AWS:

- Amazon web services (AWS) AWS offers comprehensive cloud IaaS services ranging from virtual
- compute, storage, and networking to complete computing stacks
- AWS is mostly known for its compute and storage-on demand services, namely
- Elastic Compute Cloud (EC2)

  EC2 (Elastic Compute Cloud) provides the virtualized platforms to the host VMs where the cloud application can run.

  It is possible to choose from a large variety of virtual hardware configurations, including GPU and cluster instances.

  S3:

  S3 (Simple Storage Service) provides the object-oriented storage service for users.

  EBS:

  EBS (Elastic Block Service) provides the block storage interface which can be used to support traditional applications.

  SQS:

  SQS stands for Simple Queue Service, and its job is to ensure a reliable message service between two processes. The message can be kept reliably even when the receiver processes are not running.

Users can access their objects through SOAP with either browsers or other client programs which support the SOAP standard.

EC2 also provides the capability to save a specific running instance as an image, thus allowing users to create their own templates for deploying systems.

These templates are stored into S3 that delivers persistent storage on demand.

S3 is organized into buckets;

these are containers of objects that are stored in binary form and can be enriched with attributes.

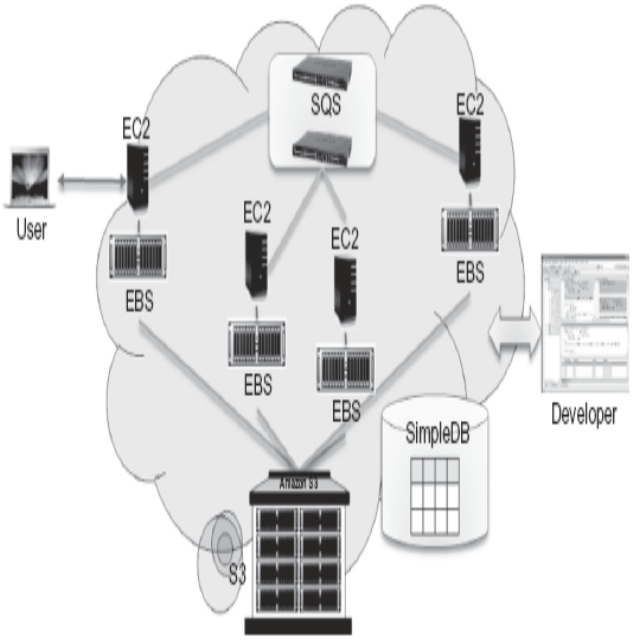

FIGURE 4.21

Amazon cloud computing infrastructure (Key service is identified here; many more are listed in Table 4.5).

(*Courtesy of Kang Chen, Tsinghua University, China*)

Compute services:

- Compute services constitute the fundamental element of cloud computing systems.
- The fundamental service in this space is Amazon EC2, which delivers an IaaS solution and it provides the virtualized platforms to the host VMs where the cloud application can run.
- Amazon EC2 allows deploying servers in the form of virtual machines created as instances(EC2 instances) of a specific image(AMI).
- Images come with a preinstalled operating system and a software stack, and instances can be configured for memory, number of processors, and storage.
- Users are provided with credentials to remotely access the instance and further configure or install software if needed

Amazon machine images:

- Amazon Machine Images (AMIs) are templates from which it is possible to create a virtual machine.
- They are stored in Amazon Simple Storage Service S3
- An AMI contains a physical file system layout with a predefined operating system installed.
- These are specified by the
- Amazon Ramdisk Image (ARI, id: ari-yyyyyy) and
- Amazon Kernel Image (AKI, id: aki-zzzzzz),
- which are part of the configuration of the template.
- AMIs are either created from scratch or “bundled” from existing EC2 instances.

Prepare a new AMIs?

- To create an instance from a preexisting AMI,
- Log into it once it is booted and running, and install all the software needed.
- Using the tools provided by Amazon, we can convert the instance into a new image.
- Once an AMI is created, it is stored in an S3 bucket
- it is also possible to associate a product code with a given AMI, thus allowing the owner of the AMI to get revenue every time this AMI is used to create EC2 instances

EC2 instances:

- EC2 instances represent virtual machines.
- They are created using AMI as templates,
- The processing power is expressed in terms of virtual cores and EC2 Compute Units (ECUs).
- EC2 instances are priced hourly according to the category they belong to.
- EC2 instances can be run either by
- using the command-line tools provided by Amazon or via the AWS console, which allows the management of other services, such as S3.

