z/VM Storage Update: DS6000 and DS8000

Tuesday, May 24, 2005

Steve Wilkins
wilkinss@us.ibm.com

Agenda

- DS8000 Series
  - ► Overview
  - ► Invocation and Use
  - ► Large Volume Support
  - ► Summary Unit Check Support
  - ► SET/QUERY EDEVICE
  - ► Dependencies
  - ► Tips
- DS6000 Series
  - ► Overview
  - ► Invocation and Use
  - ► Preferred Pathing Support
  - ► QUERY PATHS
  - ► SET/QUERY EDEVICE
  - ► Dependencies
  - ► Tips
- Future Development
- Last Call ... for Q&A
- Trademarks

DS8000 Overview

- New storage standard delivering massive scalability and performance

- Up to 6 times the throughput of the Enterprise Storage Server (ESS) Model 800 for increased response time

- Up to 192 TB of storage

- 20% smaller footprint than the ESS Model 800

- Storage system LPARs allow two independent storage systems within a single array; can be used for completely separate production or test environments.

- Same advanced features as ESS models 750 and 800 (e.g., Flashcopy, PPRC, PAV, etc.)

- Supported by z/VM 4.4.0 and 5.1.0

Invocation and Use

- VM supports ESCON/FICON attachment as 3990 Model 3 or 6 Controller with:
  - ► 3390 Model 2, 3, and 9 DASD
  - ► 3390 Model 2 and 3 DASD in 3380 track compatibility mode

- VM supports FCP-attached SCSI LUNs as emulated 9336 Model 20 DASD (on z/VM 5.1.0)

- Certain components of VM support the DS8000's 3390 Model 9 with 65,520 cylinders
  - Known to many as the 3390 Model 54
  - Also available with latest microcode level of the ESS Model 800
  - ESS Model 800 variety supported on z/VM 3.1.0 and up

- VM also supports the new Summary Unit Check architecture provided by the DS8000

Large Volume Details

- VM supports the 65,520 cylinder 3390 Model 9:

  - ► CMS and GCS only support the 32,760 cylinder 3390 Model 9 (a.k.a. Model 27)
  - ► CMS / GCS will not ACCESS or FORMAT a volume greater than 32,767 cylinders
  - ► This affects software functions dependent on CMS functions such as DIRMAINT MOVE, COPY, ERASE and DFSMS MOVE, COPY, CHECK
  - ► CMS keeps its file status and control information below the 16MB-line. Virtual storage errors are possible if too many files reside on a large volume and/or too many disks are accessed at one time.

- VM supports a 1TB (2,147,483,640 512-byte blocks) SCSI LUN as an emulated 9336:

  - ► PAGE, SPOL, and DRCT allocations must remain below the 64GB mark (Page 16,777,215) on a CP formatted LUN
  - ► CMS and GCS LUNs are restricted to 381GB subject to the above *16MB-line file information* restriction and the *software dependent on CMS* restriction

Summary Unit Check Support

- DS8000 contains *2* storage subsystem *clusters* where each cluster handles I/O requests for *1/2* of the defined logical *control units*

- When 1 storage subsystem cluster crashes, the DS8000 will *Failover* I/O operations to the other cluster

- The DS8000 failover process includes rejecting active I/O requests and presenting 1 Summary Unit Check error to each control unit in the failed cluster:
  - ► Summary Unit Checks alert VM of Failover on each associated control unit so VM can clean up 'in-flight' I/O requests (i.e., requests issued by VM but haven't made it to the DS8000 yet)
  - ► DS8000 architecture allows for growth of a storage subsystem cluster with 32K devices (64K with both clusters)
  - ► Failover capabilities of the ESS Model 800 present a unit check to every device in a cluster to alert VM of in-flight I/O (up to 2K devices per cluster)
  - ► *DS8000 cluster size would flood VM with 32K I/O interrupts without Summary Unit Checks*

