Module 3. Core AWS Services

This module covers the following subjects:

- **Overview of Services and Categories**: In this module section, we lay out the most important categories of services you should be familiar with at this stage of your AWS education.

- **Introduction to the AWS Global Infrastructure**: This part of the module ensures you understand the various components that make up all of the Amazon-owned data resources worldwide.

- **Introduction to Virtual Private Cloud**: The Virtual Private Cloud (VPC) is your virtual network inside the AWS Cloud. This part of the module ensures that you understand the possible components you can use.

- **Introduction to Security Groups**: Security groups are built-in (and often mandatory) firewalls you use to control traffic into and out of AWS resources. This part of the module describes them in detail for you.

- **Introduction to Compute Services**: The primary compute resource in AWS is the EC2 instance. This module section teaches you about EC2 and discusses an exciting alternative—serverless computing with Lambda in AWS.

- **Introduction to EBS**: This part of the module teaches you the basics of AWS's Elastic Block Store (EBS) storage. This storage is used frequently for the disks required by EC2 instances.

- **Introduction to S3**: This section of the module teaches you detailed information on AWS's Simple Storage Service (S3). S3 is flexible, object-based reliability storage that can serve many purposes.

- **Introduction to AWS Database Solutions**: Database technology is vital as it continues to drive IT. AWS provides many different options for cloud-based database solutions, and this part of the module covers the most important ones as of this writing.

It is essential to realise that AWS always adds services to its existing lineup of cloud-based IT offerings. With that said, we must consider certain services in AWS as core. This is because they are mature, heavily used, and extensively used by Amazon customers. This module seeks to get you very familiar with these core services. This will not only help you

when you implement your AWS solutions but also help you on your exam, where you are expected to know these details very well.

FOUNDATION TOPICS

OVERVIEW OF SERVICES AND CATEGORIES

AWS includes many different services, and more are being added constantly. As a result, there is also a growing list of AWS service categories. For example, machine learning would not have been a category within AWS years ago. Still, it is garnering much attention now, and soon, it will have its speciality certification in the AWS certification lineup. Figure 3-1 shows just some of AWS's service categories and services.
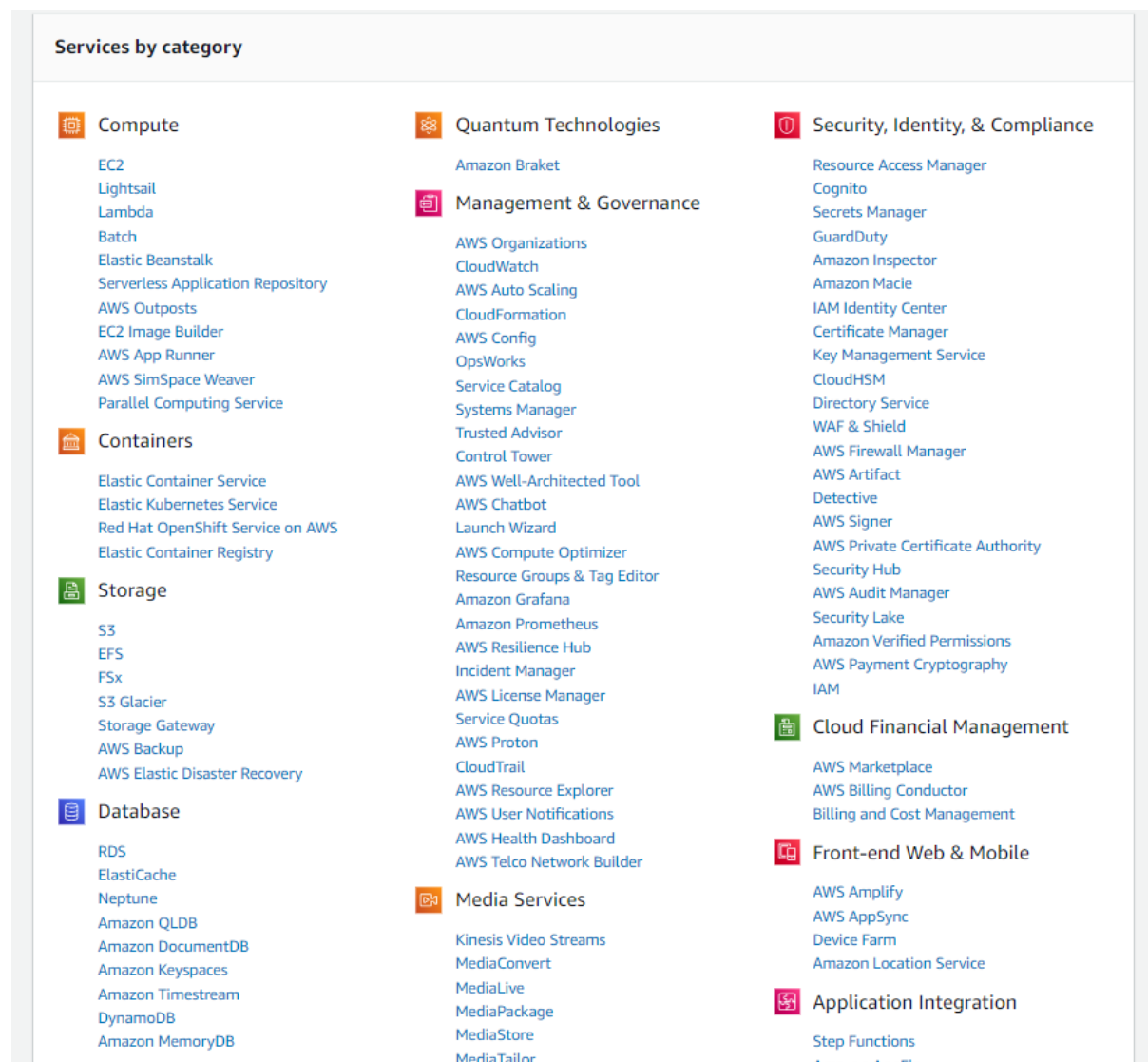
Services by category

