V1

RCE Platform Technology (RPT)

Shelf Configuration, Networking, and SDK Tools

Sergio Maldonado (smaldona@slac.stanford.edu)

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

PPA Particle Physics & Astrophysics

RPT

Overview

We will cover:

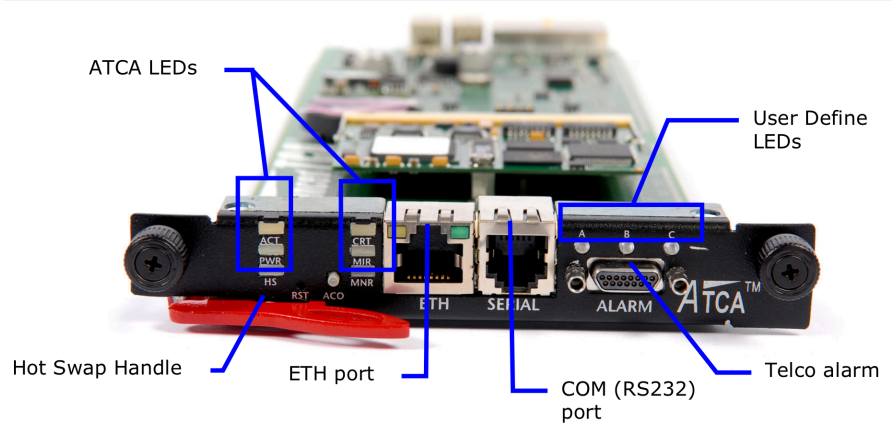
- Shelf Networking and Configuration
- DHCP Setup
- COB Network Topology
- SDK Tools
- Diskless RCE Node Using NFS

Go here!

https://confluence.slac.stanford.edu/display/RPTUSER/Home

Shelf Manager Networking

- The details of Shelf Manager configuration are vendor specific
  - Read your Shelf Manager documentation!
- Many allow various networking options
  - Static or Dynamic IP Address
- Our suggestions
  - Use a static IP address for your Shelf Manager
  - Don't place the Shelf Manager on the same network as the RCEs



Configuring an ASIS Shelf Manager

- BEWARE! Configuring a Shelf Manager is vendor specific
  - That said, most vendors base their systems on a Pigeon Point System solution, so something very like this will probably work
  - Read the documentation for your shelf
  - We'll use the Shelf Manager console port
- Set a static IP address for the Shelf Manager and store it in the Shelf FRU Information

```
shmm500 login: root
Password:
# clia setlanconfig 1 3 “172.21.6.94”
# clia setlanconfig 1 6 “255.255.255.0”
# clia setlanconfig 1 12 “172.21.6.1”
```

- Set the shelf Address
  - Intended to indicate the physical location of the shelf
  - Used by the RCE to identify its location

```
shmm500 login: root
Password:
# clia shelfaddress “shasta”
```

Shelf Networking Overview

When considering how to integrate the Shelf of COBs into an external network, you should consider

- How much bandwidth out of the Shelf do you require?
  - How many external network connections are required
    - If more than one connection is to be used, the external network must be configured in such a way as to prevent Layer 2 loops
- How will the Cluster Element IP addresses be assigned?
  - All RCE IP addresses are assigned via DHCP
    - Exception: the DHCP server DTM does not use DHCP
  - The DHCP server can reside in ONE of the DTMs or external to the Shelf
  - The internal DHCP server is enabled with a DIP switch on the COB
    - This state is indicated by the Amber LED on the Faceplate
- Best practice: Isolate the Shelf LAN from the outside world
  - Keep unintended external traffic from impacting the function of the clusters
  - Keep unintended internal traffic from impacting the wider network
  - RPT provides means of achieving this

