# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>A</th>
<th>4</th>
<th>M</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>5</td>
<td>N</td>
<td>14</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>O</td>
<td>15</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>P</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>R</td>
<td>16</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>S</td>
<td>18</td>
</tr>
<tr>
<td>G</td>
<td>10</td>
<td>T</td>
<td>19</td>
</tr>
<tr>
<td>H</td>
<td>10</td>
<td>U</td>
<td>20</td>
</tr>
<tr>
<td>I</td>
<td>11</td>
<td>V</td>
<td>20</td>
</tr>
<tr>
<td>J</td>
<td>12</td>
<td>W</td>
<td>21</td>
</tr>
<tr>
<td>K</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Amazon Web Services’ official glossary contains hundreds of entries. While we certainly don’t expect you to memorize them all, it’s worth acquainting yourself with some of the key terms used in cloud computing.

This abridged AWS glossary identifies and defines the most important AWS products, services and technical terms. Use it as a primer on all things AWS, as a quick-reference guide in a pinch or as light reading on a slow day.
Access Control List
The document that defines what each type of user can do, such as write and read permissions.

Access Key/Secret Key
Generally used for interactive with the AWS APIs, but can also be used for:

- Command line interfaces (CLIs).
- Temporary access to an AWS account.
- Launching EC2 (Elastic Cloud Compute) instances.
- Storing data in S3 (Simple Storage Service).

You can have up to two access key pairs for a single AWS user at any given time.

Amazon API Gateway
Amazon API Gateway: A managed service that makes it easy for applications to access backend features such as data storage, compute workloads or certain lines of code.

Amazon Aurora
A managed MySQL and PostgreSQL - compatible relational database that is faster, more reliable and more secure than the standard alternatives for one-tenth the cost. Aurora delivers five times the throughput of standard MySQL, is three times faster than standard PostgreSQL databases, and promises 99.99 percent availability. It allows up to 15 low-latency read replicas and significantly improves ease of scalability by auto-scaling up to 64TB per database instance.

Amazon Machine Image
A virtual appliance that contains static data such as the operating system and applications as well its configurations. A virtual machine such as an EC2 (Elastic Compute Cloud) will run this data once launched.
Amazon Resource Name (ARN)
Uniquely identifies AWS resources. An ARN is required to specify a resource unambiguously across all of AWS, such as in IAM policies, Amazon Relational Database Service (Amazon RDS) tags and API calls.

Auto Scaling
An AWS service that automatically increases or decreases instance capacity in order to maintain consistent performance at the lowest cost. The Auto scaling web service can be configured for Amazon EC2 instances, Amazon ECS tasks, Amazon DynamoDB tables and indexes, and Amazon Aurora Replicas.

Auto Scaling Group
The logical grouping of multiple EC2 instances based on certain shared characteristics for the purposes of simplifying resource provisioning through Auto Scaling.

Availability
Broadly refers to anytime, anywhere access to service. More specifically, it can be represented as a percentage of the total amount of time that a cloud service must be available in a given service-level agreement (e.g., 99.99 percent).

Block Device
Any data storage hardware that supports reading and/or writing in fixed block amounts, usually in multiples of 512 bytes.

Blue/Green Deployments
A core function of AWS CodeDeploy that allows developers to test a new application version before sending production traffic to it. In effect, the original instances are replaced by a new set of instances. If necessary, developers can easily roll back to the previous version which exists separately in the original environment.

Bucket
An Amazon S3 container for stored objects. Buckets must be created in a specific AWS region before objects can be stored (i.e., before you can upload any data). They are similar to a URL domain in that every bucket name is globally unique, and that objects can be classified within them. For example, the bucket in “http://johnsmith.s3.amazonaws.com/photos/puppy.jpg,” is “johnsmith” and the object is “photos/puppy.jpg.”
AWS Burstable Instance
Also known as EC2 T2 instances, these are general purpose instances that run at a baseline CPU performance but have the ability to “burst” above that baseline to support transient loads. T2 instances are the lowest cost EC2 instances.