Compute
EC2
Lightsail
Lambda
Batch
Elastic Beanstalk
Serverless Application Repository
AWS Outposts
EC2 Image Builder
AWS App Runner
AWS SimSpace Weaver
Parallel Computing Service

Containers
Elastic Container Service
Elastic Kubernetes Service
Red Hat OpenShift Service on AWS
Elastic Container Registry

Storage
S3
EFS
FSx
S3 Glacier
Storage Gateway
AWS Backup
AWS Elastic Disaster Recovery

Database
RDS
ElastiCache
Neptune
Amazon QLDB
Amazon DocumentDB
Amazon Keyspaces
Amazon Timestream
DynamoDB
Amazon MemoryDB

Quantum Technologies
Amazon Braket

Management & Governance
AWS Organizations
CloudWatch
AWS Auto Scaling
CloudFormation
AWS Config
OpsWorks
Service Catalog
Systems Manager
Trusted Advisor
Control Tower
AWS Well-Architected Tool
AWS Chatbot
Launch Wizard
AWS Compute Optimizer
Resource Groups & Tag Editor
Amazon Grafana
Amazon Prometheus
AWS Resilience Hub
Incident Manager
AWS License Manager
Service Quotas
AWS Proton
CloudTrail
AWS Resource Explorer
AWS User Notifications
AWS Health Dashboard
AWS Telco Network Builder

Media Services
Kinesis Video Streams
MediaConvert
MediaLive
MediaPackage
MediaStore
MediaTailor

Security, Identity, & Compliance
Resource Access Manager
Cognito
Secrets Manager
GuardDuty
Amazon Inspector
Amazon Macie
IAM Identity Center
Certificate Manager
Key Management Service
CloudHSM
Directory Service
WAF & Shield
AWS Firewall Manager
AWS Artifact
Detective
AWS Signer
AWS Private Certificate Authority
Security Hub
AWS Audit Manager
Security Lake
Amazon Verified Permissions
AWS Payment Cryptography
IAM

Cloud Financial Management
AWS Marketplace
AWS Billing Conductor
Billing and Cost Management

Front-end Web & Mobile
AWS Amplify
AWS AppSync
Device Farm
Amazon Location Service

Application Integration
Step Functions
Amazon AppFlow

Figure 3-1 Some Service Categories in AWS

Note

Remember, you can always check out my AWS web site (https://aws.amazon.com/certification) for the latest versions of AWS certifications and their requirements.

Here are the leading service categories you should be aware of at this point:

- Compute
- Storage
- Database
- Networking and Content Delivery
- Security, Identity, and Compliance

INTRODUCTION TO THE AWS GLOBAL INFRASTRUCTURE

The AWS Global Infrastructure is just that—it is global. It is made up of the latest and greatest networking and data center technologies, and it spans (almost) the entire globe.

To understand it, you should break the AWS Global Infrastructure down into three main components:

- **Regions**: Geographic locations that host two or more Availability Zones (AZs). You choose regions based on availability requirements for your resource locations, and you can even host resources in multiple regions to serve different customers in different locations. Regions are treated as separate entities, and, by default, information does not transfer between regions. Finally, not all regions offer all AWS services.

- **Availability Zones (AZs)**: A collection of regional data centres. The data centres are separated as much as possible for high availability but are connected using high-speed links. Each AZ in a region uses its power provider (or multiple power providers) and has its own backup generators; some even have power substations.

- **Edge Locations**: These locations host cached content from your architecture for fast client delivery. AWS's caching technology is CloudFront. Edge Locations add even more low-latency reach to your clients (beyond the already amazing reach of regions and AZs). They

are also entry points into the AWS network when using CloudFront or S3 Transfer Acceleration for data ingestion.

- **AWS Local Zones** are a new addition to Amazon Web Services' global infrastructure that brings AWS services closer to users, reducing latency for various applications. These zones are particularly beneficial for industries requiring real-time data processing, such as healthcare, financial services, and gaming. By deploying resources in a Local Zone, businesses can ensure low-latency access to their applications, meeting performance and data residency requirements. AWS Local Zones also simplify hybrid cloud migrations, allowing organisations to maintain the latency-sensitive aspects of their operations while leveraging the scalability of AWS.

- **AWS Wavelength** is designed to bring AWS services to the edge of the mobile network, minimising latency to deliver ultra-fast application performance. It enables developers to build applications that serve end-users with single-digit millisecond latencies over 5G networks. By deploying AWS compute and storage services closer to the end-users, Wavelength reduces the time it takes for data to travel, enhancing user experiences in gaming, live streaming, and other latency-sensitive tasks. This innovative approach transforms how applications interact with the network, pushing the boundaries of modern technology and opening new possibilities for developers and businesses alike.

