Welcome Back to CISSP Bootcamp Day 5

Michael J. Shannon

Class will begin at 10:00 am Central Standard Time

Operating Detection and Preventative Measures

Objectives

- Examine firewalls and next-generation intrusion detection systems (IDSs)/intrusion prevention systems (IPSs)
- Compare whitelisting and blacklisting
- Describe third-party security services, sandboxing, honeypots, and honeynets
- Know about antimalware
- Learn about machine learning (ML) and AI-based tools
- Describe patch and vulnerability management

Firewalls

- A firewall is a metaphor representing software and/or hardware controls that can limit the damage spreading from one subnet, virtual local area network (VLAN), zone, or domain to another
- It is typically deployed as a barrier (zone interface point) between an internal (trusted) network and an external (untrusted) network
- They are integrated systems of threat defense functioning at layers 2-7 and can be categorized as network or application firewalls

Next Generation Firewall Features

1. Layer 5-7 policies – deep packet inspection (DPI) or application visibility and control (AVC)

2. Authentication proxies – software-defined perimeter (SDP) access gateways

3. Identity services front-end for identity management (IdM) and 802.1X PNAC

Next Generation Firewall Features

4 Integrated IDS and IPS sensors

5 Unified threat management with data loss prevention (DLP)

6 Advanced malware protection (cloud-based)

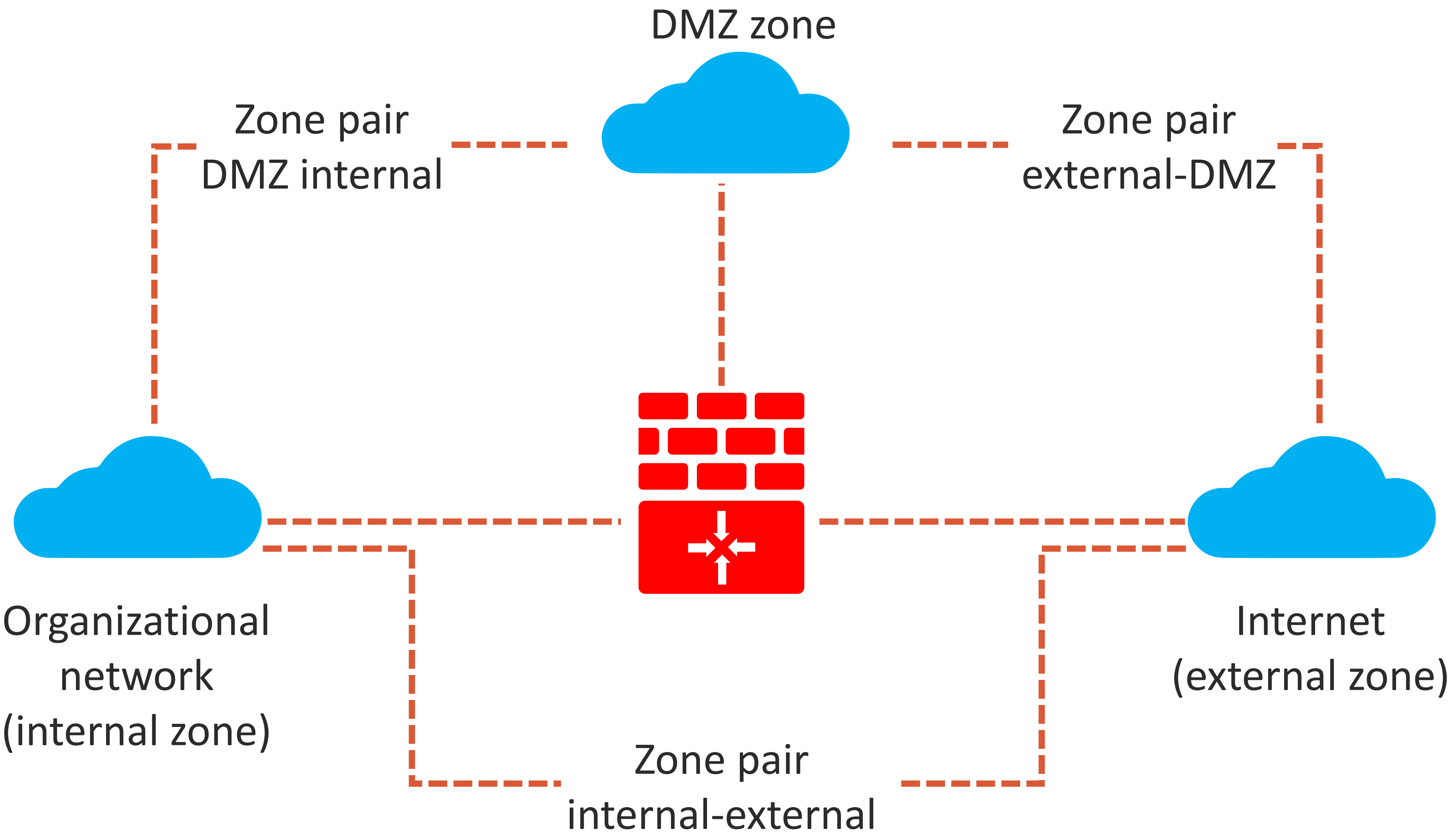
Next Generation Firewall Features

7 URL and content filtering

8 Botnet filtering for active defense

9 Cloud correlation and reputation filtering

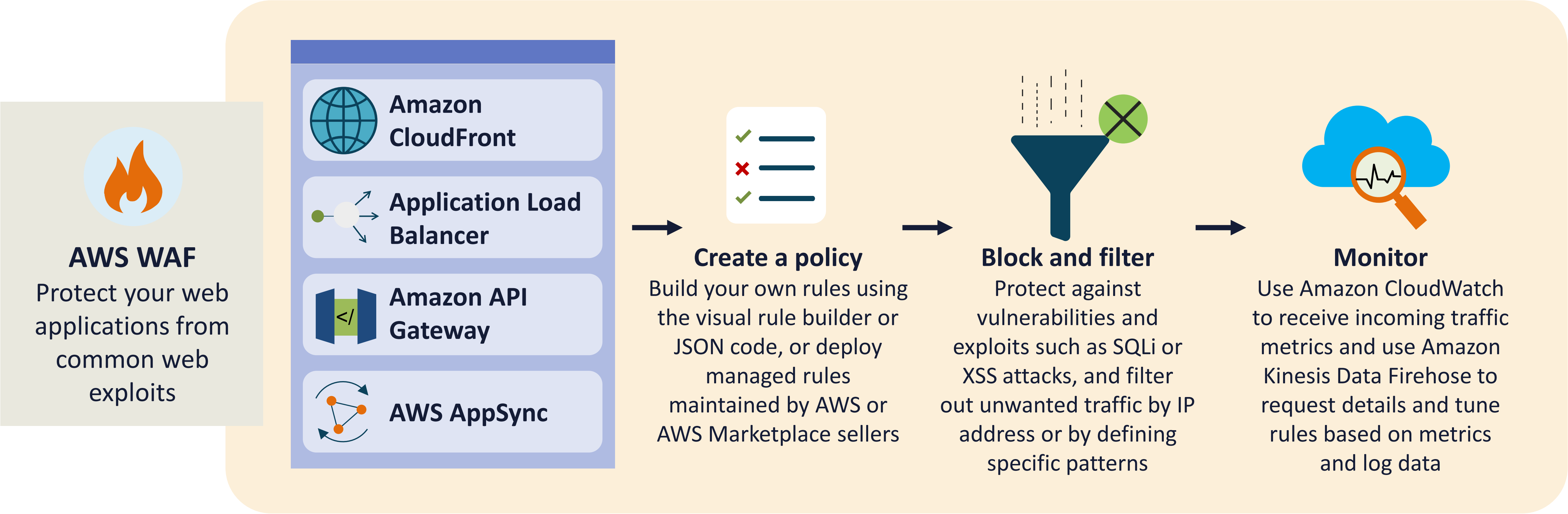
Zone-based Firewalls



Web Application Firewall (WAF)

- Also called a web security gateway, these are appliances (usually virtual), server plugins, or filter engines that apply a set of rules (web access control lists) to an HTTP or HTTPS conversation
- These rules cover common OWASP Top 10 web attacks like cross-site scripting (XSS), request forgery, and various injection attacks
- They are often managed cloud services or managed security service provider (MSSP)/cloud access security broker (CASB) solutions from different vendors
- **The AWS WAF can be deployed on an elastic application load balancer, content delivery network (CDN) distribution, or application programming interface (API) gateway**

Cloud-based WAF Solution



AWS WAF

Protect your web applications from common web exploits

Amazon CloudFront

Application Load Balancer

Amazon API Gateway

AWS AppSync

→

Create a policy

Build your own rules using the visual rule builder or JSON code, or deploy managed rules maintained by AWS or AWS Marketplace sellers

→

Block and filter

Protect against vulnerabilities and exploits such as SQLi or XSS attacks, and filter out unwanted traffic by IP address or by defining specific patterns

→

Monitor

Use Amazon CloudWatch to receive incoming traffic metrics and use Amazon Kinesis Data Firehose to request details and tune rules based on metrics and log data

Next Generation IPS (NGIPS)

- IPS solutions have evolved rapidly over the last ten years
- NGIPS is a security implementation technology that delivers intelligence and visibility to protect the infrastructure from cyber attacks and related threats
- NGIPS is commonly integrated into unified threat management appliances and next generation firewalls

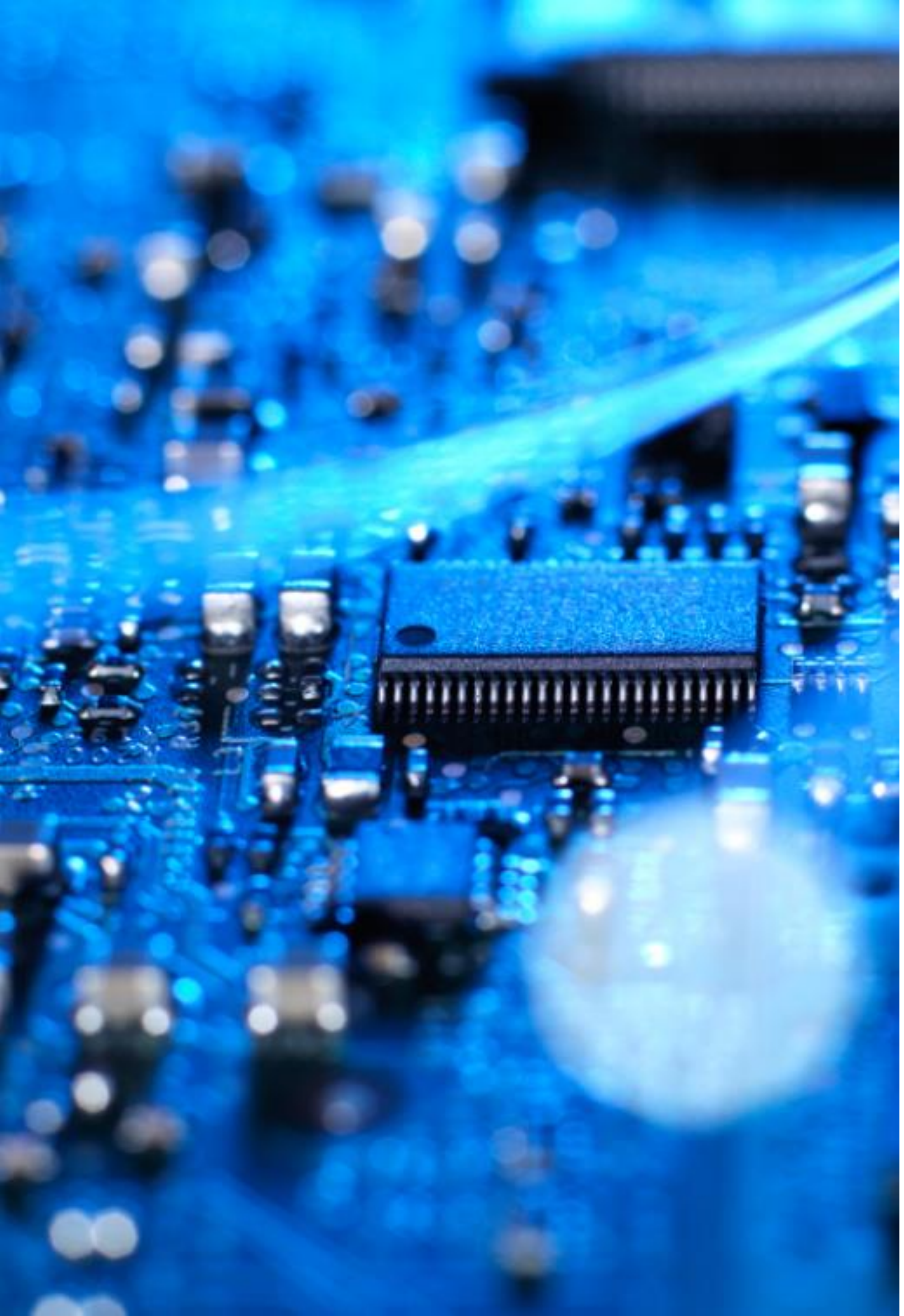
```
.data)
esponse.data[0]['create
datetime.strptime(ltdate,'%a %b %d
tetime.now()
(today-ltdate2).days
g < daywindow:
 i.screen_name, 'has          d in the
tweets += len(response.data)
 in response.data:
if j.entities.urls:
  for k in j.entities.urls:
   newurl = k['expanded_url']
```

