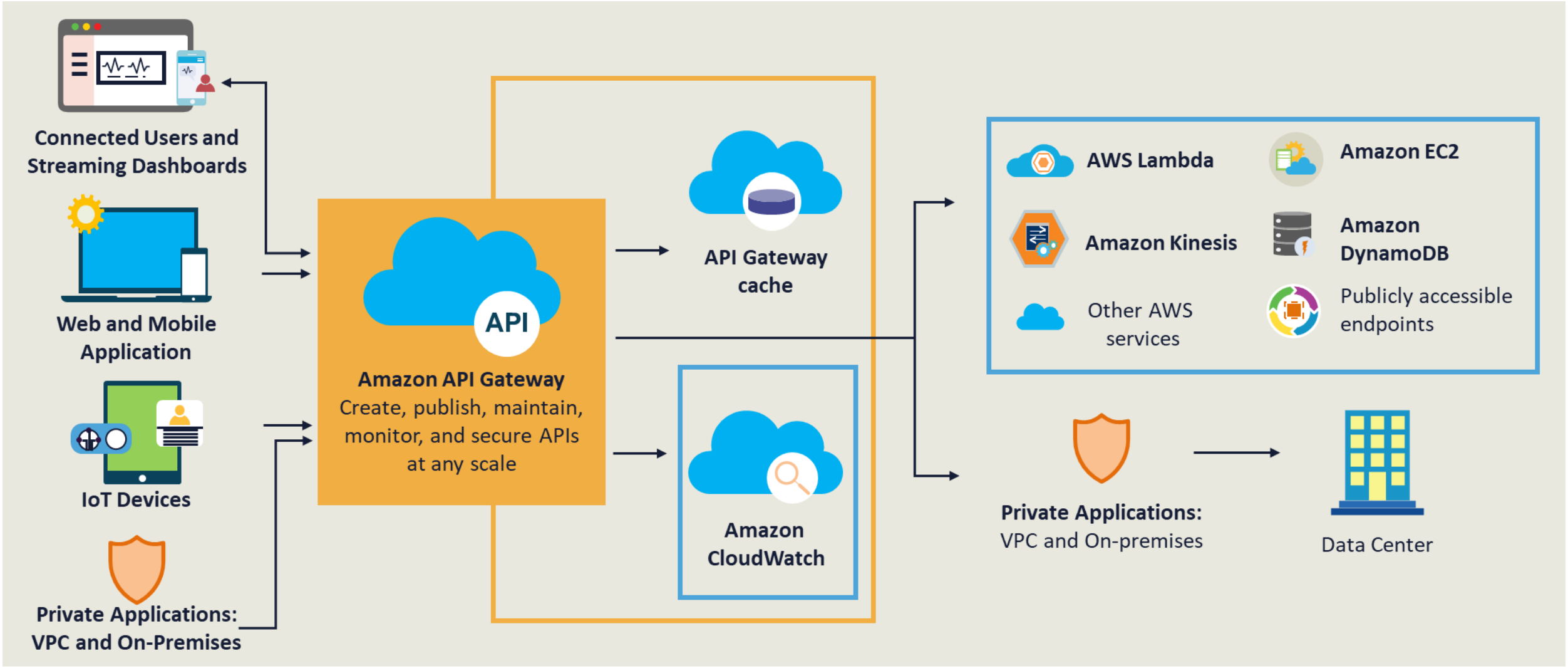
- **Erasure coding** is like using parity bits for RAID striping and helps you recover missing data if cloud data is unavailable/lost while your data is dispersed
- **Bit splitting** is like adding encryption to RAID where the data is first encrypted, then separated into chunks, and the pieces are distributed across several storage areas