- VM appropriately clears, rejects, and/or redrives currently active I/O requests as a result of Summary Unit Checks
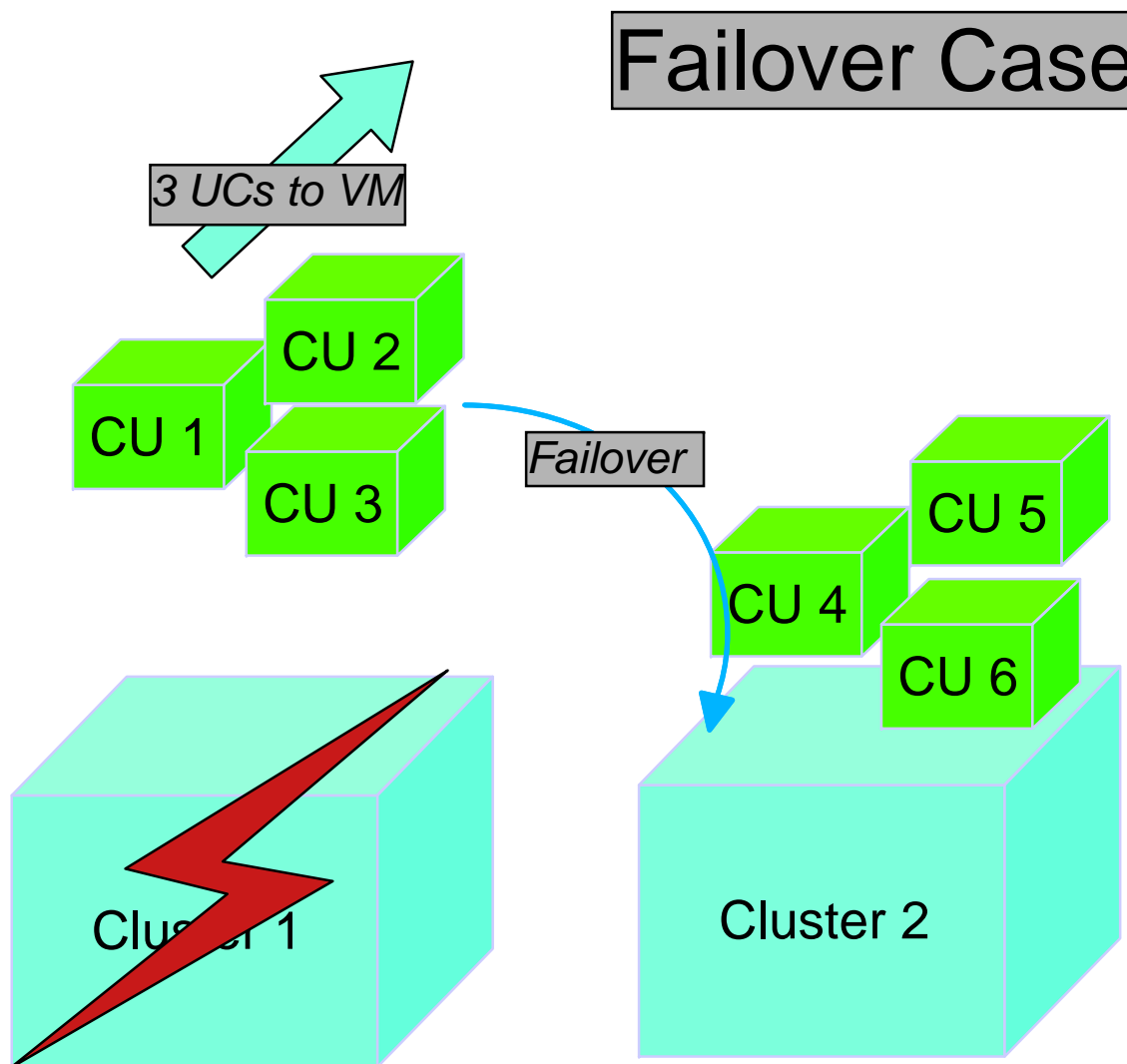
Summary Unit Check Support

