**CHAPTER 1**

**Oracle Cloud Infrastructure Concepts**

In this chapter, you will learn how to

•   Discuss OCI and cloud computing models

•   Describe the features and components of OCI

•   Describe the OCI concepts and terms

Oracle has reinvented its infrastructure cloud as the best platform for hosting Oracle technology and one of the leading cloud vendors for non-Oracle workloads. The first-generation cloud offering, known as Oracle Cloud Infrastructure Classic, offers a limited set of compute, storage, and networking offerings. The second-generation cloud offering, and the focus of this book, is Oracle Cloud Infrastructure, or OCI. Learning from both its initial foray into the cloud space and its competitors, Oracle has designed a unique infrastructure cloud offering that is open and extensible with some great features.

This chapter introduces the underlying concepts supporting OCI and provides a broad overview of the various components and features while highlighting how the puzzle pieces fit together. We discuss core concepts and terms to provide the foundation required to master the topics related to Oracle Cloud Infrastructure.

Whether you are an experienced Oracle professional or you are just entering the game, have no illusions—cloud computing is here to stay and it is the future. With the advent of OCI, there has never been a more exciting time to embrace Oracle technology. Coupling your understanding of Oracle Cloud Infrastructure with your experience with Oracle software can be both invigorating and exciting to your career.

**Introduction to OCI**

OCI has been engineered to be simple yet powerful. Modern computer systems, whether hosted on your desktop or in a data center, consist of four primary elements:

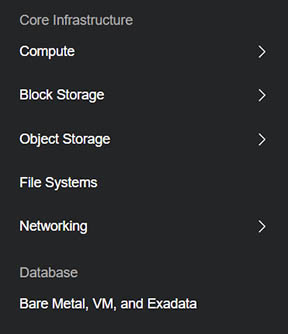
•   Non-volatile storage components that retain data after power is recycled and are usually some form of persistent disk storage.

•   Volatile storage components that are typically memory modules or RAM. These modules tend to be faster or offer higher IOs per second (IOPS) than persistent storage and are generally more expensive. Volatile storage also does not survive a power cycle event.

•   Networks that support communications between systems that rely on interfaces to connect to network routing infrastructure.

•   CPUs, or compute processing components, are the brains of the outfit, performing logical processing operations.

Infrastructure as a Service or IaaS (frequently pronounced *i-as*) is an abstraction of infrastructure components, including the four mentioned earlier. It is available in an online marketplace, enabling you to choose the most appropriate combination of these elements to meet your computing requirements. OCI and other IaaS vendors provide this basic service. [Figure 1-1](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig1) shows a subset of the infrastructure components that may be provisioned.



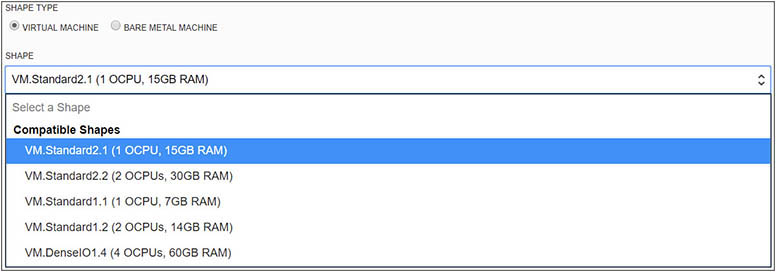
**Figure 1-1**   IaaS options from OCI

Virtualization of resources is the underlying philosophy behind IaaS. On premises, virtualization technologies like Oracle Virtual Machine (OVM) paved the way for consolidation and pooling of resources and sharing these between VMs to optimize hardware and infrastructure efficiency. This approach is very prevalent but presents several challenges. On premises, virtualization requires an initial capital outlay for hosting servers and setting up the VM environments. Provisioning a new VM system is generally faster than traditional methods, but ongoing maintenance is expensive and ecosystem growth is limited by the size of the physical virtual servers. OCI solves these problems by making a practically inexhaustible supply of resources accessible to you on demand, absolving you of the responsibility of maintaining storage, network, and servers, while refreshing the hardware transparently to ensure your systems are running on a modern supported kit.

An important feature provided by OCI relates to disaster recovery and high availability. On-premises solutions include hosting servers in multiple data centers and ensuring high-speed network connectivity between them. Backups are usually also taken and are sometimes moved to offsite storage locations.

OCI represents a collection of resources, both virtualized and bare-metal systems grouped in data centers known as *availability domains* (ADs). One or more ADs are grouped into a region. You get access to multiple Disaster Recovery (DR) locations and access to archival storage for offsite backups with just a few clicks. The value proposition is staggering.

With OCI, you choose the processing or compute power for a system you want to spin up. Oracle was the first cloud vendor among other IaaS providers to offer both virtual and bare-metal machines. Virtual machine instances may be provisioned based on available compute shapes, which are preset bundles of RAM, and compute power called OCPU (Oracle CPUs). [Figure 1-2](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig2) shows several VM shapes, highlighting the VM.Standard2.1 shape with one OCPU and 15GB RAM. Other instance provisioning items allow you to specify network configuration and to allocate storage. The infrastructure options available allow rapid provisioning of a wide range of systems from Windows desktops to Exadata Engineered systems.



**Figure 1-2**   VM shape types

Various infrastructure options exist for most components and more options are added as new technology is bolted onto OCI. For example, the first set of bare-metal x86 servers available with OCI was the X5 generation. These were replaced with the X7 generation of servers. OCI infrastructure follows a classical hardware refresh cycle and is periodically updated with modern replacements. At the storage layer, there are options for block storage, usually used for server file systems or database storage. These vary from slower devices to high-speed Non-Volatile Memory (NVMe) storage.

Traditionally, a new project implementation lifecycle would reach a stage where computing power resources are estimated and suitable server hardware is identified and ordered. After delivery and installation in one or more data centers, the system is ready for software implementation. Delivery and installation times vary widely and are inconsistent and unpredictable. This approach is inefficient and time-consuming, and these systems are usually over-provisioned.

Historically, the trend has been to design and architect systems for the worst-case scenario. It is common for enterprise software to be licensed based on the number of processors active on a server. A common trend is that many more CPU cores are licensed than actually required to cater to increased load on a few days of the year (for example, Black Friday and Boxing Day for retailers or Admissions week at colleges). During these peak times, there is a spike in processing power demand, while the system usually needs only a fraction of these resources for the vast majority of the year. OCI resolves this inefficiency by offering bursting-on-demand or the dynamic provisioning of processing power during peak loads, which is another very compelling reason you should consider OCI.



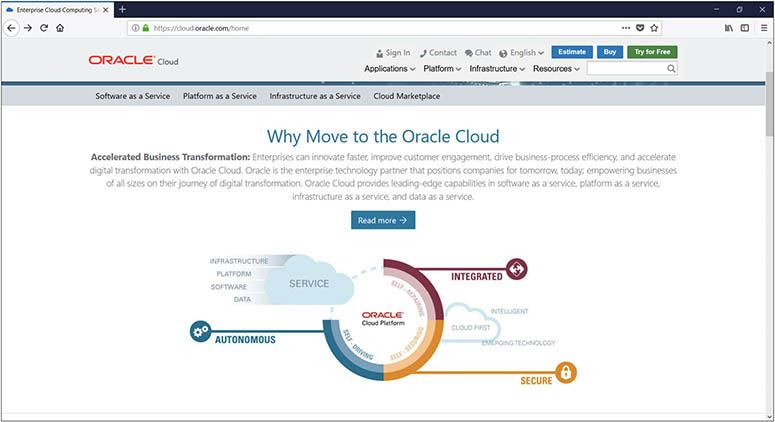
**NOTE**    OCI-Classic, launched in 2014, was known as Oracle Public Cloud (OPC). The first-generation IaaS offering supports only virtual machines (VMs) and was based on cloud infrastructure management software called Nimbula Director (acquired in 2013 by Oracle), which was a Xen-based hypervisor. OPC was renamed OCI-Classic in 2017. Cloud at Customer currently utilizes a similar hypervisor to OCI-Classic.