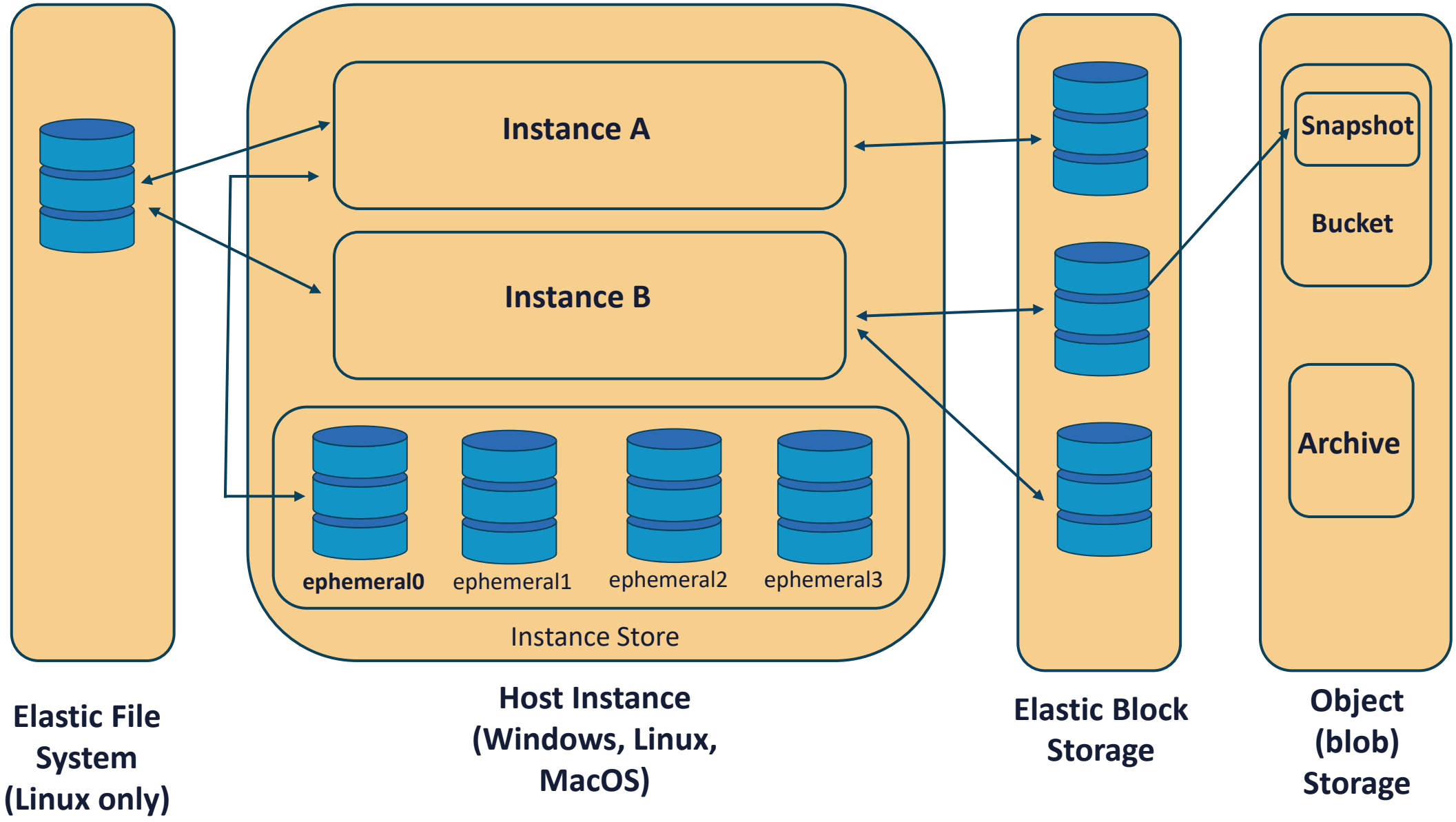
# Cloud Data Flow



# Cloud Data Flow



# Volume (Block Storage) (HDD and SDD)



# Threats to Storage Types

- The most common threat is unauthorized access to data in storage in order to modify, exfiltrate, or delete
- Storage systems can also be subject to DDoS botnet attacks
- Object storage does not have a runtime environment, so malware risk is reduced when compared to file, volume, and database storage
- Objects are usually accessed using API requests or URLs which both have their own list of vulnerabilities

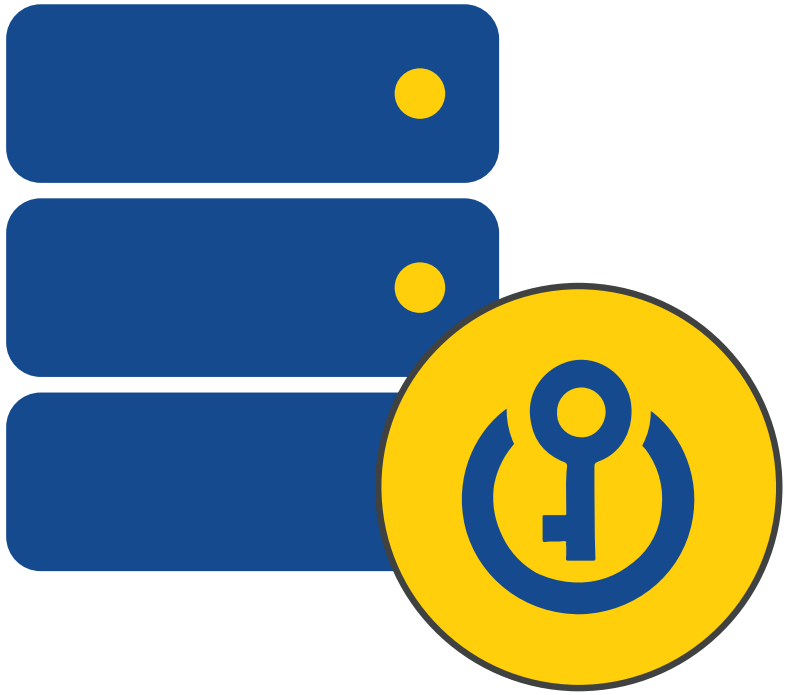
# Threats to Storage Types

- Loss based on physical disk failure in the cloud datacenter remain regardless of the storage or database type
- Ransomware attacks are becoming more common against data stored as objects or blobs at the CSP (S3, Azure Blob, GCS)
- Data containing PHI, PII, IP, and other sensitive content is a prime target for advanced persistent threats (APT) and data remanence attacks

# Encryption

- Cloud providers have the massive resources to employ teams of advanced cryptographers and cryptanalysts to develop bleeding edge crypto solutions such as post-quantum cryptography and homomorphic encryption
- Responsibility for cryptographic protection depends on the service deployment type (IaaS vs. PaaS)
- Newer mechanisms such as **AES-GCM-256**, **SHA-3 hashing**, and **elliptic curve** cryptography are commonly utilized

