Focus On Quality

Validation of Spectrometry Software

Part IX of the series discusses backup and recovery, which are essential to the process of ensuring data security and integrity.

R.D. McDowall

When was the last time you backed up your computer? The road to hell is paved with good intentions: "I should have backed up my data, but I didn't." This installment of Focus on Quality covers an essential part of the process of ensuring data security and integrity — backup and recovery. This topic deals with the first stages of the disaster recovery process: the use of tape storage for data files, and system and application software. This is not to be confused with the long-term archive, storage and restoration of data on more permanent media such as CD-ROM or DVD. The organization of data usually is different in the two cases; backup deals with all data and application programs on a disk while archive covers data organized around specific work packages.

The importance of the data stored on the computerized system or network determines the frequency and extent of your backup and recovery procedure. All of the following issues need to be considered when designing your backup and recovery process:

- How critical are the data? For critical data, the intervals between backups and the type of backups performed will be higher than in low-priority systems where backups can be made less frequently.
- How often do the data change or are new data acquired? Spectrometry systems that acquire new files regularly or that involve extensive data manipulation will need more frequent backup than systems, which change less frequently.
- What speed of recovery is required? Can the system be restored within a working day with little impact or does it need to be restored within four hours? This will affect both the frequency and nature of the backup and recovery schemes as well as the linkage with database transaction logging.

R.D. McDowall
is a visiting senior lecturer in the Department of Chemistry at the University of Surrey, principal of McDowall Consulting (Bromley, UK), and "Questions of Quality" column editor for *LCGC Europe, Spectroscopy's* sister magazine. Address correspondence to him at 73 Murray Avenue, Bromley, Kent, BR1 3DJ, UK.

Defining Backup and Recovery

Observations have been made during inspections such as:

"Failure to comply with network system backup procedures in that not all required backup procedures were documented as scheduled, showing lack of documented evidence that tape replacements were done;" and

"There were no daily incremental backups; backups were performed every two weeks. If the system fails, data acquired between backups will be lost. The firm did not have contingency plans in case of a system failure."

Backup and recovery is focused on storing and restoring system, application and user files. Storing means the copying from a source device, such as a disk on your workstation or network disk drive, to a target medium — usually a magnetic tape. Restoring means copying from media containing stored files to the primary location of the files — usually high-speed disk.

There are two main roles involved with backup and recovery in a client-server environment in most organizations — that of end users and IT personnel.

End-users are responsible for backup and recovery, because it is their system and their data; these facts often are overlooked when a backup schedule is developed. Responsibility often can be abrogated and a default schedule given that is not appropriate to the system or the data held on it. Users must be aware of their responsibilities in this area.

The IT department usually will carry out the backup and recovery of data. The schedule for the backups will be worked out in discussions with the end users and this should be (but rarely is) recorded in a service level agreement (SLA) that outlines the roles and responsibilities of all parties and the schedule. In addition, a standard operating procedure (SOP) will outline the schedule of backups, and electronic records will be generated during the backup process and any recoveries performed; these must be maintained for a validated application. For standalone systems, the user also assumes the role of the IT department.

Hardware For Data Security and Integrity

In many cases, data and application files are stored on a single computer disk; this means that there is a single point of failure that could mean the loss of data files. This is why you also should consider the use of hardware options for improv-

ing data integrity and fault tolerance within the system. Usually, these hardware options are grouped under the acronym, RAID (Redundant Array of Inexpensive Disks). Three options are available to computers and servers commonly used to hold regulatory data:

RAID Level 0: Data striping. This involves two separate drives where any data written to disk are broken into data blocks called stripes, which are written in sequence to both drives. The advantage of this configuration is speed, but the disadvantage is that if a hard disk fails the strip set will be lost and the data will have to be restored from backup tape. Apart from the speed gains, there are few other advantages of RAID 0. Thus, for data security and integrity, one of the other two options should be selected.

RAID Level 1: Disk mirroring. Data are written to two drives that are configured identically. The difference between RAID 0 and 1 is that when a RAID 1 drive fails, the other drive contains an exact copy of the data and can be used immediately (see Figure 1). However, to replace the defective drive the computer must be shut down. There is a single point of failure as there usually is a single disk controller for the two disk drives; however, if two controllers are used then this is termed disk duplexing. One disadvantage of RAID 1 (and that is exacerbated in RAID 5) is that 10 Mb of data requires 20 Mb of disk space; however, given the cost of disk drives versus the value of the data stored on them, this usually is a minor problem.

