Module 2. Advantages of the AWS Cloud

This module covers the following subjects:

- **Cloud Advantages**: This section includes just some of the many advantages of cloud technologies in general. This section is not specific to Amazon Web Services (AWS), but AWS does provide all the benefits covered. The AWS-specific advantages are elaborated on in the next section.
- **AWS Cloud Advantages**: This part of the module focuses on the specific advantages AWS brings and references many of the specific technologies that make them a reality.

Why are cloud engineers in such high demand? In fact, why is the cloud so popular? This module ensures you understand the many advantages we realize with cloud technology adoption, and then the module gets very specific to AWS. This module examines AWS-specific benefits and some of the many technologies that make them a reality.

CLOUD ADVANTAGES

It is no major surprise that various public cloud vendors (led by AWS) are experiencing more success than ever before. The list of advantages continues to grow! Here are just some:

- **OpEx replaces CapEx**: Using public cloud technologies enables startups and existing organisations to provide new features and services with minimal capital expenditures (CapEx). Instead, public cloud expenses revolve around monthly operating expenses (OpEx). For most organisations, OpEx represents significant advantages compared to substantial CapEx investments.
- **Lack of contractual commitments**: Many public cloud vendors charge hourly (if not less). For most services, there is no long-term commitment to an organisation. You can roll out new projects or initiatives and, if needed, roll back with no long-term contractual obligations. This lack of contractual commitment helps increase the agility of IT operations and lowers financial risks associated with innovative technologies.
- **Reduction of required negotiations**: Establishing new accounts with public cloud vendors is simple, and prices for the major public cloud vendors continuously reduce. This reduction in rates and the ease of account setup minimises the need for cost negotiations, as might have existed early in service-provider interactions.
- **Reduced procurement delays**: Additional resources can be set up with most cloud implementations within seconds.
- **"Pay as you go" model**: If more resources are needed to support a growing cloud presence, you can get these resources on-demand and pay for them only when needed. Conversely, if fewer resources are required, you can run less and pay for only what you need. Figure 2-1 shows an example of a cost dashboard in AWS. Notice how each service incurs a monthly cost, and the charges are broken down, like a utility bill.

