ACADEMY

GTC InfiniBand Lab Access

Access LAB Site Using...

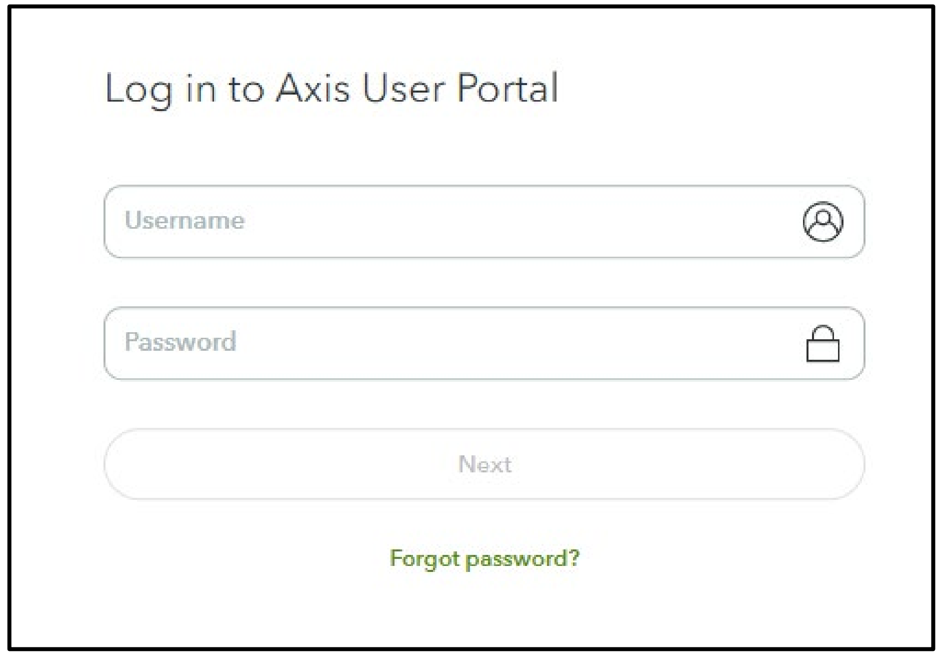
1. Open a **chrome** web browser

2. Access the lab using the following link:

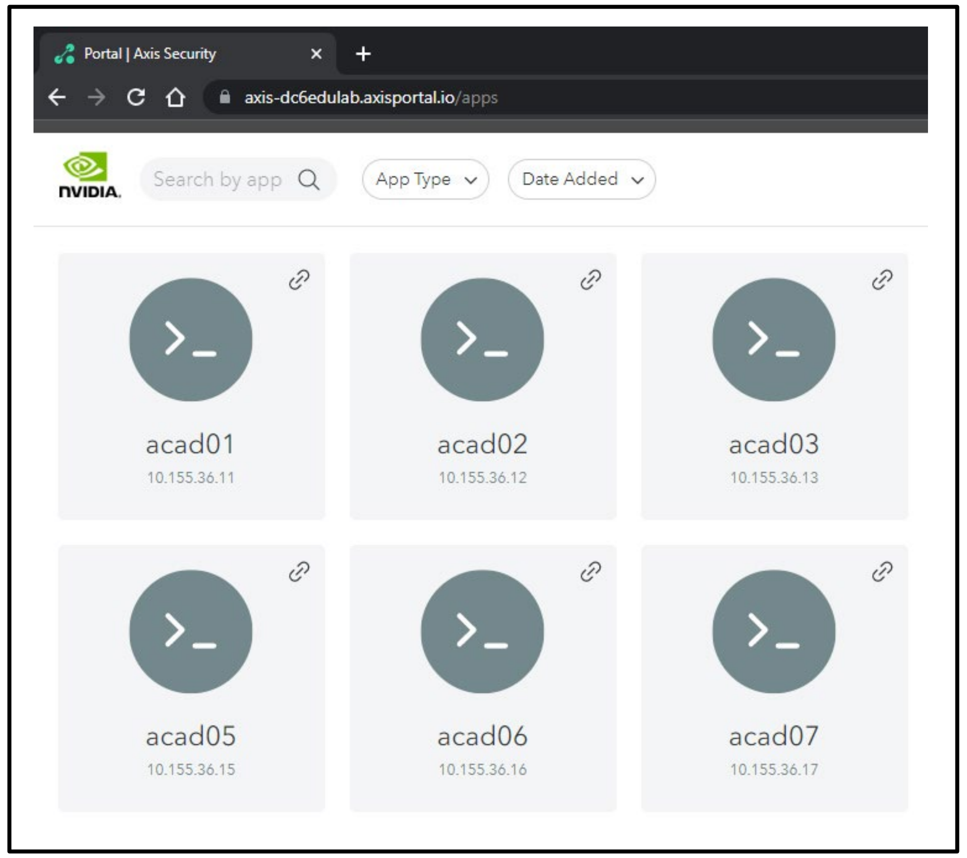
   https://axis-dc6edulab.axisportal.io/apps

   - User : gtcuser
   - Password : Welcome123!

3. Click next to see all available servers.



Log in to Axis User Portal



acad01 10.155.36.11

acad02 10.155.36.12

acad03 10.155.36.13

acad05 10.155.36.15

acad06 10.155.36.16

acad07 10.155.36.17

Access LAB Site Using...

1. Choose the first servers you are going to use, then choose “randomly“ from acad01-acad12 to optimize lab performance.

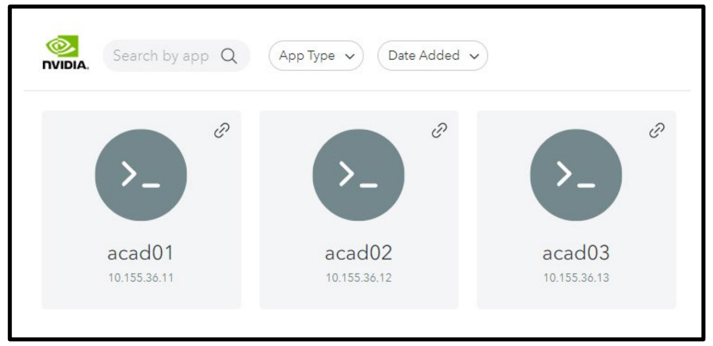
2. Click the web SSH ICON icon to login to your server.

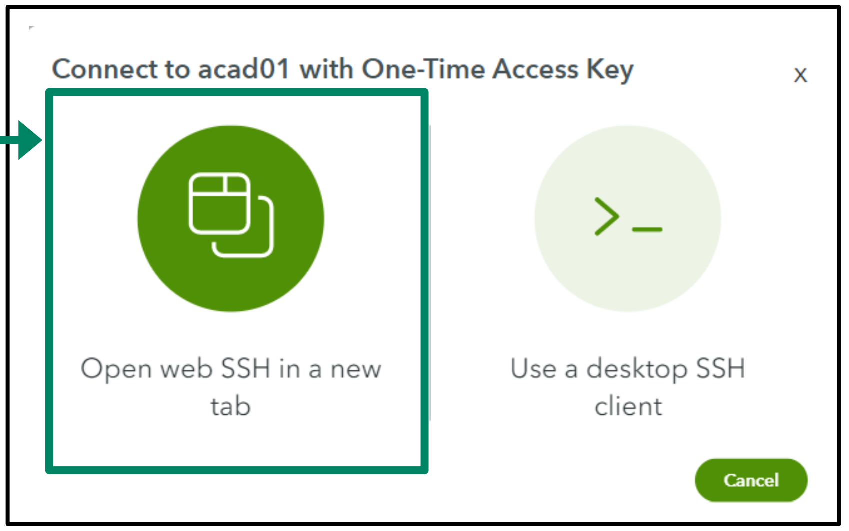
3. Login to the server using the following:

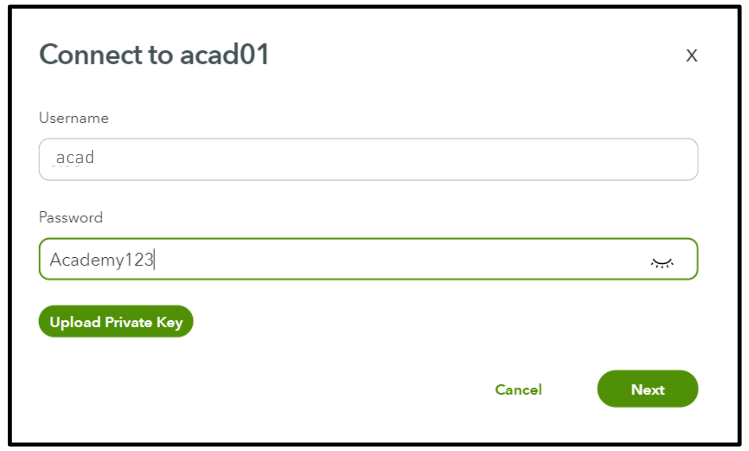
   - Username: acad
   - Password: Academy123

   and click next to reach the server display.

4. The server prompt is now displayed, you may run InfiniBand commands as depicted in the next session, Lab practices




Connect to acad01 with One-Time Access Key
x
Open web SSH in a new tab
Use a desktop SSH client
Cancel



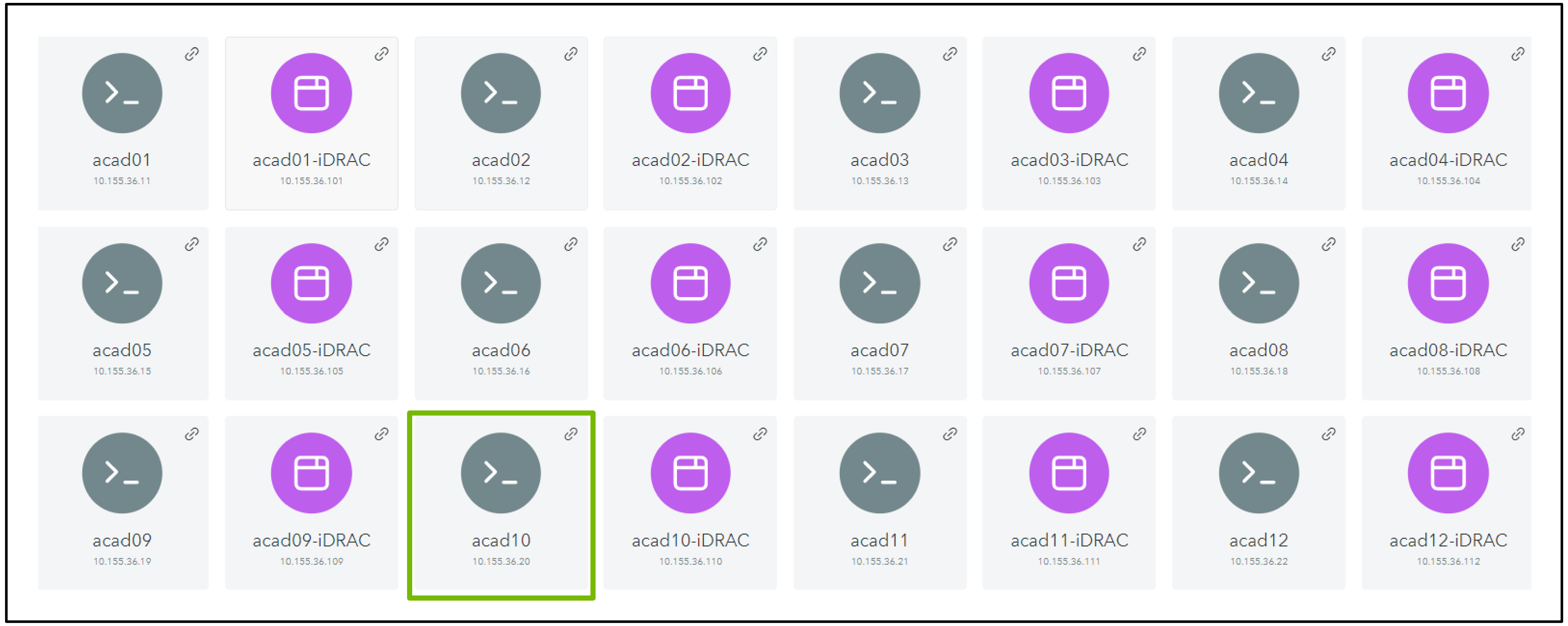
Username: acad

Password: Academy123

```
Last login: Tue Jul 12 06:17:51 2022 from 10.155.36.24
stud@acad0x:~$
```

