Vess™ R2000 Tech Brief

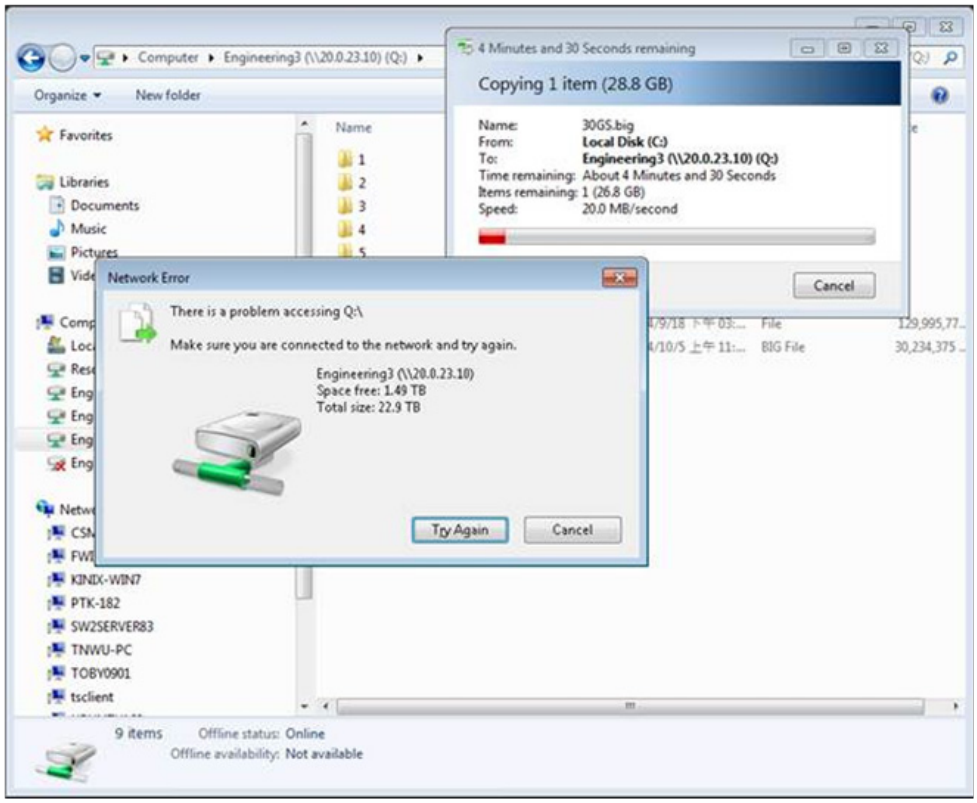
NAS Big File Copying Issue

The Problem

Copying very large files to NAS by CIFS protocol often fails.

The Symptoms

1. File copy failure occurs for large files, typically sized greater than 10 Gigabytes. It is not known to happen for small files.
2. The problem happens mostly when the system is busy, i.e. the level of system loading is unusually high.
3. As the Share Disk usage increases, there is a higher likelihood the problem will occur.
4. The problem is unpredictable.
5. When the problem happens, it is preceded by a significant drop in transfer speed. Then the file copying fails.
6. Windows Explorer reports the error in the pop-up alert pictured here:



The Root Cause

This issue might be caused by a defect of the GFS2 file system. In a clustered file system, the file locking mechanism is used when managing multiple nodes simultaneously. However, the file lock has known issues, including the following:

- listing entries in a folder while it is being written to
- large file writing and deletion
- large Share Disk (file system)capacity support

  Note that each Share Disk on the Vess R2000 has its own file system.

- memory resource allocation while writing large files (files larger than about 10 Gb)

Red Hat documentation recommends using a smaller file system. On the Vess R2000, Share Disk of about 1TB up to 10 TB are recommended. In addition, it describes a known issue when the amount of free capacity is low. The following statement is excerpted from Red Hat user documentation:

> ### 2.3. Block Allocation Issues
>
> This section provides a summary of issues related to block allocation in GFS2 file systems. Even though applications that only write data typically do not care how or where a block is allocated, a little knowledge about how block allocation works can help you optimize performance.
>
> #### 2.3.1. Leave Free Space in the File System
>
> When a GFS2 file system is nearly full, the block allocator starts to have a difficult time finding space for new blocks to be allocated. As a result, blocks given out by the allocator tend to be squeezed into the end of a resource group or in tiny slices where file fragmentation is much more likely. This file fragmentation can cause performance problems. In addition, when a GFS2 is nearly full, the GFS2 block allocator spends more time searching through multiple resource groups, and that adds lock contention that would not necessarily be there on a file system that has ample free space. This also can cause performance problems.
>
> For these reasons, it is recommended that you not run a file system that is more than 85 percent full, although this figure may vary depending on workload.
>
> .....
>
> #### 2.3.3. Preallocate, If Possible
>
> If files are preallocated, block allocations can be avoided altogether and the file system can run more efficiently. Newer versions of GFS2 include the **`fallocate`**(1) system call, which you can use to preallocate blocks of data.

As the documentation states, when the files system capacity is nearly used up (greater than 85% usage), allocation of available blocks slows down because more time is spent in the search process. Large files be written under these circumstances are more likely to have problems with file locking. If the file system supports the fallocate system, writing large files is more efficient since data blocks are pre-allocated, it is not necessary to use system resources to allocate blocks.

Validation

In order to replicate the problem, create several large Share Disks and fill them to 80% or greater capacity. The rate of occurrence increases as capacity usage percentage increases.

The test reports below were produced using Share Disks at 80 to 95% usage capacity. In each testing round, a 30 Gb file is copied to the Share Disks. The results are listed in the tables below. Note that '1' indicates file copy completion (success), and 'X' indicates the copy failed.

1. ≒ 80%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	80%	1		1			1	
8.0T	79%	1	1	1		1	1	
7.0T	80%	1	1	1	1	1	1	
23T	79%	1	1	1	1	1		1
23T	74%	X	X		X			

2. ≒ 85%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	84%	1		1			1	
8.0T	84%	1	1	1		1	1	
7.0T	85%	1	1	1	1	1	1	
23T	83%	1	1	1	1	1		
23T	84%	X	X		X			X

3. ≒ 90%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	89%	X		X			X	
8.0T	89%	1	1	1		1	1	
7.0T	91%	1	1	1	1	1	1	
23T	88%	1	1	1	1	1		
23T	89%	X	X		X			X

4. ≒ 95%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	95%	X		1			X	
8.0T	95%	X	X	1		X	X	X
7.0T	95%	1	1	1	1	1	1	1
23T	94%	X	X	1	X	X		
23T	95%	X	X		X			

From the above results, the following conclusions can be made:

- The rate of occurrence increases with the size of the file (over 10Gb)
- The rate of occurrence increases as available capacity decreases

Comparison Test on XFS

To help determine if the problem is caused by a defect in the GFS2 file system, the same test is run using the XFS file system. The results of the comparison test are listed in the tables below.

1. ≒ 80%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	79%	1		1			1
8.0T	79%	1	1	1		1	1
7.0T	79%	1	1	1	1	1	1
23T	78%	1	1	1	1	1	
23T	78%	1	1		1		

2. ≒ 85%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	84%	1		1			1
8.0T	85%	1	1	1		1	1
7.0T	86%	1	1	1	1	1	1
23T	84%	1	1	1	1	1	
23T	84%	1	1		1		

3. ≒ 90%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	90%	1		1			1
8.0T	90%	1	1	1		1	1
7.0T	90%	1	1	1	1	1	1
23T	90%	1	1	1	1	1	
23T	89%	1	1		1		

4. ≒ 95%

size	usage	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB	CP 30GB
8.0T	96%	1		1			1
8.0T	97%	1	1	1		1	1
7.0T	98%	1	1	1	1	1	1
23T	95%	1	1	1	1	1	
23T	95%	1	1		1		

The comparison test appears to confirm that the problem is an issue with the GFS2 file system.

Short Term Workaround

- Keep free space of at least 15% or more.
- Avoid copying more than 3 big (> 10Gb) files at the same time.

Preferred Solution

- Use XFS file system to replace GFS2 in next FW version (SR2.4).
- Migrate data from GFS2 to XFS file system if it is needed. Steps:

1. Upgrade FW to SR2.4
2. Create new Share Disk
3. Configure Share Disk Clone from old Share Disk to new one
4. Delete old Share Disk and get free capacity
5. Repeat steps 2~4