RAID Level 5: Fault tolerance, achieved by disk striping with parity. This essentially is an extension of RAID 1. If a single drive fails then the data on it can be recovered from the other two by using the parity checksum information. However, if two disks fail then it's over to the tape backups to recover the data. To implement RAID 5, three identical drives usually are needed and the operating system is set up to manage the set. In the normal operation of the computer, data will be copied across the disks with the parity checksums. If a disk fails, some vendors offer a hot swap option where the old disk can be replaced with a new (but empty) disk and the data on the failed disk reconstructed using the parity checksums on the remaining two disks.

Normally RAID 1 or 5 would be used to store data; however, if your data are very critical, then you must use RAID 5 with the fault tolerance aspects to reduce data loss. If the computer application needs to operate with greater than 99% of uptime, then you should consider additional hardware features such as dual processors and uninterruptible power supplies (UPS).

Options to Consider for Backup

There are a number of options that can be used in developing a backup strategy in addition to using hardware to mitigate the effects of a disk crash.

Full backup. A regular backup of the system and complete copy of all system, application and user files on all disks to tape.

Incremental backup. A regular, but partial copy of system, application and user files, identified by backup profile, to tape.

Differential backup. A regular, but partial copy of files that have changed since the last normal backup.

Special backup. Specifically requested copy of explicitly identified files to tape.

Most readers will be aware of the nature of a full backup but will be less aware of what the differences are between incremental and differential backups. Let's look at a typical example of a backup:

- A full system backup will be done on Friday evening or during the weekend when there are no users on the system. This will include all data and also can include the application and operating system, although the latter two usually are performed on a less frequent schedule, because there is less change. The operating system and application software can be separated physically from the data using separate disk drives.
- During the week, incremental backups will be made on Monday, Tuesday, Wednesday and Thursday.
- Each incremental backup will contain the files that have changed since the last incremental or full backup.

Assume a system disk fails on Friday afternoon before the next full backup is scheduled. To recover the disk, the last full backup is recovered and then the successive incremental tapes are recovered. Thus to recover the disk back to Thursday evening, the full backup and four incremental backups need to be installed. A failure in one of the early incremental tapes will result in lost data even if the later backup tapes are perfect, amplifying the impact of any data loss.

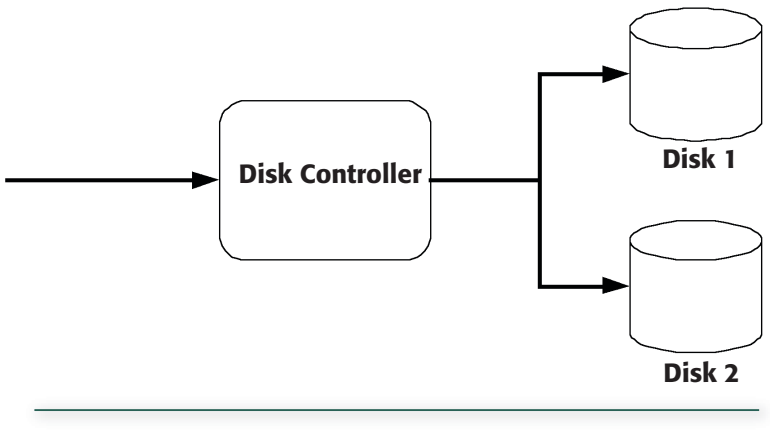

Figure 1. Level 1 RAID hardware. When a RAID 1 drive fails, the other drive contains an exact copy of the data and can be used immediately.

The differential backup, by contrast, contains all changes since the last full backup on the previous Friday night. Thus, Thursday's tape will contain all changes in files since last Friday's full backup. After disk failure it only requires the full backup and the latest differential tape to bring the system back to Thursday evening. The differential tape backup will grow in size over a week, in contrast to an incremental tape that is relatively small in size. The greatest advantage is that only two tapes are required to recover data from a differential backup as opposed to a maximum of five with an incremental backup.

Fundamental Backup Activities

The first part of the process is to schedule backups for the spectrometer system. How vital are the data stored on the computer? Only the users can tell you. If you work in IT, don't anticipate what the users might think or say — get them to commit to the backup schedule themselves. The user department usually will be paying, as is the case in most pharmaceutical companies today: the cost of IT services are recovered by cross charging for services provided to an individual system.