LAB-accessible servers for GTC

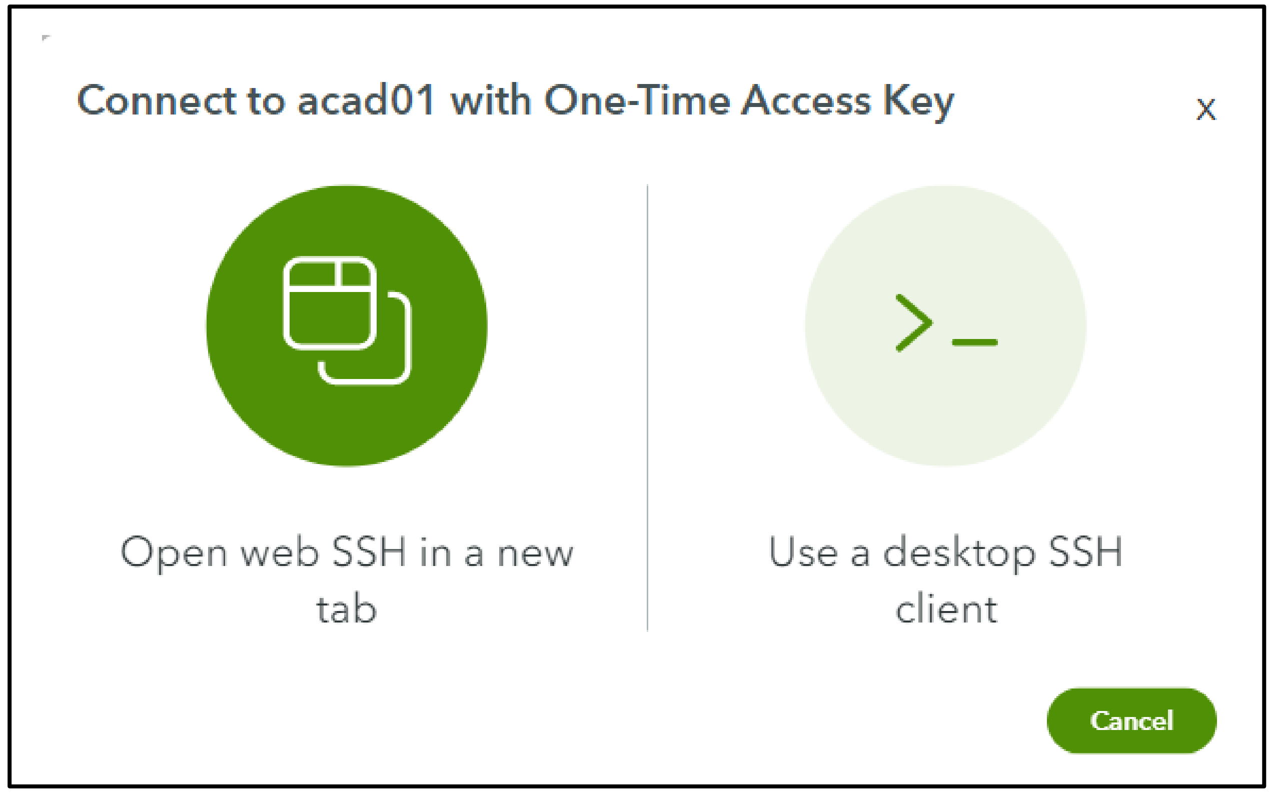
1. Choose the first servers you are going to use.
   Choose “randomly“ from
   acad01-acad12
   to optimize lab
   performance.



acad01 10.155.36.11	acad01-iDRAC 10.155.36.101	acad02 10.155.36.12	acad02-iDRAC 10.155.36.102	acad03 10.155.36.13	acad03-iDRAC 10.155.36.103	acad04 10.155.36.14	acad04-iDRAC 10.155.36.104
acad05 10.155.36.15	acad05-iDRAC 10.155.36.105	acad06 10.155.36.16	acad06-iDRAC 10.155.36.106	acad07 10.155.36.17	acad07-iDRAC 10.155.36.107	acad08 10.155.36.18	acad08-iDRAC 10.155.36.108
acad09 10.155.36.19	acad09-iDRAC 10.155.36.109	acad10 10.155.36.20	acad10-iDRAC 10.155.36.110	acad11 10.155.36.21	acad11-iDRAC 10.155.36.111	acad12 10.155.36.22	acad12-iDRAC 10.155.36.112

Access LAB Site

2. Click the web SSH ICON icon to login to your server.



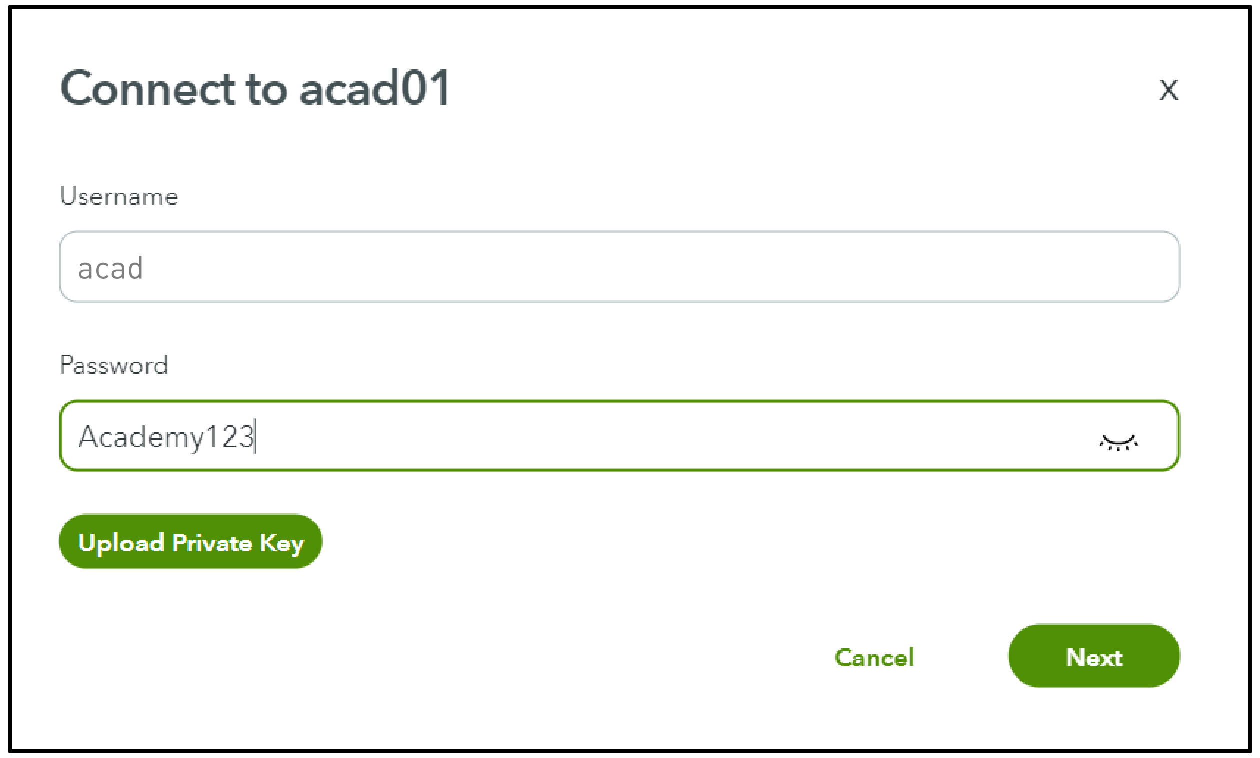
Connect to acad01 with One-Time Access Key

Access LAB Site

3. Login to the server using the following:

   - Username: acad
   - Password: Academy123

   and click next to reach the server display.



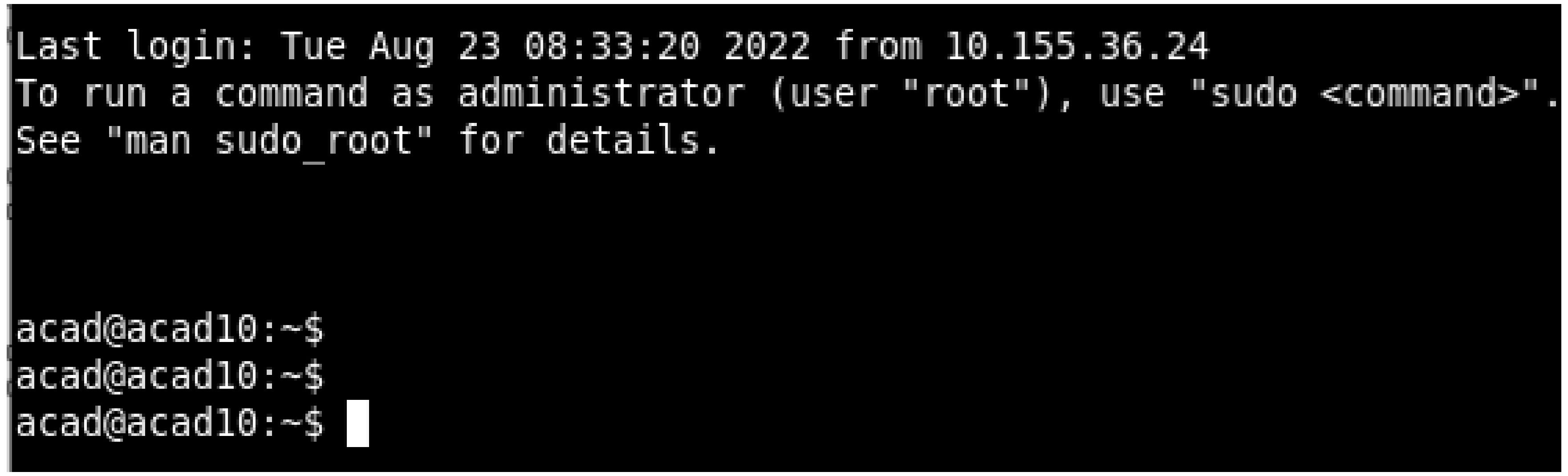
Connect to acad01

Username: acad

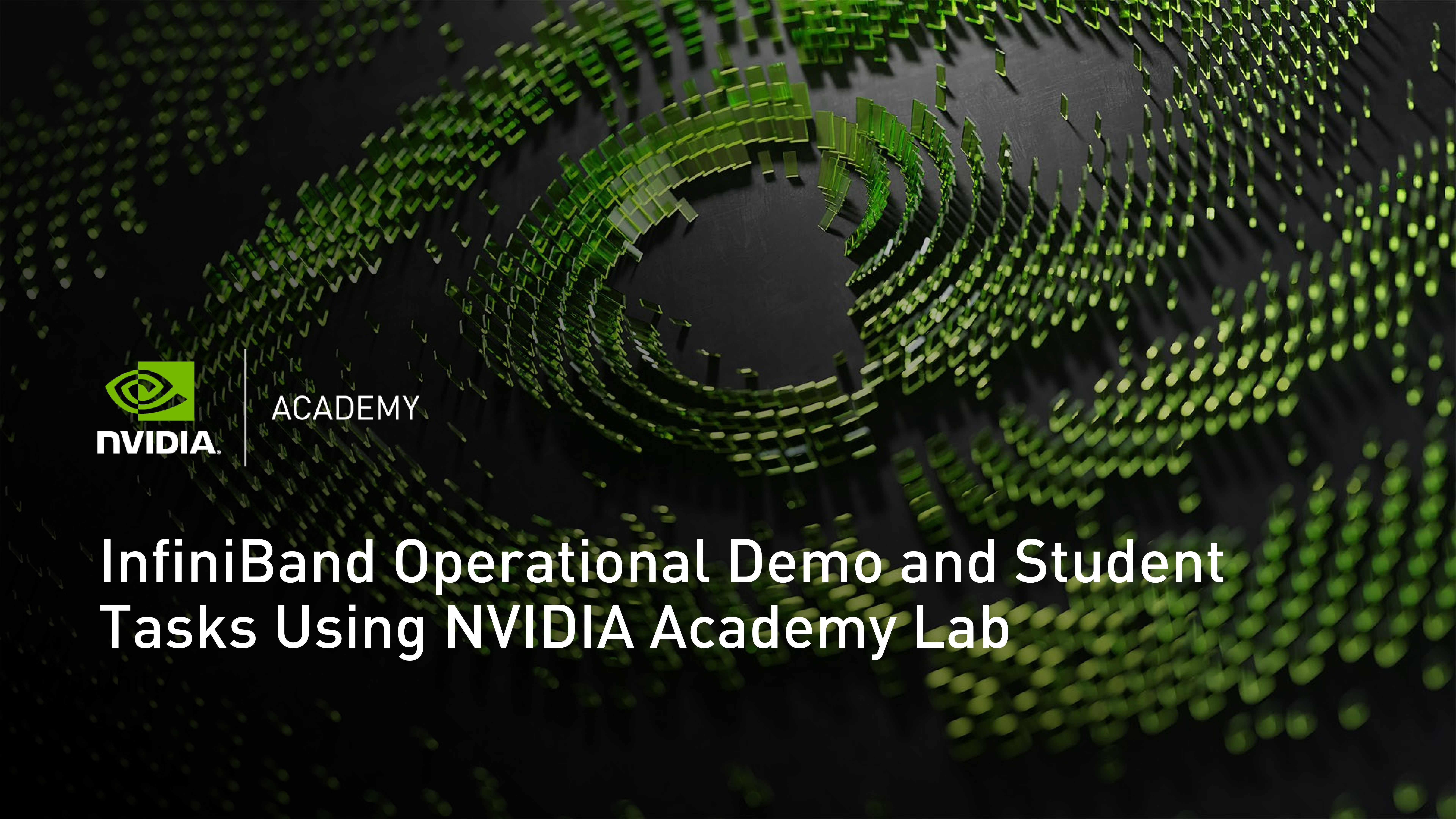
Password: Academy123

Access LAB Site

4. The server prompt is now displayed;
   you may run InfiniBand commands
   as depicted in the next session on lab practices



```
Last login: Tue Aug 23 08:33:20 2022 from 10.155.36.24
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.



acad@acad10:~$
acad@acad10:~$
acad@acad10:~$
```