INTRODUCTION TO VIRTUAL PRIVATE CLOUD

When you think about virtual networking in the public cloud of AWS, think of Virtual Private Clouds (VPCs). A VPC allows you to create private virtual networks and use the same concepts you already know from traditional networking. With a VPC, you have complete control of your network configuration. You can isolate resources from or expose resources to the public Internet or your private host systems inside your corporation.

VPC has several layers of security controls (as you would expect). For example, you have security groups and network access control lists (ACLs) that act as firewalls, to name just two. Security groups control traffic in and out of your Elastic Compute Cloud (EC2) instances, whereas network ACLs permit the control of traffic in and out of your subnets in the VPC.

Note

While it is true that you use security groups to control traffic in and out of your EC2 instances, understand that they do this by managing traffic in and out of the Elastic Network Interfaces (ENIs) used by your EC2 instances. For example, if you were to create a security group protecting one of your EC2 instances and move its ENI to another EC2 instance, the security group protections would also move.

When you are architecting solutions in AWS, you deploy various services and resources into your VPC to make up the complete solution. You can even be very specific with service and resource placement, so you know exactly where the resources reside in your virtual network. These services also conveniently inherit the security you have built into your network. Examples of services you would deploy into your VPC include Elastic Compute Cloud, Elastic File System, Relational Database Services, Elastic Load Balancing, and many more.

Here are the main features of VPC you should be aware of:

- **High availability**: VPCs build upon the high availability built into AWS regions and Availability Zones (AZs). Your VPCs live within a region, and you can have multiple VPCs per account.

- **Subnets**: Just like in your private network infrastructures, VPCs are made up of subnets you can use to provide segmentation at Layer 3 (the Network layer).

- **Route tables**: Route tables can route traffic entering and exiting your subnets. This familiar model allows you to avoid worrying about the physical routers themselves.

- **Internet Gateway (IGW)**: Permits easy-to-configure access to the Internet for your VPC; Figure 3-2 shows an Internet Gateway in AWS.
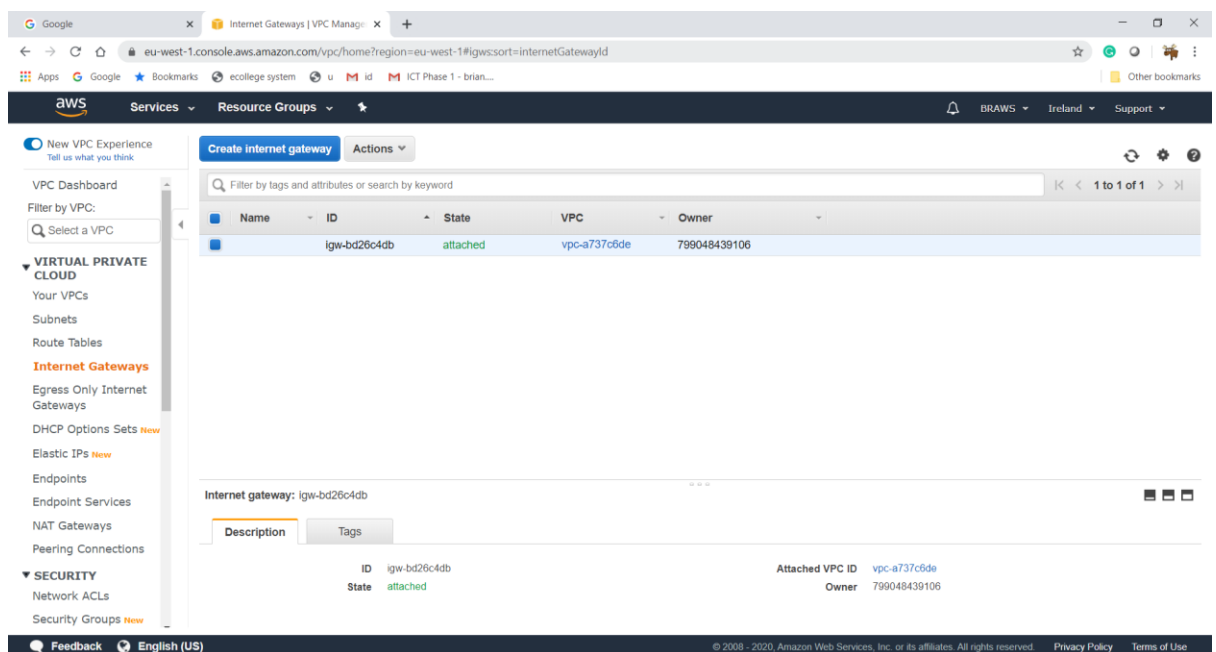

Figure 3-2 An Internet Gateway

- **NAT Gateway:** Translates your privately addressed VPC resources to access the Internet using public IP addresses.
- **NACLs**: Network access control lists allow you to control access to your VPC subnets; these are stateless constructs, meaning you must configure inbound and outbound rules, as there is no automatic recognition of state with traffic flows and no automated access entries.

