

Product Model Name: CS3102 and FS3102
subsystems

Date: March 2006

Reference No. RTS-CB 018

SUBJECT: Volumes greater than 2TB on Windows OS

Overview

This document explores how different Windows operating systems handle disk volumes sizes greater than 2 TB. The main hurdle in accessing logical volumes greater than 2TB is that most computers operate in 32-bit mode which means accessing data with an address greater than 2^{32} is not natively supported.

Two issues must be solved before IO operations can happen above this range, firstly the MBR must be overridden with a table that can hold values greater than 2^{32} and secondly the Operating system must be capable of handling addresses greater than 2^{32} .

Windows Server 2003 SP1 supports recognition and partitioning of volumes with capacities in excess of 2TB, through the use of the Globally unique identifier Partition Table or "GPT" format. This is particularly relevant with the availability of higher disk drive capacities and associated RAID LUN availability to the Host Operating System. Although the *Raidtec FS3102* and *CS3102* RAID subsystems support the 16-byte SCSI command set which allows for >2TB LUN addressing, the Operating System disk parameters must be properly configured in order to use these large LUNs.

With Windows Server 2003 Service Pack 1 (SP1) and Windows XP 64-bit Edition (x64), Microsoft have added support for 64-bit block numbers in the disk/class layer, using the new SCSI Commands included in the *SCSI-3 Block Commands-2* command set. Microsoft have also enabled GPT support for all Windows Server 2003 SP1 platforms. The absolute file system limit is actually 2^{64} 512-byte blocks. However, NTFS reduces this, because it only supports a maximum of (32-bit cluster number * 64K per cluster), which equals 256TB. If a smaller cluster (Allocation Size) is chosen, the maximum file system size will be reduced accordingly as shown in table 1. The default cluster size is 4K which allows a volume of 16TB.

The limits are as follows:

- Basic or dynamic volume size: 2^{64} blocks = 2^{73} bytes
- Maximum NTFS file system size that can be realised on Windows: 256TB

NTFS Allocation Size	<i>Maximum Volume Capacity</i>
512 bytes	2 terabyte
1k bytes	4 terabyte
2k bytes	8 terabyte
4k bytes	16 terabyte
8k bytes	32 terabyte
16k bytes	64 terabyte
32k bytes	128 terabyte
64k bytes	256 terabyte

Table 1 : Max NTFS Volume Capacities

Partition Tables

Basic Disk

This is the default partition table for DOS, Windows 95/98 and Millennium Edition. This table can only handle logical volumes of 2 TB and these operating systems cannot work with volumes greater than 2TB. In 32-bit Windows, a basic disk is called an MBR disk.

Dynamic Disk

When you convert a basic disk to a dynamic disk, the partition information is transferred to a Logical Disk Manager (LDM) database located on the disk. All dynamic disks on a machine have an identical copy of the LDM database. This permits them to participate in shared volume configurations such as RAID 1, RAID 0, RAID 5 and volume spanning.

Globally Unique Identifier (GUID) Partition Table

The GUID Partition Table or GPT can have a partition of up to 18 exabytes in size and 128 separate partitions. This compares favourably to an MBR disk (basic disk) which can only access volumes of 2TB in size and allows a maximum of 4 primary partitions. All Windows 64-bit Operating Systems must be installed on a GPT partition.

Dynamic disks and GPT

When a GPT disk is also a dynamic disk, the GPT information is transferred to the LDM table database. Windows automatically stores the LDM in a Microsoft Reserved Partition (MSR). This partition is 64MB in size on disks up to 16GB and 128MB on larger disks.

Operating System Support

Windows NT 4.0

NT cannot handle disks greater than 2 TB, and cannot handle dynamic disks. It is restricted by the MBR and the fact that it's a 32-bit OS.

Windows 2000 (service pack 4) x86

Windows 2000 cannot handle logical volumes greater than 2 TB, however, it can handle dynamic disks.

Windows XP (service pack 6a) x86

XP cannot handle logical volumes greater than 2 TB, however, it can handle dynamic disks.

Windows 2003 server (service pack 1) x86

Windows 2003 server claims support for disk's greater than 2TB but this can be misleading since, while the logical volume will show up in Windows disk management tool, it will automatically partition the volume into separate pieces of no larger than 2TB, and allow no access to the partitions above address 2^{32} . In order to use volumes greater than 2TB, it is necessary to create a GPT partition table for the volume. GPT imposes some restrictions, which are listed below.