NVIDIA ACADEMY

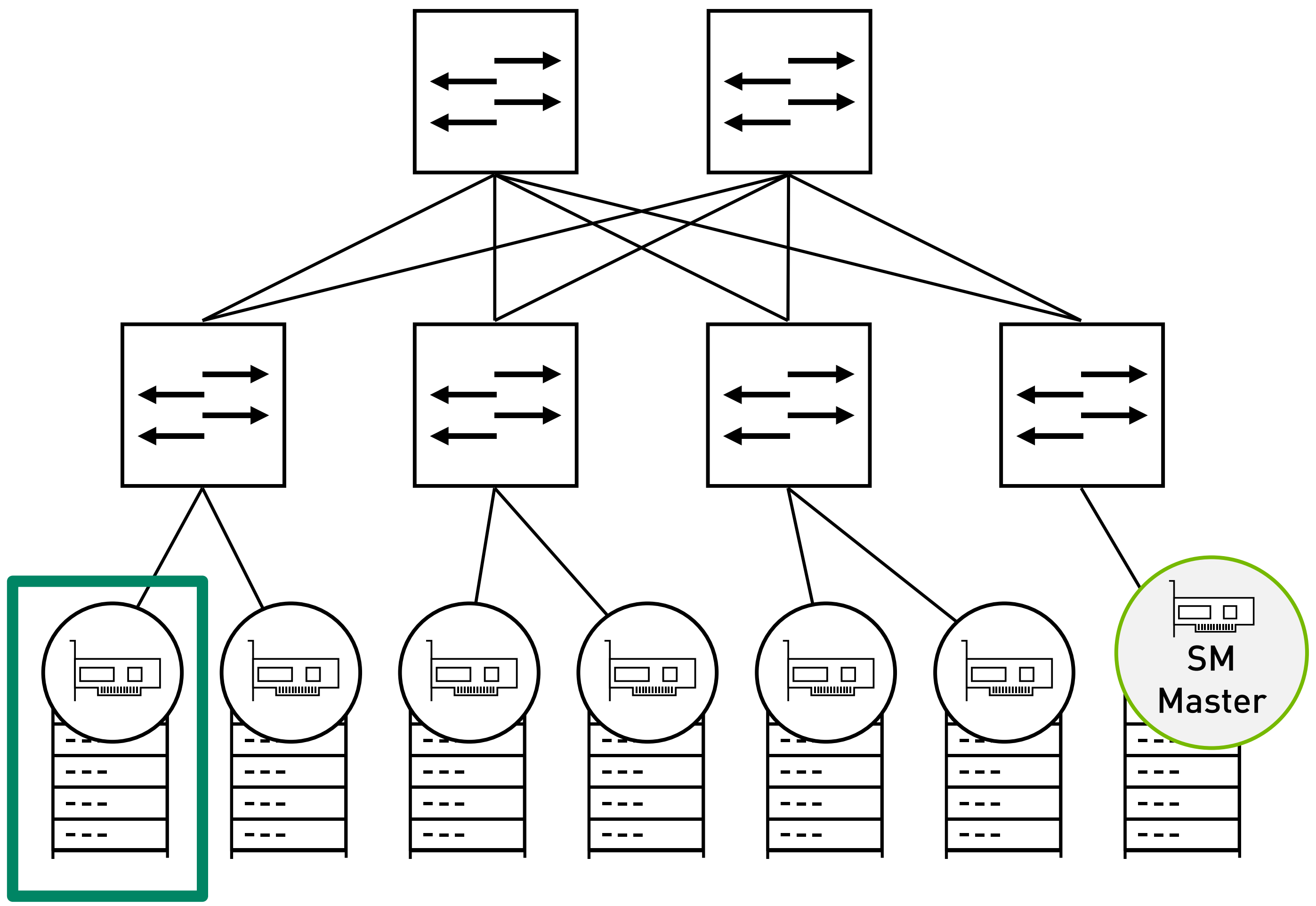
InfiniBand Operational Demo and Student Tasks Using NVIDIA Academy Lab

ofed_info -s

Verify OFED driver on your servers

```
acadadmin@acad10:~$ ofed_info -s

MLNX_OFED_LINUX-5.6-2.0.9.0:
```

```
acadadmin@acad10:~$ ofed_info
MLNX_OFED_LINUX-5.6-2.0.9.0 (OFED-5.6-2.0.9):
hcoll:
sharp:
/sw/release/mlnx_ofed/IBHPC/MLNX_OFED_LINUX-5.6-1.0.3/SRPMS/sharp-
2.7.0.MLNX20220426.703                             f9a40-1.56103.src.rpm
ucx:
/sw/release/mlnx_ofed/IBHPC/MLNX_OFED_LINUX-5.6-1.0.3/SRPMS/ucx-1.13.0-
1.56103.src.rpm
Installed Packages:
-------------------
amd64         InfiniBand diagnostics library
ii  libibumad-dev:amd64                    56mlnx40-1.56209
amd64         Development files for libibumad
ii  libibumad3:amd64                       56mlnx40-1.56209
amd64         InfiniBand Userspace Management Datagram (uMAD) library
ii  libibverbs-dev:amd64                   56mlnx40-1.56209
amd64         Development files for the libibverbs library
ii  libibverbs1:amd64                      56mlnx40-1.56209
amd64         Library for direct userspace use of RDMA (InfiniBand/iWARP)
ii  libibverbs1-dbg:amd64                  56mlnx40-1.56209
amd64         Debug symbols for the libibverbs library
```

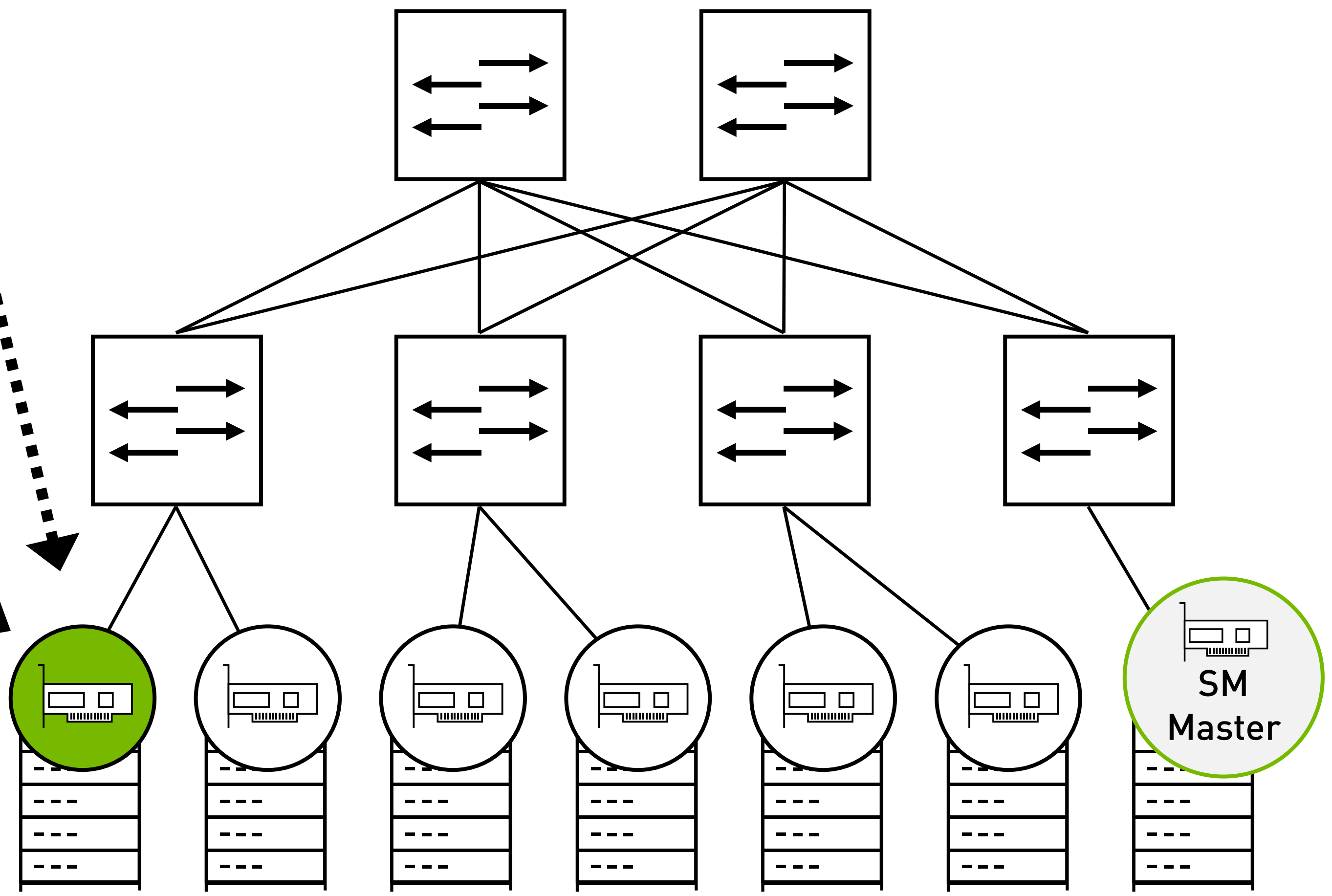


lspci and ibv_devices

Check HCA devices on the server

```
acadadmin@acad10:~$ sudo lspci | grep MT

4b:00.0 Infiniband controller: Mellanox Technologies MT28908 Family [ConnectX-6]
4b:00.1 Infiniband controller: Mellanox Technologies MT28908 Family [ConnectX-6]
```

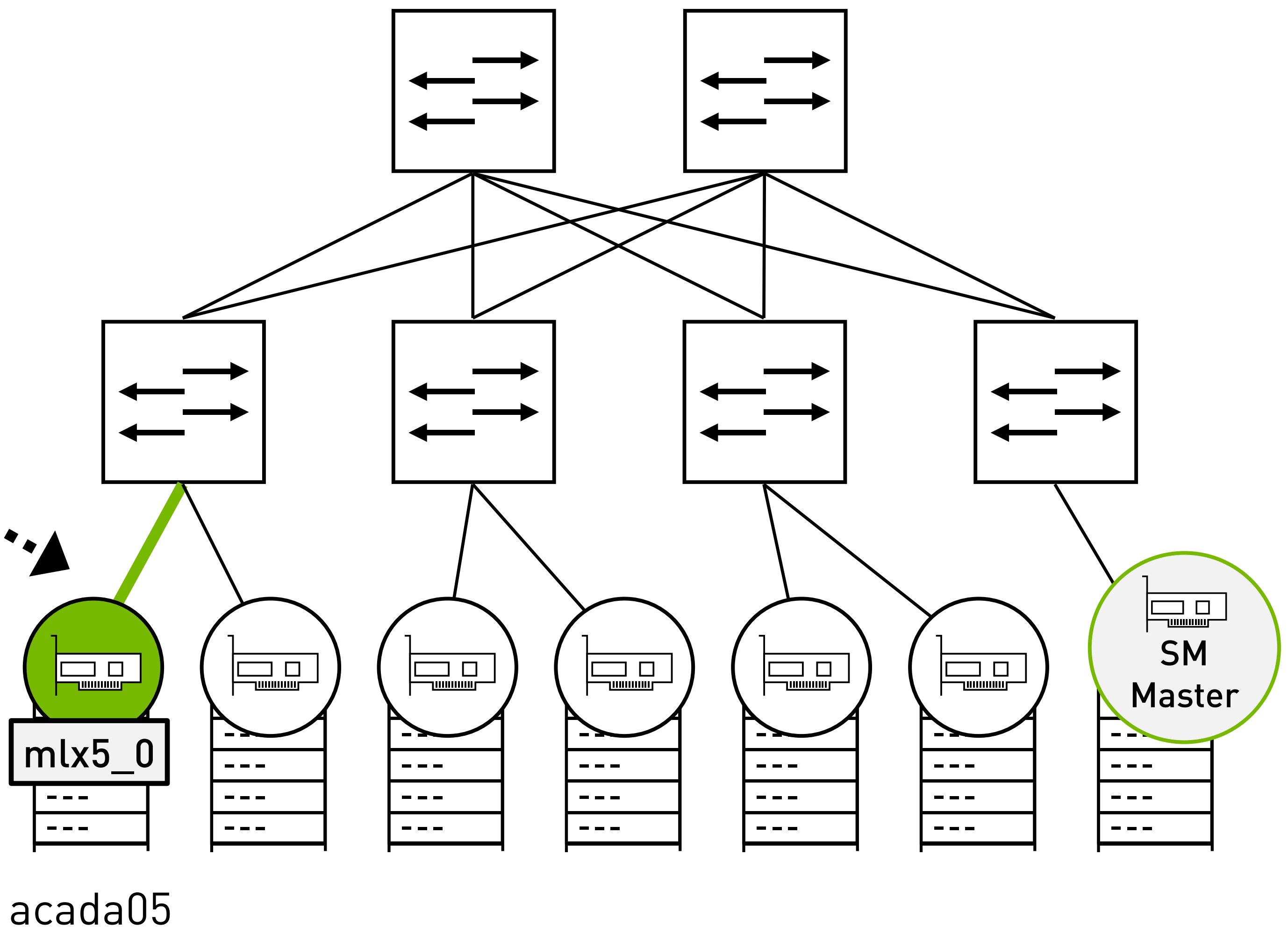
```
acadadmin@acad05:~$ ibv_devices
    device                     node GUID
    ------                  ----------------
    mlx5_0                  b8599f0300f707e4
    mlx5_1                  b8599f0300f707e5
```