COB DHCP Overview

- An RPT defined Record placed in the Shelf FRU Information defines the range of IP addresses available to the Shelf
  - The DHCP server DTMs self-assign their IP address based on their physical slot number and the IP address range found in the Shelf IP Info Record
  - If the DHCP server is internal to the Shelf, the IP addresses it assigns are taken from the range specified in the Shelf IP Info Record
- Best practice: Don't use the RCE MAC addresses to assign IP addresses
  - Retain the flexibility to move/replace boards without having to reconfigure the DHCP server
    - Don't underestimate the administrative burden
    - Why add an unnecessary step to swapping in a spare board?
  - The actual IP addresses assigned are not so important
    - The SDK provides tools which allow you to interact with the RCEs based on their location

Shelf IP Information Record

```
rddev110:bin$ display_shelf_ip_info --shelf shasta-sm

Shelf FRU Info previous commit timestamp is Wed Dec 31
16:00:00 1969

=====================================================
Shelf IP Info
-----------------------------------------------------
SLAC Shelf IP Information (shasta-sm)
-----------------------------------------------------
VLAN ID (Valid): 001 (1)
Discard Untagged: 0
Discard Tagged: 1
Discard Boundary Viol: 0
Group Base: 192.168.204.1
Group End: 192.168.204.254
Subnet Mask: 255.255.255.0
Gateway: 0.0.0.0
-----------------------------------------------------
```

- The Shelf IP Information Record is an OEM record which is loaded into the Shelf FRU Info
  - Utilities for loading, displaying, and deleting this record are included in the SDK
- Used to specify the pool of IP addresses available to the onboard DHCP server
  - Group Base – Group End
- Used to specify the DHCP server DTM IP addresses
  - DHCP server DTM addresses are filled from the “top” of the range following the rule
    - IP address = (Group End) - (Physical Slot)
- Used to identify VLAN used by the shelf and rules for discarding packets based on VLAN ID
  - Discard all VLAN tagged packets
  - Discard all VLAN untagged packets
  - Discard all packets whose VLAN tag doesn’t match the shelf VLAN

Setting the Shelf IP Info Record

```
Usage is: set_shelf_ip_info --shelf <shelf_ip> [OPTIONS]

Required arguments: -s,--shelf=SHELF_IP            IP address of shelf
                                                   with FRU info to set

Optional arguments:
When specified w/o parameters, the optional arguments will assume  their
default values. When omitted, the values present in the Shelf FRU Information
will be retained.

-i, --vlan=VLAN_ID                  VLAN ID (12-bits)
-u, --untagged=DISCARD              When DISCARD is 1, incoming
                                    untagged frames will be discarded.
-t, --tagged=DISCARD                When DISCARD is 1, incoming
                                    tagged frames will be discarded.
-x, --boundary-violations=DISCARD   When DISCARD is 1, incoming
                                    boundary violations are discarded.
-b, --group-base=BASE_IP            BASE_IP is the base address for
                                    the block of addresses available
                                    to the shelf. Host ID of all ones
                                    or all zeros are not permitted.
-e, --group-end=END_IP              END_IP is the end address for
                                    the block of addresses available
                                    to the shelf. Host ID of all ones
                                    or all zeros are not permitted.
-m, --subnet-mask=SUBNET_MASK       SUBNET_MASK is the subnet mask
                                    for the shelf. It defines the
                                    network assigned to the shelf.
-g, --gateway=GATEWAY_IP            GATEWAY_IP is the IP address
                                    of the  gateway
-d, --dry-run                       When this option is used, the
                                    nothing is written back to the shelf.
-f, --force                         When this option is used, write to
                                    the shelf even if there are no changes.
                                    When there is no record present in the
                                    shelf, this can be used w/o any other
                                    configuration options to create a
                                    default record
-v, --verbose=VERBOSITY             VEBOSITY can range from 1 to 3.

Return Value:
0 if OK
1 if there is a problem with the arguments provided
2 if there are problems communicating with the shelf
```

