目 录

1 概述

本文主要介绍 H3C Uniserver B16000 刀箱内 IB 交换模块（BX1020B）与 IB 网卡（IB1040i）的配置方法及常见问题处理。

由于 IB 交换模块没有管理串口，管理员需要 IB 网卡配合才能管理 IB 交换模块。因此对于 IB 交换模块的管理操作，需要先将 IB 交换模块和带有 IB 网卡的刀片服务器安装到刀箱中，然后在刀片服务器的 Linux 操作系统下进行（本文以服务器安装 Redhat 7.5 系统为例）。集群（包含 IB 网卡和 IB 交换模块在内的整个刀箱环境）中只需要有一台满足要求的刀片服务器，即可管理所有 IB 交换模块。

刀箱 IB 配置是基于 IB 集群化的概念，通过启用 SM（子网管理器，后面章节有详细描述），来统一管理当前集群下的所有 IB 设备（本文指 IB 交换模块和 IB 网卡）。IB 集群分为物理层和逻辑层，刀箱内 IB 交换模块与 IB 网卡均在位连接时，物理层为 linkup 状态，只有启用 SM，逻辑层链路才能正常使用。

2 配置准备

本章节主要介绍了配置子网管理器的方法，子网管理器是将整个 IB 集群的设备进行统一管理，运行在刀箱内的一台安装了 IB 网卡的刀片服务器上，在配置子网管理器之前， 需要在当前的服务器上对 IB 网卡进行驱动升级，及 MFT 工具包进行安装，MFT 是一套固件管理工具，用于生产标准化或定制化 Mellanox 固件镜像、查询固件信息、烧录固件镜像。为了保证子网管理器正常使用，需要确认 IB 交换模块和 IB 网卡正常物理连接。根据业务需求可给 IB 网卡端口设置 IP 地址。

2.1 IB网卡驱动下载（IB1040i）

请联系技术支持，根据版本配套表下载 IB 网卡的驱动安装包。

说明

IB 网卡的驱动安装包的文件名请以实际情况为准，本文中以“MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64.tgz”举例。

2.2 IB网卡驱动安装

(1) 使用 ftp 或者 SSH 工具将驱动包上传到节点服务器的/home 目录下。

```
[root@wlp-node8 ~]# cd /home/
[root@wlp-node8 home]# ls
hpcx-v2.5.0-gcc-MLNX_OFED_LINUX-4.7-1.0.0.1-redhat7.5-x86_64
hpcx-v2.5.0-gcc-MLNX_OFED_LINUX-4.7-1.0.0.1-redhat7.5-x86_64.tbz
mft-4.13.0-104-x86_64-rpm.tgz
MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64
MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64.tgz
test128
```

(2) 在当前/home 路径下对 IB 网卡进行解压：

参考命令：**tar –xvf** *MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64.tgz*

```
[root@wlp-node8 home]# tar -xvf MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64.tgz
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/src/
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/src/MLNX_OFED_SRC-4.7-1.0.0.1.tgz
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/RPMS/
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/RPMS/MLNX_LIBS/
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/RPMS/MLNX_LIBS/infiniband-diags-gu
1.47100.x86_64.rpm
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/RPMS/MLNX_LIBS/ibutils2-2.1.1-0.11
_64.rpm
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/RPMS/MLNX_LIBS/ibacm-devel-41mlnx1
./MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/RPMS/MLNX_LIBS/ibutils-1.5.7.1-0.1
```

(3) 进入 MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64/目录下，执行命令：**./mlnxofedinstall**，进行 IB 网卡的驱动安装。

```
[root@wlp-node9 MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64]# ./mlnxofedinstall
Logs dir: /tmp/MLNX_OFED_LINUX.26484.logs
General log file: /tmp/MLNX_OFED_LINUX.26484.logs/general.log
Verifying KMP rpms compatibility with target kernel...
This program will install the MLNX_OFED_LINUX package on your machine.
Note that all other Mellanox, OEM, OFED, RDMA or Distribution IB packages will be removed.
Those packages are removed due to conflicts with MLNX_OFED_LINUX, do not reinstall them.

Do you want to continue?[y/N]:y

Uninstalling the previous version of MLNX_OFED_LINUX

rpm --nosignature -e --allmatches --nodeps mft-oem mft mft-oem

Starting MLNX_OFED_LINUX-4.7-1.0.0.1 installation ...

Installing mlnx-ofa_kernel RPM
Preparing...                          ########################################
Updating / installing...
mlnx-ofa_kernel-4.7-OFED.4.7.1.0.0.1.g########################################
Installing kmod-mlnx-ofa_kernel 4.7 RPM
Preparing...                          ########################################
kmod-mlnx-ofa_kernel-4.7-OFED.4.7.1.0.########################################
Installing mlnx-ofa_kernel-devel RPM
Preparing...                          ########################################
Updating / installing...
mlnx-ofa_kernel-devel-4.7-OFED.4.7.1.0########################################
```

(4) 驱动安装成功后，重启服务器。

2.3 网络配置

安装完成 IB 网卡驱动并重启，需要为 IB 端口配置一个 IP 地址，配置方法和以太网端口的 IP 配置方法一样，步骤如下：

(1) 在 IB 网卡所在的刀片服务器的/etc/sysconfig/network-scripts/下修改配置文件 ifcfg-ib0，配置 IP 地址和子网掩码。

```
[root@wlp-node8 network-scripts]# cat /etc/sysconfig/network-scripts/ifcfg-ib0
CONNECTED_MODE=no
TYPE=InfiniBand
PROXY_METHOD=none
BROWSER_ONLY=no
BOOTPROTO=static
IPADDR=15.15.15.15
NETMASK=255.255.255.0
DEFROUTE=yes
IPV4_FAILURE_FATAL=no
IPV6INIT=yes
IPV6_AUTOCONF=yes
IPV6_DEFROUTE=yes
IPV6_FAILURE_FATAL=no
IPV6_ADDR_GEN_MODE=stable-privacy
NAME=ib0
UUID=8b59e617-d027-4ae6-938a-da5792ca8934
DEVICE=ib0
ONBOOT=yes
```