https://aws.amazon.com/pricing

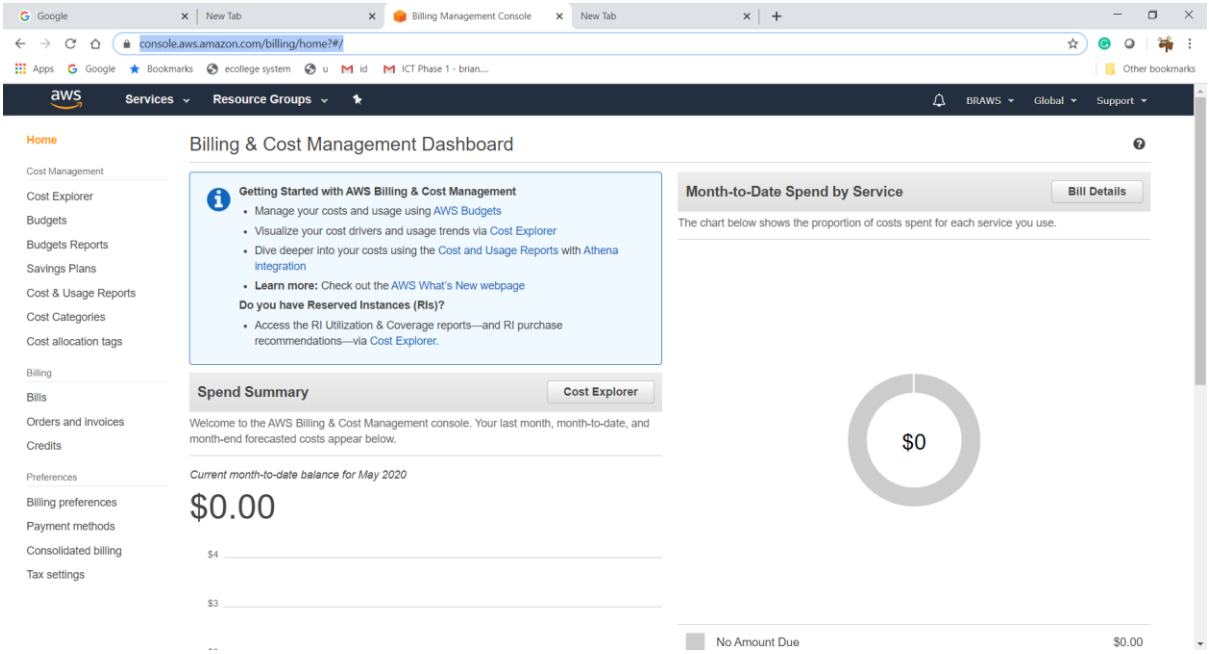

Figure 2-1 Costs in the Cloud Are "Pay as You Go"

- **High levels of security possible**: Because you can focus on the security of your resources and the cloud provider can concentrate on its security responsibilities (such as physical security and hypervisor security), the resulting infrastructure can meet stringent levels of security. This security model is called the *Shared Responsibility model*.
- **Flexibility**: With the features of public cloud platforms like AWS, you can quickly scale the cloud-based infrastructure up and down as well as out and in. This advantage is called *elasticity*. Auto-scaling functionality inside AWS allows the dynamic creation and destruction of resources based on client demand. Such scaling can occur with little to no administrator interaction. When discussing scaling the resources of service, we are scaling those resources horizontally (out and in with elasticity). In contrast, the service made up of those resources is being scaled up and down (vertically because the single service is getting bigger or smaller). A single service scales up and down, out and in, depending on the context.
- **A massive global infrastructure**: Most public cloud vendors now offer resources located all over the globe. This global dispersion of resources serves large multinational organisations very well since resources needed for certain parts of the world can be stored and optimised for access in those regions. Also, companies with clients all over the world can enjoy similar access advantages when servicing their clients' needs.

- **SaaS, PaaS, and IaaS offerings**: Cloud technologies have become so advanced that organizations can choose to give applications to clients, development environments, or even entire IT infrastructures using cloud technologies. Since cloud can offer about any component of IT these days, many refer to cloud as an *Everything as a Service (XaaS)* opportunity.
- **Emphasis on API support:** Increasingly, cloud vendors are taking an application programming interface (API) first approach. Making the same configuration possible with REST APIs (typically used) that would be possible with a software development kit (SDK), command-line interface (CLI), or graphical user interface (GUI). The API first approach means no interface (CLI or GUI) changes are made until API calls are made first. Thus, there is nothing that cannot be automated! Figure 2-2 shows how the vast amount of AWS service can be accessed from a simple GUI called the AWS Management Console.