Capacity
The amount of available compute power.

Certificate
A public key infrastructure (PKI) credential used to authenticate AWS accounts and users for certain AWS products.

Amazon CloudFront
A global content delivery network (CDN) that helps deliver content to worldwide users with the lowest possible latency.

Cloud Native
A program designed specifically for cloud architecture. Tasks in a cloud-native application are usually broken down into separate services that can run on different servers in different locations, and are backed by hardware redundancy.

Cloud Service Provider (CSP)
A company (e.g., AWS) that makes internet-hosted computing, storage and software services accessible to subscribers.

Amazon CloudWatch
A web service that lets users monitor metrics associated with AWS resources (e.g., CPU utilization, data transfer and disk usage for EC2) and configure alarm actions based on data from those metrics.

Cluster
The connection of two or more container instances to logically group tasks or services. It also refers to a grouping of servers that effectively function as a hive in which individual nodes are instructed to complete specific tasks.
CNAME
Canonical Name Record; this allows you to alias distinct domain names to one another.

Consolidated Billing
The grouping of multiple AWS accounts into a single billing statement. This feature allows for both an individual view of costs per account and a combined view.

Container
A virtualized compartment within an operating system. Containerization allows different applications to share the same OS kernel, thereby avoiding the duplication of that resource. Containers also enable applications to be more easily moved between environments, making them easier to distribute and strip down into specific functions.

Containerized Solution
An application that exists within a Docker container rather than in its own partition on a virtual machine.

Content Delivery Network (CDN)
A global network of servers that speeds up the delivery of content. When a user requests certain content from anywhere in the world (e.g., attempts to load a webpage) that request is routed to edge servers in a data center that has the lowest latency. Amazon CloudFront is a CDN.

Continuous Delivery
A DevOps application development practice in which changes to software are automatically built into the existing code and then tested in a staging environment before being deployed.

Continuous Integration
The continuous merging of code changes in a central repository in order to automate the build and testing of an application. Continuous integration is a precursor to continuous delivery.

Cross-Region Replication (CRR)
A client-side solution that automatically replicates objects within S3 buckets across different AWS regions.
**Data Consistency**
Describes the circumstance by which, when data is written or updated, all copies of that data will follow suit it in all other AWS locations.

**Dedicated Instance**
An Amazon EC2 instance that runs in a virtual private cloud (VPC) on hardware dedicated to a single customer.

**Delegation**
The act of giving one or more users access to certain resources within your AWS account; alternatively, granting users of a separate account access to a resource within your own account.

**Distribution**
The process of linking origin servers (where the original content is stored) to local edge servers that are part of a CDN in order to speed up the delivery of static and dynamic content to users making traffic requests.

**DNS (Domain Name System)**
Routes traffic to and from websites by translating domain names into IP addresses.

**Amazon DynamoDB**
A fully managed NoSQL database that provides consistent single-digit millisecond latency regardless of the scale. DynamoDB provides automatic scaling, encryption at rest and other administrative features that simplify database management and configuration.

**EC2 (Amazon Elastic Compute Cloud)**
Secure, resizable compute capacity in the cloud.
**EC2 Instance**
Any compute deployment within the Amazon EC2 service.

**Edge Location**
A site at the edge of the network that a CDN such as CloudFront will use to cache copies of content for lower-latency delivery over great distances. This is in contrast to an origin location, which is where the original content is housed.

**Elastic**
The quality of being able to automatically provision and deprovision compute and storage resourcing in order to support fluctuating workloads. Elasticity is a core competency of cloud computing.

**Elastic IP Address**
AN IP address tied to your account rather than a specific instance.

**Amazon Elastic File System (EFS)**
Provides elastic file storage for AWS and on-premises resources alike. EFS can be mounted on EC2 instances and proprietary servers via the NFSv4 protocol. Accordingly, applications that scale beyond a single instance can share a file system. EFS is highly scalable and affordable, and delivery low-latency and high throughput.

**Elastic Network Interface**
A logical networking interface that contains certain attributes such as private primary IP addresses.