(2) 配置完成后重启网络服务使 IP 生效，执行命令：**systemctl restart network.service**

```
[root@wlp-node8 ~]# systemctl restart network.service
[root@wlp-node8 ~]# ifconfig ib0
ib0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 4092
        inet 15.15.15.15  netmask 255.255.255.0  broadcast 15.15.15.255
        inet6 fe80::8198:f6d9:3e24:3b29  prefixlen 64  scopeid 0x20<link>
Infiniband hardware address can be incorrect! Please read BUGS section in ifconfig(8).
        infiniband 20:00:07:C4:FE:80:00:00:00:00:00:00:00:00:00:00:00:00:00:00  txqueuelen 256  (InfiniBand)
        RX packets 181  bytes 61745 (60.2 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 106  bytes 14285 (13.9 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

(3) 配置子网管理器

IB 网络使用子网管理器(Subnet Manager，下文简称 SM)管理网络路由，SM 可以运行在刀片服务器或者具有管理功能的 IB 交换模块上（本文介绍运行在服务器节点上）。SM 启用时，会统一将集群中的设备根据 GUID（globally unique identifier，全局唯一标识符）进行 LID 号（Local Identifier，端口标识符）分配，LID 号是唯一的，无特殊情况下不进行回收。

MLNX_OFED（官方驱动名称）驱动中集成了子网管理器 SM，安装 MLNX_OFED 驱动后，SM 已默认安装。SM 可以运行在一台或者多台服务器节点上，但同时只有一台处于 Active 状态，为了确保当前集群的正常运行。当主的 SM 处于 down 的状态，备的 SM 会接替主的 SM。主备切换间隙暂不会影响业务，但当 SM 未启用时，集群内的所有的 IB 网卡端口的逻辑层会处于初始化状态，会影响业务运行。

(4) 启用子网管理器

在服务器节点上开启子网管理功能

- 方法一：执行命令：**/etc/init.d/opensmd start**

```
[root@wlp-node10 ~]# /etc/init.d/opensmd start
Starting opensmd (via systemctl):                          [  OK  ]
[root@wlp-node10 ~]# /etc/init.d/opensmd status
opensm (pid 7821) is running...
```

此方法会默认将当前服务器上第一个 IB 网卡 Active 的端口作为 SM 管理口，如果集群中需要备的 SM，可以在集群中其他服务器中执行**/etc/init.d/opensm start**。

- 方法二：执行 **opensm –B –g** <*GUID*> **-p** <*sm_priority*>命令为 IB 网卡端口设置优先级，优先级最高且处于 Active 状态的端口将被选择为主 SM，其余端口为备 SM。其中<*GUID*>表示 IB 网卡的 GUID 号，可以通过 **ibstat** 命令查询；<*sm_priority*>表示 SM 的优先级，取值范围为 0～14，数值越大优先级越高。

```
[root@wlp-node8 ~]# opensm  -B -g 0xe41d2d0030570ceb -p 10
-------------------------------------------------
OpenSM 5.5.0.MLNX20190923.1c78385
Command Line Arguments:
 Daemon mode
 Guid <0xe41d2d0030570ceb>
 Priority = 10
 Log File: /var/log/opensm.log
-------------------------------------------------
[root@wlp-node8 ~]# /etc/init.d/opensmd status
opensm (pid 298713) is running...
```

2.4 MFT工具下载

MFT 是一套固件管理工具，用于生产标准化或定制化 Mellanox 固件镜像、查询固件信息、烧录固件镜像。请自行上网获取 MFT 工具。

说明

MFT 工具包的文件名请以实际情况为准，本文中以“mft-4.13.0-104-x86_64-rpm.tgz”举例。

2.5 MFT工具安装

(1) 通过 FTP 或者 SSH 工具将 MFT 工具包上传到刀片服务器的/home 目录下

```
[root@wlp-node8 ~]# cd /home/
[root@wlp-node8 home]# ls
hpcx-v2.5.0-gcc-MLNX_OFED_LINUX-4.7-1.0.0.1-redhat7.5-x86_64
hpcx-v2.5.0-gcc-MLNX_OFED_LINUX-4.7-1.0.0.1-redhat7.5-x86_64.tbz
mft-4.13.0-104-x86_64-rpm.tgz
MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64
MLNX_OFED_LINUX-4.7-1.0.0.1-rhel7.5-x86_64.tgz
test128
```

(2) 执行命令：**tar -xvf mft-4.13.0-104-x86_64-rpm.tgz** 解压 IB 通用的 MFT 工具包。

```
[root@wlp-node8 home]# tar -xvf mft-4.13.0-104-x86_64-rpm.tgz
mft-4.13.0-104-x86_64-rpm/
mft-4.13.0-104-x86_64-rpm/RPMS/
mft-4.13.0-104-x86_64-rpm/RPMS/mft-4.13.0-104.x86_64.rpm
mft-4.13.0-104-x86_64-rpm/RPMS/mft-oem-4.13.0-104.x86_64.rpm
mft-4.13.0-104-x86_64-rpm/SRPMS/
mft-4.13.0-104-x86_64-rpm/SRPMS/kernel-mft-4.13.0-104.src.rpm
mft-4.13.0-104-x86_64-rpm/install.sh
mft-4.13.0-104-x86_64-rpm/uninstall.sh
mft-4.13.0-104-x86_64-rpm/old-mft-uninstall.sh
mft-4.13.0-104-x86_64-rpm/LICENSE.txt
```

(3) 进入/home/ mft-4.13.0-104-x86_64-rpm /，执行命令：**./install.sh --oem**，安装 MFT 工具

```
[root@wlp-node8 home]# cd mft-4.13.0-104-x86_64-rpm/
[root@wlp-node8 mft-4.13.0-104-x86_64-rpm]# ls
install.sh  LICENSE.txt  old-mft-uninstall.sh  RPMS  SRPMS  uninstall.sh
[root@wlp-node8 mft-4.13.0-104-x86_64-rpm]# ./install.sh --oem
-I- Removing all installed mft packages: mft
-I- Removing the packages:  kmod-kernel-mft-mlnx-4.13.0-1.rhel7u5.x86_64...
-I- Building the MFT kernel binary RPM...
-I- Installing the MFT RPMs...
Preparing...                          ################################# [100%]
Updating / installing...
   1:kernel-mft-4.13.0-3.10.0_862.el7.################################# [100%]
Preparing...                          ################################# [100%]
Updating / installing...
   1:mft-4.13.0-104                   ################################# [100%]
Preparing...                          ################################# [100%]
Updating / installing...
   1:mft-oem-4.13.0-104               ################################# [100%]
-I- In order to start mst, please run "mst start".
```

(4) MFT 工具安装成功后，执行命令：**mst start**，确认工具已安装成功，并且能运行。

```
[root@wlp-node8 mft-4.13.0-104-x86_64-rpm]# mst start
Starting MST (Mellanox Software Tools) driver set
Loading MST PCI module - Success
Loading MST PCI configuration module - Success
Create devices
Unloading MST PCI module (unused) - Success
```

2.6 IB设备管理

子网管理器设置成功后，就可以对当前集群下的所有的 IB 设备进行统一管理。

(1) 节点服务器启动 mst 服务，执行命令：**mst start**。如已执行请跳过。

```
[root@wlp-node10 ~]# mst start
Starting MST (Mellanox Software Tools) driver set
Loading MST PCI module - Success
[warn] mst_pciconf is already loaded, skipping
Create devices
Unloading MST PCI module (unused) - Success
[root@wlp-node10 ~]#
```

(2) 自动将集群中的所有 IB 设备加入子网管理器中统一管理，执行命令：**mst ib add**

(3) 查看集群中添加的 IB 设备，命令 **mst status –v**（截图中的 Inband devices 表示已添加的设备）