Normal Case

CU 2

CU 1

CU 3

CU 5

CU 4

CU 6
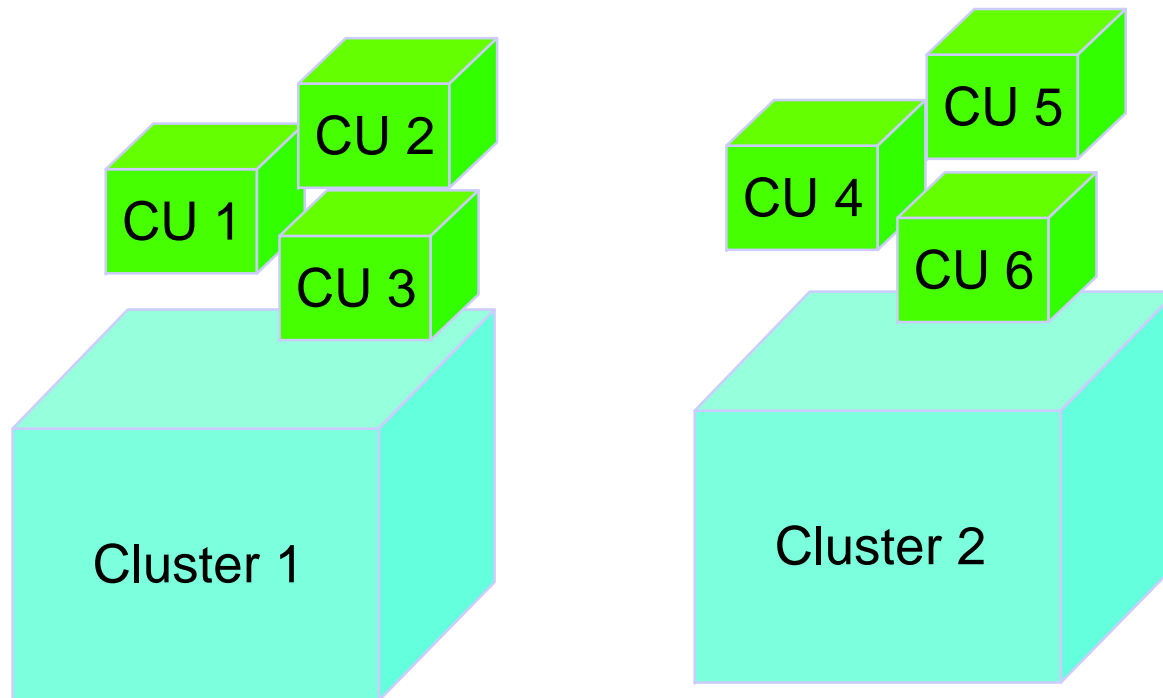
Cluster 1

Cluster 2

- Cluster 1 handles I/O for volumes in logical controllers 1 - 3
- Cluster 2 handles I/O for volumes in logical controllers 4 - 6

Summary Unit Check Support

Failover Case

3 UCs to VM

CU 2

CU 1

CU 3

Failover

CU 5

CU 4

CU 6

Cluster 1

Cluster 2

- Cluster 1 fails!
- I/O for logical controllers 1 - 3 fails over to Cluster 2
- 1 Summary Unit Check per controller is presented to VM for failing cluster

SET EDEVICE

- VM emulates SCSI LUNs as 9336 FBA DASD
- Issue SET EDEVICE command with the new 2107 attribute to map a VM emulated 9336 DASD to a DS8000 SCSI LUN
- Links LUN to VM's DS8000 SCSI driver

```
>>-SET--EDEVice--rdev------------------------------------------->
 >---TYpe--FBA--ATTRibutes--.-1750-.--.-------------.-| Paths |-><
                            |-2105-|   |-ADD PATH----|
                            |-2107-|   '-DELete PATH-'
Paths:                      '-SCSI-'
 <-------------------------------------------------------------<
|--FCP_DEV--nnnn--WWPN--wwwwwwwwwwwwwwww--LUN--llllllllllllllll--|
```