- Command included in the SDK `set_shelf_ip_info` is used to write/ modify the Shelf IP Information Record
- `group-base` and `group-end` cannot be the lowest or highest addresses in the network
- The number of addresses between `group-base` and `group-end` must be large enough for each configuration
  - Internal DHCP
    - All RCEs in the Shelf and any external
- Won't let you specify an inconsistent set of IP parameters
  - All addresses must be on the same network
  - Consistent with the subnet mask

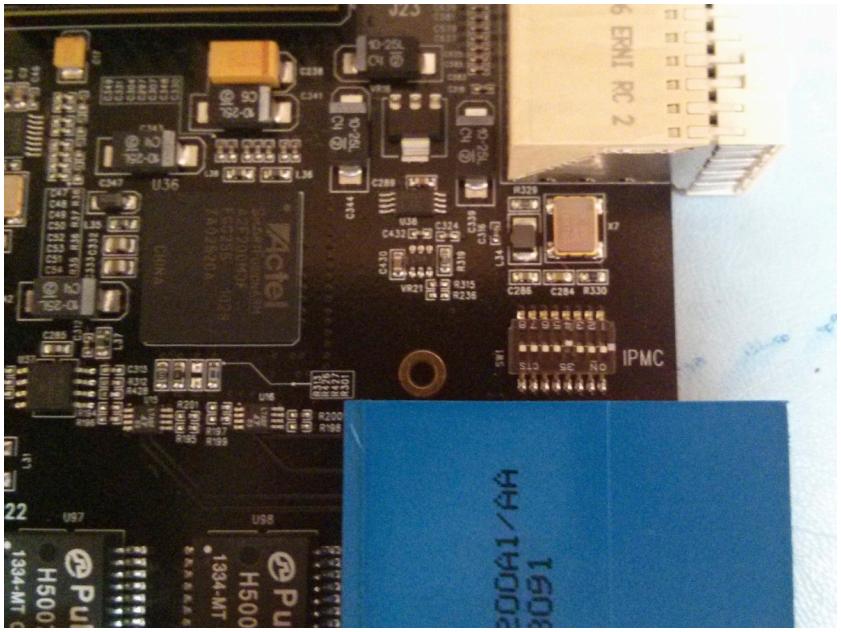
Setting the Shelf IP Info Record

- Use the `set_shelf_ip_info` command from the SDK
  - IMPORTANT!!!! After updating the Shelf IP Record you will need to reboot the Shelf Manager
    - To update the cached Shelf FRU Information
  - IMPORTANT!!! After rebooting the Shelf Manager you should send a Cold Data Reset command to all the COBs
    - To update the cached Shelf IP Record in the IPMC

```
$ set_shelf_ip_info --shelf=”shasta-sm" --vlan=1 --untagged=0 --tagged=1 \
--boundary-violations=0 --group-base=192.168.204.1 --group-end=192.168.204.254 \
--subnet-mask=255.255.255.0 --gateway=0.0.0.0 –v

=====================================================
New Shelf IP Info
-----------------------------------------------------
SLAC Shelf IP Information (shasta-sm)
-----------------------------------------------------
      VLAN ID (Valid): 001 (1)
     Discard Untagged: 0
       Discard Tagged: 1
Discard Boundary Viol: 0
           Group Base: 192.168.204.1
            Group End: 192.168.204.254
          Subnet Mask: 255.255.255.0
              Gateway: 0.0.0.0
-----------------------------------------------------
Shelf FRU Info Lock timestamp is Fri Oct  1 08:45:33 1971
Shelf FRU Info Commit timestamp is Fri Oct  1 08:45:41 1971
```