ECU:

- The ECU is a measure of the computing power of a virtual core.

- By using compute units instead of real frequency values, Amazon can change over time the mapping of such units to the underlying real amount of computing power allocated, thus keeping the performance of EC2 instances consistent with standards set by the times.

Standard instances-

- This class offers a set of configurations that are suitable for most applications.
- EC2 provides three different categories of increasing computing power, storage, and memory.

Micro instances-

- This class is suitable for those applications that consume a limited amount of computing power and memory and occasionally need bursts in CPU cycles to process surges in the workload.
- Micro instances can be used for small Web applications with limited traffic.

High-memory instances

- This class targets applications that need to process huge workloads and require large amounts of memory.
- Three-tier Web applications characterized by high traffic are the target profile.
- Three categories of increasing memory and CPU are available, with memory proportionally larger than computing power.

High-CPU instances-

- This class targets compute-intensive applications.
- Two configurations are available where computing power proportionally increases more than memory.

Cluster Compute instances-

- This class is used to provide virtual cluster services.
- Instances in this category are characterized by high CPU compute power and large memory and an extremely high I/O and network performance, which makes it suitable for HPC applications.

Cluster GPU instances-

- This class provides instances featuring graphic processing units (GPUs) and high compute power, large memory, and extremely high I/O and network performance.
- This class is particularly suited for cluster applications that perform heavy graphic computations, such as rendering clusters.
- This class is suitable for HPC applications.

EC2 environment:

- EC2 instances are executed within a virtual environment. The EC2 environment is in charge of
- allocating addresses,
- attaching storage volumes, and

- configuring security in terms of access control and network connectivity
- By default, instances are created with an internal IP/Elastic IP address, which makes them capable of communicating within the EC2 network and accessing the Internet as clients.
- An Elastic IP address is a static public IPv4 address associated with your AWS account in a specific Region.
- Unlike an auto-assigned public IP address, an Elastic IP address is preserved after you stop and start your instance in a virtual private cloud (VPC).
- EC2 instances are also given a domain name that generally is in the form
- ec2-xxxxxx-xxx.compute-x.amazonaws.com
- Instance owners can partially control where to deploy instances.
- Instance owners can associate a key pair to one or more instances which allows the owner to remotely connect to the instance once this is running and gain root access to it.
- Amazon EC2 controls the accessibility of a virtual instance with basic firewall configuration.Security groups and firewall rules constitute a flexible way of providing basic security for EC2 instances

Amazon elastic MapReduce:

- Amazon Elastic MapReduce provides AWS users with a cloud computing platform for MapReduce applications.

- It utilizes Hadoop as the MapReduce engine, deployed on a virtual infrastructure composed of EC2 instances, and uses Amazon S3 for storage needs.

Storage services AWS:

- S3 is a distributed object store that allows users to store information in different formats.
- The core service in this area is represented by Amazon Simple Storage Service (S3).
- Accessible through a Representational State Transfer (REST) interface
  Features:
- **The storage is organized in a two-level hierarchy:**
  Buckets & Objects
- S3 organizes its storage space into buckets that cannot be further partitioned.
- This means that it is not possible to create directories or other kinds of physical groupings for objects stored in a bucket.
- **Stored objects cannot be manipulated like standard files.**
- S3 has been designed to essentially provide storage for objects that will not change over time.
- Therefore, it does not allow
- renaming, modifying, or relocating an object.
- Once an object has been added to a bucket,
- its content and position is immutable, and
- the only way to change it is to remove the object from the store and add it again.

- **Content is not immediately available to users**
- The main design goal of S3 is to provide an eventually consistent data store.
- S3 uses replication to provide redundancy and efficiently serve objects across the globe; this practice introduces latencies when adding objects to the store—especially large ones—which are not available instantly across the entire globe
- **Requests will occasionally fail.**
- Due to the large distributed infrastructure being managed,
- requests for object may occasionally fail.
- Under certain conditions, S3 can decide to drop a request by returning an internal server error.
- Therefore, it is expected to have a small failure rate during day-to-day operations, which is generally not identified as a persistent failure.

Core Components of S3:

- **Buckets:**
- A bucket is a container of objects and do not support nesting.
- It can be thought of as a virtual drive hosted on the S3 distributed storage.
- Users can select the location at which to create buckets,
- By default buckets are created in Amazon's U.S. datacenters.
- Buckets, objects, and attached metadata are made accessible through a REST interface .Canonical form: http://s3.amazonaws.com/bukect_name/
- Users create a bucket by sending a PUT request to

http://s3.amazonaws.com/bukect_name/

- The content of a bucket can be listed by sending a GET request specifying the name of the bucket.
- Once created, the bucket cannot be renamed or relocated.
- The deletion of a bucket is performed by a DELETE request, which can be successful if and only if the bucket is empty.