Note: EDEVICE system configuration file statement also updated

QUERY EDEVICE

- Use QUERY EDEVICE command to display what attribute (i.e., device driver) has been defined for an emulated 9336 DASD

- Ensure that attribute 2107 is displayed for emulated 9336 DASD mapped to a DS8000 LUN

Example:

q edev 607
EDEV 0607 TYPE FBA ATTRIBUTES 2107
Ready;

Dependencies

- Hardware
  - ► An ESCON, FICON, or FCP attached DS8000 at the latest microcode level

- Software
  - ► VM CP APARs
    - – VM63534 for base support; PTFs UM31301(440), UM31302 (510)
    - – VM63702 if placing PAGE or SPOL allocations above cylinder 32,767; PTFs UM31400 (440), UM31401 (510)

  - ► CMS APAR
    - – VM63653 to cap SCSI LUN size to 381GB on FORMAT and ACCESS; PTF pending for 510

  - ► DIRMAINT APAR
    - – VM63700 for 3390-54 and SCSI-1TB support; PTFs pending for 440 & 510

  - ► DFSMS/VM APAR
    - – VM63664 for SCSI-1TB support; PTF pending for FL221

Tips

- VM CP APAR VM63534 **must** be applied.   Advanced functions, such as FLASHCOPY, will not work without it.

- VM63702 must be applied if placing PAGE or SPOL allocations above cylinder 32,767 on a CP Formatted volume

- Upgrade DS8000 microcode to latest level possible since hardware fixes continue to go in

DS6000 Overview

- Exceptional price and performance in a modular package

- Same advanced features as ESS Model 750, ESS Model 800, and DS8000 (Flashcopy, PPRC, PAV, etc.)

- Supports 65,520 cylinder 3390 Model 9 (a.k.a. model 54) and 1TB SCSI LUN, same as DS8000

- Up to 4 times the throughput of the ESS Model 750 for increased response time

- Supported by z/VM 4.4.0 and 5.1.0

Invocation and Use

- VM supports FICON attachment as 3990 Model 3 or 6 Controller with:
  - ► 3390 Model 2, 3, and 9 DASD
  - ► 3390 Model 2 and 3 DASD in 3380 track compatibility mode

- DS6000 doesn't provide ESCON attachment

- VM supports FCP-attached SCSI LUNs as emulated 9336 Model 20 DASD (on z/VM 5.1.0)

- VM fully supports the preferred pathing anomaly of the DS6000

Preferred Pathing Support

- DS6000 hardware structure is such that not all D6000 ports (FICON or FCP) have equivalent access to the volumes (DASD or LUNs)
  - ► Half the volumes need to be accessed by crossing a PCI-X bus
  - ► This bus crossing can add 30% to the response time and reduce bandwidth by 50%

- For FICON connections, VM dynamically discovers which ports (paths or CHPIDs) are preferred versus non-preferred when a volume is varied on-line

- For FCP connections, preferred versus non-preferred paths must be defined on the SET EDEVICE command (or EDEVICE system configuration file statement)

- For both FICON and FCP, VM contains logic in its I/O scheduler (or SCSI device driver) to use preferred paths over non-preferred paths

- In general, I/O will failover (i.e., be redriven) on non-preferred paths only if no preferred paths are available or functioning