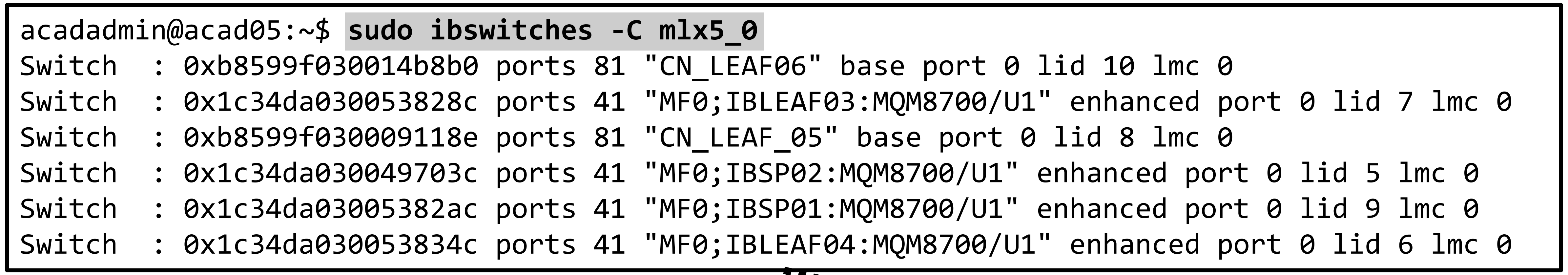
ibstat

Check HCA Link and Port Features Addresses Operational Status

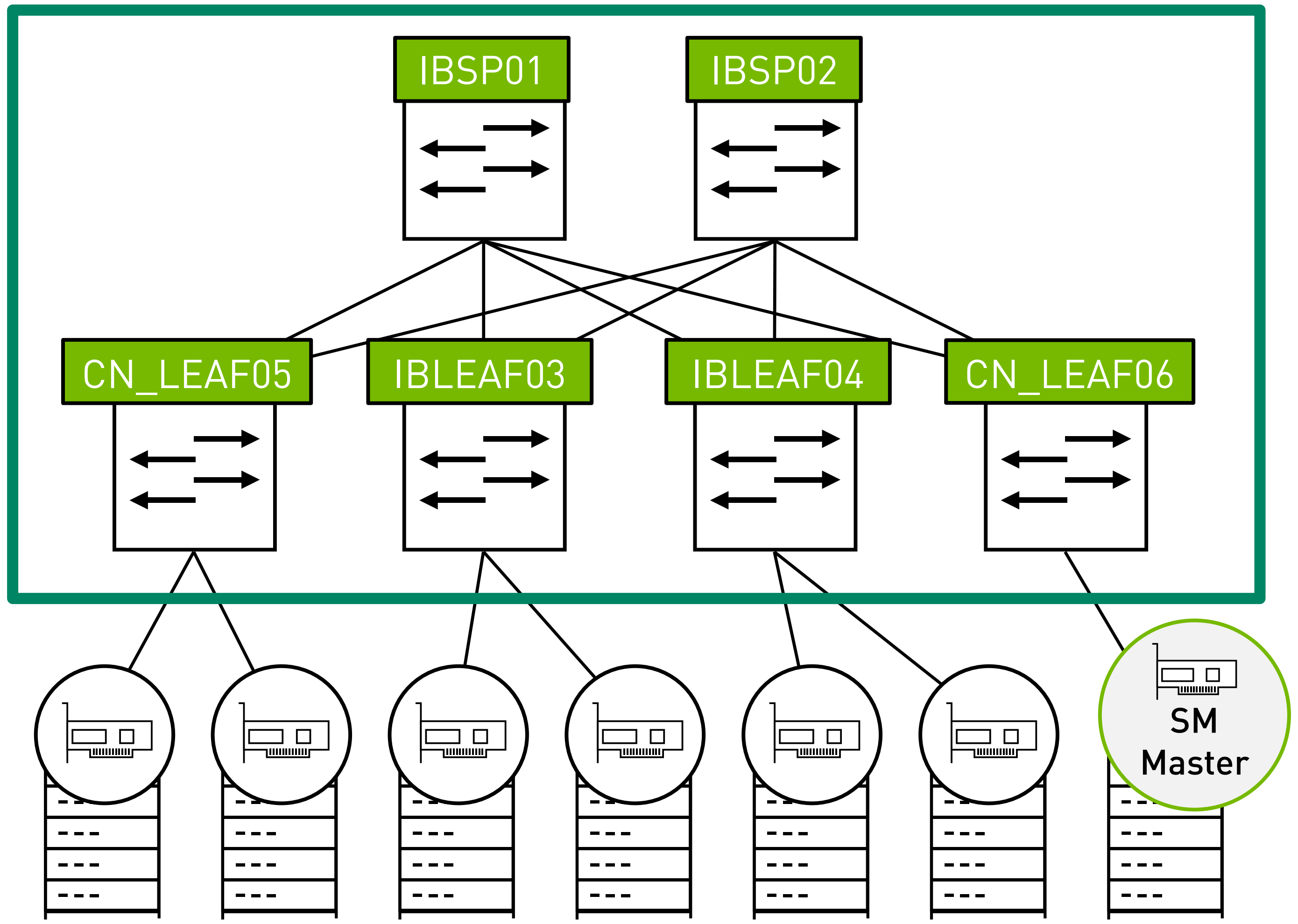
```
acadadmin@acad05:~$ ibstat
CA 'mlx5_0'
        CA type: MT4123
        Number of ports: 1
        Firmware version: 20.33.1048
        Hardware version: 0
        Node GUID: 0xb8599f0300f707e4
        System image GUID: 0xb8599f0300f707e4
        Port 1:
                State: Active
                Physical state: LinkUp
                Rate: 200
                Base lid: 24
                LMC: 0
                SM lid: 23
                Capability mask: 0xa651e848
                Port GUID: 0xb8599f0300f707e4
                Link layer: InfiniBand
```



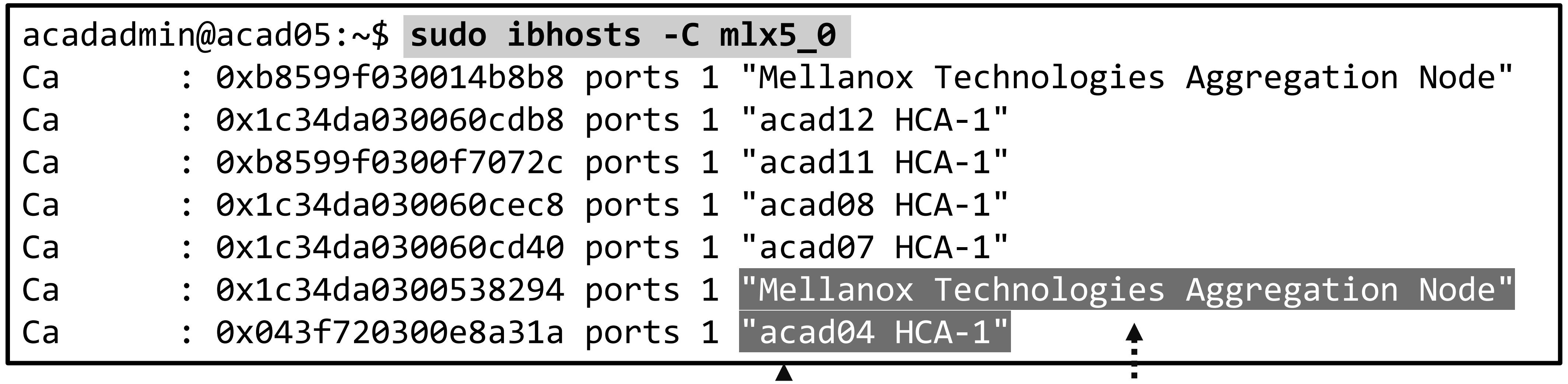
ibswitches—Display All Cluster Switches



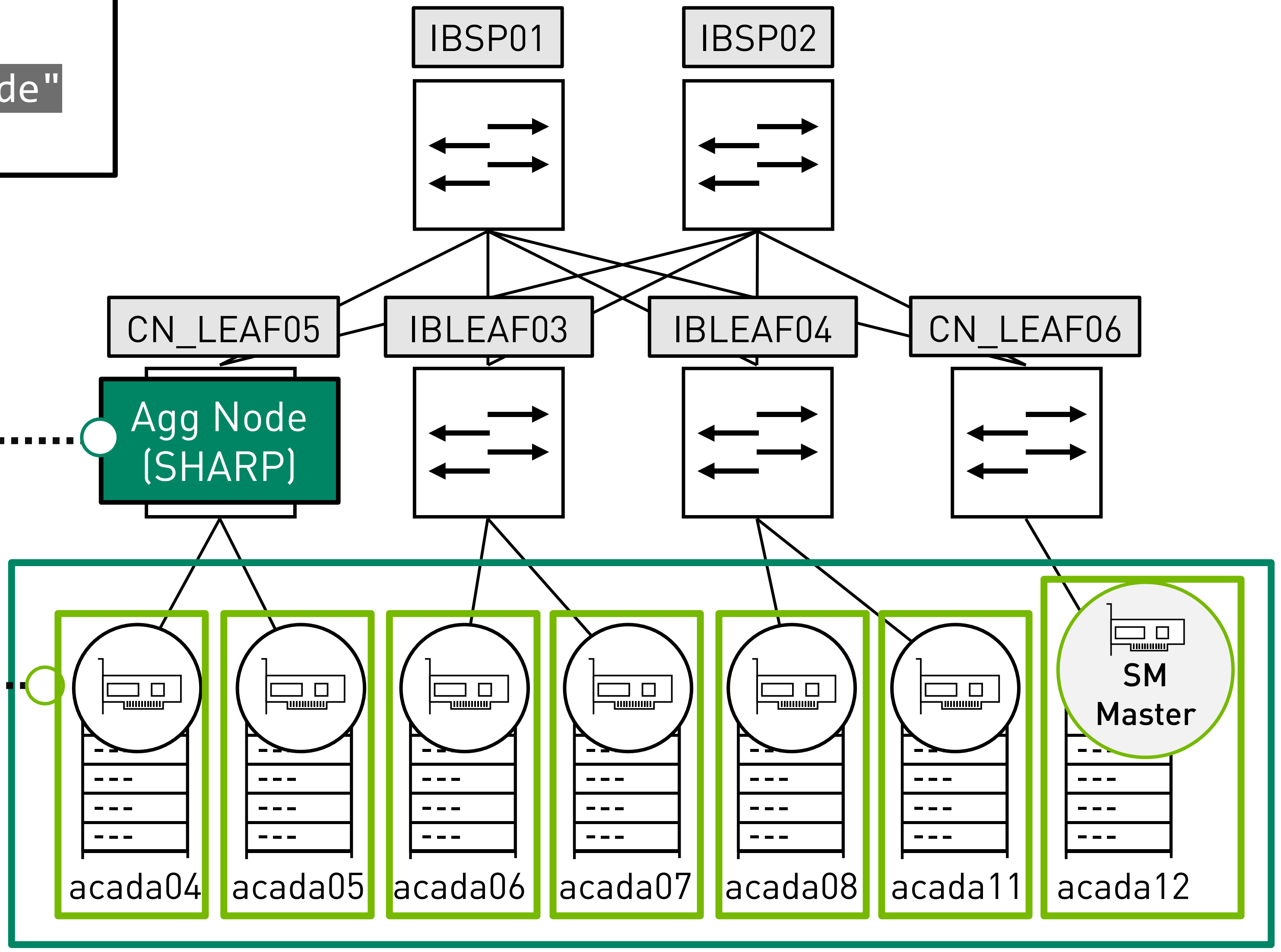
```
acadadmin@acad05:~$ sudo ibswitches -C mlx5_0
Switch  : 0xb8599f030014b8b0 ports 81 "CN_LEAF06" base port 0 lid 10 lmc 0
Switch  : 0x1c34da030053828c ports 41 "MF0;IBLEAF03:MQM8700/U1" enhanced port 0 lid 7 lmc 0
Switch  : 0xb8599f030009118e ports 81 "CN_LEAF_05" base port 0 lid 8 lmc 0
Switch  : 0x1c34da030049703c ports 41 "MF0;IBSP02:MQM8700/U1" enhanced port 0 lid 5 lmc 0
Switch  : 0x1c34da03005382ac ports 41 "MF0;IBSP01:MQM8700/U1" enhanced port 0 lid 9 lmc 0
Switch  : 0x1c34da030053834c ports 41 "MF0;IBLEAF04:MQM8700/U1" enhanced port 0 lid 6 lmc 0
```



