Release Notes for the

Grid Director^{TM} 4000 Series
Software Version 3.0.0

May 13, 2010

Table of Contents

Overview

These release notes pertain to the Grid Director 4000 Series, software version 3.0.0.

Voltaire's Grid Director™ 4000 series of smart switches addresses the growing size and complexity of clusters by providing high interconnect bandwidth, advanced carrier-class management and a unique HyperScale™ stackable architecture.

Grid Director 4700/4200 – Modular Switches

With configurations of up to 324 ports or double-sided 648 ports of 40 Gb/s per port InfiniBand connectivity, the Voltaire Grid Director 4700 delivers an impressive 51.8 Tbps of non-blocking bandwidth and the lowest latency in the industry. The Grid Director's 4700 smart design provides unprecedented levels of performance and makes it easy to build clusters that can scale out to thousands of nodes. The Voltaire Grid Director 4700 is a high performance, ultra low latency and fully non-blocking InfiniBand switch for high performance clusters. The switch's HyperScale™ architecture provides a unique inter-switch link capability for stacking multiples of 324 ports to form highly scalable, cost effective, and low latency fabrics. The Grid Director 4200 is a smaller modular switch, scaling from 18 to 162 40 Gb/s InfiniBand ports, and sharing many components and all software with the 4700. The Grid Director 4200 optimizes the price and space required for small to mid-size (70-160 nodes) clusters. The Grid Director 4700/4200 switches are both fully redundant with no single point of failure.

Grid Director 4036/2036 – Fixed-port Switches

The Voltaire Grid Director 2036/4036 is a high performance, low latency, and fully non-blocking switch for high performance computing clusters. Perfect for use as an edge or leaf switch, the Grid Director 2036/4036 offers full bisectional bandwidth of 1.44/2.88 Tb/s. With built-in high availability and a design that is easy to maintain, the Grid Director 2036/4036 is a cost-effective alternative to proprietary interconnect technologies. With 36 20/40 Gb/s ports in a 1U chassis, the Grid Director 2036/4036 delivers high bandwidth and low latency at an affordable price.

Grid Director 4036E

The Voltaire Grid Director 4036E is a high performance, low latency, and fully non-blocking InfiniBand switch, which includes a built-in Ethernet gateway for bridging traffic to and from Ethernet-based networks or storage. With thirty four 40 Gb/s InfiniBand ports (delivering 2.72 Tb/s), under 100 nanoseconds of port-to-port latency, and two 1/10Gb Ethernet ports bridging traffic in under two microseconds, I/O bottlenecks are removed to allow applications to operate at

their maximum efficiency. The Voltaire Grid Director 4036E was designed as a self-contained full solution including an InfiniBand switch, an embedded subnet manager, and a built-in Ethernet gateway in a compact 1U device. The efficient Grid Director 4036E's smart design makes it easily useful for both small and large clusters.

Key Features

Grid Director 4000 Series

The following lists the key features of the Grid Director 4000 Series:

- Scalable 40Gb/s InfiniBand switching family ranging from 36 ports to 648 ports

- 40/20/10 Gbps InfiniBand (QDR/DDR/SDR) auto-negotiating QSFP

- Ultra-low latency: between 100 and 300 nanoseconds port-to-port

- Unique HyperScale™ architecture allows scaling to thousands of nodes with a single tier of switches (under 400 nanoseconds of latency)

- Up to 51.8 Tbps of non-blocking bandwidth

- Support for longest and most varied range of cables (copper and optic)

- Zero down time with no single point of failure and real-time fault notifications

- InfiniBand specification 1.2 compliant

- Full device management via CLI or standard SNMP interface

- Embedded subnet manager for quick, simple deployment

- Advanced fabric management via Voltaire Unified Fabric Manager (UFM).

Grid Director 4036E built-in Ethernet Gateway.

The following lists the key features of the Grid Director 4036E built-in Ethernet Gateway.