Configuring a COB to run a DHCP server

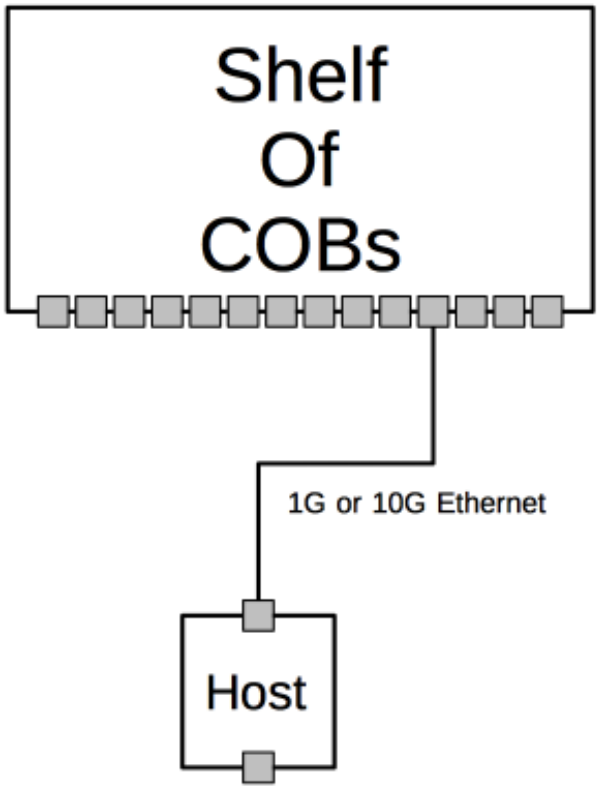


- To enable the DHCP server on one COB set the DIP switch #4 to OFF
  - DIP switch is located near the Zone 1 connector
- The Amber LED should light immediately
- The DHCP server will be started after the DTM is rebooted

COB Networking: Example Topology 1

Single External Connection:

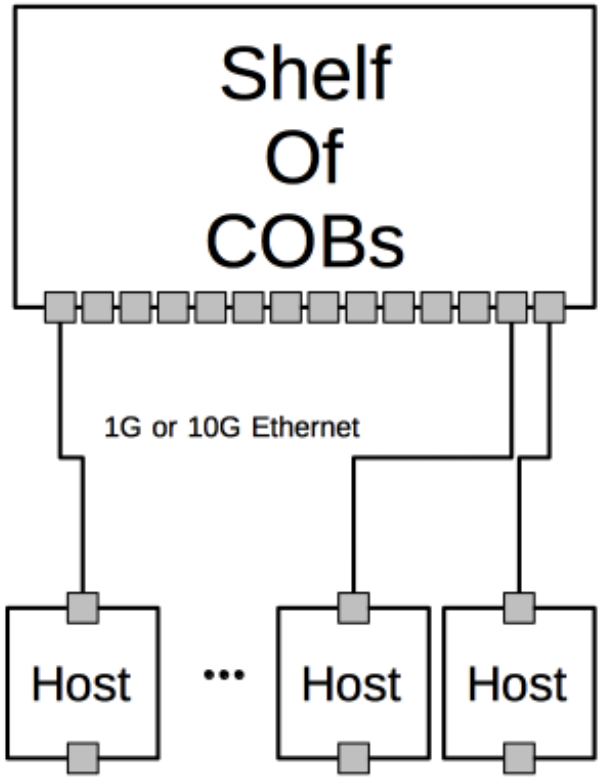
- In this configuration one SFP on one COB is connected to an external host
- This host is dual homed
  - Packets pulled from Shelf can undergo processing/steering to the outside world
  - Don't bridge the two interfaces!
    - Defeats the purpose of dual homing, may as well put the Shelf on the wider network
- Shelf network can reside on a VLAN for extra protection against unnecessary broadcast traffic and flooding



COB Networking: Example Topology 2

Multiple External Connections:

- If more bandwidth out of the shelf is required connect more COBs
- Most straightforward configuration is one external host per external connection
  - Again dual homed
  - Now more critical that the two interfaces are not bridged to avoid Layer 2 Loops if the outside interfaces reside on the same network
- A switch could be substituted for the set of hosts, but great care must be used in configuring it to avoid Layer 2 Loops
  - For now, you'll have to know what you're doing!