Oracle provides free credits to get you started with OCI. As you work your way through this book, we strongly recommend that you do so with access to your own OCI account.

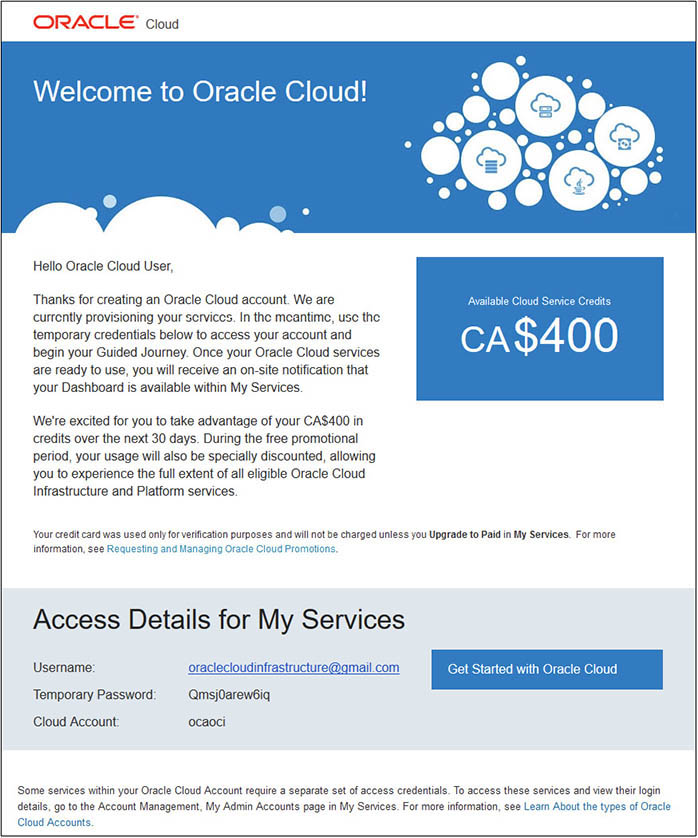
**Exercise 1-1: Create a New OCI Account**

In this exercise, you will create a new OCI account. If you already have access to an OCI account, you can skip this exercise. To register for the free credits, you need an email address that, preferably, has not been used previously to register for OCI. You will also need a credit card. Note that unless you upgrade to the paid option, you will not be charged. It is still safest to use a credit card with a low credit limit. As of this writing, the workflow to create a new OCI account is as follows. These steps may vary as OCI interfaces are updated.

**1.**   Navigate to [http://cloud.oracle.com](http://cloud.oracle.com/) and choose Try For Free.

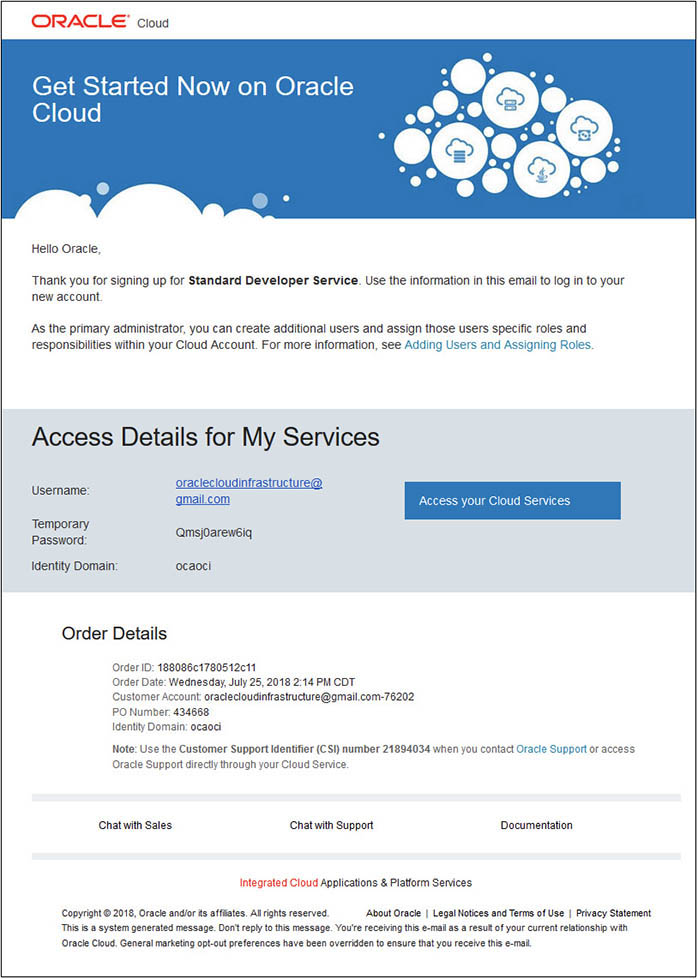


**2.**   Complete the required fields to sign up. Pay attention to the cloud account name. This must be unique, and because this will appear in some URLs related to this account, you may wish to choose a recognizable name. The email address you provide will also be your username. After providing name and address information, the second section verifies that you have a mobile phone from a supported country. You are required to enter a verification code texted to your mobile device and click the Verify button. The third section validates your credit card information. If you accept the terms and conditions, check the box and click Complete. You should land at the getting-started portal where Oracle thanks you for signing up and begins provisioning your account. There are some interesting videos on this page, well worth a watch while you wait for the provisioning email to show up in your Inbox.

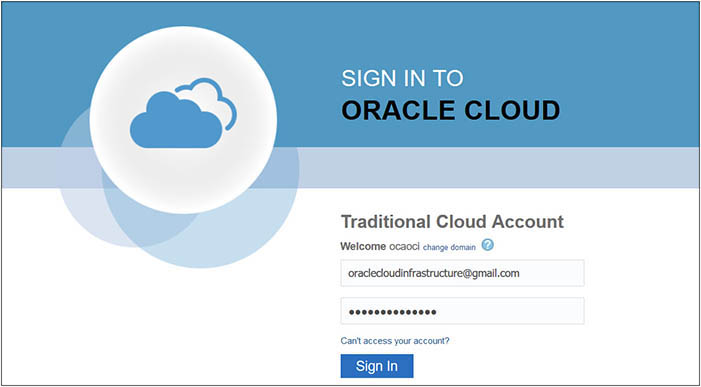


**3.**   After signing up, you should receive several emails initially. One provides credentials, including a temporary password for My Services. Take note of the cloud account name that you provided when registering. In this example, the cloud account is called ocaoci. Clicking the Get Started With Oracle Cloud button takes you to a login screen labeled Identity Cloud Service. Sign in with the credentials in the email, and you will be prompted to set a new password, language, and time zone preferences.

