AWS 101

Chih-Chieh Hung 洪智傑

aws academy

Chih-Chieh Hung 洪智傑

- **Amazon Academy Program (Accredited Instructor)**
  2018-Current
- **Tamkang University (Assistant Professor)**
  2016-Current
- Rakuten Inc., Japan (Data Scientist)
  2013-2015
- Yahoo! Inc., Taiwan (Research Engineer)
  2011-2013

Cloud Computing Concept

What is Cloud Computing?

Cloud computing is the **on-demand** delivery of compute power, database storage, applications, and other IT resources through a cloud services platform **via the internet** with **pay-as-you-go** pricing.

Before Cloud Computing

Cloud computing enables you to **stop thinking of your infrastructure as hardware**, and instead **think of it (and use it) as software.**

Before Cloud Computing

- Hardware solutions are **physical**. This means they require:
  - Space
  - Staff
  - Physical security
  - Planning
  - Capital expenditure
- Guess at theoretical maximum peaks
  - Is there enough resource capacity?
  - Do we have sufficient storage?

What if your needs change?

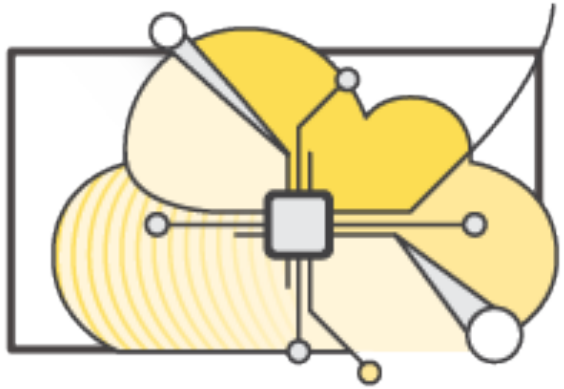
You have to go through the **time, effort, and cost** required to change all these.

Utilizing Cloud Computing

Software is flexible.

If your needs change, your software can change much more **quickly, easily, and cost-effectively** than your hardware.

Three Cloud Deployment Models



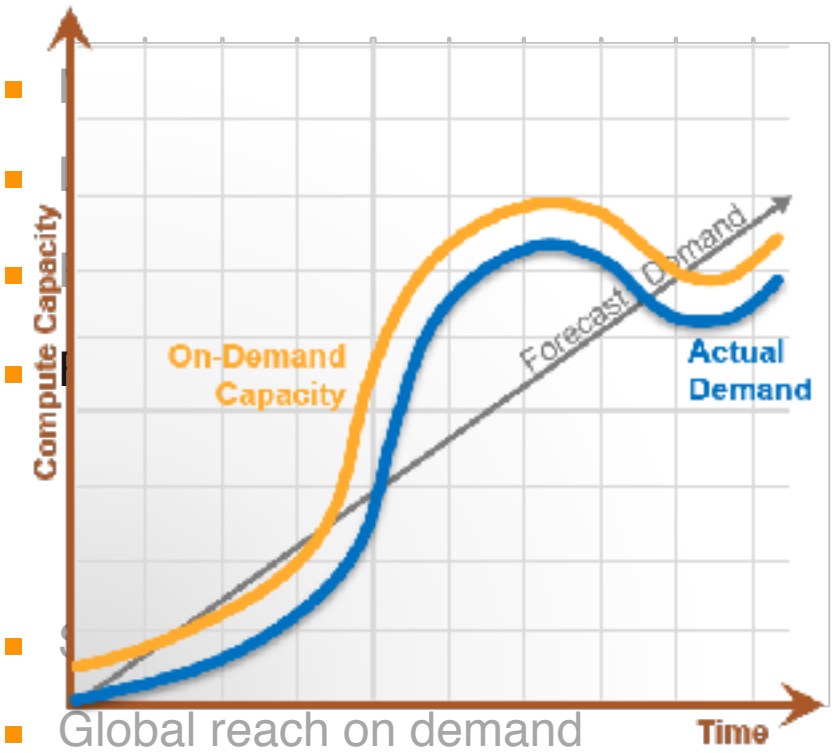
All-In Cloud

Hybrid

Private Cloud (On-premises)

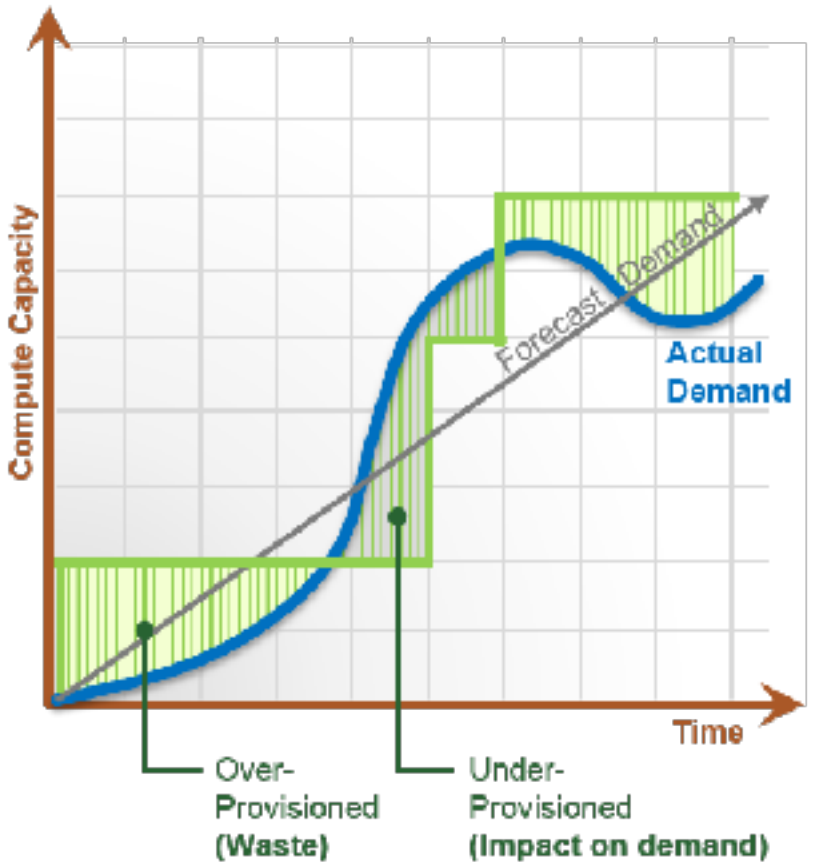
All-In Cloud versus On-Premises

All-In Cloud

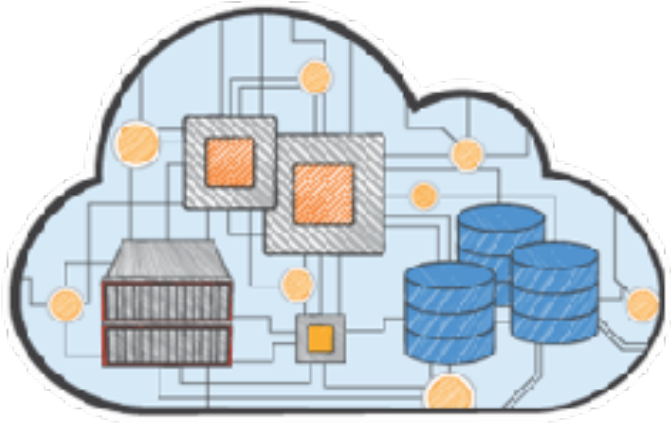


- Global reach on demand
  - ✓ No waste
  - ✓ Meets demand

On-Premises



All-In Cloud versus On-Premises



All-In Cloud

- No upfront investment
- Low ongoing costs
- Focus on innovation
- Flexible capacity
- Speed and agility
- Global reach on demand

On-Premises

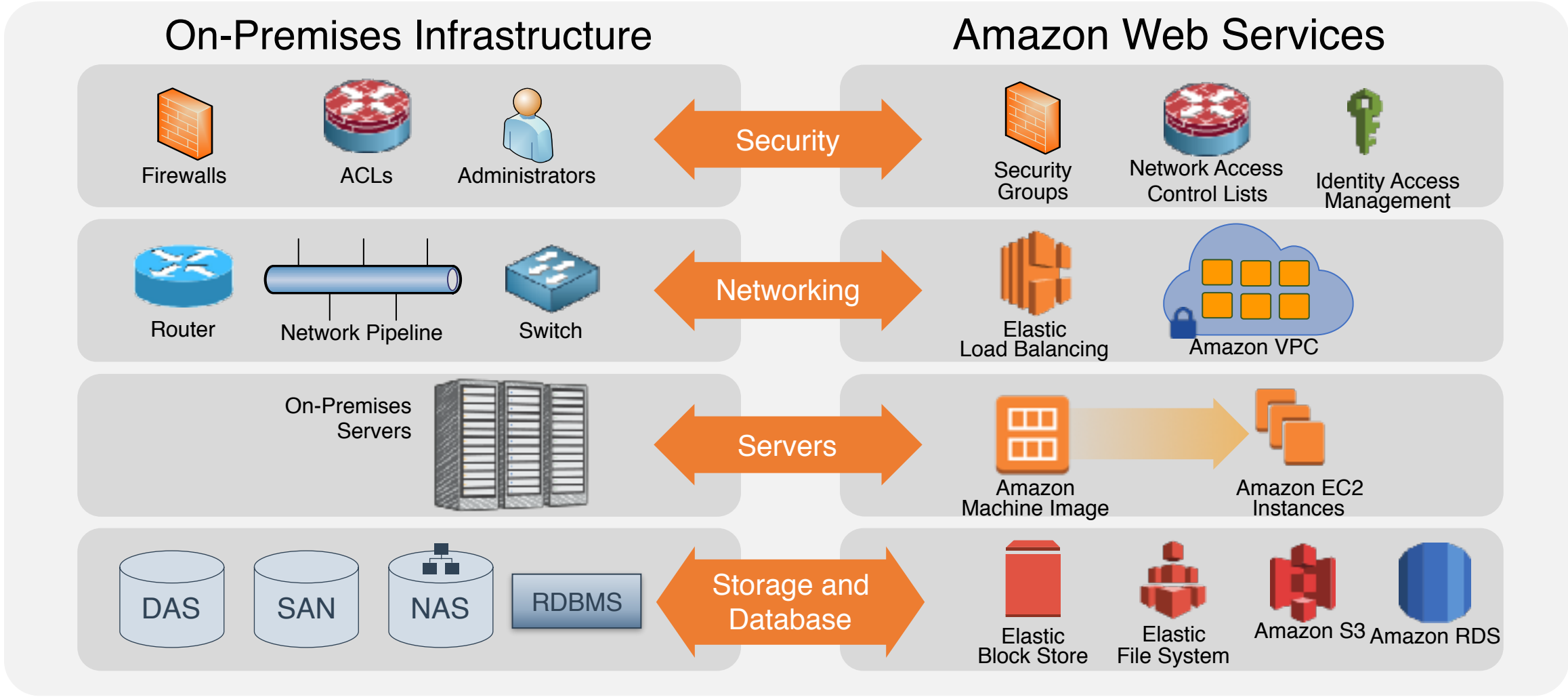
- Large initial purchase
- Labor, patches, and upgrade cycles
- Systems administration
- Fixed capacity
- Long procurement cycle and setup
- Limited geographic regions

What can you do in the cloud?

You can use a cloud computing platform for:

- Application Hosting
- Backup and Storage
- Content Delivery
- Websites
- Enterprise IT
- Databases

On-Premises and AWS Comparison



Important Cloud Terminology

- **High Availability (Highly Available):**
  - Accessible when you need it
- **Fault Tolerance (Fault Tolerant):**
  - Ability to withstand a certain amount of failure and still remain functional
- **Scalability (Scalable):**
  - Ability to easily grow in size, capacity, and/or scope when required
  - Growth is (usually) based on demand
- **Elasticity (Elastic):**
  - Ability to grow (scale) when required and to reduce in size when resources are no longer needed

Advantage #1: Capex to Variable Expense

Trade **capital expense** for **variable expense**.

Advantage #2: Economies of Scale

Benefit from **massive economies of scale**.

Advantage #3: Capacity Planning

Eliminate guessing on your capacity needs.

Advantage #4: Speed and Agility

Increase **speed** and **agility**.

Advantage #5: Spend Strategically

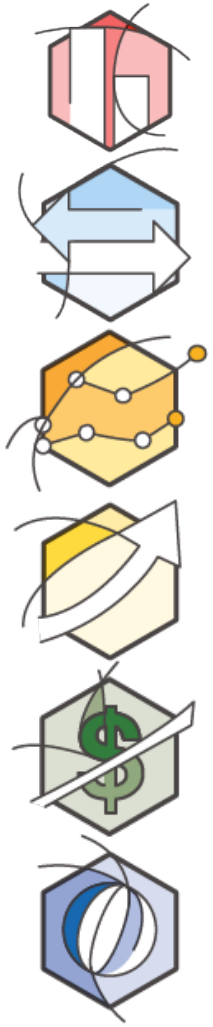
Stop spending money on running and maintaining data centers.

Advantage #6: Ease of Deployment

Go global in minutes.

Go Global in Minutes

Summary



Trade **capital expense** for **variable expense**.

Benefit from **massive economies of scale**.

Eliminate guessing on your capacity needs.

Increase **speed** and **agility**.

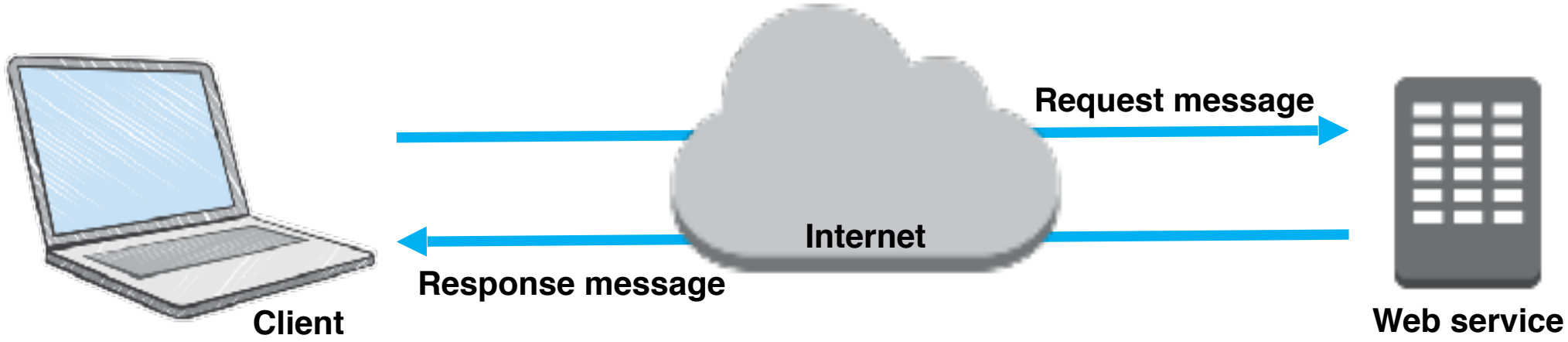
Stop spending money to run and maintain data centers.

Go global in minutes.

What is Amazon Web Services (AWS)?