Windows 2003 Server (service pack 1) Itanium

Itanium's 64-bit capabilities make it possible to see disks greater than 2 TB. To solve the MBR 32-bit addressing problem, GPT is used.

Windows XP 64-bit

XP 64-bit OS is capable of handling logical volumes greater than 2TB. Its native support for 64-bit addressing, along with its support for GUID partitioning tables means that it can handle logical volume sizes of up to 18 exabytes.

Restrictions on using GPT

- 32-bit versions of previous Windows operating systems (Windows 9x, Windows NT 4.0, Windows 2000, and Windows XP) do not support GPT disks.
- 32-bit versions of Windows Server 2003 Service Pack 1 (SP1) and Windows XP 64-bit Edition (x64) cannot boot from GPT disks.
- On Itanium-based computers, the operating system loader and boot partition must reside on a GPT disk. Other hard disks can be either MBR or GPT.
- You cannot move GPT disks to x86-based computers running Windows NT 4.0, Windows 2000, Windows XP, or Windows Server 2003. However, you can move GPT disks from either x86-based computers running Windows Server 2003 with SP1 or x64-based computers to Itanium-based computers running Windows Server 2003 or Windows XP, or vice versa.
- You cannot move a GPT disk with an Itanium-based version of Windows from Itanium-based computers to either x86-based computers running Windows Server 2003 with SP1 or x64-based computers and then boot that Operating System. GPT disks that are used in non-Itanium-based computers must be used for data storage only
- You can have both MBR and GPT disks in a single dynamic disk group. You can also have a mix of basic GPT and MBR disks, which are not part of disk groups. You can create mirrored, striped, spanned and RAID-5 volumes using a combination of MBR and GPT disks, but the MBR cylinder alignment restriction may cause some difficulties in creating mirrored volumes. Always mirror the MBR disk to the GPT disk to avoid cylinder alignment difficulties.
- You can convert an MBR disk to a GPT disk and vice versa only if the disk is empty.

- Mirroring the EFI System partition is not supported. Instead, you must clone the EFI System partition using the 'bootcfg' command.
- You cannot use the GPT partitioning style on removable media, or on cluster disks that are connected to shared SCSI or Fibre Channel buses used by the Cluster service.

Notes

- For the test system under Windows Server 2003 SP1, changes were not made to the SCSI or Fibre Channel HBA drivers, because none were needed. However, system designers should check with their adapter vendors for any applicable BIOS or firmware updates.
- The HBA (SCSI or FC) may need to be configured to run larger volumes.
 - For some HBAs this is easily achieved using the configuration tools that come with them.
 - For others, it is necessary to manually change some of the registry keys to allow larger volumes.
 - For the rest (older HBAs - more so with SCSI) it is not possible.
- A huge disk volume can present long delays with 'chkdsk' and searching, especially with billions of files. There is no substitute for a considered layout of storage and partitioning.

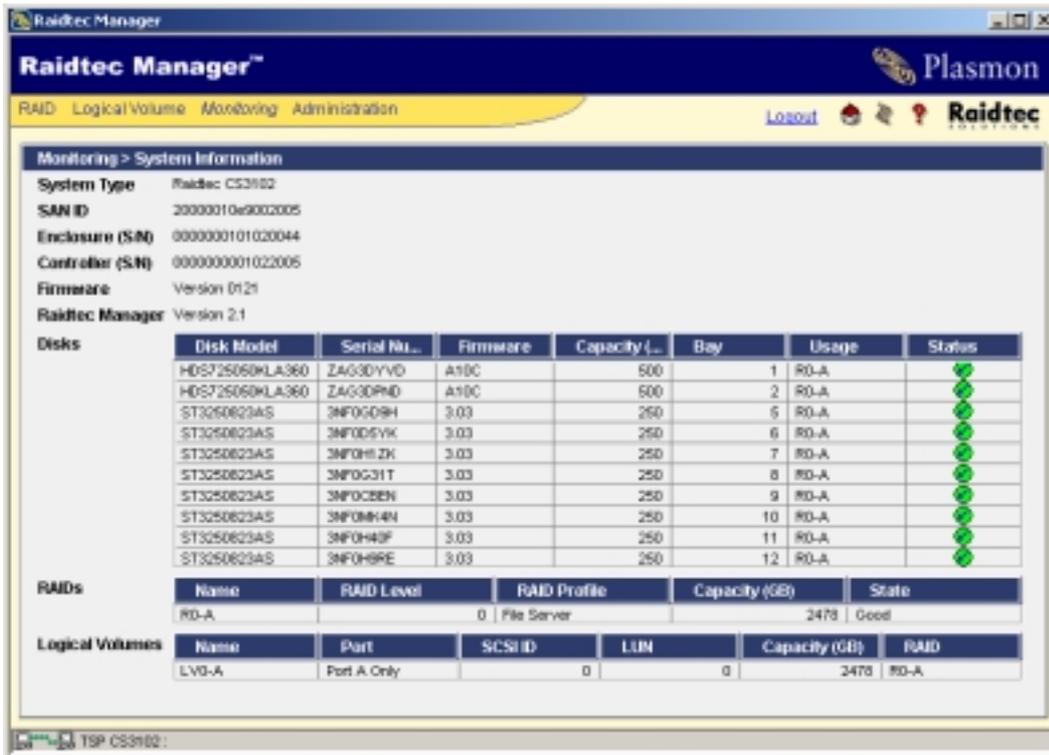
Example : How to create partitions greater than 2TB