Next Generation IPS (NGIPS)

- Another trend for NGIPS is to connect with cloud-based providers and Software as a Service (SaaS) and CASB network services
- NGIPS forwards and transports Internet Protocol (IP) packets to all information and services
- NGIPS also has built-in sandboxing and advanced malware protection components

NGIPS Features

- **Visibility:**
  - NGIPS allows security managers to see more contextual data from your network and fine-tune your security
  - NGIPS can see all applications, indicators of compromise, file disposition
  - It also offers sandboxing, vulnerability data, and device-level OS views
- **Effectiveness:**
  - Secure IPS solutions typically get new policy rules and signatures every couple of hours, so that security is up to date
  - Solutions leverage global threat detection networks to bring security effectiveness to the products



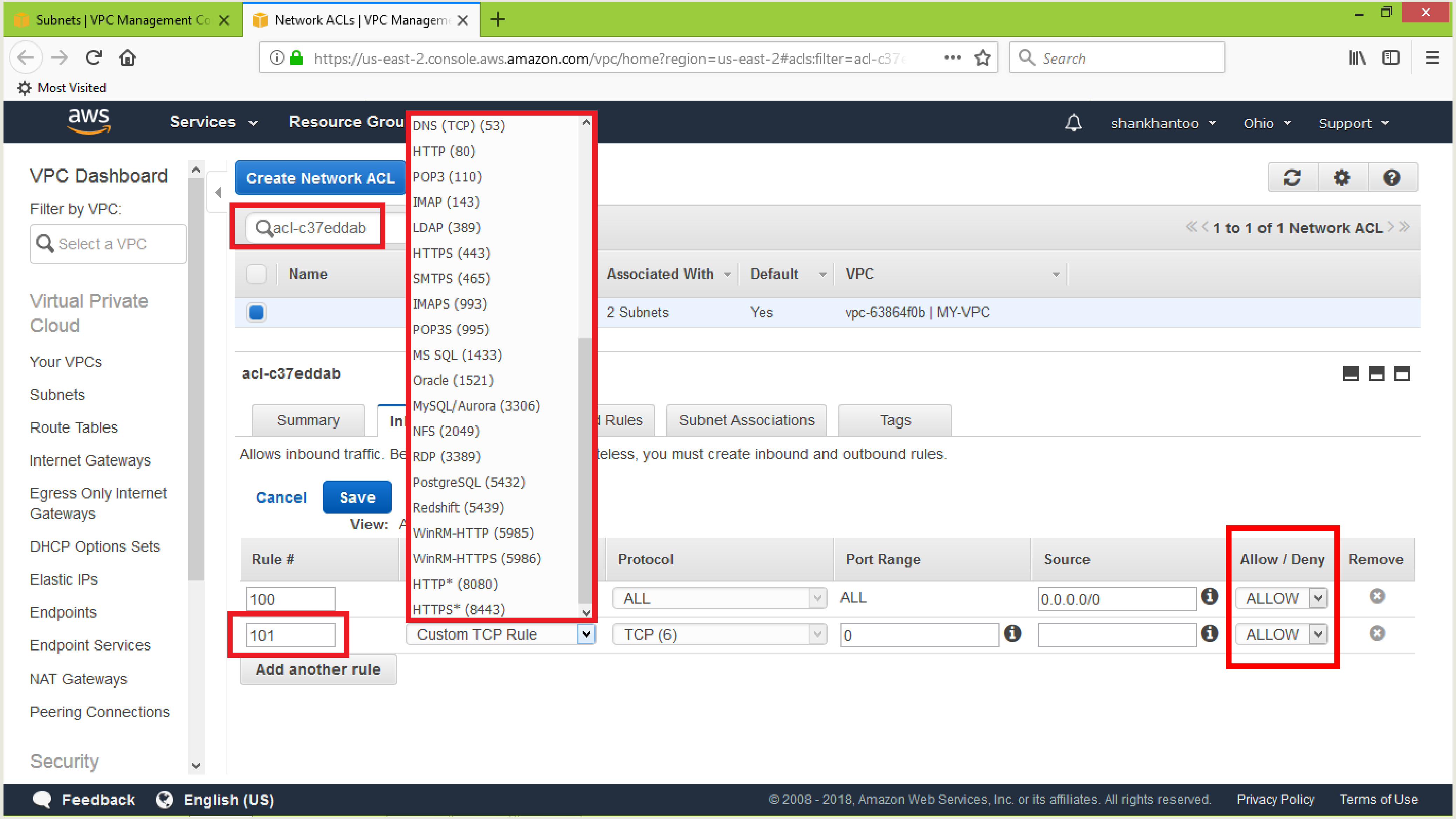
NGIPS Features

- **Cost optimization:**
  - Automation enhances operational productivity and lowers overhead by separating actionable events from noise
- **Flexibility and agility:**
  - NGIPS has flexible deployment options
  - It can be deployed at the corporate edge, in the data center distribution/core, or behind the firewall
  - Secure IPS can be deployed for inline prevention or passive detection
- **Integration:**
  - NGIPS integrates into modern networks without major hardware modifications or substantial downtime
  - They have a single unified portal with a single pane in the security operations center

Managed Security Service Providers (MSSPs)

- A common next-generation IPS solution comes from MSSPs offering outsourced security monitoring and management for security systems and devices
- Common MSSP services are
  - IDS/IPS
  - Managed layer 3-7 firewalls
  - Endpoint response and detection
  - Virtual private networking support
  - Vulnerability scanning and antiviral services

Blacklisting



Subnets | VPC Management Co ✕ | Network ACLs | VPC Managem ✕ | +

https://us-east-2.console.aws.amazon.com/vpc/home?region=us-east-2#acls:filter=acl-c37

Search

Most Visited

aws | Services | Resource Grou | shankhantoo | Ohio | Support

VPC Dashboard

Filter by VPC:

Select a VPC

Virtual Private Cloud

- Your VPCs
- Subnets
- Route Tables
- Internet Gateways
- Egress Only Internet Gateways
- DHCP Options Sets
- Elastic IPs
- Endpoints
- Endpoint Services
- NAT Gateways
- Peering Connections

Security

Create Network ACL

acl-c37eddab

1 to 1 of 1 Network ACL

☐	Name		Associated With	Default	VPC
☑			2 Subnets	Yes	vpc-63864f0b | MY-VPC

acl-c37eddab

Summary | In | Rules | Subnet Associations | Tags

Allows inbound traffic. Be | eless, you must create inbound and outbound rules.

Cancel | Save

View: A

- DNS (TCP) (53)
- HTTP (80)
- POP3 (110)
- IMAP (143)
- LDAP (389)
- HTTPS (443)
- SMTPS (465)
- IMAPS (993)
- POP3S (995)
- MS SQL (1433)
- Oracle (1521)
- MySQL/Aurora (3306)
- NFS (2049)
- RDP (3389)
- PostgreSQL (5432)
- Redshift (5439)
- WinRM-HTTP (5985)
- WinRM-HTTPS (5986)
- HTTP* (8080)
- HTTPS* (8443)

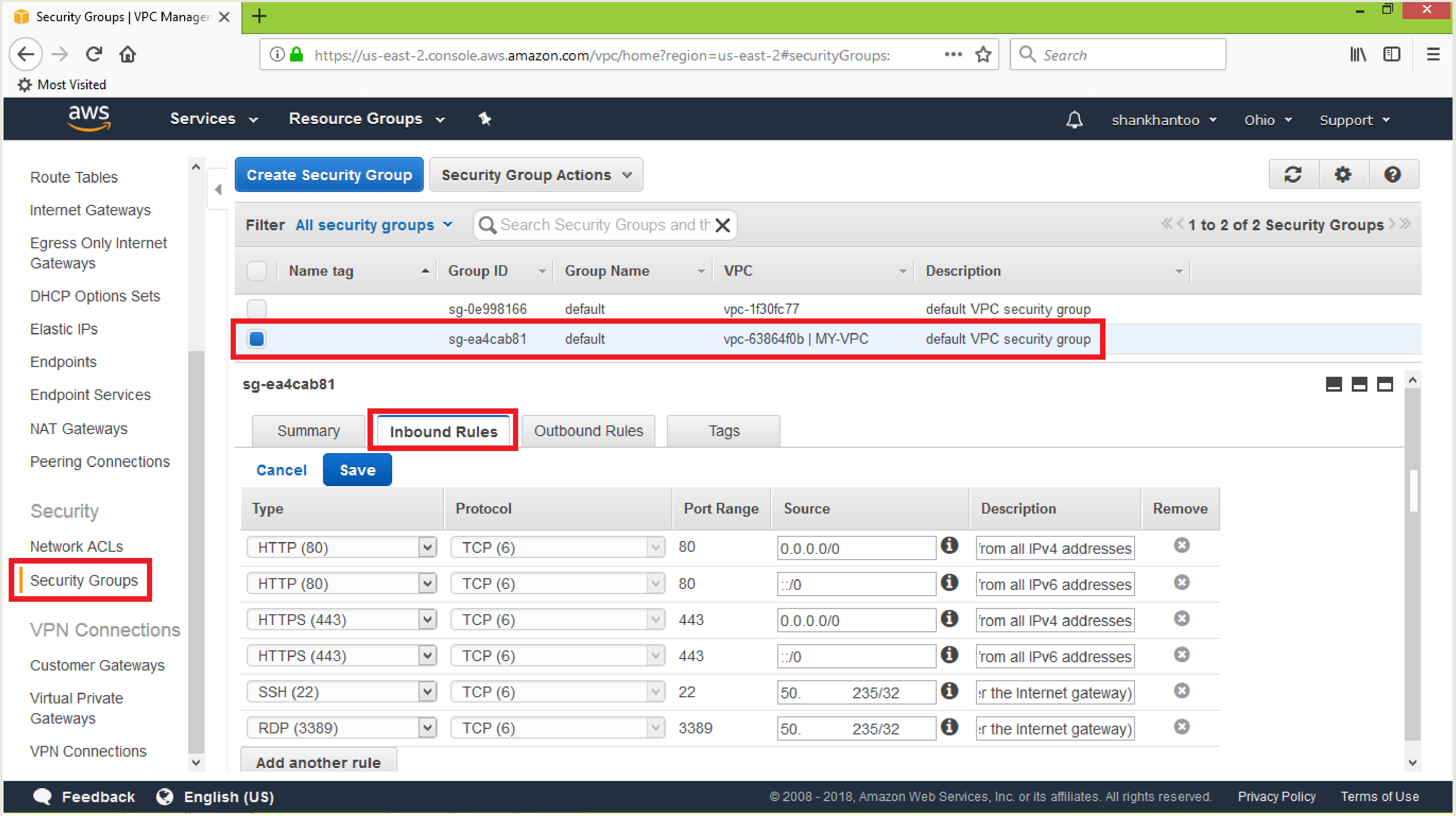
Rule #		Protocol	Port Range	Source	Allow / Deny	Remove
100		ALL	ALL	0.0.0.0/0	ALLOW	⊗
101	Custom TCP Rule	TCP (6)	0		ALLOW	⊗

Add another rule

Feedback | English (US)

© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Whitelisting



	Name tag	Group ID	Group Name	VPC	Description
□		sg-0e998166	default	vpc-1f30fc77	default VPC security group
☑		sg-ea4cab81	default	vpc-63864f0b | MY-VPC	default VPC security group

sg-ea4cab81

Summary | **Inbound Rules** | Outbound Rules | Tags

Cancel | Save

Type	Protocol	Port Range	Source	Description	Remove
HTTP (80)	TCP (6)	80	0.0.0.0/0	'rom all IPv4 addresses	⊗
HTTP (80)	TCP (6)	80	::/0	'rom all IPv6 addresses	⊗
HTTPS (443)	TCP (6)	443	0.0.0.0/0	'rom all IPv4 addresses	⊗
HTTPS (443)	TCP (6)	443	::/0	'rom all IPv6 addresses	⊗
SSH (22)	TCP (6)	22	50. 235/32	:r the Internet gateway)	⊗
RDP (3389)	TCP (6)	3389	50. 235/32	:r the Internet gateway)	⊗

