Backup and Recovery

Backup

- A data warehouse is a complex system and it contains a huge volume of data. Therefore it is important to back up all the data so that it becomes available for recovery in future as per requirement. In this chapter, we will discuss the issues in designing the backup strategy.

Backup Terminologies

- Before proceeding further, you should know some of the backup terminologies discussed below.
- **Complete backup** − It backs up the entire database at the same time. This backup includes all the database files, control files, and journal files.
- **Partial backup** − As the name suggests, it does not create a complete backup of the database. Partial backup is very useful in large databases because they allow a strategy whereby various parts of the database are backed up in a round-robin fashion on a day-to-day basis, so that the whole database is backed up effectively once a week.

- **Cold backup** − Cold backup is taken while the database is completely shut down. In multi-instance environment, all the instances should be shut down.
- **Hot backup** − Hot backup is taken when the database engine is up and running. The requirements of hot backup varies from RDBMS to RDBMS.
- **Online backup** − It is quite similar to hot backup.

Hardware Backup

- It is important to decide which hardware to use for the backup.
- The speed of processing the backup and restore depends on the hardware being used, how the hardware is connected, bandwidth of the network, backup software, and the speed of server's I/O system.
- Tape Technology
  - Disk Backups
  - Tape Technology
- The tape choice can be categorized as follows −
  - Tape media
  - Standalone tape drives
  - Tape stackers
  - Tape silos

Disk Backups

- Methods of disk backups are −
  - Disk-to-disk backups
  - Mirror breaking

Disk-to-Disk Backups

- Here backup is taken on the disk rather on the tape. Disk-to-disk backups are done for the following reasons −
  - Speed of initial backups
  - Speed of restore

Mirror Breaking

- The idea is to have disks mirrored for resilience during the working day. When backup is required, one of the mirror sets can be broken out. This technique is a variant of disk-to-disk backups.

Software Backups

- There are software tools available that help in the backup process. These software tools come as a package.
- These tools not only take backup, they can effectively manage and control the backup strategies.

- The criteria for choosing the best software package are listed below
  - How scalable is the product as tape drives are added?
  - Does the package have client-server option, or must it run on the database server itself?
  - Will it work in cluster and MPP environments?
  - What degree of parallelism is required?
  - What platforms are supported by the package?
  - Does the package support easy access to information about tape contents?
  - Is the package database aware?
  - What tape drive and tape media are supported by the package?