PART I

Introduction to Amazon Web Services

CHAPTER 1: Introduction to Cloud Computing and Amazon Web Services

CHAPTER 2: Regions, Availability Zones, and Edge Locations

CHAPTER 3: AWS Identity and Access Management

CHAPTER 4: Amazon EC2

CHAPTER 5: Amazon S3

CHAPTER 6: Amazon DynamoDB

CHAPTER 7: AWS Lambda

Online Bonus Chapters:

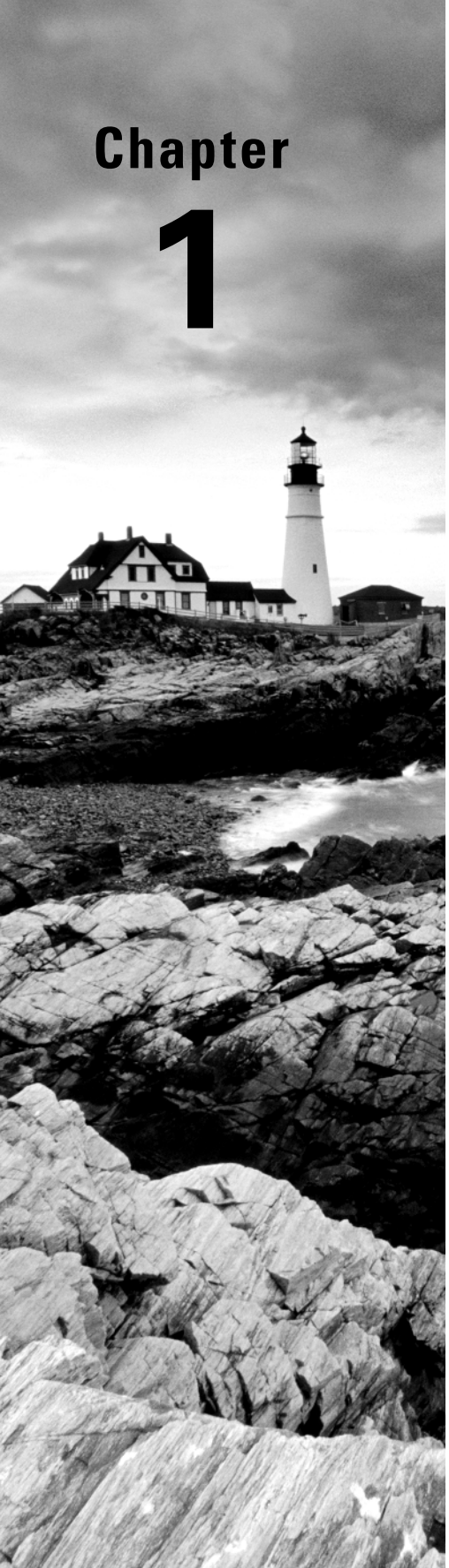
BONUS CHAPTER 1: Amazon Simple Notification Service

BONUS CHAPTER 2: Amazon Mobile Analytics

BONUS CHAPTER 3: AWS Device Farm

BONUS CHAPTER 4: Installing WordPress on Amazon EC2

COPYRIGHTED MATERIAL



Chapter 1

Introduction to Cloud Computing and Amazon Web Services

WHAT'S IN THIS CHAPTER

- ✓ Introduction to the basics of cloud computing.
- ✓ Introduction to the AWS Ecosystem and key services covered in this book.
- ✓ Learn to sign up for an account under the AWS free tier.

Hello, and welcome to the exciting world of cloud computing with Amazon Web Services (AWS). In this chapter, you learn about what cloud computing is, read about common models of abstraction used when discussing cloud based services, and discover a high-level overview of Amazon's offerings in the cloud computing space. The chapter wraps up by walking you through signing up for an AWS account.

Broadly speaking, the word *cloud*, when used in the context of cloud computing, implies some type of shared computing resource accessed over the Internet. Cloud computing as we know it today was born in 2006 when Amazon launched its Elastic Compute Cloud (EC2) service, which allowed users to create a virtual server in the cloud and deploy their software onto this server within minutes. Amazon EC2 is one of the core services offered by AWS today. Soon after in 2008, Microsoft launched its Azure service. This was followed by competing offers from other players, including Rackspace, Google, Oracle, and Apple.

If you have never heard of AWS, you may be surprised to learn that Amazon, a company that started out as an online book retailer, is the leading provider of cloud computing services.

Before diving in to the anatomy of AWS, the following sections introduce you more fully to cloud computing and its service and deployment models.

What Is Cloud Computing?

Cloud computing is defined by the U.S National Institute of Standards and Technology (NIST) as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Peter Mell and Timothy Grance, "The NIST Definition of Cloud Computing," NIST Special Publication 800-145. September 2011. (`http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf`)

NIST defines five essential characteristics in this model. Each of these is briefly examined next:

- **Broad network access:** A consumer should be able to access services from anywhere.
- **Resource pooling:** A provider's computing resources are pooled to support multiple customers.