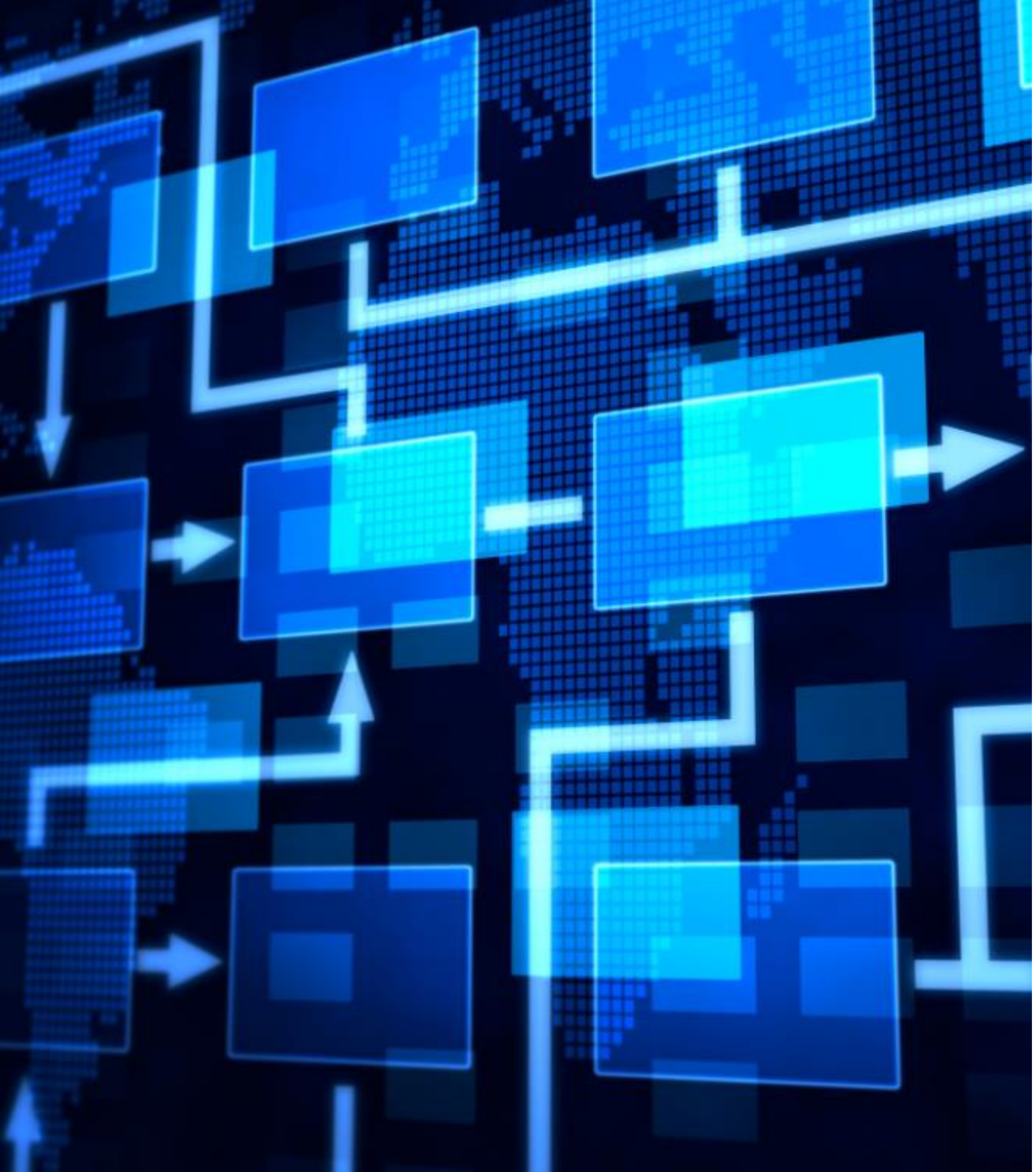
Add another rule

Sandboxing

- Sandboxing is a practice where professionals run code or applications in an insulated environment such as a type 1 or type 2 hypervisor
- This detonation chamber is used to test code or applications that may potentially be dangerous
- Sandboxing environments are often for threat modeling and reverse-engineering of malware
- Sandboxing is designed to reduce threats from entering the network by analyzing and eliminating them

Types of Sandboxing

- Manual
- Automatic
- Hybrid



Sandbox Environment Features

- **Virtual machines** running in on-site and CSP private cloud deployments are leveraged to simulate entire computer systems
- **Emulators** imitate certain hardware or software components, allowing for granular testing
- **System-level sandboxes** are fully isolated systems for observing possibly malicious code that affects system-level processes
- **Application-level sandboxes** confine an application's access to specific system resources to test potentially malicious code while stopping any widespread damage

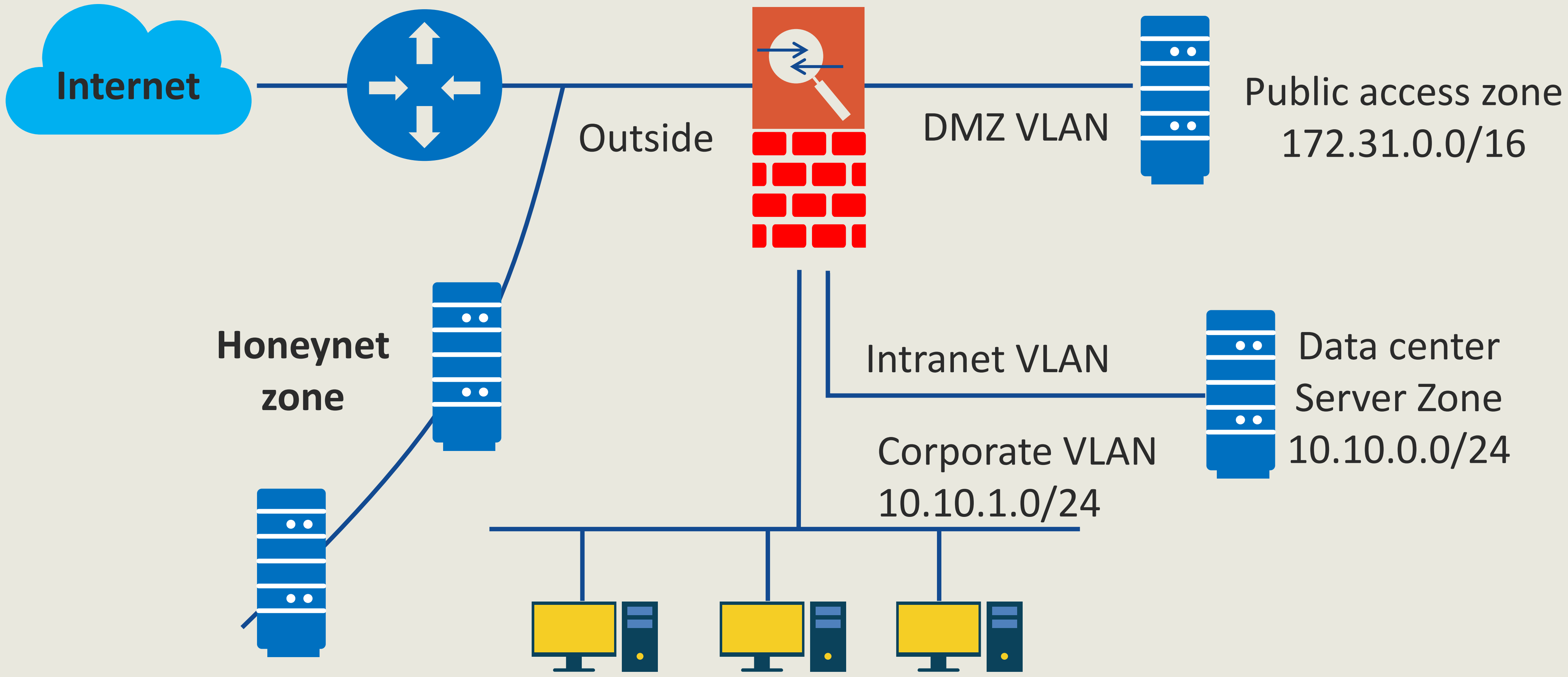
Sandboxing Analysis Types

- **Static analysis** looks at questionable code without executing it:
  - This exposes known malicious patterns without placing the code in a runtime state
- **Dynamic/runtime analysis** examines the code in real-time as it runs within the sandbox environment:
  - This helps to discover the malicious actions that only happen during execution
- **Memory dump analysis** comprises inspecting the memory image from the sandbox to detect malicious code running in volatile memory:
  - This practice is critical for exposing newer threats that reside only in memory and with few other trace behaviors

Honeypots and Honeynets

- Honeypots and honeynets are isolated systems, sites, and services with data that appear to be legitimate and valuable to an advanced persistent threat actor
- They entice potential malicious users to connect (internal or external)
- IDS and security information and event management (SIEM) systems track and log all traffic to and from the honeypot
- Blue teams will perform active defense procedures
- Security orchestration, automation, and response (SOAR) runbooks can also be run for counterattack/active defense initiatives

Honeynet Zones



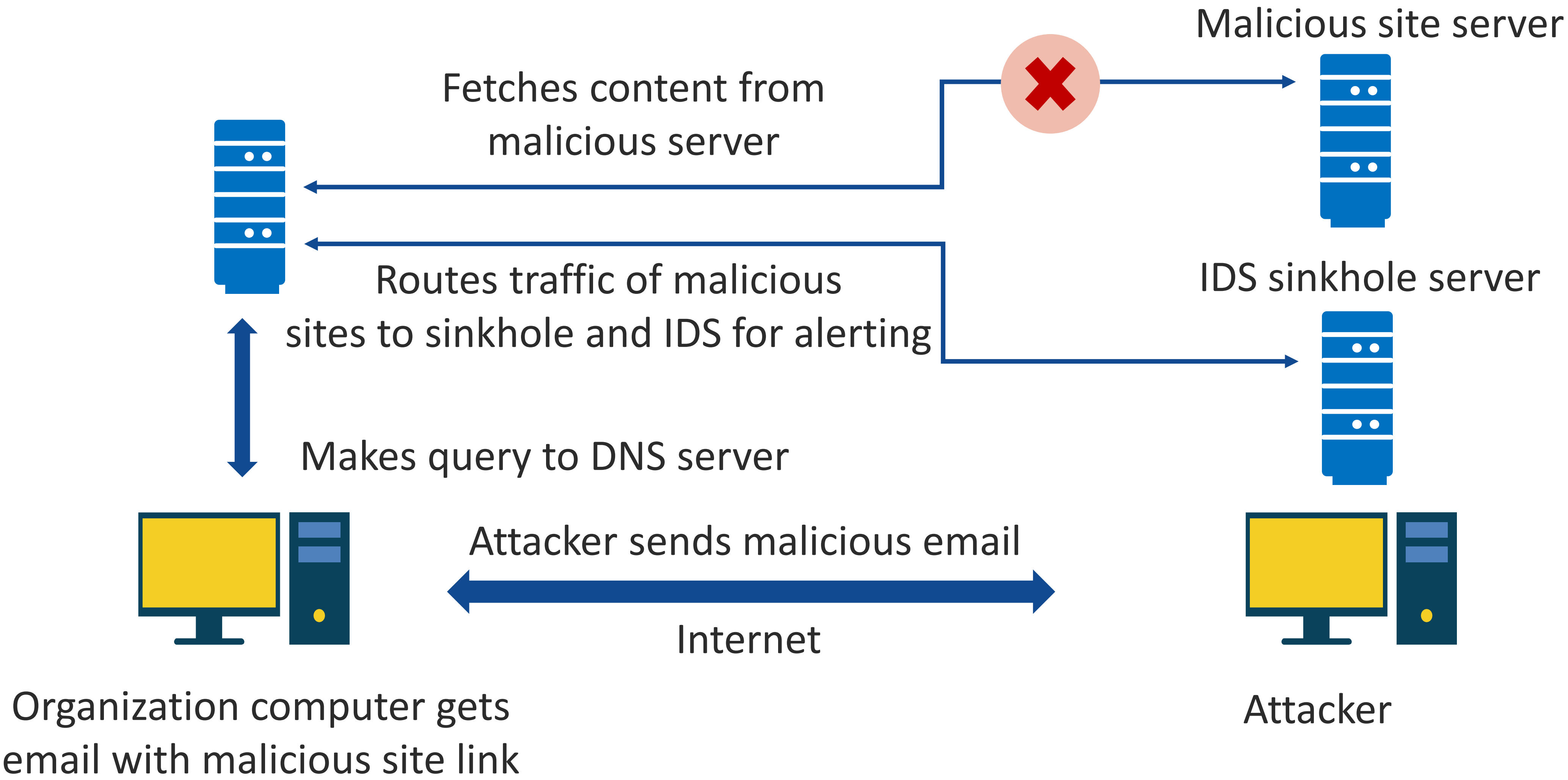
Active Defense Deception: Fake Telemetry

- Deception is the first and most common phase of active defense used by many organizations
- Fake telemetry involves augmenting existing enterprise tools to offer critical threat intelligence for early breach detection and high-fidelity alerting
- It involves making tools available on honeypots and honeynets for attackers to use to attribute and attack back

Active Defense Deception: DNS Sinkholes

- DNS sinkhole (or black hole DNS) is used to spoof DNS servers to prevent resolving hostnames of specified URLs
- This can be accomplished by configuring the DNS forwarder to return a false IP address to a specific URL
- It can be used against attackers to slow them down in a honeynet deployment and then possibly perform active defense attribution techniques
- Can also be used to prevent access to malicious URLs at an enterprise level for internal users