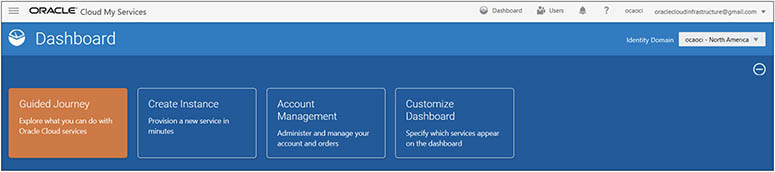
**4.**   You have just signed into your cloud account, authenticated with Oracle Identity Cloud Service (IDCS).



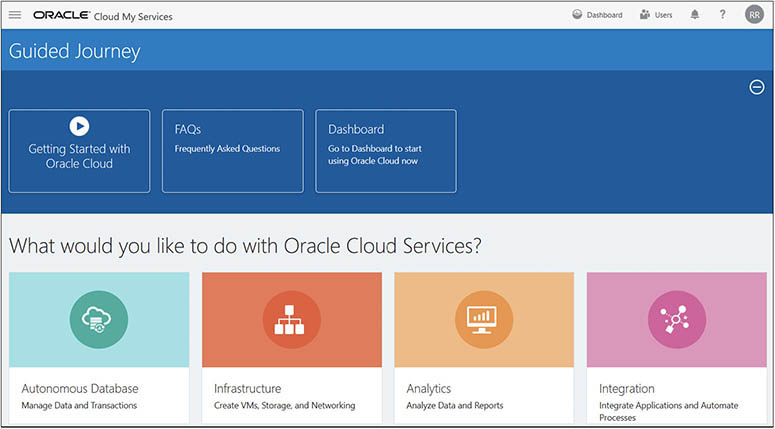
**5.**   You will receive another email to confirm that you signed up for the Standard Developer Service. It also provides additional account information, such as your order details and even a Customer Support Identifier (CSI) number you may use when accessing Oracle Support on [http://support.oracle.com](http://support.oracle.com/). Notice that an Identity Domain has been created with the same name as the cloud account name. Choosing to Access Your Cloud Services strangely takes you to a different login screen, labeled Traditional Cloud Account.



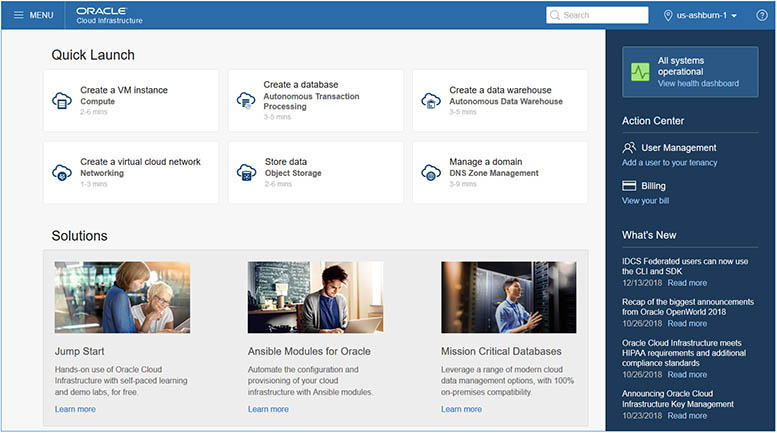
**6.**   It gets tricky at this point. If you have been following along, you would have chosen a new password that is different from the temporary password emailed to you in an earlier step. However, when signing in to the Traditional Cloud Account, you have to use the temporary password from the email. Once authenticated, you are prompted by Oracle Identity Self Service to set a new password and to register challenge questions for your account. You now have two different types of cloud account: the traditional account and the Identity Cloud Service account. These different account types will be explored in the next chapter. For most of the exercises in this book, you will use your IDCS authenticated account.



**7.**   Explore the Dashboard interface. It may appear quite simple, but you have access to many powerful services. Note the Identity Domain drop-down list related to the two types of cloud accounts. The traditional and standard identity domains let you access the OCI-Classic and OCI features respectively. Choose Guided Journey to access various guided services.



**8.**   There are several ways to launch the OCI console from the My Services interface, including using the hamburger menu and choosing the Compute option in the Services menu or choosing Infrastructure from the Guided Journey screen and selecting Create Compute Instance. Various OCI console options are explored in this book, so prepare to become very familiar with this interface. You primarily navigate the OCI console using the burger menu. Choose Menu | Administration | My Services Dashboard to return to the dashboard or choose Oracle Cloud Infrastructure next to the burger menu to access the OCI console Quick Launch page.

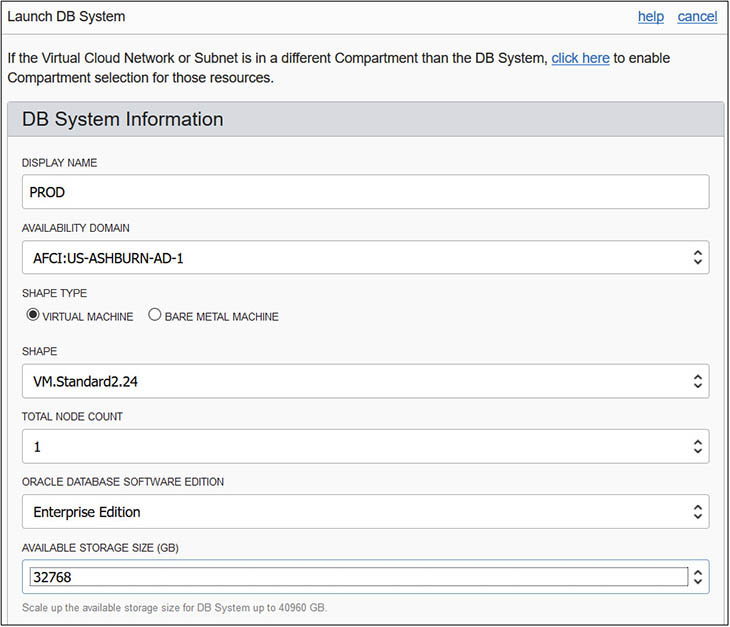


**Cloud Computing Models**

Modern cloud computing refers to a collection of applications, the platform, servers, storage, and network infrastructure resident in a data center, exposed as services. Three primary models of cloud computing typically implemented as services are Infrastructure, Platform, and Software as a Service—or IaaS, PaaS and SaaS.

•   **IaaS**   A collection of servers, storage, and network infrastructure onto which you deploy your platform and software. This is most akin to provisioning your own hardware in an on-premises data center. Teams of hardware engineers, storage specialists, network specialists, system administrators, and database administrators are usually involved in installing and configuring on-premises infrastructure. With IaaS, no hardware engineers or storage specialists are required. A good cloud architect (like you) is all that is required to design and provision this infrastructure. The cloud vendor provides the hardware engineers and storage specialists.

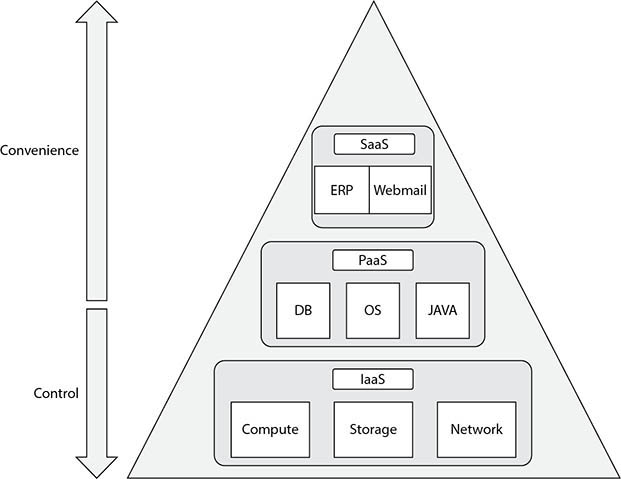
•   **PaaS**   A collection of one or more preconfigured infrastructure instances usually provided with an operating system, database, or development platform onto which you can deploy your software. The primary benefit of PaaS is convenience as the cloud vendor provides and supports the underlying infrastructure and platform. A subset of PaaS is Database as a Service (DBaaS). [Figure 1-3](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig3) shows how you could provision an Oracle database instance with the required compute, memory shape, version, and block storage with a few clicks. Behind the scenes, OCI creates a machine, deploys an OS, installs the requested database version, and provides a functional platform in minutes. You may then connect to the machine where an Oracle database is ready and waiting. There is no need for explicit hosting and maintaining hardware infrastructure, operating systems, and databases.