```
[root@wlp-node9 ~]# mst status -v
MST modules:
------------
    MST PCI module is not loaded
    MST PCI configuration module loaded
PCI devices:
------------
DEVICE_TYPE             MST                           PCI       RDMA            NET                     NUMA
ConnectX5(rev:0)        /dev/mst/mt4119_pciconf0.1    18:00.1   mlx5_1          net-ib1                 0

ConnectX5(rev:0)        /dev/mst/mt4119_pciconf0      18:00.0   mlx5_0          net-ib0                 0

Inband devices:
---------------
/dev/mst/CA_MT4119_MT4119_ConnectX5___Mellanox_Technologies_lid-0x000A
/dev/mst/CA_MT4119_MT4119_ConnectX5___Mellanox_Technologies_lid-0x0006
/dev/mst/CA_MT4119_wlp-node8_HCA-1_lid-0x0004
/dev/mst/CA_MT4119_wlp-node8_HCA-3_lid-0x0008
/dev/mst/CA_MT4119_wlp-node9_HCA-2_lid-0x0007
/dev/mst/CA_MT4119_wlp-node8_HCA-2_lid-0x0005
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x000E
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0017
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0014
/dev/mst/CA_MT4119_wlp-node10_HCA-2_lid-0x0002
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001
```

(4) 设备加入集群中后可以对当前的设备进行升级、查询，具体使用方法可以参考 3.11 IB 交换模块和 IB 网卡固件升级。

3 IB 网卡常用命令(IB1040i)

3.1 查看网卡端口状态

本命令用于查看 IB 网卡的基本信息。

【命令】

ibstat

【举例】

```
[root@wlp-node9 ~]# ibstat
CA 'mlx5_0'
        CA type: MT4119
        Number of ports: 1
        Firmware version: 16.25.8000
        Hardware version: 0
        Node GUID: 0xe41d2d0030570ce6
        System image GUID: 0xe41d2d0030570ce6
        Port 1:
                State: Down
                Physical state: Polling
                Rate: 10
                Base lid: 65535
                LMC: 0
                SM lid: 0
                Capability mask: 0x2651e848
                Port GUID: 0xe41d2d0030570ce6
                Link layer: InfiniBand
CA 'mlx5_1'
        CA type: MT4119
        Number of ports: 1
        Firmware version: 16.25.8000
        Hardware version: 0
        Node GUID: 0xe41d2d0030570ce7
        System image GUID: 0xe41d2d0030570ce6
        Port 1:
                State: Active
                Physical state: LinkUp
                Rate: 100
                Base lid: 7
                LMC: 0
                SM lid: 5
                Capability mask: 0x2651e848
                Port GUID: 0xe41d2d0030570ce7
                Link_layer: InfiniBand
```

【信息说明】

CA ‘mlx5_0’：网卡端口名称，用于指定端口打流时多带的参数。

Firmware version：当前网卡的固件版本。

State：逻辑层网卡连接状态。

Physical state:物理层网卡连接状态。

Rate：速率。

Base lid：SM 分配的 LID 号。

SM lid：当前 SM 的 LID 号。

Port GUID：网卡每个端口的 GUID 号，可用来指定 SM。

3.2 查看网卡对应端口名称及状态

本命令用于查看当前系统下网卡名称与系统下 ib 端口名称的对应关系及状态。

【命令】

ibdev2netdev

ibdev2netdev -v

【举例】

```
[root@wlp-node9 ~]# ibdev2netdev
mlx5_0 port 1 ==> ib0 (Down)
mlx5_1 port 1 ==> ib1 (Up)
[root@wlp-node9 ~]# ibdev2netdev -v
0000:18:00.0 mlx5_0 (MT4119 -            )                fw 16.25.8000 port 1 (DOWN  ) ==> ib0 (Down)
0000:18:00.1 mlx5_1 (MT4119 -            )                fw 16.25.8000 port 1 (ACTIVE) ==> ib1 (Up)
```

3.3 查看集群中所有的IB网卡设备

本命令用于将集群中所有的 IB 网卡设备进行罗列。

【命令】

ibhosts

【举例】

```
[root@wlp-node9 ~]# ibhosts
Ca      : 0x0004c903003ef608 ports 1 "Mellanox Technologies Aggregation Node"
Ca      : 0xe41d2d0030570ce8 ports 1 "MT4119 ConnectX5    Mellanox Technologies"
Ca      : 0xe41d2d0030570cea ports 1 "wlp-node8 HCA-1"
Ca      : 0xe41d2d0030570ceb ports 1 "wlp-node8 HCA-2"
Ca      : 0x0004c903002ef610 ports 1 "Mellanox Technologies Aggregation Node"
Ca      : 0x506b4b0300f52050 ports 1 "wlp-node8 HCA-3"
Ca      : 0x0004c903002ef612 ports 1 "Mellanox Technologies Aggregation Node"
Ca      : 0xe41d2d0300570fc5 ports 1 "localhost mlx5_0"
Ca      : 0xe41d2d0030570ce7 ports 1 "wlp-node9 HCA-2"
```

【信息说明】

0xe41d2d0030570ce6 ：IB 网卡的端口 GUID 号。

wlp-node8 HCA-1：服务器系统下 IB 网卡的名称。

3.4 查看网卡端口的详细信息

本命令用于查看网卡端口的详细信息，与 ibstat 功能类似，多了 PSID（board_id）信息。

【命令】

ibv_devinfo

【举例】

```
[root@wlp-node9 ~]# ibv_devinfo
hca_id: mlx5_1
        transport:                      InfiniBand (0)
        fw_ver:                         16.25.8000
        node_guid:                      e41d:2d00:3057:0ce7
        sys_image_guid:                 e41d:2d00:3057:0ce6
        vendor_id:                      0x02c9
        vendor_part_id:                 4119
        hw_ver:                         0x0
        board_id:                       H3C0000000008
        phys_port_cnt:                  1
        Device ports:
                port:   1
                        state:                  PORT_ACTIVE (4)
                        max_mtu:                4096 (5)
                        active_mtu:             4096 (5)
                        sm_lid:                 5
                        port_lid:               7
                        port_lmc:               0x00
                        link_layer:             InfiniBand

hca_id: mlx5_0
        transport:                      InfiniBand (0)
        fw_ver:                         16.25.8000
        node_guid:                      e41d:2d00:3057:0ce6
        sys_image_guid:                 e41d:2d00:3057:0ce6
        vendor_id:                      0x02c9
        vendor_part_id:                 4119
        hw_ver:                         0x0
        board_id:                       H3C0000000008
        phys_port_cnt:                  1
        Device ports:
                port:   1
                        state:                  PORT_DOWN (1)
                        max_mtu:                4096 (5)
                        active_mtu:             4096 (5)
                        sm_lid:                 0
                        port_lid:               65535
                        port_lmc:               0x00
                        link_layer:             InfiniBand
```

【信息说明】

board_id：PSID 厂家信息编码。

3.5 查看服务器下网卡端口的GUID号

本命令用于查看服务器系统下网卡的 Node GUID 号。

【命令】

ibv_devices

【举例】