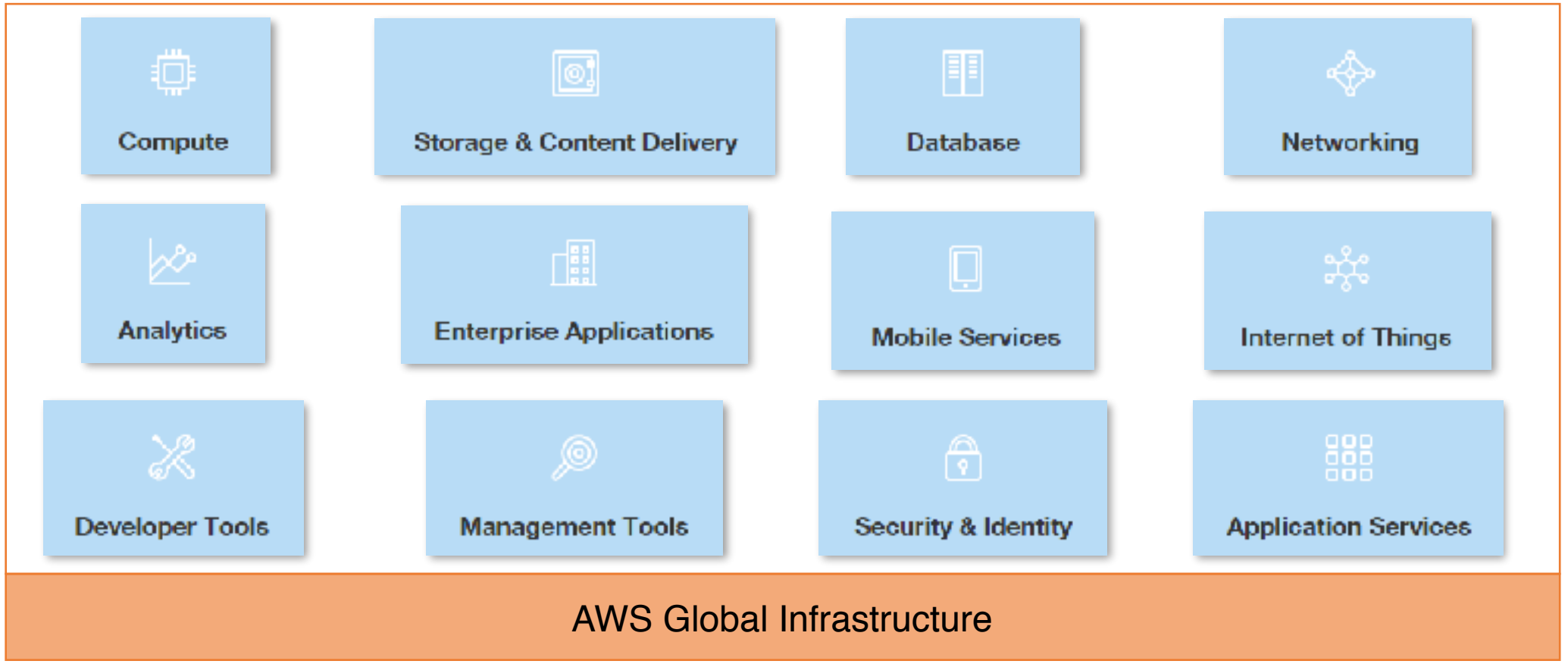
What are Web Services?

A **web service** is any piece of software that makes itself available over the internet and uses a **standardized format** (XML or JSON) for the request and the response of an **API interaction**.



What is AWS?

AWS is a secure cloud services provider with more than 50 different services that include solutions for:



AWS by Category: Core Services

Compute


Amazon EC2


AWS Lambda


Auto Scaling


AWS Elastic Beanstalk


Amazon Elastic Container Registry


Amazon Elastic Container Service


Amazon Lightsail


AWS Batch

Networking


Amazon VPC


Amazon Route 53

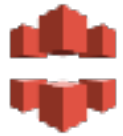

AWS Direct Connect


Elastic Load Balancing

Storage


Amazon S3


Amazon EBS


Amazon CloudFront


Amazon Glacier


Amazon Elastic File System


AWS Snowball


Storage Gateway


AWS Snowmobile

Database


Amazon RDS


Amazon DynamoDB


Amazon Redshift

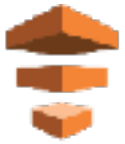

AWS Database Migration Service


Amazon ElastiCache

AWS by Category: Foundational Services

Analytics


Amazon EMR


AWS Data Pipeline


Amazon Elasticsearch


Amazon Kinesis


Amazon Machine Learning


Amazon QuickSight


Amazon Redshift


Amazon Athena

Enterprise Apps


Amazon WorkSpaces


Amazon WorkMail


Amazon WorkDocs

Mobile Services


AWS Mobile Hub


Amazon SNS


Amazon Cognito


AWS Device Farm


Amazon Mobile Analytics


AWS Mobile SDKs


Amazon Pinpoint

Internet of Things


AWS IoT


AWS Greengrass

AWS by Category: Developer and Operations Services

aws academy

Developer Tools


AWS CodeCommit


AWS CodeDeploy


AWS CodePipeline


AWS CodeBuild


AWS X-Ray

Management Tools


Amazon CloudWatch


AWS CloudFormation


AWS CloudTrail


AWS Config


AWS OpsWorks


AWS Service Catalog


AWS Trusted Advisor


AWS Organizations

Security & Identity


AWS Identity and Access Management


AWS Directory Service


Amazon Inspector


AWS CloudHSM


AWS Key Management Service


AWS WAF


AWS Certificate Manager


AWS Shield

App Services


Amazon API Gateway


Amazon AppStream


Amazon CloudSearch


Amazon Elastic Transcoder


Amazon SES


Amazon SNS


Amazon SQS


Amazon SWF

Core Services: The Basics

aws academy

Compute


Amazon EC2


AWS Lambda

Auto Scaling


AWS Elastic Beanstalk

Amazon Elastic Container Registry

Amazon Elastic Container Service

Amazon Lightsail

AWS Batch

Networking


Amazon VPC


Amazon Route 53

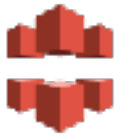
AWS Direct Connect

Elastic Load Balancing

Storage


Amazon S3


Amazon Glacier


Amazon CloudFront

Amazon EBS

Amazon Elastic File System

AWS Snowball

Storage Gateway

AWS Snowmobile

Database


Amazon RDS


Amazon DynamoDB

Amazon Redshift

AWS Database Migration Service

Amazon ElastiCache

Core Services: The Basics

Management Tools

Amazon CloudWatch

AWS CloudFormation

AWS CloudTrail

AWS Config

AWS OpsWorks

AWS Service Catalog

AWS Trusted Advisor

Security & Identity

AWS Identity and Access Management

AWS Directory Service

Amazon Inspector

AWS CloudHSM

AWS Key Management Service

AWS WAF

AWS Organizations

AWS Certificate Manager

AWS Shield

Access to AWS Services

- AWS Management Console
  - Access on the go with AWS Console Mobile App
- AWS Command Line Interface (AWS CLI)
- Software Development Kits (SDK)

AWS Global Infrastructure

AWS Global Infrastructure

Applications

- Virtual Desktops
- Collaboration and Sharing

Platform Services

Databases	Analytics	App Services	Deployment and Management	Mobile Services
Relational	Cluster Computing	Queuing	Containers	Identity
No SQL	Real-time	Orchestration	Dev/ops Tools	Sync
Caching	Data Warehouse	App Streaming	Resource Templates	Mobile Analytics
	Data Workflows	Transcoding	Usage Tracking	Notifications
		Email	Monitoring and Logs	
		Search		

Foundation Services

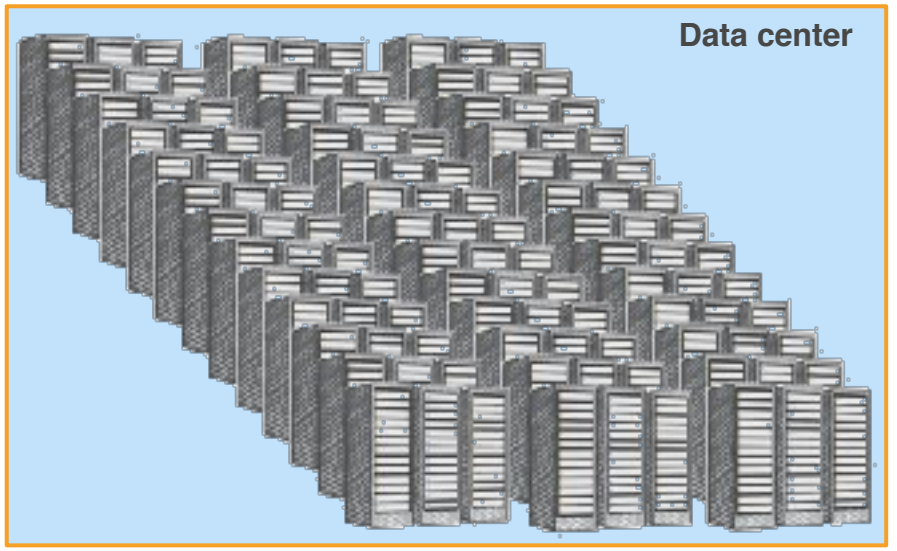
- Compute (Virtual, Auto-scaling and Load Balancing)
- Networking
- Storage (Object, Block and Archive)

Infrastructure

- Regions
- Availability Zones
- Edge Locations

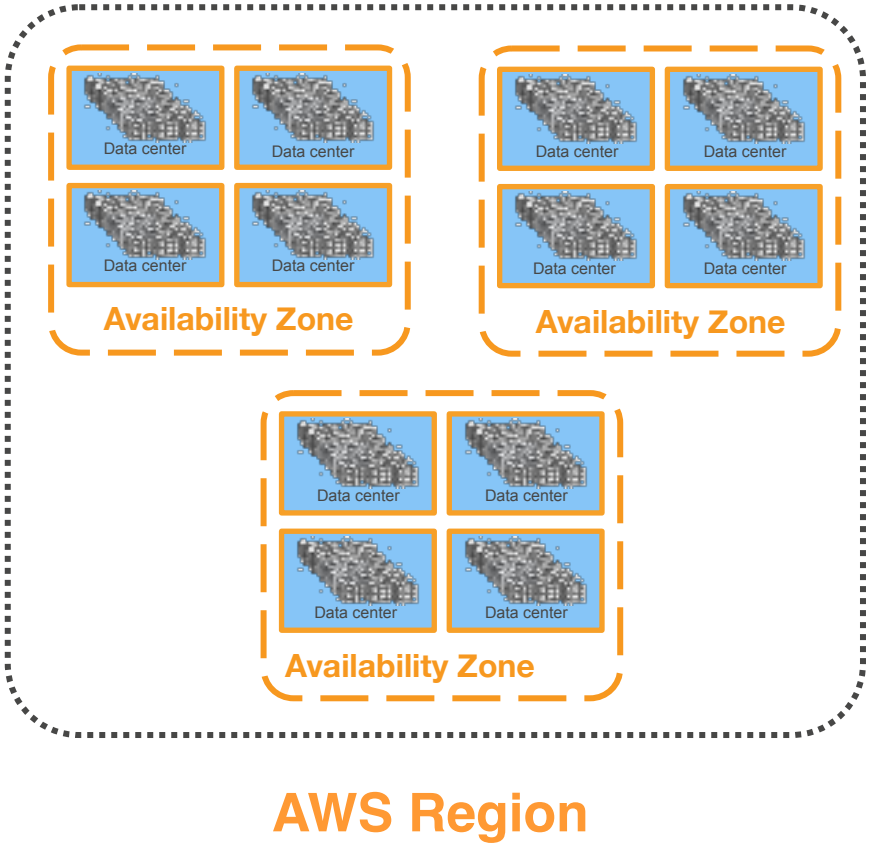
AWS Data Centers

- Data centers are securely designed
- A datacenter is a location where actual physical data resides
- A data center typically has 50,000 to 80,000 physical servers
- All data centers are online. No data center is “cold”
- AWS custom network equipment:
  - Multi-ODM sourced
  - Amazon custom network protocol stack



AWS Regions

- An AWS Region is a **geographical area**.
- Each Region is made up of **two or more Availability Zones**.
- AWS has **18 Regions** worldwide.
- You enable and control **data replication** across Regions.
- Communication between Regions uses **AWS backbone network** connections infrastructure.

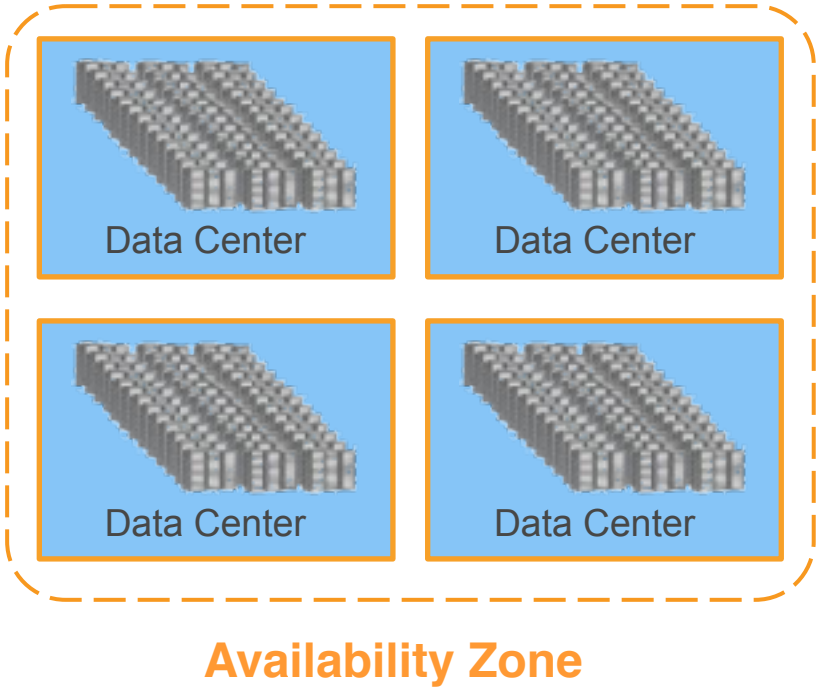


AWS Global Infrastructure: Current Regions

AWS Global Infrastructure: Regions Coming Soon

AWS Availability Zones

- Each Availability Zone is:
  - Made up of **one or more** data centers.
  - Designed for **fault isolation**.
  - Interconnected with other Availability Zones using high-speed **private** links.
- You choose your Availability Zones.
- AWS recommends replicating across Availability Zones for resiliency.

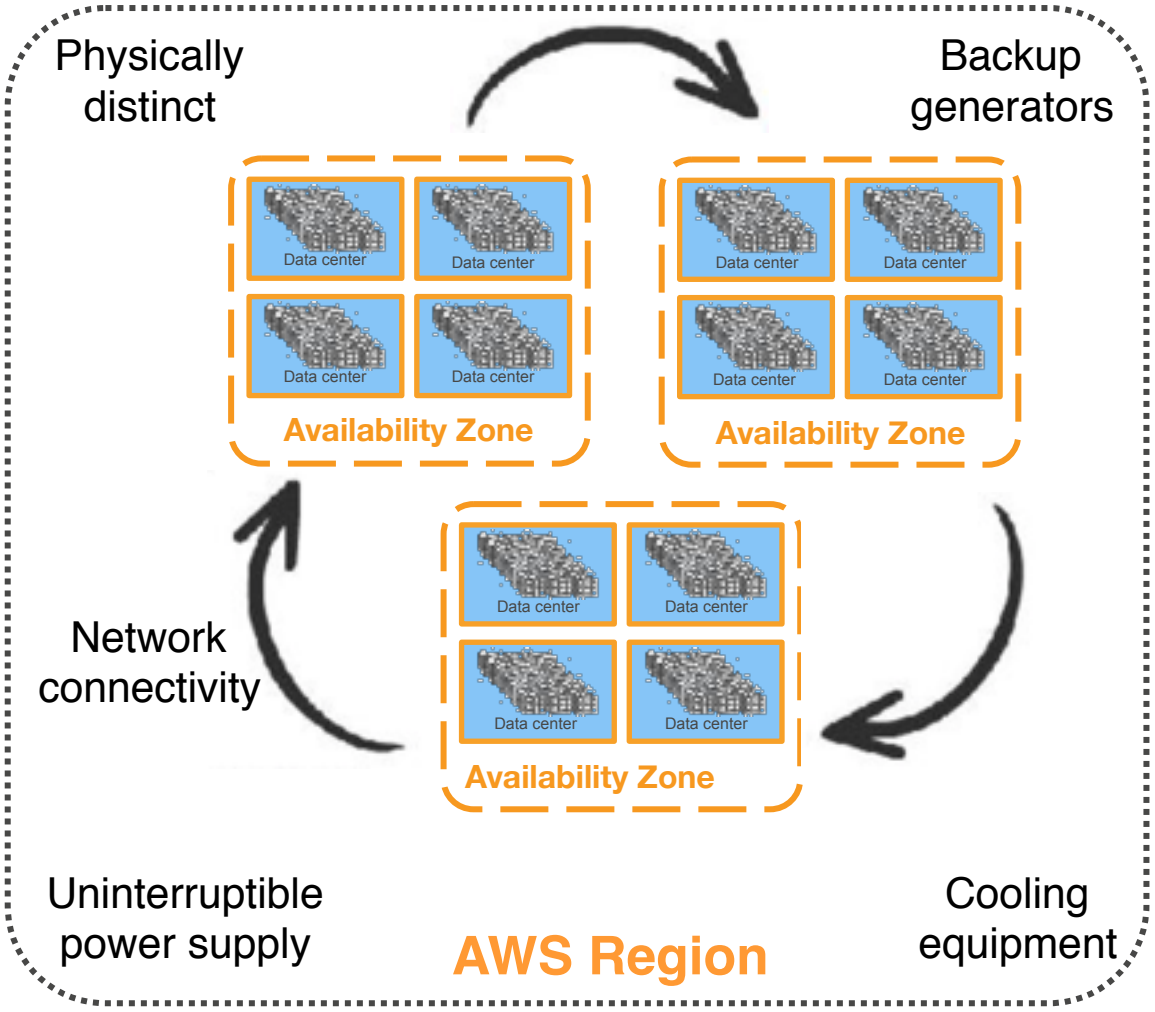


AWS Edge Locations

- An **Edge Location** is where users access AWS services.
- It is a global network of 114 points of presence (103 Edge Locations and 11 regional Edge Caches) in 56 cities across 24 countries.
- Specifically used with Amazon CloudFront, a **Global Content Delivery Network (CDN)**, to deliver content to end users with reduced latency.
- Regional edge caches used for content with infrequent access.

