

IBM Statement of Support: Clarification of supported and unsupported use of Flashcopy backups of mounted AIX filesystems.

December 21, 2004

Why is FlashCopy of a mounted file system unsupported?

The specific problem is this: the file system logging mechanism was designed to recover from a power failure or a system crash. In those situations the I/O to the separate file system and log volumes stops at the same time. This point-in-time image of the file system and log guarantees that when the log is replayed all of the metadata will be consistent. In situations where writing continues to either the log or the file system while writes to the other have been stopped that guarantee cannot be made.

The result of replaying a log in the situation where all I/O does not stop at the same point-in-time is a file system that may contain corruption. There will be no indication of the corruption at the time the copy is mounted – mount will replay the log successfully and mount the file system. However, at some later time when the inconsistent metadata is accessed there will be problems. If the corruption is recognized as such then the system may crash. If it is not recognized when it is used then loss or corruption of user data may result.

When is FlashCopy of a mounted file system a supported operation?

The latest levels of AIX 5.2 and 5.3 include a freeze/thaw function for the JFS2 file system. FlashCopy of a mounted file system is supported when the JFS2 freeze/thaw function is used.

Note: The freeze/thaw function is currently available for AIX 5.2 with APARs IY59928 and IY59770. The 5.3 version will be available in the first maintenance level, due shortly.

Some storage systems provide a means for holding all I/O to a set of volumes to produce a consistent image across those volumes. This function is generally referred to as “consistency groups”. It is possible to produce a valid point-in-time image of the file system and log when the storage system provides this function and it is properly set up by the administrator. However, this has not been explicitly tested by AIX development.

What steps are recommended for a consistent copy?

The recommended, supported environment for producing valid point-in-time copies is:

1. AIX version 5.2 or 5.3
2. JFS2 file systems
3. Use of “consistency groups” where multiple volumes must be copied

The usual scenario would then follow this script:

1. Set the application to on-line backup mode, if possible.
2. Issue a sync command.
3. Issue the file system freeze command:
`chfs -a freeze=<timeout in seconds> /fsname`

4. Issue the point-in-time copy command to the required LUNs of the underlying storage subsystem. Wait for the point-in-time copies to complete.
5. Issue the file system thaw command:
`chfs -a freeze=off /fsname`
6. Set the application back to normal mode.

What if logredo fails?

Replaying the log should succeed even in unsupported configurations. If logredo produces error messages then the service team should examine the log to look for evidence of system errors. One instance where logredo will fail is in the case of log wrap. If the log has wrapped it cannot bring the file system to a consistent state. When log wrap occurs the log needs to be extended.

What about user data?

The discussion so far has been about the effect of FlashCopy on the integrity of file system metadata. It should be pointed out that application data within the file system must also be treated carefully to insure its integrity. The primary concern here is the buffering of data. Most user data is cached either at the application level or in the kernel before it is written to disk. The sync command or fsync system call will force data cached in the kernel to be written from the cache to the disk, and it will force modified metadata to be logged. Applications should ideally provide an interface to users that allow the application to be quiesced. At the time of quiesce the application should issue fsync calls for all open files. The administrator should also issue sync commands to force out data that is not associated with well-behaved applications. The sync command may not be able to write all data for files that are actively in use at the time of the sync. A short delay followed by another sync command might help, but if the applications continue to write new data some will not be copied. Please refer to the appropriate documentation for the sync command and fsync system call for more in-depth information.

Many applications and middleware include controls for doing online backup. These facilities need to be used in addition to the sync command and fsync system call.

The JFS2 freeze/thaw will guarantee a point-in-time view of user data as well as the metadata. It is the most effective way (short of unmount) to ensure that all data that has been written is committed to disk.

What can be done if a supported environment is not available?

*** IBM strongly recommends that supported methods be used to backup data. In the event supported methods cannot be used, the following information is provided to assist customers to in attempting to produce valid, consistent copies. These procedures in no way guarantee that valid, consistent copies will result nor does it imply that IBM will provide support for errors that may occur if these procedures are used. ***

Note: Other than using the supported Freeze/Thaw command as discussed previously, the only other supported procedure in all other AIX filesystem environments for producing

consistent point-in-time copies is to unmount the file system prior to invoking the point-in-time copy function.

If it is not possible to do that then the following actions should be taken:

1. Use JFS2 with inline log so that all data is in a single volume. If not using JFS2 or inline logs, don't log multiple file systems to a single log.
2. Take action to quiesce applications in order to prevent I/O from continuing during the copy process. (Use FlashCopy consistency groups or equivalent, use application function to quiesce, use sync/fsync to schedule/force writes to disk and delay some time prior to initiating the FlashCopy operation.)
3. Do not attempt to mount the copied file system until the following steps are done:
 - Run "logredo /dev/<logname>". If the "log wrap" error is displayed then the log for the source volume needs to be extended and the copy recreated.
 - Run a full fsck on the copied file system after it is created. Any error produced from this indicates that there was a problem with the copy and the copied file system should not be used.

The purpose of these steps is to minimize the potential for corruption. If these procedures are followed but fsck on the copied file system produces errors then it is the result of other factors in the environment. It is not a system error.