- **On-demand self-service:** A consumer should be able to provision computing resources (such as virtual servers) as needed, with minimal human interaction.
- **Measured service:** A consumer should be able to use computing resources on a pay-as-you-use basis.
- **Elasticity:** A consumer should be able to provision additional resources automatically and on demand. To ensure this, the provider pools computing resources to provide horizontal scalability to the consumer.

Cloud-computing solutions provide two major advantages to businesses:

- **Costs:** The cloud-computing paradigm is based on sharing and optimal utilization of hardware resources. A business need only pay for the time during which it utilizes a resource. When a resource is not needed, the business can relinquish it and make it available for someone else to use. This reduces both the upfront hardware investment cost for a business and also ongoing maintenance costs. The cloud service provider, not the consumer, handles the maintenance of the underlying hardware.
- **Availability:** The time to provision a ready-to-use resource in the cloud is significantly lower than having to set up a similar resource in-house. For instance, a business could provision a virtual server with a cloud provider within seconds, whereas the actual process of procuring new server hardware and software usually takes a few months in most medium- to large-sized organizations.

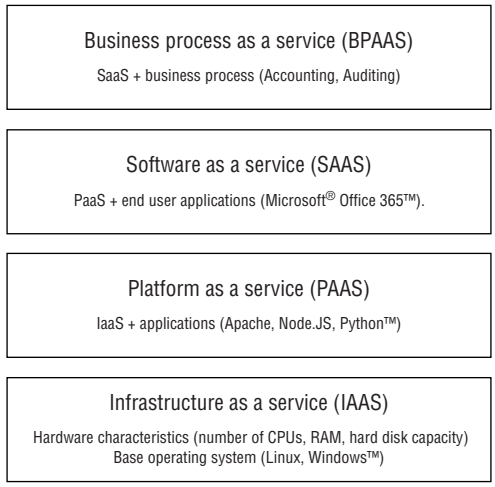
Cloud Service Models

Cloud computing is built on virtualization technology. Fundamentally, there are two types of virtualization:

- **Application virtualization:** A single machine hosts one or more applications that are delivered to one or more users over the Internet.
- **Hardware virtualization:** Also known as server virtualization, in this model a single physical machine hosts multiple virtual machines. Each virtual machine can have its own operating system (different from the operating system of the underlying physical machine) and its own unique set of applications.

As an end user, you expect to consume one or more services from your cloud computing provider over the Internet. These services can range from bare-bones virtual machines with a basic operating system to entire suites of applications. The four common models for cloud services are shown in Figure 1.1. These can be conceptualized using a layered model, with higher layers building upon the services offered by lower layers.

FIGURE 1.1 Common cloud service models



- **Infrastructure as a service (IaaS):** You specify the nitty-gritty of the virtual server you require, including the number of CPUs, RAM, hard disk space, networking capabilities, and operating system. The cloud provider offers a virtual machine to match these requirements. A scenario in which you want to store files on a cloud-based virtual hard disk will come under the IaaS service model.
- **Platform as a service (PaaS):** You specify not only the infrastructure but also the precise version of all the software you require. The cloud provider offers a virtual machine with preinternal applications to match these requirements.
- **Software as a service (SaaS):** You specify the kind of software application you want to use, such as a word processor. The cloud provider provisions the required infrastructure, operating system, and applications to match your requirement. Most SaaS cloud providers include limited choice of the hardware characteristics that run the application.
- **Business process as a service (BPaaS):** You specify a business process that you want to outsource to a cloud provider. The cloud service provider provisions the hardware, operating system, support software, and web applications to provide the required service. A good example of BPaaS would be a cloud-based service to compute quarterly value-added tax (VAT) returns and submit these returns to the relevant tax authority on your behalf. Such a service could present you with a browser-based front end in which you upload your invoices and business bank statements. The service could then extract relevant information from the uploaded documents (using OCR perhaps), fill up the relevant tax authority's forms, ask you to review the results, and submit the return on your behalf after you are happy with the numbers.

Cloud Deployment Models

A deployment model answers the following questions:

- Who can access a computing resource?
- How can a user access a computing resource?
- Where is the physical hardware?

