Third Party Software: How do I configure BeeGFS to run on an InfiniBand interface?

How do I configure BeeGFS to do native IB rather than IP over IB?

Preliminary: BeeGFS Installation

By default BeeGFS is not installed on Bright Cluster Manager. Setting it up is straightforward; cm-beegfs-setup works as described in the administrator manual:

http://support.brightcomputing.com/manuals/8.1/admin-manual.pdf#page=387

Configuration Of BeeGFS Native IB Support

The following steps rely on the BeeGFS documentation at https://www.beegfs.io/wiki/NativeInfinibandSupport

After cm-beegfs-setup installation is finished, communication between management BeeGFS elements defaults to internalnet, as can be verified by running the commands:

```
# beegfs-ctl --listnodes --nodetype=storage --details

# beegfs-ctl --listnodes --nodetype=meta --details

# beegfs-ctl --listnodes --nodetype=client --details
```

For BeeGFS version 7.1 and above, BeeGFS communications can be made to switch over to the IB interface as follows:

The file:

```
/cm/images/default-image/etc/beegfs/beegfs-client-autobuild.conf
```

should be edited.

The line:

```
buildArgs=-j8
```

should be changed to:

```
buildArgs=-j8 BEEGFS_OPENTK_IBVERBS=1
```

Third Party Software: How do I configure BeeGFS to run on an InfiniBand interface?

The package libbeegfs-ib should be installed into the image that is used by the BeeGFS nodes:

```
# chroot /cm/images/default-image
# yum install libbeegfs-ib
```

Verifying

At this point, the beegfs-ctl commands that were run earlier on in this article should output that BeeGFS is using the IB interface:

```
[root@goofy default-image]# beegfs-ctl --listnodes --nodetype=meta --details
node1 [ID: 1]
 Ports: UDP: 8005; TCP: 8005
 Interfaces: ib0(RDMA) br0:vxlan(TCP) br0(TCP) ib0(TCP)
```

The text "RDMA" here means that the associated interface is enabled for the native Infiniband protocol (IB verbs).

Additional configuration: disabling the ibacm service

A typical source of trouble having the `ibacm` service (`/etc/init.d/ibacm`) still running on the machines. This service causes RDMA connection attempts to stall. It should be disabled in all nodes:

```
# systemctl stop ibacm.service
```

```
# systemctl disable ibacm.service
```

Additional Notes:

- More configuration examples can be seen at: https://www.beegfs.io/wiki/NativeInfinibandSupport#hn_59ca4f8bbb_4
- In an RDMA-capable cluster, there may still be some BeeGFS communication (especially communication with the management service, which is not performance-critical) that still uses TCP/IP and UDP/IP transfer. On some hardware the default "`connected`" IP-over-IB mode of InfiniBand and Omni-Path does not seem to work well and results in spurious problems. If that seems to be the case, then switching the IPoIB mode to "`datagram`" on all hosts should be tried.

Unique solution ID: #1442

Third Party Software: How do I configure BeeGFS to run on an InfiniBand interface?

Author: Giorgi Gabunia
Last update: 2018-12-04 17:22