DNS Sinkhole



Malicious site server

Fetches content from malicious server

Routes traffic of malicious sites to sinkhole and IDS for alerting

IDS sinkhole server

Makes query to DNS server

Attacker sends malicious email

Internet

Organization computer gets email with malicious site link

Attacker

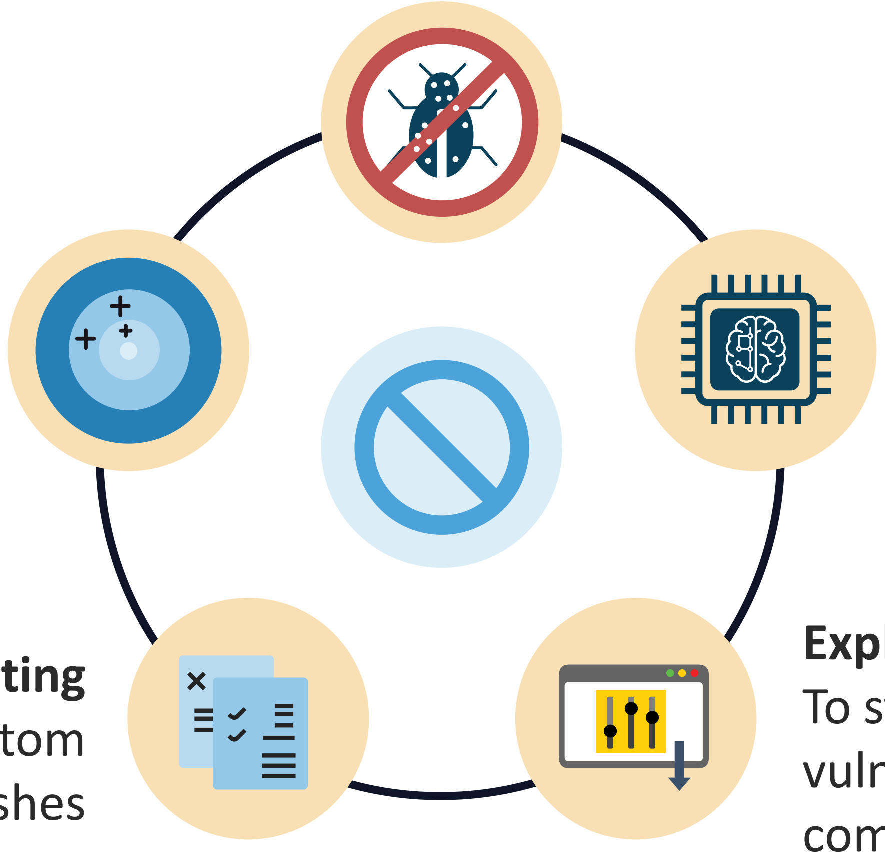
Honeytokens

- Honeytokens are a piece of fake or deliberately misleading data used to locate and expose compromised privileged insiders
- These are artifacts placed in strategic locations to be discovered and used by suspects in an unauthorized manner
- Common examples are
  - Fake privileged accounts to a database system or directory service
  - Access keys to dummy cloud provider accounts
  - Additional balances added to petty cash or discretionary expense funds

Antimalware

- Antimalware is a form of software that is intended to proactively and dynamically protect applications and information systems and individual computers from dangerous malware
- Antimalware programs will persistently scan a system to thwart, detect, quarantine, and eradicate potentially unwanted programs (PUPs) and code

NGAV



Known malware prevention
To weed out the obvious

Machine learning
To detect and prevent known and unknown malware – whether endpoints are on or off the network

Exploit mitigation
To stop attacks that exploit vulnerabilities to compromise hosts

Whitelisting/blacklisting
To block or allow custom hashes

Indicators of attacks
To correlate endpoint events to detect stealthy activities that indicate malicious activity

ML and AI-based Tools

- ML security tools can identify patterns and forecast threats in huge data sets at machine speed
- By automating the analysis, blue security teams can quickly discover active parts of the kill chain, incursion artifacts, and indicators of compromise
- ML tools then isolate scenarios in hypervisors or the cloud that need further human analysis

ML and AI-based Tools

- Machine learning detects threats by continually monitoring the activities of the network for anomalies
- ML engines process huge volumes of data in near real-time to expose critical incidents and vulnerabilities
- ML and AI techniques enable advanced visibility into insider threats, unknown malware, and policy violations
- Cloud services like AWS GuardDuty use ML and AI to perform pre-crime activities against domains and IP prefixes 7-10 days in advance of the release of zero days

Business Continuity Planning and Exercises

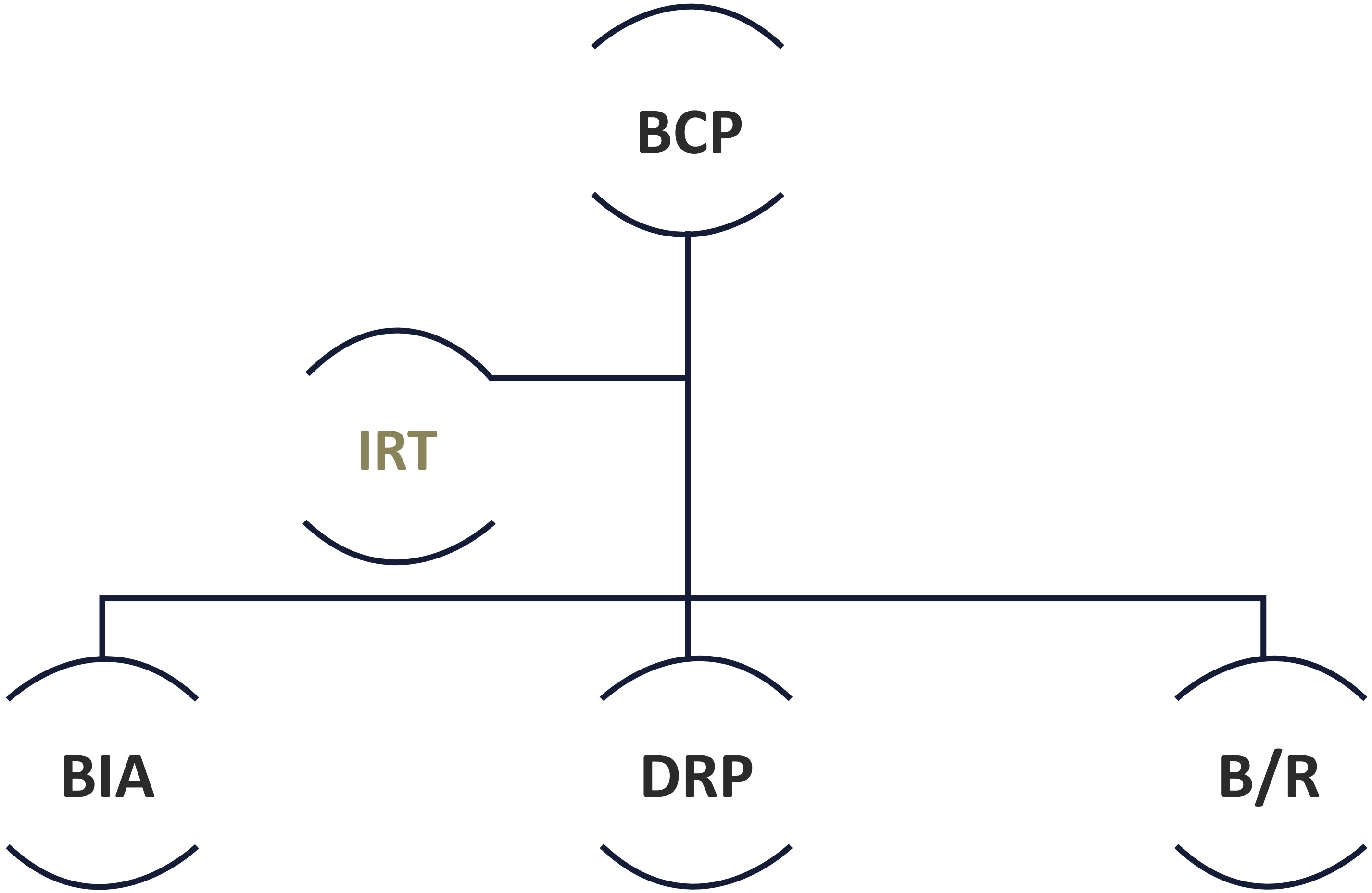
Objectives

- Describe business continuity requirements and business impact analysis (BIA)
- Explore backup storage and recovery site strategies
- Compare multiple processing sites
- Examine system resilience, high availability, quality of service, and fault tolerance
- Explain business continuity planning and exercises

Business Continuity Planning (BCP)

- For government agencies and other non-commercial entities, the term continuity of operations planning (COOP) is often used instead
- BCP involves the development and preparation of all activities and procedures deployed to avert the loss of critical business functions and services for a pre-determined acceptable amount of time
- It is often driven by an initial gap analysis that is an aspect of corporate security strategy and governance
- **The first step is to define the scope of the BCP**

Business Continuity Planning



BCP – NIST SP 800-34, Rev 1

1. Develop a continuity planning policy statement
2. Conduct the business impact analysis
3. Identify preventive controls
4. Create contingency strategies
5. Develop an information system contingency plan
6. Ensure plan testing, training, and exercises
7. After-action report
8. Ensure plan maintenance

Threats to Continuity

Environmental

- Earthquakes
- Wildfires
- Flooding
- Snow
- Tsunamis
- Hurricanes
- Tornadoes
- Landslides
- Asteroids

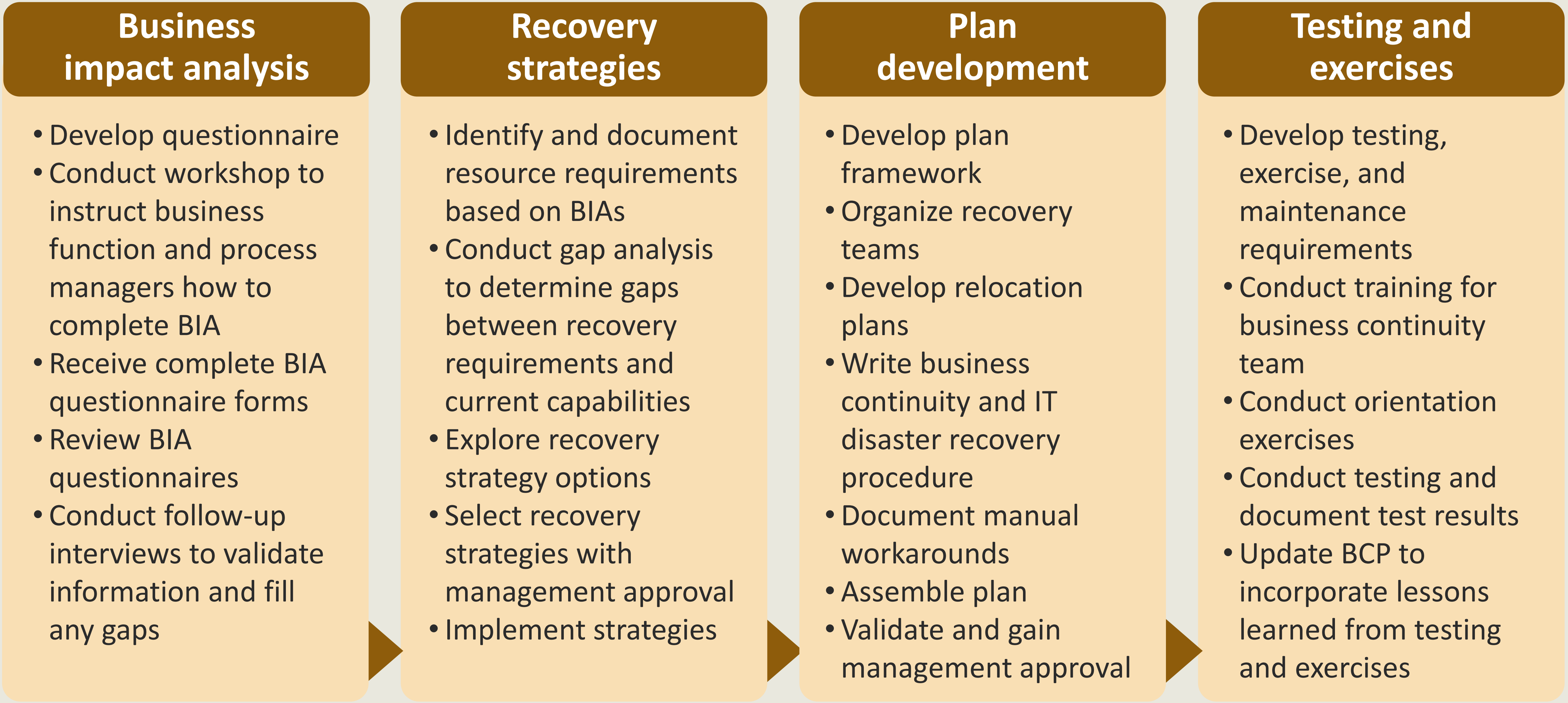
Man-made intentional

- Arson
- Terrorist
- Political
- Break-ins
- Theft
- Damage
- File destruction
- Information disclosure