- High performance InfiniBand-to-Ethernet connectivity based on Voltaire-developed silicon

- Support for both 1 and 10GbE connectivity based on SFP+

- Extremely low latency (< 2uS IB-ETH)

- Transparent mapping between IB partitions and ETH VLANs

- Hardware-based multicast acceleration

- Hardware-based traffic flow management (TFM)

- Standard-based (IPoIB) implementation – "plug and play"

- Link aggregation

- Bandwidth Aggregation and High Availability across multiple gateways.

New and Changed Information

This section lists the new features for version 3.0.0. Other than fixes for known issues identified in the "Resolved Issues" section, there are no additional changes for version 3.0.0.

New Features

This section lists the new features in software version 3.0.0.

- Support for new products – Grid Director 4200 and Grid Director 4036E

- Embedded Subnet Manager on the Grid Director 4700/4200, supporting up to 648 nodes.

- Performance management for the Grid Director 4700/4200 tool reporting fabric counters and status.

- Flow-based Congestion Control Management (CCM)

- Support for Voltaire Fabric Collective Accelerator (FCA).

Installation Notes

Use the information in this section as a general guideline for installation requirements and procedures. Follow the specific installation instructions contained within the Installation Manual for the specific Grid Director 4000 Series switch. Refer to Product Documentation for more information.

Software Revision

All modules installed in a single chassis should be running the latest software revision. The build number ID appears after the software revision number.

In case one of the modules (sFB-4700, sLB-4018, sMB, etc) in the chassis does not have the latest software version, you will need to apply a software update for the entire chassis. Refer to the Software Update section for more information.

InfiniScale IV Firmware Revision

The Voltaire switch software runs on the InfiniScale IV (IS4) chip. IS4 firmware revision 7.3.0 is the basic requirement for this version of software.

To update the *Grid Director* firmware, refer to the Firmware Update section for more information.

The specific configuration of each firmware image is specified by the firmware physical parameter ID (PSID). The PSID is an indicator associated with each firmware image.

The following table lists the required firmware PSID per module.

Table 1. Firmware PSID Versions

Board	PSID
4036	VLT1210031100
4036E	VLT1280031100
2036	VLT1220031100
sLB-4018	VLT1240031100
sFB-4700	VLT1250031100
sFB-4200	VLT1270031100
sFB-4700-x2 (IS-4 1)	VLT1260031100
sFB-4700-x2 (IS-4 2)	VLT1260031100

InfiniBand-Ethernet Gateway Firmware Revision

The InfiniBand to Ethernet gateway module in the 4036E runs its own firmware that supports InfiniBand to Ethernet bridging.

Firmware revision 3.0.0_4 is the basic requirement for this version of the software.

Firmware/Software Revisions Verification

To verify the most updated firmware revision, firmware PSID and software revision, run the '`module firmware show`' command. Refer to the *Grid Director 4000 Series User Manual* for a full description of the command. All modules installed in a single *Grid Director* should be running the latest software revision, build ID, and firmware revision.

Software Update

This section describes the procedure for updating the *Grid Director* software revision.

Starting from software version V3.0.0, all chassis software and firmware will be updated to the latest version during software update, including all chassis modules (sFB-4700, sLB-4018, sMB, etc).

On *Grid Directors* running software version prior to V3.0.0, firmware update should be performed manually after the software update. Refer to the Software Update Scenarios section.

If any one of the modules (sFB-4700, sLB-4018, sMB, 4036, 4036E, etc) does not have the latest software or firmware revision, you must update the chassis software as follows:

Before you begin to update the chassis software:

■ Be certain that the new software file, `upgrade_all.tgz`, is located on a remote FTP/SCP server.

■ The server details should be configured at the Grid Director CLI via the 'remote' command. Refer to the *Grid Director 4000 Series User Manual* for more details.

■ The software update can be invoked from a privileged CLI login via the 'update software chassis command.

Software Update Scenarios

This section provides specific software update instructions for different scenarios.

4700/4200 Update from Software Revision 1.0.1/2.5.0

IMPORTANT:

Before upgrading from V2.5.0, verify that SNMP mode is set enabled via the `config->snmp-> snmp mode show` CLI command.