ibhosts—Display All Cluster HCAs and ANs

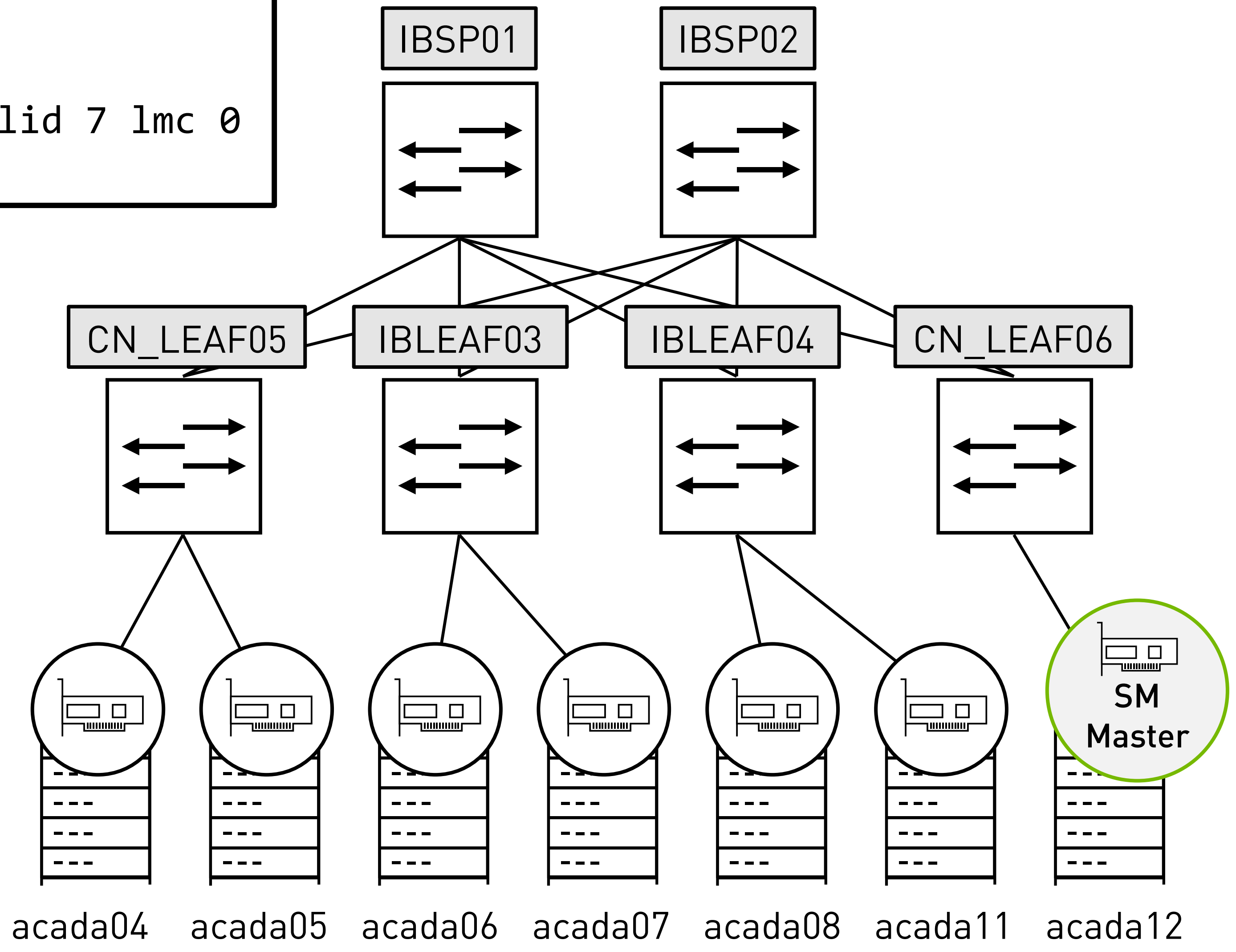


```
acadadmin@acad05:~$ sudo ibhosts -C mlx5_0
Ca      : 0xb8599f030014b8b8 ports 1 "Mellanox Technologies Aggregation Node"
Ca      : 0x1c34da030060cdb8 ports 1 "acad12 HCA-1"
Ca      : 0xb8599f0300f7072c ports 1 "acad11 HCA-1"
Ca      : 0x1c34da030060cec8 ports 1 "acad08 HCA-1"
Ca      : 0x1c34da030060cd40 ports 1 "acad07 HCA-1"
Ca      : 0x1c34da0300538294 ports 1 "Mellanox Technologies Aggregation Node"
Ca      : 0x043f720300e8a31a ports 1 "acad04 HCA-1"
```



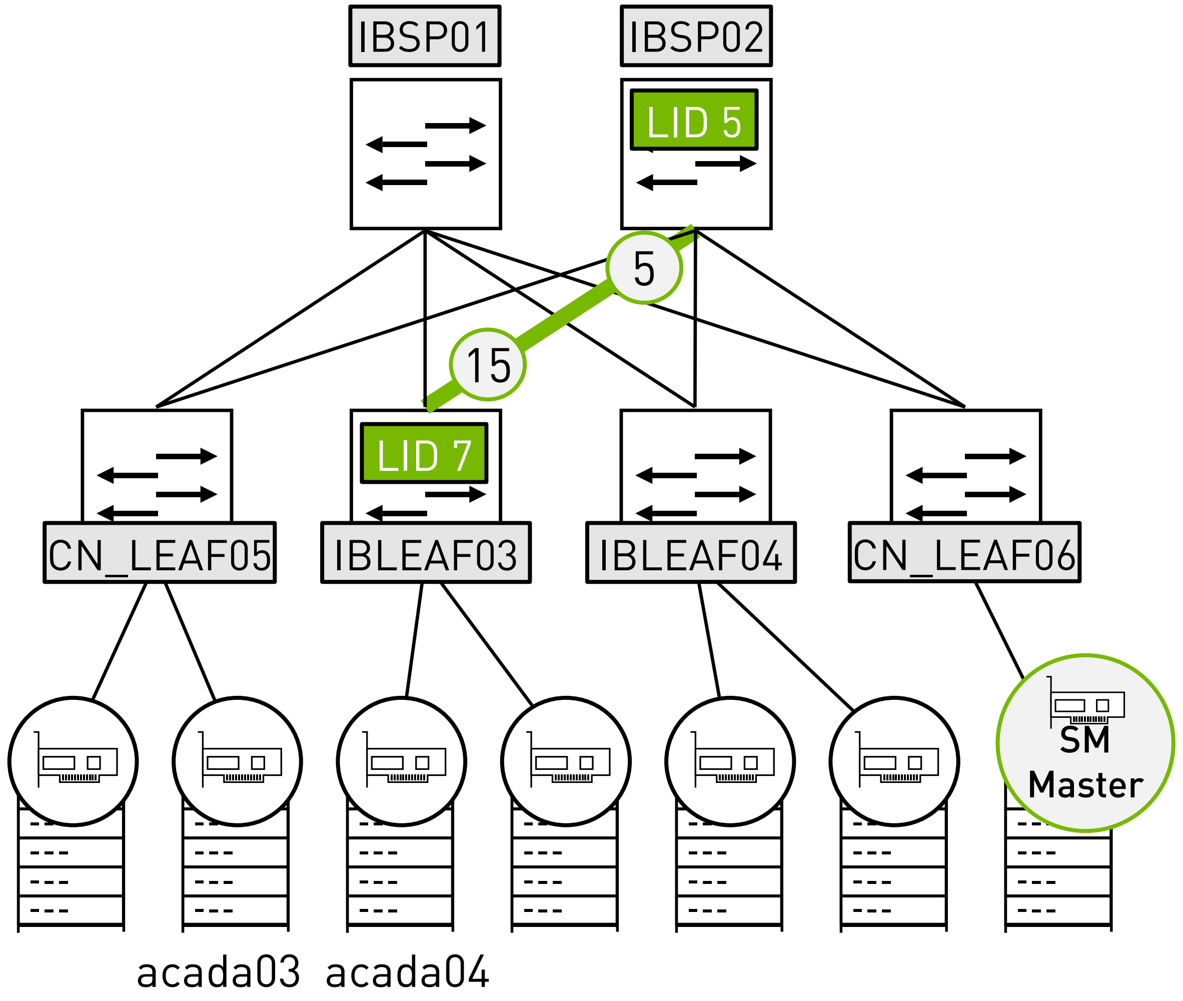
ibnodes—Display All Cluster Nodes Switches and Hosts

```
acadadmin@acad05:~$ sudo ibnodes -C mlx5_0

Ca       : 0xb8599f030014b8b8 ports 1 "Mellanox Technologies Aggregation Node"
Ca       : 0x1c34da030060cdb8 ports 1 "acad12 HCA-1"
Ca       : 0xb8599f0300f7072c ports 1 "acad11 HCA-1
Ca       : 0x1c34da030060cd30 ports 1 "acad06 HCA-1"
Ca       : 0xb8599f0300f707e4 ports 1 "acad05 HCA-1"
Switch   : 0xb8599f030014b8b0 ports 81 "CN_LEAF06" base port 0 lid 10 lmc 0
Switch   : 0x1c34da030053828c ports 41 "MF0;IBLEAF03:MQM8700/U1" enhanced port 0 lid 7 lmc 0
Switch   : 0xb8599f030009118e ports 81 "CN_LEAF_05" base port 0 lid 8 lmc 0
```



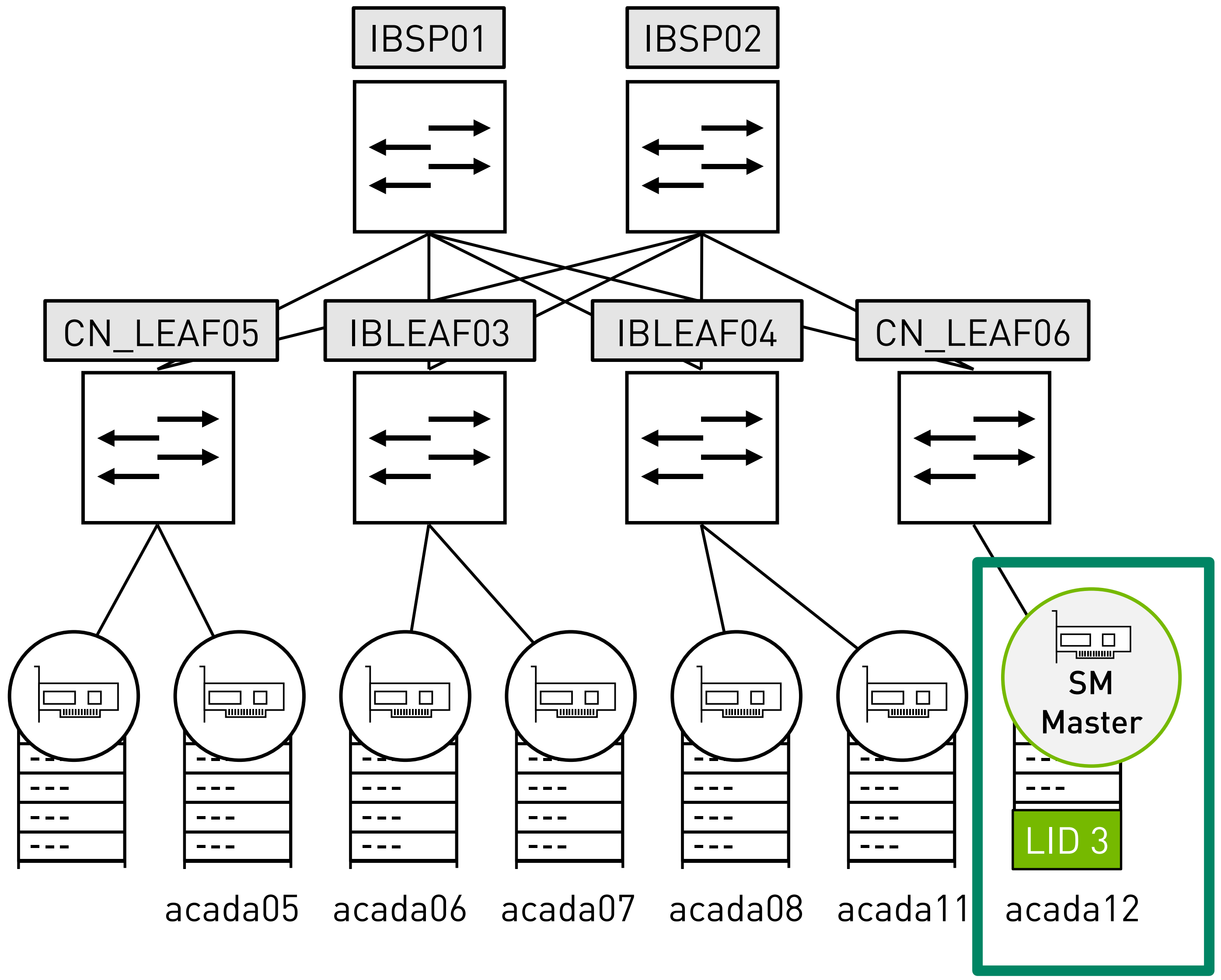
iblinkinfo—Displays Full Topology Peer to Peer Link Information

```
sudo iblinkinfo
Switch: 0x1c34da030053828c MF0;IBLEAF03:MQM8700/U1:
           7    1[  ] ==( 4X        53.125 Gbps Active/  LinkUp)==>      2    1[  ] "acad03 HCA-1" ( )
           7    2[  ] ==(                Down/ Polling)==>          [  ] "" ( )
           7    3[  ] ==(                Down/ Polling)==>          [  ] "" ( )
           7    4[  ] ==( 4X        53.125 Gbps Active/  LinkUp)==>      3    1[  ] "acad04 HCA-1" ( )

           7    9[  ] ==(                Down/ Polling)==>          [  ] "" ( )
           7   10[  ] ==(                Down/ Polling)==>          [  ] "" ( )
           7   11[  ] ==( 4X        53.125 Gbps Active/  LinkUp)==>      9    5[  ] "MF0;IBSP01:MQM8700/U1" ( )

           7   14[  ] ==(                Down/ Polling)==>          [  ] "" ( )
           7   15[  ] ==( 4X        53.125 Gbps Active/  LinkUp)==>      5    5[  ] "MF0;IBSP02:MQM8700/U1" ( )
```



sminfo—Display Cluster Master Subnet Manager

```
acadadmin@acad05:~$ sudo sminfo

sminfo: sm lid 3 sm guid 0x1c34da030060cdb9, activity count 742718 priority 15 state 3 SMINFO_MASTER
```