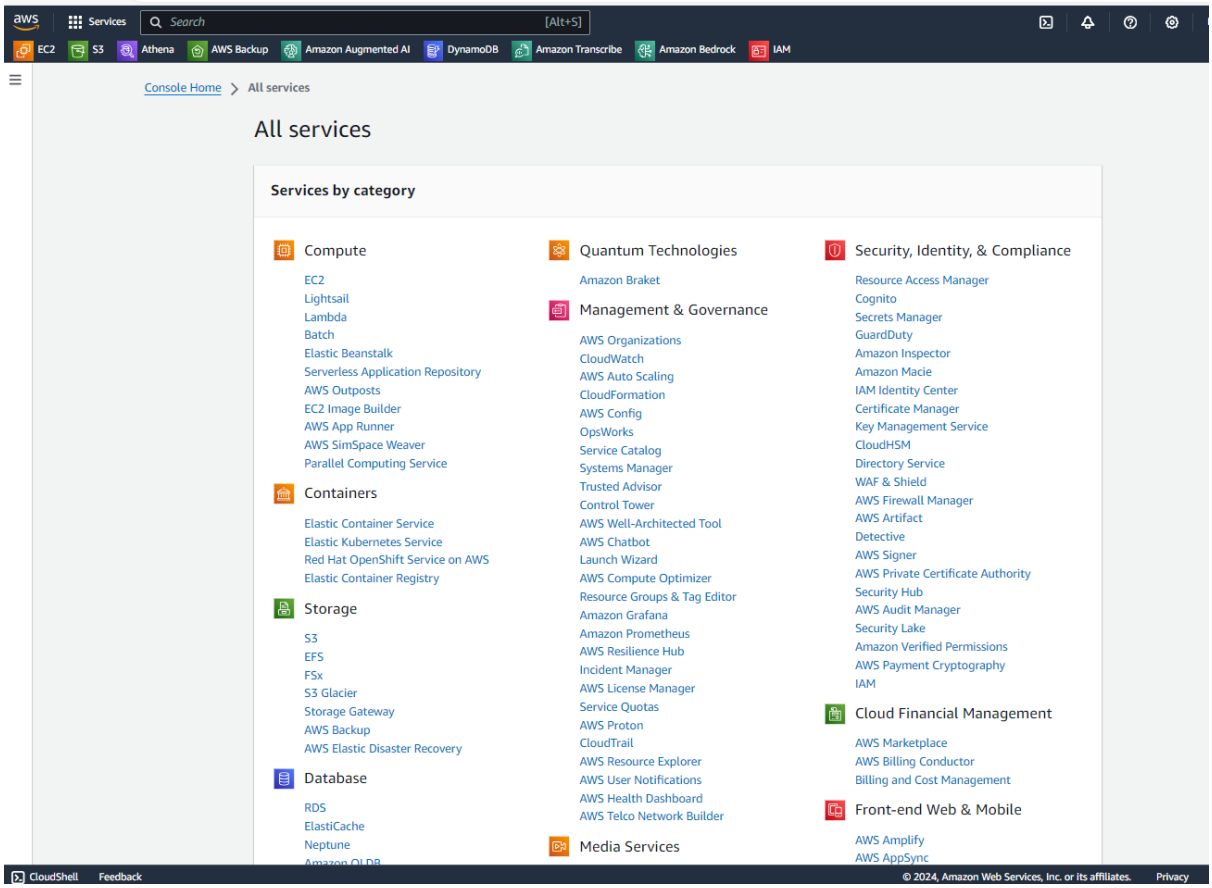

Figure 2-2 Managing Vast Cloud Resources in a Simple GUI

AWS CLOUD ADVANTAGES

The “pay as you go” model followed by AWS is a massive revolution in budgeting and affording the latest technological innovations. This simple model allows engineers to focus on innovation and new business solutions instead of worrying about infrastructure and other resource shortfalls.

Before the broad adoption of AWS solutions, engineers would waste time and money over-provisioning resources to attempt to provide reliability and performance, even under peak load conditions. AWS allows cloud engineers to “spin up” new resources in seconds and view these resources as temporary and disposable. It is not uncommon for AWS customers to deploy massive amounts of infrastructure for a short period to test new technologies without paying enormous upfront costs. Amazon has published many examples of the benefits of AWS, including examples of usage. One was about a large US university needing to conduct some artificial intelligence testing. The university turned to AWS for help and spun up millions of CPUs across hundreds of thousands of Elastic Cloud Compute (EC2) virtual machines (VMs). Sure enough, they only required this horsepower for a weekend. All of the resources were “terminated” (the AWS term for deleting an EC2 VM) once the testing was complete. Amazon did not tell us what the bill was for this weekend of work, but you can rest assured that the cost of all of those resources if purchased as a capital expense, would have been massive. And what would the company do with the resources once the testing was complete?

Amazon Web Services truly enables an organization to be flexible in efficiently and cost-effectively provisioning resources because there are fewer constraints.

The most significant advantage companies see in moving to AWS is the ability to increase their agility. Three main aspects of AWS accomplish this:

- **Speed**: The AWS Global Infrastructure spans the entire globe. This global reach ensures you can place resources geographically close to those that need to consume them. The efficient location of resources helps reduce network latency and fosters excellent performance. As described earlier, the AWS cloud can allocate massive amounts of resources within seconds.
- **Experimentation**: With AWS, you can implement your IT operations as code. In addition to running with administrative ease and error-

free, this fosters the ease of experimentation and testing. Templates are available thanks to services like AWS CloudFormation that permit you to instantly create complex networks and IT resources for testing and experimenting. Once the experimentation is complete, you can dispose of the resources and save money. Figure 2-3 shows an example of CloudFormation.