INTRODUCTION TO SECURITY GROUPS

Because the security of your resources in the cloud is a prime concern for you and Amazon, it is no big surprise that AWS provides you with built-in firewalls with your compute resources. These security groups help you easily control the accessibility of your EC2 resources, for example. <u>Figure 3-3</u> shows an example of a security group in AWS.
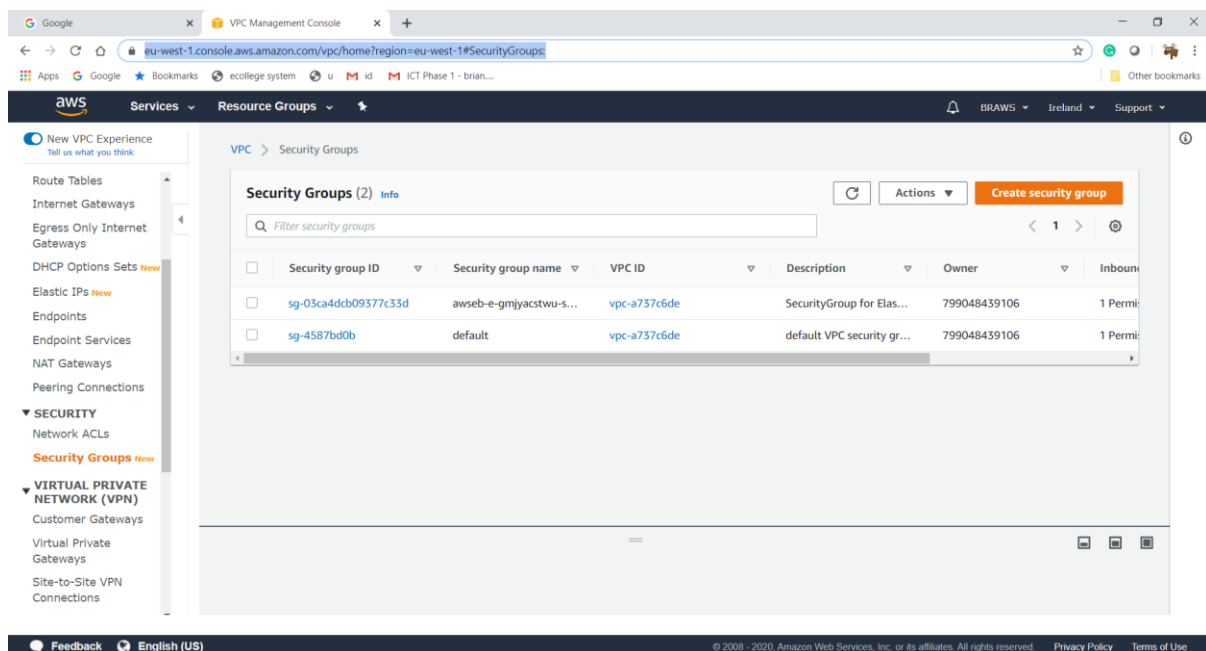

Figure 3-3 An AWS Security Group

Perhaps you have a Web tier in your AWS architecture. You can configure the security group for this tier to permit HTTP and HTTPS traffic from customers using the Web tier. At the same time, you can allow your team of support engineers to access the Web tier using SSH and RDP. The security group denies all other protocol attempts at accessing the Web tier.

INTRODUCTION TO COMPUTE SERVICES

Because computing resources are so significant to our IT solutions, it is no surprise that Amazon has many services that center around computing resources. Although the primary job of this section is to provide you with valuable information regarding the primary computing resource in AWS, Elastic Compute Cloud (EC2), you first need an understanding of the other computing resources available to you.

Lambda

AWS Lambda is an exciting alternative to EC2 instances that you must operate and maintain. Lambda provides computing resources in a fully managed (by AWS) serverless computing cloud. You send compute requirements to Lambda in various ways (such as a call from a web app), and Lambda takes care of the compute requirements for you. Sub-second metering is used for your cost calculations, so providing the computing resources you require is often inexpensive. It also supports many different programming languages for ease of use. With Lambda, a typical workflow follows these steps:

1. You upload your code to AWS Lambda.
2. You set up your code to trigger from other AWS services, HTTP endpoints, or in-app activity.
3. Lambda runs your code when triggered, only consuming the resources needed; it is essential to realise that, like most AWS services, Lambda provides continuous scaling as required.
4. You pay for just the compute time required.

Fargate

AWS Fargate is a serverless compute engine that simplifies deploying containers without the need to manage servers or clusters. It integrates with Amazon ECS and Amazon EKS, offering a pay-as-you-go model that allows for flexible application scaling and management. With AWS Fargate, developers can focus on building and deploying applications, as it handles the operational aspects such as server provisioning, scaling, and monitoring. This service supports a variety of CPU architectures and operating systems, catering to a wide range of use cases, from web applications to AI and machine learning workloads.

Elastic Beanstalk

Elastic Beanstalk offers a rapid and straightforward method for getting your applications into the AWS Cloud. It is a Platform as a Service (PaaS) offering. The infrastructure and platform are quickly built for you in the cloud. This permits the quick deployment of your applications. Elastic Beanstalk also reduces the ongoing management complexity of your deployment. Importantly, you maintain control of the platform. For example, you have complete control if you want to scale your applications more aggressively. Another significant aspect of this service is that it supports a wide variety of languages and platforms, such as Go, Java SE, PHP, Python, and Node.js, to name a few. Application upgrades are simple, as you deploy them to Elastic Beanstalk as needed.

