The development of efficient cloud applications inherits the challenges posed by the natural imbalance

among computing, I/O, and communication bandwidths of physical systems. These challenges

are greatly amplified due to the scale of the system, its distributed nature, and the fact that virtually

all applications are data-intensive. Though cloud computing infrastructures attempt to automatically

distribute and balance a load, the application developer is still left with the responsibility of placing the

data close to the processing site and identifying optimal storage for the data.

One of the main advantages of cloud computing, the shared infrastructure, could also have a negative

impact. Performance isolation1 is nearly impossible to reach in a real system, especially when the system

is heavily loaded. The performance of virtual machines fluctuates based on the load, the infrastructure

services, and the environment, including the other users. Security isolation is also challenging on multitenant

systems.

Reliability is also a major concern; node failures are to be expected whenever a large number of

nodes cooperate for the computations. Choosing an optimal instance (in terms of performance isolation,

reliability, and security) from those offered by the cloud infrastructure is another critical factor to be

considered. Of course, cost considerations also play a role in the choice of the instance type.

Many applications consist of multiple stages; in turn, each stage may involve multiple instances

running in parallel on the systems of the cloud and communicating among them. Thus, efficiency,

consistency, and communication scalability are major concerns for an application developer. Indeed, due to shared networks and unknown topology, cloud infrastructures exhibit internode latency and

bandwidth fluctuations that affect application performance.

Data storage plays a critical role in the performance of any data-intensive application; the organization

of the storage, the storage location, and the storage bandwidth must be carefully analyzed to lead to

optimal application performance. Clouds support many storage options to set up a file system similar

to the Hadoop file system discussed in Section 8.6; among them are off-instance cloud storage (e.g.,

S3), mountable off-instance block storage (e.g., EBS), and storage persistent for the lifetime of the

instance.

Many data-intensive applications use metadata associated with individual data records; for example,

the metadata for an MPEG audio file may include the name of the song, the singer, recording information,

and so on. Metadata should be stored for easy access, and the storage should be scalable and

reliable.

Another important consideration for the application developer is logging. Performance considerations

limit the amount of data logging, whereas the ability to identify the source of unexpected results

and errors is helped by frequent logging. Logging is typically done using instance storage preserved only

for the lifetime of the instance. Thus, measures to preserve the logs for a postmortem analysis must be

taken. Another challenge awaiting resolution is related to software licensing, discussed in Section 3.11.