e1350 November Hints and Tips:

Symptom:
Cisco OFED on SLES 10 SP1 PPC64 with Cisco_OFED-1.2.5-fcs.iso does not install kernel modules correctly.

Explanation:
This problem occurs because the Cisco ISO includes a duplicate set of 2.6.16.21-0.8 IB kernel modules in the base directory. All RPMs in this directory are installed with --force --nodeps by the ofedinstall script, which causes the wrong version of the kernel modules to be installed.

Action:
Remove the kernel-ib rpm packages from <iso root>/sles10/ppc64 as follows:

```
mkdir /tmp/cisco_ib
mount -o loop Cisco_OFED-1.2.5-fcs.iso /tmp/cisco_ib
cd /tmp/cisco_ib/sles10/ppc64
rm kernel-ib-*
cd /tmp/cisco_ib
./ofedinstall
```

The relevant line from the ofedinstall script is:

```
535  rpm -Uvh --force --nodeps $RPM_ROOT_DIR/$DISTKW/$BASE_ARCH/*.rpm
```

This line causes the forced update of all rpms in the directory, including the duplicated kernel-ib rpms, overwriting the drivers that are supposed to be there. Editing the ofedinstall script is not recommended.

Symptom:
ConnectX blade daughtercards connected through an InfiniBand High-speed Pass-Through Module to either a Cisco 7012D or Cisco 7024D switch do not link at DDR speed.

Explanation:
The currently released firmware for the Cisco 7012D and 7024D switches do not support DDR with the ConnectX blade daughtercards.

Action:
The Cisco 7000D switch does not have this issue, so it can be used as a leaf switch, uplinking to Cisco 7012D/7024D switches at the core of the InfiniBand fabric.

Updated firmware for the Cisco 7012D and 7024D switches is planned to be released soon to address this issue.

Symptom:
The OFED-1.2.5.1 installation script aborts when building the mpitests rpm.

Explanation:
Building the included mpitests rpm in the OFED 1.2.5.1 stack is currently not supported.

Action:
When running the OFED 1.2.5.1 installation script, choose the "Basic" option to avoid compiling mpitests. Afterwards, for MPI functionality, build MPI separately using OpenMPI 1.2.3 from http://www.open-mpi.org.

Symptom:
When starting an MPI job using ConnectX HCAs and OpenMPI, "WARNING: No HCA parameters were found for the HCA that Open MPI detected" is reported.

Explanation:
OpenMPI version 1.2.3 was released before the ConnectX HCAs and thus does not contain the information in the *.ini file for them.

Action:
This is not a critical error and is just a warning. It can be ignored but may effect performance. The correct settings are as follows,

```
# A.k.a. ConnectX
[Mellanox Hermon]
vendor_id = 0x2c9,0x5ad,0x66a,0x8f1,0x1708
vendor_part_id = 25408,25418,25428
use_eager_rdma = 1
mtu = 2048
```

which should be added to the
<openmpi root>/share/openmpi/mca-btl-openib-hca-params.ini file.

Symptom:
Updating the BIOS or BMC firmware using lflash under RHEL 5 fails on some systems with "tail: cannot open `+49' for reading: No such file or directory" reported.

Explanation:
The lflash binary is is comprised of a shell script with the actual firmware code appended. The script script portion of the file calls "tail" with the syntax "tail +<n>", which is deprecated, and in fact not supported in RHEL 5.

Action:
Either use hexedit to modify the lflash binary and change the tail command from:

```
tail +$SKIP $0
```

to

```
tail -n +49 $0
```

or move the lflash file you are trying to run to oldfile.sh, and run the following commands:

```
head -n 48 oldfile.sh > newfile.sh
sed -ei s/"+$SKIP"/"-n +49"/g newfile.sh
tail -n +49 oldfile.sh >> newfile.sh
chmod 755 newfile.sh
```

Then run newfile.sh to update the system BIOS or BMC firmware as follows:

```
./newfile.sh -s
```

Symptom:
GotoBLAS version 1.12 or greater fails to compile with "copy_sse_core2.S:241: Error: no such instruction: `palignr $4,%xmm0,%xmm1'" on SLES 10 x86_64.

Explanation:
This problem is caused by an incompatible binutils package released with SLES 10 x86_64, as well as with SLES 10 SP1 x86_64.

Action:
While in the base build directory for GotoBLAS run the following commands:

```
cp level1/copy/x86_64/copy_sse.S level1/copy/x86_64/copy_sse_core2.S
cp level1/copy/x86_64/zcopy_sse.S level1/copy/x86_64/zcopy_sse_core2.S
cp level1/dot/x86_64/dot_sse.S level1/dot/x86_64/dot_sse_core2.S
```

Then re-run quickbuild.64bit.

Symptom:
The driver for the MegaRAID 8480 adapter fails to initialize when option ROM execution is disabled in the system BIOS for the slot in which the MegaRAID adapter is installed.

Explanation:
The adapter's firmware does not initialize properly unless its option ROM is allowed to run during system POST.

Action:
Enable option ROM execution in the system BIOS for the slot in which the MegaRAID 8480 adapter is installed.

Symptom:
When building OpenMPI version 1.2.3 on SLES 10 SP1 PPC64, linking fails with "libstdc++.so: could not read symbols: File in wrong format" reported.