## Cloud computing has a huge dependency on encryption

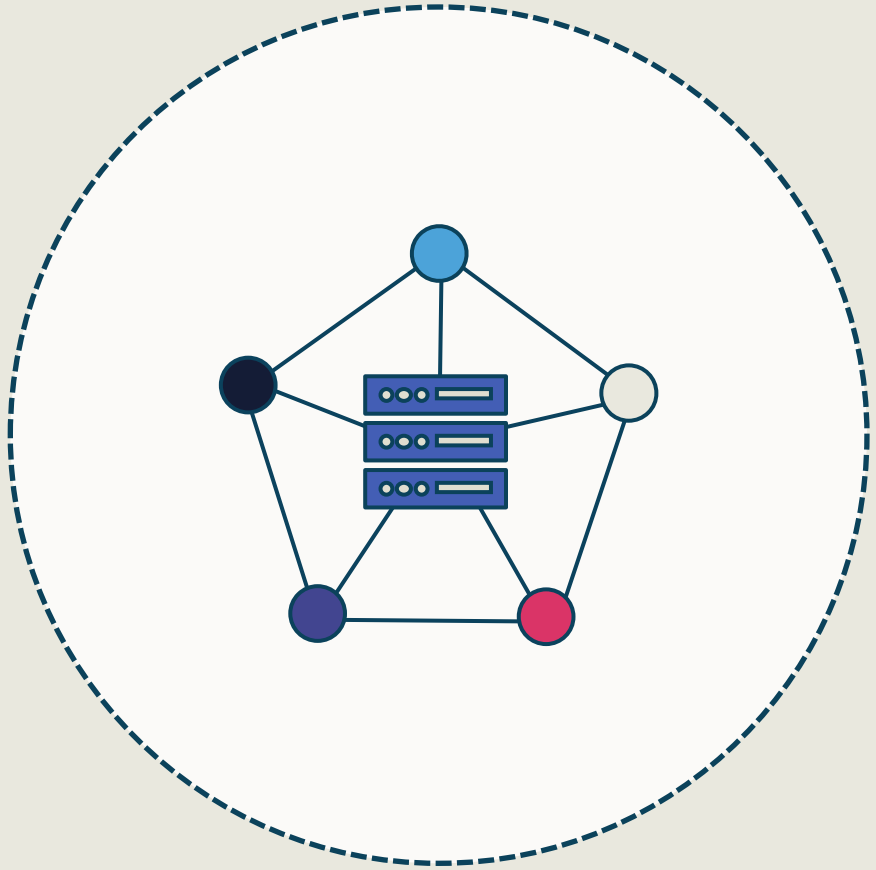


# Hashing

#

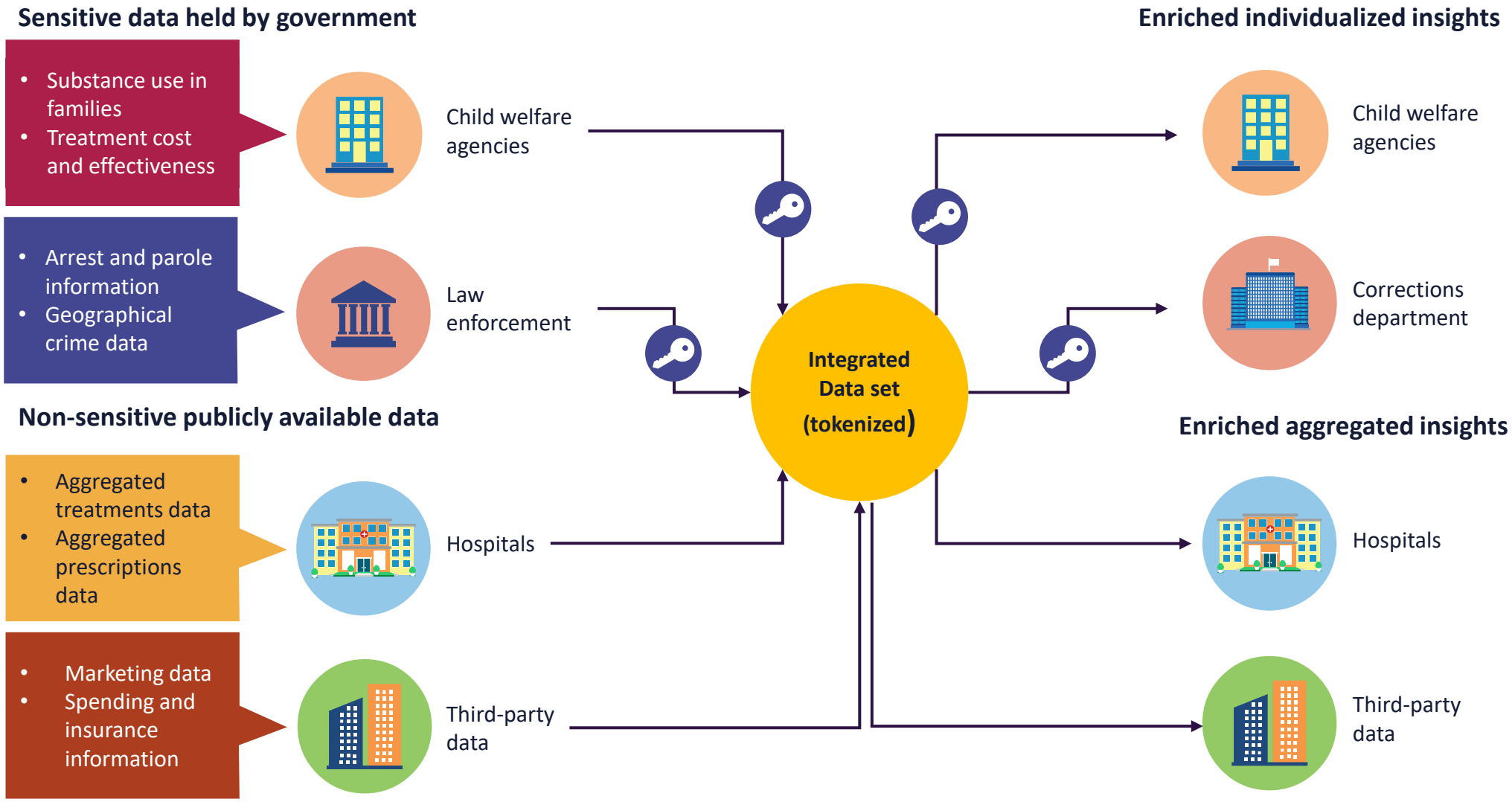
- In a database management system, hashing transforms a string of characters into a typically shorter fixed-length value or key that represents the original string
- Hashing is often used to index and retrieve items in a database because it is faster to find the data item using the shorter hashed key than using the original value

# Database Security: Tokenization

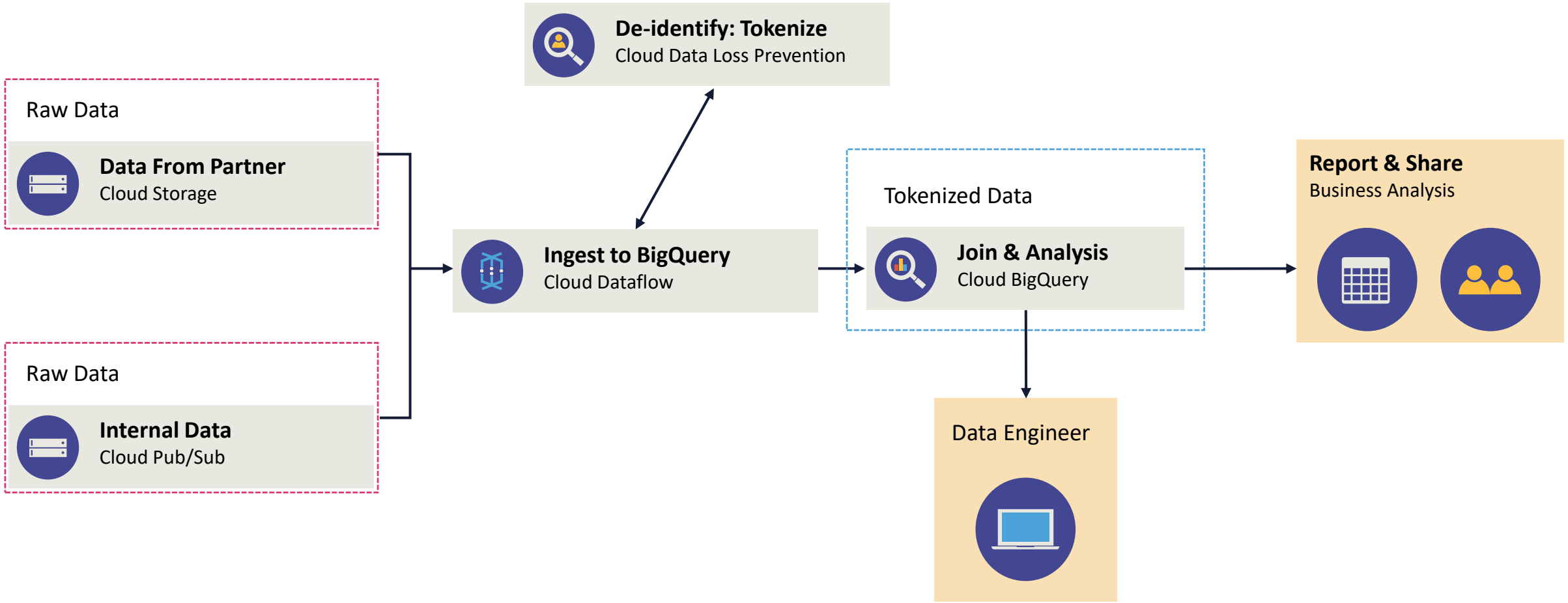


- Tokenization involves sending sensitive data through an API call (or batch file) to a provider that replaces the data with non-sensitive placeholders called tokens
- The practice involves two distinct databases
  - One with the actual sensitive data
  - One with tokens mapped to each chunk of data
- Unlike encrypted data, the tokenized data is irreversible and unintelligible