**Environment**
May refer broadly to your total AWS footprint (compute, storage, network, database) or more specifically to a networking environment (CDN), or to an application environment on a server or virtual machine.

**Elastic Load Balancing**
A web service that automatically distributes application traffic across EC2 instances, containers and IP addresses in one or more availability zones. Elastic Load Balancing supports hybrid load balancing, meaning on-premises resources and AWS resources can share a load balancer.
Amazon Elastic Container Service (ECS)
Simplifies the deployment and management of containerized applications on AWS. Amazon ECS makes it easy to run, stop and manage Docker containers on a cluster.

Function as a Service (FaaS)
A category of cloud computing that lets developers create application functions without worrying about the infrastructure typically needed for deployment. FaaS is a core function of serverless computing such as AWS Lambda. In effect, a developer can run code (functions) that will be automatically loaded into containers when a client-side request is made. This reduces the amount of server-side work.

Gateway
Broadly describes hardware or software that bridges networks, i.e., connecting a customer’s home router to remote cloud storage. Specific types of gateways include customer gateways, internet gateways and NAT gateways.

Health Check
Something that checks the health of a system. A health check can be done on several specific endpoints to ensure the health of the system.

High Availability
A characteristic of a service that has very low rates of failure. Represented as 100 percent being “never failing,” 99.999 percent continuity or more is considered high availability.
Horizontal Scaling
Increasing capacity by adding more hardware or software components. Horizontal scaling tends to be more efficient than vertical scaling since it doesn’t require a complete replacement of existing hardware, and can be executed without need for downtime.

Hypervisor
A resource used to create and launch virtual machines (VMs). AWS, which formerly used Xen for this purpose, now uses KVM (Kernel Virtual Machine).

AWS Identity and Access Management (IAM)
An AWS service that lets the customer create identities (could be groups, individual users or specific endpoints) and manage the level of access those “identities” have to certain cloud resources.

IAM Group
A collection of IAM users.

IAM Role
A tool that provisions temporary access to a resource for a user or group of users.

IAM User
A person or application associated with his or her/its own AWS identity.

Infrastructure as a service (IaaS): The foundation of cloud computing. An IaaS provider like AWS supplies the computer power, storage and networking upon which other cloud services are constructed.

Instance
Amazon-specific nomenclature for running something.

Instance Type
Defined by the memory, CPU, storage capacity and usage cost of a particular instance.
**Intrusion Detection System (IDS)**
Software used to monitor for indicators of an attack or intrusion against a network or application.

**In-Memory-Caching**
Random access memory (RAM) that a microprocessor can access more quickly than it does regular RAM. An in-memory key-value data store like Redis can deliver sub-millisecond response times, which improve application response times.

**JSON**
Stands for JavaScript Object Notation; format that uses readable text to represent simple data structures and objects.

**Key**
Key has multiple applications in AWS:
- Public and private key pairings (used as security credentials in IAM)
- Access key ID and secret access key pairings (used to cryptographically authenticate programmatic AWS requests)
- Customer master keys (used via AWS Key Management Service to encrypt or decrypt data)
- Primary keys that take the form of partition keys or sort keys (used to identify each item in an Amazon DynamoDB table)
- A key prefix (logical grouping of key pairings in a bucket)

**Key Pair**
Refers to the public and private keys that are used as credentials to verify a user’s identity electronically.
AWS Key Management Service (AWS KMS)
Managed service that simplifies the creation and control of keys for data encryption. AWS KMS hardware security modules that adhere to FIPS 140-2 (a U.S. Government-imposed benchmark for implementing cryptographic software).

Load Balancer
Combination of DNS name and ports that, together, distribute requests among application instances within a region. Load balancers are used to increase capacity (concurrent users) and reliability of applications. They improve the overall performance of applications by decreasing the burden on servers associated with managing and maintaining application and network sessions, as well as by performing application-specific tasks.

Logical Name
A customer-generated string of text within an AWS CloudFormation template that stands in for the physical ID of a resource, mapping, parameter or output.