Cloud-computing solutions have four distinct deployment models:

- **Public cloud:** A public cloud provides services over the Internet to a consumer located anywhere in the world. The physical resources utilized by the provider to supply these services can also be anywhere in the world. This type of service could represent potential challenges to organizations such as banks that are prevented by regulatory requirements from storing customer data in other countries.
- **Private cloud:** A private cloud offers services to a single organization. Services are provided over a secure internal network and are not accessible to the general public over the Internet. The organization owns the physical hardware that supplies underlying services. Because of the large infrastructure costs associated with this model, only larger corporations can afford to have their own private clouds.
- **Community cloud:** A community cloud provides services to a small group of entities (individuals or corporations) over a secure network. The underlying resources used to supply the services are owned by the entities that the community cloud serves. In essence, this type of cloud service can be thought of as something between a public cloud and a private cloud. The service is not publicly accessible to any user and does not put a significant drain on an entity's finances. The entities involved usually share a common goal or provide services in a common industry sector.
- **Hybrid cloud:** A hybrid cloud is essentially a cloud service that is composed of other types of cloud services. For example, a hybrid cloud could consist of both public and private clouds. The public subcloud could provide services that are intended for consumption by any user over the Internet. A private cloud could offer services that are sensitive to the business.

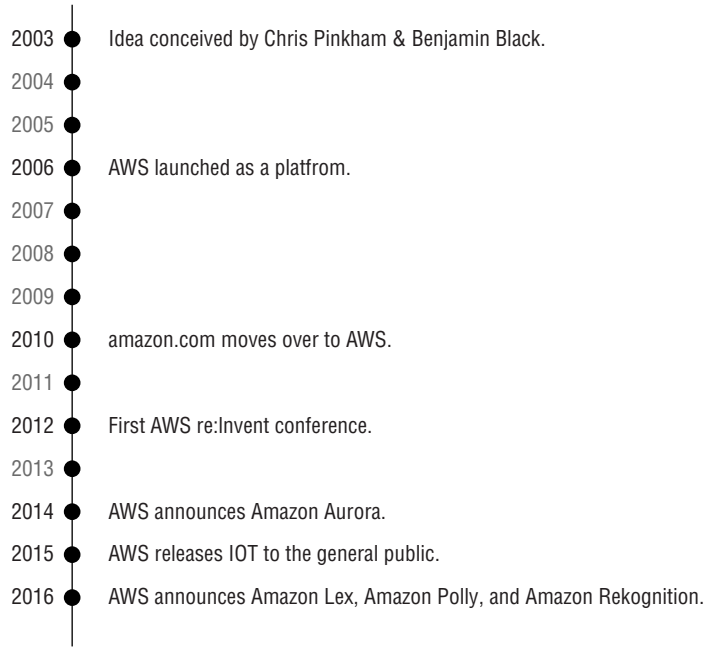
The AWS Ecosystem

Amazon Web Services is the most rapidly evolving cloud-computing service in the market. The current AWS offering consists of several dozen services offered in locations around the world, with new services being added every year.

New additions to AWS are announced at AWS re:Invent, which is the official annual AWS conference. You can find more information on AWS re:Invent at `https://reinvent.awsevents.com`.

Figure 1.2 describes a brief timeline of AWS. AWS was borne in a paper presented by Chris Pinkham & Benjamin Black to Jeff Bezos in 2003.

FIGURE 1.2 Brief timeline of AWS services



This paper suggested selling Amazon's internal infrastructure as a service to the world. In 2006 AWS was officially launched as a platform with a few key services including EC2 and Amazon S3.

In November 2010, Amazon announced that all of amazon.com had migrated to AWS. The first AWS re:Invent conference took place from November 27 to November 29, 2012 at the Venetian Hotel in Las Vegas. In 2014, Amazon Aurora was announced. Aurora is a MySQL-compatible high-performance database.

At the time this book was written, AWS offered 97 services, each of which resided in one of 18 different service categories. The service categories are as follows:

- Compute
- Storage
- Database
- Migration
- Networking & Content Delivery
- Developer Tools
- Management Tools
- Security, Identity & Compliance
- Analytics

- Artificial Intelligence
- Mobile Services
- Application Services
- Messaging
- Business Productivity
- Desktop & App Streaming
- Internet of Things
- Contact Center
- Game Development

You can find details on all AWS services within these categories at `https://aws.amazon.com/products/`.

The services that are covered in this book are listed below:

- **EC2:** Amazon Elastic Compute Cloud (EC2) provides secure, scalable, on-demand hardware virtualization in the cloud. Using Amazon EC2 eliminates the need to invest in server hardware up front so that users can focus on, develop, and deploy applications faster. EC2 instances come in various hardware configurations and include an operating system. EC2 is covered in Chapter 4.
- **Lambda:** AWS Lambda lets users run snippets of code without provisioning an infrastructure. This service is billed on a pay-as-you-go model, with users only paying for the execution time of their lambda code. There is no charge when code is not running. Lambda code can be set up to automatically trigger from other AWS services or called directly from any web or mobile app. AWS Lambda is covered in Chapters 7, 14, 21.
- **S3:** Amazon Simple Storage Service is a secure, durable, and scalable cloud-based object store. Using this service, you can store your files in the cloud. Amazon S3 is covered in Chapters 5, 13, 20.
- **DynamoDB:** Amazon DynamoDB is a high-performance, scalable cloud-based NoSQL database service. Amazon DynamoDB is covered in Chapters 6, 12, 19.
- **IAM:** Amazon Identity and Access Management (IAM) lets you securely control who can access your AWS resources, what resources they can access, and what they can do with these resources. IAM is covered in Chapter 3.
- **Amazon Cognito:** Amazon Cognito allows you to create identity profiles for your app's users and allows them to sign in to the app with their Amazon, Facebook, Twitter, or Google accounts. After users have authenticated from the app, the app is given a token that can be used to access AWS cloud resources securely. Amazon Cognito also offers a service that allows authenticated users to sync their app data on different devices. Amazon Cognito is covered in Chapters 9, 10, 11, 16, 17, 18.