AWS Infrastructure Features

- Elastic and Scalable:
  - Elastic infrastructure; dynamic adaption of capacity
  - Scalable infrastructure; adapts to accommodate growth
- Fault-tolerant:
  - Continues operating properly in the presence of a failure
  - Built-in redundancy of components
- High availability:
  - High level of operational performance
  - Minimized downtime
  - No human intervention



AWS Foundational Services

Applications

- Virtual Desktops
- Collaboration and Sharing

Platform Services

Databases	Analytics	App Services	Deployment and Management	Mobile Services
Relational	Cluster Computing	Queuing	Containers	Identity
No SQL	Real-time	Orchestration	Dev/ops Tools	Sync
Caching	Data Warehouse	App Streaming	Resource Templates	Mobile Analytics
	Data Workflows	Transcoding	Usage Tracking	Notifications
		Email	Monitoring and Logs	
		Search		

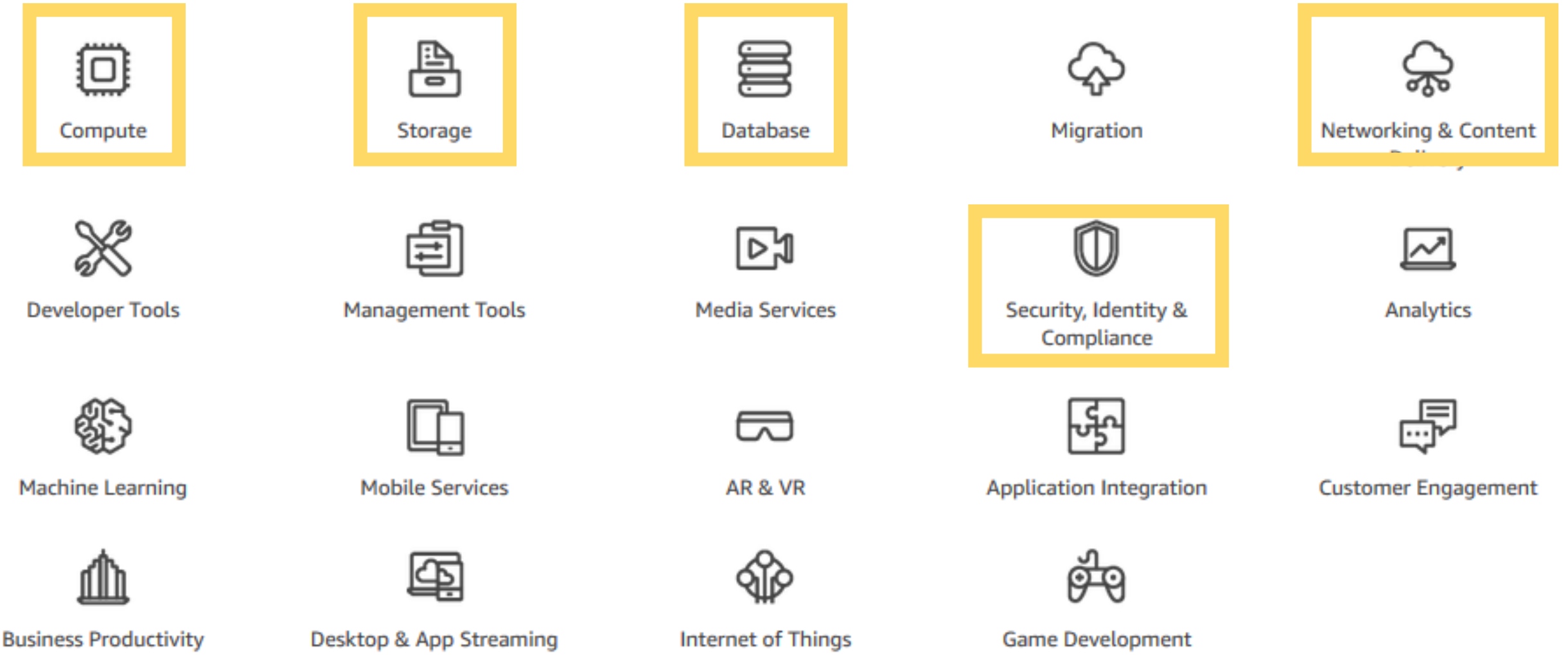
Foundation Services

- Compute (Virtual, Auto-scaling and Load Balancing)
- Networking
- Storage (Object, Block and Archive)

Infrastructure

- Regions
- Availability Zones
- Edge Locations

AWS Services and Categories



AWS Core Services - Compute Services

aws academy

Compute Services Overview

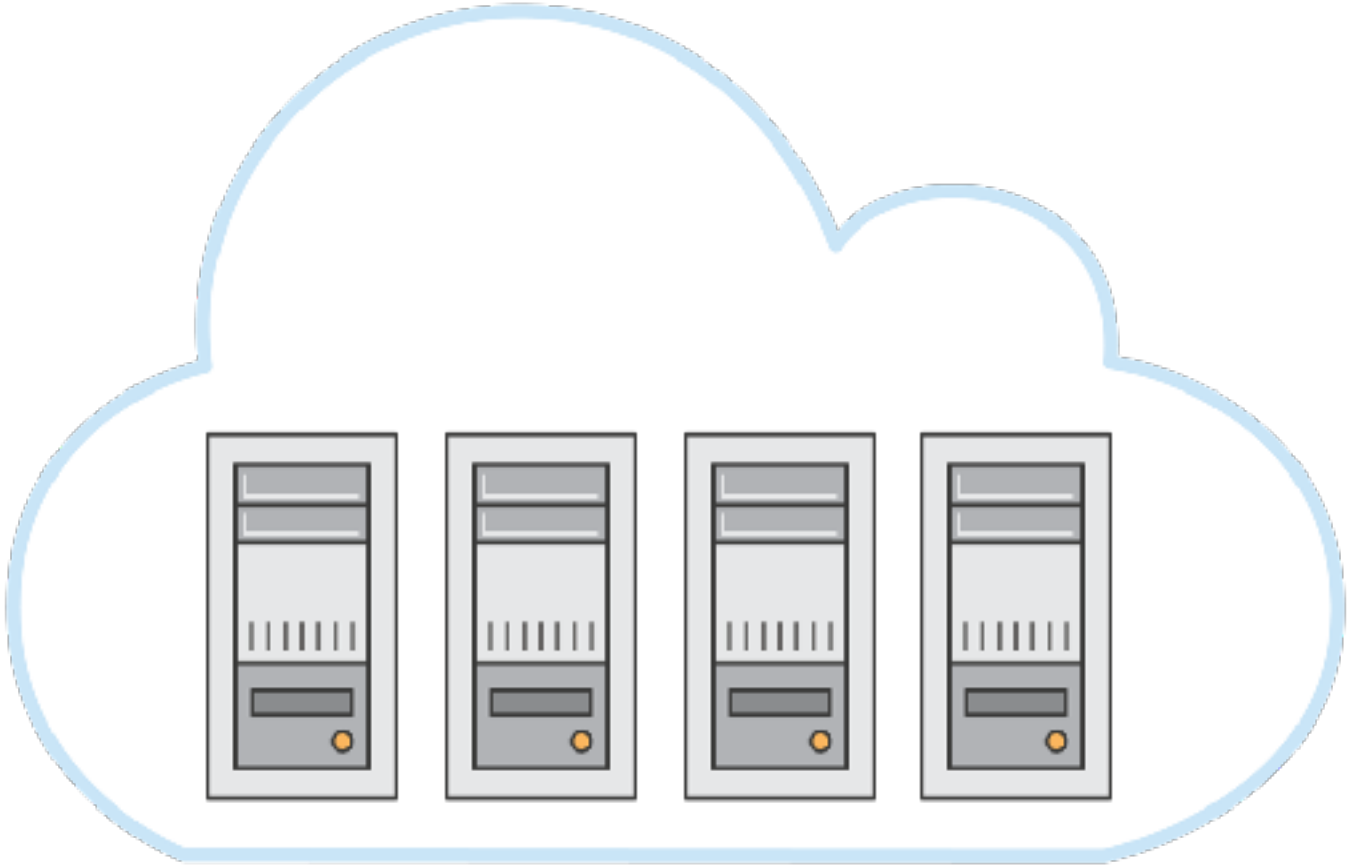
- Amazon Elastic Compute Cloud (Amazon EC2):
  - Virtual computing environment in the cloud
- AWS Lambda:
  - Fully managed serverless compute
- Automatic Scaling:
  - Scales EC2 capacity as needed
  - Improves availability
- Elastic Load Balancer:
  - Distributes incoming traffic
  - Helps achieve higher levels of fault tolerance
- AWS Elastic Beanstalk:
  - Quickly deploys, scales, and manages web apps
  - No charge for Elastic Beanstalk – pay only for the underlying AWS services used

Additional Compute Services

- Amazon Lightsail:
  - Everything needed to jump start a project
  - Manage simple web and application servers
- Amazon Elastic Container Services (ECS):
  - Highly scalable, high-performance container management service
  - Eliminates need to manage cluster management infrastructure
- AWS Fargate:
  - Containers without server or cluster management
- Amazon Elastic Container Service for Kubernetes (EKS):
  - Run Kubernetes without managing Kubernetes clusters

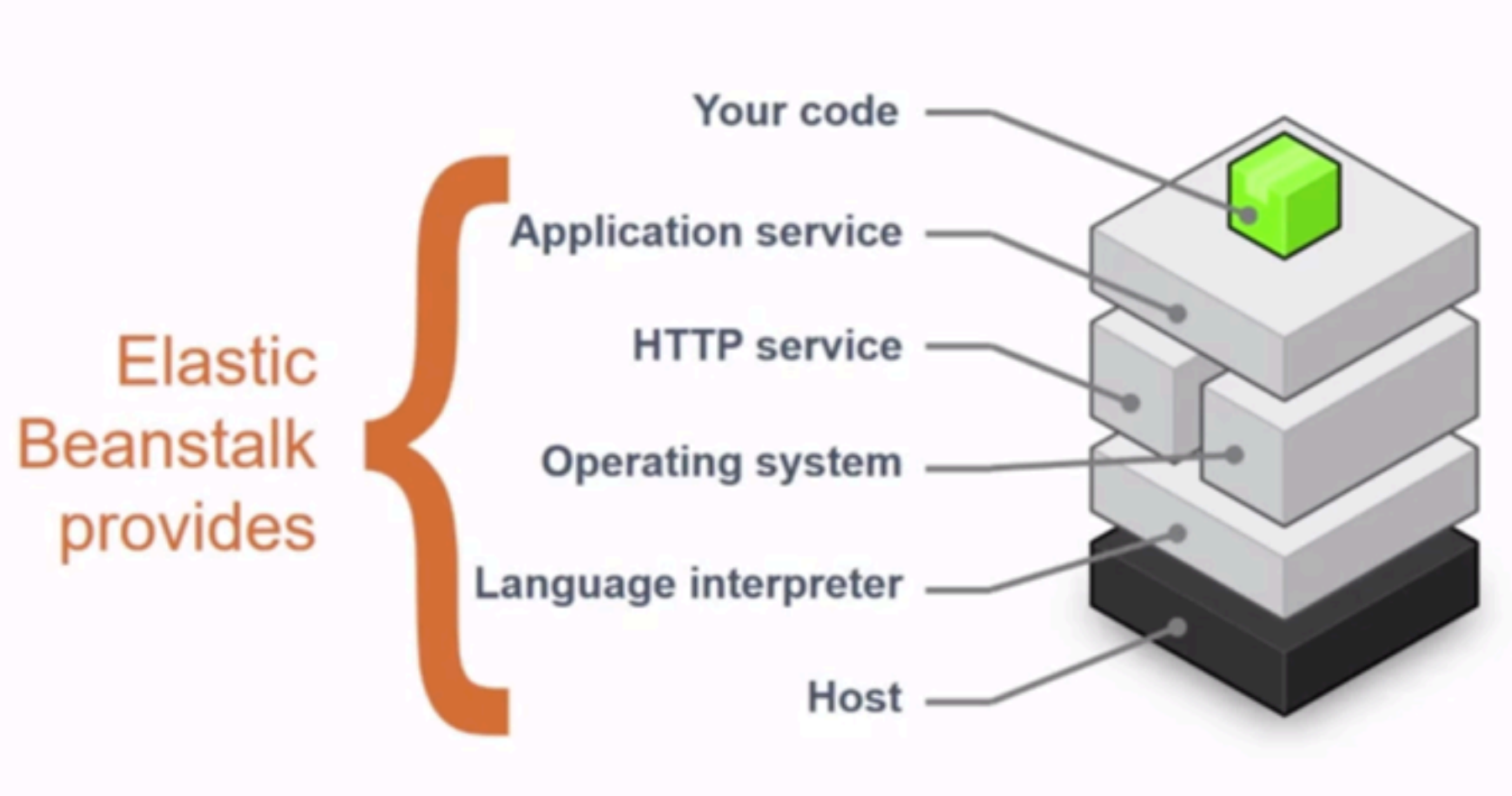
Amazon EC2?

Elastic Compute Cloud

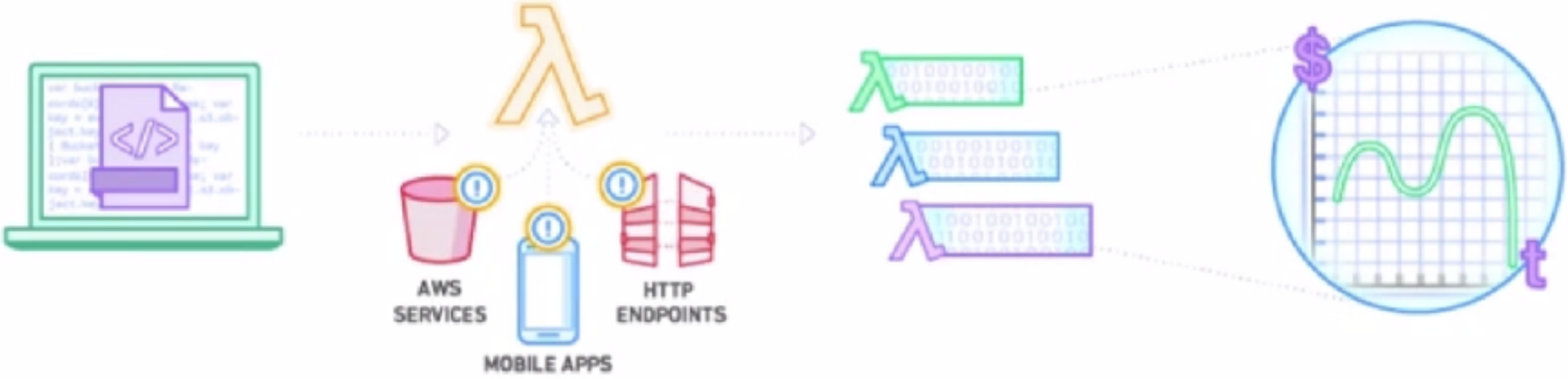


- ✓ Application server
- ✓ Web server
- ✓ Database server
- ✓ Game server
- ✓ Mail server
- ✓ Media server
- ✓ Catalog server
- ✓ File server
- ✓ Computing server
- ✓ Proxy server

Elastic Beanstalk Components



AWS Lambda



Upload your code to AWS Lambda

Set up your code to trigger from other AWS services, HTTP endpoints, or in-app activity

Lambda runs your code only when triggered using only the compute resources needed