```
[root@wlp-node9 ~]# ibv_devices
    device                 node GUID
    ------              ----------------
    mlx5_1              e41d2d0030570ce7
    mlx5_0              e41d2d0030570ce6
```

3.6 查看当前SM的运行的guid号

本命令用于查询当前集群下 SM 的 GUID 号，表明当前 SM 设备已经被指定。

【命令】

sminfo

【举例】

```
[root@wlp-node9 ~]# sminfo
sminfo: sm lid 13 sm_guid 0xe41d2d0030570ce6, activity count 4057 priority 0 state 3 SMINFO_MASTER
```

3.7 启用mst服务功能

本命令用于集群内开始 mst 服务，该命令用于启动注册访问程序，列出可用 mst 设备。

【命令】

mst start

【举例】

```
[root@wlp-node9 ~]# mst start
Starting MST (Mellanox Software Tools) driver set
Loading MST PCI module - Success
[warn] mst_pciconf is already loaded, skipping
Create devices
Unloading MST PCI module (unused) - Success
```

3.8 IB设备加入集群

本命令用于将集群中所有的 IB 设备加入到当前的集群中，进行统一管理。

【命令】

mst ib add

【举例】

```
[root@wlp-node9 ~]# mst ib add
-I- Discovering the fabric - Running: ibdiagnet -skip all
-I- Added 6 in-band devices
```

【信息说明】

Added 6 ib-band devices：当前集群中可以加入的设备数量（包含交换 IB 交换模块和 IB 网卡）。

3.9 查看当前集群内的所有设备

本命令用于查看集群中所有 IB 设备。

【命令】

mst status –v

【举例】

```
[root@wlp-node9 ~]# mst status -v
MST modules:
------------
    MST PCI module is not loaded
    MST PCI configuration module loaded
PCI devices:
------------
DEVICE_TYPE             MST                           PCI       RDMA            NET                       NUMA
ConnectX5(rev:0)        /dev/mst/mt4119_pciconf0.1    18:00.1   mlx5_1          net-ib1                   0

ConnectX5(rev:0)        /dev/mst/mt4119_pciconf0      18:00.0   mlx5_0          net-ib0                   0

Inband devices:
---------------
/dev/mst/CA_MT4119_wlp-node9_HCA-1_lid-0x000D
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x000E
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0017
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0014
/dev/mst/CA_MT4119_wlp-node10_HCA-1_lid-0x0003
/dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001
```

【信息说明】

DEVICE_TYPE：IB 网卡设备的型号、设备全称、PCI 号、相关联的 CPU。

Inband devices：此处查看设备分配的 lid 号为 16 进制，需要转化成十进制。

3.10 IB交换模块和IB网卡设备信息查询

本命令用于对当前集群下的设备进行管理使用，查询网卡或者 IB 交换模块的基本信息。

【命令】

flint -d <*IB device*> **query**

【参数说明】

<*IB device*>：表示 IB 交换模块和 IB 网卡名称，该名称可以通过 **mst status –v** 命令查询，如 /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001。IB 交换模块名称以 SW 开头，IB 网卡以网卡名称开头，例如 mt4119_pciconf0。

【举例】

- IB 交换模块固件版本查询举例：

```
[root@wlp-node9 ~]# flint -d /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001 query
Image type:            FS4
FW ISSU Version:       1
FW Version:            27.2000.1600
FW Release Date:       27.6.2019
Description:           UID                GuidsNumber
Base GUID:             0004c903002ef600        8
Base MAC:              0004c92ef600            8
Image VSD:             N/A
Device VSD:            N/A
PSID:                  H3C0000000009
Security Attributes:_  N/A
```

- IB 网卡固件版本查询举例：

```
[root@wlp-node9 ~]# flint -d /dev/mst/mt4119_pciconf0 query
Image type:            FS4
FW Version:            16.25.8000
FW Release Date:       31.7.2019
Product Version:       16.25.8000
Rom Info:              type=UEFI version=14.18.22 cpu=AMD64
                       type=PXE version=3.5.702 cpu=AMD64
Description:           UID                GuidsNumber
Base GUID:             e41d2d0030570ce6        4
Base MAC:              e41d2d570ce6            4
Image VSD:             N/A
Device VSD:            N/A
PSID:                  H3C0000000008
Security Attributes:_  N/A
```

3.11 IB交换模块和IB网卡固件升级

本命令用于 IB 网卡和 IB 交换模块的固件升级。

【命令】

flint -d <*IB device*> **-i** < *fw version*> **burn**。

【参数说明】

<*IB device*>：表示 IB 交换模块和 IB 网卡名称，该名称可以通过 **mst status –v** 命令查询，如 /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001。IB 交换模块名称以 SW 开头，IB 网卡以网卡名称开头，例如 mt4119_pciconf0。

< *fw version*>：表示固件版本。

详细参数可以输入 **flint --help** 进行参考。

【举例】

- IB 交换模块固件版本升级，升级完成后 IB 交换模块需要断电重启生效。

```
[root@wlp-node9 ~]# flint -d /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0017 -i fw-Quantum-rel-27_2
000_1600-H3C_Quantum_UIS9000_Ax-2019_09_19.bin burn

    Current FW version on flash:  27.2000.1600
    New FW version:               27.2000.1600

    Note: The new FW version is the same as the current FW version on flash.

 Do you want to continue ? (y/n) [n] : y

Burning FW image without signatures - OK
Burning FW image without signatures - OK
Restoring signature                 - OK
-I- To load new FW run reboot machine.
```

- IB 网卡固件版本升级，升级完成后请重启服务器生效。

```
[root@wlp-node10 ~]# flint -d /dev/mst/mt4119_pciconf0  -i fw-ConnectX5-rel-16_25_8000-H3C_2P_MEZZ_100G_Ax-UEFI-14.18.22-FlexBoot-3.5.702.bin
 burn

    Current FW version on flash:  16.25.8000
    New FW version:               16.25.8000

    Note: The new FW version is the same as the current FW version on flash.