SDK Tools – Overview

- Low level utilities to support:
  - network connectivity tests
  - health and status reports
  - remote administration
  - software updates
- Can be used at the command line
- Meant to be integrated with higher level applications

SDK Tools – RCE Addressing

Addressing of RCEs is based on physical or ATCA location

shelfname/slot/bay/element

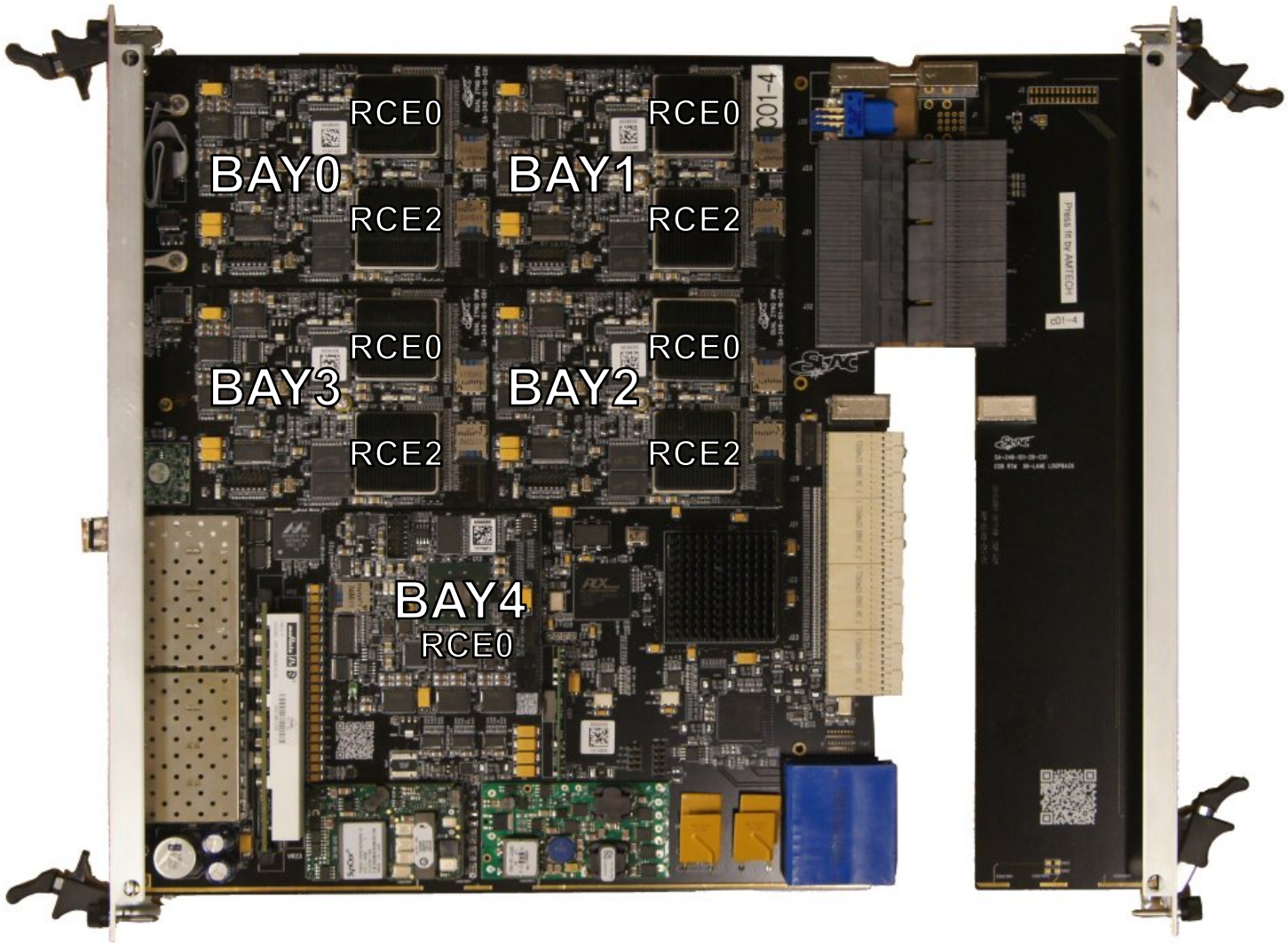
- shelfname – The shelf name (i.e. shasta)
- slot – The physical slot number (1 – 16)
- bay – The bay number (0 – 4)
- element – The specific RCE on the DPM (0, 2)