Man-made unintentional

- Configuration errors and mistakes
- Power outage
- Illness
- Epidemics
- Information disclosure
- Damage
- File destruction
- Coding errors

BCP from ready.gov (U.S. DHS)



Business impact analysis

- Develop questionnaire
- Conduct workshop to instruct business function and process managers how to complete BIA
- Receive complete BIA questionnaire forms
- Review BIA questionnaires
- Conduct follow-up interviews to validate information and fill any gaps

Recovery strategies

- Identify and document resource requirements based on BIAs
- Conduct gap analysis to determine gaps between recovery requirements and current capabilities
- Explore recovery strategy options
- Select recovery strategies with management approval
- Implement strategies

Plan development

- Develop plan framework
- Organize recovery teams
- Develop relocation plans
- Write business continuity and IT disaster recovery procedure
- Document manual workarounds
- Assemble plan
- Validate and gain management approval

Testing and exercises

- Develop testing, exercise, and maintenance requirements
- Conduct training for business continuity team
- Conduct orientation exercises
- Conduct testing and document test results
- Update BCP to incorporate lessons learned from testing and exercises

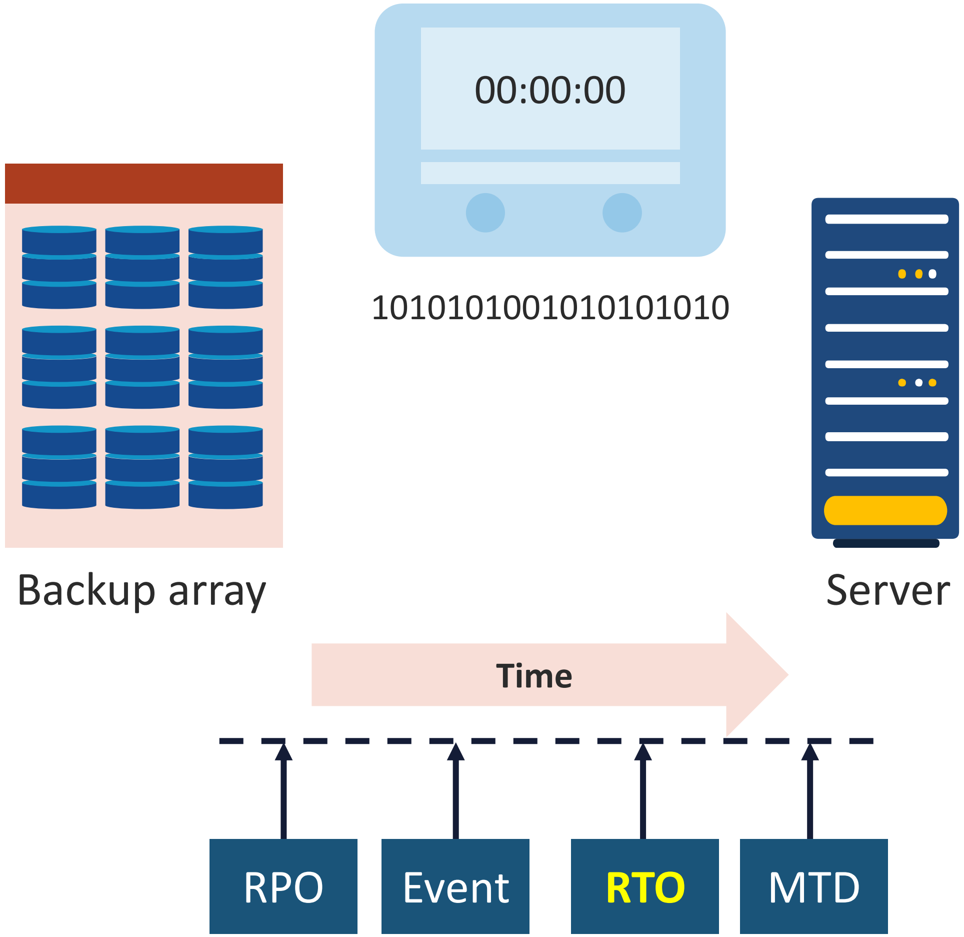
Business Impact Analysis

- The BIA is the quantitative risk assessment aspect of the BCP
- The primary goal is to identify critical functions to the business and prioritize them based on need for survival and continuity
- Groups must gather meaningful metrics and indicators and categorize the risks associated with the critical functions
  - The probability of the risk occurring (likelihood)
  - The impact the risk will have (magnitude)
  - Identify how to eliminate or reduce the risk

Key BIA Metrics and Risk Indicators

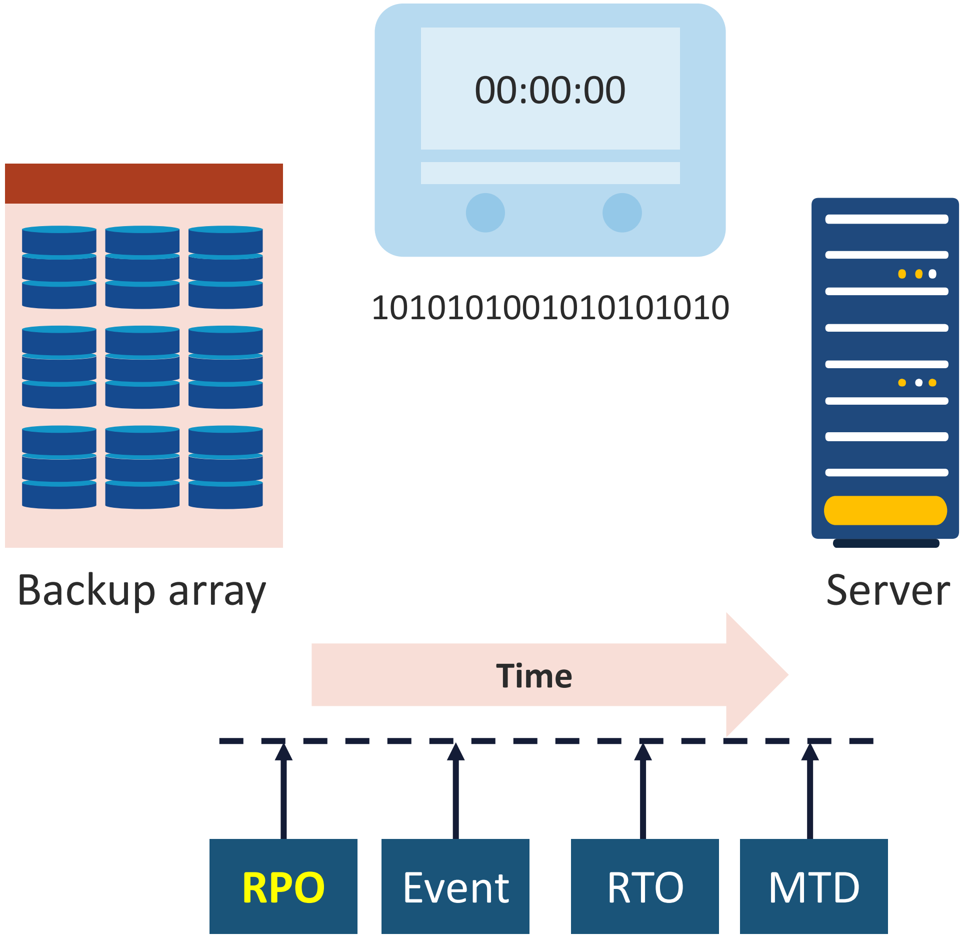
- **Recovery time objective (RTO)**
  - The target amount of time within which a process must be restored after disruption
- **Maximum Tolerable Downtime (MTD)**
  - Absolute maximum amount of time that a resource, service, or function can be unavailable before we start to experience a loss
- **Recovery point objective (RPO)**
  - The maximum targeted point where a recovery activity was conducted relative to the incident timestamp
- **Mean time to repair (MTTR)**
  - The average time needed to repair or replace a failed system or module
- **Mean time between failures (MTBF)**
  - The number of failures per million hours for a product

Recovery Time Objective (RTO)



- The amount of time available to recover the resource, service, and function
  - IT must be equal to or less than MTD
- Any BIA solutions must be accomplished within this time frame, or it is considered loss
- Ways to lower the impact of disasters include:
  - Hardening physical security
  - Adding redundancy
  - Purchasing insurance
  - Investing in bigger generators
  - Utilizing faster and more robust supply chains
  - Safeguarding media off-site or in the cloud

Recovery Point Objective (RPO)



- The point in time, relative to a disaster, where the recovery process begins
- In IT systems, it is often the point in time when the last successful backup or transaction log was done before the disruptive event occurs
- Also including:
  - Last Known Good configuration
  - Image or instance snapshot
  - Recovery volume update
  - State Machine

Mean Time between Failure

- MTBF is a measure of the dependability of a component or system
  - The average time that it will operate before it fails
- MTBF is a maintenance metric, represented in hours usually gathered from the OEM or reliable reviewer
  - It is calculated as the arithmetic mean time between failures of a system
  - For example, a solid-state drive (SSD) drive may have a mean time between failures of 10 years
- MTBF is also used to evaluate performance, safety, and component design

Mean Time to Replace (or Repair)

- MTTR considerations
  - How is this metric affected by supply chain disruptions?
  - How long does it take to repair, fix, or repair a system, device, or component?
- It is the average value predicted based on experience and documentation
- MTTR = (Total down time)/(number of breakdowns)

BIA External Dependencies

- Supply chain
- Vendors and manufacturers
- Warehousers
- Distributors
- Large customers
- Strategic partners (B2B)
- Service-level (master) agreement
- Reciprocal agreement
- Regulators

Full Backups

- The process backs up everything regardless of whether the archive bit is set or not
    - Clears the archive bit once the backup completes
- This method takes the longest to back up and the time depends on how much must be backed up
- A full backup is quickest to restore as only the most recent full backup is required
- A full backup should be scheduled, automated, and tested although it is common to perform this manually

Incremental Backups

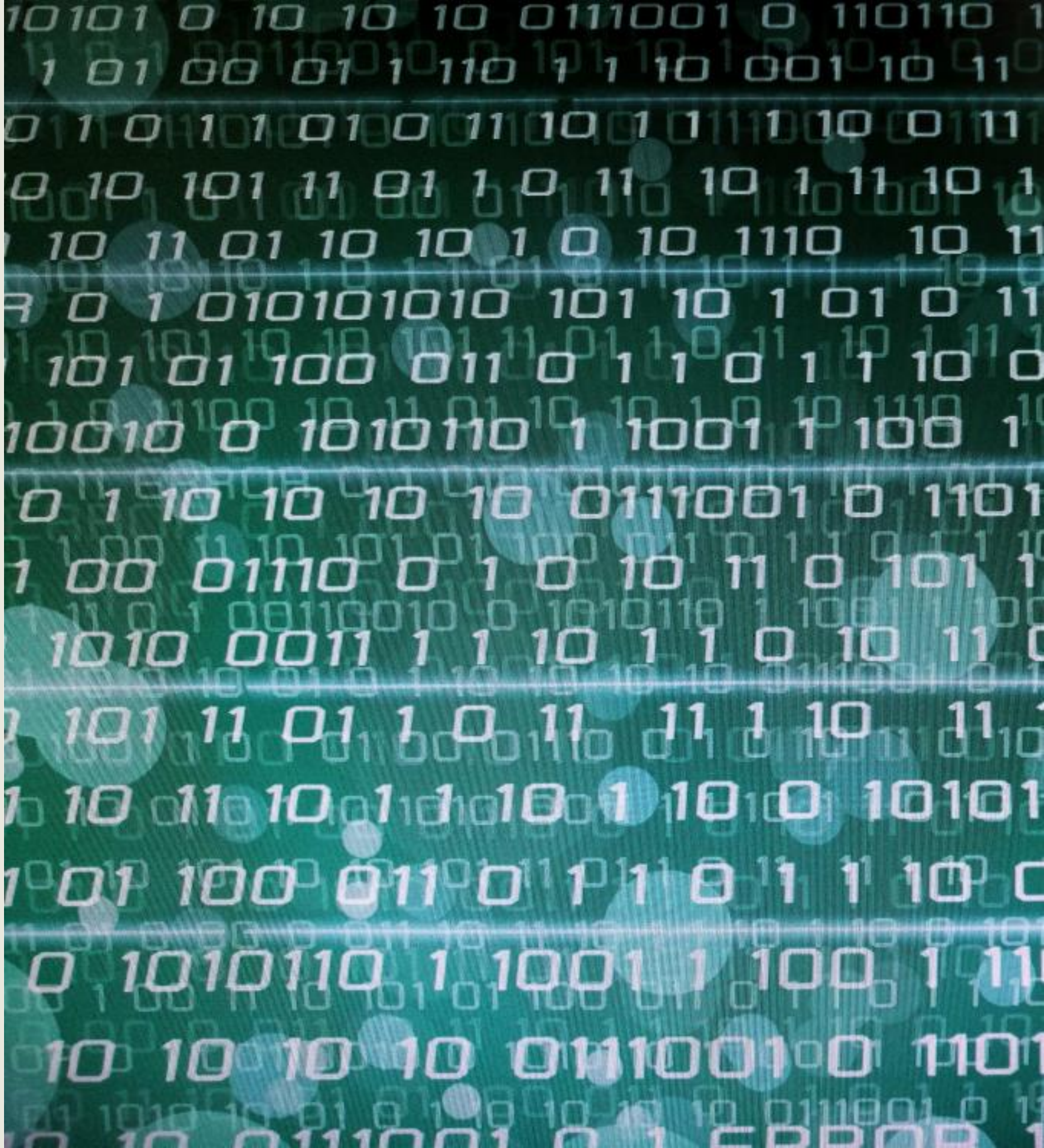
- This method backs up any new file or any file that has changed since
  - The last full backup
  - The last incremental backup
- Subsequent backups only store changes that were made since the previous backup
- An incremental backup clears the archive bit once the backup completes
- The process of restoring lost data from an incremental backup is longer, but the backup process is much quicker
- It is not recommended to perform incremental backups manually