 Do you want to continue ? (y/n) [n] : y

Initializing image partition -   OK
Writing Boot image component -   OK
-I- To load new FW run mlxfwreset or reboot machine.
```

3.12 端口速率及状态查询

本命令用于查询 IB 交换模块内部及外部口端口状态以及 IB 网卡本身的端口状态。

【命令】

查看 IB 网卡的端口状态：

mlxlink -d *<IB device>*

查看 IB 交换模块上端口状态：

mlxlink -d *<IB device>* **-port** *<port number>*

【参数说明】

<IB device>：表示 IB 交换模块和 IB 网卡名称，该名称可以通过 **mst status –v** 命令查询，如 /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001。IB 交换模块名称以 SW 开头，IB 网卡以网卡名称开头，例如 mt4119_pciconf0。

<port number>：表示端口号，取值 1～41。

详细参数可以输入 **mlxlink --help** 进行参考。

【举例】

- 查询当前 IB 网卡的端口状态及端口速率。

```
[root@wlp-node9 ~]# mlxlink -d /dev/mst/mt4119_pciconf0

Operational Info
----------------
State                              : Active
Physical state                     : LinkUp
Speed                              : IB-EDR
Width                              : 4x
FEC                                : Standard LL RS-FEC - RS(271,257)
Loopback Mode                      : No Loopback
Auto Negotiation                   : ON

Supported Info
--------------
Enabled Link Speed                 : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)
Supported Cable Speed              : 0x00000000 ()

Troubleshooting Info
--------------------
Status Opcode                      : 0
Group Opcode                       : N/A
Recommendation                     : No issue was observed.
```

- 查看 IB 交换模块 33 号端口状态及速率。

```
[root@wlp-node9 ~]# mlxlink -d /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001 -p 33

Operational Info
----------------
State                              : Active
Physical state                     : LinkUp
Speed                              : IB-EDR
Width                              : 4x
FEC                                : Standard LL RS-FEC - RS(271,257)
Loopback Mode                      : No Loopback
Auto Negotiation                   : ON

Supported Info
--------------
Enabled Link Speed                 : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)
Supported Cable Speed              : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)

Troubleshooting Info
--------------------
Status Opcode                      : 0
Group Opcode                       : N/A
Recommendation                     : No issue was observed.
```

【信息说明】

Speed：表示当前网卡支持的最大带宽速率。

Enabled Link Speed：表示当前网卡支持的协商速率。

Support Cable Speed：表示与 IB 交换模块连接的线缆支持的协商速率（红色字体表示为内部连接无 cable）。

3.13 设置IB网卡的up/down

本命令用于设置 IB 网卡的端口的 up/down，仅适用于当前服务器下的 IB 网卡。

【命令】

mlxlink -d *<IB device>* **-a** *<up/down>*

【参数说明】

<IB device>：表示 IB 网卡名称，该名称可以通过 **mst status –v** 命令查询，如 mt4119_pciconf0。

<up/down>：端口的 up/down 状态，取值为：UP 或 DN，DN 表示 down。

【举例】

- 设置 IB 网卡端口为 down 状态。

```
[root@wlp-node10 ~]# mlxlink -d /dev/mst/mt4119_pciconf0 -a DN

Operational Info
----------------
State                                : Active
Physical state                       : LinkUp
Speed                                : IB-EDR
Width                                : 4x
FEC                                  : Standard LL RS-FEC - RS(271,257)
Loopback Mode                        : No Loopback
Auto Negotiation                     : ON

Supported Info
--------------
Enabled Link Speed                   : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)
Supported Cable Speed                : 0x00000000 ()

Troubleshooting Info
--------------------
Status Opcode                        : 0
Group Opcode                         : N/A
Recommendation                       : No issue was observed.

Configuring Port State (Down)...
```

```
Configuring Port State (Down)...

[root@wlp-node10 ~]# mlxlink -d /dev/mst/mt4119_pciconf0

Operational Info
----------------
State                                : Disable
Physical state                       : ETH_AN_FSM_ENABLE
Speed                                : N/A
Width                                : N/A
FEC                                  : N/A
Loopback Mode                        : N/A
Auto Negotiation                     : ON

Supported Info
--------------
Enabled Link Speed                   : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)
Supported Cable Speed                : 0x00000000 ()

Troubleshooting Info
--------------------
Status Opcode                        : 1
Group Opcode                         : PHY FW
Recommendation                       : The port is closed by command. Please check that the interface is enabled.
```

- 设置 IB 网卡端口为 UP 状态。

```
[root@wlp-node10 ~]# mlxlink -d /dev/mst/mt4119_pciconf0 -a UP

Operational Info
----------------
State                           : Disable
Physical state                  : ETH_AN_FSM_ENABLE
Speed                           : N/A
Width                           : N/A
FEC                             : N/A
Loopback Mode                   : N/A
Auto Negotiation                : ON

Supported Info
--------------
Enabled Link Speed              : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)
Supported Cable Speed           : 0x00000000 ()

Troubleshooting Info
--------------------
Status Opcode                   : 1
Group Opcode                    : PHY FW
Recommendation                  : The port is closed by command. Please check that the interface is enabled.

Configuring Port State (Up)...
```

```
[root@wlp-node10 ~]# mlxlink -d /dev/mst/mt4119_pciconf0

Operational Info
----------------
State                           : Active
Physical state                  : LinkUp
Speed                           : IB-EDR
Width                           : 4x
FEC                             : Standard LL RS-FEC - RS(271,257)
Loopback Mode                   : No Loopback
Auto Negotiation                : ON

Supported Info
--------------
Enabled Link Speed              : 0x0000003f (EDR,FDR,FDR10,QDR,DDR,SDR)
Supported Cable Speed           : 0x00000000 ()