**Figure 1-3**   Database as a Service

•   **SaaS**   Applications are deployed on a cloud and all you do is access them through your browser. These could range from webmail to complex ERP and BI Analytic systems.

Oracle Cloud encompasses the Oracle Public Cloud, which represents a collection of infrastructure, platforms, and applications exposed as services on [cloud.oracle.com](http://cloud.oracle.com/). [Figure 1-4](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig4) highlights two important factors on the same spectrum when considering cloud computing models: control and convenience. The convenience offered by SaaS and PaaS comes at the cost of less control. IaaS offers the most control but requires the most effort. With IaaS, you have complete access to the infrastructure you provision. For example, when you provision a compute instance, you have the option to choose the operating system image. With PaaS, if you provision a database, you will have administrator privileges, but you have fewer options. For example, when you provision a database, you cannot choose the operating system on the compute instance created. SaaS offers maximum convenience, but you are dependent on the cloud vendor for maintenance and support.



**Figure 1-4**   IaaS, PaaS, and SaaS

The placement of the OS (operating system) in the IaaS tier is sometimes debated. Some argue that IaaS refers to just the virtualized or physical infrastructure and therefore the OS forms a part of the PaaS layer. However, the NIST definition (<https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>) specifies IaaS as “The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.” OCI appears to be congruent with this definition.

Some highly regulated and sensitive environments cannot consume public cloud services because of data residency (data must reside in a particular locale) and data sovereignty (data may only be managed by resources located in a particular region). For organizations in countries without local OCI data centers, Oracle offers IaaS, PaaS, and SaaS through an on-premises solution such as the Cloud at Customer products discussed in [Chapter 8](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch8.xhtml#ch8). Oracle Cloud comprises SaaS, PaaS, OCI-Classic (OCI-C), and OCI. OCI is all about IaaS and the focus of this book and exam.



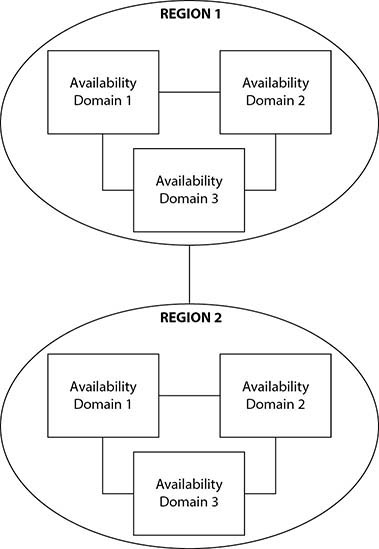
**EXAM TIP**    The three primary cloud-computing models are Infrastructure, Platform, and Software as a Service, better known as IaaS, PaaS, and SaaS. Compute, Storage, and Network resources fall under IaaS. Java and Database Cloud Services are examples of PaaS while ERP, SCM, and Analytics Cloud applications are examples of SaaS.

**OCI Features and Components**

OCI was engineered from the ground up using expertise and learning derived from OCI-Classic as well as other mainstream cloud vendors. OCI shares many features common to most mainstream cloud vendors. However, several features that are unique to OCI are key differentiators. Amazon Web Services (AWS) offers a plethora of services targeting the entire spectrum of cloud consumers. OCI focuses on enterprise consumers and has specifically catered to hosting Oracle workloads while still providing a general-purpose IaaS framework. For example, OCI regions and availability domains have AWS parallels called regions and availability zones, while OCI has off-box networking that is not provided by most other cloud vendors. This constantly evolving feature set is explored in the next section.

**Regions and Availability Domains**

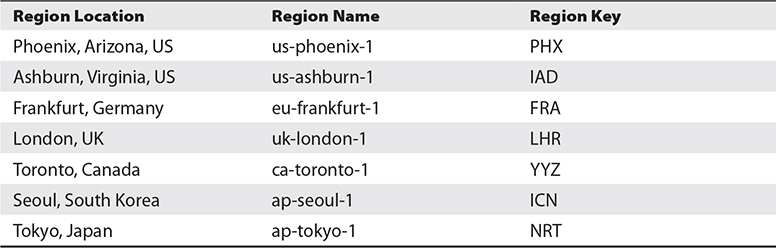
Cloud infrastructure consists of servers, storage, and networking equipment. These reside in data centers. Data centers with resilient and redundant components that do not have a single-point of failure are referred to as fault-tolerant data centers. In OCI *cloud-speak*, a fault-tolerant data center is an availability domain (AD). One or more availability domains located in a metropolitan area connected with high-speed networks are grouped into a region. [Figure 1-5](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig5) describes two regions, each comprising three availability domains.



**Figure 1-5**   Regions and availability domains

Each AD contains three fault domains. A fault domain is an infrastructure grouping that allows your instances to be distributed so they do not reside on the same physical hardware within an AD. When creating instances you may optionally specify the fault domain to control instance placement. Fault domains offer additional protection at a physical server level against unexpected hardware failures and improves availability during planned outages.

The predictable consistent network performance available in OCI is a key differentiator due to the off-box network architecture that will be discussed in the next section. [Table 1-1](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1tab1) lists the initial five OCI regions generally available, in various countries across several continents.

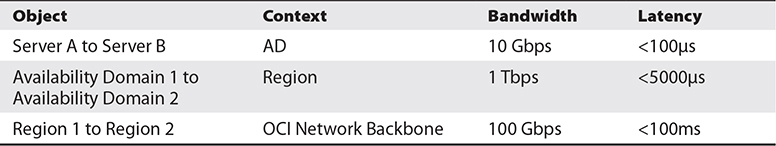


**Table 1-1** OCI Regions

Regions including Brazil and India as well as several government regions in the United States form the second wave of provisioned OCI regions. You have access to all regions, or you can subscribe to specific regions for data residency compliance or performance requirements. Placing your primary infrastructure in a geographical region nearest to your users reduces network latency.

Network performance is typically measured by the metrics *bandwidth* and *latency*. Bandwidth refers to the throughput or volume of data that can be transferred over time. For example, 5Gb/s means that five gigabits of data will be transferred over the network from one endpoint to another per second. Latency refers to the delay or time taken for a data packet to traverse a network. It is usually measured as the Round Trip Time (RTT).

Oracle claims the network bandwidth between servers in each AD is 10 Gbps with a latency of less than 100 microseconds. This network is a flat high-speed non-oversubscribed *Clos* network that provides around one million network ports per AD. Oracle promises a maximum of two network hops between Compute and Storage resources regardless of the size and scale of the estate. The bandwidth between ADs in each region is 1 Tbps with a latency of less than 5,000 microseconds. Finally, the bandwidth between regions, which are geographically vast distances apart, is 100 Gbps with a latency of less than 100 milliseconds. These network performance metrics are listed in [Table 1-2](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1tab2).