Objects:

- Objects constitute the content elements stored in S3.
- An object is identified by a name that needs to be unique within the bucket in which the content is stored.
- The name cannot be longer than 1,024 bytes when encoded in UTF-8, and it allows almost any character.
- Since buckets do not support nesting, even characters normally used as path separators are allowed.
- Users create an object via a PUT request that specifies
- the name of the object together with the bucket name, its contents, and additional properties.
- The maximum size of an object is 5 GB.
- Once an object is created, it cannot be modified, renamed, or moved into another bucket.
- It is possible to retrieve an object via a GET request;
- Deleting an object is performed via a DELETE request.

Metadata:

- Objects can be tagged with metadata, which are passed as properties of the PUT request.
- Meta Data can be retrieved either with a GET request or with a HEAD request, which only returns the object's metadata without the content.

Access control and security of Buckets and Objects:

- Amazon S3 allows controlling the access to buckets and objects by means of Access Control Policies (ACPs).(expressed by means of an XML configuration file).
- Currently, five different permissions can be used:
- READ allows the grantee to retrieve an object and its metadata and to list the content of a bucket as well as getting its metadata.
- WRITE allows the grantee to add an object to a bucket as well remove it.
- READ_ACP allows the grantee to read the ACP of a resource.
- WRITE_ACP allows the grantee to modify the ACP of a resource.
- FULL_CONTROL grants all of the preceding permissions.

Amazon elastic block store:

- The Amazon Elastic Block Store (EBS) allows AWS users to provide EC2 instances with persistent storage.
- They accommodate up to 1 TB of space and are accessed through a block device interface

- The content of an EBS volume survives the instance life cycle and is persisted into S3.
- EBS volumes can be cloned, used as boot partitions, and constitute durable storage.

	S3	EBS
Accessibility	The files within an S3 bucket are stored in an unstructured manner and can be retrieved using HTTP protocols(of its accessibility to HTTP clients) and even with BitTorrent.	EBS is specifically meant for EC2 (Elastic Computing Cloud) instances and is not accessible unless mounted to one.
Storage type	Object storage designed for storing large numbers of user files and backups	Block storage for Amazon EC2 compute instances. It is used for virtualized instances.A volume can be mounted on an EC2 instance and it would appear just like a hard disk partition.
Size of data	The standard limit is of 100 buckets and each bucket has got an unlimited data capacity.	EBS has a standard limit of 20 volumes and each volume can hold data up to 1TB.
Limitation	S3 can have multiple images of its contents so it can be used by many at the same time. Side effect of this is	Volume images cannot be shared among instances, But it is possible to attach multiple volumes to a single

	‘eventual consistency’. With S3, the changes are not written immediately so if you write something, it may not be the data that a read operation returns.	instance. A limitation of EBS is its inability to be used by multiple instances at once. Once it is mounted by an instance, no other instance can use it.
Application	Amazon S3 is a simple storage service offered by Amazon and it is useful for hosting website images and videos, various management features,websites, mobile applications, backup and restore, archive, enterprise applications	Use cases include business continuity, transactional and NO SQL database, software testing, etc.
Security	Amazon S3 can prevent unauthorized accessing of data using its access management tools and encryption policies	No such feature is present in EBS. In EBS, if any user gets unauthorized access to the instance then he/she can easily access the attached EBS.

Amazon ElastiCache:

- ElastiCache is an implementation of an elastic in-memory cache based on a cluster of EC2 instances.

- Automatic patch management and failure detection and recovery of cache nodes allow the cache cluster to keep running without administrative intervention from AWS users.
- An ElastiCache cluster can be dynamically resized according to the demand of the client applications.
- ElastiCache nodes are priced according to the EC2 costing model, with a small price difference due to the use of the caching service installed on such instances.

Structured storage solutions:

- Rely on databases to store data in a structured form.(RDBMS traditionally)
- Amazon provides applications with structured storage services in three different forms:
- **Preconfigured EC2 AMIs:**
- Preconfigured EC2 AMIs are predefined templates featuring an installation of a given database management system.
- EC2 instances created from these AMIs can be completed with an EBS volume for storage persistence.
-
-
- **Amazon Relational Data Storage (RDS):**
- **Amazon SimpleDB:**