NVIDIA IndeX
System settings

20 February 2024
Version 2.3

NVIDIA IndeX – System settings

Copyright Information

Contents

1 BIOS settings

Ensure power save mode is disabled.

2   GPU driver

Disable ECC on all GPUs for performance improvement:

1. Get a list of GPUs:

```
/usr/bin/nvidia-smi -L
GPU 0: Tesla M2070 (S/N: 0323910065135)
GPU 1: Tesla M2070 (S/N: 0323910067128)
```

2. Disable ECC (where `<id>` is the GPU ID):

```
/usr/bin/nvidia-smi -i <id> -e 0
```

3. Reset the GPU state to effect your changes (where `<id>` is the GPU ID):

```
/usr/bin/nvidia-smi -i <id> -r
```

Set the amount of GPU memory that NVIDIA IndeX can use for volumes. Set this value in `index::CUDA_volume_memory` in the `project.prj` file.

> **Note:** A value of 80-90% of the total available GPU memory is recommended. For example, if the amount of available GPU memory is 12GB, a safe value for `index::CUDA_volume_memory` is 11000.

If the Nouveau GPU driver[1] is installed, installation of the original NVIDIA GPU driver will fail. To uninstall the Nouveau driver:

1. Add `rdblacklist=nouveau` to kernel attributes in `/boot/grub/menu.lst`
2. Execute:

```
echo "blacklist nouveau" > /etc/modprobe.d/nouveau.conf
```

3. Reboot.

---

1. Nouveau is an open-source NVIDIA GPU driver. For more information, see the freedesktop.org Nouveau webpage.)

3   Network

The following sections describe network configuration requirements and recommendations.

3.1 Hardware

For cluster rendering a Gigabit Ethernet or better is recommended. Using slower networks also works correctly, but can slow down cluster rendering significantly. Using Infiniband networks is also supported and can result in faster network communication.

In the default network configuration DiCE may generate large amount of multicast traffic. Sophisticated switches allow enabling a feature called multicast snooping or multicast spoofing. This will allow the switch to forward multicast traffic only to hosts that have registered interest in those multicast groups. Consumer grade network switches are usually not able to handle this amount of network traffic. The Cisco 3750G and 2960G switches, and the Foundry SuperX switch have been found to work well.

DiCE also supports cluster formation in TCP mode. TCP mode is intended only for environments where multicast is not available.

Note that DiCE built-in load balancing and resource allocation strategies assume that all memory, all CPU cycles, and the entire network bandwidth are at its disposal. It will dynamically ramp up its usage to optimally use all available resources. In general, it is not a good idea to run other large or interactive programs (or additional DiCE services) on a DiCE host.

3.2 Software configuration

3.2.1 UDP Multicast mode (default)

To set the network interface address, use the inet address displayed by running the `/sbin/ifconfig` command in a shell. In the options file, an address is specified by following the 4-tuple address with the subnet mask, expressed as 8, 16, or 24.

For example, if the inet address displayed by the `ifconfig` command is 192.21.120.2 and the subnet mask is 255.255.0.0, then the network interface should be set using 16 for the mask in the `project.prj` options file:

```
dice::network::cluster_interface_address = 192.21.0.0/16
```

If the subnet mask value is 255.255.255.0, then the mask value should be set to 24 in the options files with this statement:

```
dice::network::cluster_interface_address = 192.21.120.0/24
```

3.2.2 TCP Mode

TCP mode works by using the viewer node as the discovery node. All worker nodes will connect to it. As opposed to the UDP multicast mode, the user has to know the viewer node's IP or hostname. Besides the discovery address, the cluster interface should also be set for systems with multiple interfaces.

The cluster interface address is configured as in the UDP section. The discovery address is the IP of the viewer node. If the head node is on ip 192.21.120.5, then the discovery address should be set in the `project.prj` options file:

```
dice::network::discovery_address = 192.168.120.5
```

3.2.3 System configuration

1. Adjust network buffer size by editing `/etc/sysctl.conf`:

```
net.core.rmem_max = 8388608
net.core.wmem_max = 8388608
```

2. Check if UDP multicast is enabled on all cluster nodes. In a shell, executing:

```
/sbin/ifconfig
```

should show the following line:

```
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
```

Additional tests can be performed running `net_test` (part of the NVIDIA IndeX delivery).

1. On a chosen node start `net_test`.
2. Start `net_test -m` on all other cluster nodes. If multicast is working correctly, `net_test` statistics should show no multicast package loss.

The software utility `net_test_dice` emulates functionality of DiCE to test networking.

3.3 Infiniband (optional support)

For maximum performance set all nodes to "connected mode":

1. In file `/etc/infiniband/openib.conf`, enter:

```
set SET_IPOIB_CM=yes
```

2. Check by executing:

```
cat /sys/class/net/ib#/mode
```

4 User settings

A local user to run the NVIDIA IndeX software must be created on all nodes. Ensure the UID of the user is the same on all nodes.