**Table 1-2** Network Performance Between Servers, ADs, and Regions

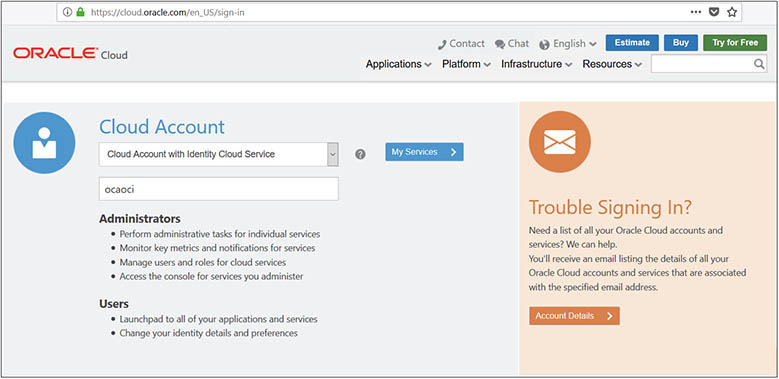
The high-performance AD and region network architecture provides two key benefits:

•   **Isolation**   Each AD is self-contained, contains highly fault-tolerant computing resources, and is isolated from the other ADs in the region. ADs in a region do not share power, cooling, or the internal AD network. Failure of one AD is very unlikely to impact the availability of other ADs in the region. High availability and DR zero data loss architectures such as Oracle Maximum Availability Architecture (MAA) may be built within a single region utilizing the low-latency, high-bandwidth network, supporting near real-time data replication between availability domains.

•   **Disaster recovery**   It is unlikely that multiple ADs within a region will fail simultaneously. However, natural disasters and other compliance requirements may necessitate an architecture (usually DR strategy) spanning multiple regions. This is supported by a cloud backbone and direct peering that enables private connections between regions.

**Exercise 1-2: Explore Your Availability Domain and Region**

**1.**   Navigate to [http://cloud.oracle.com](http://cloud.oracle.com/) and choose Sign In. Provide the cloud account name you specified when registering, and choose My Services.

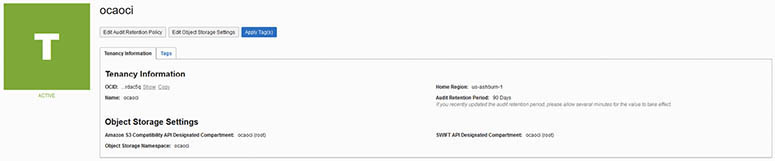


**2.**   Upon login, find your way to the OCI interface. One route is to choose the Create Instance from the My Services dashboard. At the OCI interface, note your home region to the right of the search box.

Images

**3.**   When registering your cloud account in the previous exercise, you specified a cloud account name as well as a Default Data Region.

**4.**   When registering the *ocaoci* cloud account, North America was selected as the Default Data Region. In the OCI console, navigate to Administration| Tenancy Details. The cloud account name is the OCI tenant name and is listed under Tenancy Information. OCI randomly allocates a home region close to the Default Data Region specified. OCI allocated the us-ashburn-1 region as the home region for this tenancy. Note the OCI console URL. In this tenancy, the URL is [https://console.us-ashburn-1.oraclecloud.com](https://console.us-ashburn-1.oraclecloud.com/). Your OCI console URL should begin with your home region.



**5.**   Note the Home Region, tenant name, and OCID (sometime pronounced *o-sid*). Your OCID uniquely identifies your OCI environment. Copy your OCID and keep this safe. Choose another available region, such as uk-london-1, and select Subscribe To This Region.

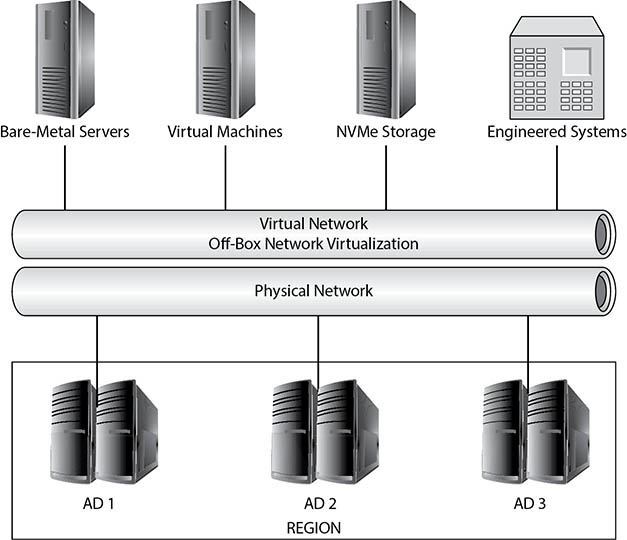


**6.**   The encircled R icon adjacent to the newly selected region turns from gray to orange, indicating work-in-progress, and a short time later turns green. You have just added your tenancy to another OCI region across the world. Choose the new region and examine your Tenancy OCID. This value has not changed and uniquely identifies your tenancy across regions. Each Oracle Cloud Infrastructure resource has a unique, Oracle-assigned identifier called an Oracle Cloud ID (OCID).

**Off-Box Network Virtualization**

One of the key design decisions taken by the OCI development architects was to implement off-box network virtualization. Traditional cloud vendors use hypervisors on servers to run multiple workloads through virtualized operating system and network functions, including I/O calls. In OCI, I/O virtualization has been relocated from the hypervisor layer to the network.

This separation has several important implications. A common cloud performance issue occurs when I/O from multiple VM workloads overwhelms the hypervisor, which leads to server performance degradation. This *noisy neighbor* situation cannot occur when I/O virtualization is off-box (off the server). Relocating the network and I/O virtualizations out of the software stack and into the network layer allows dedicated physical servers with a full software-defined layer 3 network topology to be defined. [Figure 1-6](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig6) shows this layer as the Virtual Network. Off-box network virtualization enables OCI to support VMs and bare-metal compute instances, RAC (clustered) databases, and engineered systems.



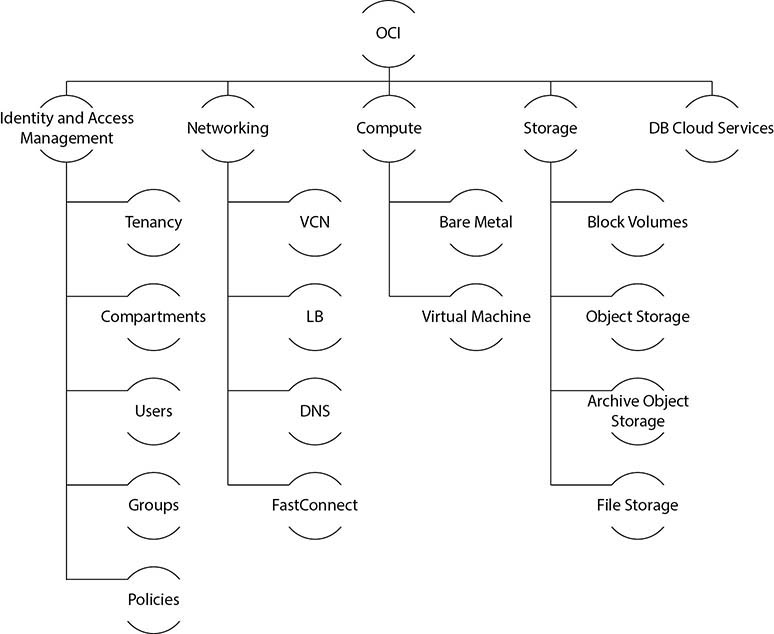
**Figure 1-6**   Off-box network virtualization



**EXAM TIP**    The noisy neighbor phenomenon occurs when VMs sharing a physical server are impacted by one or more VM workloads overwhelming the hypervisor. This situation is avoided in OCI through off-box network virtualization.

