The technological capacity to store information has grown over time at an accelerated pace [164,178]:

• 1986: 2.6 EB; equivalent to less than one 730 MB CD-ROM of data per computer user.

• 1993: 15.8 EB; equivalent to four CD-ROMs per user.

• 2000: 54.5 EB; equivalent to 12 CD-ROMs per user.

• 2007: 295 EB; equivalent to almost 61 CD-ROMs per user.

Though it pales in comparison with processor technology, the evolution of storage technology is

astounding. A 2003 study [251] shows that during the 1980–2003 period the storage density of hard

disk drives (HDD) increased by four orders of magnitude, from about 0.01 Gb/in2 to about 100 Gb/in2.

During the same period the prices fell by five orders ofmagnitude, to about 1 cent/Mbyte.HDDdensities

are projected to climb to 1,800 Gb/in2 by 2016, up from 744 Gb/in2 in 2011.

The density of Dynamic Random Access Memory (DRAM ) increased from about 1 Gb/in2 in 1990

to 100 Gb/in2 in 2003. The cost of DRAM tumbled from about $80/MB to less than $1/MB during the

same period. In 2010 Samsung announced the first monolithic, 4 gigabit, low-power, double-data-rate

(LPDDR2) DRAM using a 30 nm process.

These rapid technological advancements have changed the balance between initial investment in

storage devices and system management costs. Now the cost of storage management is the dominant

element of the total cost of a storage system. This effect favors the centralized storage strategy supported

by a cloud; indeed, a centralized approach can automate some of the storage management functions,

such as replication and backup, and thus reduce substantially the storage management cost.

While the density of storage devices has increased and the cost has decreased dramatically, the

access time has improved only slightly. The performance of I/O subsystems has not kept pace with the

performance of computing engines, and that affects multimedia, scientific and engineering, and other

modern applications that process increasingly large volumes of data.

The storage systems face substantial pressure because the volume of data generated has increased

exponentially during the past few decades; whereas in the 1980s and 1990s data was primarily generated

by humans, nowadays machines generate data at an unprecedented rate. Mobile devices, such

as smart-phones and tablets, record static images, as well as movies and have limited local storage

capacity, so they transfer the data to cloud storage systems. Sensors, surveillance cameras, and digital

medical imaging devices generate data at a high rate and dump it onto storage systems accessible via

the Internet. Online digital libraries, ebooks, and digital media, along with reference data,2 add to the

demand for massive amounts of storage.

As the volume of data increases, new methods and algorithms for data mining that require powerful

computing systems have been developed. Only a concentration of resources could provide the CPU cycles along with the vast storage capacity necessary to perform such intensive computations and

access the very large volume of data.

Although we emphasize the advantages of a concentration of resources, we have to be acutely aware

that a cloud is a large-scale distributed system with a very large number of components that must

work in concert. The management of such a large collection of systems poses significant challenges

and requires novel approaches to systems design. Case in point: Although the early distributed file

systems used custom-designed reliable components, nowadays large-scale systems are built with offthe-

shelf components. The emphasis of the design philosophy has shifted from performance at any

cost to reliability at the lowest possible cost. This shift is evident in the evolution of ideas, from the

early distributed file systems of the 1980s, such as the Network File System (NFS) and the Andrew File

System (AFS), to today’s Google File System (GFS) and the Megastore.