T10/08-156 revision 0

Date: March 13, 2008

To: T10 Committee (SCSI)

From: George Penokie (LSI)

Subject: SBC-3 SPC-4: Non-volatile cache becoming volatile

1 Overview

In devices that contain non-volatile cache that non-volatile cache may become volatile. This may occur, for example, if the battery voltage goes too low. If this happens there is currently nothing defined in the SCSI standards that notifies application clients that the cache is no longer non-volatile. This could lead to data corruption.

This proposal adds a new Warning ASCQ that would be reported using IEC definitions.

2 SPC-4 changes

Add the following ASCQ into SPC-4:

0Bh 06h DT LPWROMAEBKVF WARNING - NON-VOLATILE CACHE NOW VOLATILE

3 SBC-3 changes

4.11 Caches

Direct-access block devices may implement caches. A cache is an area of temporary storage in the direct-access block device with a fast access time that is used to enhance performance. Cache exists separately from the medium and is not directly accessible by the application client. Use of cache for write or read operations may reduce the access time to a logical block and increase the overall data throughput.

Cache stores user data and protection information, if any.

Cache may be volatile or non-volatile. Volatile caches do not retain data through power cycles. Non-volatile cache memories retain data through power cycles. There may be a limit on the amount of time a non-volatile cache is able to retain data without power.

Non-volatile caches may become volatile (e.g., battery voltage becomes to low too sustain cache contents when power is lost). In this case operations that use the force unit access non-volatile cache (FUA_NV) bit in the CDB of commands performing write or read operations may bypass the cache resulting in a decrease in overall data throughput.

If a non-volatile caches becomes volatile, then the device server shall set the REMAINING NON-VOLATILE TIME field to zero (see 6.2.4x).

If:

    a) an Extended INQUIRY Data VPD page (see SPC-4) indicates that the device server contains non-volatile cache (i.e., NV_SUP bit set to one):
    b) the non-volatile cache become volatile: and
    c) reporting of informational exceptions control warnings is enabled (i.e., EWASC bit is set to one in the Information Exceptions Control mode page (see SPC-4)),

then the device server shall report the change in the cache as specified in the Information Exceptions Control mode page with an additional sense code set to WARNING - NON-VOLATILE CACHE NOW VOLATILE.

If:

    a) an Extended INQUIRY Data VPD page indicates that the device server contains non-volatile cache (i.e., NV_SUP bit set to one):
    b) the non-volatile cache become volatile: and

c) the reporting of informational exceptions control warnings is disabled (i.e., EWASC bit is set to zero in the Information Exceptions Control mode page),

then the device server shall establish a unit attention condition (see SAM-4) for the initiator port associated with every I_T nexus, with the additional sense code set to WARNING - NON-VOLATILE CACHE NOW VOLATILE.

If:

a) a power on or hard reset occurs;
b) an extended INQUIRY Data VPD page indicates that the device server contains non-volatile cache (i.e., NV_SUP bit set to one): and
c) the non-volatile cache is currently volatile,

then the device server shall establish a unit attention condition for the initiator port associated with every I_T nexus, with the additional sense code set to WARNING - NON-VOLATILE CACHE NOW VOLATILE.

During read operations, the device server uses the cache to store logical blocks that the application client may request at some future time. The algorithm used to manage the cache is not part of this standard. However, parameters are provided to advise the device server about future requests, or to restrict the use of cache for a particular request.

During write operations, the device server uses the cache to store data that is to be written to the medium at a later time. This is called write-back caching. The command may complete prior to logical blocks being written to the medium. As a result of using a write-back caching there is a period of time when the data may be lost if power to the SCSI target device is lost and a volatile cache is being used or a hardware failure occurs. There is also the possibility of an error occurring during the subsequent write operation. If an error occurred during the write operation, it may be reported as a deferred error on a later command. The application client may

request that write-back caching be disabled with the Caching mode page (see 6.3.4) to prevent detected write errors from being reported as deferred errors. Even with write-back caching disabled, undetected write errors may occur. The VERIFY commands and the WRITE AND VERIFY commands may be used to detect those errors.

When the cache becomes full of logical blocks, new logical blocks may replace those currently in the cache. The disable page out (DPO) bit in the CDB of commands performing write, read, or verify operations allows the application client to influence the replacement of logical blocks in the cache. For write operations, setting the DPO bit to one specifies that the device server should not replace existing logical blocks in the cache with the new logical blocks being written. For read and verify operations, setting the DPO bit to one specifies that the device server should not replace logical blocks in the cache with the logical blocks that are being read.

> NOTE 1 - This does not mean that stale data is allowed in the cache. If a write operation accesses the same LBA as a logical block in the cache, the logical block in the cache is updated with the new write data.

Application clients may use the force unit access (FUA) bit in the CDB of commands performing write or read operations to specify that the device server shall access the medium. For a write operation, setting the FUA bit to one causes the device server to complete the data write to the medium before completing the command. For a read operation, setting the FUA bit to one causes the device server to retrieve the logical blocks from the medium rather than from the cache.

When the DPO and FUA bits are both set to one, write and read operations effectively bypass the cache.

Application clients may use the force unit access non-volatile cache (FUA_NV) bit in the CDB of commands performing write or read operations to specify that the device server may access a non-volatile cache, if any, rather than the medium, if the FUA bit is set to zero. For a write operation, an FUA_NV bit set to one with the FUA bit set to zero allows the device server to complete the data write to non-volatile cache rather than the medium before completing the command. For a read operation, an FUA_NV bit set to one with the FUA bit set to zero allows the device server to retrieve the logical blocks from the non-volatile cache rather than the medium.

When a VERIFY command or a WRITE AND VERIFY command is processed, both a force unit access and a synchronize cache operation are implied, since the logical blocks are being verified as being stored on the

medium. The DPO bit is defined in the VERIFY command since the VERIFY command may cause the replacement of logical blocks in the cache.

Commands may be implemented by the device server that allow the application client to control other behavior of the cache:

a) the PRE-FETCH commands (see 5.4 and 5.5) cause a set of logical blocks requested by the application client to be read into cache for possible future access. The logical blocks fetched are subject to later replacement;

b) the SYNCHRONIZE CACHE commands (see 5.20 and 5.21) force any write data in cache in the requested set of logical blocks to be written to the medium. These commands may be used to ensure that the data is written and any detected errors reported;

c) the Caching mode page (see 6.3.4), writable by the MODE SELECT commands, allows control of cache behavior.