Because protection of electronic records are mandated by 21 CFR 11 (3), the data generated by the spectrometer is backed up as the first stage of protection. Criticality of the data will dictate the type of backup and the frequency so that any potential data loss is minimized.

There are two basic approaches to backup of computerized systems:

- **Hot or on-line backup** takes place while the system is still operating.
- **Cold or off-line backup** occurs when the system has been stopped and users have been logged off the system.

The cold backup generally is thought of as the safest type of backup because hot backup requires the system to be buffered while the backup occurs and the system updated when complete. The option you select should depend upon the use of the system and the value of data.

For instance, if you have a data system that is required to be available 24 hours per day, seven days per week, then a hot backup of the system would be required. Alternatively, if the system only was required to be available 95% of the time, then a cold or off-line backup approach could be devised where the system would be backed up when there were no users on the system.

A cold backup is scheduled out of working hours and is performed automatically with the logs of the activity reviewed the next morning to confirm that all has gone well. The process is as follows:

Remove Users from System. If you need to backup the system during normal working hours, warn the users of impending downtime, then ensure that all users are logged off and have their files closed. The system manager can disable further user access and make sure the appropriate processes for backup are running.

For normal hours or off-hours backup the common process is:

- **Copy Files to Media.** Using a software tool provided explicitly for the purpose of backups (either with the operating system or purchased specifically for the purpose), files are copied from their primary location on disk to the backup media, usually tape. Typical backup applications are Backup-Exec, Networker, or ArcServe.
- **Verify Readability of Backed up Files.** You are placing your trust and your data in the hands of a magnetic medium; therefore the quality of the backup is essential. Verification confirms that the files have been copied to the tape correctly and that the tape can be read again. Verifying readability of files backed up can include a comparison of the files on the backup media with the originals on disk. This gives the best confirmation, but can be time consuming. For offline backups this step usually occurs before users are allowed access to their applications following the backup.
- **Allow User Access To System.** If the backup is verified as readable, the system can be returned to normal operations.

What happens if a backup fails?

Circle XXX

Unsuccessful backups either must be rescheduled for the next backup window or a backup is performed as soon as the problem is known. However, a cold backup requires that users must be logged off the system and this could result in down time. The system owner has to balance the potential loss of data against system downtime; in this instance it might be worthwhile to consider a hot backup schedule.

Hot backups require a fast tape system to transfer the data to tape while the system is operational. There might be a slight degradation of performance, but the availability of the system overrides this issue.

One way to overcome this is to have a second disk that is empty and the same size as the data disk on the server. Transfer the data from the operational disk to the empty disk; this is a relatively quick operation as the original disk reading, transfer via the internal bus, and disk writing is far quicker than the disk to tape transfer. When complete, the image of the data on the second disk is backed up as if it were off-line. When the backup is complete and has been verified, the data on the second disk can be deleted. The disadvantage is the cost of the additional disk and any associated service costs from the IT department. However, in my view, the benefits of this approach outweigh the disadvantages.

Media Management

Media typically are tapes — either digital audio tape (DAT) or digital linear tape (DLT). The former is cheaper and slower compared to the latter and if you need speed, then DLT currently is the optimal choice.

Media management is defined as the activities necessary to ensure that backups and restores have reliable media, where they need it and when they need it. This can take a number of forms, such as:

Media Identification. Ensuring that all tapes are identified uniquely with a number, color, and if an automated robot is used, a bar code.

Media Rotation. Regular cycling of media used for backups; this includes replenishing supply and disposing of

Regulatory Requirements

OECD Consensus Guide for Computerized Systems – Section 6d, Back-up (1)
It is standard practice with computerized systems to make back-up copies of all software and data to allow for recovery of the system following any failure which compromises the integrity of the system e.g., disk corruption. The implication, therefore, is that the back-up copy may become raw data and must be treated as such.

GMP 21 CFR 211 – 211.68(b) (2)
…. A backup file of data entered into the computer or related system shall be maintained ... except where certain data, such as calculations performed in connection with laboratory analysis, are eliminated by computerisation or other automated processes. In such instances a written record of the programme shall be maintained along with appropriate validation data. Hard copy or alternative systems, such as duplicates, tapes, or microfilm, designed to assure that backup data are exact and complete and that it is secure from alteration, inadvertent erasures, or loss, shall be maintained.