Pay just for the compute time you use

Introduction to Amazon EC2

Amazon EC2

Amazon VPC

Amazon EC2

Amazon S3

Amazon EBS

Amazon EFS

Amazon Glacier

Storage

Amazon RDS

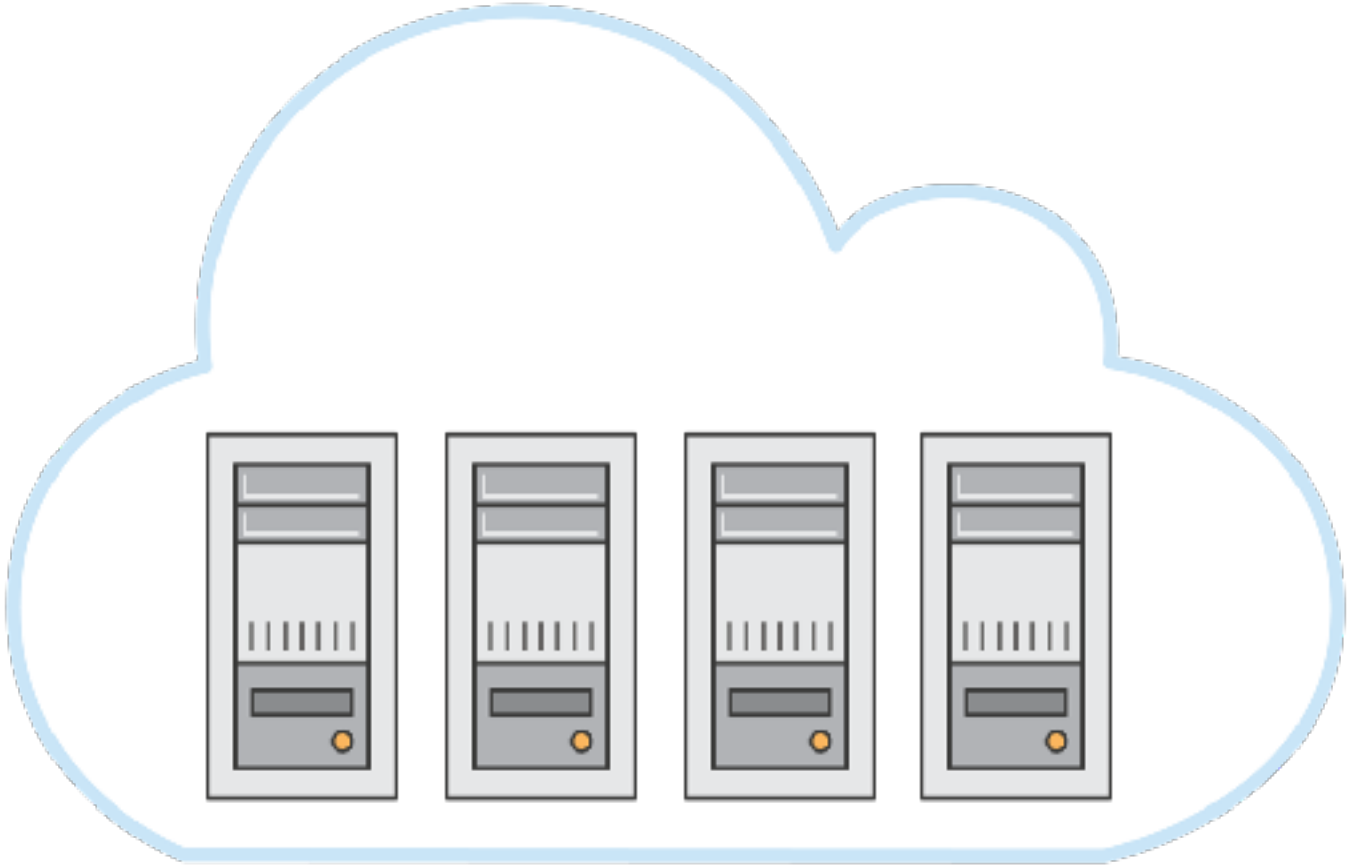
Amazon DynamoDB

Database

AWS IAM

What is Amazon EC2?

Elastic Compute Cloud



- ✓ Application server
- ✓ Web server
- ✓ Database server
- ✓ Game server
- ✓ Mail server
- ✓ Media server
- ✓ Catalog server
- ✓ File server
- ✓ Computing server
- ✓ Proxy server

Amazon EC2

Amazon Elastic Compute Cloud (EC2) offers **virtual computing environments** , known as **instances** you can launch and manage with a few clicks of a mouse or a few lines of code.

- Most server operating systems are supported.
- Create, save, and reuse your own server images as Amazon Machine Images (AMIs).
- Add more instances when you need them; terminate when you do not
- Launch one instance at a time or launch a whole fleet.
- CPU, memory, storage, networking, graphics, and general purpose instance types are available.
- Amazon EC2 instances in Amazon VPC now offer native support for the IPv6 protocol.
- Use security groups to control traffic to and from instances.

Choosing the Right Amazon EC2 Instance

- AWS uses Intel Xeon processors, providing customers with high performance and value.
- Amazon EC2 instance types are optimized for different use cases and workload requirements. They come in multiple sizes.

- Consider the following when choosing your instances:
  - **Core count**
  - **Memory size**
  - **Storage size & type**
  - **Network performance**
  - **CPU technologies**

Amazon Machine Image (AMI)

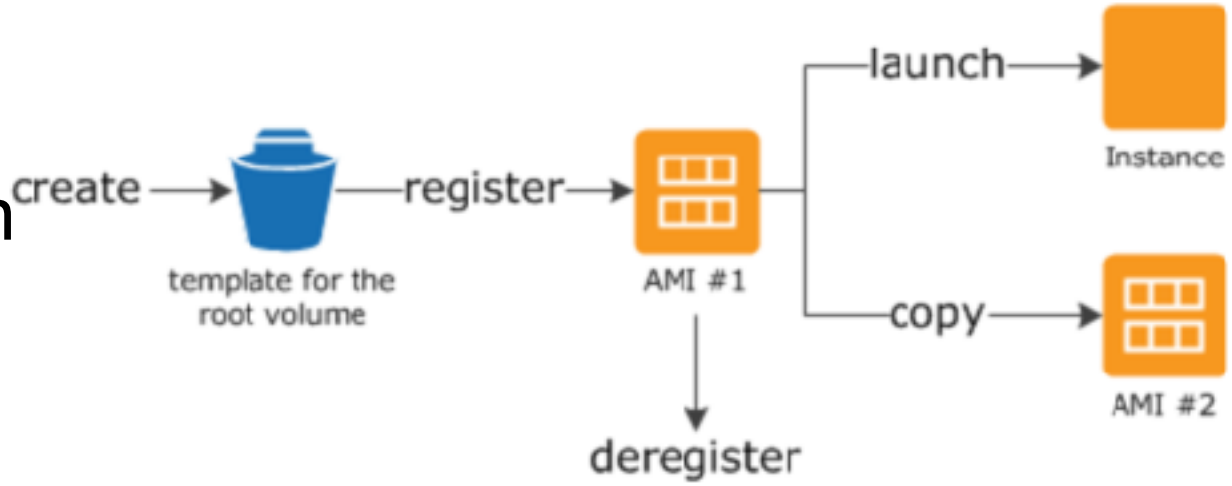
Amazon Machine Image (AMI) defines the **initial software that will be on an instance** when it is launched serving as the basic unit of deployment for services delivered using Amazon EC2 and defines every aspect of the software state at instance launch including:

- The Operating System (OS) and its configuration
- The initial state of any patches
- Application or system software

All AMIs are based on X86 Oss, either Linux or Windows.

AMI Lifecycle and Uses

- Create and register an AMI.
- Uses:
  - Launch new instance.
  - Copy within the same region or to different regions.
- De-register the AMI when no longer required.

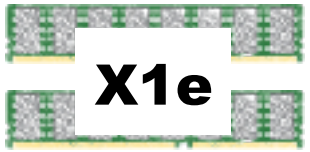


Broad Set of Compute Instance Types

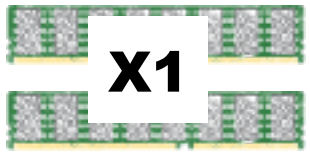
General Purpose

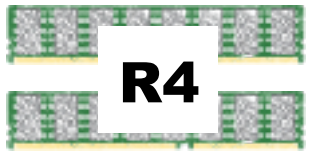
Compute Optimized

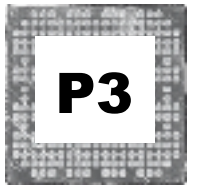
Storage and I/O Optimized



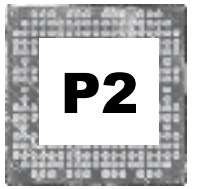
Memory Optimized

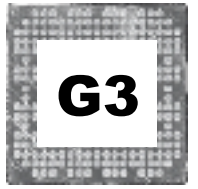


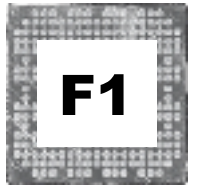




GPU- or FPGA-Enabled







Burstable Performance

Graphics Processing Unit (GPU)
Field Programmable Gate Array (FPGA)

Amazon EC2 Pricing

On-Demand

Spot Instances

Reserved Instances

- Per-second billing (Amazon Linux and Ubuntu only)
- Per-hour billing (All other OSs)

Dedicated Hosts

- Per-hour billing

Per Second Billing

- Pay for only what you use
- On-Demand, Reserved, and Spot forms
- Instances running for irregular periods of time
- Allow customers to focus on their application instead of maximizing usage to the hour
- All AWS Regions and Availability Zones
- Amazon Linux and Ubuntu

Amazon EC2 Pricing: Costs

On-Demand Instances	Spot Instances	Reserved Instances	Dedicated Hosts
• Pay for what you use • Per-second billing	• Spot price based on supply and demand • Per-second billing	• Pay low or no upfront fee; overall cost is lower • Per-second billing	Pay the On-Demand rate for every hour the host is active in the account

Amazon EC2 Pricing Options: Benefits

On-Demand Instances	Spot Instances	Reserved Instances	Dedicated Hosts
Low cost and flexibility	Large scale, dynamic workload	Predictability ensures compute capacity is available when needed	• Save money on licensing costs • Help meet compliance and regulatory requirements

Amazon EC2 Pricing Options: Use Cases

On-Demand Instances	Spot Instances	Reserved Instances	Dedicated Hosts
• Short-term, spiky, or unpredictable workloads • Application development or testing	• Applications with flexible start and end times • Applications only feasible at very low compute prices • Users with urgent computing needs for large amounts of additional capacity	• Steady state or predictable usage workloads • Applications that require reserved capacity, including disaster recovery • Users able to make upfront payments to reduce total computing costs even further	• Bring your own license (BYOL) • Compliance and regulatory restrictions • Usage and licensing tracking • Control instance placement

Amazon EC2: Billing and Instance Configuration

1. Clock Hours of Server Time for Second/Hourly Billing:

- Resources incur charges only when running

2. Instance Configuration:

- Physical capacity of the instance
- Pricing varies with:
  - AWS region
  - OS
  - Number of cores
  - Memory

Amazon EC2: Purchase Types

3. Ways to purchase Amazon EC2 Instances:

- **On-demand Instances:**
  - Compute capacity by the hour & second
  - Minimum of 60 seconds
- **Reserved Instances:**
  - Full, partial, or no up-front payment for instances reserved
  - Discount on hourly charge for that instance
  - 1 or 3 year term
- **Spot Instances:**
  - Bid for unused Amazon EC2 capacity
  - Price based on supply and demand
  - Instances can be lost if you are outbid
  - instances can be interrupted if Spot price exceeds maximum
- **Dedicated Hosts:**
  - Can be purchased On-Demand (hourly)
  - Can be purchased as a Reservation for

Amazon EC2: Number of Instances and Load Balancing

4. **Number of Instances:**
   - Provision multiple instances to handle peak loads and shut them down when they are no longer needed. Pay for only the capacity that you actually use.

5. **Load Balancing -** Uses Elastic Load Balancing to distribute traffic among Amazon EC2 instances
   - Calculates monthly cost based on:
     - Hours load balancer runs
     - Data load balancer processes

Amazon EC2: Detailed Monitoring

6. **Use Amazon CloudWatch to Monitor Instances:**
   - Basic monitoring (default, no additional cost)
   - Detailed monitoring
     - Fixed monthly rate for seven preselected metrics recorded once a minute
     - Prorated partial months

Amazon EC2

7. Auto Scaling:

- Automatically adjusts number of Amazon EC2 instances in your deployment.
- Incurs no additional charge beyond CloudWatch fees.

8. Elastic IP Addresses:

- No charge for one Elastic IP address associated with a running instance.

Amazon EC2: OS and Software

9. Pricing for operating systems and software packages:

- Includes OS prices in instance prices
- Partner with other vendors for certain software
- Requires licenses from vendors for other software
- Bring your existing license through specific vendor programs

Spot Instance Hibernation

- Hibernate Amazon EBS-backed instances in the event of an interruption.
- Resume instances when capacity is available.
- Use an encrypted Amazon EBS volume as the root volume.
- Hibernation agent required.
- Check the documentation for requirements.

Amazon EC2 Cost Optimization

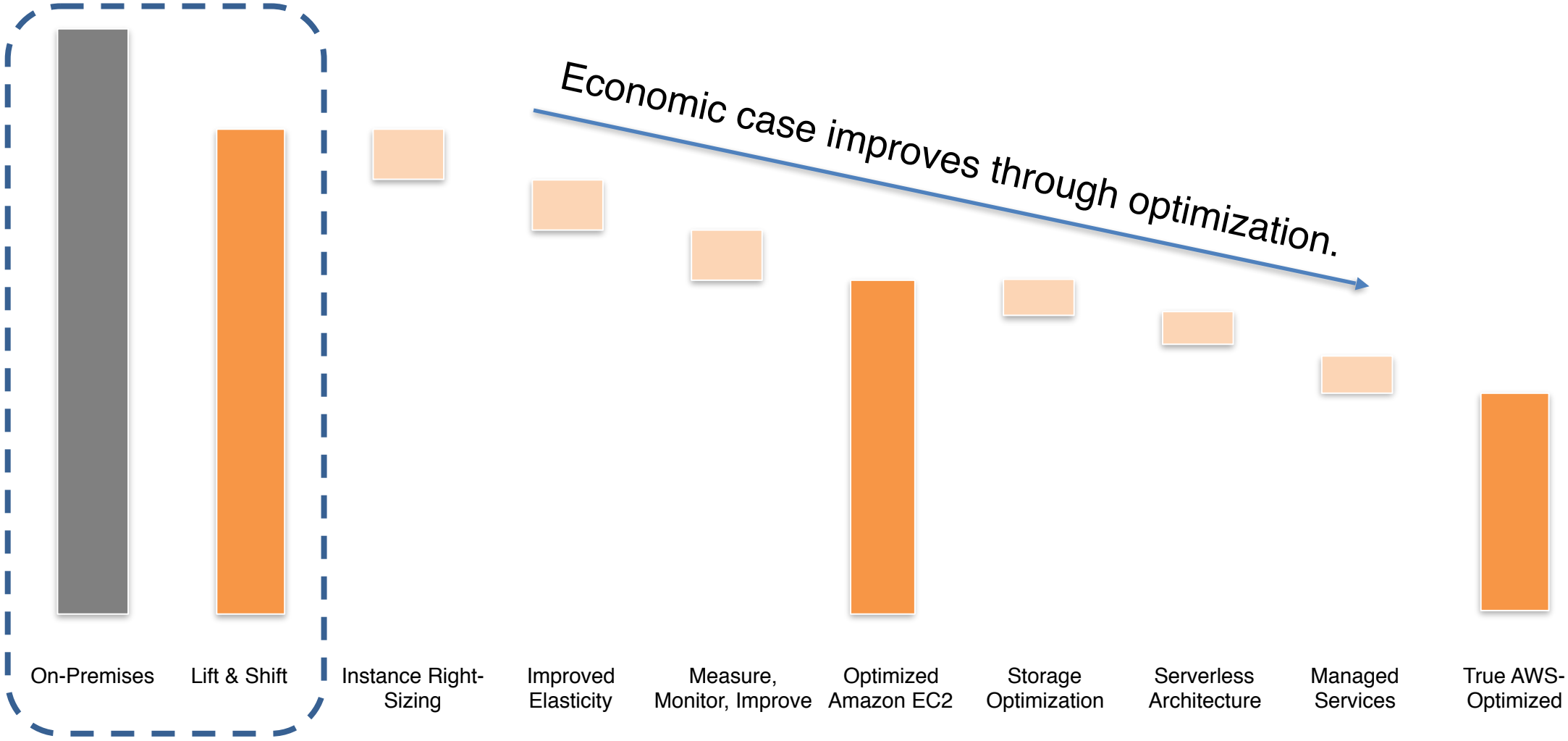
What is Cost Optimization?

