

4.2. Programming on Amazon EC2

- ✓ Amazon was the first company to introduce VMs in application hosting. Customers can rent VMs instead of physical machines to run their own applications. By using VMs, customers can load any software of their choice.
- ✓ The elastic feature of such a service is that a customer can create, launch, and terminate server instances as needed, paying by the hour for active servers. Amazon provides several types of preinstalled VMs.
- ✓ Instances are often called Amazon Machine Images (AMIs) which are preconfigured with operating systems based on Linux or Windows, and additional software.

his sequence is supported by public, private, and paid AMIs.

Image Type	AMI Definition
Private AMI	Images created by you, which are private by default. You can grant access to other users to launch your private images.
Public AMI	Images created by users and released to the AWS community, so anyone can launch instances based on them and use them any way they like. AWS lists all public images at http://developer.amazonwebservices.com/connect/kbcategory.jspa?categoryID=171 .
Paid QAMI	You can create images providing specific functions that can be launched by anyone willing to pay you per each hour of usage on top of Amazon's charges.

- ✓ Standard instances are well suited for most applications.

- ✓ Micro instances provide a small number of consistent CPU resources and allow you to burst CPU capacity when additional cycles are available. They are well suited for lower throughput applications and web sites that consume significant compute cycles periodically.
- ✓ High-memory instances offer large memory sizes for high-throughput applications, including database and memory caching applications.
- ✓ High-CPU instances have proportionally more CPU resources than memory (RAM) and are well suited for compute-intensive applications.
- ✓ Cluster compute instances provide proportionally high CPU resources with increased

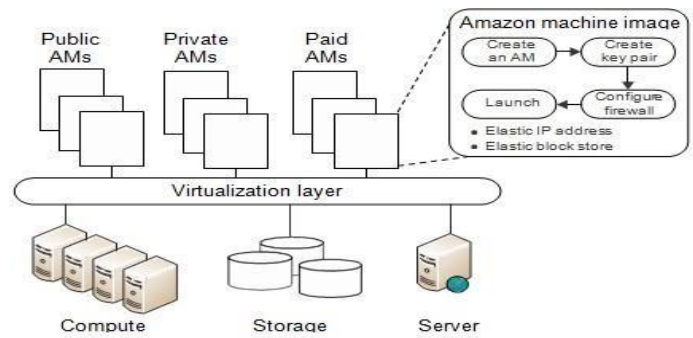


FIGURE 6.23 Amazon EC2 execution environment.

network performance and are well suited for high-performance computing (HPC) applications and other demanding network-bound applications. They use 10 Gigabit Ethernet interconnections.

Table 6.13 Instance Types Available on Amazon EC2 (October 6, 2010)

Compute Instance	Memory GB	ECU or EC2 Units	Virtual Cores	Storage GB	32/64 Bit
Standard: small	1.7	1	1	160	32
Standard: large	7.5	4	2	850	64
Standard: extra large	15	8	4	1690	64
Micro	0.613	Up to 2		Only EBS	32 or 64
High-memory	17.1	6.5	2	420	64
High-memory: double	34.2	13	4	850	64
High-memory: quadruple	68.4	26	8	1690	64
High-CPU: medium	1.7	5	2	350	32
High-CPU: extra large	7	20	8	1690	64
Cluster compute	23	33.5	8	1690	64

Amazon Elastic Block Store (EBS) and SimpleDB

- ✓ The Elastic Block Store (EBS) provides the volume block interface for saving and restoring the virtual images of EC2 instances.
- ✓ Traditional EC2 instances will be destroyed after use. The status of EC2 can now be saved in the EBS system after the machine is shut down. Users can use EBS to save persistent data and mount to the running instances of EC2.
- ✓ Note that S3 is “Storage as a Service” with a messaging interface. EBS is analogous to a distributed file system accessed by traditional OS disk access mechanisms. EBS allows you to create storage volumes from 1 GB to 1 TB that can be mounted as EC2 instances.
- ✓ Multiple volumes can be mounted to the same instance. These storage volumes behave like raw, unformatted block devices, with user-supplied device names and a block device interface.
- ✓ You can create a file system on top of Amazon EBS volumes, or use them in any other way you would use a block device (like a hard drive). Snapshots are provided so that the data can be saved incrementally. This can improve performance when saving and restoring data. In terms of pricing, Amazon provides a similar pay-per-use schema as EC2 and S3.

Amazon SimpleDB Service

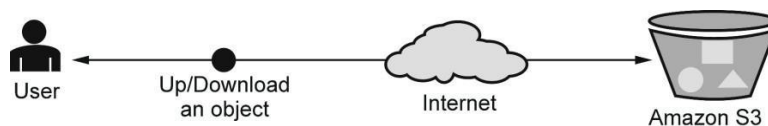
- ✓ SimpleDB provides a simplified data model based on the relational database data model. Structured data from users must be organized into domains. Each domain can be considered a table. The items are the rows in the table.
- ✓ A cell in the table is recognized as the value for a specific attribute (column name) of the corresponding row. This is similar to a table in a relational database. However, it is possible to assign multiple values to a single cell in the table.

