

## 12.3 Fault Tolerance

At some point, all equipment fails which means you must be tolerant towards fault. Backup is the most obvious way of having fault tolerance for a server. If the server fails, did you back up the data so you can restore it? Although database administrators might use a number of different types of data backups, from a security point of view the three-primary backup types are:

- Full:** All changes
- Differential:** All changes since last full backup
- Incremental:** All changes since last backup of any type

Consider a scenario where you carry out a full back up at 2 a.m. each morning. However, you are concerned about the possibility of a server crash before the next full backup. Therefore, you want to complete a backup every two hours. The type of backup you choose will determine the efficiency of completing these frequent backups and the time needed to restore. Let us consider each type of backup in a crash scenario and what would happen if the system crashes at 10:05 a.m.

●**Full:** In this scenario you complete a full back up at 4 a.m., 6 a.m., 10 a.m., and then the system crashes. All you would have to do is restore the last full backup, which was done at 10 a.m. This makes restoration much simpler. However, running a full back up every 2 hours is very time consuming and resource intensive and will have a significant negative impact on your server's performance.

●**Differential:** In this scenario you would carry out a differential backup at 4 a.m., 6 a.m., 10 a.m., and then the system crashes. You need to restore the last full backup done at 2 a.m., and the most recent differential backup done at 10 a.m. This is a little more complicated than the full backup strategy. However, those differential backups are going to get larger each time you do them, and thus more time consuming and resource intensive. Although they will not have the same impact as doing full backups, they will still slow down your network.

●**Incremental:** In this scenario you do an incremental backup at 4 a.m., 6 a.m., ...10 a.m., and then the system crashes. You need to restore the last full backup done at 2 a.m., and then each

incremental backup done since then, and they must be restored in order. This is a much more complex restore, but each incremental backup is small and does not take much time nor consume many resources.

There is no “best” backup strategy. Which strategy you select will depend on your organisation’s needs. Whatever backup strategy you choose, you must periodically test it. The only effective way to test your backup strategy is to actually restore the backup data to a test machine.

The other fundamental aspect of fault tolerance is RAID, or redundant array of independent disks. RAID allows your servers to have more than one hard drive, so that if the main hard drive fails, the system keeps functioning. The primary RAID levels are described here:

- RAID 0** (striped disks) distributes data across multiple disks in a way that gives improved speed at any given instant. This offers NO fault tolerance.

- RAID 1** mirrors the contents of the disks, making a form of 1:1 ratio real-time backup. This is also called mirroring.

- RAID 3 or 4** (striped disks with dedicated parity) combines three or more disks in a way that protects data against loss of any one disk. Fault tolerance is achieved by adding an extra disk to the array and dedicating it to storing parity information. The storage capacity of the array is reduced by one disk.

- RAID 5** (striped disks with distributed parity) combines three or more disks in a way that protects data against the loss of any one disk. It is similar to RAID 3 but the parity is not stored on one dedicated drive; instead parity information is interspersed across the drive array. The storage capacity of the array is a function of the number of drives minus the space needed to store parity.

- RAID 6** (striped disks with dual parity) combines four or more disks in a way that protects data against loss of any two disks.

- RAID 1+0** (or 10) is a mirrored data set (RAID 1) that is then striped (RAID 0), hence the “1+0” name. A RAID 1+0 array requires a minimum of four drives: two mirrored drives to hold half of the striped data, plus another two mirrored for the other half of the data.