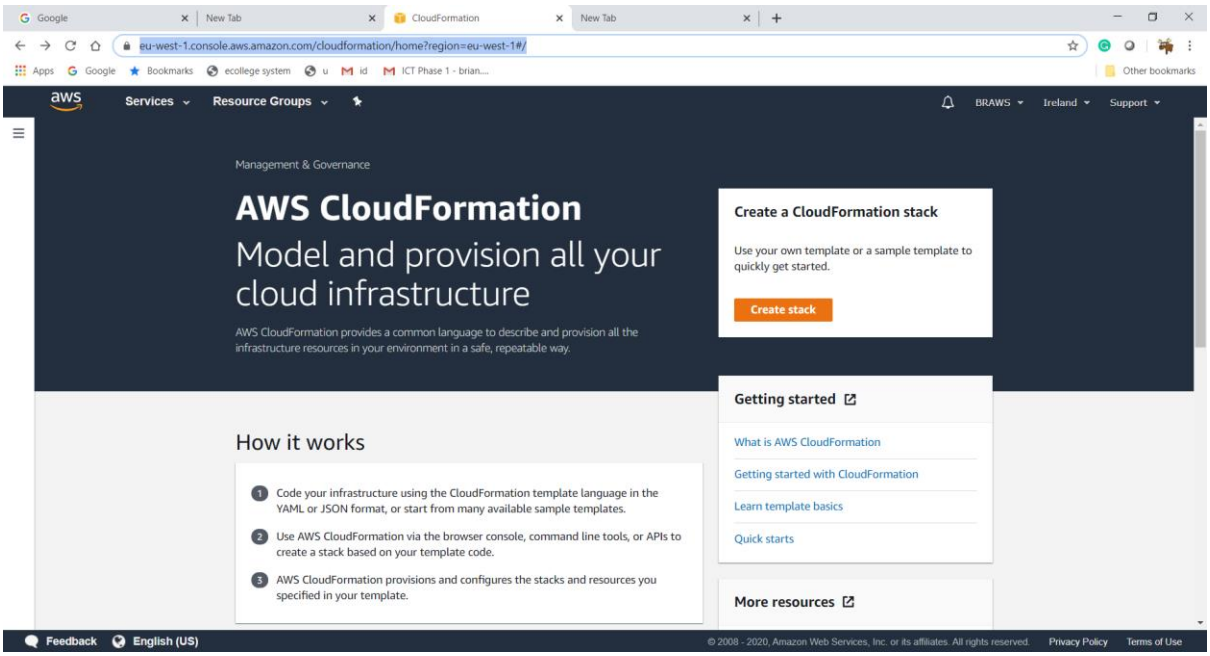

Figure 2-3 CloudFormation in AWS

- **Culture of innovation**: These enablers of agility previously listed also help foster a culture of innovation in your enterprise. An increasing number of companies participate in AWS functions because of this. It is not just about saving money. They love the ability to experiment with new technologies with a low risk to their organization. Innovations are thought of as very possible thanks to AWS.

Perhaps at the very core of AWS is what makes it all possible—the AWS Global Infrastructure. This vast network is what makes the incredible elasticity, scalability, and reliability possible across a vast number of IT services. The AWS Global Infrastructure is made up of many components, but here are the two you should master now:

- **Regions**: Regions are physical locations in geographically dispersed parts of the globe. For example, there are US West regions and US East regions, just as there are regions in Europe and Asia. Inside each region are multiple Availability Zones (AZs). These AZs house the

data centres that contain massive amounts of physical network resources and data.

- **Availability Zones**: AZs consist of one or more data centres. These data centres are filled with redundancy at every level, from network connections to physical devices. They are also physically distant from one another, which helps mitigate the effects of localised disasters. They feature high availability (HA) and fault tolerance (FT).
- **Local Zones** AWS Local Zones are a revolutionary step in cloud computing, bringing AWS services closer to users and reducing latency for various applications. This infrastructure deployment allows for running applications requiring single-digit millisecond latency, such as real-time gaming, live streaming, and AR/VR experiences. It also facilitates hybrid cloud migrations and helps meet data residency requirements in various sectors, including healthcare and financial services. With AWS Local Zones, businesses can deploy applications closer to end users, ensuring a seamless and responsive user experience.

