



National Energy Research Scientific Computing Center (NERSC)

Highly Scalable Networking Configuration Management For Highly Scalable Systems

Nicholas P. Cardo
NERSC Center Division, LBNL
CUG 2008, Helsinki





Setting the Stage

- Systems are getting large
- Systems are getting more complex
- Networking capability is increasing
- Systems have many nodes
- Systems have many network interfaces
- Systems have many different interface types





Cardo's Corollary #3

If it can change, it will change.





Network Files

- `/etc/sysconfig/network/ifcfg-eth0`
- `/etc/sysconfig/network/ifcfg-eth1`
- `/etc/sysconfig/network/ifcfg-ss`
- `/etc/sysconfig/network/scripts/`





Standard Interface Configuration

- `cd /etc/sysconfig/network`
- `vi ifcfg-eth0`
 - Set IPADDR
 - Set NETMASK
 - Set MTU
 - Set other parameters

For each interface.

On each node.



Now on the Cray

- ssh root@boot
- xtopview -n <nid>
- cd /etc/sysconfig/network
- xtspec ifcfg-eth0
- vi ifcfg-eth0
 - Set IPADDR
 - Set NETMASK
 - Set MTU
 - Set other parameters
- exit
- c
- <some text>
- .

**Repeat per
interface**

**Repeat per
node**



Lets Do the Math

56 Nodes

1 SS per node

5 minutes per change

20 10Gb

16 1Gb

92 Interfaces @ 5 minutes = 460 minutes = 7 hours 40 minutes





Goals

- Single points of data entry for routing and interface configurations.
- Scalable to support all service nodes.
- Flexible to handle complex routing if necessary.
- Minimal specialization.





The Setup

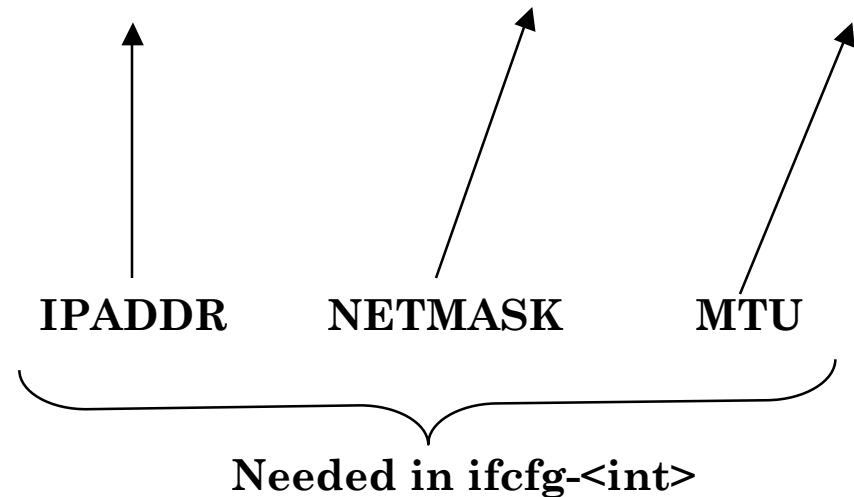
- `ssh root@boot`
- `xtopview`
- `mkdir /etc/NERSC`
- `cd /etc/sysconfig`
- `mv ifcfg-ss /etc/NERSC`
- `mv ifcfg-eth0 /etc/NERSC`
- `ln -s /etc/NERSC/ifcfg-eth0 ifcfg-eth0`
- `ln -s /etc/NERSC/ifcfg-ss ifcfg-ss`





Interface Configuration File

```
# Node      Location  Type   Dev  Address          Netmask          MTU
nid04100    c1-0c0s1n0 login01 eth0  128.55.81.34     255.255.248.0    9000
nid04100    c1-0c0s1n0 login01 ss    128.55.42.134    255.255.255.192 NA
```



Route Configuration File

Do this in POST_UP_SCRIPT

#	Node	Location	Type	Dev	Route	Gateway	Netmask	MTU	Fwd
	nid04100	c1-0c0s1n0	login01	eth0	default	128.55.80.1	NA	NA	0
	nid04100	c1-0c0s1n0	login01	ss	128.55.32.0	128.55.42.130	255.255.224.0	9000	0

```
/sbin/route add $TYP $ADR gw $GWY $MSK dev $DEV $MTU
```

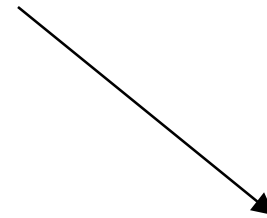
-net or -host

```
echo $FOR >/proc/sys/net/ipv4/ip_forward
```



ARP Table Pre-Load

```
/sbin/arp -f /etc/NERSC/arp-ss -i ss
```



```
128.55.42.130 00:00:10:00
```

Do This in POST_UP_SCRIPT





Summary

- One file text describes all interfaces
- One file text file describes all routes
- One time setup





Questions?