- **Device Farm:** Amazon Device Farm is an app testing service that allows you to upload your app and run a set of automated tests and a set of real-time tests on physical devices hosted by Amazon. Both Android and iOS apps are supported. Amazon Device Farm is covered in Online Bonus Chapter 3.
- **Mobile Analytics:** Amazon Mobile Analytics allows you to collect and visualize app usage data. A set of reports is generated based on the collected data. You can export the analytics data to other AWS cloud services for further processing. Amazon Mobile Analytics is covered in Online Bonus Chapter 2.
- **SNS:** Amazon Simple Notification Service (SNS) enables you to send push messages. These messages can be pushed to mobile application, e-mail inboxes, web servers, and other AWS cloud services. SNS supports various protocols including SMS, e-mail, and HTTP/S. SNS is covered in Online Bonus Chapter 1.

Sign Up for an AWS Free Tier Account

To use AWS services you need to sign up for an AWS account. If you do not already have one, you can sign up for an account under the AWS free tier. An AWS account under the free tier is designed to enable you to try some of the AWS offerings free for 12 months, subject to certain usage limits. Go to `https://aws.amazon.com/free/` to obtain information on what is included in an AWS free tier account.

To create an AWS account under the free tier, you need to go through a five-step process. Some of the steps have multiple substeps:

- Contact Information
- Payment Information
- Identity Verification
- Support Plan Selection
- Confirmation

Step 1: Contact Information

To start the signup process for an AWS account under the free tier, visit `aws.amazon.com` and click on the Create an AWS Account link on the top-right corner of the page (see Figure 1.3).

Amazon frequently tries out new user experiences with its customers, so this page may look different from the screen shot. However, you should still be able to find the relevant option to create an AWS account on the page.

Type in a valid e-mail address on the Sign In or Create an AWS Account screen (see Figure 1.4), select I am a new user, and click on the Sign in using our secure server button.

FIGURE 1.3 Amazon Web Services home page

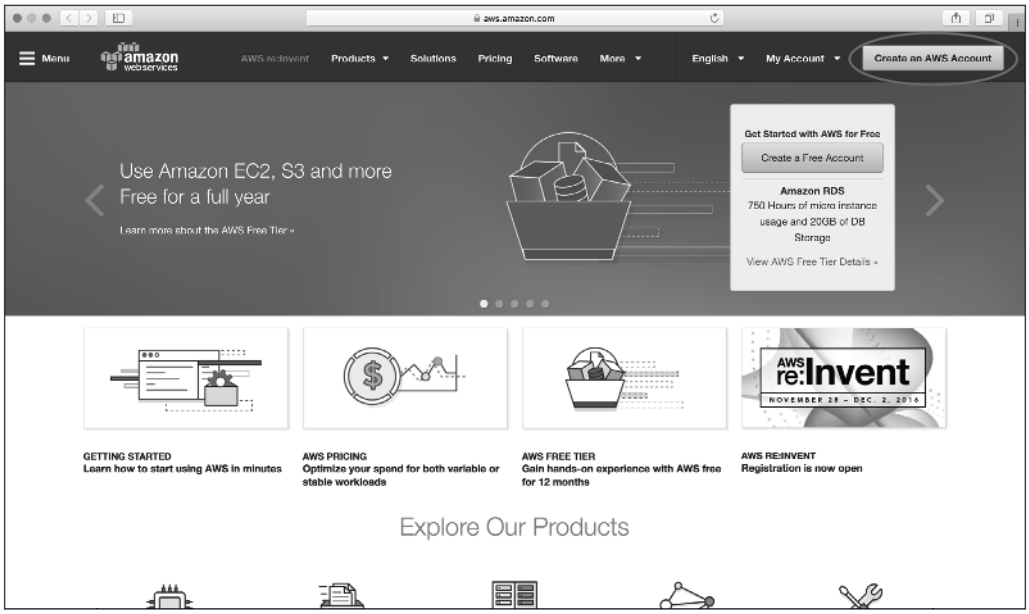
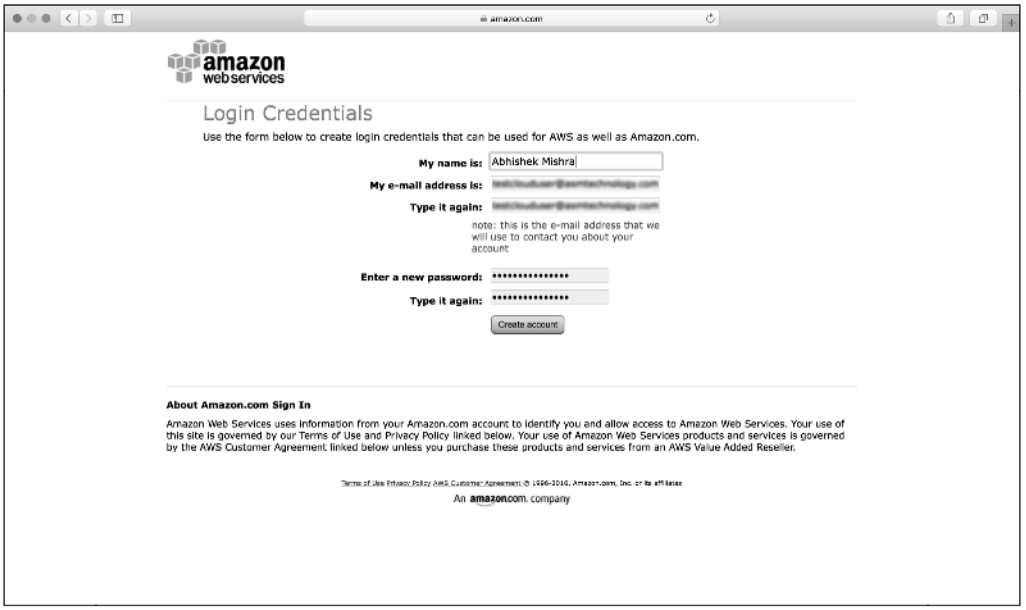