To update the software to 3.0.0:

Step 1 Extract the Voltaire provided upgrade_all.tgz in a directory on the SCP\FTP remote server. In Linux, run: `tar –zxvf upgrade_all.tgz`
The extracted files are `upgrade.tgz, firmware.tar` and `versions.txt`

Step 2 With a single sMB installed, run the 'update software chassis' CLI command via the sMB.

With two sMBs installed, extract the right sMB and run 'update software chassis' CLI command via the left sMB.
After the chassis reboots, extract the left sMB and place the other sMB in the left slot. Wait untill the system boots, and run the 'update software chassis' CLI command via the left sMB.

Note: Throughout the operation, the other sMB will not be installed.
Wait utill the system boots again.
Switch the 2 sMBs back to their original places. The left sMB is now the same module as before the update.

Step 3 Chassis firmware 7.3.0 is a basic requirement for software V3.0.0.
When updating from previous software versions, the chassis firmware should be manually updated following the software update, as described in the "Firmware Update" section below.

Non-sMB Modules with Software other than V3.0.0

When upgrading from sMB with software V3.0.0 where at least one of the non-sMB modules (sLB-4018, sFB-4700, etc) is on a software version other than V3.0.0, the software upgrade should be applied as described in steps 1-2 above.

During the update procedure, the following error message may display:

```
"Upgrade timeout. Some modules failed to upgrade. Upgraded modules
    should be rebooted manually - Upgrade Failure Error occured
    during firmware burning - Upgrade Failure"
```

The above message indicates that the chassis firmware upgrade *did not* complete successfully and that the software update *was* applied.

To complete the upgrade procedure:

Step 1 Execute the 'reset all' command from the 'hardware' menu in the CLI.

Step 2 After the system reboots after the hardware reset, follow 'Step 3' above.

Upgrading from V1.0.1 Only

If you are upgrading from V1.0.1 and you are updating the software via FTP, you can store the new software file in any directory on the FTP server except for the FTP home directory.

2036/4036 Software Update from Version 2.X.X

To update the software to 3.0.0:

Step 1 Extract the Voltaire provided upgrade_all.tgz:
In Linux, run: tar -zxvf upgrade_all.tgz
You will get: upgrade.tgz, firmware.tar and versions.txt files

Step 2 Create symbolic link to the software file upgrade.tgz, to

upgrade_2036_4036.tgz as follows:

    ln -s upgrade.tgz upgrade_2036_4036.tgz

For V2.0.0: Run the update software CLI command
For V2.0.1: Run the update software chassis CLI command.

Step 3 4036/2036 firmware 7.3.0 is a basic requirement for software V3.0.0.
When updating from previous software versions, the chassis firmware should be updated as instructed in the "Firmware Update" section below.

2036/4036 Software Update from Version 1.0.0

If you are upgrading from V1.0.0, perform the upgrade procedure to V2.0.0 as explained in V2.0.0 RN Appendix A, and then perform V3.0 upgrade.

Firmware Update

From Grid Director V3.0.0 and up, the chassis firmware will be updated automatically during software update.

On occasion, the chassis firmware may require a separate update without the software being updated.

To update for the latest firmware only, place the `firmware.tar` file on remote FTP/SCP location and run the `update firmware` command.

The remote FTP/SCP account should be configured in advance and accessible during the update.

Refer to the *Grid Director 4000 Series User Manual* for a description of the `update firmware` command.

Interoperability/Compatibility

The Grid Director 4000 Series v3.0.0 is fully interoperable and compatible with the following products:

- ISR 2012/2004 and sRB-20210G
- OFED 1.4 & 1.5
- HBSM 1.5.2
- UFM 2.1 & 2.2
- ConnectX, ConnectX2 HCAs (DDR + QDR)

For additional details, refer to the interoperability matrix.

Grid Director 4036E Ordering Options

Grid Director 4036E is available in two different ordering options:

- 4036E (VLT-30034): Full functionality
- 4036E-LM (VLT-30035): Low-memory version with the following limitations (by design):
  - › Only up to 4 VLANs/partitions are supported
  - › Only up to 4 multicast groups are supported
  - › No support for VMA (Voltaire Messaging Accelerator)
  - › No support for TFM (Traffic Flow Management).