The Raidtec FS3102 FC-SATA Raid subsystem will operate with >2TB GPT partitioning under the default Windows2003 Server (SP1 or R2) Microsoft Qlogic 23xx HBA driver. Similarly, the CS3102 SCSI-SATA Raid subsystem will operate on the default driver for the LSI U320 adapter (at BIOS 5.07.03 on 'factory defaults').

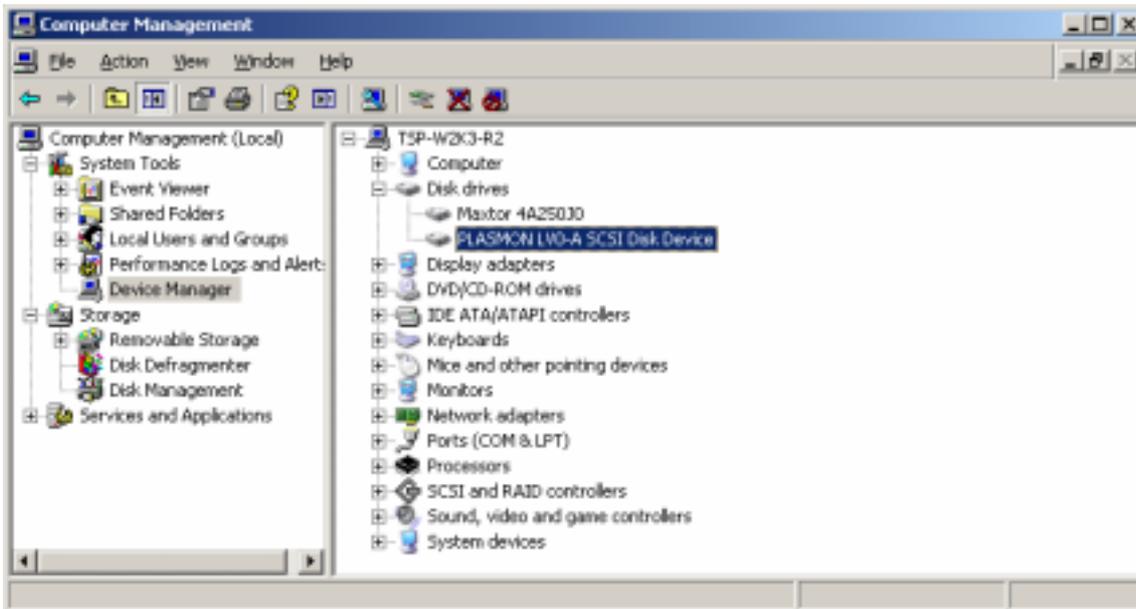
CS3102 subsystem >2TB LUN recognition

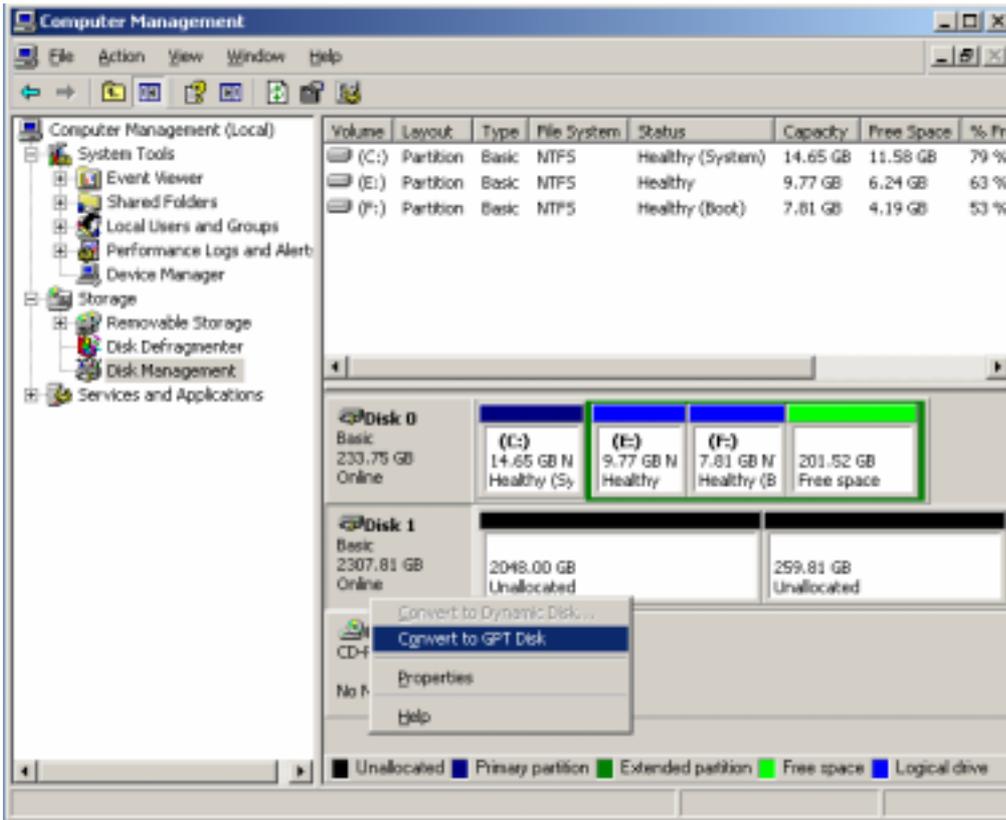
A sample CS3102 subsystem is configured with a single LUN of over 2TB, and integrated into the Windows 2003 SP1 operating system, via the LSI Logic U320 and Microsoft driver.

Logical Drive "LV0-A" is mapped to SCSI ID-0 LUN-0 on Port-A.