FIGURE 1.4 AWS sign-in screen

You will be asked to provide your full name and a password on the Login Credentials screen (see Figure 1.5). Click on Create account to move on to the next screen in the account creation process.

FIGURE 1.5 Creating login credentials



You will be asked to provide contact information (including a phone number) on the Contact Information screen (see Figure 1.6). You must provide a phone number that you have immediate access to and can receive a call on. Scroll down to the bottom of the page if necessary, read and accept the terms and conditions of the AWS customer agreement, and click on Create Account and Continue to move to the next step.

Step 2: Payment Information

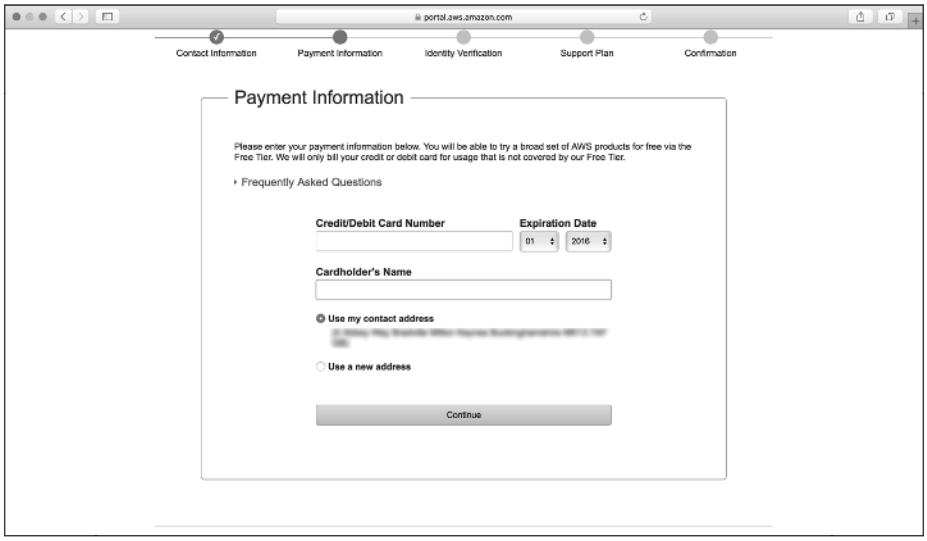
You need to provide credit/debit card details (see Figure 1.7). Although an account under the free tier provides access to some AWS services for free, not all services are included in the free tier. If you use services outside the free tier account, the card you provide is charged.

The precise services and options that are available under the free tier can change from time to time. Every effort will be made in this book to inform you whether an example utilizes AWS features outside those available in the free tier. To get up-to-date information on what is included in the free tier, visit `https://aws.amazon.com/free/`.

Type your credit/debit card details and click Continue to move on to the identity verification step.