- ✓ This is not permitted in a traditional relational database which wants to maintain data consistency.
- ✓ Many developers simply want to quickly store, access, and query the stored data. SimpleDB removes the requirement to maintain database schemas with strong consistency.
- ✓ SimpleDB is priced at \$0.140 per Amazon SimpleDB Machine Hour consumed with the first 25 Amazon SimpleDB Machine Hours consumed per month free (as of October 6, 2010).
- ✓ SimpleDB, like Azure Table, could be called “LittleTable,” as they are aimed at managing small amounts of information stored in a distributed table; one could say BigTable is aimed at basic big data, whereas LittleTable is aimed at metadata.

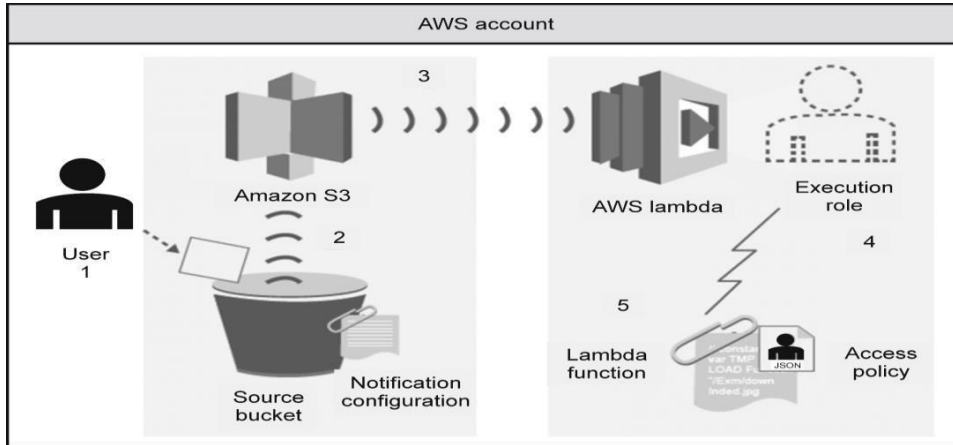
Simple Storage Service (S3)

Amazon S3 offers a simple web services interface that can be used to store and retrieve any amount of data from anywhere, at any time on the web. It gives any developer access to the same scalable, secure, fast, low - cost data storage infrastructure that Amazon uses to operate its own global website network. S3 is an online backup and storage system. The high - speed data transfer feature known as AWS Import/Export will exchange data to and from AWS using Amazon’s own internal network to another portable device.

Amazon S3 is a cloud - based storage system that allows storage of data objects in the range of 1 byte up to 5 GB in a flat namespace. The storage containers in S3 have predefined buckets, and buckets serve the function of a directory, though there is no object hierarchy to a bucket, and the user can save objects to it but not files. Here it is important to note that the concept of a file system is not associated with S3 because file systems are not supported, only objects are stored. In addition to this, the user is not required to mount a bucket, as opposed to a file system. Fig. 4.7 shows an S3 diagrammatically.



S3 system allows buckets to be named (Fig. 4.8), but the name must be unique in the S3 namespace across all consumers of AWS. The bucket can be accessed through the S3 web API (with SOAP or REST), which is similar to a normal disk storage system.



Source bucket

The performance of S3 is limited for use with non-operational functions such as data archiving, retrieval and disk backup. The REST API is more preferred to SOAP API because it is easy to work with large binary objects in REST.

Amazon S3 offers large volumes of reliable storage with high protection and low bandwidth access. S3 is most ideal for applications that need storage archives. For example, S3 is used by large storage sites that share photos and images.

The APIs to manage the bucket has the following features:

- Create new, modify or delete existing buckets.
- Upload or download new objects to a bucket.
- Search and identify objects in buckets.
- Identify metadata associated with objects and buckets.
- Specify where the bucket is stored.
- Provide public access to buckets and objects.

The S3 service can be used by many users as a backup component in a 3-2-1 backup method. This implies that your original data is 1, a copy of your data is 2 and an off-site copy of data is 3. In this method, S3 is the 3rd level of backup. In addition to this, Amazon S3 provides the feature of versioning.

In versioning, every version of the object stored in an S3 bucket is retained, but for this, the user must enable the versioning feature. Any HTTP or REST operation, namely PUT, POST, COPY or DELETE will create a new object that is stored along with the older version. A GET operation retrieves the new version of the object, but the ability to recover and undo actions are also available.

Versioning is a useful method for reserving and archiving data.

Amazon Glacier

Amazon glacier is very low - price online file storage web service which offer secure, flexible and durable storage for online data backup and archiving. This web service is specially designed for those data which are not accessed frequently. That data which is allowed to be retrieved within three to five hours can use amazon glacier service.

You can virtually store any type of data, any format of data and any amount of data using amazon glacier. The file in ZIP and TAR format are the most common type of data stored in amazon glacier.

Some of the common use of amazon glacier are:

- Replacing the traditional tape solutions with backup and archive which can last longer.
- Storing data which is used for the purposes of compliance.

Glacier Vs S3

Both amazon S3 and amazon glacier work almost the same way. However, there are certain important aspects that can reflect the difference between them. Table 6.10.1 shows the comparison of amazon glacier and amazon S3:

Amazon Glacier	Amazon S3
It supports 40 TB archives	It supports 5 TB objects
It is recognized by archive IDs which are system generated	It can use “friendly” key names

It encrypts the archives automatically	It is optional to encrypt the data
It is extremely low - cost storage	Its cost is much higher than Amazon Glacier

Amazon Glacier Vs Amazon S3

You can also use amazon S3 interface for availing the offerings of amazon glacier with no need of learning a new interface. This can be done by utilizing Glacier as S3 storage class along with object lifecycle policies.