Reduce Costs…

Pay only for ***what*** you need

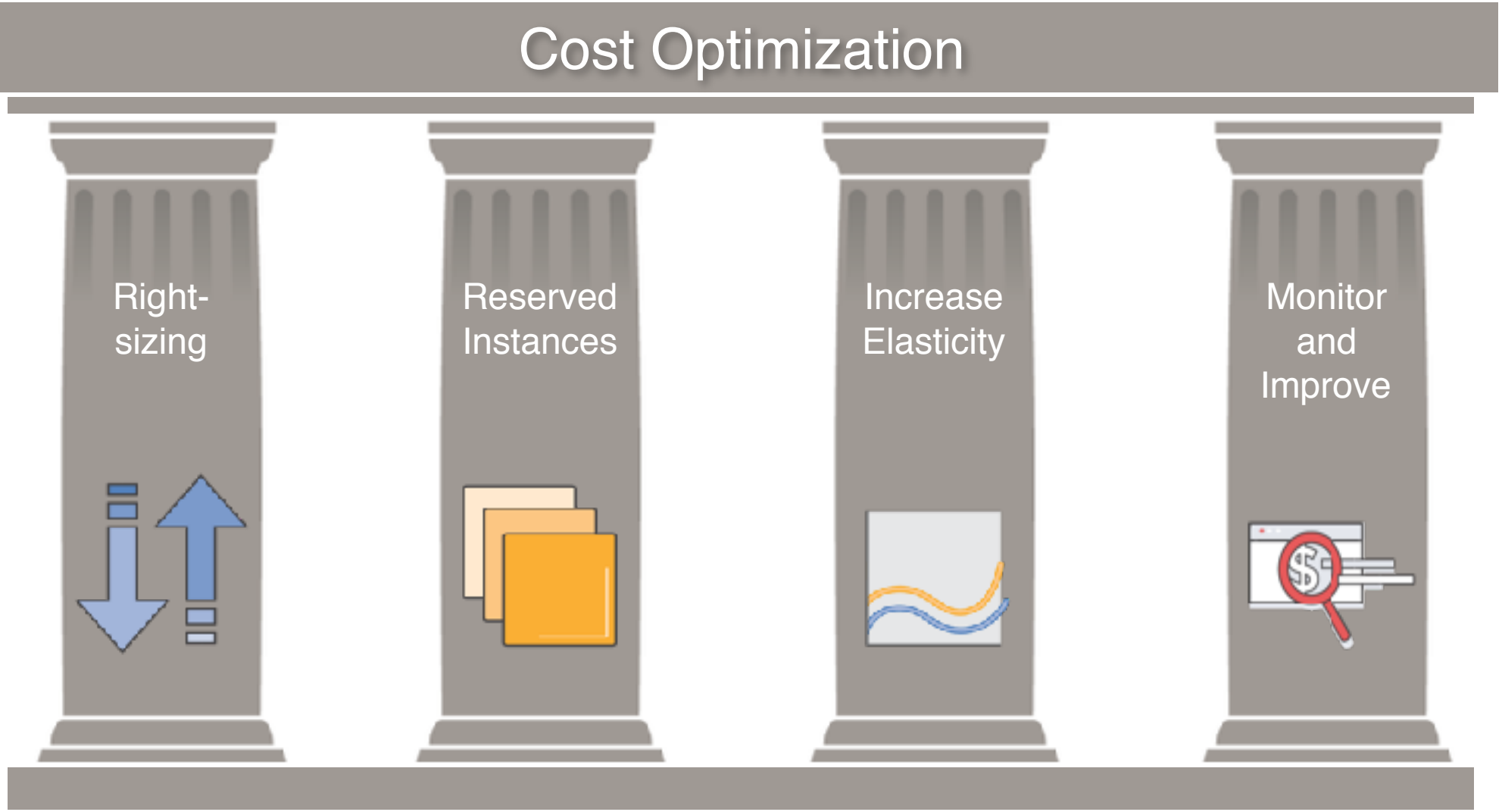
when you need it.

Lowering TCO Through Cost Optimization



Traditional TCO Comparisons

The Four Pillars of Cost Optimization



Driver 1: Right-Sizing

Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

- Select the appropriate instance types
- Downsize instances
- Leverage Amazon CloudWatch metrics

Best practice:

- Right size, then reserve

Optimize and Combine Amazon EC2 Purchase Types

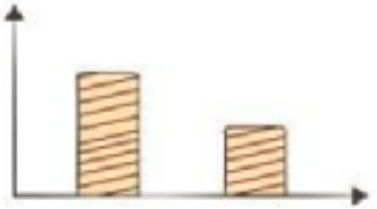
Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

On-Demand

Reserved

Spot



Dedicated

Optimize and Combine Amazon EC2 Purchase Types

Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

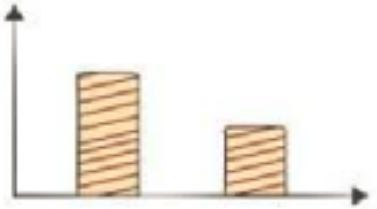
On-Demand

Spiky Workloads

- Pay by the hour.
- No long-term commitments

Reserved

Spot



Dedicated

Optimize and Combine Amazon EC2 Purchase Types

Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

On-Demand

Spiky Workloads

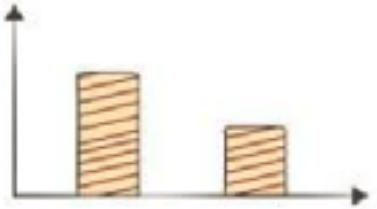
- Pay by the hour.
- No long-term commitments

Reserved

Steady-state Workloads

- Pay upfront
- 50-75% lower hourly rate

Spot



Dedicated

Optimize and Combine Amazon EC2 Purchase Types

Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

On-Demand

Spiky workloads

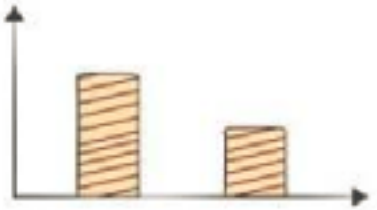
- Pay by the hour.
- No long-term commitments

Reserved

Steady-state workloads

- Pay upfront
- 50-75% lower hourly rate

Spot



Time-insensitive workloads

- Bid for unused Amazon EC2 capacity

Dedicated

Optimize and Combine Amazon EC2 Purchase Types

Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

On-Demand

Spiky Workloads

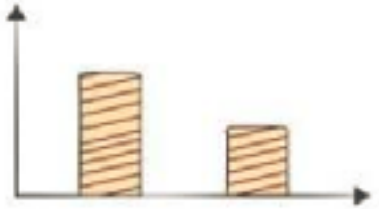
- Pay by the hour.
- No long-term commitments

Reserved

Steady-state Workloads

- Pay upfront
- 50-75% lower hourly rate

Spot



Time-insensitive Workloads

- Bid for unused Amazon EC2 capacity

Dedicated

Highly sensitive Workloads

- In your VPC
- Isolated, steady-state workloads

Optimize and Combine Amazon EC2 Purchase Types

Driver 1:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

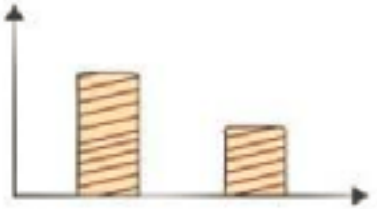
On-Demand

Spiky Workloads

Reserved

Steady-state Workloads

Spot



Time-insensitive Workloads

Dedicated

Highly sensitive Workloads

- ✓ Pay only for what you use
- ✓ On-demand, elastic provisioning
- ✓ Control and security

Driver 2: Reserved Instance Capacity

Driver 2:

Right-Sizing
Reserved
Instances (Ris)
Increase Elasticity
Monitor & Improve

Reserved Instances (RIs)/

Capacity

- Amazon Elastic Compute Cloud (EC2)
- Amazon Relational Database Service (RDS)
- Amazon DynamoDB
- Amazon Redshift
- Amazon ElastiCache

Commitment level

- 1 year
- 3 years

Up to 75%+ savings

* Dependent on specific AWS service, size/type, and region

Reserved Instances

Driver 2:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

Step 1: RI Coverage

- Cover always-on resources
- Target 70–80% always-on coverage

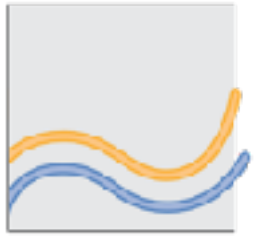
Step 2: RI Utilization

- Leverage RI flexibility to increase utilization
- Merge and split RIs as needed
- Target 95% RI utilization rate

Driver 3: Increase Elasticity

Driver 3:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

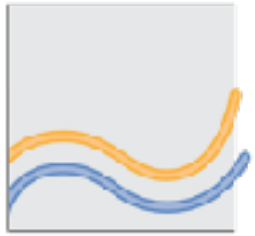


- **Elasticity**
  Using an instance when you need, turning it off when you don't
- **Turn off non-production instances**
  Example: Dev/test
- **Auto scale production**
  Use Auto Scaling to scale up and down based on demand and usage (e.g., spikes)
- **Target: 20-30% of Amazon EC2 instances**
  Run in On-demand or as Spot

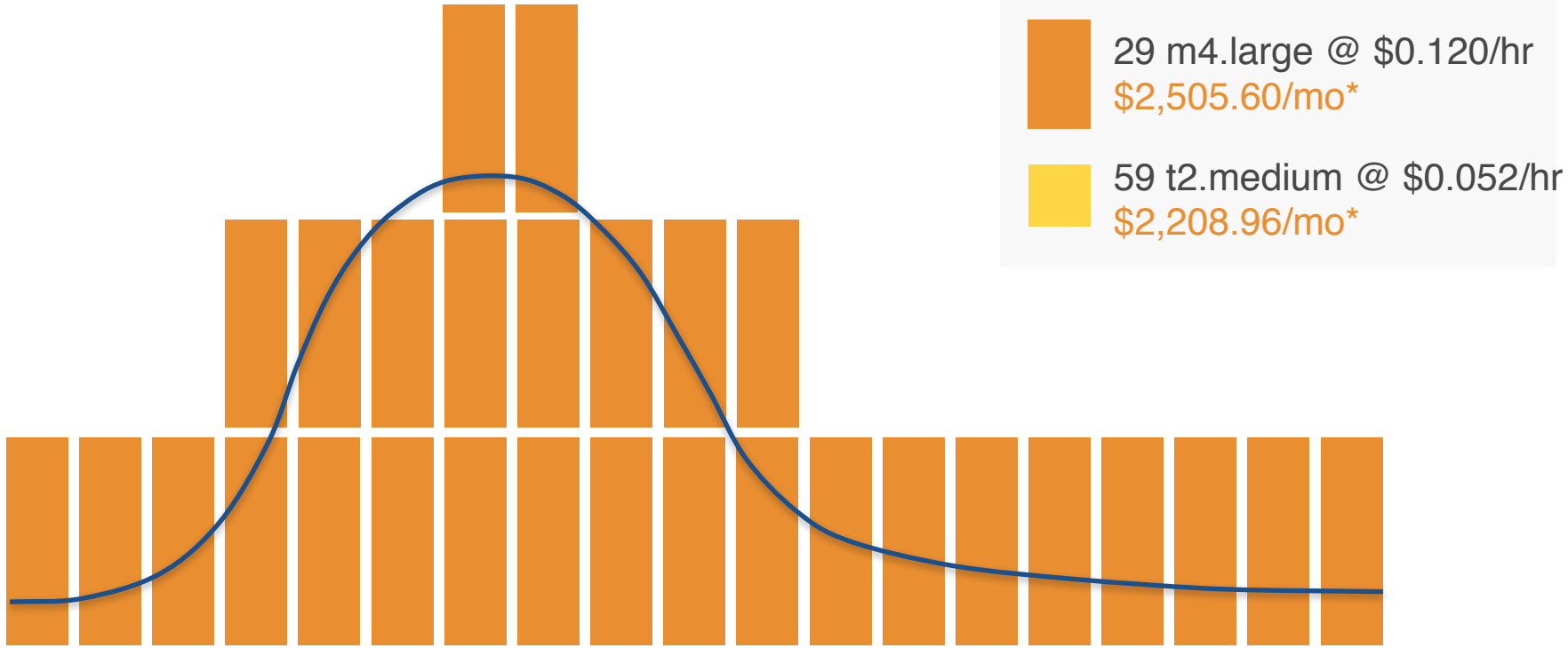
Using Right-sizing and Elasticity to Lower Cost

Driver 3:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve



More, smaller instances vs. fewer, larger instances



*Assumes Linux instances in the US-East (N. Virginia) Region at 720 hours per month

Driver 4: Measure, Monitor, and Improve

Driver 4:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

Cost Optimization Opportunities:

1. Auto-tag resources
2. Identify always-on non production systems
3. Identify instances to downsize
4. Recommend Reserved Instance (RIs) to purchase
5. Dashboard your status
6. Consolidate your billing
7. Report on savings

Measure, Monitor, and Improve

Driver 4:

Right-Sizing
Reserved Instances
Increase Elasticity
Monitor & Improve

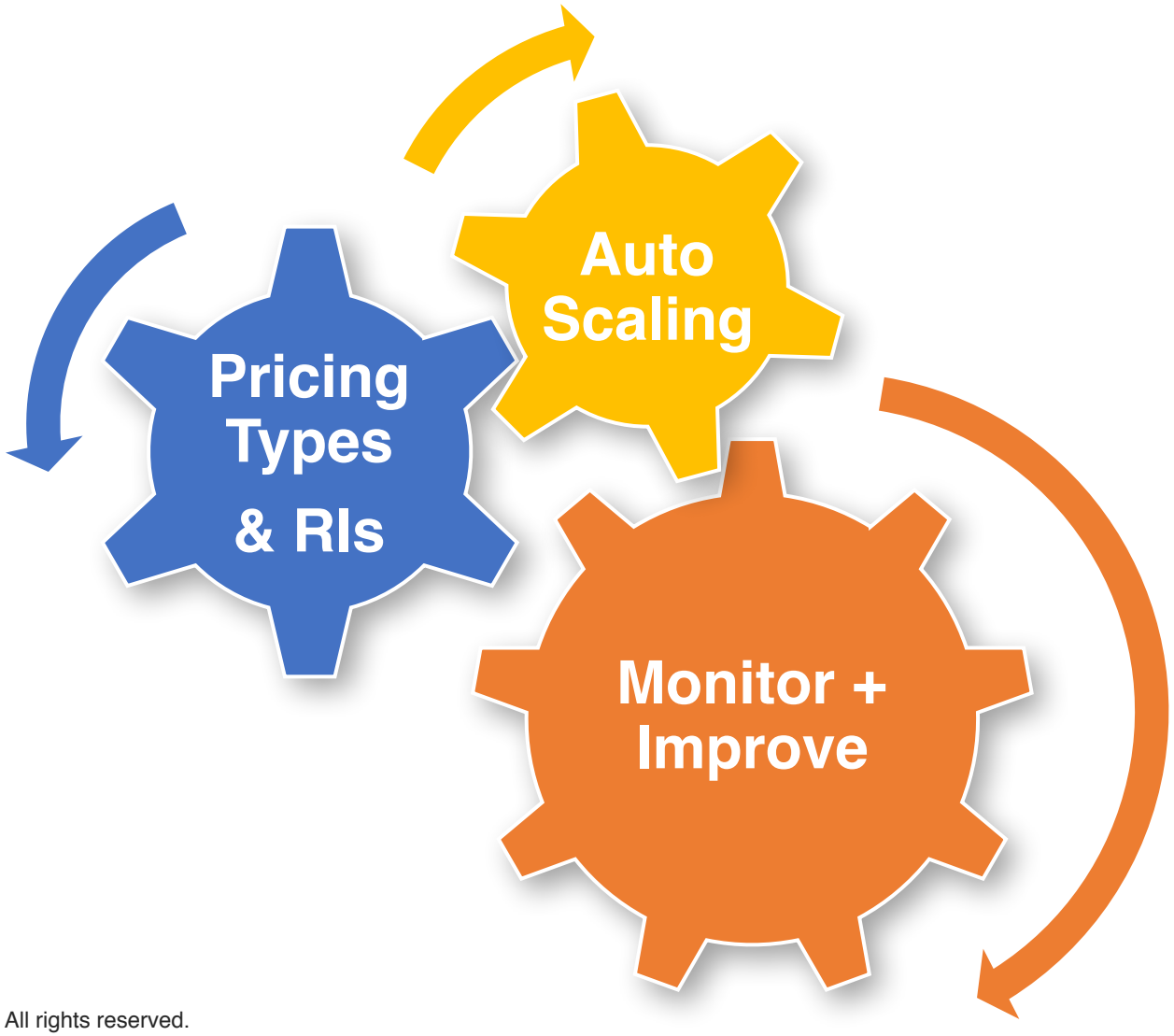
AWS Trusted Advisor

- Optimize your AWS environment
- Reduce cost, increase performance, and improve security

Cost Explorer

- View graphs of your costs: the last 13 months
- Forecast your likely costs: the next 3 months
- View time data by day or month

Continual Process of Cost Optimization



AWS Core Services - Storage

aws academy

Core AWS Services

Amazon VPC

Amazon EC2

Amazon S3

Amazon EBS

Amazon EFS

Amazon Glacier

Storage

Amazon RDS

Amazon DynamoDB

Database

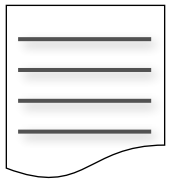
AWS IAM

Amazon Elastic Block Store (EBS)

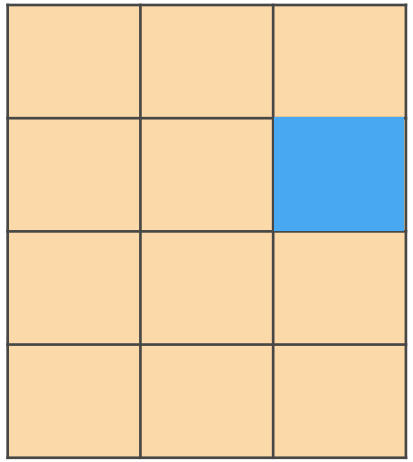
Storage

Amazon Elastic Block Store (Amazon EBS)

AWS Storage Options: Block vs. Object Storage

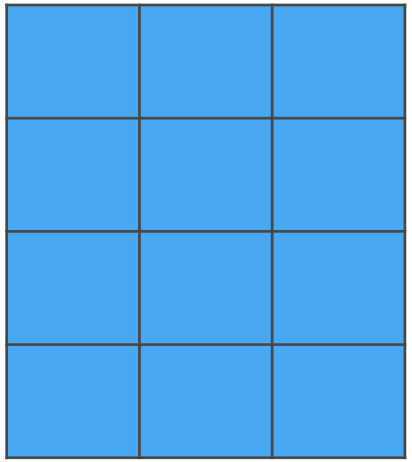


What if you want to **change** one character in a 1-GB file?