# Data Tokenization Use Cases



# Data Tokenization on Google Cloud Platform




Source: https://cloud.google.com/blog/products/identity-security/take-charge-of-your-data-how-tokenization-makes-data-usable-without-sacrificing-privacy

# Masking

- Masking often involves using characters like “X” to hide some or all data
- Example is to only display the last 4 digits of:
  - Social Security number
  - Credit card number
  - National ID number
  - Bank account number
  - Username or email address

## Hiding the data with random or useless characters

# De-identification

**Also “anonymization”**

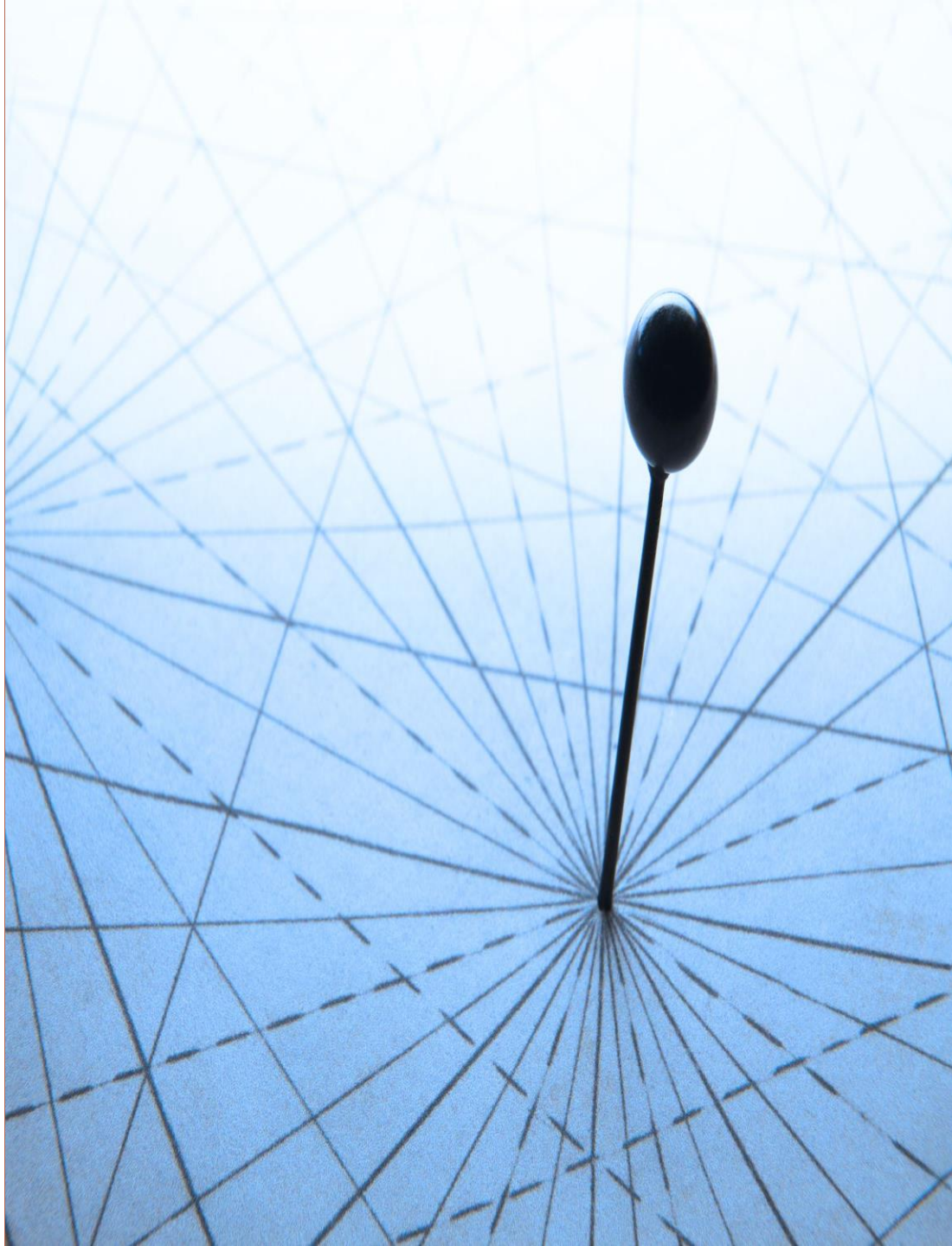
- Anonymization and de-identification involves removing personally identifiable information from a data set that may be used for:
  - Medical research
  - Auditing and testing
  - Financial inquiry
  - Statistical analysis
  - Forensic investigations