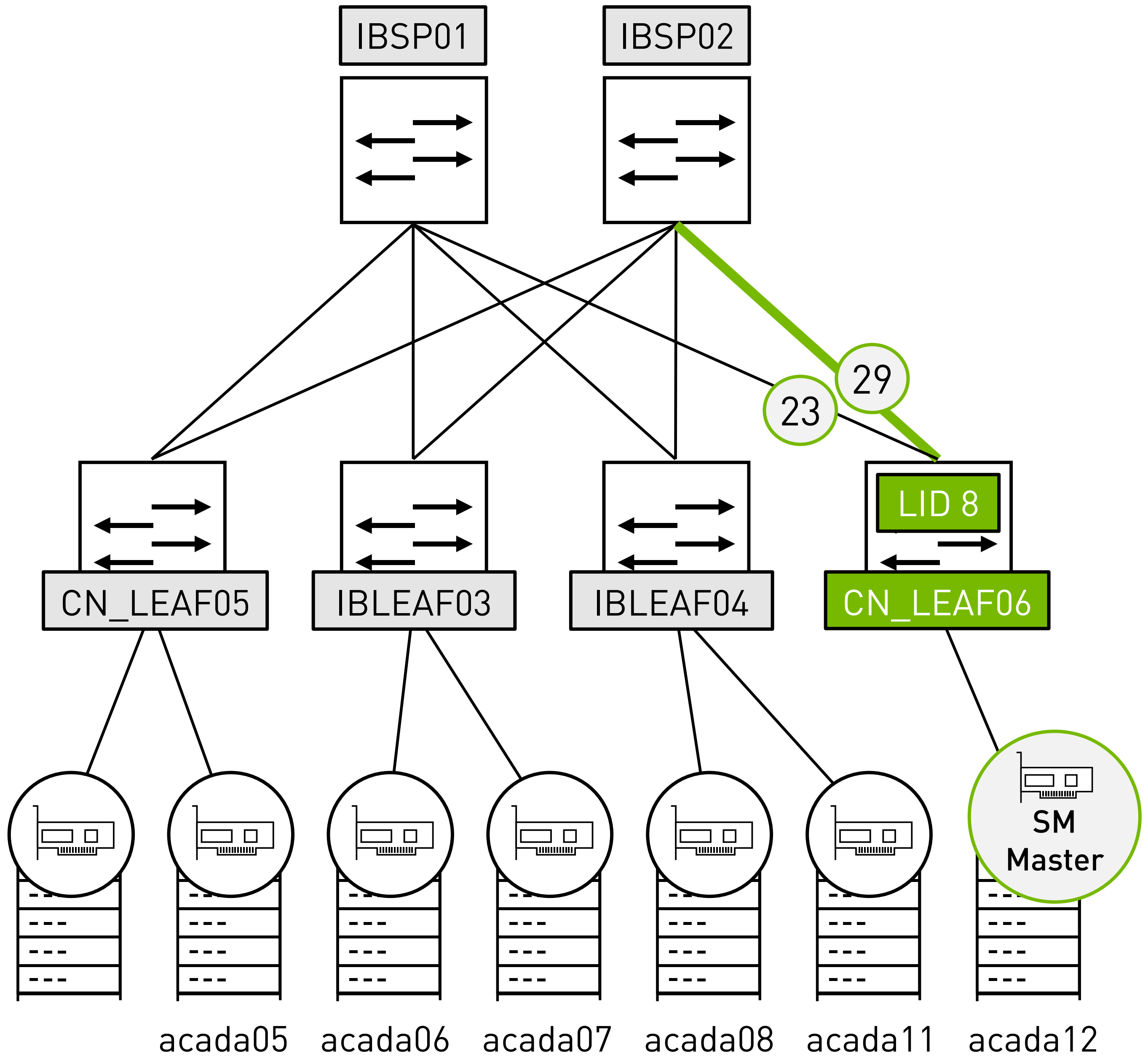


ibportstate

Display any port status within the cluster

```
sudo ibportstate -C mlx5_0  <switch LID>  < Port>
```

```
acadadmin@acad05:~$ sudo ibportstate -C mlx5_0 8  29
Switch PortInfo:
# Port info: Lid 8 port 29
LinkState:.......................Active
PhysLinkState:...................LinkUp
Lid:.............................0
SMLid:...........................0
LMC:.............................0
LinkWidthActive:.................4X
LinkSpeedExtSupported:...........14.0625 Gbps or 25.78125 Gbps or
53.125 Gbps Gbps
```



ib_write_lat (Test Process)

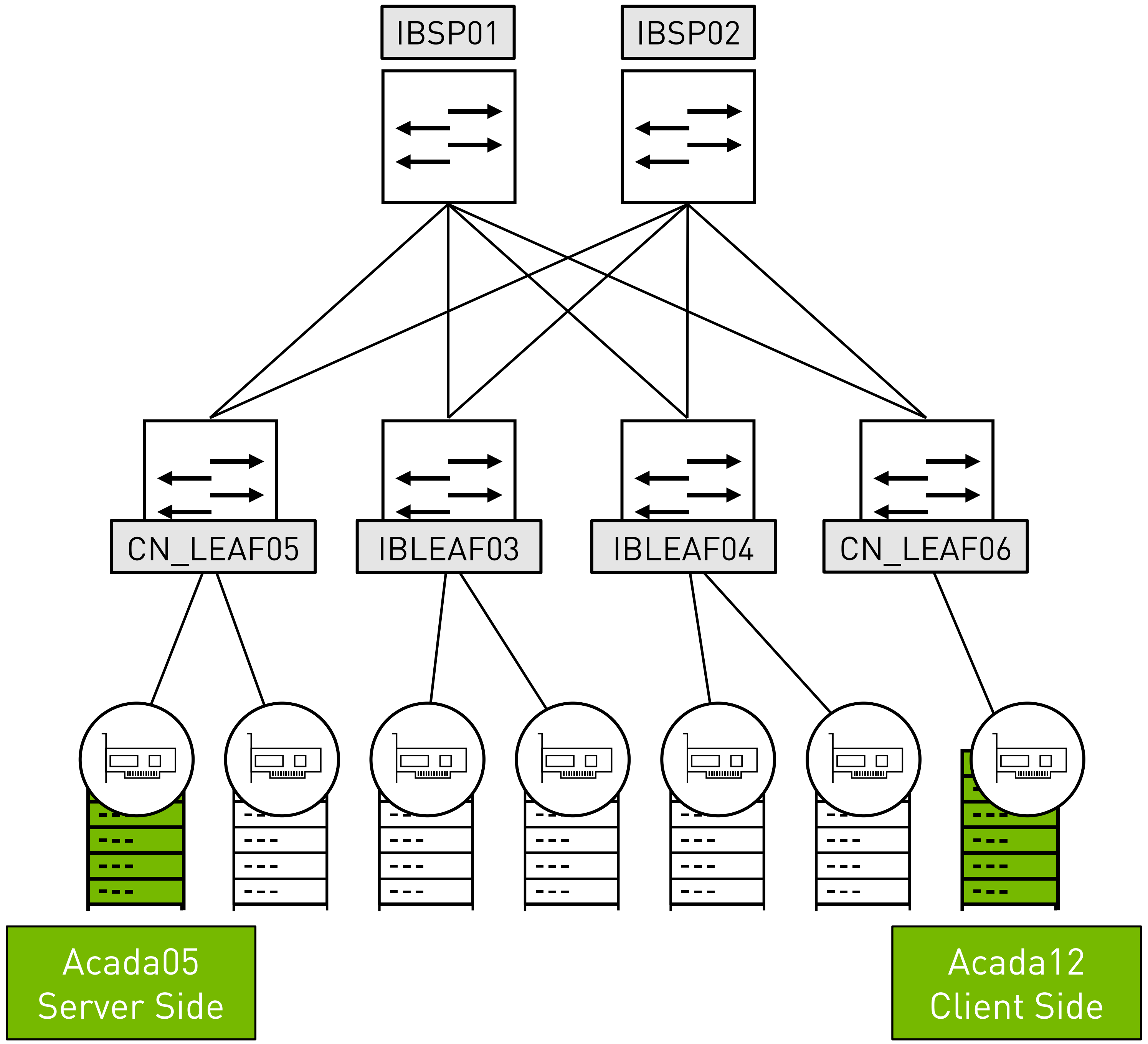
Create RDMA session between 2 nodes and check latency

Server Side

```
[root@acad05 ~]# sudo ib_write_lat -d mlx5_0 -F

************************************
* Waiting for client to connect... *
************************************
```

Client Side

```
[root@macad12 ~]# sudo ib_write_lat -d mlx5_0 acad05 -F

************************************
* Waiting for client to connect... *
************************************
```