Known Issues

This section lists the known Issues that are specific for the, Grid Director version 3.0.0.

Table 2. Known Issues

#	Module	Description	Platform
1	sFB-4700-x2 Support	The following information and actions are not available for the HyperScale fabric card (sFB-4700-x2): • External port status and change trap • External port counters • External port enable/disable This information can be obtained from UFM or any external Subnet Manager.	4700
2	Cable info support	Currently, QSFP & CXP information presentation for sLB & sFB-4700-x2 is not supported.	4700, 4200
3	Device Management	sysObjectID MIB variable returns incorrect data for the Grid Director 4036E. This data can be obtained via InfiniBand MADs. This issue will be resolved in a future software release.	4036E
4	System	When inserting the QSFP cable into an InfiniBand port, it could take the physical link up to 20 seconds to come up.	All
5	System	On occasion, a "JFFS2 warning" message might pop up using serial connection. These messages can be safely ignored.	All
6	SM	On setting max_op_vls to `0` or `1`, a single VL, the minimum available with InfiniBand, will be set.	All
7	Performance Management (PM)	When PM is enabled, HCA port data counters are reset each PM polling interval (default 60 sec. range 30-360 sec) to minimize the event of counters overflow. IS4 data counters are not reset since they are reported as 64bit extended counters, not 32bit as in the HCA.	All
8	Fabric Management	4700, 4200, 4036E are not supported by GridVision V5.4 (running on Voltaire's DDR 2012/2004 switches)	4700, 4200, 4036E
9	Fabric Management	IB multicast traffic over 4036E is not supported by GridVision V5.4 (running on Voltaire's DDR 2012/2004 switches)	4036E

#	Module	Description	Platform
10	Fabric Management	All fabric related traps reporting consistent failures are repeated every 1 minute. In future software releases those failures will be reported only once.	2036, 4036, 4036E
11	Fabric Management	The 4700/4200 do not support SNMP traps for fabric changes. This issue will be resolved in a future software release.	4700, 4200
12	IB Diagnostics	Perfquery utility –x option on IS4 chip port will fail unless counters mask 0x00ff is provided (ASIC layer bug).	All
13	IB Diagnostics	ibtracert utility: the '-D' (direct route) option is not working.	All
14	IB Diagnostics	Smpquery -c does not work. This option returns SMP info (such as NodeInfo, PortInfo, etc) with a combined path – LID and Direct Path together.	All
15	4036E Eth ports	The ethernet ports on 4036E currently work only in 10GbE mode. Support for 1GbE will be added in the next software release.	4036E
16	4036E Eth ports	Plugging an optic module in the 4036E ETH port when the port is configured to 'disabled' is not supported. In this case, traffic will not flow even when you set the port back to 'enabled'	4036E
17	Management by UFM	Currently, UFM does not support the switches Reboot, Software/Firmware upgrade and device management functions. Workaround: Perform these actions via SSH.	All
18	4036E Eth ports System/configuration	Changing the LAG mode requires a reboot.	4036E
19	4036E Eth ports Multicast Configuration	In some rare scenarios in which a multicast group is alternately created and immediately destroyed, the 4036E InfiniBand to Ethernet gateway fails to join the group. Workaround: To recover, close and reopen the MC group.	4036E
20	4036E Eth ports PMTU	The PMTU mechanism is not Plug and Play. Workaround: Configure an IP address in the subnet for the appropriate interface on the 10G ports.	4036E