While it is easy to implement, it is also robust. You supply the application code, and AWS provides components such as the following:

- Application services
- HTTP services
- Required operating systems
- Required language interpreters
- The physical hosts required

EC2

Amazon Elastic Compute Cloud (EC2) is a web service that provides secure and resizable compute resources in the AWS Cloud. The EC2 service allows you to provision and configure capacity with minimal effort, allowing you to easily control your computing resources.

EC2 reduces the time required to obtain and boot new servers (EC2 instances) to just minutes. This efficiency allows you to scale capacity vertically (up and down, making your server resources bigger or smaller) and horizontally (out and in, adding more capacity in more instances), as your computing requirements change. As you might recall from previous Modules, this property is called *elasticity*.

The many benefits of EC2 in AWS include the following:

- EC2 allows for controlled expenditures as your business expands; you pay only for the resources you use as your business grows.
- EC2 provides the tools to build failure-resilient applications that isolate themselves from common failure scenarios.

- EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously.

- You have complete control of your EC2 instances. You have root access to each one, and you can interact with them as you would any traditional virtual machine.

- You can stop your EC2 instance while retaining the data on your boot partition and then subsequently restart the same instance using web service APIs. Instances can be stopped and started remotely using web service APIs.

- You can choose among multiple instance types, operating systems, and software packages. Instance types inside AWS permit the choice of emphasis on CPU, RAM, and/or networking resources.

- EC2 integrates with most AWS services, such as Simple Storage Service (S3), Relational Database Service (RDS), and Virtual Private Cloud (VPC). This tight integration allows you to use EC2 for a wide variety of compute scenarios.

- EC2 offers a reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers. AWS offers as much as 99.95 percent availability for each region.

- Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources:

  - Your compute instances are located in a VPC with an IP address range that you specify.

  - You decide which instances are exposed to the Internet and which remain private.

  - Security groups and network access control lists (ACLs) allow you to control inbound and outbound access to and from your network interfaces.

  - You can connect your existing IT infrastructure to resources in your VPC using industry-standard encrypted IPsec virtual private network (VPN) connections, or you can take advantage of a private AWS Direct Connect option.

- You can provision your Amazon EC2 resources as dedicated instances. Dedicated instances are Amazon EC2 instances that run on

hardware dedicated to a single customer for additional isolation. Alternatively, you can provision your Amazon EC2 resources on dedicated hosts, physical servers with EC2 instance capacity entirely dedicated to your use. Dedicated hosts can help you address compliance requirements and reduce costs by allowing you to use your existing server-bound software licenses.

Several pricing models exist, including the following:

- **On-demand instances**: With this model, you pay for compute capacity by the hour (or even by the second with some AMIs) with no long-term commitments. You can increase or decrease your computing capacity depending on your application's demands and pay the specified hourly rate only for the instances you use. Using on-demand instances frees you from planning, purchasing, and maintaining hardware costs and complexities. As mentioned in the first section, this model also transforms commonly substantial fixed costs into much smaller variable costs.

- **Reserved instances**: This model provides a significant discount (up to 75 per cent) compared to on-demand instance pricing. You can change families, operating system types, and tenancies while benefitting from reserved instance pricing when you use convertible reserved instances. Figure 3-4 shows the beginning steps of configured reserved instances in AWS.
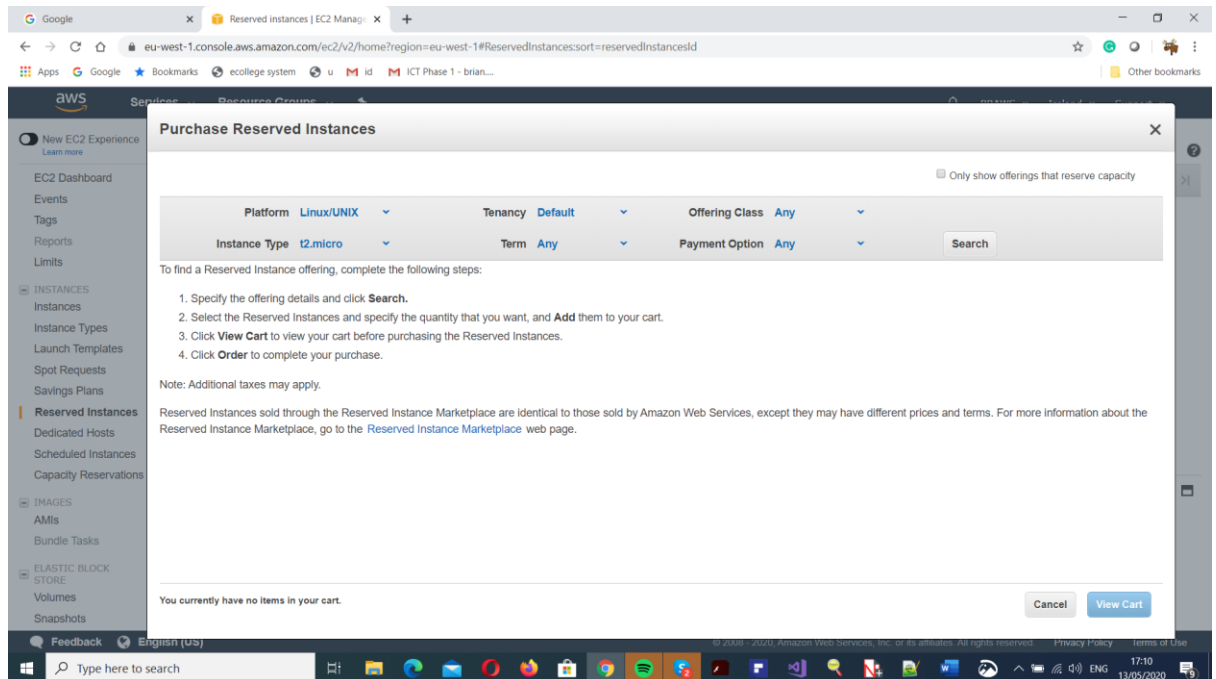
Figure 3-4 Purchasing Reserved Instances in AWS