21 CFR 11 Requirements
The impact of electronic record and electronic signature regulations (3) also means that data must be backed up effectively to avoid data loss as 21 CFR 11 has specific requirements that involve backup and recovery of your data:
11.10(c): *Protection of records to enable their accurate and ready retrieval throughout the records retention period.*

Thus backup and recovery are very important functions that need to be validated to demonstrate that the overall process works and continues to do so after upgrades of software.

PIC/S Guidance – Section 19.5 (4)
The validated back-up procedure including storage facilities and media should assure data integrity. The frequency of back up is dependent on the computer system functions and the risk assessment of a loss of data. In order to guarantee the availability of stored data, back-up copies should be made of such data that are required to re-construct all GXP-relevant documentation (including audit trail records).

Section 19.6
There should be written procedures for recovery of the system following a breakdown; these procedures should include documentation and record requirements to assure retrieval and maintenance of GXP information. The examination of the procedures and records should assure that the following basic back up and disaster recovery requirements are satisfied:
- *There should be procedures to assure routine back-up of data to a safe storage location, adequately separated from the primary storage location, and at a frequency based on an analysis of risk to GXP data.*
- *The back-up procedure including storage facilities and media used should assure data integrity. There should be a log of backed up data with references to the media used for storage. Media used should be documented and justified for reliability.*
- *All GXP related data, including audit trails should be backed-up.*

Procedure for regular testing, including a test plan, for back up and disaster recovery procedures should be in place.
- *A log of back up testing including date of testing and results should be kept. a record of rectification of any errors should be kept.*

unreliable media (media are considered unreliable when they have been used beyond their normal supported life, when they have been found to be unreadable, when they have flaws making them unusable, or when an error is reported during backup). This is key: do not think you can save money by reusing suspect media. You will pay a much higher price in the long run through data loss.

Logical Media Library. A catalog of the backup media with retrieval index, contents, and location for each system.

Media Audit. Verification that media can be found at the location specified, are readable, contain the data specified, and are listed in the logical media library. Audits can be scheduled at regular frequencies to confirm that there would be no problem locating the appropriate tapes.

Dual Locations. Once every two weeks or a month, full backup duplicate tapes are made and they are stored in a separate location either on the site or offsite as a disaster recovery measure.

Manage backup media generations. Depending upon retention policies, determine which backup generations can be reused and which must be saved. For example, if retention indicates that three months of the first full backup of each month are to be saved, and it is the middle of September, the media from June can be reused.

Determine additional needs for new media. If media use is increasing due to higher volumes of data being backed up, more frequent backups, or other changes in the backup profile, there might be a need for additional media. Plan proactively for this rather than run out of tapes and have no cover.

Despite all the efforts of designing fault-tolerant hardware there will be a time during the operation of any system that anything from a single file to a whole disk will be need to be restored. This is where the appropriate tape is invaluable, assuming the backup has been done correctly and the tape can be read.

Request Media from Library. You'll need to identify the tape or tapes that the data are on and bring them to the tape unit for the system. A restore request usually will indicate the file(s) to be restored and the date of the backup. Thus, the media request resulting from this process will identify the media to be used.

Execute and Verify Restore. Using the correct tape — identified through your effective library catalogue that you validated before it became operational — the file or data requested are restored to your system. Of course, we do not forget to verify that the recovery has worked. Database recoveries can be a little more complex than simple files. For example, a database recovery might entail recovering the log files (for example, redo logs) following a restore.

Return Media. Tape(s) are returned to the library.

Summary

We've looked at the rationale and process for backup and recovery as well as some steps using hardware that you can take to make your computer system more resilient. However, we have not considered the records generated by backup and recovery and the written procedures required under GXP regulations. The next installment of Focus on Quality will address these two aspects.

References

1. Organization for Economic Development (OECD), Paris, 1995.
2. Current Good Manufacturing Practice regulations (21 CFR 211), Food and Drug Administration, Rockville MD, 1978.
3. Electronic Records; Electronic Signatures Final Rule (21 CFR 11), Food and Drug Administration, Rockville MD, 1997.
4. Computerized Systems in GXP Environments, Pharmaceutical Inspection Convention / Scheme (PIC/S), Geneva, 2003. ■

Circle XXX