Logical Path Illustration
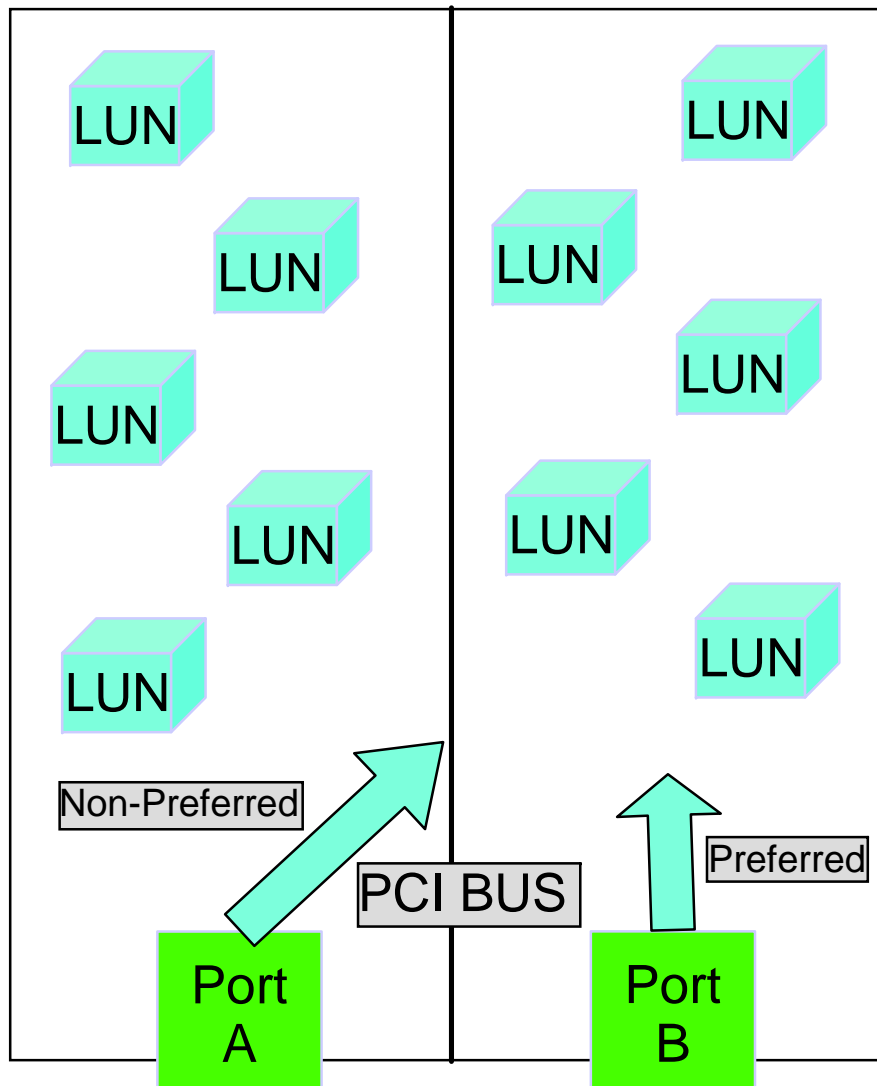
- Port B is preferred path for LUNs on right side of PCI bus; Port A is non-preferred path
- Port A is preferred path for LUNs on left side of PCI bus; Port B is non-preferred

QUERY PATHS

- QUERY PATHS command will display which FICON paths are preferred versus non-preferred when connected to a DS6000

```
Example:


q paths to 991
Device 0991, Status ONLINE
CHPIDs to Device 0991 (PIM)  : 28 14 43 81
Physically Available  (PAM)  : +  +  +  +
Online                (LPM)  : +  +  +  +
Preferred                    : +  +  -  -
                        Legend    + Yes - No
```

Note: Preferred output applicable to only DS6000

SET EDEVICE

- VM emulates SCSI LUNs as 9336 SCSI DASD
- Issue SET EDEVICE command with the new 1750 attribute to map a VM emulated 9336 DASD to a DS6000 SCSI LUN
- Links LUN to VM's DS6000 SCSI driver
- PREFerred and NOTPREFerred parameters define preferred versus non-preferred SCSI paths

```
>>-SET--EDEVice--rdev------------------------------------------->
 >---TYpe--FBA--ATTRibutes--.-1750-.--.-------------.-| Paths |---><
                            |-2105-|  |-ADD PATH----|
                            |-2107-|  '-DELete PATH-'
Paths:                      '-SCSI-'
 <-------------------------------------------------------------<
|--FCP_DEV--nnnn--WWPN--wwwwwwwwwwwwwwww--LUN--llllllllllllllll-->
 >----------------------.-------------.-----------------------|
                        |-PREFerred----|
                        '-NOTPREFerred-'
```