Differential Backups

- This method backs up any file that has the archive bit set
- Backs up any new file or any file that has changed since the last full backup
- A differential back up DOES NOT clear the archive bit when the backup completes
- It is slow to back up but quick to restore
- The last full backup and the most recent differential backup are needed for restoration
- It is not recommended to perform differential backups manually

Snapshots

- Snapshots are immediate point-in-time virtual copies of the source data
- Offers easier and faster backups and restores
- Should be replicated to another medium or cloud storage to be considered a backup
- Time to back up does not increase with amount of data
- Improved RTO and RPO
- Restores are fast
- Less data is lost with an outage
- Can easily be encrypted and decrypted



Backup Frequency

- Backup frequency is often based on the business impact analysis metric known as RPO
  - RPO is the maximum amount of data loss that you can tolerate in case of a disaster
  - The lower the RPO, the more frequently you need to back up your data
- The type of database management system (DBMS), data volume, data change rate, and performance needs all contribute to deciding the best backup strategy

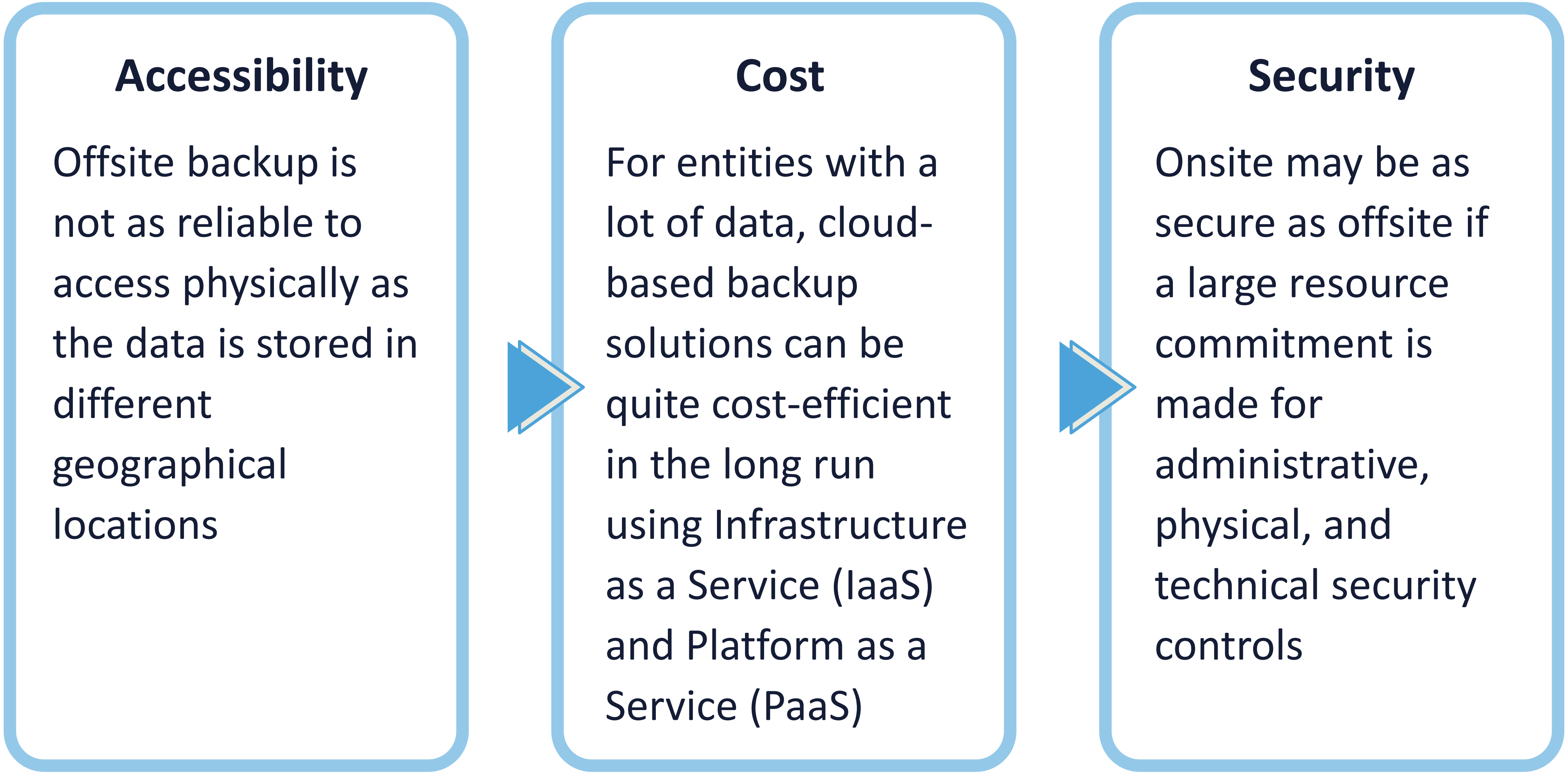
Backup Frequency

- Commonly, full backups are conducted automatically or manually at least once a week, or more frequently depending on the criticality or latency of the data
- Differential backups should be done daily if the RPO is low or the data changes regularly
- Incremental backups should be done hourly if the RPO is very low or the data changes very rapidly
- Snapshots are common techniques for virtual data and should also be automated and scheduled based on various recovery points and time objectives

RAID Comparisons

Features	RAID 0	RAID 1	RAID 5	RAID 6	RAID 10
Minimum number of drives	2	2	3	4	4
Fault tolerance	None	Single-drive failure	Single-drive failure	Two-drive failure	Up to one disk failure in each sub-array
Read performance	High	Medium	Low	Low	High
Write performance	High	Medium	Low	Low	Medium
Capacity utilization	100%	50%	67-94%	50-88%	50%

Onsite vs. Offsite Backup Strategies



Accessibility

Offsite backup is not as reliable to access physically as the data is stored in different geographical locations

Cost

For entities with a lot of data, cloud-based backup solutions can be quite cost-efficient in the long run using Infrastructure as a Service (IaaS) and Platform as a Service (PaaS)

Security

Onsite may be as secure as offsite if a large resource commitment is made for administrative, physical, and technical security controls

Onsite vs. Offsite Backup Strategies

Scalability

Scalability is one of the huge advantages of offsite data backup where the cloud service provider (CSP) is responsible for providing the storage

Support and maintenance

With on-premises solutions, the organization has the most control with their own support team responsible for data backup

Reliability

Offsite data backup is more reliable because the data is not stored in the same place as the original data

Recovery and Restoration

- Without a comprehensive well-tested recovery and restoration practice there is no real backup strategy
- Many organizations have relied on regular automated backups when suffering a ransomware attack only to find out there were configuration errors or gaps that were not discovered through ongoing recovery testing
- The team that performs recovery is often different that the backup operators due to Separation of Duties (SOD)

Multipath Connectivity

- Uninterrupted service and constant access are critical to the daily operation and productivity of the enterprise
- Since downtime leads directly to loss of income, datacenters must be designed for redundant, fail-safe reliability and availability
- Datacenter reliability is also defined by the performance of the infrastructure

Multipath Connectivity

- Cabling and connectivity backed by a trustworthy vendor service-level agreement (SLA) with guaranteed error-free performance will help avoid poor data transmission in the datacenter
- There should be redundant connectivity from multiple providers into the datacenter
  - This will help prevent a single point of failure for network connectivity
- The redundant paths should deliver the minimum expected connection speeds (10GB/100GB) for datacenter operations

Resilience

- Enterprise resilience relates to an organization's ability to anticipate and respond to change, survive catastrophic events, and to learn from the experiences
- Resilience is the capacity for weathering systemic interruptions and adjusting to the introduction of emerging risks and vulnerabilities
- It practically refers to an organization's ability to withstand, adapt, and mature in the face of unexpected changes

Design Resilient

	P1	P2	P3	P4	P5
	Multi-AZ deployment	Static stability in region	Application portfolio distribution	Multi-AZ deployment (regional DR)	Multi-region active-active deployment
Design complexity	Low	Medium	Medium	High	Very High
Cost to implement	Low	High	Medium	High	High
Operational effort	Low	Medium	Medium	Medium	Very High
Effort to secure	Low	Medium	Medium	High	High
Environmental impact	Low	Medium	Medium	High	High

Lowest → Highest

Availability

Redundancy

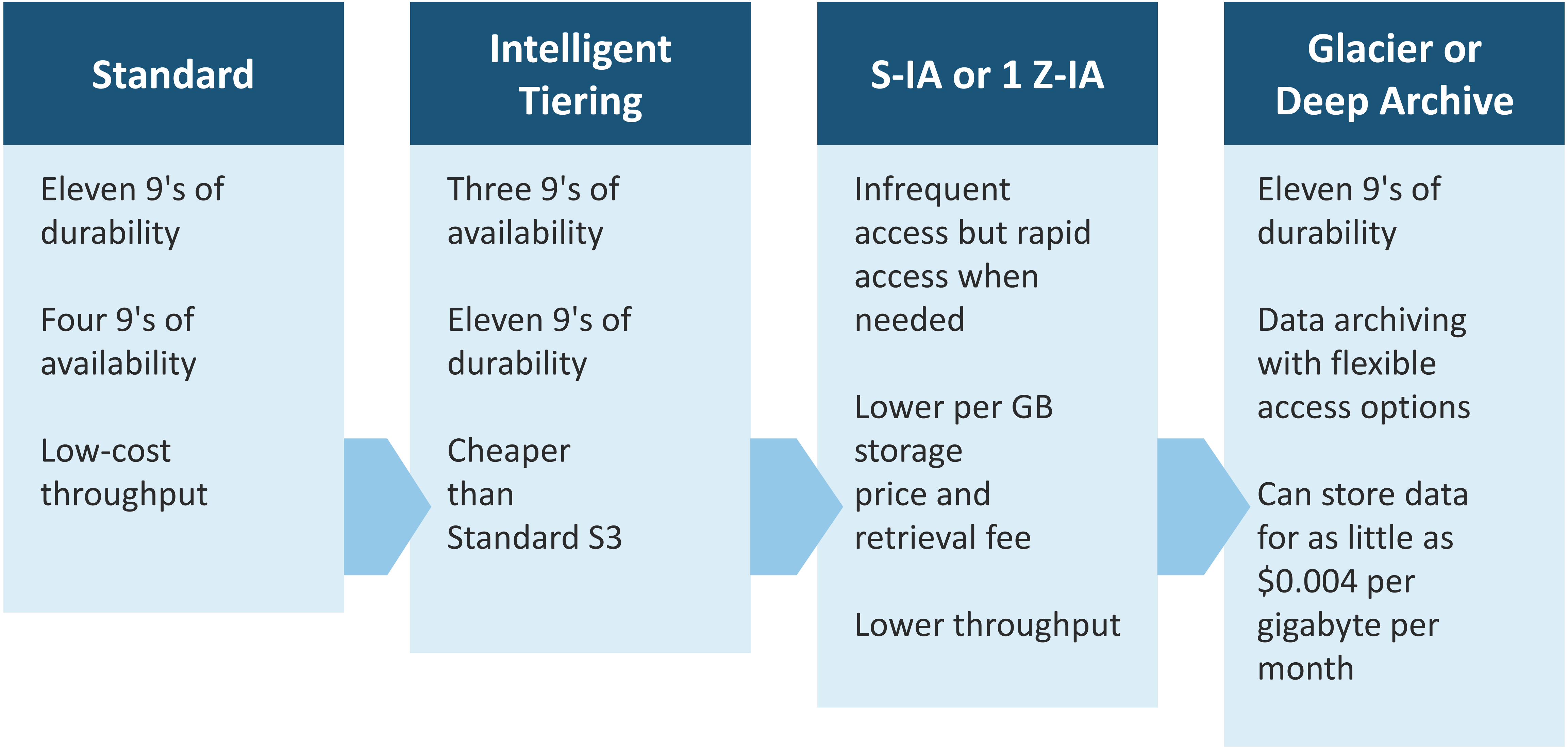
- **Passive** redundancy uses additional capacity to reduce the impact of component failures
  - Active/passive failover
  - Hot spares
  - Snapshots
- **Active** redundancy eliminates performance problems by having simultaneous capacity in use
  - Active/active failover
  - Hot, mirrored, or parallel sites