Lambda
AWS’ serverless computing offering.

Log Aggregation
The act of centralizing the log data created by your various IT systems and software. Log data refers to the time-stamped documentation of events. Every IT system/application has a log. With log aggregation, the management and monitoring of those logs is simplified.

MFA (Multi-Factor Authentication)
An AWS account security feature that uses two forms of authentication: knowledge (e.g., a password) and possession (e.g., a one-time password sent to a mobile device).

Multi-AZ Deployment
Configuration of AWS services to span multiple Availability Zone (AZ), either through failover or redundancy.
**Multitenancy**
The ability to use a single software instance installed on multiple servers to serve multiple customers (or tenants).

**Microservices**
Refers to the development of applications as a suite of independently launched modular services, each of which performs its own specific function.

**Multicloud**
An IT environment that uses more than one cloud type (private and public) or cloud vendor.

**Managed Service Provider**
Any vendor that remotely manages a client's IT infrastructure, or provides access to a fully managed application, for a subscription fee.

**Multiregion**
The availability of cloud instances in multiple geographic regions simultaneously. In AWS, each region contains a set of more localized availability zones. AWS contains 18 regions and 54 availability zones.

**Memcached**
An open-source memory-object caching system that minimizes the frequency by which a database or API needs to be externally accessed. This helps to speed up load times for dynamic web content.

**NAT**
Network address translation; the remapping of one or more IP addresses to a different IP address while data packets are in transit across a traffic routing device.

**Node**
An Elasticsearch instance, e.g., a data instance or dedicated master instance.
DEFINITIONS: N-P

NoSQL
Highly available, scalable and high-performance non-relational database systems (e.g., Amazon DynamoDB) that rely on key-value pairs or document storage for data management.

Network File System (NFS)
A distributed file system protocol that users easily access a remote network of files as though it were on a local machine. An example is Amazon Elastic File System (EFS).

Object
Any entity type stored in Amazon S3.

Origin Server
A server that listens for and processes incoming internet requests. In edge computing, an origin server communicates with users through an intermediary set of edge servers that are part of a content delivery network (CDN).

On-Premises
IT infrastructure or applications that are stored and managed in the same location as the people who access it.

Permission
A statement in a policy that permits or prohibits access to a specific resource.

Policy
In IAM, a policy is a set of permissions that define what a user, group or role can do in AWS. In Auto-scaling, certain instances can be launched or terminated according to user-defined policies.

Private IP
A non-internet facing IP address that networked devices use for internal communications.
PV Virtualization
Paravirtual virtualization; virtualization in which a guest operating system runs on host systems without being able to issue hardware commands.

Platform as a Service (PaaS)
A cloud computing model in which the hardware and software needed for application hosting is provided for the customer. With PaaS, middleware, servers, storage, networking, runtime, etc. are all the vendor’s responsibility. The client is free to focus on management of his or her applications. AWS offers a combination of IaaS and PaaS resources. AWS Elastic Beanstalk is an example of the latter.

Private Cloud
A cloud model whereby a single client has access to an isolated set of managed IT resources. AWS’ private cloud offering is called Amazon Virtual Private Cloud (Amazon VPC).

Public Cloud
A cloud model whereby multiple clients can access a set of shared IT resources that are leased on a pay-as-you-go basis.

Provider
In cloud computing, may refer to a CSP (cloud service provider) such as Amazon that provides internet-hosted computing, storage, and software services. There are generally three categories of cloud provider: Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS).

Production
Refers to the state whereby an application has been launched and made available for public use.

RDS (Relational Database Service)
Managed AWS service that simplifies setup, management and scaling of a cloud-based relational database.
Redshift
A fully managed, petabyte-scale data warehouse primarily used to run data analysis via existing Business Intelligence (BI) tools.

Reserved Instance
Discounted on-demand EC2 instance usage, provided those instances meet specified parameters.

Resource
Any entity that users can work with in AWS including but not limited to an EC2 instance, an Amazon DynamoDB table and an Amazon S3 bucket.