Troubleshooting Info
--------------------
Status Opcode                   : 0
Group Opcode                    : N/A
Recommendation                  : No issue was observed.
```

4 IB 交换模块常用命令(BX1020B)

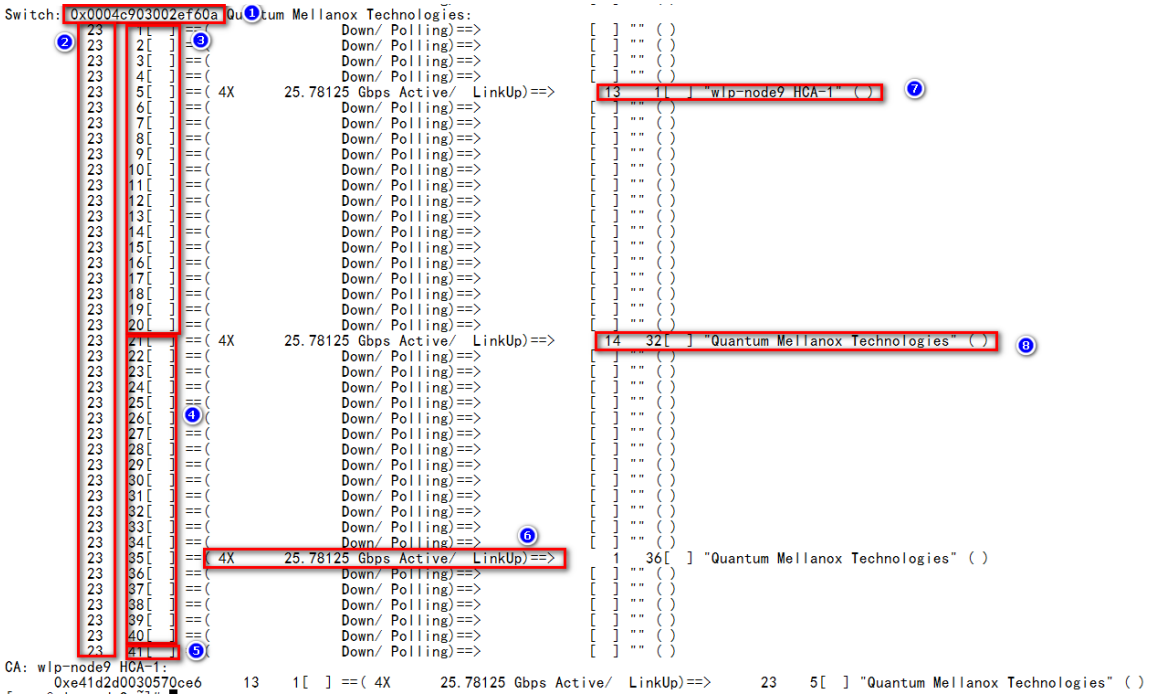
4.1 IB交换模块端口查询

本命令用于查看 IB 交换模块的所有端口的连接状态。此命令会将集群内所有的 IB 交换模块都进行列举。

【命令】

iblinkinfo

【举例】



【信息说明】

① 表示这台交换机的 GUID 号，唯一区分交换机的标识。

② LID 号，集群创建时 SM 分配的唯一便于管理的 lid 号。

③ 表示 1-20 号口为 IB 交换模块与网卡连接的内部端口，其中 11-12、17-18 为预留口。

④ 表示 21-40 为 IB 交换模块外部连接口。

⑤ 表示 41 号口为虚拟机口，为集群跑压力的自由路由端口。

⑥ 表示当前端口的连接速率：4X*25.78125=100Gpbs(EDR)。

⑦ 表示网卡端口的 lid 号及所对应的服务器。

⑧ 表示 IB 交换模块外部连接口的 lid 号、端口序列号、对端连接设备的名称。

4.2 IB交换模块和IB网卡设备信息查询

本命令用于对当前集群下的设备进行管理使用，查询网卡或者 IB 交换模块的基本信息，详细介绍请参见 3.10 IB 交换模块和 IB 网卡设备信息查询。

4.3 IB交换模块和IB网卡固件升级

本命令用于 IB 网卡和 IB 交换模块的固件升级，详细介绍请参见 3.11 IB 交换模块和 IB 网卡固件升级。

4.4 查看集群下各IB交换模块对应的设备

本命令用于查看集群内 IB 交换模块设备的情况。

【命令】

ibnetdiscover

【举例】

```
[root@wlp-node9 ~]# ibnetdiscover
#
# Topology file: generated on Wed Nov 20 02:28:22 2019
#
# Initiated from node e41d2d0030570ce6 port e41d2d0030570ce6

vendid=0x2c9
devid=0xd2f0
sysimgguid=0x4c903003ef600
switchguid=0x4c903003ef600(4c903003ef600)
Switch  41 "S-0004c903003ef600"          # "Quantum Mellanox Technologies" base port 0 lid 20 lmc 0
[36]    "S-0004c903002ef600"[33]                 # "Quantum Mellanox Technologies" lid 1 4xEDR

vendid=0x2c9
devid=0xd2f0
sysimgguid=0x4c903002ef600
switchguid=0x4c903002ef600(4c903002ef600)
Switch  41 "S-0004c903002ef600"          # "Quantum Mellanox Technologies" base port 0 lid 1 lmc 0
[5]     "H-e41d2d0300570fc3"[1](e41d2d0300570fc3)                # "wlp-node10 HCA-1" lid 3 4xEDR
[33]    "S-0004c903003ef600"[36]                 # "Quantum Mellanox Technologies" lid 20 4xEDR
[36]    "S-0004c903002ef60a"[35]                 # "Quantum Mellanox Technologies" lid 23 4xEDR

vendid=0x2c9
devid=0xd2f0
sysimgguid=0x4c903002ef608
switchguid=0x4c903002ef608(4c903002ef608)
Switch  41 "S-0004c903002ef608"          # "Quantum Mellanox Technologies" base port 0 lid 14 lmc 0
[32]    "S-0004c903002ef60a"[21]                 # "Quantum Mellanox Technologies" lid 23 4xEDR

vendid=0x2c9
devid=0xd2f0
sysimgguid=0x4c903002ef60a
switchguid=0x4c903002ef60a(4c903002ef60a)
Switch  41 "S-0004c903002ef60a"          # "Quantum Mellanox Technologies" base port 0 lid 23 lmc 0
[5]     "H-e41d2d0030570ce6"[1](e41d2d0030570ce6)                # "wlp-node9 HCA-1" lid 13 4xEDR
[21]    "S-0004c903002ef608"[32]                 # "Quantum Mellanox Technologies" lid 14 4xEDR
[35]    "S-0004c903002ef600"[36]                 # "Quantum Mellanox Technologies" lid 1 4xEDR

vendid=0x2c9
devid=0x1017
sysimgguid=0xe41d2d0300570fc3
caguid=0xe41d2d0300570fc3
Ca      1 "H-e41d2d0300570fc3"           # "wlp-node10 HCA-1"
[1](e41d2d0300570fc3)    "S-0004c903002ef600"[5]          # lid 3 lmc 0 "Quantum Mellanox Technologies" lid 1 4xEDR
```

【信息说明】

① 表示交换机的 GUID 号，用于区别集群中的 IB 交换模块。

② 表示与当前交换机连接的设备信息及连接端口。

4.5 查看在位IB交换模块

本命令用于查看当前集群的 IB 交换模块，此命令会将集群中所有的设备一一列举。

【命令】

ibswitches

【举例】

```
[root@wlp-node9 ~]# ibswitches
Switch  : 0x0004c903003ef600 ports 41 "Quantum Mellanox Technologies" base port 0 lid 20 lmc 0
Switch  : 0x0004c903002ef600 ports 41 "Quantum Mellanox Technologies" base port 0 lid 1 lmc 0
Switch  : 0x0004c903002ef608 ports 41 "Quantum Mellanox Technologies" base port 0 lid 14 lmc 0
Switch  : 0x0004c903002ef60a ports 41 "Quantum Mellanox Technologies" base port 0 lid 23 lmc 0
```

【信息说明】

显示当前 IB 交换模块的总端口数量以及集群下 lid 号。

4.6 设置IB交换模块端口up/down

本命令用于将 IB 交换模块端口 up/down/reset。

【命令】

ibportstate *<LID> <Port> <port state>*

【参数说明】

<LID>：表示端口所在设备的 LID 号。

<Port>：端口号。

<port state>：包括 enable、disable 等状态，enable 表示端口 UP，disable 表示端口 DOWN。

详细参数可以输入 **ibportstate --help** 进行参考。

【举例】

- 查看 LID 为 1 的设备的 33 号端口状态。