Explanation:
The build files in OpenMPI 1.2.3 reference

```
/usr/lib/gcc/powerpc64-suse-linux/4.1.2/libstdc++.so
```

rather than

```
/usr/lib/gcc/powerpc64-suse-linux/4.1.2/64/libstdc++.so.
```

Action:
Change the symbolic link in /usr/lib/gcc/powerpc64-suse-linux/4.1.2 from:

```
/usr/lib/gcc/powerpc64-suse-linux/4.1.2/libstdc++.so --> /usr/lib/libstdc++.so
```

to:

/usr/lib/gcc/powerpc64-suse-linux/4.1.2/libstdc++.so --> /usr/lib64/libstdc++.so

In order to make the changes, do the following:

```
mv /usr/lib/gcc/powerpc64-suse-linux/4.1.2/libstdc++.so
/usr/lib/gcc/powerpc64-suse-linux/4.1.2/libstdc++.so.old
ln -s /usr/lib/gcc/powerpc64-suse-linux/4.1.2/libstdc++.so
/usr/lib64/libstdc++.so
```

Symptom:
After building OpenMPI 1.2.3 on a system with Myrinet MX installed, the resulting OpenMPI code does not contain MX support.

Explanation:
The OpenMPI source RPM requires special parameters in order to build MX support.

Action:
Install the OpenMPI 1.2.3 source RPM, run

```
cd /usr/src/<pkgdir>/SPECS
```

where <pkgdir> is "redhat" for RHEL 5 and "packages" for SLES 10 SP1, then run

```
rpmbuild -bb --define 'configure_options --with-mx=/opt/mx
--with-libdir=/opt/mx/lib64' openmpi-1.2.3.spec
```

Symptom:
Warewulf diskless systems with more than two interfaces exhibit network connectivity problems.

Explanation:
The configuration for network interfaces for a diskless node managed by xCAT and Warewulf is passed on the kernel command line. The linux kernel only supports up to 256 characters on the kernel command line, any additional characters are truncated. This means that those additional characters cannot be accessed by the Warewulf startup scripts. In practice, given a typical kernel command line in the xCAT/Warewulf environment, the truncation of characters occurs on the third network interface. This can have the result that that interface comes up, but does not have the correct subnet mask. In a network environment where this subnet mask resolves to a network address appearing to be accessible locally, rather than having to go through a router, then that network would be unreachable from the diskless node. For example, if the network on the third interface comes up as 10.0.0.0/8 rather than its intended address of 10.y.0.0/16, and one or more of the first two interfaces is on a 10.x.0.0/16 network, and a real 10.0.0.0/16 network does exist (accessible through a router on one of the local networks), then when the diskless node attempts to send a packet to an address of the form 10.0.n.m, it will attempt to do so through the third network interface without going through a router, which will fail.

Action:
  Because this truncation of the kernel command line in an xCAT/Warewulf
  environment will practically always occur in the clause with the settings
  for the third network interface, only two network interfaces should be
  specified on the kernel command line.  Setup of additional network
  interfaces on Warewulf diskless nodes should be done by other means
  (modifying the network interface configuration files/startup scripts in
  the diskless image as appropriate).  To keep the third and higher network
  interfaces from being misconfigured, modify the appropriate pxelinux
  configuration file (normally /tftpboot/pxelinux.cfg/<nodename>) on the
  appropriate management or staging node from (the middle portion of the
  APPEND line in the pxelinux configuration file has been abbreviated with
  elipses below), changing it from the form:

    APPEND ...
warewulf=l5,10.16.0.2,10.16.0.1,compute_x86_64,eth0=10.16.24.5:255.255.0.0,eth1=10.1
7.24.5:255.255.0.0,eth2=10.26.24.5:255.255.0.0,eth3=10.27.24.5:255.255.0.0

  to:

    APPEND ...
warewulf=l5,10.16.0.2,10.16.0.1,compute_x86_64,eth0=10.16.24.5:255.255.0.0,eth1=10.1
7.24.5:255.255.0.0

  removing all network interfaces definititions that cannot be completely
  specified within the 256 character kernel command line limit.

Symptom:
  The network installation source is not setup properly when running copycds
  using SLES 10 physical media on xCAT 1.3.0.

Explanation:
  The NUMISO variable needs to be set to make the CD1 directory, this doesn't
  happen when using physical media.

Action:
  The following patches address this issue:

    /opt/xcat/lib/copyds/SLES10-GA.sniff

    Index: SLES10-GA.sniff
    ===================================================================
    --- SLES10-GA.sniff (revision 829)
    +++ SLES10-GA.sniff (working copy)
    @@ -12,6 +12,7 @@
    then
    if grep "^VERSION 10" content > /dev/null
    then
    + NUMISO=$(tail -n 1 media.1/media)
    CDT="SLES10-$MYARCH"
    echo "Detected SLES10-$MYARCH"
    return 0

    /opt/xcat/lib/copyds/SLES10-GA.copycd

    Index: SLES10-GA.copycd
    ===================================================================

```
--- SLES10-GA.copycd (revision 829)
+++ SLES10-GA.copycd (working copy)
@@ -1,8 +1,8 @@
function ccd {
-if [ $ISO == 0 ]; then
+#if [ $ISO == 0 ]; then
#This only works for ISOs
- return 1;
-fi
+# return 1;
+#fi
OSVER=sles10
ARCH=`echo $CDT | awk -F- '{print $2}'`
SP=`echo $CDT | awk -F- '{print $3}'`
```

Another option is to use the ISO files, passing them as command line arguments to copycds.