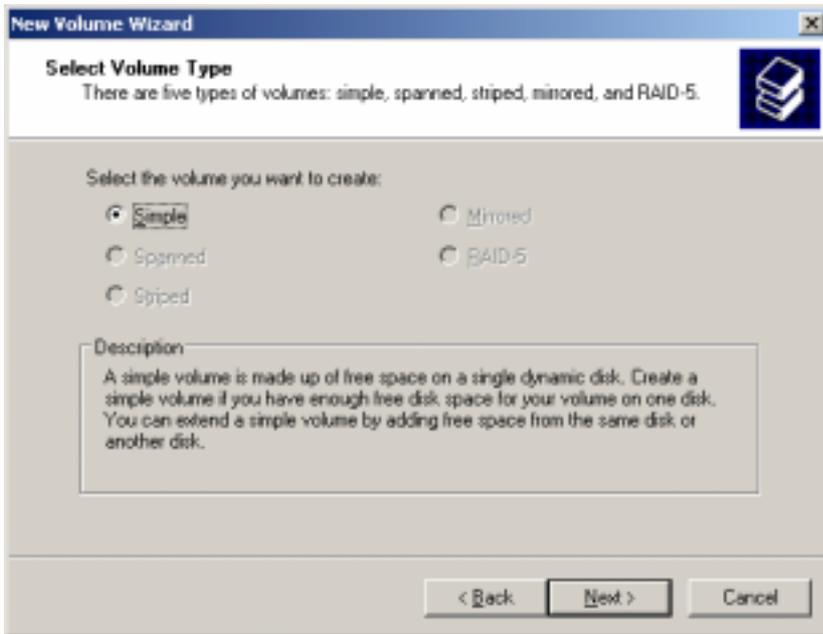
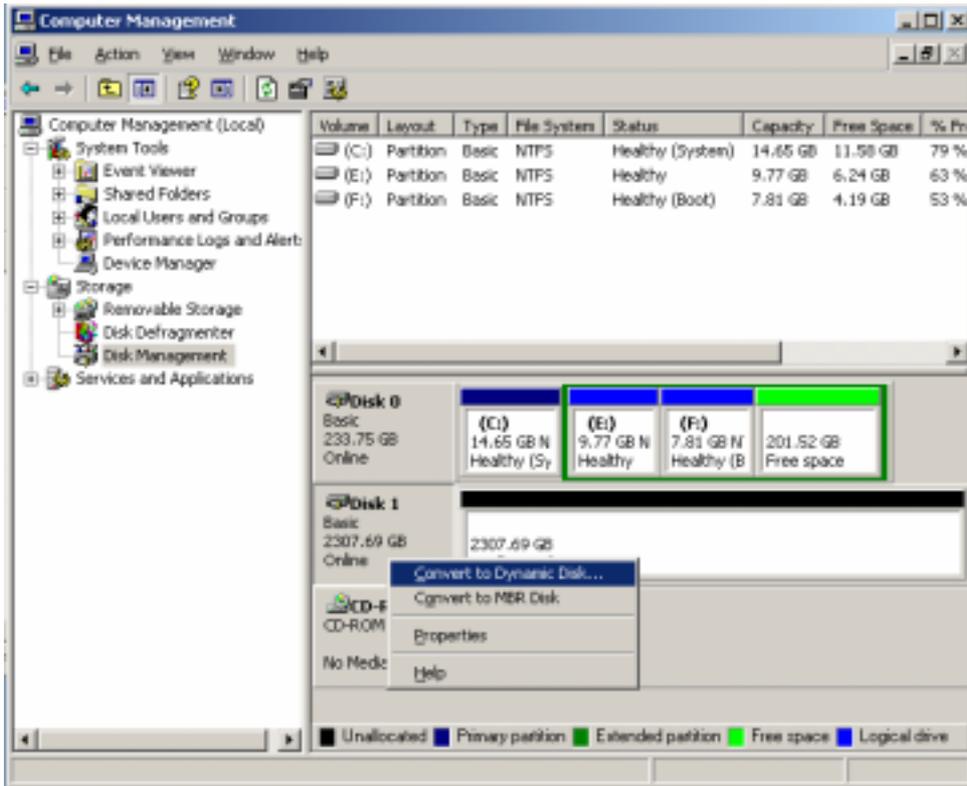
Windows2003 Device Manager recognition and Disk Management selection options