Block Storage

Change one block (piece of the file) that contains the character



Object Storage

Entire file must be updated

Amazon EBS

Amazon EBS allows you to **create individual storage volumes** and **attach them** to an Amazon EC2 instance.

- Amazon EBS offers block-level storage.
- Volumes are automatically replicated within its Availability Zone.
- Can be backed up automatically to Amazon S3.
- Uses:
  - Boot volumes and storage for Amazon EC2 instances
  - Data storage with a file system
  - Database hosts
  - Enterprise applications

Amazon EBS Volume Types

	Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
	General Purpose	Provisioned IOPS	Throughput-Optimized	Cold
Max volume size	16 TiB	16 TiB	16 TiB	16 TiB
Max IOPS/volume	10,000	32,000	500	250
Max throughput/volume	160 MiB/s	500 MiB/s	500 MiB/s	250 MiB/s

Amazon EBS Volume Types

	Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
	General Purpose	**Provisioned IOPS**	**Throughput-Optimized**	**Cold**
Use Cases	• Recommended for most workloads • System boot volumes • Virtual desktops • Low-latency interactive apps • Development and test environments	• I/O-intensive workloads • Relational DBs • NoSQL DBs • • •	• Streaming workloads requiring consistent, fast throughput at a low price • Big data • Data warehouses • Log processing • Cannot be a boot volume	• Throughput-oriented storage for large volumes of data that is infrequently accessed • Scenarios where the lowest storage cost is important • Cannot be a boot volume

Amazon EBS

Snapshots:

- Point-in-time snapshots
- Recreate a new volume at any time

Encryption:

- Encrypted Amazon EBS volumes
- No additional cost

Elasticity:

- Increase capacity
- Change to different types

Amazon EBS: Volumes and IOPS

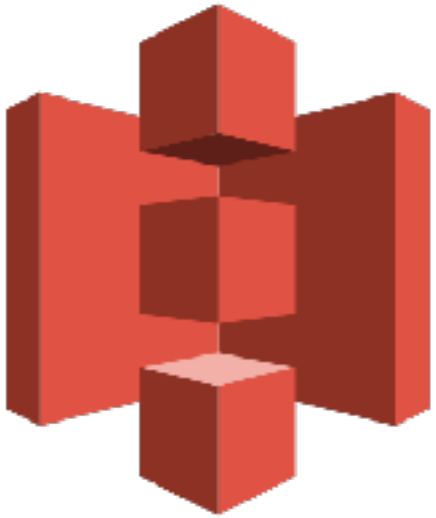
1. **Volumes:**
   - Amazon EBS volumes persist independently from the instance.
   - All volume types are charged by the amount provisioned per month.
2. **Input Output Operations per Second (IOPS):**
   - General Purpose (SSD)
     - Charged by the amount your provision in GB per month until storage is released
   - Magnetic
     - Charged by the number of requests to volume
   - Provisioned IOPS (SSD)
     - Charged by the amount you provision in IOPS (by % of day / month used)

Amazon EBS: Snapshots and Data Transfer

3. **Snapshots:**
   - Added cost of Amazon EBS snapshots to Amazon S3 is per GB-month of data stored.

4. **Data Transfer:**
   - Inbound data transfer is free.
   - Outbound data transfer charges are tiered.

Amazon Simple Storage Service (S3)



Amazon Simple Storage Service (Amazon S3)

Amazon S3

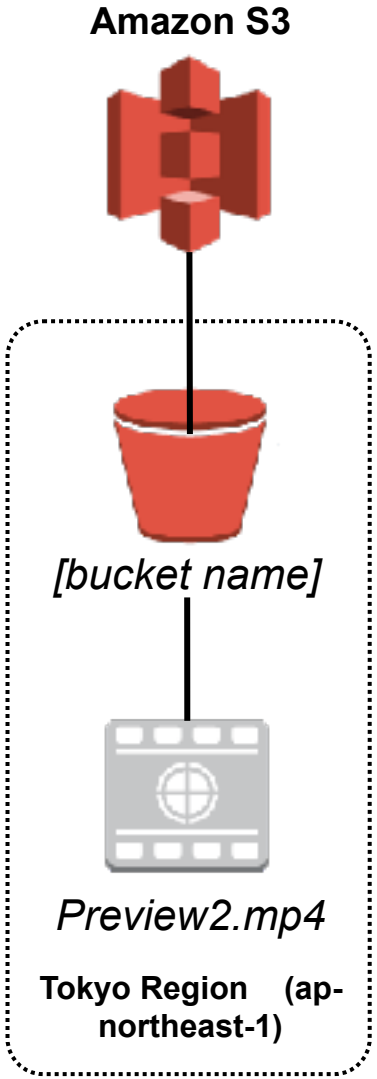
Managed cloud storage solution designed to **scale seamlessly and** provide **99.999999999% durability.**

- Store as many objects as you want.
- Bucket names must be unique across all existing bucket names in Amazon S3.
- Amazon S3 cannot be used as a bootable drive.
- Data is stored redundantly.
- Access Amazon S3 with the AWS Management Console, one of the Software Development Kits (SDKs), or a third-party solution.
- Object uploads or deletes can trigger notifications, workflows, or even scripts.
- Data in transit and at rest can be encrypted automatically.
- Storage class analysis (Amazon S3 Analytics) to analyze storage access patterns and transition the right data to the right storage class.

Amazon S3 Storage Classes

- Amazon S3 provides four classes of object-level storage:
  - Amazon S3 Standard
  - Amazon S3 Standard-IA
  - Amazon S3 One Zone-IA
  - Amazon Glacier

Amazon S3



To upload your data (photos, videos, documents, etc.):

1. Create a **bucket** in one of the AWS Regions.
2. Upload any number of **objects** to the bucket.

Bucket

https://s3-ap-northeast-1.amazonaws.com/[bucket name]/

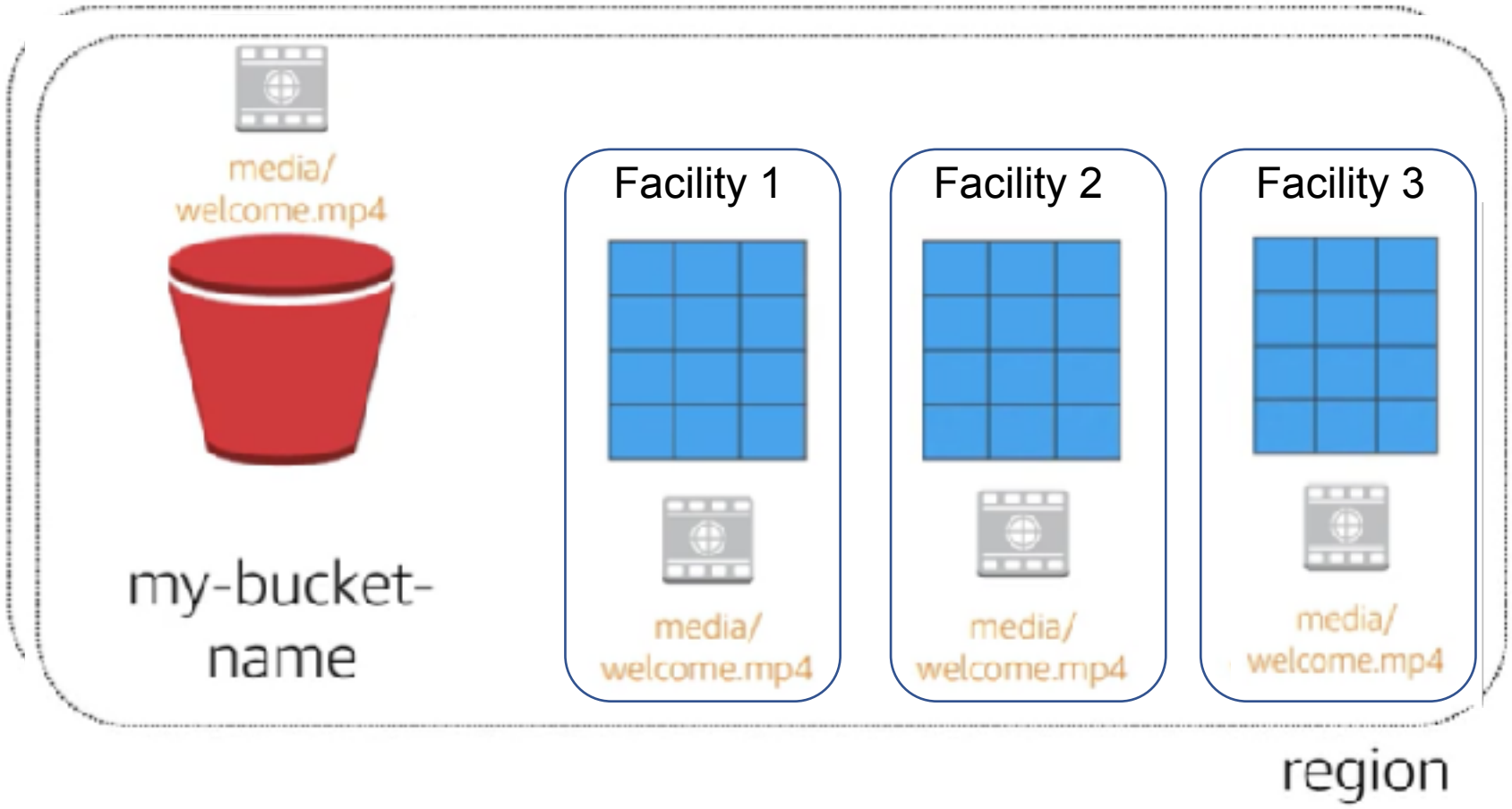
Region code — **Bucket name**

Object

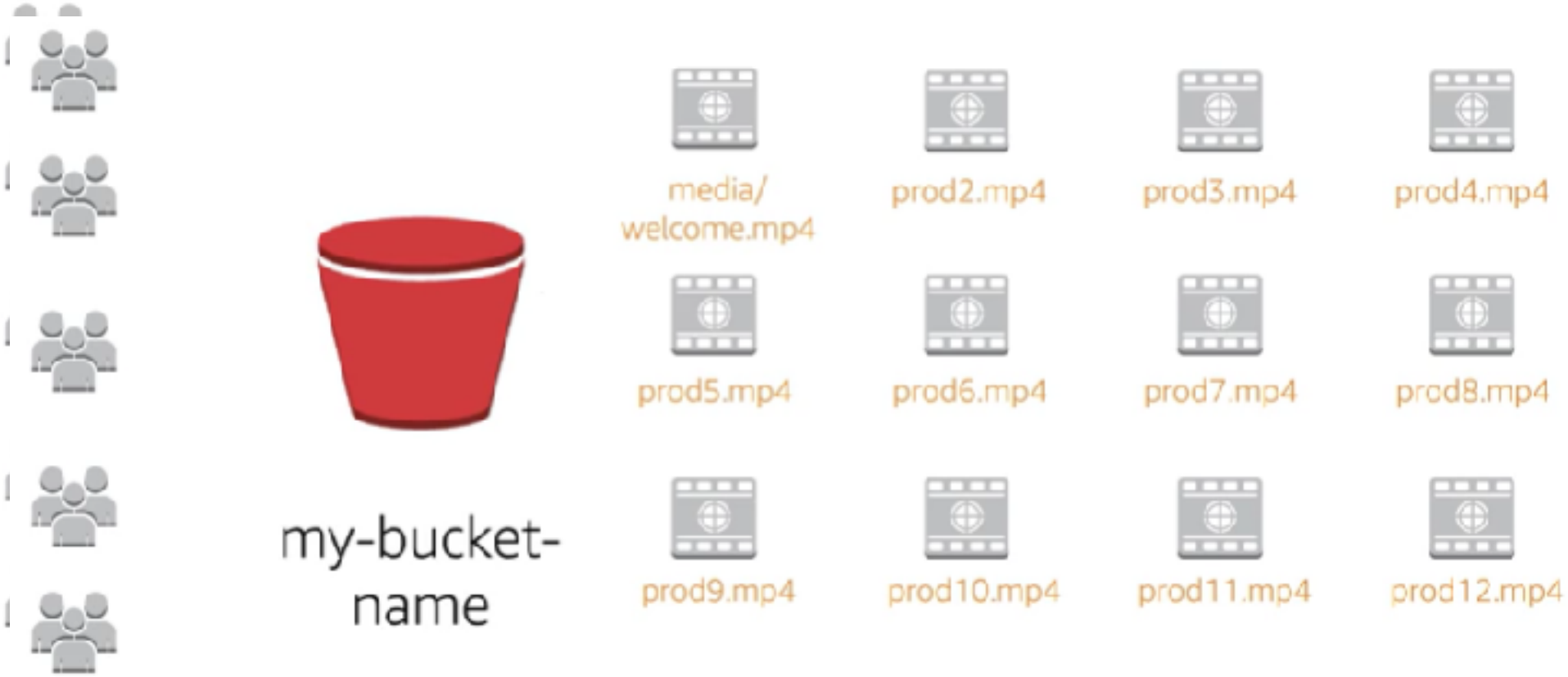
https://s3-ap-northeast-1.amazonaws.com/[bucket name]/Preview2.mp4

Key

Data Redundantly Stored in Region



Designed for Seamless Scaling



Access the Data Anywhere

AWS Management Console

AWS CLI

AWS SDKs

Common Use Cases

- Storing application assets
- Static web hosting
- Backup and Disaster Recovery (DR)
- Staging area for big data
- *Many more….*

Amazon S3 Pricing

- Pay only for what you use, including:
  - GBs per month
  - Transfer OUT to other regions
  - PUT, COPY, POST, LIST, and GET requests

- You do NOT have to pay for:
  - Transfers IN to Amazon S3.
  - Transfers OUT from Amazon S3 to Amazon CloudFront or Amazon EC2 in the same region.

Amazon S3: Storage Pricing

To estimate Amazon S3 costs, consider the following:

1. **Types of storage classes:**
   - Standard Storage
     - 99.999999999% durability
     - 99.99% availability
   - Standard-Infrequent Access (SIA)
     - 99.999999999% durability
     - 99.9% availability
2. **Amount of storage:**
   - The number and size of objects
   - Type of storage

Amazon S3: Storage Pricing

3. **Requests:**
   - The number of requests **(GET, PUT, COPY):**
   - Type of requests
     - Different rates for GET requests than other requests.
4. **Data Transfer:**
   - Pricing based on the amount of data transferred out of the Amazon S3 region
     - Data transfer in is free, but charges for data transfer out.