- **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. It is crucial for latency-sensitive applications like game streaming, real-time analytics, and machine learning inference at the edge. This innovative service transforms how applications are delivered, making it possible to offer new experiences previously impossible due to latency constraints.

Figure 2-4 demonstrates how easy it is to drop a menu at the top of the management console to select a new region of the world to initialise localised resources for that region.

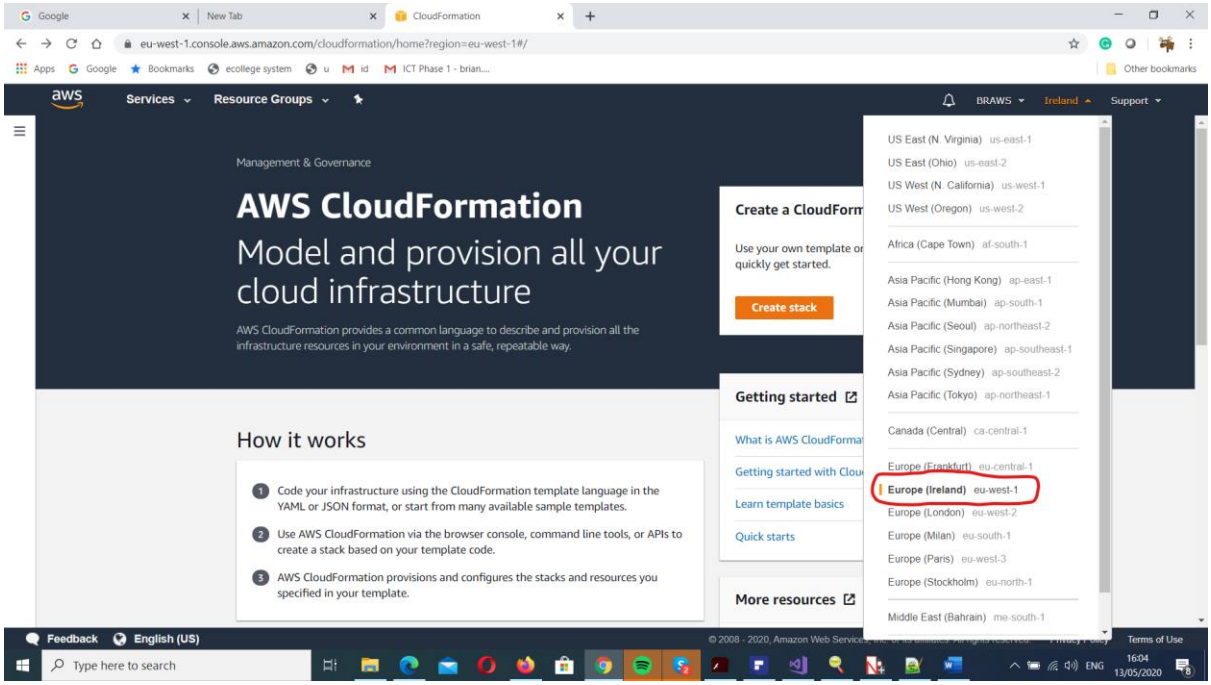

Figure 2-4 Selecting a Region in AWS

Another massive advantage that AWS brings is using easy-to-use tools that foster overall cloud benefits such as elasticity. There are two tools, in particular, you should be aware of:

- **Auto Scaling**: Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost. Thanks to this powerful tool, you can enable application scaling for multiple resources across multiple services in minutes. The service provides a simple, powerful user interface that lets you build scaling plans for resources like Amazon EC2 instances.
- **Elastic Load Balancing**: Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones. Elastic Load Balancing offers three types of load balancers that all feature the high availability, automatic scaling, and robust security necessary to make your applications fault-tolerant. These three types are the Application Load Balancer, the Network Load Balancer, and the Gateway Load Balancer. Figure 2-5 shows the configuration of the Network Load Balancer in AWS.