- **Spot instances**: These instances allow you to bid on spare EC2 computing capacity. Because spot instances are often available at a discount compared to on-demand pricing, you can significantly reduce the cost (up to 90 per cent) of running your applications.

Savings Plans

AWS Savings Plans are flexible pricing models that offer significant cost savings on Amazon Web Services (AWS) usage in exchange for a commitment to a consistent amount of usage (measured in USD per hour) over a one- or three-year term. Depending on the plan and commitment level, these plans allow customers to save up to 72% compared to On-Demand pricing.

Types of AWS Savings Plans:

1. **Compute Savings Plans:**
   - Offers the most flexibility and applies to a wide range of AWS compute services, including Amazon EC2, AWS Lambda, and AWS Fargate.
   - Applicable across any region, instance family, operating system, and tenancy.

      o Ideal for customers who want the flexibility to change instance types, sizes, regions, or even move workloads between EC2, Lambda, and Fargate without losing savings.

2. **EC2 Instance Savings Plans:**

      o Offer lower discounts than Compute Savings Plans but provide additional savings for customers with predictable workloads.

      o Commit to using a specific EC2 instance family within a chosen region.

      o The plan allows changes within the same instance family, such as switching between sizes or operating systems, but is less flexible compared to Compute Savings Plans.

Key Features:

- **Commitment Periods:** Available for one- or three-year terms, with three-year plans offering higher discounts.

- **Payment Options:** Flexible payment options include No Upfront, Partial Upfront, and All Upfront, affecting the discount rate.

- **Coverage:** Automatically applies savings to eligible usage, prioritizing the highest discount available under the plan.

Benefits of AWS Savings Plans:

- **Cost Savings:** Reduces costs significantly on compute usage.

- **Flexibility:** This option allows changes in how resources are used (depending on the plan type), providing freedom to optimize resources based on evolving needs.

- **Predictability:** Helps in planning and forecasting cloud costs more effectively.

AWS Savings Plans are ideal for businesses looking to reduce their cloud computing expenses while maintaining flexibility in their AWS usage patterns. They are especially beneficial for organisations with steady-state or predictable workloads that can commit to consistent compute usage over time.

INTRODUCTION TO EBS

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability.

EBS volumes offer the consistent and low-latency performance needed to run your workloads. With Amazon EBS, you can scale your usage up or down within minutes—all while paying a low price for only what you provision.

Features of EBS include the following:

- **High-performance volumes**: Choose between solid-state disk (SSD)-backed or hard disk drive (HDD)-backed volumes that can deliver the performance you need for your most demanding applications.

- **Availability**: Each Amazon EBS volume is designed for 99.999 percent availability and automatically replicates within its Availability Zone to protect your applications from component failure.

- **Encryption**: Amazon EBS encryption provides seamless support for data at rest and data in transit between EC2 instances and EBS volumes.

- **Access management**: Amazon's flexible access control policies allow you to specify who can access which EBS volumes, ensuring secure access to your data.

- **Snapshots**: You can protect your data by creating point-in-time snapshots of EBS volumes, which are backed up to Amazon S3 for long-term durability.

INTRODUCTION TO S3

Amazon Simple Storage Service (Amazon S3) is object storage with a simple web service interface to store and retrieve any amount of data from anywhere on the web. It is designed to deliver 99.999999999 percent durability. Figure 3-5 shows two storage buckets in S3.
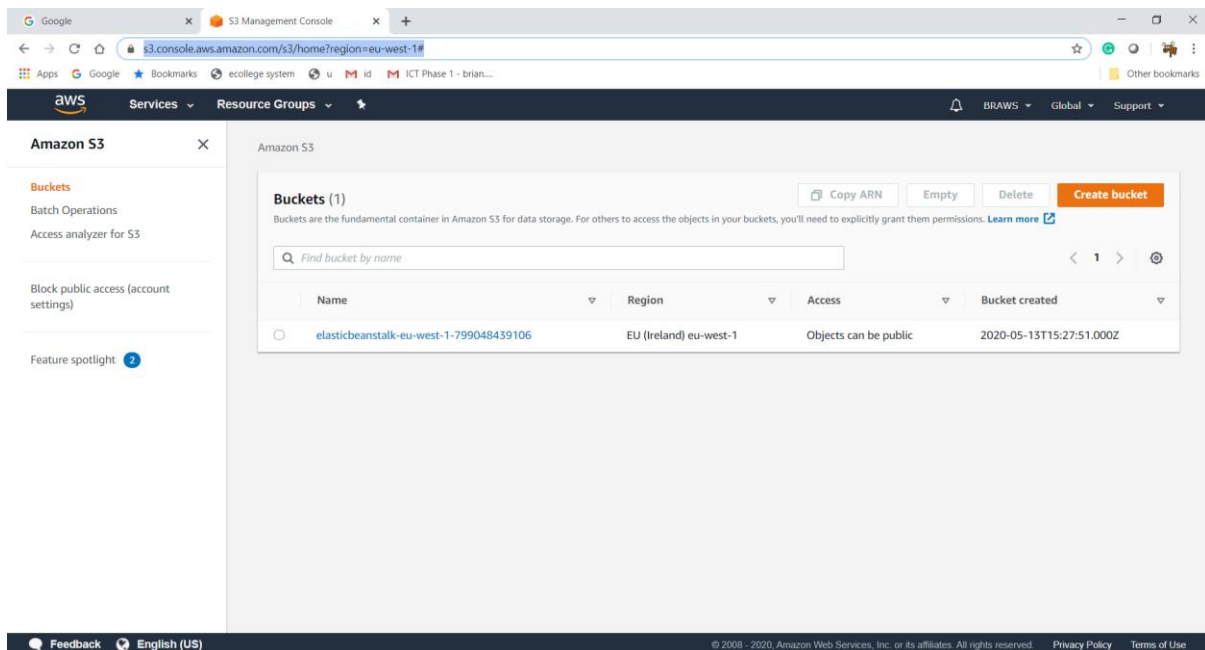