FIGURE 1.6 Contact Information screen

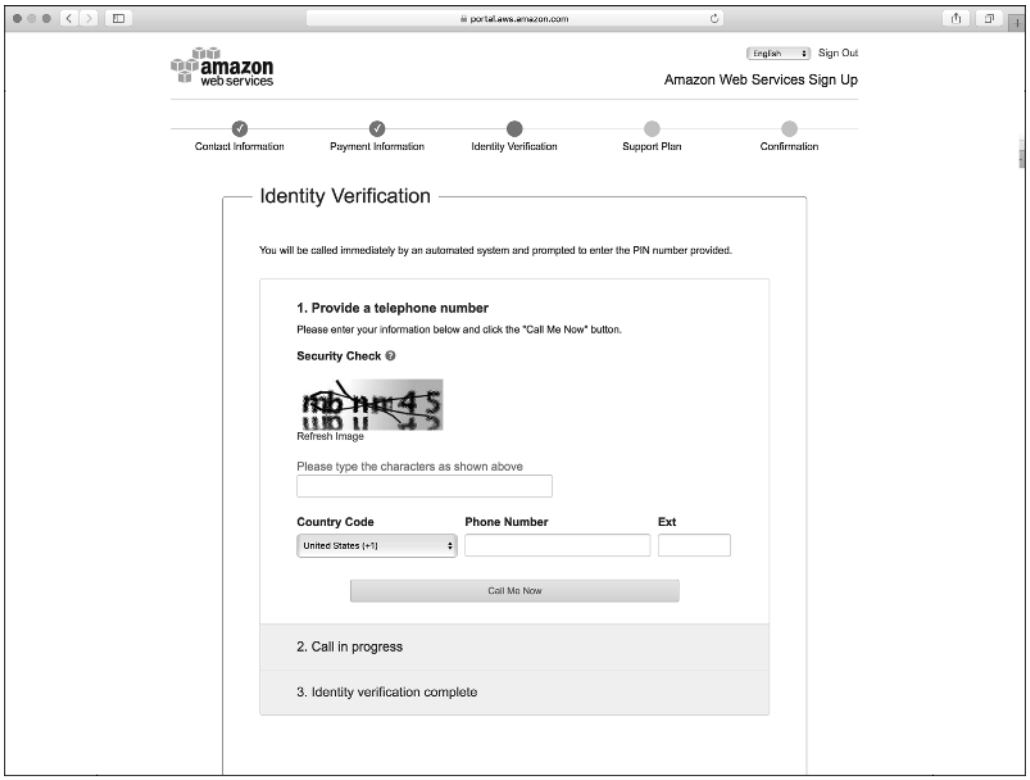
FIGURE 1.7 Payment Information screen



Step 3: Identity Verification

The identity verification process involves receiving a call from an automated system on a number you provide and entering a four-digit pin number into your phone when prompted. Type a telephone number and click Call Me Now (see Figure 1.8).

FIGURE 1.8 Identity Verification screen



A four-digit pin number then appears on the web page (see Figure 1.9). You receive a call on the telephone number you have provided and are asked to enter the four-digit pin you see on the web page.

The identity verification process completes after you key in the four-digit pin over the telephone call. The web page refreshes to reflect this (see Figure 1.10).

Click on Continue to select your Support Plan to move on to the next step of the account creation process.