# Obfuscation

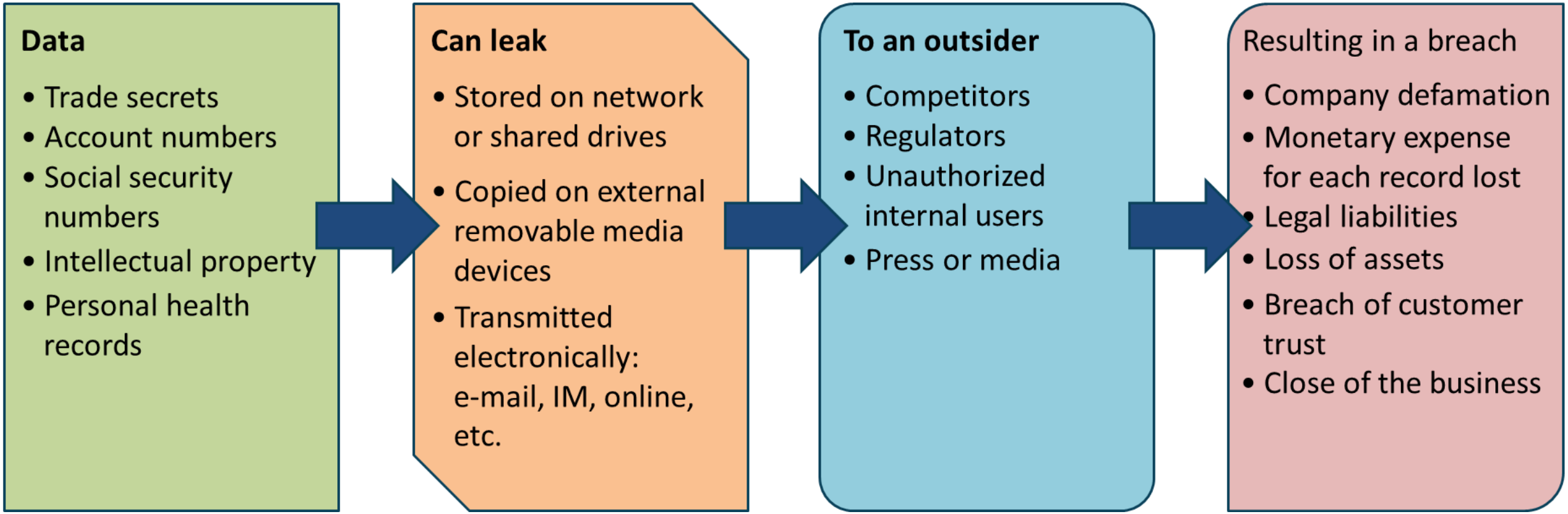
- A general term that applies to any or all mechanisms to present less decipherable or meaningful data
- The goal is to render data unreadable or to hide aspects of personally identifiable, personal health, or corporate intellectual property information
  - **“Obscuring”** is a similar concept where static or dynamic techniques are used on the original data or a representational data set
  - **“Shuffling”** is a term that describes utilizing characters from the same data set to further present the data
  - **“Randomization”** is when all or some of the data is replaced with indiscriminate characters
- **Exam Tip: It is critical that the chosen method prevent inference whereby one could extrapolate the original data set - additional abstraction may be necessary**

# Data Mapping

- Cloud mapping allows you to catalog application resources, such as databases, queues, microservices, and other cloud assets with custom names
- Cloud services can continuously check the health of resources to make sure the location is up-to-date
- For example, the state field in a source system may show Texas as “Texas" but the destination may store it and map it as “TX”

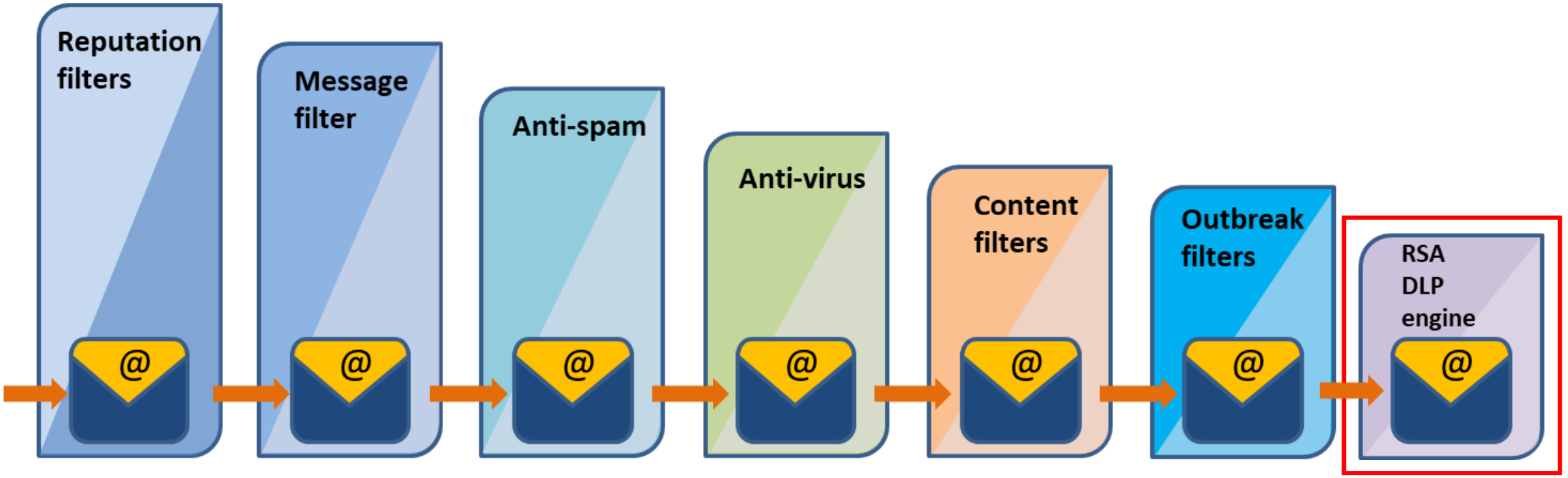


# Data Loss Prevention (DLP)

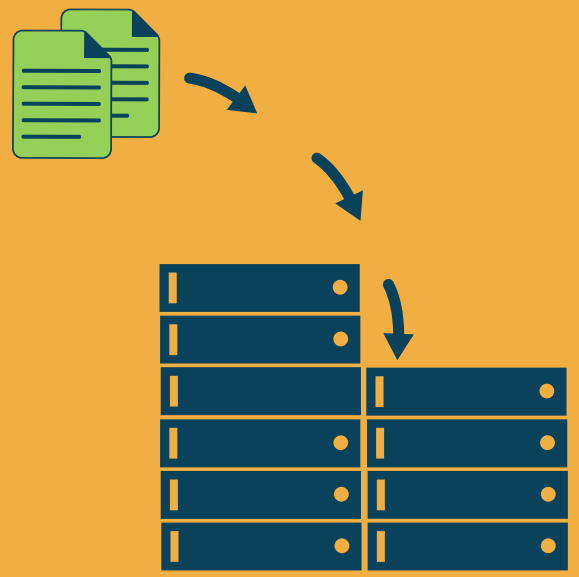


# DLP Egress Webmail Monitoring Example

- The process of examining data as it exits a zone or administrative domain to prevent data loss or leakage of IP, PII, PHI and more. It typically has 4 key goals:
  - Enhanced security controls as part of layered defense-in-depth
  - Enforcement of corporate security policy to accompany the AUP
  - Heightening the monitoring and visibility of egress data and information flow
  - Adherence to governance and regulatory compliance (HIPAA, SOX, PCI, GDPR)



# Data Structures



- **Structured** data follows a pre-defined data model and is therefore straightforward to analyze since it conforms to a tabular format with relationship between the different rows and columns (Excel files or SQL databases)
- **Unstructured** data – information that does not have a predefined data model or is not organized in a pre-defined way (audio, video files or No-SQL databases)
- **Semi-structured data** - does not conform with the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contain tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data (JSON, YAML, XML)
- **Metadata** – data about data that supports Big Data analysis and big data solutions to provides deeper analysis regarding a specific set of data