```
[root@wlp-node10 ~]# ibportstate 1 33
Switch PortInfo:
# Port info: Lid 1 port 33
LinkState:.....................Active
PhysLinkState:.................LinkUp
Lid:...........................0
SMLid:.........................0
LMC:...........................0
LinkWidthSupported:............1X or 4X or 2X
LinkWidthEnabled:..............1X or 4X or 2X
LinkWidthActive:...............4X
LinkSpeedSupported:............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedEnabled:..............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedActive:...............10.0 Gbps
LinkSpeedExtSupported:.........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtEnabled:...........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtActive:............25.78125 Gbps
# MLNX ext Port info: Lid 1 port 33
StateChangeEnable:.............0x00
LinkSpeedSupported:............0x01
LinkSpeedEnabled:..............0x01
LinkSpeedActive:...............0x00
OOOSLMask:.....................0x0000
Peer PortInfo:
# Port info: Lid 1 DR path slid 7; dlid 65535; 0,33 port 36
LinkState:.....................Active
PhysLinkState:.................LinkUp
Lid:...........................0
SMLid:.........................0
LMC:...........................0
LinkWidthSupported:............1X or 4X or 2X
LinkWidthEnabled:..............1X or 4X or 2X
LinkWidthActive:...............4X
LinkSpeedSupported:............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
```

- 设置 LID 为 1 的设备的 33 号端口状态为 Down。

```
[root@wlp-node10 ~]# ibportstate 1 33 disable
Initial Switch PortInfo:
# Port info: Lid 1 port 33
LinkState:.....................Active
PhysLinkState:.................LinkUp
Lid:...........................0
SMLid:.........................0
LMC:...........................0
LinkWidthSupported:............1X or 4X or 2X
LinkWidthEnabled:..............1X or 4X or 2X
LinkWidthActive:...............4X
LinkSpeedSupported:............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedEnabled:..............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedActive:...............10.0 Gbps
LinkSpeedExtSupported:.........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtEnabled:...........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtActive:............25.78125 Gbps
# MLNX ext Port info: Lid 1 port 33
StateChangeEnable:.............0x00
LinkSpeedSupported:............0x01
LinkSpeedEnabled:..............0x01
LinkSpeedActive:...............0x00
OOOSLMask:.....................0x0000
Disable may be irreversible

After PortInfo set:
# Port info: Lid 1 port 33
LinkState:.....................Initialize
PhysLinkState:.................LinkUp
Lid:...........................0
SMLid:.........................0
LMC:...........................0
```

```
[root@wlp-node10 ~]# ibportstate 1 33
Switch PortInfo:
# Port info: Lid 1 port 33
LinkState:.....................Down
PhysLinkState:.................Disabled
Lid:...........................0
SMLid:.........................0
LMC:...........................0
LinkWidthSupported:............1X or 4X or 2X
LinkWidthEnabled:..............1X or 4X or 2X
LinkWidthActive:...............4X
LinkSpeedSupported:............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedEnabled:..............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedActive:...............Extended speed
LinkSpeedExtSupported:.........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtEnabled:...........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtActive:............No Extended Speed
# MLNX ext Port info: Lid 1 port 33
StateChangeEnable:.............0x00
LinkSpeedSupported:............0x01
LinkSpeedEnabled:..............0x01
LinkSpeedActive:...............0x00
OOOSLMask:.....................0x0000
[root@wlp-node10 ~]# ibportstate 1 33 enable
Initial Switch PortInfo:
# Port info: Lid 1 port 33
LinkState:.....................Down
PhysLinkState:.................Disabled
```

- 设置 LID 为 1 的设备的 33 号端口状态为 UP。

```
[root@wlp-node10 ~]# ibportstate 1 33 enable
Initial Switch PortInfo:
# Port info: Lid 1 port 33
LinkState:.....................Down
PhysLinkState:.................Disabled
Lid:...........................0
SMLid:.........................0
LMC:...........................0
LinkWidthSupported:............1X or 4X or 2X
LinkWidthEnabled:..............1X or 4X or 2X
LinkWidthActive:...............4X
LinkSpeedSupported:............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedEnabled:..............2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedActive:...............Extended speed
LinkSpeedExtSupported:.........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtEnabled:...........14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtActive:............No Extended Speed
# MLNX ext Port info: Lid 1 port 33
StateChangeEnable:.............0x00
LinkSpeedSupported:............0x01
LinkSpeedEnabled:..............0x01
LinkSpeedActive:...............0x00
OOOSLMask:.....................0x0000
```

```
[root@wlp-node10 ~]# ibportstate 1 33
Switch PortInfo:
# Port info: Lid 1 port 33
LinkState:.....................Active
PhysLinkState:.....................LinkUp
Lid:.....................0
SMLid:.....................0
LMC:.....................0
LinkWidthSupported:.....................1X or 4X or 2X
LinkWidthEnabled:.....................1X or 4X or 2X
LinkWidthActive:.....................4X
LinkSpeedSupported:.....................2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedEnabled:.....................2.5 Gbps or 5.0 Gbps or 10.0 Gbps
LinkSpeedActive:.....................10.0 Gbps
LinkSpeedExtSupported:.....................14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtEnabled:.....................14.0625 Gbps or 25.78125 Gbps
LinkSpeedExtActive:.....................25.78125 Gbps
# MLNX ext Port info: Lid 1 port 33
StateChangeEnable:.....................0x00
LinkSpeedSupported:.....................0x01
LinkSpeedEnabled:.....................0x01
LinkSpeedActive:.....................0x00
OOOSLMask:.....................0x0000
Peer PortInfo:
# Port info: Lid 1 DR path slid 7; dlid 65535; 0,33 port 36
LinkState:.....................Active
PhysLinkState:.....................LinkUp
Lid:.....................0
SMLid:.....................0
LMC:.....................0
LinkWidthSupported:.....................1X or 4X or 2X
LinkWidthEnabled:.....................1X or 4X or 2X
LinkWidthActive:.....................4X
LinkSpeedSupported:.....................2.5 Gbps or 5.0 Gbps or 10.0 Gbps
```

4.7 检查当前环境下光纤物理链路的健康情况

本命令用于检测当前集群下物理链路的健康情况，利用快速查看环境下链路状态。

执行此命令之后 log 信息都在/var/tmp/ibdiagnet2/目录下。

【命令】

ibdiagnet

【参数说明】

详细参数可以输入 **ibdiagnet --help** 进行参考。

【举列】

```
[root@wlp-node9 ~]# ibdiagnet
----------
Load Plugins from:
/usr/share/ibdiagnet2.1.1/plugins/
(You can specify more paths to be looked in with "IBDIAGNET_PLUGINS_PATH" env variable)

Plugin Name                                   Result     Comment
libibdiagnet_cable_diag_plugin-2.1.1          Succeeded  Plugin loaded
libibdiagnet_phy_diag_plugin-2.1.1            Succeeded  Plugin loaded

---------------------------------------------
Discovery
-I- Discovering ... 10 nodes (4 Switches & 6 CA-s) discovered.
-I- Fabric Discover finished successfully

-I- Discovered 10 nodes (4 Switches & 6 CA-s).