**OCI Concepts and Terms**

The sheer volume of offerings from cloud vendors can be overwhelming. [Figure 1-7](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig7) offers a framework that may be useful for classifying some of the popular OCI offerings. This is not an exhaustive list but provides a context for readers new to OCI to frame the plethora of cloud services. The concepts are described in a typical order of provisioning. Identity and access management is addressed first, then networking because resources associated with these concepts are usually provisioned first in cloud architectures. You could then create compute instances, allocate storage, or take advantage of the Database Cloud Service, a popular platform service. Strictly speaking, compute, storage, and networking services are pure IaaS offerings. In the Oracle world, PaaS offerings such as Database Cloud Service and Java Cloud Service are so prevalent that these, along with Container services and other popular offerings, are available on OCI.

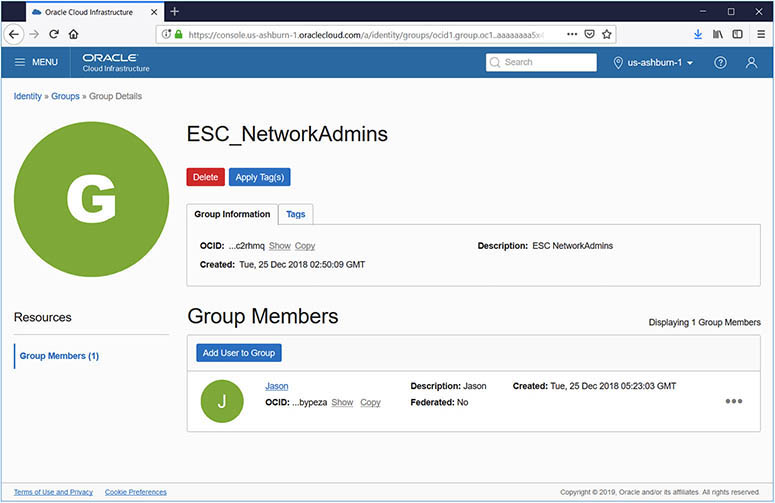


**Figure 1-7**   OCI taxonomy

**Identity and Access Management**

When starting with OCI, some of the first decisions you usually make relate to Identity and Access Management (IAM) entities. If you have registered an OCI account, you already have defined your tenancy, which is your cloud account. One of the next IAM constructs you create is called a compartment, which is a logical collection of related cloud resources. Compartments are used to isolate resources.

Users may be created, deleted, and have their accounts managed. [Figure 1-8](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig8) shows a user named Jason who belongs to the ESC\_NetworkAdmins group. Groups may be created with collections of Users. One of the really neat things that IAM on OCI offers is the natural, descriptive language used by IAM policies to grant access:



**Figure 1-8**   IAM resources

Allow group Developers to manage all-resources in compartment Lab

IAM resources have global scope and are available in each Compartment, AD, and Region. IAM Compartments, Groups, Users, and Policies are discussed in detail in the next chapter.



**EXAM TIP**    As new users are signed up for OCI, the cloud administrator should make use of IAM policies to grant appropriate access to Groups and then create Users.

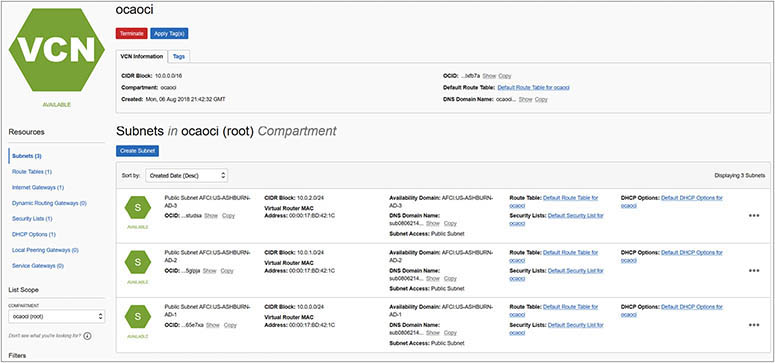
**Networking**

OCI networking is elegantly simple yet powerful and has direct parallels to networking in an on-premises data center.

**Virtual Cloud Network**

The first networking element you set up is your virtual cloud network (VCN), which works much like a traditional private on-premises network. The VCN is defined at a Compartment level so you could have multiple VCNs. At least one VCN must be set up before instances may be provisioned. There are simplified options to choose while creating a VCN that provisions related resources including subnets, a gateway, and a set of route rules to start you off with a working VCN. If you create your VCN manually, you have to provide a Classless Inter-Domain Routing (CIDR often pronounced *cider*) block, which specifies a range of IP addresses that may be allocated in that VCN.

When the VCN named ocaoci was provisioned, three subnets were automatically created, one for each availability domain. The VCN spans all ADs in a region. Notice in [Figure 1-9](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig9) that the CIDR blocks allocated to each subnet do not overlap. A default route table is created along with an Internet Gateway, allowing your compute instances to connect to the Internet if desired. A default security list is also created with several default ingress rules, one of which permits SSH access on port 22 to provide remote login and access to your compute instances. An ingress rule permits incoming traffic, while an egress rule permits outgoing traffic. The Internet Gateway provides a network path for traffic between the VCN and the public Internet.



**Figure 1-9**   Virtual cloud network

**Load Balancers**

A Load Balancer (LB) is a network device you may provision that receives incoming traffic on an IP address and routes the traffic to one or more underlying instances. The OCI LB service is a regional service that distributes traffic to instances either within the same availability domain or across multiple availability domains.

The protocol and ports being serviced by an LB are specified in an entity called the Listener. When creating an LB, you specify the VCN in which incoming traffic is accepted as well as whether it will be a private or public LB. You also choose the shape of the LB, which limits the speed at which network traffic is routed. LBs are commonly used to support high availability and scaling out of web servers.

LBs distribute traffic to backend servers based on a set of policies known as a *backend set*. Routing algorithms, including Weighted Round-Robin, IP Hash, and Least Connections, are specified when creating the backend set. Path Route Sets specify a set of rules to route requests to different backend sets but this is optional and is only used if this level of sophistication is necessary. Finally, backend sets reference one or more hostnames, which are the target compute instances, which may be running a web server.



**EXAM TIP**    OCI load balancers support three protocols: TCP, HTTP, and WebSockets.

**Domain Name Service**

OCI also provides a Domain Name Service (DNS) that lets you create and manage DNS zones, add records to zones, and allow the VCN to resolve DNS queries from your on-premises domain and vice versa. One of the primary services DNS provides is hostname resolution. For example, it is DNS that allows you to connect to [http://cloud.oracle.com](http://cloud.oracle.com/) instead of [http://23.9.97.203](http://23.9.97.203/). This abstraction provides network resiliency to underlying network changes.