FIGURE 1.9 Identity Verification PIN number

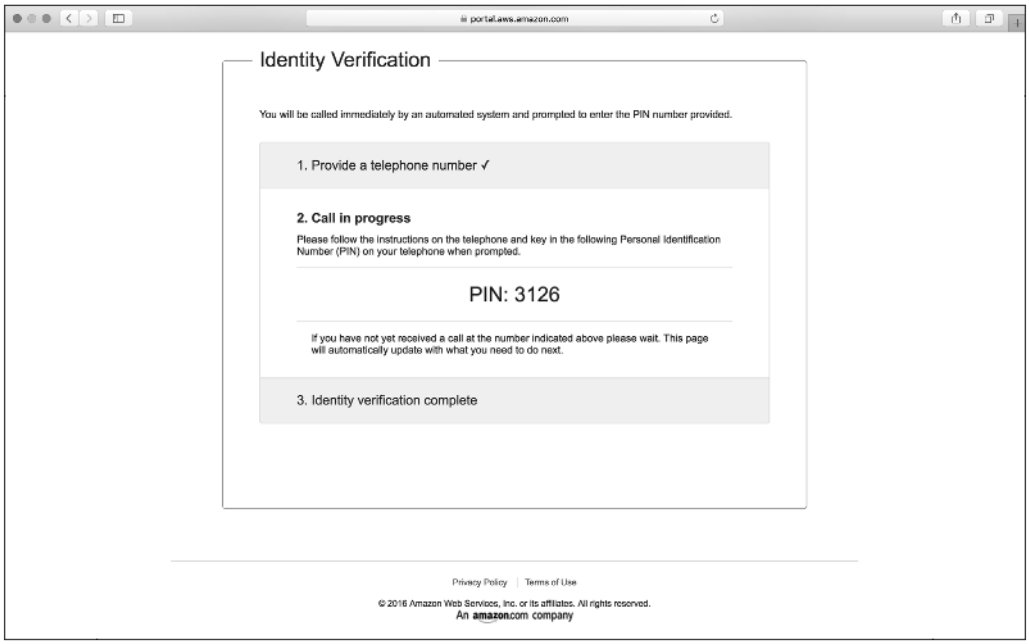
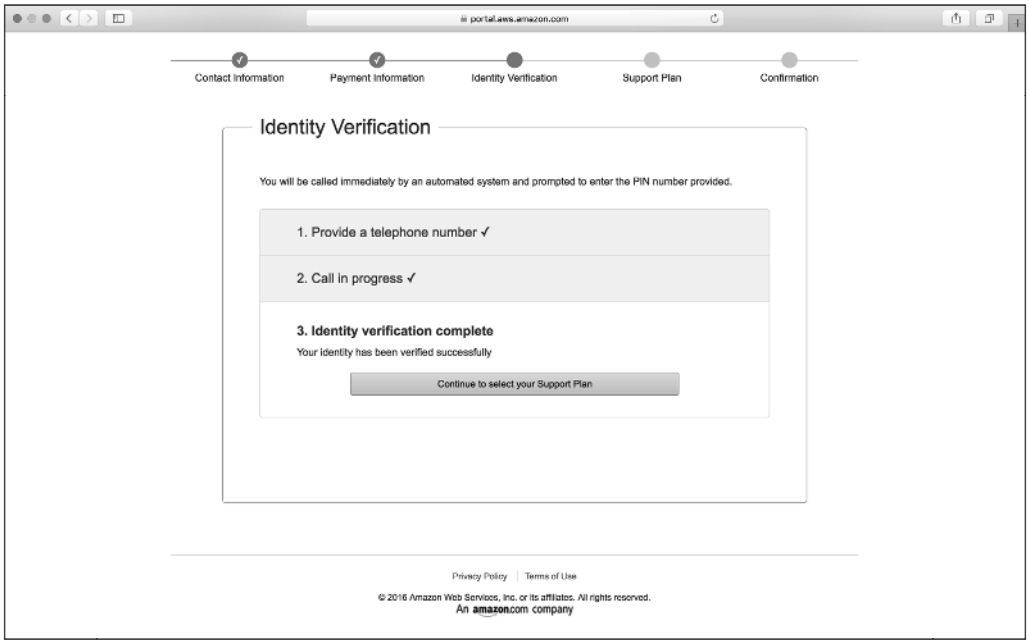