#	Module	Description	Platform
21	4036E Eth ports PMTU	Some network issues may arise when sending heavy traffic with oversized packets from a host that does not support PMTU. When working in non-homogeneous MTU environment, make sure all hosts which communicate through the 4036E gateway support the PMTU mechanism.	4036E
22	4036E Eth ports	Several 4036E GATEWAY features, such as PMTU, routing rules, and handling broadcast traffic – require a configuration of the interface's IP address. Workaround: set an IP address for every interface that is configured in the 4036E GATEWAY	4036E
23	4036E Eth ports	In some scenarios, after a Windows-based IB host (that runs WinOF) restarts, it is unable to communicate with the Ethernet side via the 4036E GATEWAY. Workaround:Ddisable the IPoIB interface and then enable it again.	4036E
24	4036E Eth ports loop-prevention configuration	Changing the "Full" loop-prevention mode when several 4036E GATEWAY units are part of the same Layer-2 domain – may cause temporary loops that may affect network stability. Workaround: Disable the Ethernet ports of the 4036E GATEWAY before changing the loop-prevention mode.	4036E
25	4036E Eth ports	In cases when the IB subnet prefix is changed (e.g. – working with different GID), the 4036E GATEWAY has to be rebooted.	4036E
26	4036E Eth ports	When connecting several 4036E GATEWAY directly to each other through their Ethernet ports, the IP broadcast traffic may be discarded. The problem occurs when the HA priority of the directly-connected 4036E GATEWAY is not synchronized. Workaround: For a stable and resilient fabric, configure a different HA priority for each pair of 4036E GATEWAY, and make sure that 4036E GATEWAY units that are directly connected have the same priority.	4036E
27	4036E Eth ports TFM	The TFM rule always applies to Unicast traffic.	4036E

#	Module	Description	Platform
28	4036E Eth ports	Calculation of the "Tx total packets" counter is erroneous and does not reflect the total packets that actually arrived on the Ethernet ports. Workaround: For the accurate counter, refer to type specific counters, such as ARP, ICMP, TCP, etc.	4036E
29	High Availability	If the remote logger was configured on the right-side sMB that served as the active sMB, while the left-side sMB was plugged into the chassis – the sMB handover occurs. In this case, it takes up to 5 minutes for the remote logger to resume its functionality on the left-side sMB.	4700, 4200

Resolved Issues

This section lists the resolved issues for the version 3.0.0.

Table 3. Resolved Issues

#	Module	Description	Platform
1	Info LED	The Info LED is only supported on the Management Board and is not supported on line/fabric cards.	4700
2	Chassis Management	There is no SNMP trap when the sFU rate changes. An error message is reported to the Log file.	4700
3	Management Interface IP Configuration	When issuing an `ip-address show` CLI command, the internal-subnet-prefix erroneously displays 10.10.10.0/16 instead of 10.10.0.0/16.	4700
4	Management by Unified Fabric Manager (UFM)	The default gateway must be configured in order for the UFM to manage the 2036/4036 properly. If not already configured, the switch IP address should be set as the default Gateway IP	All
5	Management	The Congestion Control Management (CCM) statistics link on the switch web page does not work.	4036/2036
6	Linear Forwarding Table (LFT)	The Linear Forwarding Table should be initialized to 0xff	All

Product Documentation

This section lists the product documentation for the Grid Director 4000 Series version 3.0.0.

Table 4. Product Documentation

Document Title	Description	Part Number
Grid Director 4000 Series Getting Started Guide	Provides basic hardware and software instructions on how to install your system.	LIT-00037
Grid Director 4700 Installation Manual	Hardware documentation for the Grid Director 4700,	DOC-00726
Grid Director 4200 Installation Manual	Hardware documetnation for the Grid Director 4200.	DOC-00912
Grid Director 4036E Installation Manual	Hardware documentation for the Grid Director 4036E.	DOC-00911
Grid Director 4036/2036 Installation Manual	Hardware documentation for the Grid Director 4036/2036	DOC-00467
Grid Director 4000 Series User Manual	Software documentation for the Grid Director 4000 Series.	DOC-00932
Regulatory Compliance and Safety Information Reference Guide	Compliance and safety information guidelines.	DOC-00859

Please send your documentation-related comments and feedback to: docs@voltaire.com.

Submitting a Service Request

The Voltaire Technical Assistance Center (TAC) is at your service. You may access Warranty Service through our Web Request Form by using the following link: http://www.voltaire.com/support