**Dynamic Routing Gateway and FastConnect**

Connecting your existing on-premises infrastructure with your OCI VCN is a common step in the journey to OCI. This connectivity is enabled from OCI using a Dynamic Routing Gateway (DRG) that connects to an on-premises router created in OCI as Customer Premises Equipment (CPE). Your on-premises network is then bridged to your VCN using an encrypted IPSec VPN tunnel. OCI also offers FastConnect, which provides a dedicated, high-speed, private connection between OCI and your existing on-premises infrastructure. FastConnect requires that you must be either collocated with Oracle in a FastConnect location or that you connect through a third-party FastConnect provider that is already connected to Oracle.

**Compute**

When you provision a compute instance, you can choose a virtual machine (VM) or a bare-metal (BM) server. Bare-metal servers provide your instance with exclusive use of the hardware. Not sharing hardware with other instances comes at a cost and bare-metal instances are more expensive that similarly sized virtual machines. BM servers are only available with a much higher CPU and memory footprint than entry-level VMs.

When you provision a new compute instance, you specify a name; the AD it resides in; a boot volume, which may be an Oracle-provided image such as Oracle Linux 7.5; and the shape type, which is either BM or VM. Depending on the operating system image and the shape type chosen, you can select compatible shapes. For example, you could provision a VM with the shape labeled VM.DenseIO2.16 (16 OCPUs, 240GB RAM, 12.8TB NVMe SSD) or a bare-metal instance with the BM.Standard1.36 (36 OCPUs, 256GB RAM) shape. The shape names contain several useful identifiers. Standard means that only block storage is available while DenseIO refers to local NVMe drivers being present. The last digits in the shape name refer to the number of OCPUs, or Oracle Compute Units. An OCPU provides CPU capacity equivalent to one physical core of an Intel Xeon processor with hyperthreading enabled. Each OCPU corresponds to two hardware execution threads, known as vCPUs.

You are also required to provide a secure shell (SSH) public key. SSH keys are created in pairs. The matching private key, which must be stored safely, will allow you to connect to the instance using the SSH protocol once it is provisioned.

**Storage**

Once your compute instance is provisioned, it will have a boot volume and usually no other storage. Four storage types are available on OCI:

•   Block volumes

•   Object storage

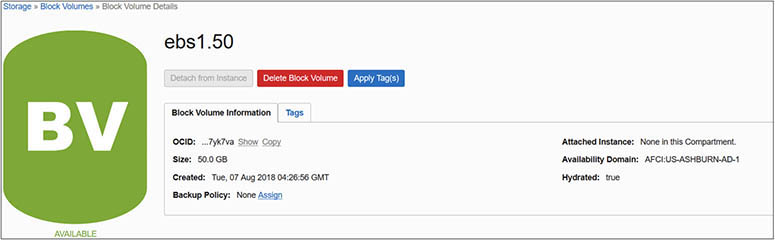
•   Archive storage

•   File storage

**Block Storage Volume**

Block volumes are often used for user-created file systems as well as database storage and reside in a specific AD.

[Figure 1-10](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig10) shows a 50GB block volume that resides in AD1 in the US-ASHBURN region. This block volume is not initially attached to any compute instance. After attaching the block volume to a compute instance using an iSCSI interface, OCI provides a handy list of commands to attach and detach the volume to your operating system, if required. I/O performance of block volume storage using iSCSI over Ethernet is comparable with modern SAN disks. Block volumes may be dismounted from one instance and mounted to another instance in the same AD without data loss.

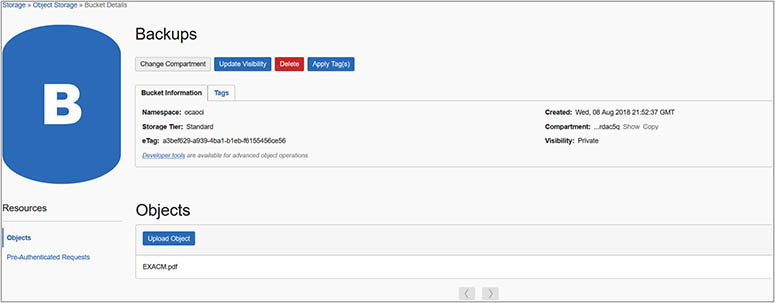


**Figure 1-10**   Block storage volume

**Object and Archive Storage**

Standard object storage (referred to as object storage hereafter) and archive storage are different tiers of the same storage solution. Using HTTP-based protocols, you can access these virtually unlimited durable and secure storage services. This storage is based on buckets, which may be private or public. Objects are uploaded to or downloaded from these buckets. There are developer tools available for interfacing with the object storage service.

Object and archive storage have parallels with AWS S3 and Glacier storage services. Object storage provides a scalable, readily available, and accessible storage service. Archive storage is durable but much slower. It is designed for extremely low-cost long-term data retention. It can take a long time to retrieve data from archive storage as the data must first be restored before it can be downloaded. You pay a slightly higher price for object storage where your objects are immediately ready for download. [Figure 1-11](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig11) describes a bucket named Backups created in the standard storage tier (as opposed to the archive storage tier). One object, a file called EXACM.pdf, has been uploaded to this bucket. Examining the details of the object reveals several unusual properties, such as the URL path (URI), <https://objectstorage.us-ashburn-1.oraclecloud.com/n/ocaoci/b/Backups/o/EXACM.pdf>, and the visibility property, which is set to private. Object and file system storage systems differ in several aspects, the primary difference being that there are no directories and subdirectories in object storage systems. Common use cases include storing daily and weekly backups using object storage while historical data that must be retained for compliance purposes is placed in archive storage.

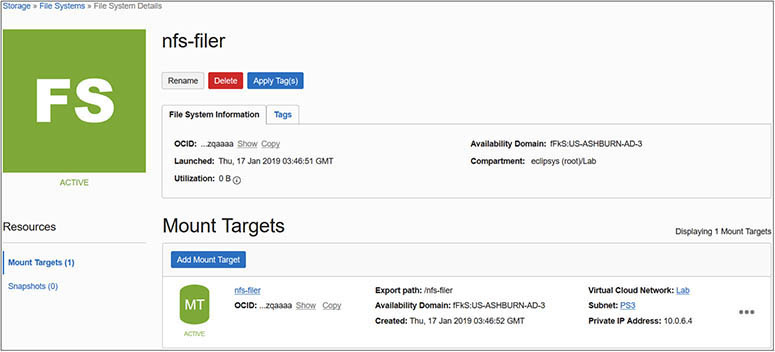


**Figure 1-11**   Object storage bucket

**File Storage Service**

Most enterprise architectures contain one or more Network File Systems (NFS), which allow users access to remote shared storage that is mounted on their computers as if they are local file systems. NFS storage is typically mounted by multiple clients and is commonly used for backups, file sharing, and sometimes even for database storage.

The OCI File Storage Service provides durable NFS storage that you can connect to from any bare metal, virtual machine, or container instance in your virtual cloud network. The nfs-filer file system in [Figure 1-12](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig12) is created in the ASHBURN-AD-1 availability domain and may be mounted by instances using the IP address and path name associated with the NFS mount target.



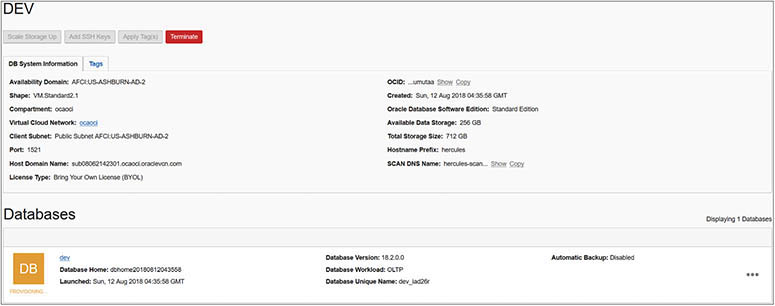
**Figure 1-12**   File storage service

**Database Cloud Service**

Oracle Database Cloud Service (DBCS) is easily the most utilized PaaS offering on OCI. It is enthralling to contemplate that, with a few mouse clicks, you can provision a fully operational Oracle Database in minutes.