IBSP01

IBSP02

CN_LEAF05

IBLEAF03

IBLEAF04

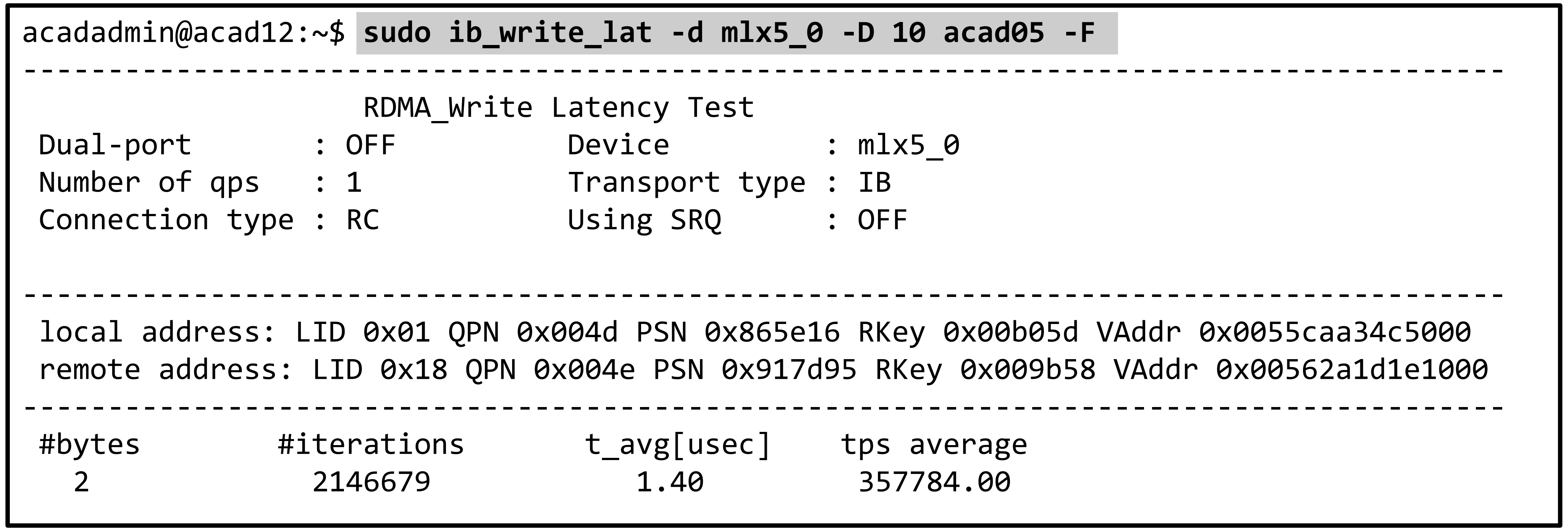
CN_LEAF06

Acada05
Server Side

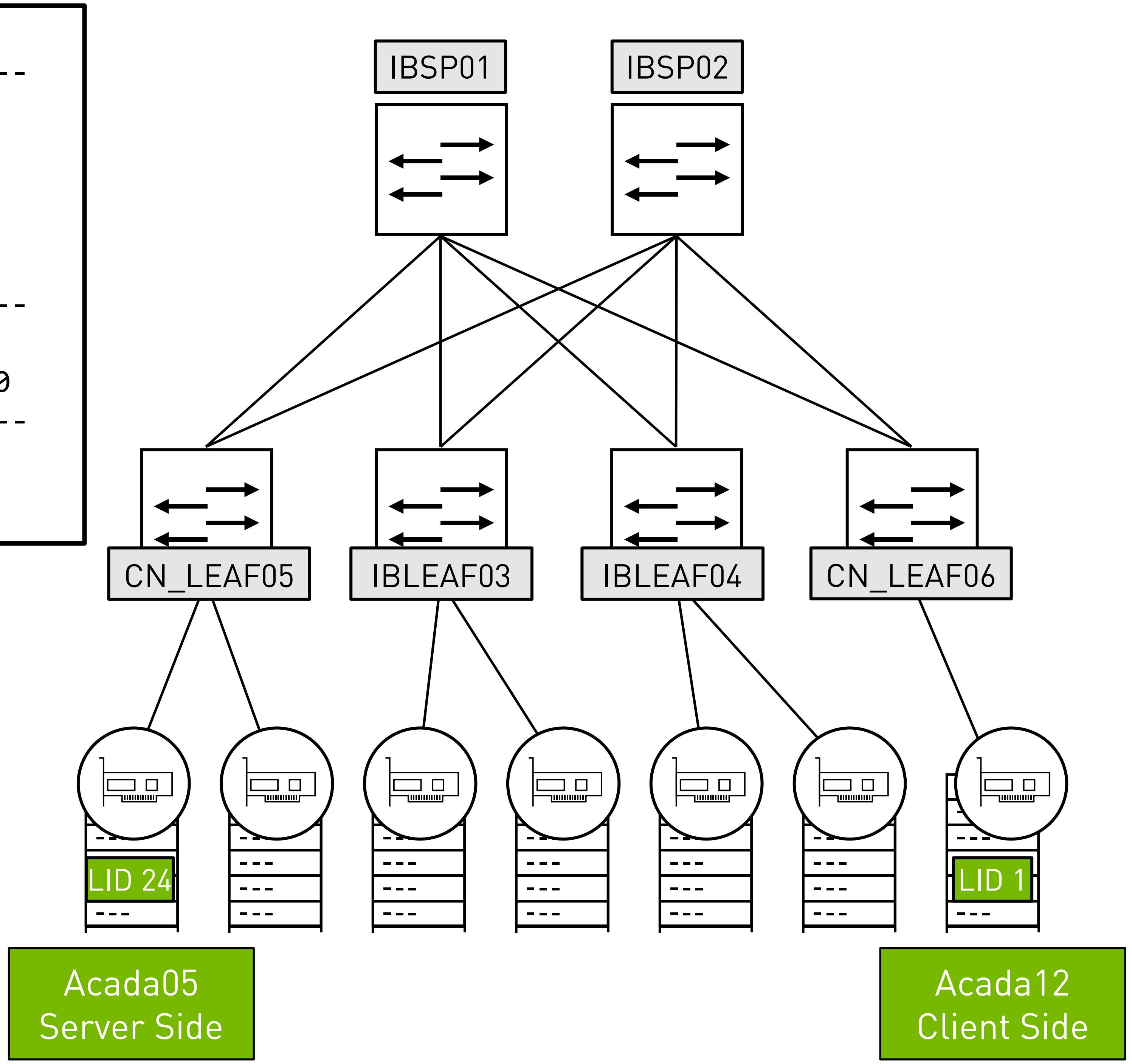
Acada12
Client Side

ib_write_lat (Test Results)

Create RDMA session between 2 nodes and check latency



```
acadadmin@acad12:~$ sudo ib_write_lat -d mlx5_0 -D 10 acad05 -F
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF


---------------------------------------------------------------------------------------
 local address: LID 0x01 QPN 0x004d PSN 0x865e16 RKey 0x00b05d VAddr 0x0055caa34c5000
 remote address: LID 0x18 QPN 0x004e PSN 0x917d95 RKey 0x009b58 VAddr 0x00562a1d1e1000
---------------------------------------------------------------------------------------
 #bytes        #iterations       t_avg[usec]    tps average
   2             2146679            1.40         357784.00
```



ib_write_bw (Test Process)

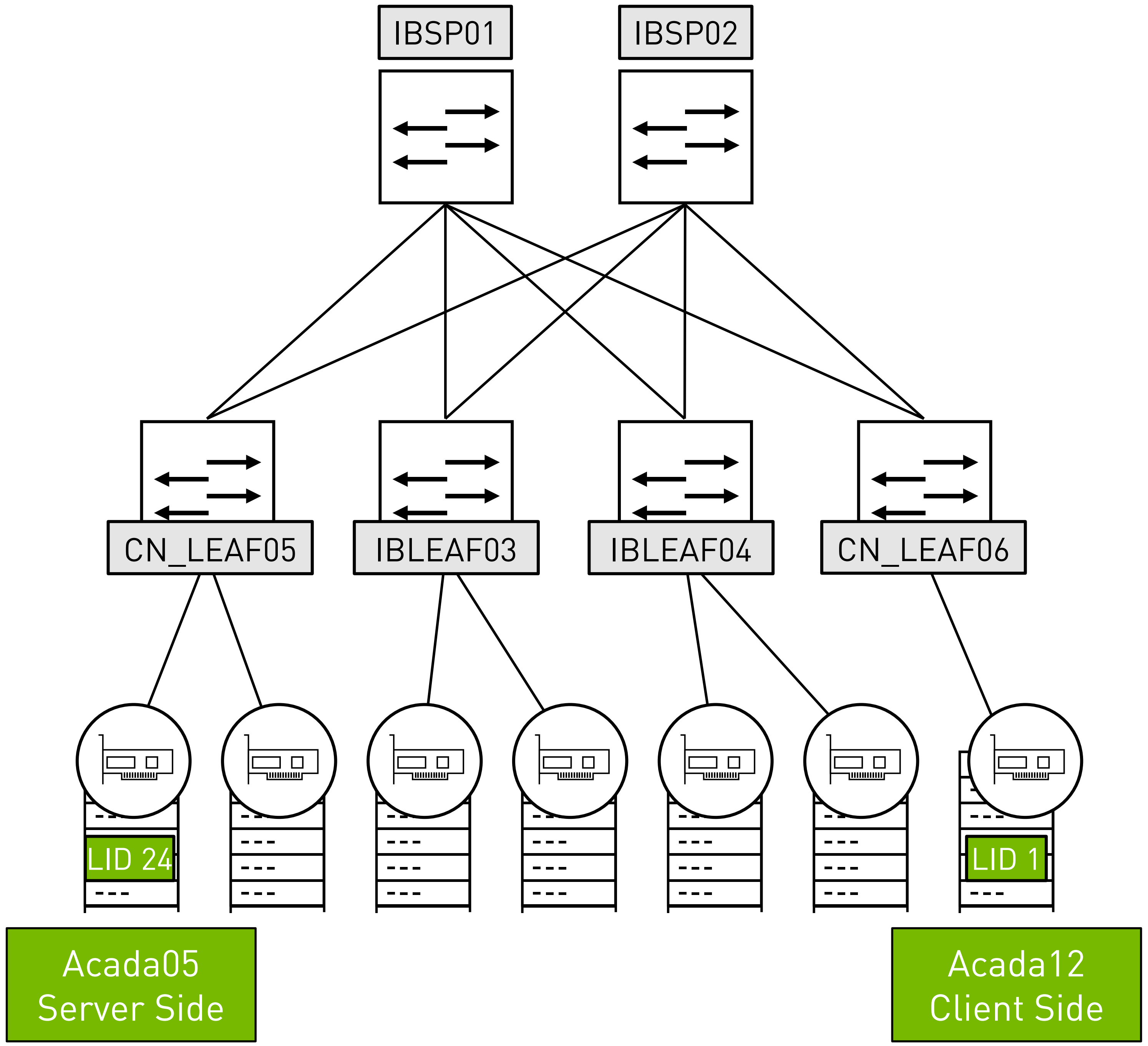
Create RDMA session between 2 nodes and check bandwidth

Server Side

```
[root@acad05 ~]# sudo ib_write_bw -d mlx5_0 --report_gbits
************************************
* Waiting for client to connect... *
************************************
```

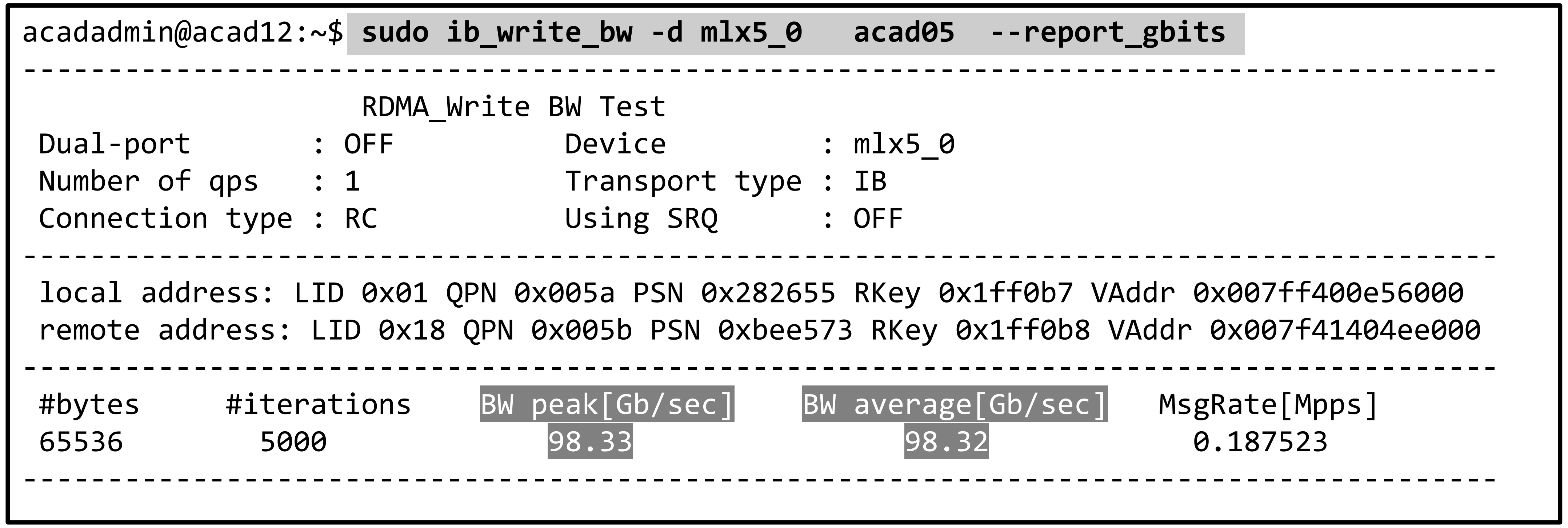
Client Side

```
[root@macad12 ~]# sudo ib_write_bw -d mlx5_0 acad05 --report_gbits

************************************
* Waiting for client to connect... *
************************************
```

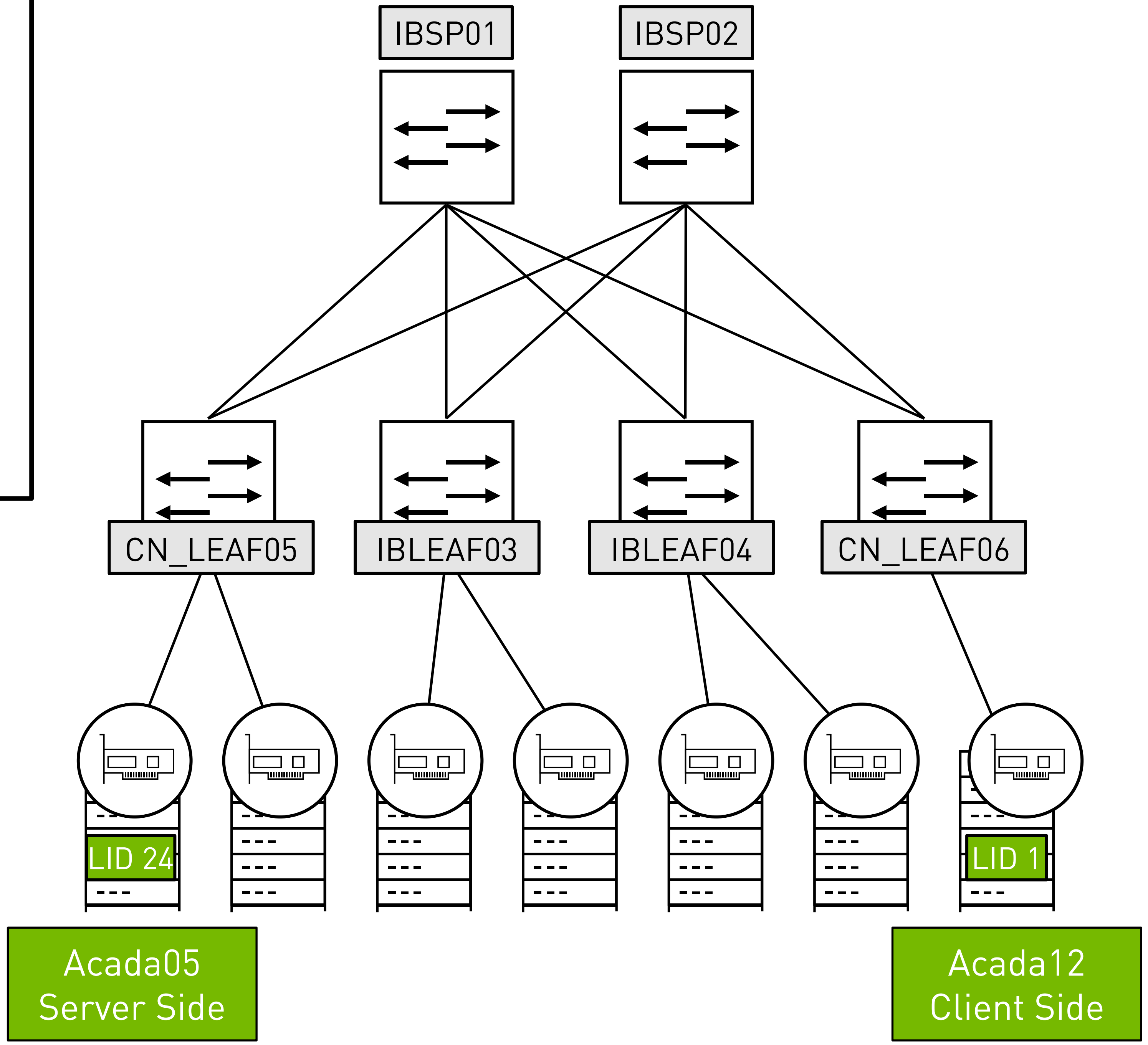


ib_write_bw (Test Results)