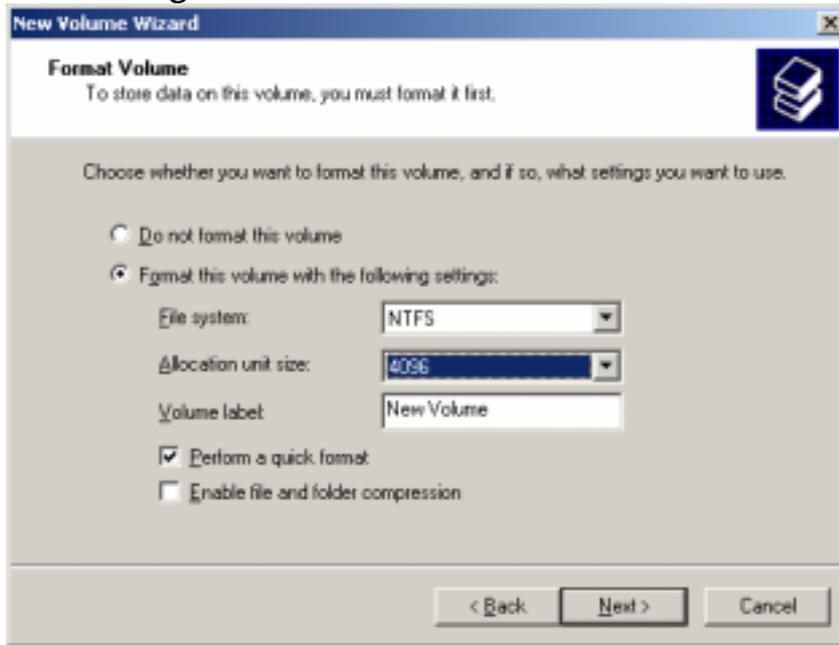


Initially the disk will be seen as multiple <2TB partitions. Select 'Convert to GPT Disk' option for >2TB disk partitioning.