FIGURE 1.10 Completing the Identity Verification process

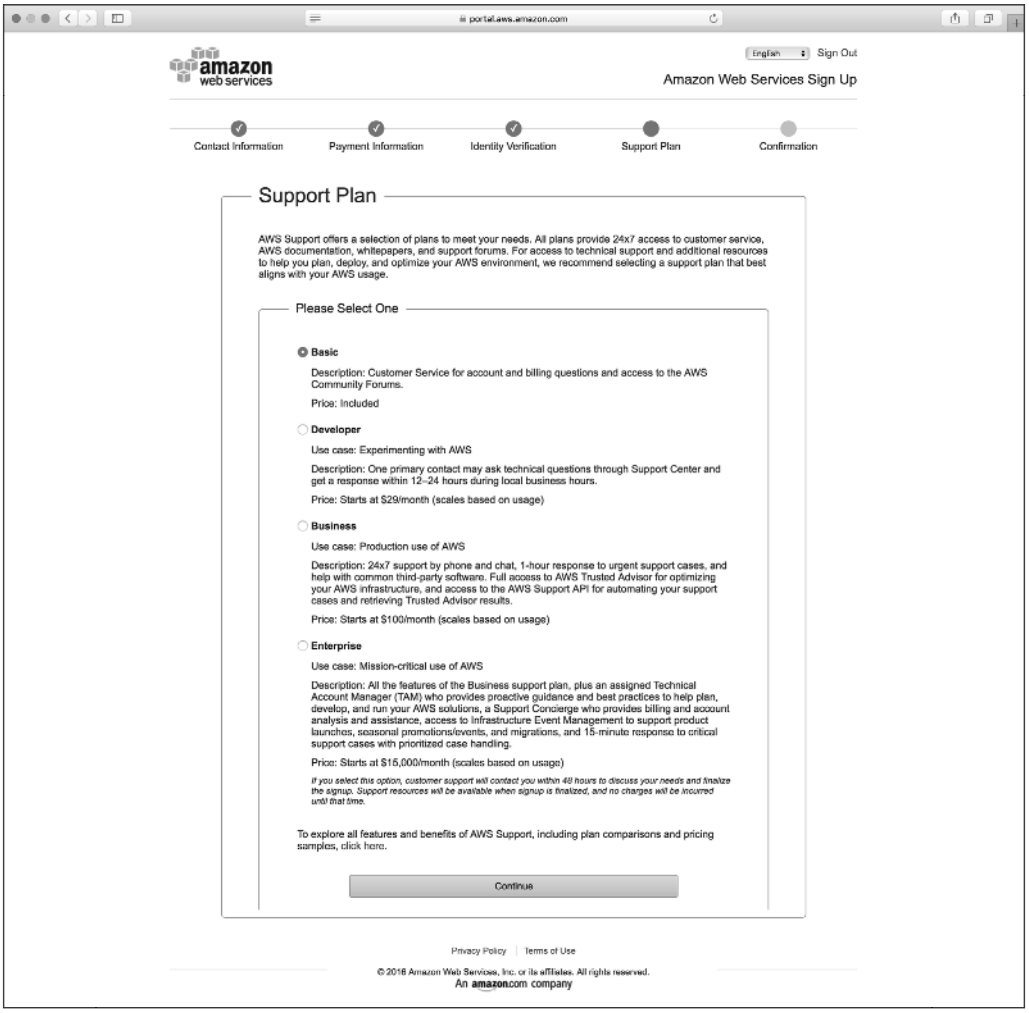


Step 4: Support Plan Selection

Select a support plan from the list of options available (see Figure 1.11). The options available follow:

- Basic
- Developer
- Business
- Enterprise

FIGURE 1.11 Support Plan selection

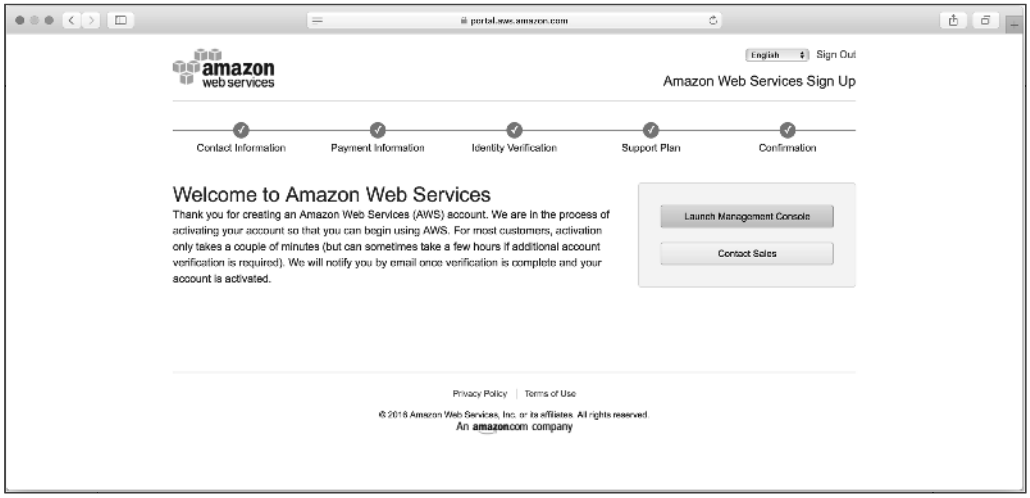


The support plans are cumulative and differ in the level of support that you receive, with the more expensive options giving you access to an Amazon employee to answer your questions. The default, selected option is Basic, and it is free. For the purposes of this book, the Basic support plan suffices. Select the default Basic support plan and click on Continue.

Step 5: Confirmation

You receive confirmation that your AWS free tier account is now set up (see Figure 1.12).

FIGURE 1.12 Completing the signup process



You also receive a confirmation message by e-mail to the address you used during the signup process. In Chapters 2 and 3 you learn different ways you can access AWS and how to secure the account you have just created.

Summary

- Cloud computing is defined by the U.S National Institute of Standards and Technology (NIST) as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."
- EC2 was the first service launched by Amazon and is one of its core service offerings today.
- Cloud computing provides both cost and availability benefits to businesses.

- Common cloud computing service models include infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and business process as a service (BPaaS).
- Cloud solutions are deployed using standard deployment models. A deployment model defines how a computing resource can be accessed, who can access the resource, and where the physical hardware is located.
- There are four distinct deployment models that are commonly used for cloud solutions. These are the private cloud, public cloud, community cloud, and hybrid cloud deployment models.
- AWS offers over 97 different cloud services, grouped into one of 18 different service categories. Amazon is continuously adding to the services that are available.
- An AWS account under the free tier is designed to allow you to try some of the AWS offerings free for 12 months, subject to certain usage limits.