Myrinet Roll Experiences

Rocks-A-Palooza I

Track 2

Case Studies of Existing Rolls

SDSC

Myrinet Roll Function

- Automatically install and configure on each compute node with Myrinet card:
  - Myricom's messaging layer (GM)
  - MPICH libraries that utilize GM
- Automatically install/configure MPICH environment on frontend
  - Compiler scripts (mpicc, etc.)
  - Libraries

Issue

- GM is a kernel module
  - The version of the module must match the version of the running kernel
- One solution:
  - Provide gm.o/gm.ko for each verison of the kernel
- Problems
  - Kernel RPM is updated multiple times per year
  - How to provide support for custom user kernels?

Dealing with the Driver Issue

- Trick: Build the gm module during the first boot after installation
- Guarantees the version gm.o/gm.ko matches that of the underlying kernel
- Note: Driver not rebuilt on a normal reboot
  - Only on first boot after installation

How We Do It

- Create a source RPM
- Place SRPM in /opt/rocks/SRPMS/ on compute node
  - All SRPMs in /opt/rocks/SRPMS/ will be rebuilt and installed during the boot
    - /etc/rc.d/rocksconfig.d/pre-10-src-install is the script that does this function
  - The script removes SRPM after building
    - This ensures the SRPM is not built on each boot

The SRPM

- Check if a Myrinet board is installed

```
from src/roll/myrinet/src/gm/gm.spec.in

%build
#
# check if a myrinet card is installed. if not, cancel this build
# by exiting the script
#
/sbin/lspci -n -d "10e8:*" >  /tmp/lspci.output
/sbin/lspci -n -d "14c1:*" >> /tmp/lspci.output
if [ ! -s /tmp/lspci.output ]
then
        rm /tmp/lspci.output
        echo "This machine doesn't have a Myrinet card installed."
        echo "The driver will not be built."
        echo

        exit -1
fi
```

Prep The Source Tree

- Another piece of code preps the source tree for device driver building

```
from /etc/rc.d/rocksconfig.d/pre-09-prep-kernel-source

KERNEL_VERSION=`uname -r | awk '{ \
        split($0, values, "."); \
        printf("%s.%s", values[1], values[2]); \
}'`

case $KERNEL_VERSION in
2.6)
        cd /usr/src/linux-2.6 ;;
2.4)
        cd /usr/src/linux-2.4 ;;
esac

.
.
.
```

```
.
.
.

if [ ! -f .config ]
then
        make mrproper 2>&1 > /dev/null

        #
        # create the name of the .config file
        #
        KERNEL_CONFIG_NAME="config-"`uname -r`

        cp /boot/$KERNEL_CONFIG_NAME .config
fi

make oldconfig 2>&1 > /dev/null
make modules_prepare 2>&1 > /dev/null
```

Prep The Source Tree

- /etc/rc.d/rocksconfig.d/pre-09-prep-kernel-source is installed (and run) on all nodes
- Benefit: any device driver that needs to be built doesn't need to prepare the kernel source for building

Other Rolls Which Use This Trick

- Infinicon's Infiniband Roll
  - Builds IB drivers
- Viz Roll
  - Builds nvidia device driver