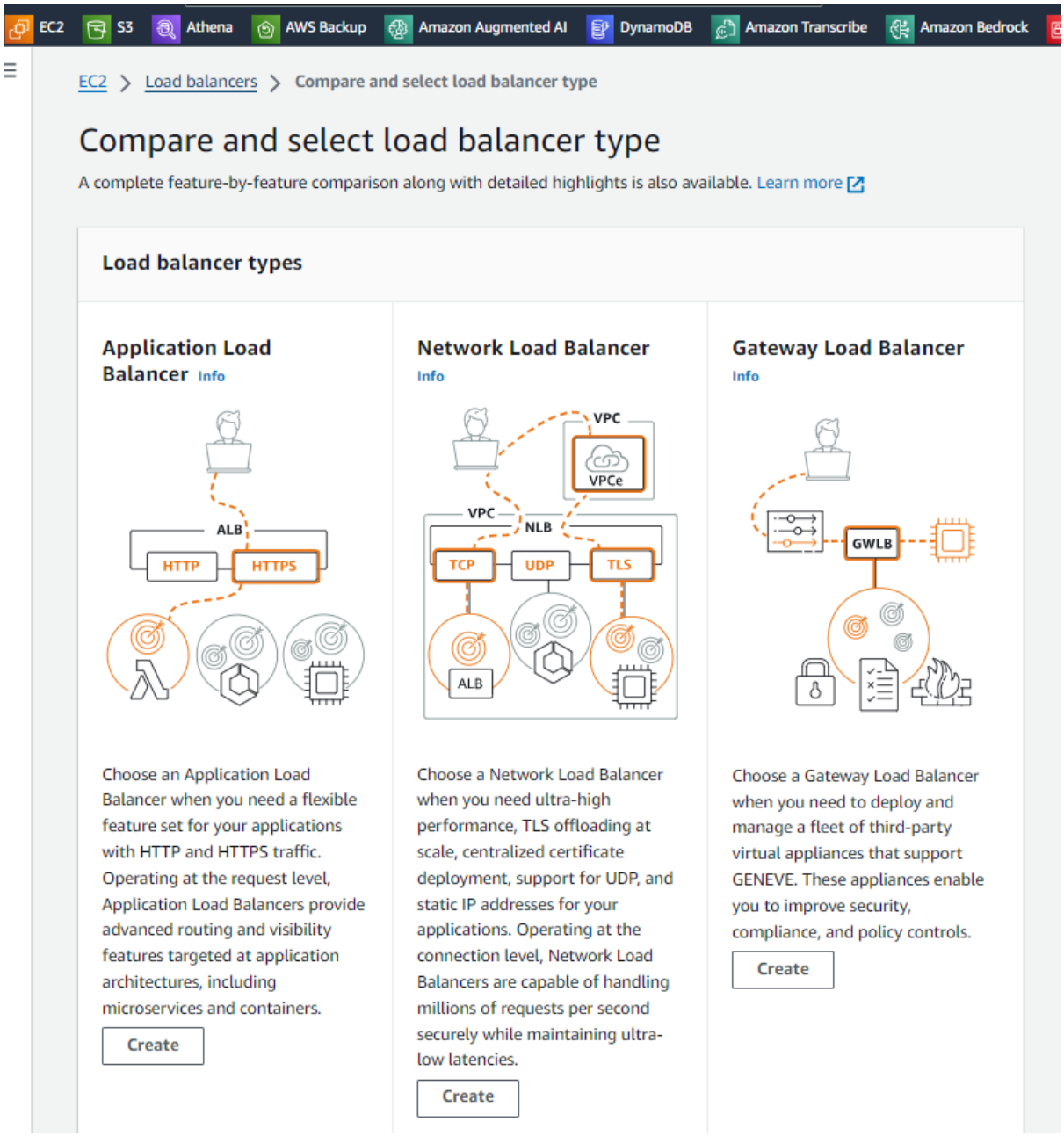

Figure 2-5a Elastic Load Balancing in AWS

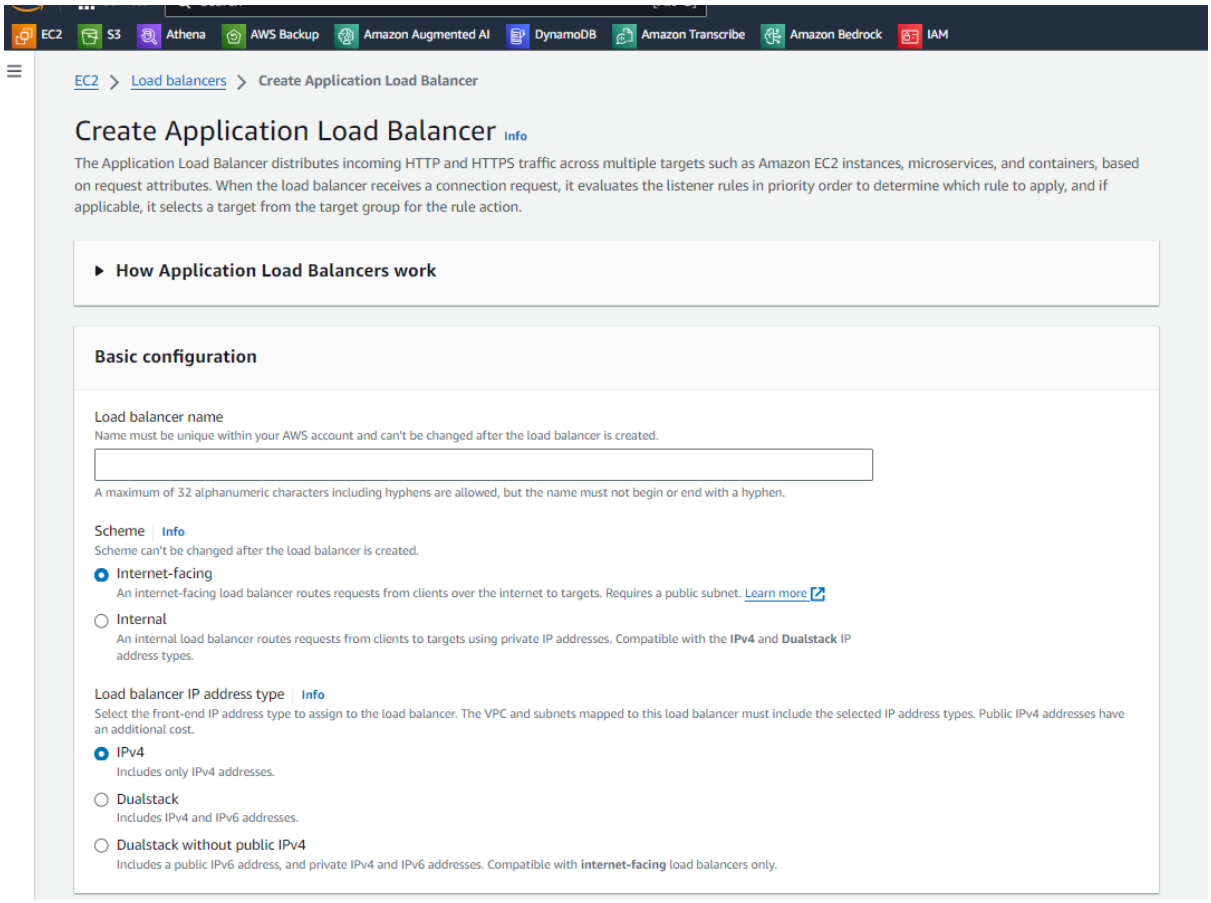

Figure 2-5b Elastic Load Balancing in AWS

Another advantage of AWS is in security and compliance. Thanks to the AWS Global Infrastructure, you maintain control of where your resources are stored geographically, making it easier for you to comply with regional governance responsibilities. AWS permits you to achieve the strictest security requirements, and you can rest at ease, knowing that AWS uses the most cutting-edge security-heavy data centres in the world.

You also enjoy vast amounts of reliability for your AWS projects. This reliability helps you deliver high performance, permits ease with failure recovery and permits you to acquire new resources, as needed, dynamically and with high speed.

EXAM PREPARATION TASKS

- **REVIEW ALL TOPICS**
- **DEFINE ALL KEY TERMS AND CHECK ANSWERS IN THE GLOSSARY.**
- **DO THE QUIZ – REPEAT UNTIL YOU PASS IT.**

DEFINE KEY TERMS

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

CapEx

OpEx

API

agility

AWS Global Infrastructure

regions

Availability Zones

Auto Scaling

Elastic Load Balancing