# Data Discovery

- Data discovery is a methodology that often serves two goals:
  - The enterprise is performing an initial asset assessment and inventory of data ownership
  - The organization is performing e-discovery as part of a digital forensic investigation
- There are three main forms of data discovery
  - **Content-based** – dataset contents such as terms and pattern-matching
  - **Label-based** – discovery is based on existing labels an/or tagging that is applied to physical and logical assets both on-prem and in the cloud
  - **Metadata-based** – leveraging the extensible metadata available on data stored as objects – i.e., using APIs against data in AWS S3, Google Cloud Storage, Azure Blob storage

# Implementing Data Classification

## A key aspect of risk management

- The earliest stages of information security life cycle involve identification, assessment (valuation), and classification of data assets
- Labeling and tagging (logical) concerns the classification of information and is used to determine the level of protection and how the asset should be handled
- Handling controls who has access to information and is based on labeling – how it has been classified
- Classification can be based on location, value, age, utility, useful lifetime, personal association, or a sensitivity label for a mandatory access control (MAC Bell-LaPadula or  Biba) model

# Sensitivity Levels for Data Classification

**Government/military:**

- Top Secret
- Secret
- Confidential
- Sensitive But Unclassified (SBU)
- Unclassified

**Commercial/private sector:**

- Confidential
- Private
- Sensitive
- Public

# PII and PHI

- Personally Identifiable Information (PII)
  - Consists of first name or initial with last name and one or more pieces of info
  - Social Security number, driver's license number, ID card, financial account number
- Protected Health Information (PHI)
  - Contains at least one piece of info
  - Name, address, birth date, phone number, mail or e-mail Address, Social Security number, URLs
  - Medical history
  - Health records and lab reports

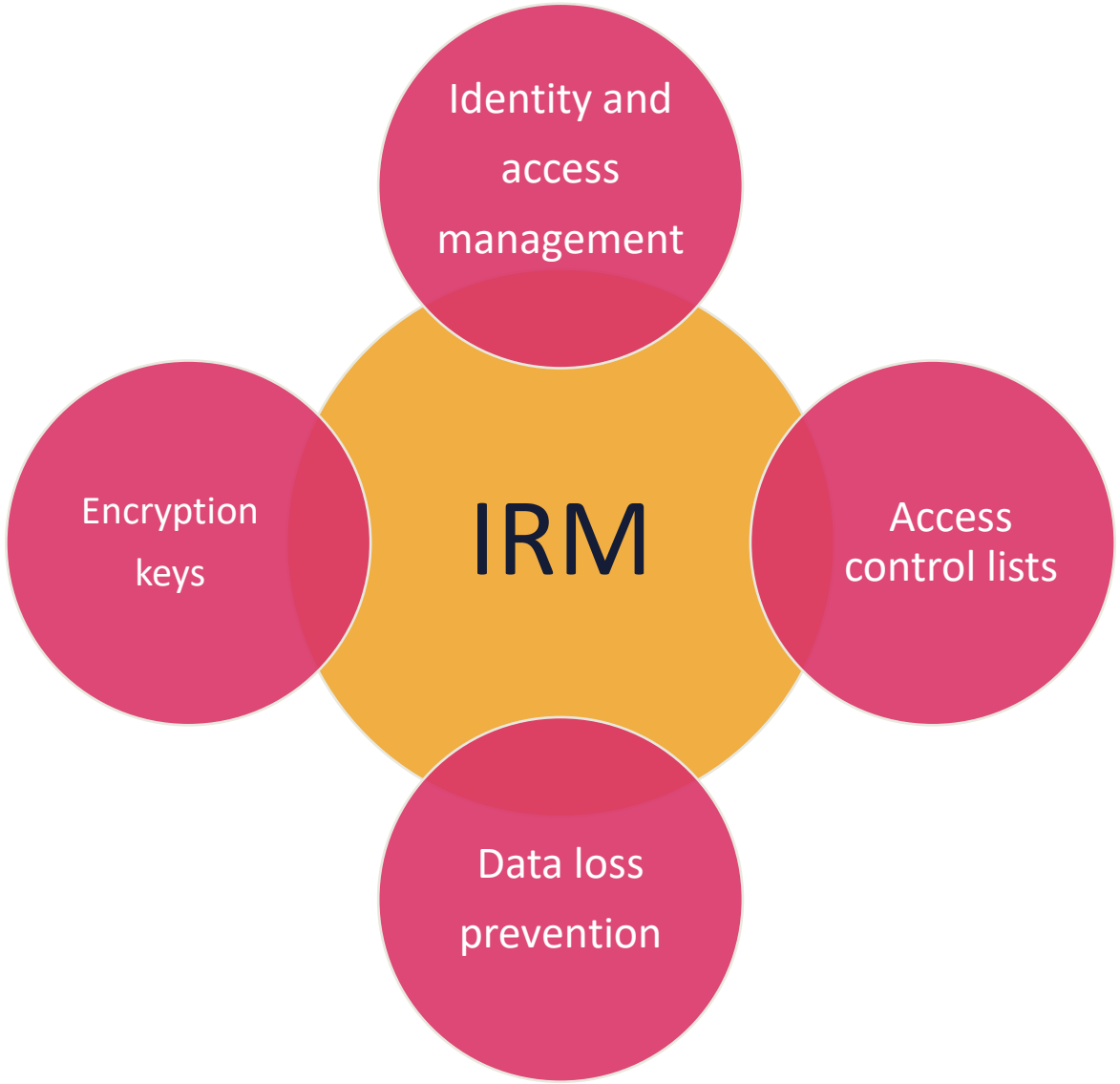
# Information Rights Management (IRM)

## Also called DRM and E-DRM

- The objectives of IRM are to employ controls that work with, or in addition to, access control security to protect data and file-level assets
- Example: Must control copying, deleting, and modifying certain PDF documents to protect intellectual property (IP) and copyrights
  - A copyright usually persists for either 70 after the death of the author or 120 years after first use of a work for hire
- Since digital signing and certificates are often used, an enterprise PKI may be part of the policy