REST
Representational state transfer. Stateless architecture that conveys textual representations of web sources rather than objects, thereby improving interoperability between disparate systems relying on multiple programming languages.

RESTful Web Service
Lightweight, highly scalable web service that adheres to REST architectural constraints.

Root Device Volume
Contains the image used to boot an instance.

Region
A geographical cluster of AWS resources; must contain at least two availability zones. There are 18 AWS regions.

Redundancy
The quality of having duplicate data and/or resources in order to ensure availability in the event of equipment downtime or other disruption.

Redis
An open-source, in-memory key-value data store that delivers sub-millisecond response times to improve application response times. (See “in-memory caching.”)
Amazon Simple Service Storage (S3)
Highly durable, scalable and available object storage for frequently accessed data.

Sandbox
A partitioned, virtual testing space used to trial application functionality without risk of disrupting any processes.

Scalability
Describes the ability of a system, process, network or application to quickly adjust to increased capacity demands. Scalability is a prominent feature of cloud computing.

Search Instance
A subset of an Amazon CloudSearch domain that indexes data and processes search requests.

Secret Access Key
The corresponding component of an access key ID, which when used together, cryptographically signs programmatic AWS requests.

Server-Side Encryption (SSE)
Encryption of Amazon S3 data at rest; decryption occurs upon access.

Serverless
A method of cloud computing in which a function of an application is automatically loaded into a container when a client-side request is made for that function. This reduces the amount of server-side work, since programming code is able to exist with minimal dependency on server middleware. (See Lambda, FaaS.)

Service Level Agreement (SLA)
A contract with a cloud vendor that identifies terms of service using specific, quantifiable metrics. These may include service availability, latency, security and other performance and reliability attributes.
Amazon Simple Email Service (SES)
A highly scalable, fully managed email platform that enables easy sending and receiving.

Amazon Simple Notification Service (SNS)
A fully managed web service that lets applications, end-users and devices send/receive notifications from the cloud.

Amazon Simple Queue Service (SQS)
A fully managed message queuing service that facilitates communication between distributed application components.

Software as a Service (SaaS)
The leasing of a software license for an application that is fully managed in a remote data center. In SaaS, the underlying infrastructure as well as the application platform is managed and/or paid for entirely by the vendor. The client pays a subscription fee for access.

Staging
Refers to a virtual environment for final testing of an application prior to production (live deployment).

Throttling
An automatic, calibrated slowdown of operations in response to certain limitations.

TLS (Transport Layer Security)
A cryptographic protocol used to secure communications over the internet.

Tunnel
A route for the transmission of encrypted private network communications over a public network.
User
A person or application associated with an account who/that needs to trigger an API action to AWS products.

Utility Computing
A type of on-demand computing in which services are provisioned and resources are made available to the customer as needed. Cloud computing is the most well-known example of utility computing.

Utility Billing
A pay-as-you-go billing model in which customers only pay for the services and resources that they use. This describes most billing methods used in cloud computing. All AWS service offerings use utility billing.

Versioning
The act of keeping every version of an object stored in an Amazon S3 bucket so that it can be easily recovered in the event of an application failure or unprompted user action.

VPC (Virtual Private Cloud)
An isolated virtual network within an AWS cloud.

VPN (Virtual private network)
Secure encrypted network connection of a VPC to public internet.

Vertical Scaling
Increasing capacity by adding more computer power (e.g., CPU or RAM) to existing servers.

Virtual Machine
A virtual partition on a server that effectively functions as its own server.
DEFINITIONS: W-Z

(W)

WorkDocs
A fully managed file storage and sharing service.

WorkMail
A managed business email and calendar service that supports seamless integration with a variety of client email applications.

Web application firewall
A security resource that enables an administrator to block certain types of web traffic requests.

(Z)

Zone Awareness
A configuration that distributed Amazon Elasticsearch instances (nodes) in two different Availability Zones (distinct location within a geographical region) in order to minimize downtime and/or data loss in the event of a note or data center outage.