AVAILABILITY OF DISK SUBSYSTEMS

Disk subsystems are assembled from standard components, which have a limited fault-tolerance. In this chapter we have shown how these standard components are combined in order to achieve a level of fault-tolerance for the entire disk subsystem that lies significantly above the fault-tolerance of the individual components. Today, disk subsystems can be constructed so that they can withstand the failure of any component without data being lost or becoming inaccessible. We can also say that such disk subsystems have no ‘single point of failure’.

The following list describes the individual measures that can be taken to increase the availability of data:

• The data is distributed over several hard disks using RAID processes and supplemented by further data for error correction. After the failure of a physical hard disk, the data of the defective hard disk can be reconstructed from the remaining data and the additional data.

• Individual hard disks store the data using the so-called Hamming code. The Hamming code allows data to be correctly restored even if individual bits are changed on the hard disk. Self-diagnosis functions in the disk controller continuously monitor the rate of bit errors and the physical variables (e.g., temperature, spindle vibration). In the event of an increase in the error rate, hard disks can be replaced before data is lost.

• Each internal physical hard disk can be connected to the controller via two internal I/O channels. If one of the two channels fails, the other can still be used.

• The controller in the disk subsystem can be realised by several controller instances. If one of the controller instances fails, one of the remaining instances takes over the tasks of the defective instance.

• Other auxiliary components such as power supplies, batteries and fans can often be duplicated so that the failure of one of the components is unimportant. When connecting the power supply it should be ensured that the various power cables are at least connected through

various fuses. Ideally, the individual power cables would be supplied via different external power networks; however, in practice this is seldom realisable.

• Server and disk subsystem are connected together via several I/O channels. If one of the channels fails, the remaining ones can still be used.

• Instant copies can be used to protect against logical errors. For example, it would be possible to create an instant copy of a database every hour. If a table is ‘accidentally’ deleted, then the database could revert to the last instant copy in which the database is still complete.

• Remote mirroring protects against physical damage. If, for whatever reason, the original data can no longer be accessed, operation can continue using the data copy that was generated using remote mirroring.

• Consistency groups and write-order consistency synchronise the copying of multiple virtual hard disks. This means that instant copy and remote mirroring can even guarantee the consistency of the copies if the data spans multiple virtual hard disks or even multiple disk subsystems.

• LUN masking limits the visibility of virtual hard disks. This prevents data being changed or deleted unintentionally by other servers.

This list shows that disk subsystems can guarantee the availability of data to a very high degree. Despite everything it is in practice sometimes necessary to shut down and switch off a disk subsystem. In such cases, it can be very tiresome to co-ordinate all project groups to a common maintenance window, especially if these are distributed over different time zones.

Further important factors for the availability of an entire IT system are the availability of the applications or the application server itself and the availability of the connection between application servers and disk subsystems. Chapter 6 shows how multipathing can improve the connection between servers and storage systems and how clustering can increase the fault-tolerance of applications.