-I- Retrieving ... 10/10 nodes (4/4 Switches & 6/6 CA-s) retrieved.
-I- VS Capability GMP finished successfully

-I- Retrieving ... 10/10 nodes (4/4 Switches & 6/6 CA-s) retrieved.
-I- VS Capability SMP finished successfully

-I- Retrieving ... 10/10 nodes (4/4 Switches & 6/6 CA-s) retrieved.
-I- VS ExtendedPortInfo finished successfully

-I- Retrieving ... 10/10 nodes (4/4 Switches & 6/6 CA-s) retrieved.
-I- Port Info Extended finished successfully

-I- Retrieving ... 10/10 nodes (4/4 Switches & 6/6 CA-s) retrieved.
-I- Switch Info retrieving finished successfully

-I- Duplicated GUIDs detection finished successfully

-W- Note: If you have unmanaged systems then duplication can occur
-W- Duplicated Node Description detection finished with errors
-W- S0004c903002ef60a/N0004c903002ef60a - Node with GUID=0x0004c903002ef60a is configured with duplicated node description - Quantum Mellanox Te
chnologies
-W- S0004c903002ef608/N0004c903002ef608 - Node with GUID=0x0004c903002ef608 is configured with duplicated node description - Quantum Mellanox Te
chnologies
-W- S0004c903002ef600/N0004c903002ef600 - Node with GUID=0x0004c903002ef600 is configured with duplicated node description - Quantum Mellanox Te
chnologies
-W- S0004c903003ef600/N0004c903003ef600 - Node with GUID=0x0004c903003ef600 is configured with duplicated node description - Quantum Mellanox Te
chnologies
```

```
-------------------------------------------------
Summary
-I- Stage                     Warnings   Errors     Comment
-I- Discovery                 4          0
-I- Lids Check                0          0
-I- Links Check               0          0
-I- Subnet Manager            0          0
-I- Port Counters             0          2
-I- Nodes Information         0          0
-I- Speed / Width checks      0          0
-I- Alias GUIDs               0          0
-I- Virtualization            0          0
-I- Partition Keys            0          0
-I- Temperature Sensing       0          0

-I- You can find detailed errors/warnings in: /var/tmp/ibdiagnet2/ibdiagnet2.log


-I- ibdiagnet database file    : /var/tmp/ibdiagnet2/ibdiagnet2.db_csv
-I- LST file                   : /var/tmp/ibdiagnet2/ibdiagnet2.lst
-I- Network dump file          : /var/tmp/ibdiagnet2/ibdiagnet2.net_dump
-I- Subnet Manager file        : /var/tmp/ibdiagnet2/ibdiagnet2.sm
-I- Ports Counters file        : /var/tmp/ibdiagnet2/ibdiagnet2.pm
-I- Nodes Information file     : /var/tmp/ibdiagnet2/ibdiagnet2.nodes_info
-I- Alias guids file           : /var/tmp/ibdiagnet2/ibdiagnet2.aguid
-I- VPorts file                : /var/tmp/ibdiagnet2/ibdiagnet2.vports
-I- VPorts Pkey file           : /var/tmp/ibdiagnet2/ibdiagnet2.vports_pkey
-I- Partition keys file        : /var/tmp/ibdiagnet2/ibdiagnet2.pkey
```

4.8 IB交换模块日志收集

本命令用于收集 IB 交换模块端口日志信息。

【命令】

mlxdump -d <*IB device*> **snapshot --mode full -o** <*log name*>

【参数说明】

<*IB device*>：表示 IB 交换模块和 IB 网卡名称，该名称可以通过 **mst status –v** 命令查询。

<*log name* >：表示日志文件的名称，如 *Quantum_mlxdump1.log*。

详细参数可以输入 **flint --help** 进行参考。

【举例】

```
[root@wlp-node9 ~]# mlxdump -d /dev/mst/SW_MT54000_Quantum_Mellanox_Technologies_lid-0x0001 snapshot --mode full -o Quantum_mlxdump1.log
-I- Dumping crspace...
-I- crspace was dumped successfully
-I- Dumping phy_uc...
-I- phy_uc was dumped successfully
-I- Dump file "Quantum_mlxdump1.log" was generated successfully
```

5 常见问题处理

5.1 IB网卡常见问题

5.1.1 IB 网卡端口不可见

诊断步骤:

(1) 排查 IB 卡与服务器的兼容性

- 如果使用非标准系统，请联系具体 OS 研发解决。
- 如果 IB 卡版本不配套，请先升级。

(2) 排查 PCIe 硬件设备是否可以见（**lspci |grep Mellanox**）

- 如果 PCIe 设备不可见，看相应 CPU 是否在位；IB 网卡是否未安装到位；更换 IB 网卡的槽位
- 如果 PCIe 设备可见，但是网口不可见，可使用 **ifconfig –a /ifconfig ibN up**，然后重新安装驱动，重启系统。

5.1.2 IB 网口不通或端口为初始化状态

诊断步骤:

(1) 排查 IB 网卡是否 up，且状态为 link。

- 如果是，查看 IP 是否设置正确网口上，相关命令：**ifconfig ibN up**，**ibstat**。
- 如果否，建议检查集群中的 opensm 功能是否开启，相关命令：**/etc/init.d/opensm start** , **/etc/init.d/opensm status**。

(2) 重新对 IB 网卡更换槽位，检查是否正常。

(3) 查看与当前网卡连接的 IB 交换模块是否在位可用。

5.1.3 IB 网卡端口速率协商异常

诊断步骤：

(1) 排查当前 IB 网卡是否可以支持当前所需要协商的速率。

- 如果否，请更换支持的 IB Mezz 网卡。
- 如果是，排查端口是否被协商降速，具体命令参考：3.12 端口速率及状态查询

(2) 更换当前的网卡槽位，检查是否正常。

5.1.4 IB 网卡查询系统中组网设备的相关命令执行失败，如 ibv_devinfo 执行报错，failed to get IB deviceslist。

该 IB 命令在驱动还未加载的情况下被调用，执行命令：**/etc/init.d/openibd start**。

5.2 IB交换模块常见问题

表5-1 IB 交换模块常见故障诊断方法

故障种类	故障现象	故障原因及处理方法
LEDs	面板端口指示灯为琥珀色常亮	这个状态标明端口仅物理层UP： 确认SM已经运行在当前网络中：/etc/init.d/opensm status； 将当前线缆进行更换，确认非线缆问题。
	面板端口指示灯闪烁（琥珀色）	这个状态可能说明线缆有问题： 将当前的线缆进行更换，确认非线缆问题； 更换面板其他端口，确认非端口问题。
	IB交换模块指示灯闪红色	这个状态说明交换机有告警信息： 确认风扇正常使用，未导致IB交换模块温度过高；
	IB交换模块系统指示灯长灭	这个状态说明可能IB交换模块未正常上电： 确认OM电源正常使用，且可正常给IB交换模块上电； 确认IB交换模块固件版本正确，GUID设备号正确烧录。