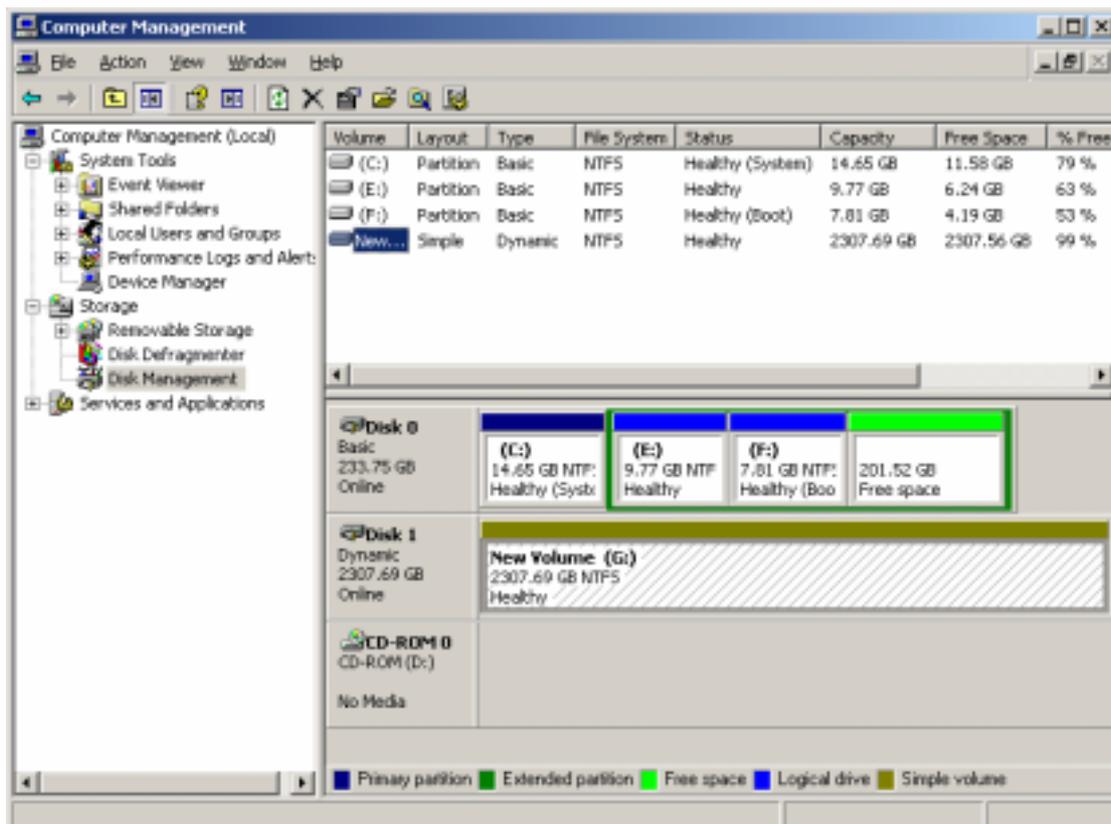
Note: With "Dynamic Disks" available, it is also possible to implement OS Striping or Spanning to create an even larger Virtual Disk using multiple RAID subsystems. However, this is **not** recommended for High Availability solutions because if one of the subsystems fails for whatever reason, it can cause data corruption and potential loss of the entire file system necessitating a restore from backup.



Formatting



Note: The maximum usable volume capacity will be limited by the Allocation Unit Size as shown in Table 1.



For more information

Windows Server 2003 service pack 1:

<http://www.microsoft.com/technet/prodtechnol/windowsserver2003/library/ServerHelp/f9172c16-0114-49dc-bc92-608ff17b76e5.mspx>

Itanium (floating point exceptions):

developer.intel.com/design/itanium/downloads/245415.htm

**Contact Information for Technical Questions:
Worldwide**

Page 9 of 9

Plasmon Technical Support Team
International Calls: +44 (0) 1763 26296
Fax: +44 (0)1763 264407
Email: support@plasmon.co.uk
Internet: www.plasmon.com