Figure 3-5 AWS S3

You can use Amazon S3 for a vast number of purposes, such as the following:

- Primary storage for cloud-native applications
- A bulk repository, or "data lake," for analytics
- A target for backup and recovery and disaster recovery
- For use with serverless computing

You can move large volumes of data into or out of Amazon S3 with Amazon's cloud data migration options. You can store data in S3 and then automatically tier the data into lower-cost, longer-term cloud storage classes like S3 Standard–Infrequent Access and Glacier for archiving. You could even utilize a new storage class that reduces high availability (One Zone Infrequent Access) when you do not require it and want to save on storage costs.

S3 offers many advantages, including the following:

- **Simple**: S3 is easy to use with a web-based management console and mobile app. Amazon S3 also provides full REST APIs and SDKs for easy integration with third-party technologies. A command-line interface (CLI) is also extremely popular for working with S3.

- **Durable**: S3 provides a durable infrastructure to store essential data. Amazon designed S3 for durability of 99.999999999 percent of objects. S3 redundantly stores your data across multiple facilities and multiple devices in each facility.

- **Scalable**: With S3, you can store as much data as you want and access it when needed. While there is a 5TB limit on the size of an individual object, there is no limit to the number of objects you can store!

- **Secure**: S3 supports data transfer over SSL and automatic encryption of your data following the upload. If you want, you can control client-side or server-side encryption. You can use Amazon-generated or customer-generated keys and have full key management capabilities/options. You can also configure bucket policies to manage object permissions and control access to your data using Identity and Access Management (IAM).

- **Available**: S3 Standard is designed for up to 99.99 percent availability of objects over a given year and is backed by the Amazon S3 service level agreement, ensuring that you can rely on it when needed. You can also choose an AWS region to optimize for latency, minimize costs, or address regulatory requirements.

- **Low cost**: S3 allows you to store large amounts of data at a small cost. Using lifecycle policies, you can configure the automatic migration of your data to different storage tiers within AWS.

- **Simple data transfer**: Amazon provides multiple options for cloud data migration and makes it simple and cost-effective for you to move large volumes of data into or out of S3. You can choose from network-optimized, physical disk-based, and third-party connector methods for import to or export from S3.

- **Integrated**: S3 is deeply integrated with other AWS services to make it easier to build solutions that use a range of AWS services. Integrations include the following:

  - CloudFront

  - CloudWatch

  - Kinesis

  - Relational Database Service (RDS)

  - Glacier

  - Elastic Block Store (EBS)

  - DynamoDB

  - Redshift

  - Route 53

  - Elastic MapReduce (EMR)

  - Virtual Private Cloud (VPC)

  - Key Management Service (KMS)

  - Lambda

- **Easy to manage**: S3 storage management features allow you to take a data-driven approach to storage optimisation, data security, and management efficiency. These enterprise-class capabilities give you data about your data so that you can manage your storage based on that personalized metadata.

While technically part of the S3 service, S3 Glacier is a secure, durable, and extremely low-cost storage service for data archiving and long-term backup. With Glacier, you can do the following:

- Reliably store large or small amounts of data for as little as $0.004 per gigabyte per month.

- Save money compared to on-premises storage options.

- Keep costs low yet suitable for varying retrieval needs.

- Choose from three options for access to archives—from a few minutes to several hours.

INTRODUCTION TO AWS DATABASE SOLUTIONS

Many types of databases are available today. The great news is that AWS supports these varieties. In fact, AWS permits several different approaches to their implementation. This section gives you an overview of these exciting technologies.

Aurora

Amazon Aurora is a MySQL- and PostgreSQL-compatible relational database engine. It offers many benefits, including the following:

- **High performance**: Aurora can provide up to five times the throughput of standard MySQL or twice the throughput of standard PostgreSQL running on the same hardware.

- **Highly secure**: Aurora provides multiple levels of security for your database. These include network isolation using a VPC, encryption of data at rest using keys you create and control through Key Management Service (KMS), and encryption of data in transit using SSL.