**The CCSP exam prefers the term IRM over DRM**

# Information Rights Management (IRM)



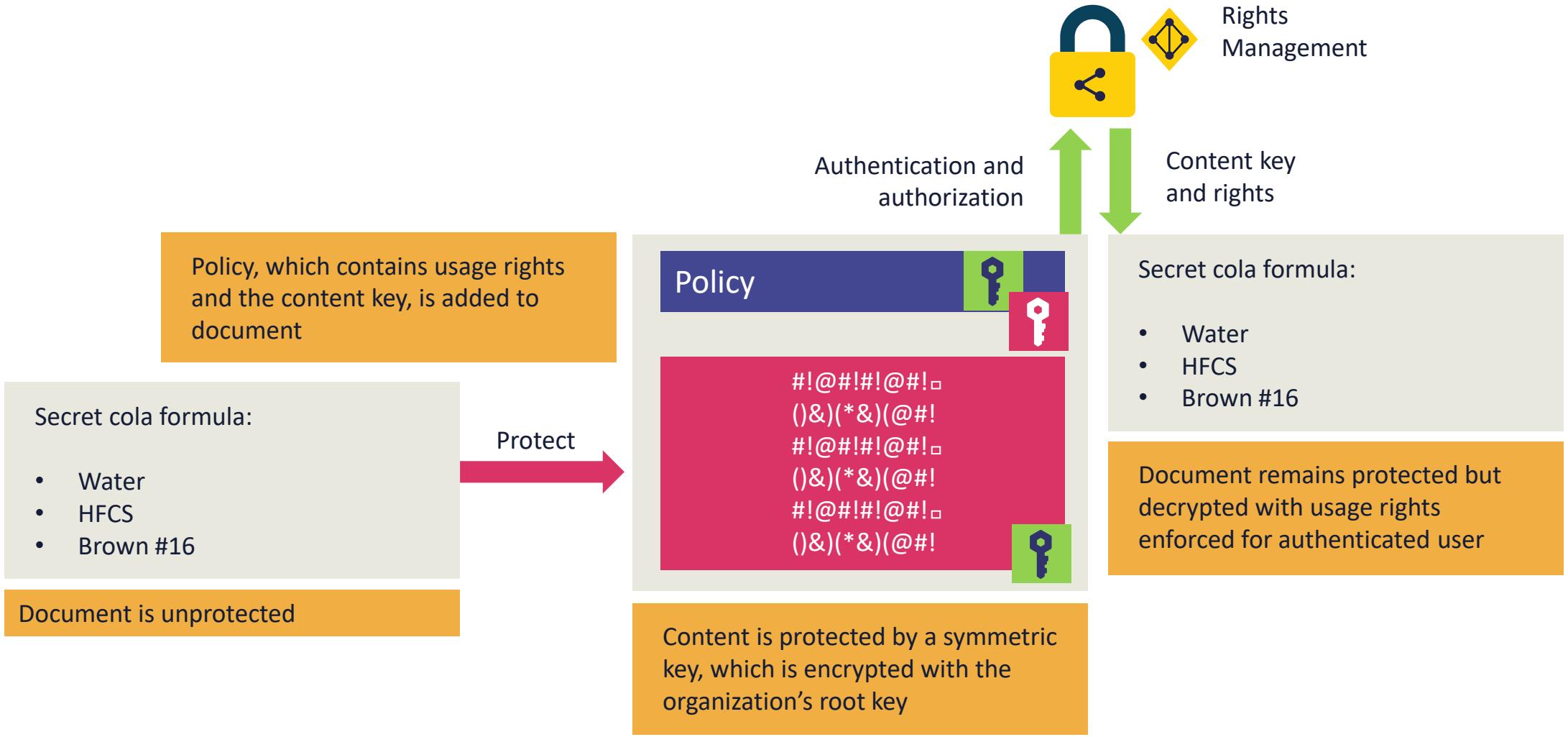
# Digital Rights Management (DRM)

- DRM is access-control technology that protects licensed digital intellectual property (IP)
- DRM is used by publishers, manufacturers, and IP owners for digital content and device monitoring
- Digital media licensees attempt to balance the rights of IP owners and Internet users by protecting rights and profits for digital product manufacturers and retailers
- DRM mechanisms are varied access control technologies used to control usage of proprietary hardware and copyrighted works

# Rights Management for PDF Files

| | | | | |
|---|---|---|---|---|
| Manage document usage | Deny unauthorized sharing | Stop screen captures or printing to files | Enforce expiration | Revoke access based on least privilege |
| Restrict to specific IP CIDR ranges | Watermark PDF files | Track document usage | Integrate with CLI for automation | Integrate with e-commerce solutions |

# EXAMPLE: AZURE RIGHTS MANAGEMENT

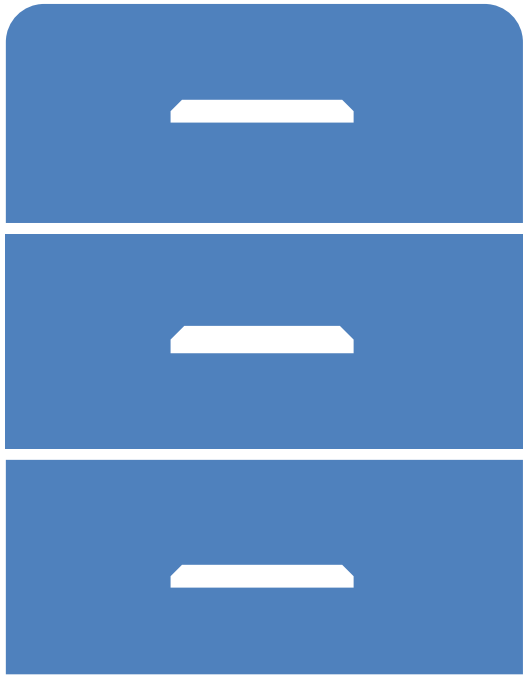


# Data Retention

- Data retention typically applies to production data in the Use and Share phases as it is kept for practical purposes as well as corporate policies and regulatory mandates
  - The policy may also mandate formatting and storage of data as well
  - A data retention policy may be closely aligned with the **Archive** phase of the cloud data lifecycle
- The retention period (usually several years) is an established time period that data remains in the archive phase before transitioning to the disposal phase
  - These period may be modified due to change of providers, SLAs, or new regulations

# Data Archiving Decisions

- There is no one-size-fits-all for archiving data however it may likely be the Backup/Restore function of BCP
  - A backup policy is only as good as the restoration testing that has been performed
- There may be different data storage and archiving policies for volume (block) data and object data
  - It is highly recommended that incremental and differential backups ALWAYS be programmed and automated
- CSPs offer cold storage and “deeper” cold storage
- Some cloud options involve different retrieval plans
  - Expedited – 1 to 5 minutes
  - Standard – 3 to 5 hours
  - Bulk – 5 to 12 hours



# Data Deletion

- As far as cloud computing is concerned, the two main methods for customer data deletion are:
  - Using random data or null pointers to **overwrite** over data sectors that stored sensitive or proprietary intellectual property
    - The acceptable methods will be driven by regulatory standards and guidelines including the number of iterations
    - Not a practical method for confidence at a cloud provider
  - The most effective and practical method in the cloud is **cryptographic shredding or erasure**