Create RDMA session between 2 nodes and check bandwidth



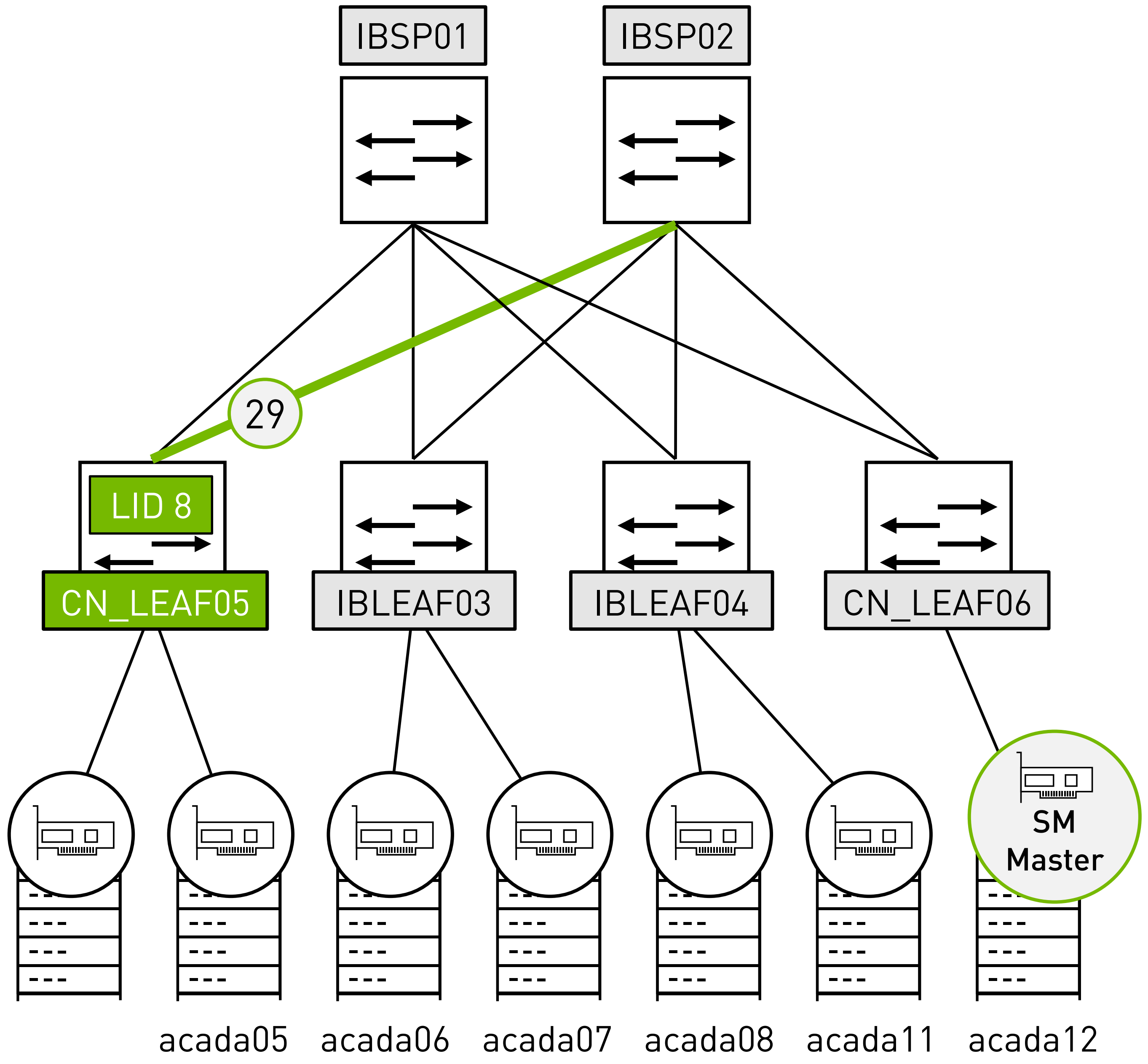
```
acadadmin@acad12:~$ sudo ib_write_bw -d mlx5_0   acad05   --report_gbits
---------------------------------------------------------------------------------------
                    RDMA_Write BW Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
---------------------------------------------------------------------------------------
 local address: LID 0x01 QPN 0x005a PSN 0x282655 RKey 0x1ff0b7 VAddr 0x007ff400e56000
 remote address: LID 0x18 QPN 0x005b PSN 0xbee573 RKey 0x1ff0b8 VAddr 0x007f41404ee000
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 65536        5000            98.33              98.32               0.187523
---------------------------------------------------------------------------------------
```



perfquery - Display Any Port Counters Record

```
sudo perfquery -x  -C mlx5_0  <switch LID>  < Port>
```

```
acadadmin@acad10:~$ sudo perfquery -x  -C mlx5_0 8 29
# Port extended counters: Lid 8 port 29 (CapMask: 0x5300 CapMask2:
0x0000002)
PortSelect:......................29
CounterSelect:...................0x0000
PortXmitData:....................157458328805003
PortRcvData:.....................25501731359306
PortXmitPkts:....................154041505906
PortRcvPkts:.....................42206502318
PortUnicastXmitPkts:.............154041505400
PortUnicastRcvPkts:..............42206502318
PortMulticastXmitPkts:...........506
PortMulticastRcvPkts:............0
CounterSelect2:..................0x00000000
SymbolErrorCounter:..............0
LinkErrorRecoveryCounter:........0
LinkDownedCounter:...............0
PortRcvErrors:...................0
PortRcvRemotePhysicalErrors:.....0
PortRcvSwitchRelayErrors:........0
PortXmitDiscards:................0
PortXmitConstraintErrors:........0
PortRcvConstraintErrors:.........0
LocalLinkIntegrityErrors:........0
ExcessiveBufferOverrunErrors:....0
VL15Dropped:.....................0
PortXmitWait:....................12302767437877
```



InfiniBand Operational commands

- Link commands
- Network commands
- MLX MFT tools

Basic InfiniBand Commands

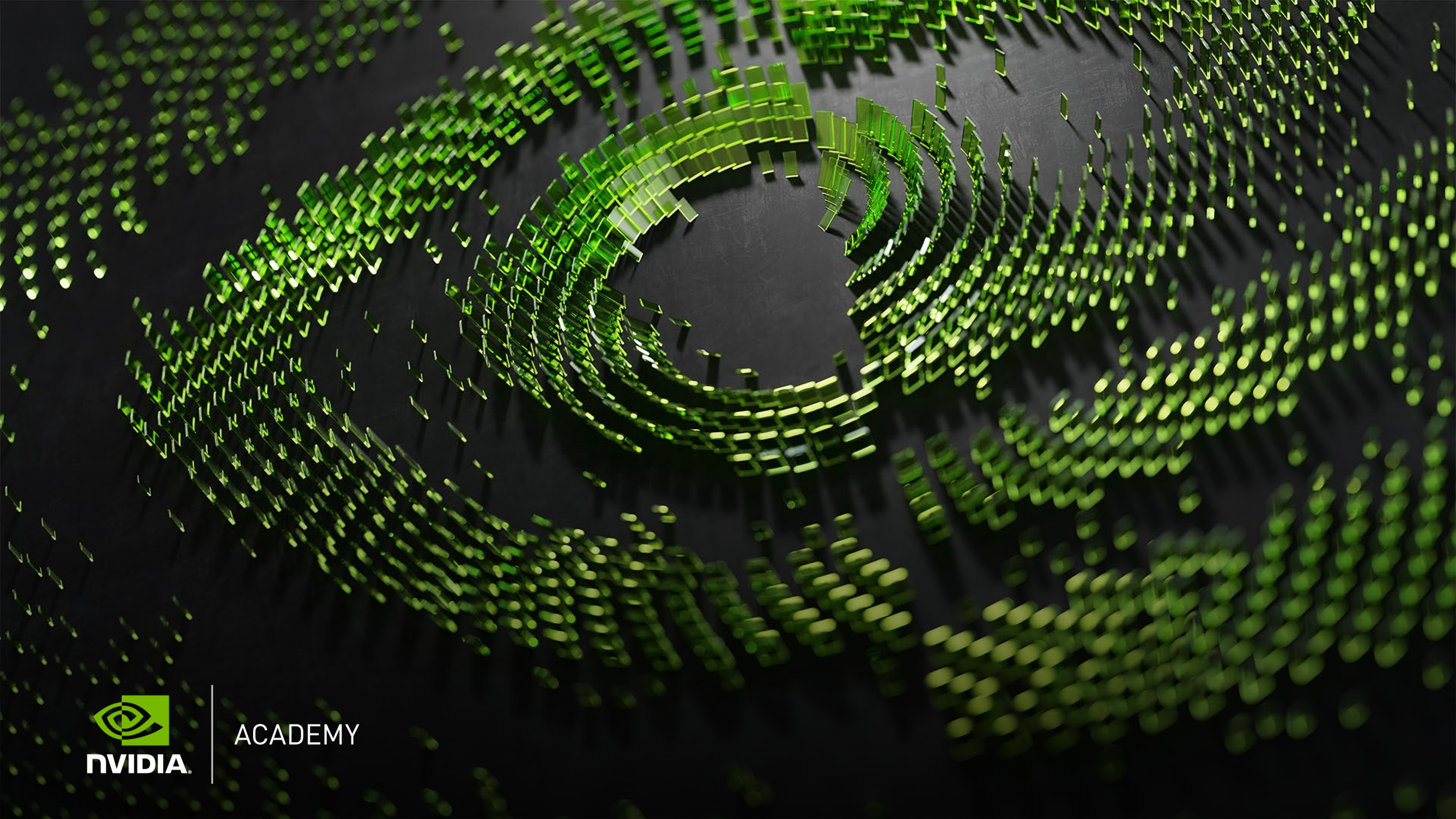
1	**ibstat**	Port information and Link operational status	**sudo ibstat**
2	**ibv_devices**	Devices supported by the OFED driver	**sudo ibv_devices**
3	**ibv_devinfo**	Ports enhanced details	**sudo ibv_devinfo**
4	**ibdev2netdev**	IPoIB ports name mapping and status	**sudo ibdev2netdev**
5	**ibhosts**	Channel adapters detected on this subnet & ANs	**sudo ibhosts**
6	**ibswitches**	Switches detected on this subnet	**sudo ibswitches –C mlx5_0**
7	**ibnodes**	Inclusive nodes detected on this subnet	**sudo ibnodes -C mlx5_0**
8	**sminfo**	Identifies the active subnet manager identifiers and priority	**sudo sminfo –C mlx5_0**

Basic InfiniBand Commands

1	**ibportstate**	Display and set local and remote switch ports	**sudo ibportstate –C mlx5_0 <switch LID> < Port>**
2	**ibtracertnfo**	Peer to peer link information for all subnet ports	**sudo iblinkinfo –C mlx5_0** **sudo iblinkinfo –C mlx5_0 -S <SWITCH GUID>**
3	**ibnetdiscover**	Displays full network topology , details end to end connections HCAs and switches ports	**sudo ibnetdiscover -C mlx5_0**
4	**ibroutes**	Displays Linear /static forwarding tables content	**sudo ibroute < Switch LID > -C mlx5_0**
5	**ibtracert**	Displays the route a packets takes between the source and destination LIDS	**sudo ibtracert < SLID> < DLID>**
6	**ibping**	Verifying InfiniBand L2 connection between 2 Hosts	**Server : sudo ibping -C mlx5_0 -S** **client : sudo ibping -C mlx5_0 < server LID>**

Unit Summary

- InfiniBand Network Stack
- IB Architecture Layers
- Data Packet Structure
- Subnet Manager (SM)
- Fabric Addressing – GUIDs, LIDs, GIDs
- Fabric Segmentation with Partitions
- OFED and OFED Utilities



NVIDIA ACADEMY