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 or services remain

- **Durability**, on the other hand, refers to long-term data protection
- 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

CSP Storage Plans (AWS)



Standard

Eleven 9's of durability

Four 9's of availability

Low-cost throughput

Intelligent Tiering

Three 9's of availability

Eleven 9's of durability

Cheaper than Standard S3

S-IA or 1 Z-IA

Infrequent access but rapid access when needed

Lower per GB storage price and retrieval fee

Lower throughput

Glacier or Deep Archive

Eleven 9's of durability

Data archiving with flexible access options

Can store data for as little as $0.004 per gigabyte per month

Quality of Service (QOS)

- QoS is commonly a metric used to evaluate the reliability of the delivery of frames, cells, and packets over wired and wireless networks
- The steady and high-quality service availability is a critical concern of network administrators for critical applications and protocols
- As applications, apps, and devices demand more bandwidth, QoS is a vital tool to assure a predictable level of availability (CIA)

Quality of Service

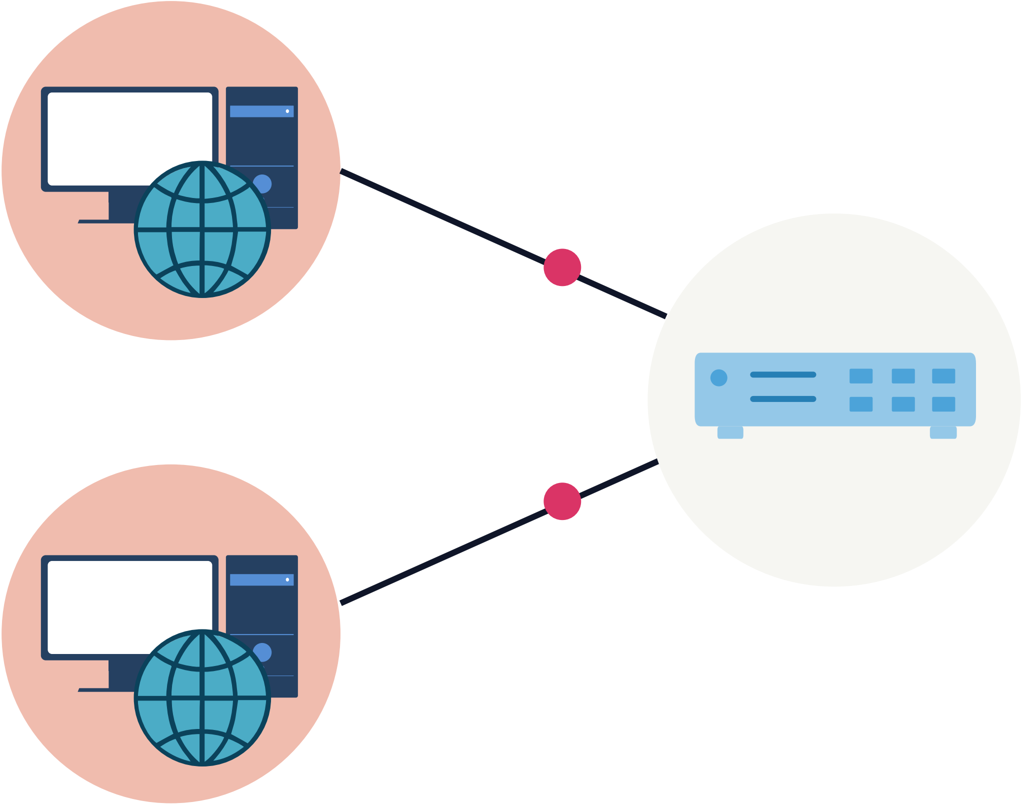
- One of the main objectives in preserving QoS is making sure that critical or time-sensitive applications receive priority over other traffic
- For example, voice over IP (VoIP), conferencing, and streaming packets should get higher priority than normal data transfers
- Users should not experience dropped packets, latency, or delays

Load Balancing

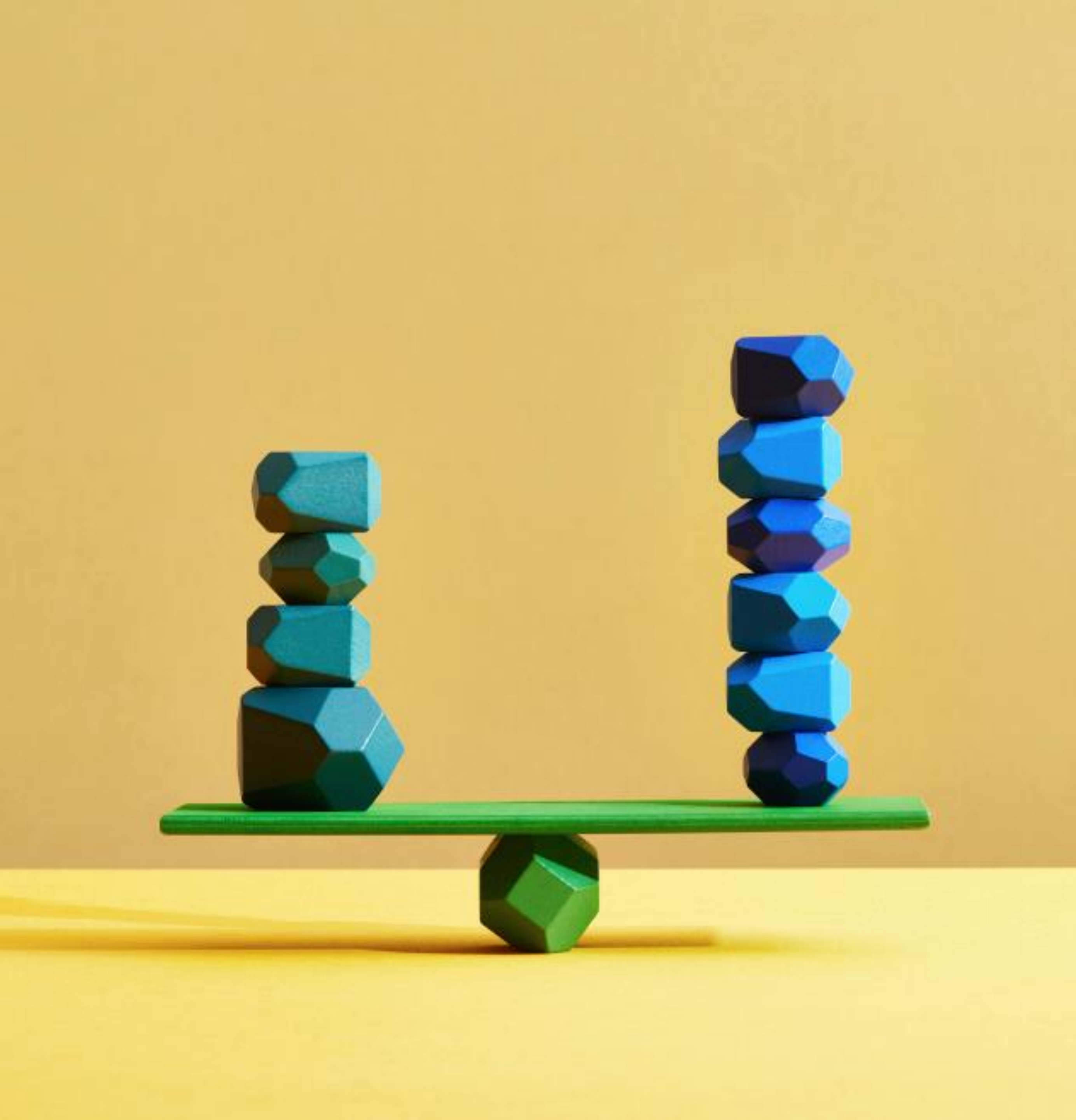
- Load balancing devices and services are popular due to the usage of data and network intensive applications and services
- They can optimize application availability and performance
- They distribute Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Hypertext Transfer Protocol (HTTP), and Transport Layer Security (TLS) traffic across multiple servers to efficiently allocate resources and offer failover solutions

Load Balancing

- Dedicated load balancing appliances and modules have become a standard component in physical and virtual networks
- All the major network equipment vendors offer load balancing solutions to basically "put traffic in its place"
- These systems can optimize application availability and performance, distribute traffic across multiple servers, and offer failover solutions



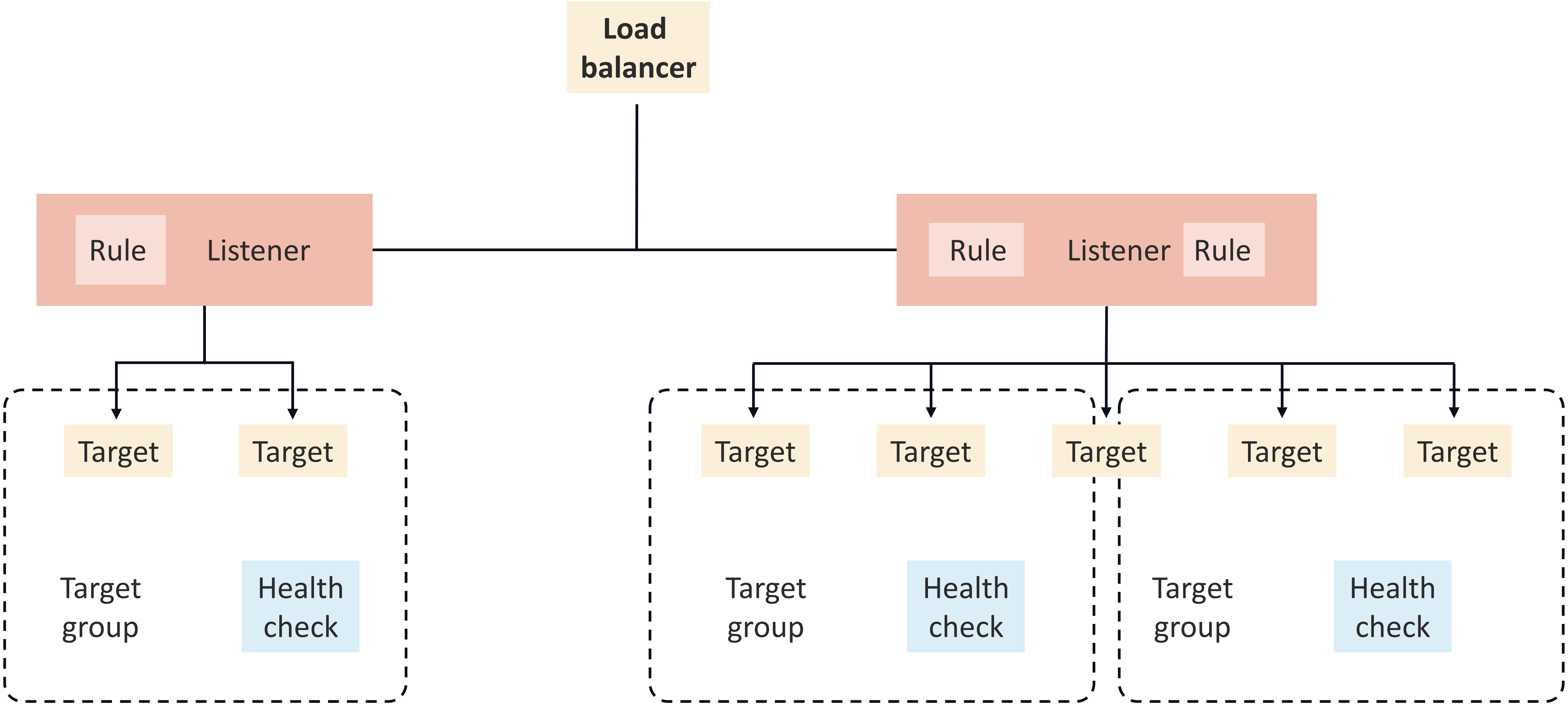
Load balancer



Load Balancing at Cloud Providers

- Network or application load balancing
- Often represents virtual network to the public based on IP address or public domain name
- Performs health checks on back-end instances and containers
- Produces flow logs for other threat management services
- Runs the TLS listener to decrypt traffic
- Can also have layer 3/4 and web application firewall (web access control list (ACL))