# Legal Holds

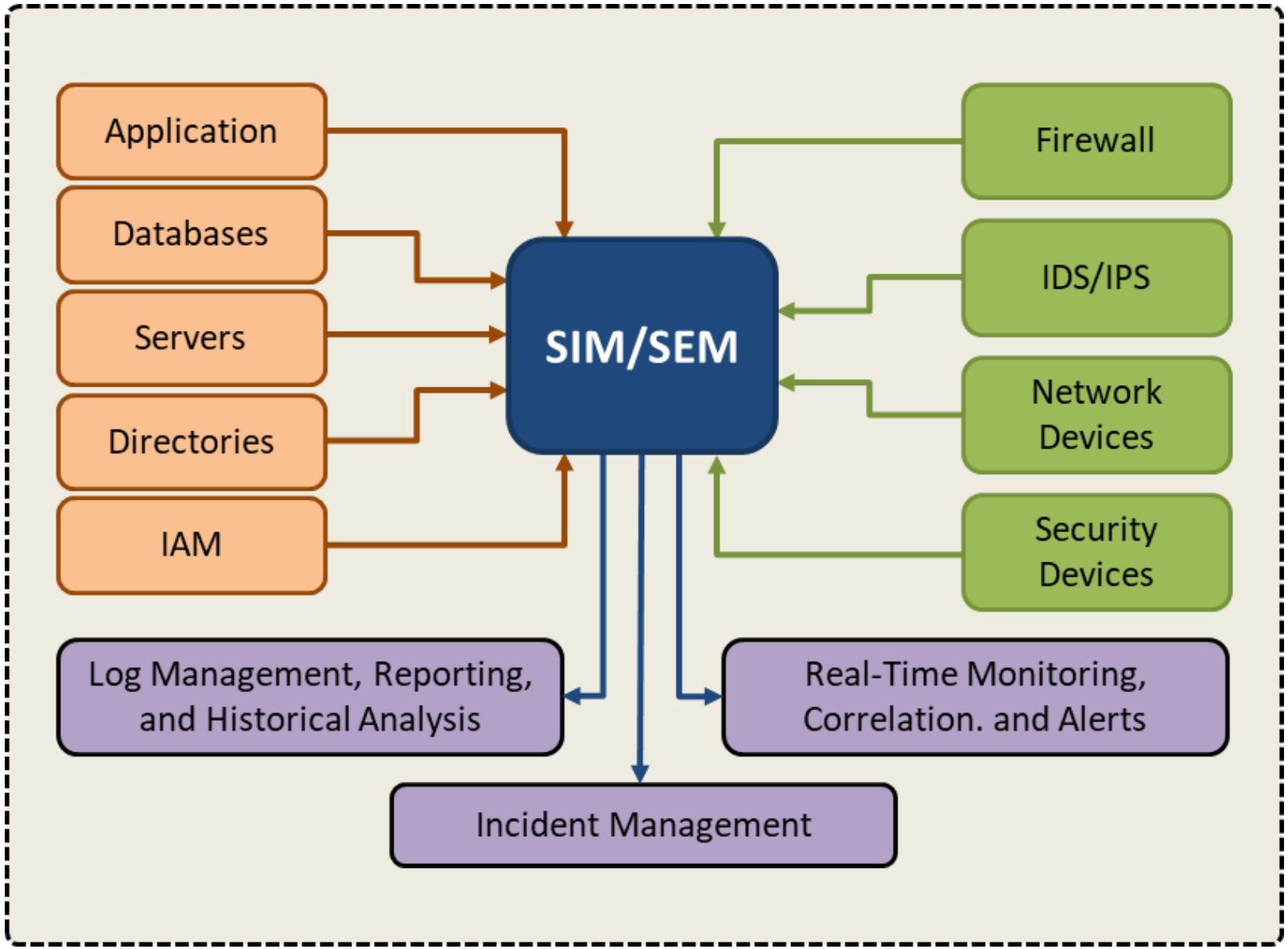
- A legal hold refers to a process which an organization uses to retain all forms of pertinent data and information when it reasonably expects some type of litigation against it, or some need for future utility in a court of law
- It can be a restriction placed on a database or set of records that exists as a result of existing or anticipated litigation, audit, government investigation or other such activity that suspends the regular usage, processing, or disposition of data.

# Data Audit Policies

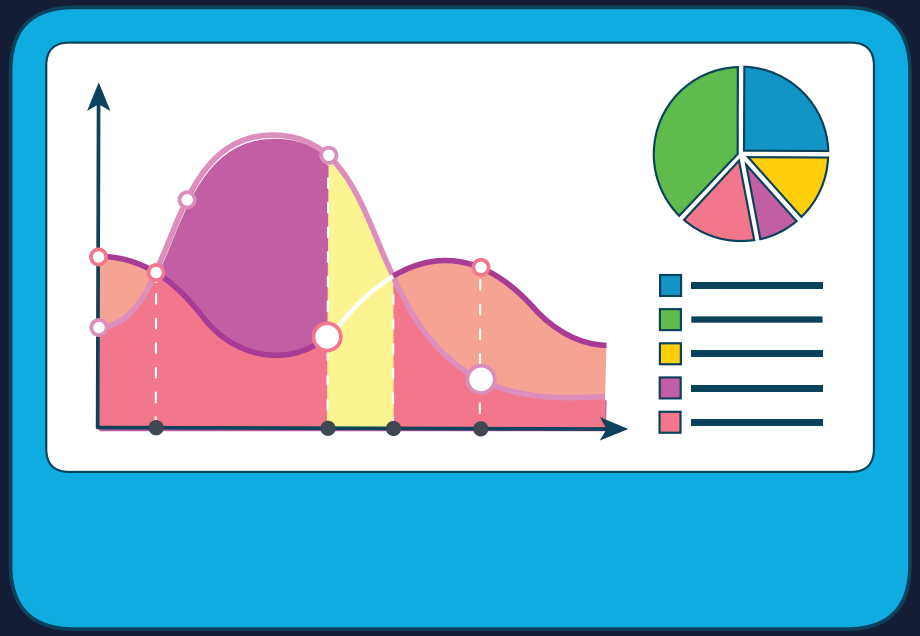
**Critical for continual improvement of data security in the enterprise**

- The audit policy should consider the following:
  - Auditing periods
  - Audit scope
  - Internal or external or both
  - Restrictions
  - Avoiding conflicts if interest
  - Established processes and procedures
  - Approved tools and techniques

# Audit Event Sources



# Handling Audit Results



- The raw auditing data that is processed by SIEM systems or RADIUS accounting services is often stored on backend databases in a highly available and secure manner
- The data will often be copied to a cloud service (such as AWS Redshift data warehousing or Google BigQuery) for further analysis – often with machine learning and artificial intelligence engines
  - API calls with the logs, records and reports should be digitally signed whenever feasible
- If the audit results are part of a legal hold or forensic investigation, the chain of custody must be maintained throughout the entire lifecycle
- Audit files and images may need to be hashed with SHA-2 or SHA-3 algorithms