

2.1.4 Saving persistent data in cloud storage

You store your luggage in the trunk of your car. Similarly, the cloud provides a place to store your machine images, your applications, and any data your applications need.

Cloud storage has also increased in popularity recently for many of the same reasons as cloud computing. Cloud storage delivers *virtualized storage on demand* over a network based on a request for a given quality of service (QoS). Unlike the long provisioning lead times required in corporate data centers, there is no need to purchase storage or in some cases even provision it before storing data. Typically, you pay for transit of data into the cloud and, subsequently, a recurring fee based on the amount of storage consumption your data uses.

You can use cloud storage in many different ways. For example, you can back up local data (such as on a laptop) to cloud storage; a virtual disk can be synched to the cloud and distributed to other computers; and you can use it as an archive to retain (under some policy) data for regulatory or other purposes.

You can use cloud storage for applications that provide data directly to their clients via the network. The application redirects the client to a location at the cloud storage provider for the data. Media such as audio and video files are examples. The network requirements for streaming data files can be made to scale in order to meet the demand without affecting the application.

The type of interface used for this is HTTP. You can fetch the file from a browser without having to do any special coding, and the correct application is invoked automatically. But how do you get the file there in the first place, and how do you make sure the storage you use is of the right type and QoS? Again, many offerings expose an interface for these operations, and it's not surprising that many of these interfaces use REST principles. This is typically a data-object interface with operations for creating, reading, updating, and deleting the individual data objects via HTTP operations.

Keeping with Amazon's APIs as good examples to study, we've outlined a simple API dealing with Amazon's S3 in table 2.3.

A cloud storage standard

The Storage Networking Industry Association has created a technical work group to address the need for a cloud storage standard. The new Cloud Data Management Interface (CDMI) enables interoperable cloud storage and data management. In CDMI, the underlying storage space exposed by the interfaces is abstracted using the notion of a *container*. A container is not only a useful abstraction for storage space, but also serves as a grouping of the data stored in it and a point of control for applying data services in the aggregate.

Table 2.3 Basic terms and operations of Amazon S3

Terms	Description
Object	Fundamental entity stored in S3. Each object can range in size from 1 byte to 5 GB. Each object has object data and metadata. <i>Metadata</i> is a set of name-value pairs that describe the data.
Bucket	Fundamental container in S3 for data storage. Objects are uploaded into buckets. There is no limit to the number of objects you can store in a bucket. The bucket provides a unique namespace for the management of objects contained in the bucket. Bucket names are global, so each developer can own only up to 100 buckets at a time.
Key	A key is the unique identifier for an object within a bucket. A bucket name plus a key uniquely identifies an object within all of S3.
Usage	<ol style="list-style-type: none"><li data-bbox="464 560 921 587">1. Create a bucket in which to store your data.<li data-bbox="464 596 921 622">2. Upload (write) data (objects) into the bucket.<li data-bbox="464 631 949 657">3. Download (read) the data stored in the bucket.<li data-bbox="464 666 878 693">4. Delete some data stored in the bucket.<li data-bbox="464 702 792 728">5. List the objects in the bucket.

In many cases, the coarse granularity and unstructured nature of cloud storage services such as S3 aren't sufficient for the type of data access required. For many applications, an alternative structured data-storage method is required. Let's explore how databases in the cloud work (and don't).