Cloud Load Balancers

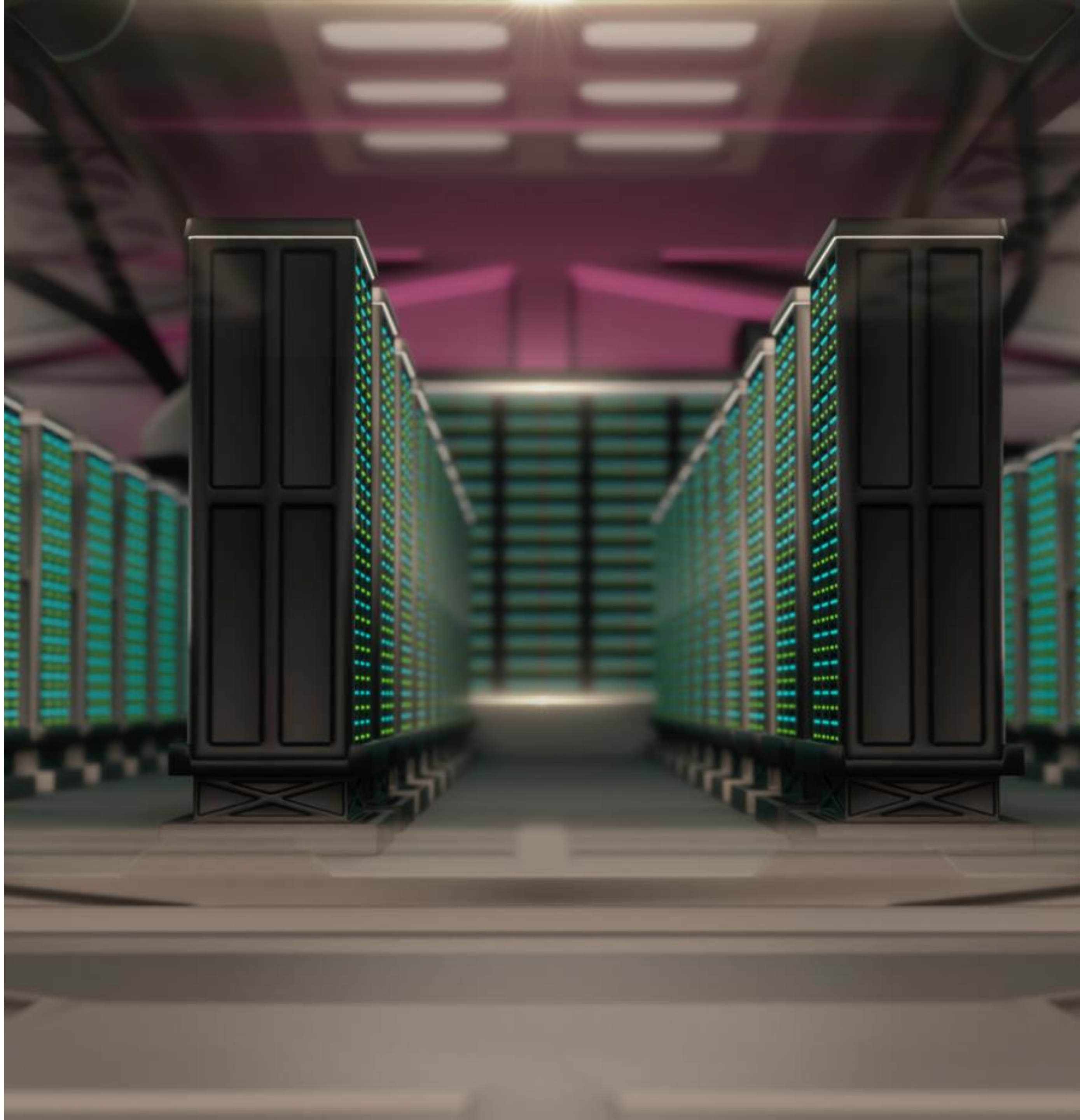


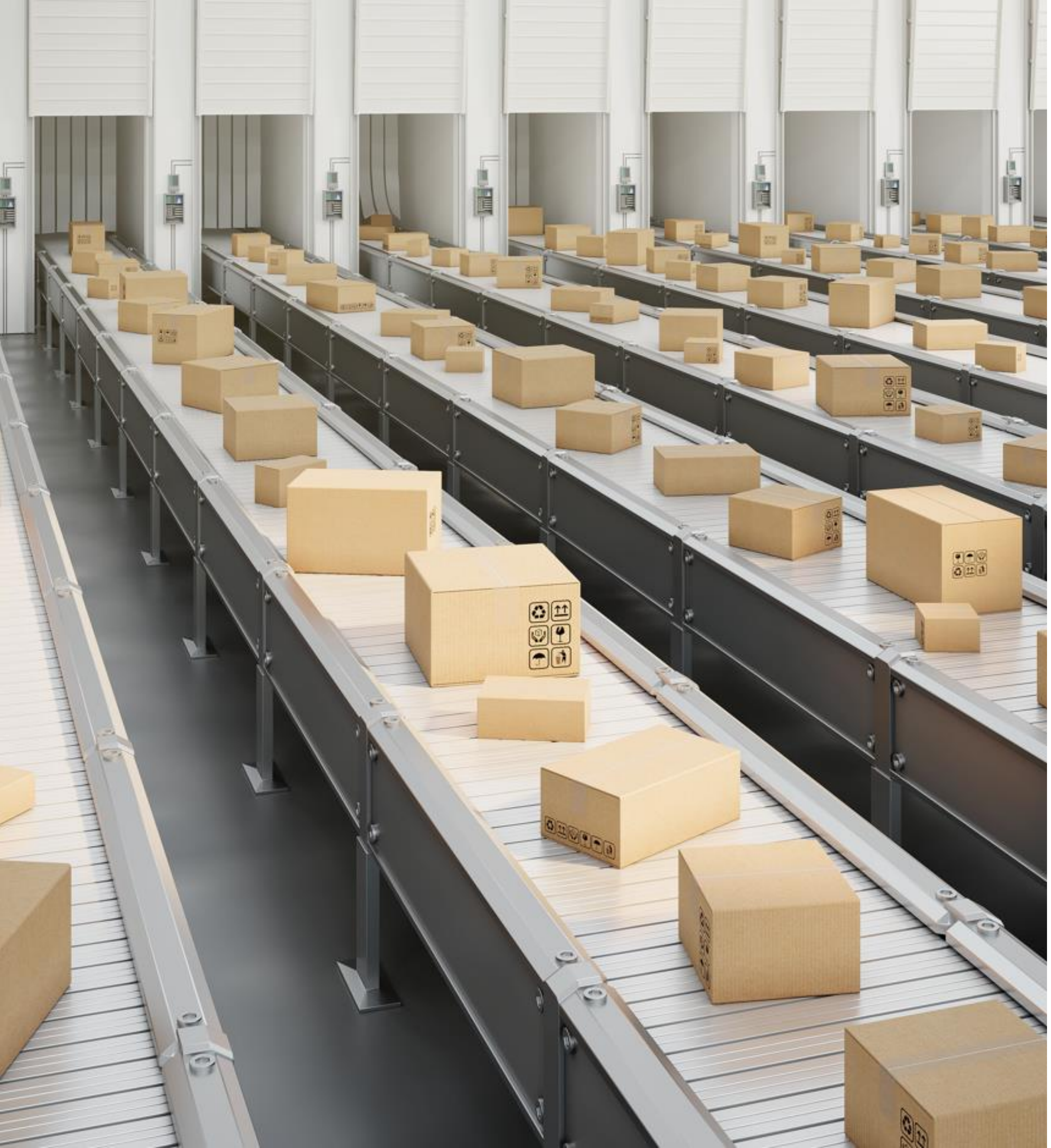
Clustering

- A primary target of modern load balancers is a cluster
- Clustering is intended to improve performance and availability of a complex physical or virtual system
- Clusters are designed to be a redundant set of service functionalities based on active-standby or active-active deployments

Clustering

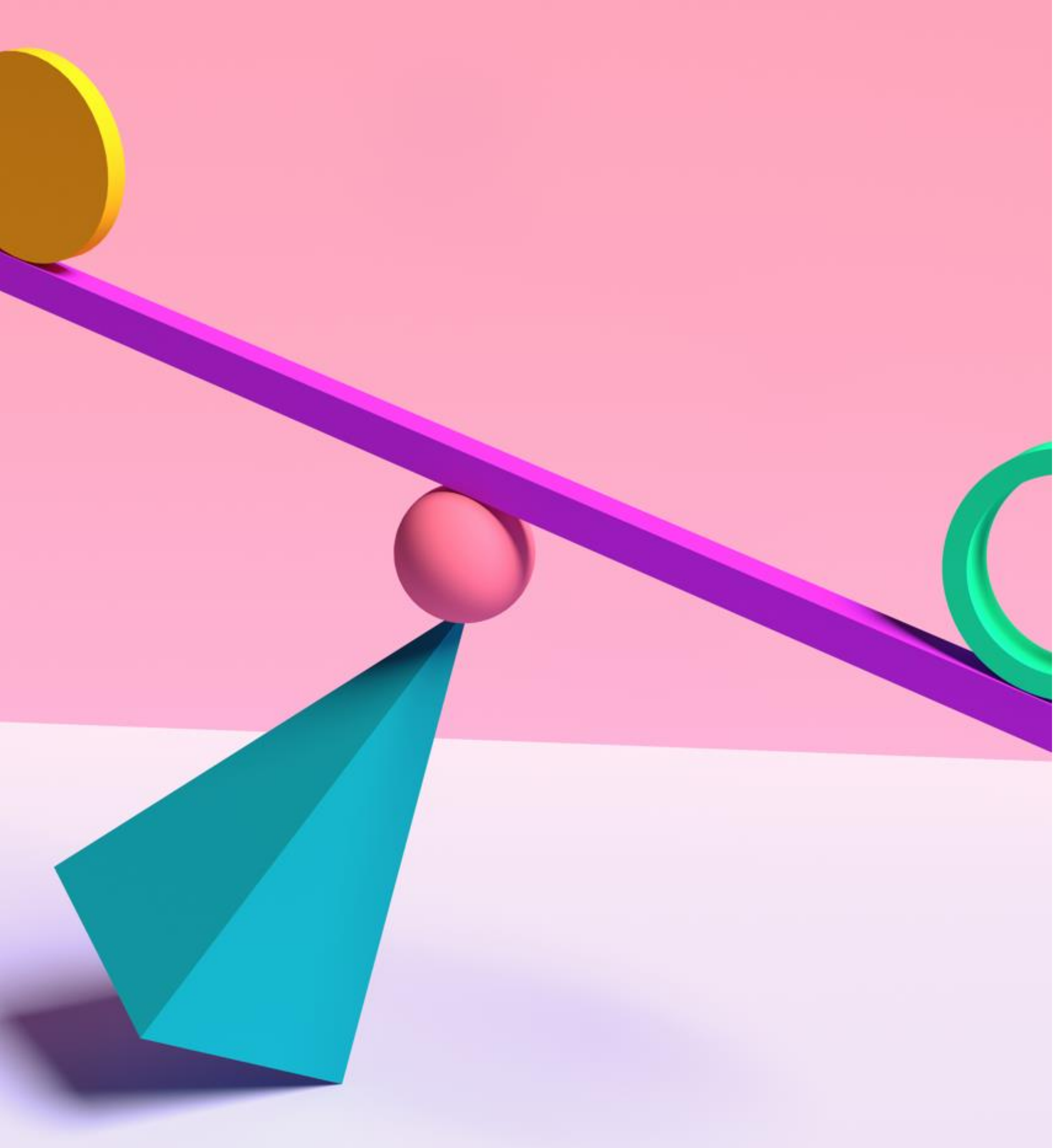
- Cluster deployments are often measured by:
  - Reliability – the ability to successfully provide responses on each incoming request
  - Availability – the uptime of the server (usually measured as % of annual uptime)
  - Performance – measured by the average of the time spent by the service to provide responses or by the throughput
  - Scalability – the ability to handle a growing amount of work in a capable manner without degradation in the quality of service





Clustering vs. Load Balancing

- Server clustering combines multiple servers and containers to operate as a single physical and/or virtual entity
- Load balancing distributes a workload across multiple servers to improve performance
- Both load balancing and server clustering technologies are often used together to coordinate multiple servers to handle a larger workload
- Server clusters typically require identical hardware and versioning to function optimally
- Load balancers can be used to distribute workload to different types of servers and can be more easily integrated into existing architecture



Clustering vs. Load Balancing

- These solutions have several common attributes:
  - To external devices, both technologies typically appear to be a single system that manages all requests
  - Both technologies often integrate reverse-proxy techniques that allow for a single IP address to redirect traffic to different IP or MAC addresses
  - Both were developed for managing a data center's physical servers but have been extended to applications, virtual servers, cloud servers, and containers

Clustering Techniques

- **High availability clusters** prioritize resilience over other advantages and can be implemented in either active-passive or active-active architecture
- **Load balancing clusters** highlight balancing the jobs among all the servers in the cluster and incorporate load balancing software in the controller node

Clustering Techniques

- **High-performance clusters** use multiple servers to execute a specific task very quickly and support data-intensive projects such as live-streaming and real-time data processing
- **Storage clusters** offer massive storage arrays, sometimes in support of high-performance clusters, but always in a support role for other servers or clusters such as storage area networking or hypervisor cluster data stores