Amazon Glacier

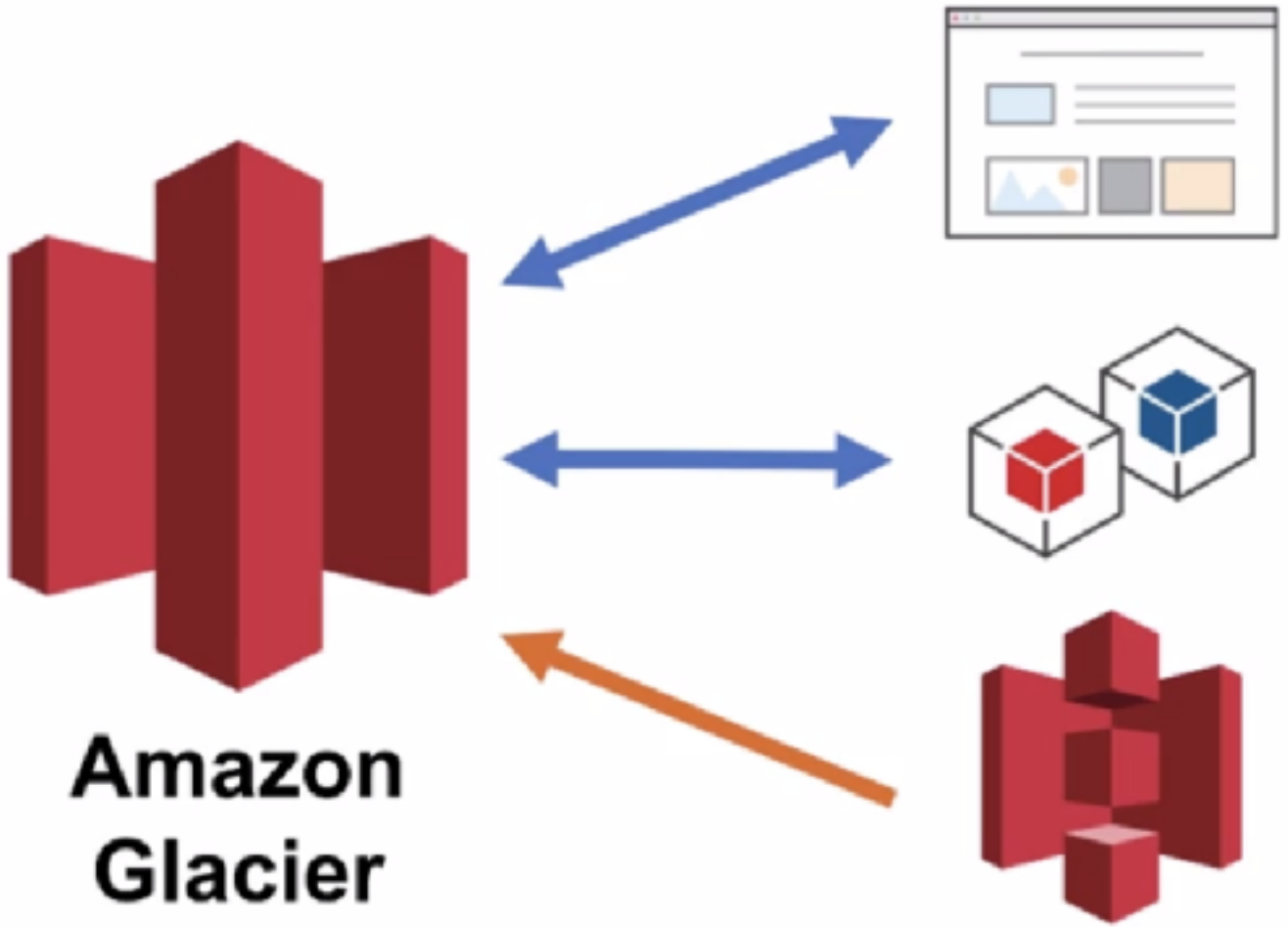
Amazon Glacier

Amazon Glacier Review

Amazon Glacier is a **data archiving service** designed for **security**, **durability**, and an **extremely low cost**.

- Designed for durability of 99.999999999% of objects.
- Supports SSL/TLS encryption of data in transit and at rest.
- The Vault Lock feature enforces compliance via a lockable policy.
- Extremely low-cost design is ideal for long-term archiving.
  - Provides three options for access to archives (Expedited, Standard, and Bulk) from a few minutes to several hours.

Using Amazon Glacier



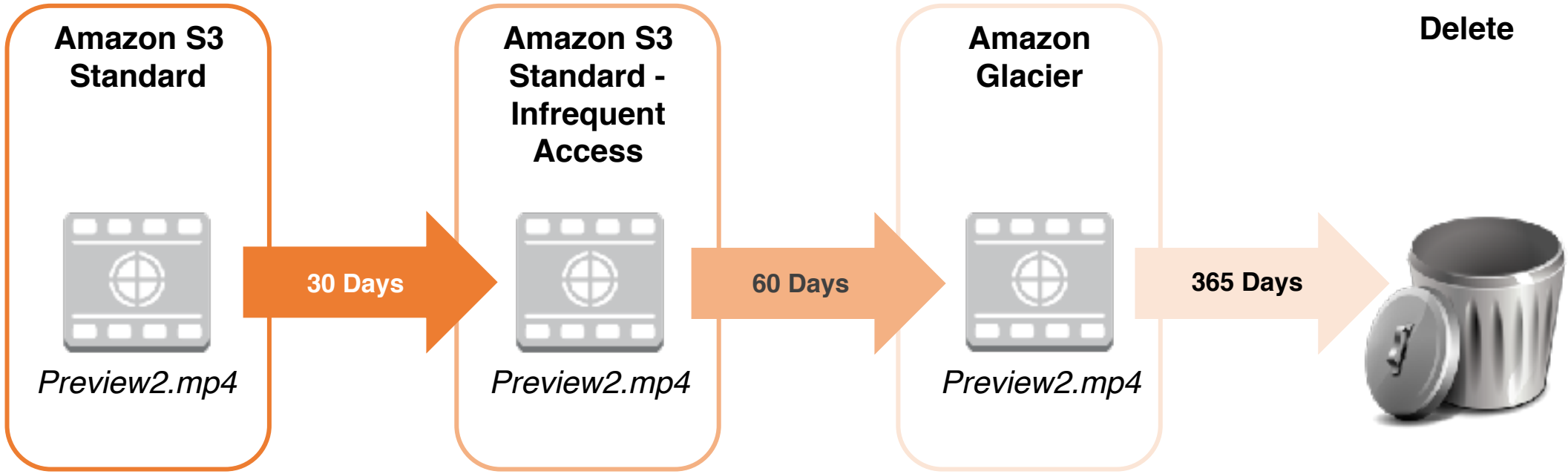
RESTful Web services

Java or .NET SDKs

Amazon S3 with lifecycle policies

Lifecycle Policies

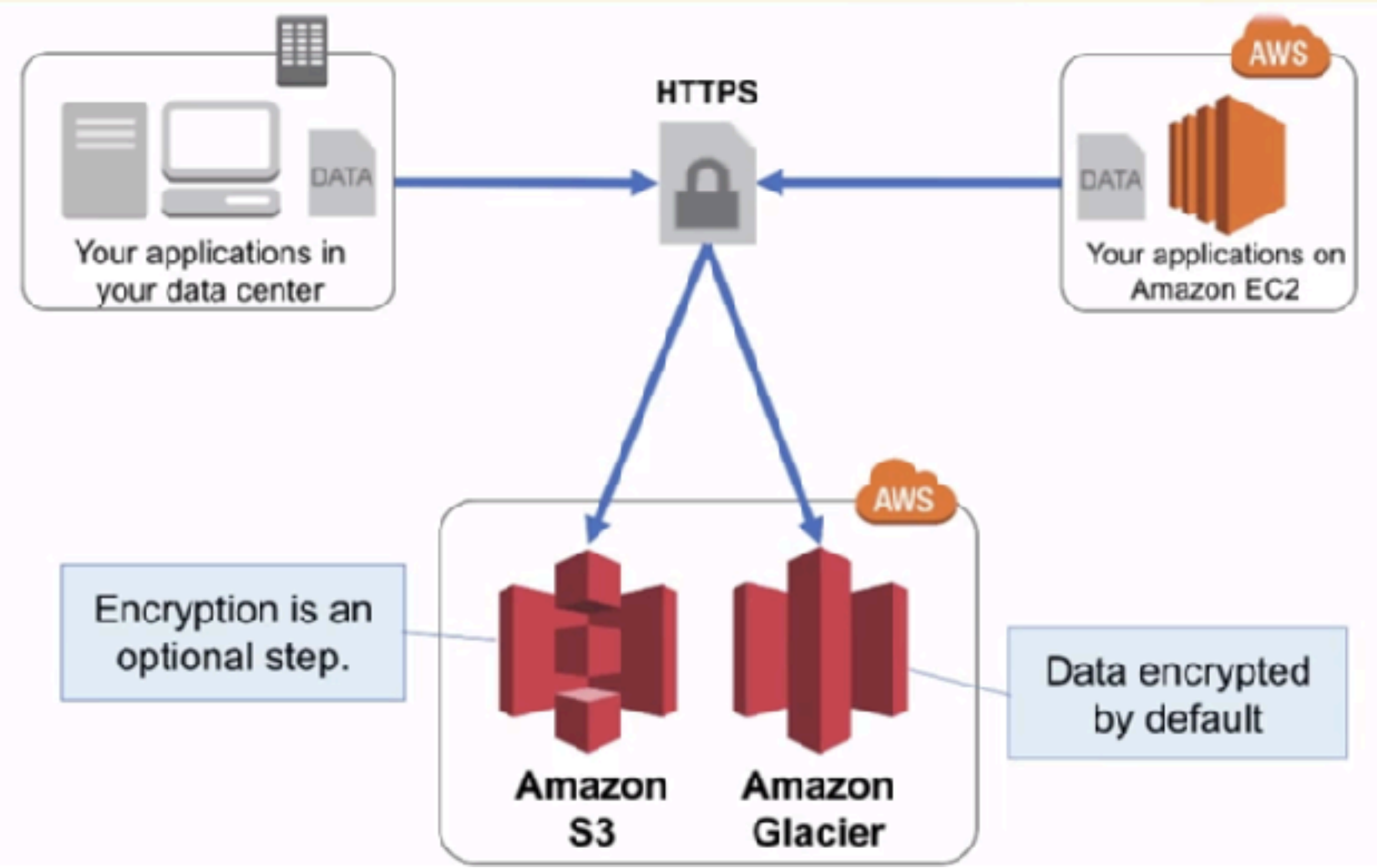
Amazon S3 lifecycle policies allow you to delete or move objects based on age.



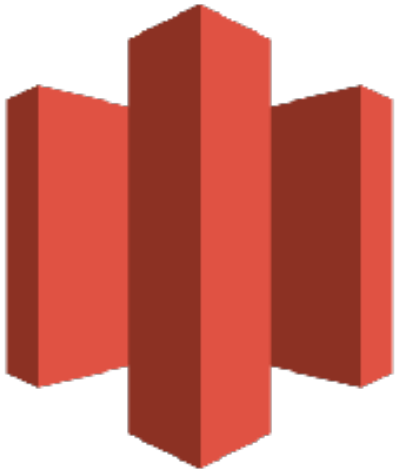
Storage Comparison

	Amazon S3	Amazon Glacier
Data Volume	No limit	No limit
Average Latency	ms	min/hrs
Item Size	5 TB max	40 TB max
Cost/GB Per Month	¢¢	¢
Billed Requests	PUT, COPY, POST, LIST, and GET	UPLOAD and retrieval
Retrieval Pricing	¢ Per request	¢¢ Per request and per GB

Server-Side Encryption



Security with Amazon Glacier



Amazon Glacier

Control access with AWS IAM

Amazon Glacier encrypts your data with AES-256

Amazon Glacier manages your keys for you

Cloud Economics

aws academy

AWS Fundamentals

- Pay for AWS fundamentals:
  - Compute
  - Storage
  - Outbound data transfer
- No charge:
  - Inbound data transfer
  - Data transfer between services within the same region
- Charge for aggregated outbound

AWS Pricing Model

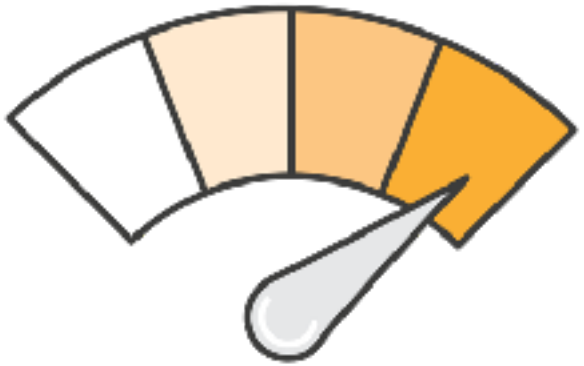
AWS Pricing Policy:

While the number and types of services offered by AWS have increased dramatically, our philosophy on pricing has not changed: at the end of each month, **you pay for what you use**, You can start or stop using a product at any time. No long-term contracts required.

- Pay for what you use
- Pay less when you reserve
- Pay less when you use more
- Pay even less as AWS grows

Pay for What You Use

Pay only for the services you consume, with no large upfront expenses.



- Lower variable costs
- Adapt to changing business needs

Benefits:

- Adapt to changing business needs
- Redirect focus on innovation and invention

Pay Less When You Reserve

Invest in Reserved Instances (RIs):

- Save up to 75%
- Options:
  - All Up-front Reserved Instance **(AURI)** → **largest discount**
  - Partial Up-front Reserved Instance **(PURI)** → **lower discounts**
  - No Upfront Payments Reserved Instance **(NURI)** → **smaller disco**

Benefits:

- Minimize risk
- Predictably manage budgets
- Comply with policies that require longer-term commitments

Pay Less By Using More

Realize volume-based discounts:

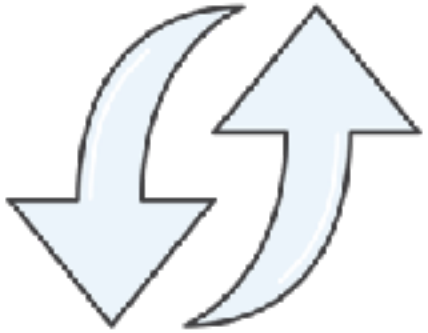
- **Savings** as usage increases.
- **Tiered pricing** for services (for example, Amazon S3, EBS, EFS) → the more you use, the less you pay per GB.
- Data transfer **IN** is always free.
- Multiple storage services deliver **lower** storage costs based on needs.

Benefits:

- Choosing the right combination of storage options **helps you reduce cost** while preserving performance, security, and durability.

Pay Even Less as AWS Grows

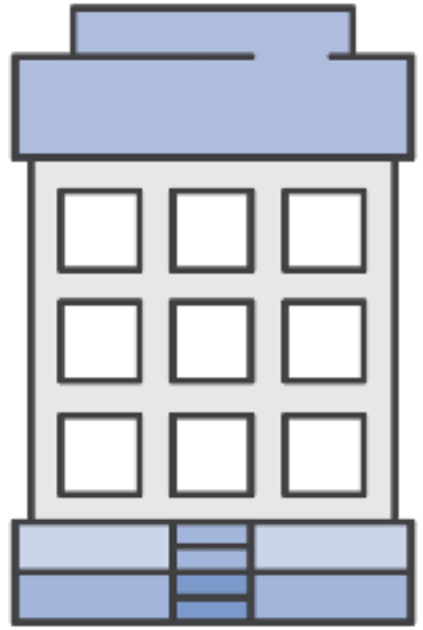
As AWS grows:



- AWS is focused on lowering cost of doing business.
- Results in AWS passing savings from economies of scale to you.
- Since 2006, AWS has **LOWERED PRICING** 61 times.
- Future higher performing resources replace current resources for no extra charge.

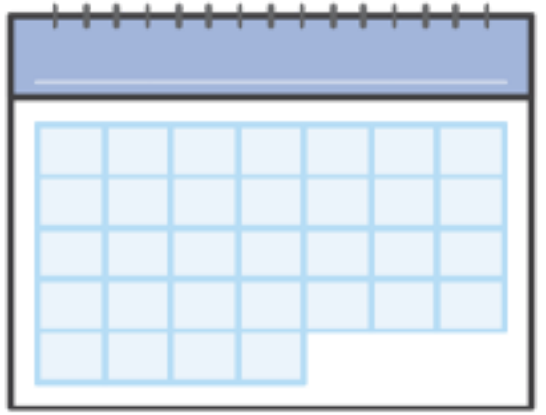
Custom Pricing

- Meet varying needs through custom pricing.
- Available for high-volume projects with unique requirements.



AWS Free Tier

AWS Free Tier helps customers get started in the cloud.