- **MySQL and PostgreSQL compatible**: The Aurora database engine is fully compatible with MySQL 5.6 and MySQL 5.7 using the InnoDB storage engine.

- **Highly scalable**: You can scale your Aurora database from an instance with two vCPUs and 4 GB of memory up to an example with 32 vCPUs and 244 GB.

- **High availability and durability**: Aurora is designed to offer higher than 99.99 per cent availability. It is also amazing regarding durability, ensuring six synchronous copies of your data running across three AZs. This becomes a huge advantage over standard RDS implementations.

- **Fully managed**: Aurora is a fully managed database service. Amazon handles hardware provisioning, software patching, setup, configuration, monitoring, and backups.

Relational Database Service

Relational Database Service (RDS) makes it easy to set up, operate, and scale a relational database in the cloud. RDS provides six database engines: Aurora, PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server.

Benefits of RDS include the following:

- **Fast and easy to administer**: You can use the AWS Management Console, the AWS RDS command-line interface, or simple API calls to access the capabilities of a production-ready relational database in minutes.

- **Highly scalable**: You can scale your database's compute and storage resources with only a few mouse clicks or an API call, often without downtime.

- **Available and durable**: RDS runs on the same highly reliable infrastructure other Amazon Web Services used. RDS synchronously replicates the data to a standby instance in a different Availability Zone (AZ) when provisioning a Multi-AZ DB instance.

- **Secure**: RDS makes controlling network access to your database easy. RDS also lets you run your database instances in a VPC, which enables you to isolate your database instances and connect to your existing IT infrastructure through an industry-standard encrypted IPsec VPN. Many RDS engine types offer encryption at rest and encryption in transit. You can also take advantage of Direct Connect.

- **Inexpensive**: You pay low rates and only for the resources you consume.

DynamoDB

Amazon DynamoDB is a fast and flexible NoSQL database service for all applications that need consistent, single-digit millisecond latency at any scale. It is a great fit for mobile, web, gaming, ad-tech, Internet of Things (IoT), and many other applications.

Benefits of DynamoDB include the following:

- **Fast, consistent performance**: DynamoDB delivers consistent, fast performance at any scale for all applications.

- **Highly scalable**: When creating a table, you specify the required request capacity. If your throughput requirements change, you update your table's request capacity using the AWS Management Console or the DynamoDB APIs. DynamoDB manages all the scaling behind the scenes, and you are still able to achieve your previous throughput levels while scaling is underway. Instant scaling and auto-scaling capabilities now exist that even assist you if you are unsure of the initial capacity you require.

- **Fully managed**: DynamoDB is a fully managed cloud NoSQL database service. You create a database table, optionally set your throughput or allow auto-scaling, and let the service handle the rest.

- **Event-driven programming**: DynamoDB integrates with Lambda to provide triggers that enable you to architect applications that automatically react to data changes.

- **Fine-grained access control**: DynamoDB integrates with IAM for fine-grained access control.

- **Flexible**: DynamoDB supports both document and key-value data structures, giving you the flexibility to design the optimal data architecture for your application.

ElastiCache

ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud. The service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases.

ElastiCache supports two open-source in-memory caching engines:

- **Redis**: A fast, open-source in-memory data store and cache. ElastiCache for Redis is a Redis-compatible in-memory service that delivers the ease of use and power of Redis along with the availability, reliability, and performance suitable for the most demanding applications.

- **Memcached**: A widely adopted memory object caching system. ElastiCache is protocol-compliant with Memcached, so tools that you use today with existing Memcached environments work seamlessly with the service.

Redshift

Redshift is a fast, fully managed, petabyte-scale data warehouse that makes it simple and cost-effective to analyze all your data using your existing business intelligence tools. Features include the following:

- High query performance on data sets ranging in size from a hundred gigabytes to a petabyte or more.

- Using columnar storage, data compression, and zone maps to reduce the amount of I/O needed to perform queries.

- Redshift has massively parallel processing (MPP) data warehouse architecture, parallelizing and distributing SQL operations to take advantage of all available resources. The underlying hardware is designed for high-performance data processing, using locally attached storage to maximize throughput between the CPUs and drives, and a 10GigE mesh network to maximize throughput between nodes.

Database Migration Service

AWS Database Migration Service helps you migrate databases to or from AWS easily and securely. Features include the following:

- The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database.

- It migrates your data to and from most widely used commercial and open-source databases. The service supports homogeneous migrations such as Oracle to Oracle, as well as various migrations between different database platforms, such as Oracle to Amazon Aurora or Microsoft SQL Server to MySQL.

- It also allows you to stream data to Redshift from any of the supported sources, including Aurora, PostgreSQL, MySQL, MariaDB, Oracle, SAP ASE, Teradata, and SQL Server, enabling consolidation and straightforward analysis of data in the petabyte-scale data warehouse.

- You can use AWS Database Migration Service for continuous data replication with high availability.

EXAM PREPARATION TASKS

- **REVIEW ALL TOPICS**

- **DEFINE ALL KEY TERMS AND CHECK ANSWERS IN THE GLOSSARY.**

- **DO THE QUIZ – REPEAT UNTIL YOU PASS IT (100% PASSMARK).**

DEFINE KEY TERMS

Define the following key terms from this module and check your answers in the Glossary:

Networking and content delivery

regions

Availability Zones

Edge Locations

security groups

on-demand instances

reserved instances

spot instances