When provisioning a DBCS instance, you choose the shape of the compute instance, the database software version, edition, and license type. The compute instances may be virtual machines, bare metal, or Exadata, which is Oracle’s leading-edge engineered system with converged infrastructure. The available database software versions vary and, as of this writing, versions 11.2.0.4, 12c, and 18c are available. DBCS software editions include the on-premises Standard and Enterprise Editions and additionally offer two new editions: Enterprise Edition High Performance and Enterprise Edition Extreme Performance (available only on Exadata).

[Figure 1-13](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1fig13) shows an 18.2 Standard Edition database being provisioned on an instance with compute shape VM.Standard2.1 (1 OCPU, 15GB memory) and 256GB storage.



**Figure 1-13**   Oracle database provisioned by DBCS

You can choose to provision a database with the license included or bring your own license (BYOL). The BYOL model is often used when migrating on-premises database workloads to OCI, allowing existing licenses to be reused. Massive license optimizations are often seen due to innovative features such as capacity on-demand bursting that allows you to license just what you need most of the time and not what you might need at peak load. Unsurprisingly, Oracle makes running its database most cost-effective on its own cloud compared to other cloud vendors.

**Chapter Review**

OCI combines enormous power and a deceptively simple user interface to offer an enterprise-grade IaaS platform. Virtualization, consolidation, and removal of data center management responsibilities are key benefits to creating virtual cloud networks and provisioning cloud architecture as an extension, and ultimately a replacement, for your on-premises workloads.

Periodic hardware refreshes to ensure supportability on current equipment is now the responsibility of the cloud vendor. Another compelling reason to consider running your Oracle database workloads on OCI is the potential cost savings on processor-based licenses. The flexibility of scaling and bursting CPU cores on some OCI-based database cloud services supports CPU sizing for what you need most of the time and not for the worst-case scenario.

High availability and disaster recovery options are provided for by arrangements of fault-tolerant data centers known as *availability domains* clustered together to form a region. Regions are geographically dispersed, some on different continents, and may be linked together to accommodate Maximum Availability Architectures.

Off-box networking is a key differentiator of OCI from most other cloud vendors and promotes predictable performance and uniquely enables OCI to support both VMs and bare-metal compute instances, RAC databases, and Engineered Systems.

An Identity and Access Management framework isolates environments using compartments and uses simple language policy statements to enforce the IAM framework on users and groups.

A full suite of virtualized network entities may be provisioned, including the virtual cloud network, internet and dynamic routing gateways, load balancers, route tables, and security lists.

Compute instances may be provisioned as virtual or bare-metal machines based on a selection of available compute shapes comprising various combinations of OCPUs and memory, operating systems, and even some PaaS options such as DBaaS. Block volume storage may be added to your compute instances and systems may be backed up to network file storage, object storage, or archive storage for longer retention.

OCI offers a continually expanding suite of enterprise-grade IaaS features and many (including this author) believe it is one of the most exciting innovations from Oracle.

**Questions**

[**1**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans1)**.**   Which cloud computing model offers the most control of your environment?

**A.**   PaaS

**B.**   SaaS

**C.**   IaaS

**D.**   DBaaS

[**2**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans2)**.**   What infrastructure resource cannot be provisioned in OCI?

**A.**   NVMe storage

**B.**   Networks

**C.**   CPU and RAM

**D.**   KVM switches

[**3**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans3)**.**   Which of the following statements is false?

**A.**   VM instances may be created based on available compute shapes.

**B.**   Block storage volumes can be added only to bare-metal machines.

**C.**   VM instances may share resources on physical servers with other tenants.

**D.**   Block storage volumes can be added to both virtual and bare-metal machines.

[**4**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans4)**.**   What is the OCI term for a fault-tolerant data center?

**A.**   Availability zone

**B.**   Region

**C.**   Virtual cloud network

**D.**   Availability domain

[**5**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans5)**.**   Which of the following statements is true?

**A.**   An availability domain is a collection of regions.

**B.**   A region is a collection of availability domains.

**C.**   An availability zone is a collection of regions.

**D.**   Two or more regions in a metro area are grouped into an AD.

[**6**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans6)**.**   Network performance is typically measured by which two metrics? (Choose two.)

**A.**   Bandwidth and latency

**B.**   Round Trip Time and latency

**C.**   Throughput and bandwidth

**D.**   Isolation and disaster recovery

[**7**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans7)**.**   Which of the following statements is false?

**A.**   Instances in one region can connect to instances in another region.

**B.**   Availability domains in a region share power and cooling.

**C.**   Each AD is self-contained and highly fault tolerant.

**D.**   Failure of one AD is very unlikely to impact the availability of other ADs in the region.

[**8**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans8)**.**   Choose four storage types available on OCI.

**A.**   Block volumes, object storage, archive storage, SSD

**B.**   Block volumes, file storage, archive storage, SSD

**C.**   Block volumes, object storage, archive storage, file storage

**D.**   Block volumes, buckets, file storage, SSD

[**9**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans9)**.**   A Load Balancing Router routes traffic to backend servers based on which backend set routing algorithms?

**A.**   IP Hash, Weighted Round-Robin

**B.**   IP Hash, Weighted Round-Robin, Least Connections

**C.**   Weighted Round-Robin, Least Connections

**D.**   IP Hash, Weighted Round-Robin, Least Connections, Shortest Path

[**10**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#ch1ans10)**.**   What is the unique, Oracle-assigned identifier for Oracle Cloud Infrastructure resources known as?

**A.**   OCID

**B.**   Tenant ID

**C.**   ACID

**D.**   Cloud account name

**Answers**

[**1**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans1)**.**   **C.** IaaS provides you access to a collection of servers, storage, and network infrastructure onto which you deploy your platform and software. You have complete control of your environment.

[**2**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans2)**.**   **D.** KVM switches are not available or relevant in OCI. OCI allows NVMe storage, networks, CPU, and RAM to be provisioned.

[**3**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans3)**.**   **B.** Block storage volumes can be added to both virtual and bare-metal machines.

[**4**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans4)**.**   **D.** An AD is a fault-tolerant data center.

[**5**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans5)**.**   **B.** A region is a collection of availability domains.

[**6**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans6)**.**   **A.** Bandwidth and latency are key metrics used to measure network performance.

[**7**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans7)**.**   **B.** Availability domains in a region do not share power and cooling.

[**8**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans8)**.**   **C.** Block volumes, object storage, archive storage, and file storage are four storage options available on OCI.

[**9**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans9)**.**   **B.** IP Hash, Weighted Round-Robin, and Least Connections are valid backend set routing algorithms that a Load Balancing Router can use to route traffic to backend servers.

[**10**](https://learning.oreilly.com/library/view/oracle-cloud-infrastructure/9781260452600/ch1.xhtml#r_ch1ans10)**.**   **A.** Each Oracle Cloud Infrastructure resource has a unique, Oracle-assigned identifier called an Oracle Cloud ID (OCID).