Note: PREF / NOTPREF input applicable only to DS6000

QUERY EDEVICE

- Use QUERY EDEVICE command to display what attribute (i.e., device driver) has been defined for an emulated 9336 DASD
- Ensure that attribute 1750 is displayed for emulated 9336 DASD defined on a DS6000
- Use DETAILS parameter of QUERY EDEVICE to display which DS6000 SCSI paths have been defined as preferred versus non-preferred

---

Examples:

q edev 607
EDEV 0607 TYPE FBA ATTRIBUTES 1750
Ready;

q edev 609 details
EDEV 0609 TYPE FBA ATTRIBUTES 1750
   PATHS:
      FCP_DEV: B100 WWPN: 4004066300C2023B LUN: 0011000000000000 PREF
      FCP_DEV: B200 WWPN: 4004066300C2023C LUN: 0011000000000000 NOTPREF
Ready;

Note: PREF / NOTPREF output applicable only to DS6000

Dependencies

- Hardware
  - ► A FICON or FCP attached DS6000 at the latest microcode level

- Software
  - ► VM CP APARs
    - VM63535 for base support; PTFs UM31393 (440), UM31394 (510)
    - VM63702 if placing PAGE or SPOL allocations above cylinder 32,767; PTFs UM31400 (440), UM31401 (510)

  - ► CMS APAR
    - VM63653 to cap SCSI LUN size to 381GB on FORMAT and ACCESS;   PTF pending for 510

  - ► DIRMAINT APAR
    - VM63700 for 3390-54 and SCSI-1TB support; PTFs pending for 440 & 510

  - ► DFSMS/VM APAR
    - VM63664 for SCSI-1TB support; PTF pending for FL221

Tips

- VM CP APAR VM63535 **must** be applied.   This APAR pre-reqs VM63534 (to *drag* along DS8000 support) and provides preferred pathing support for both FICON and FCP connections.

- See details of the SET EDEVICE command in *Memo To Users* file for APAR VM63535 on how to determine preferred versus non-preferred paths on a FCP connected DS6000

- Like DS8000, upgrade DS6000 microcode to latest level possible since hardware fixes continue to go in

Future Development

- Support of the DS6000 and DS8000 in their native control unit modes (1750 and 2107)

- Virtualization of DASD self description data will allow each guest virtual machine to *see* the control unit mode at the level the guest supports (such as 3990)

- Exploitation of the Parallel Access Volume feature (PAV) for system data and for guest data residing on VM minidisks

no commitment implied

Last Call ... for Q&A

The End

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

CICS*	Language Environment*	S/370
DB2*	MQSeries*	S/390*
DB2 Connect	Multiprise*	S/390 Parallel Enterprise Server
DB2 Universal Database	MVS	VisualAge*
DFSMS/MVS*	NetRexx	VisualGen*
DFSMS/VM*	OpenEdition*	VM/ESA*
e business(logo)*	OpenExtensions	VTAM*
Enterprise Storage Server*	OS/390*	VSE/ESA
ESCON*	Parallel Sysplex*	WebSphere*
FICON	PR/SM	z/Architecture
GDDM*	QMF	z/OS*
HiperSockets	RACF*	zSeries*
IBM*	RAMAC*	z/VM*
IBM(logo)*	RISC	

* Registered trademarks of the IBM Corporation

The following are trademarks or registered trademarks of other companies.

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation.
Tivoli is a trademark of Tivoli Systems Inc.
LINUX is a registered trademark of Linus Torvalds.
Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
Penguin (Tux) compliments of Larry Ewing

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

IBM considers a product "Year 2000 ready" if the product, when used in accordance with its associated documentation, is capable of correctly processing, providing and/or receiving date data within and between the 20th and 21st centuries, provided that all products (for example, hardware, software and firmware) used with the product properly exchange accurate date data with it. Any statements concerning the Year 2000 readiness of any IBM products contained in this presentation are Year 2000 Readiness Disclosures, subject to the Year 2000 Information and Readiness Disclosure Act of 1998.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.