- Limitations:
  - Only new customers
  - Up to one year
  - Applicable to only certain services and options

For more details: www.aws.amazon.com/free

No Charge

AWS services for no additional charge:

- Amazon VPC
- AWS Identity and Access Management (IAM)
- Consolidated Billing
- AWS Elastic Beanstalk**
- AWS CloudFormation**
- Automatic Scaling**
- AWS OpsWorks**

****Note**: There may be charges associated with other AWS services used in conjunction with these services.

On-Premises versus Cloud

Traditional Infrastructure

Equipment

Resources and Administration

Contracts

Cost

≠

AWS Cloud

No Up Front Expense -- Pay for what you use

Improve Time to Market & Agility

Scale Up and Down

Self-service Infrastructure

What is Total Cost of Ownership (TCO)?

Total Cost of Ownership (TCO) is the financial estimate to help identify direct and indirect costs of a system.

Why use TCO?

- To compare the costs of running an **entire infrastructure environment or specific workload** on-premises versus on AWS.
- To budget and **build the business case** for moving to the cloud.

TCO Considerations

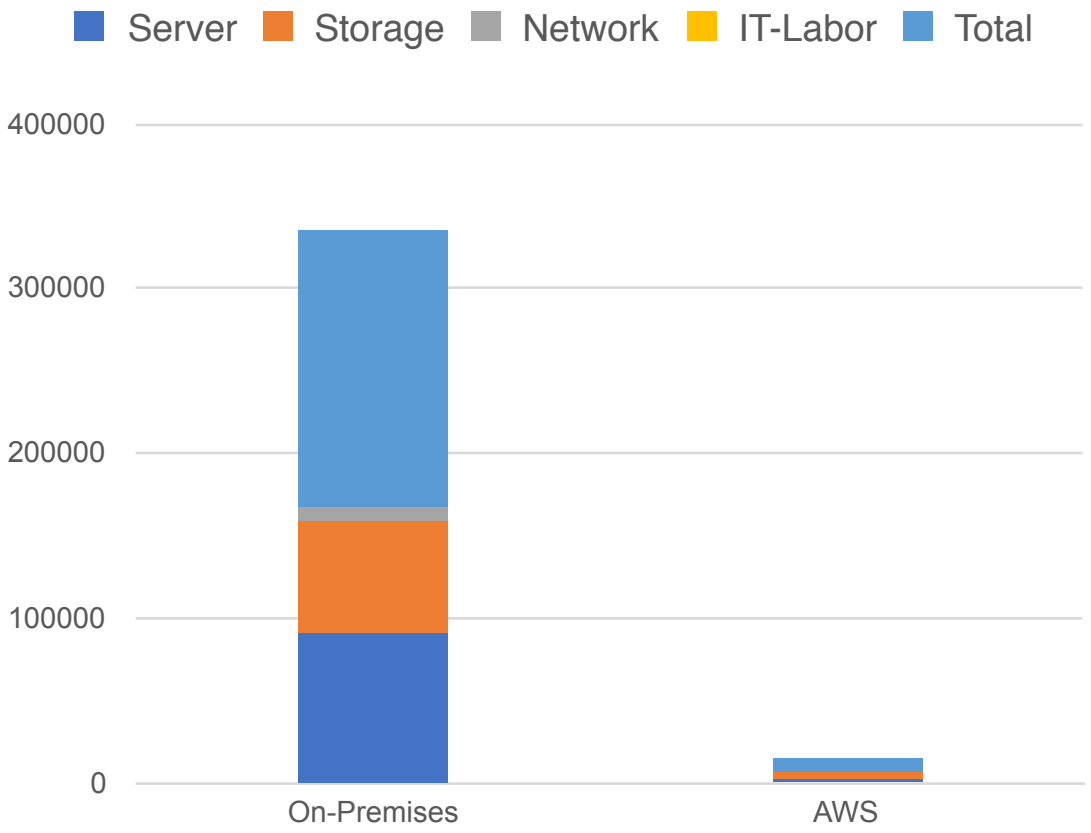
1	Server Costs	Hardware—server, Rack Chassis PDUs, Tor Switches (+Maintenance)	Software—OS, Virtualization Licenses (+Maintenance)	Facilities Cost		
				Space	Power	Cooling
2	Storage Costs	Hardware—storage Disks, San/ Fc Switches	Storage Admin Costs	Facilities Cost		
				Space	Power	Cooling
3	Network Costs	Network Hardware—LAN Switches, Load Balancer Bandwidth Costs	Network Admin Costs	Facilities Cost		
				Space	Power	Cooling
4	IT Labor Costs	Server Admin				

On-Premises versus All-In Cloud

You could save **96%** a year by moving your infrastructure to AWS.
Your three year total savings would be **$159,913**.

3 Year Total Cost of Ownership		
	On-Premises	**AWS**
Server	$91,922	$2,547
Storage	$67,840	$4,963
Network	$7,660	$-------
IT – Labor	$ ----------	$-------
Total	**$167, 422**	**$7,509**

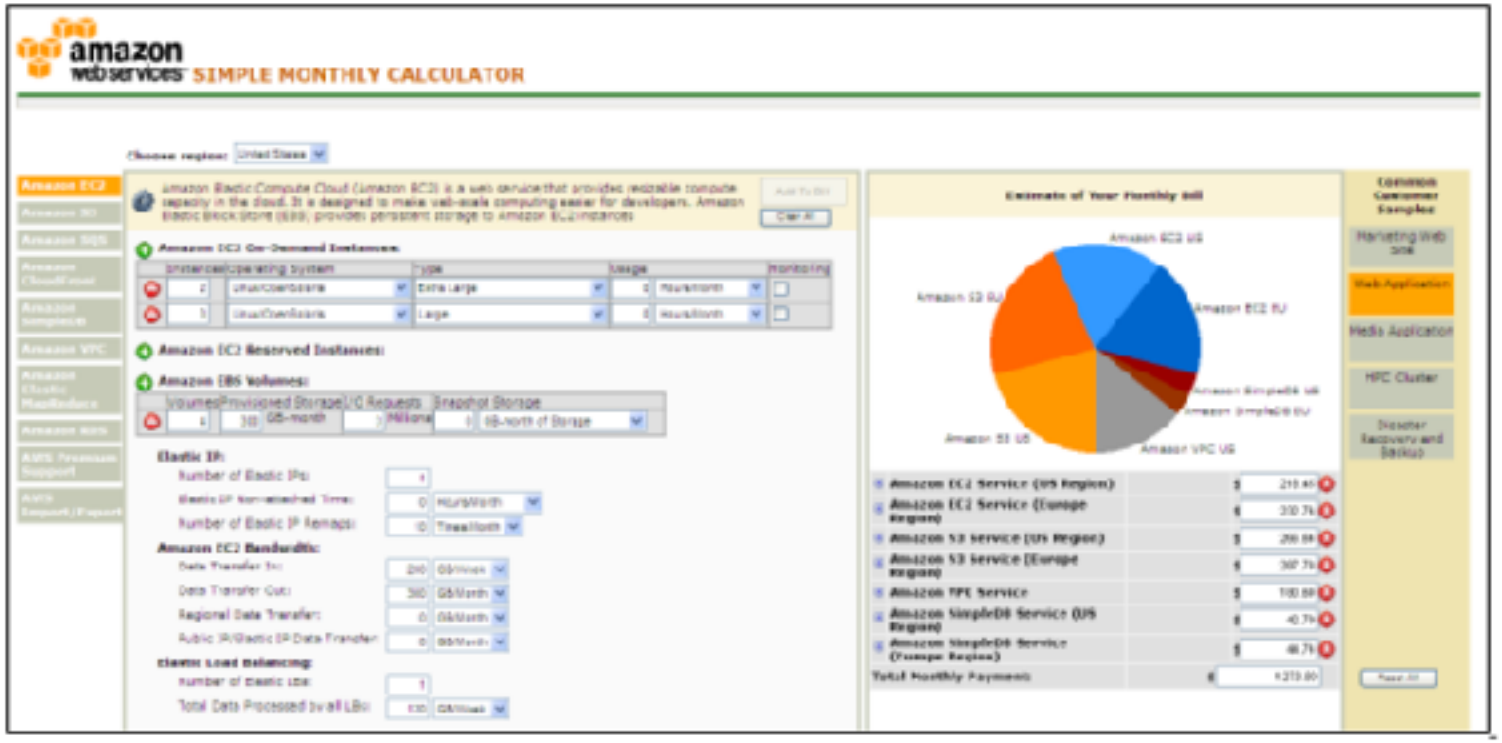
AWS cost includes business level support



AWS Simple Monthly Calculator

Use the **Simple Monthly Calculator** to:

- Estimate monthly costs
- Identify opportunities to reduce monthly costs
- Use templates to compare services and deployment models



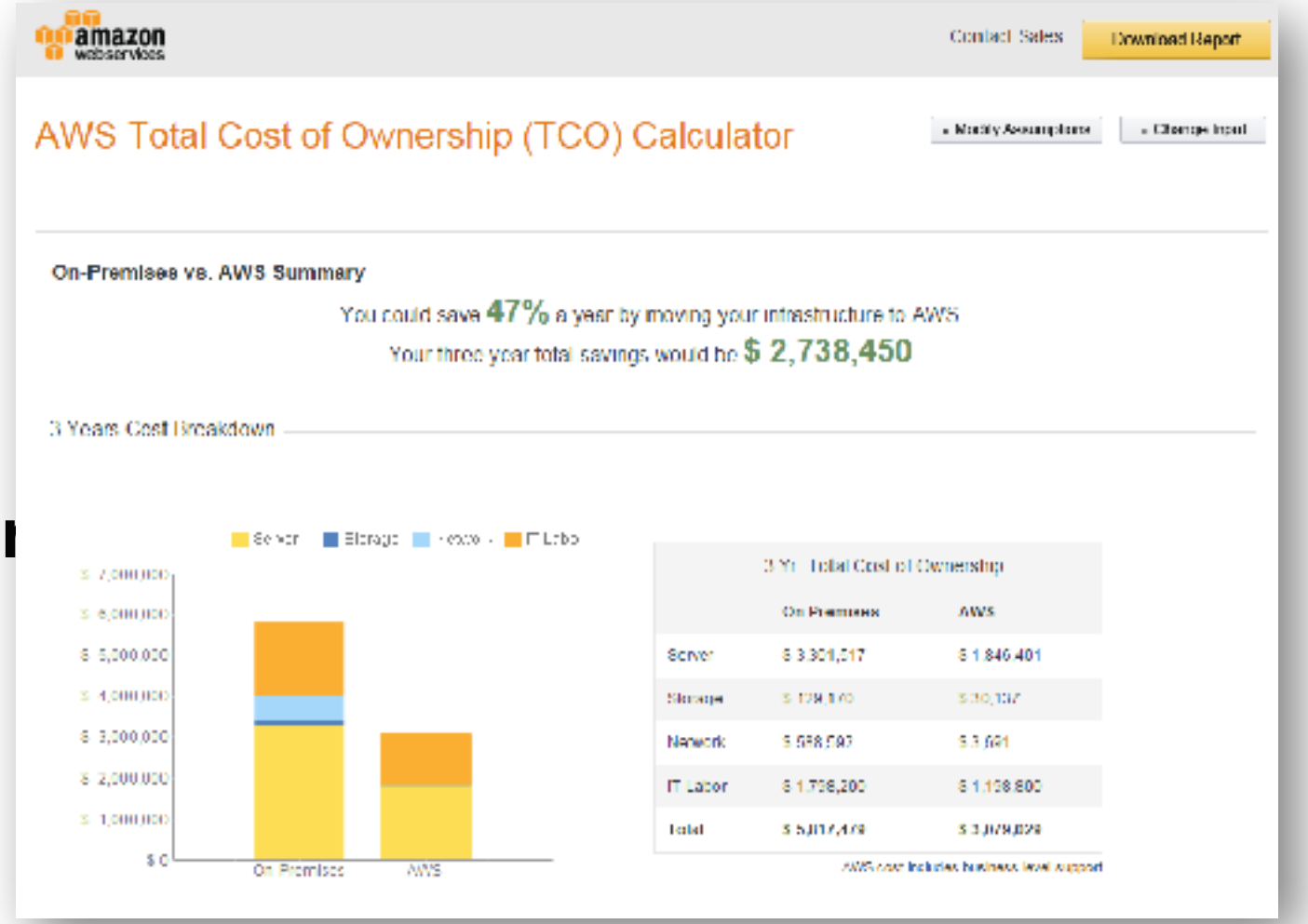
Access the Simple Monthly Calculator.

- http://calculator.s3.amazonaws.com/index.html

AWS TCO Calculator

Use the **TCO Calculator** to:

- Estimate cost savings
- Use detailed reports
- Modify assumptions



Accessing the TCO Calculator

- https://awstcocalculator.com

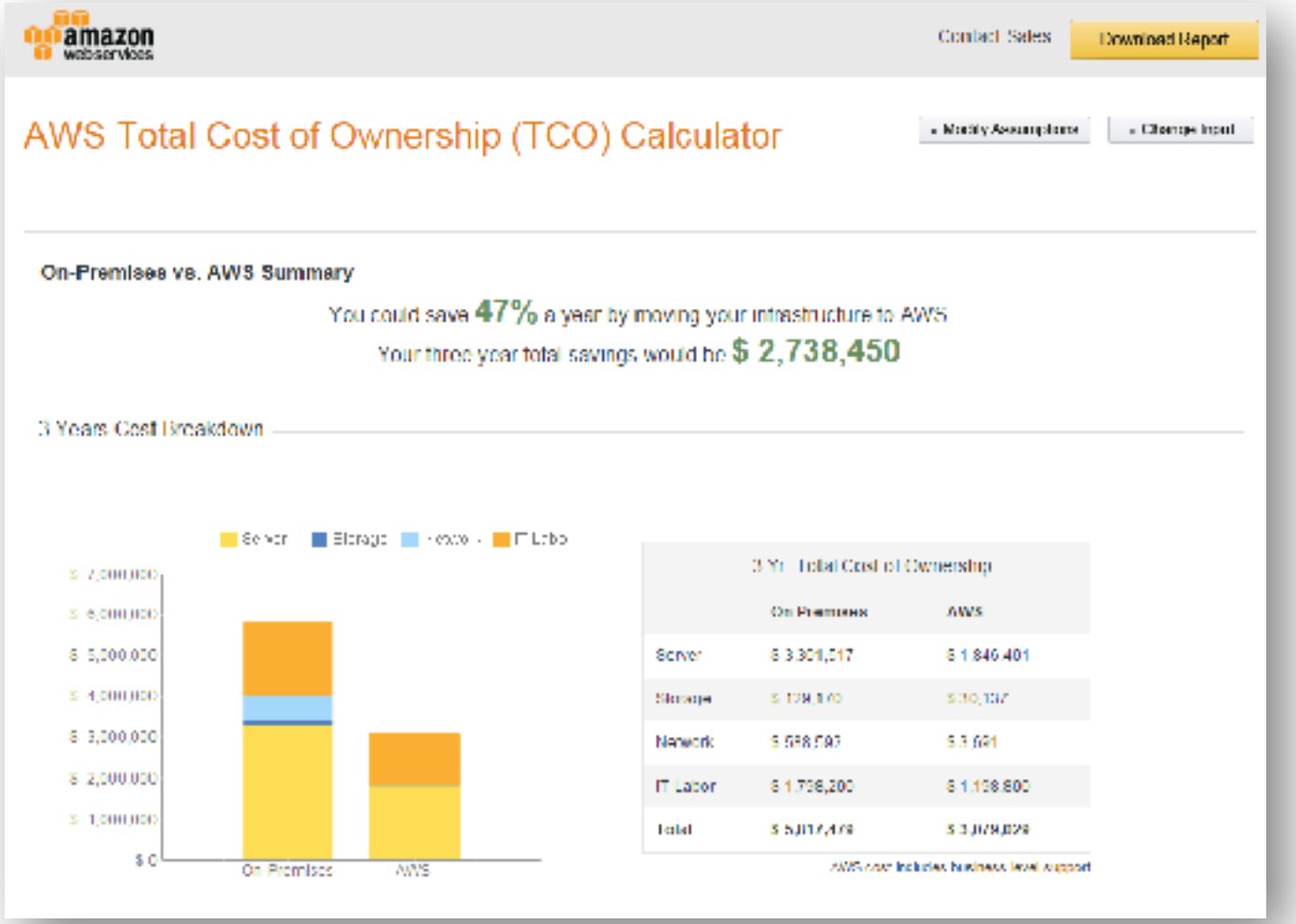
Resources to Get You Started

AWS Economics Center
http://aws.amazon.com/economics/

AWS TCO Calculator
https://awstcocalculator.com

Simple Monthly Calculator
https://calculator.s3.amazonaws.com/index.html

Case studies and research
http://aws.amazon.com/economics/



Thanks for participating!

aws academy