For shelf manager commands, this changes to:

shelfmanager_ip/slot/bay/element

- shelfmanager_ip – The shelf manager IP address or hostname (i.e. 172.21.6.94 or shasta-sm)

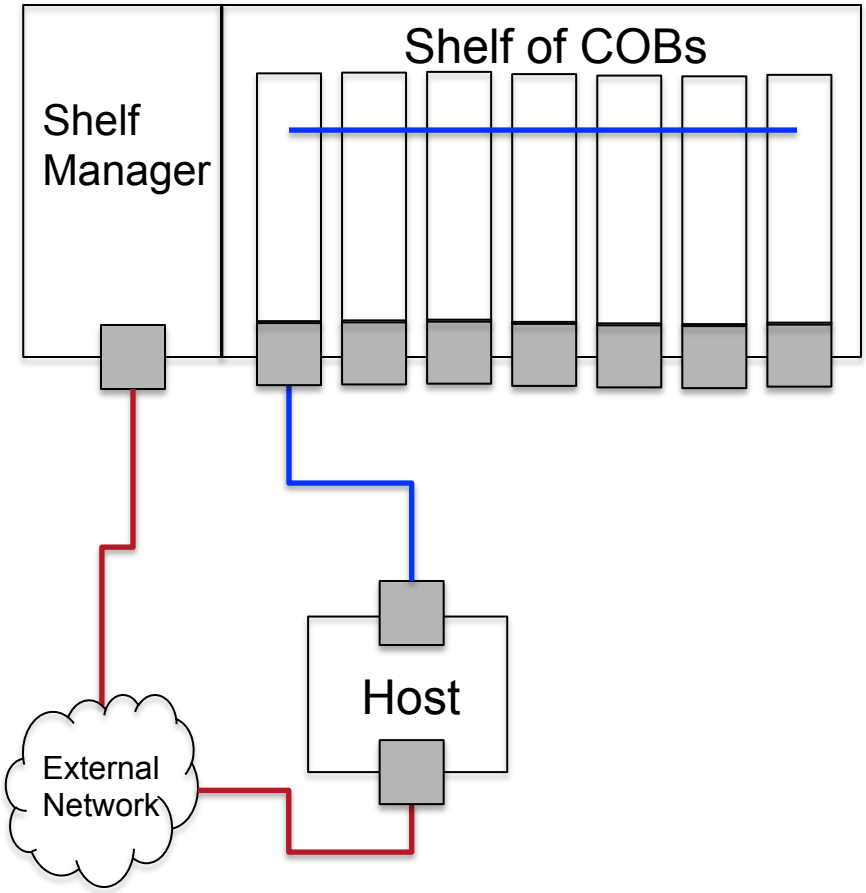
SDK Tools – RCE Addressing



RCE Addressing Topology – Single Connection

SDK Utilities:

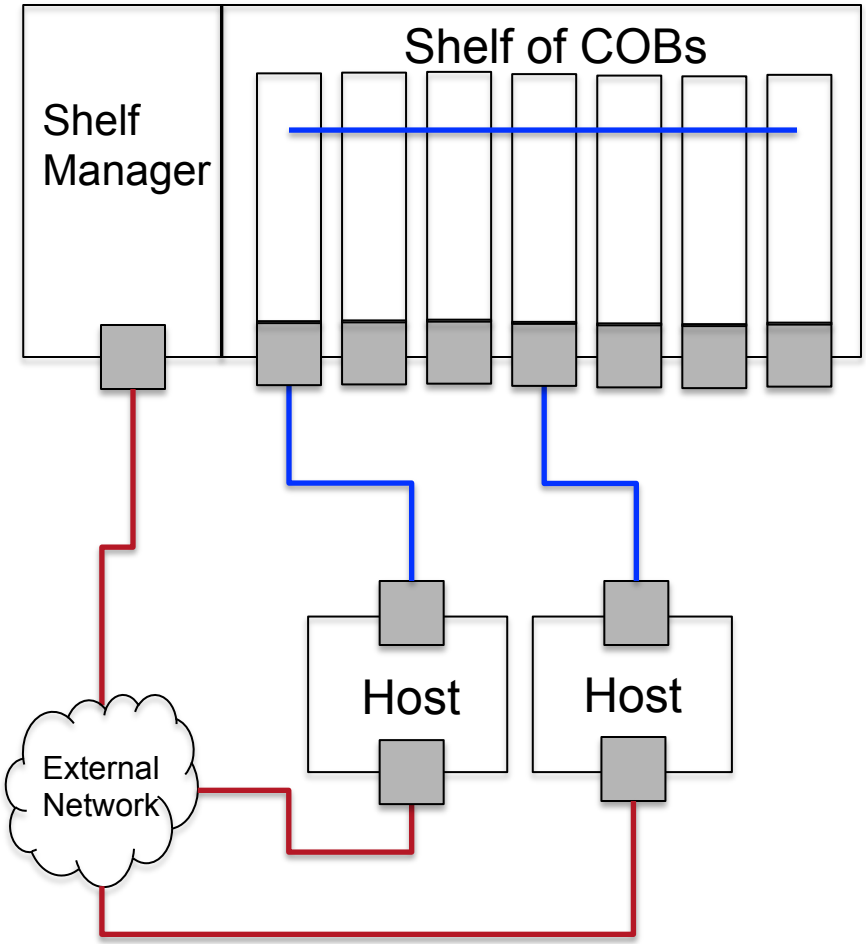
- atca
- dsl
- cob and shelf_ip_info



RCE Addressing Topology – Multiple Connections

SDK Utilities:

- atca
- dsl
- cob and shelf_ip_info



SDK Tools – RCE Network Info

Map ATCA space (shelf/slot/bay/element) to IP info

- **atca_dump**: dump the ATCA info to a terminal

```
$ atca_dump shasta/4/4/0 --ifname eth1.204
Mapped shasta/4/4/0 to this location...
  Layer 3 Port: 0 (decimal)
  Layer 3 Address: 192.168.204.254
  Layer 2 Address: 08:00:56:00:43:F2
```

- **atca_ip**: return the formatted IP address inline

```
$ ping `atca_ip shasta/4/4/0 --ifname eth1.204`
PING 192.168.204.254 (192.168.202.254) 56(84) bytes of data.
64 bytes from 192.168.204.254: icmp_seq=1 ttl=64 time=2.22 ms
```

- **atca_mac**: return the formatted MAC address inline

SDK Tools – RCE Status, Reboot, and Update

These commands interact with RCEs and use an address form that allows wild cards: shelf[/slot[/bay[/rce]]]

- **dsl_identify** address [--ifname ifname]
  Determine IP address and software version on RCEs.
- **dsl_reboot** address [--ifname ifname] [-t os_type] [-b bitfile load]
  Remotely reboot one or more RCEs, optionally to another OS (linux|rtems|rescue), optionally with or without loading the default bitfile.
- **dsl_update** address [--ifname ifname] [--ip update_server_ip]
  Remotely execute software update on one or more RCEs.
  - nfs mount remote directory, execute shell script, pull files to RCE

SDK Tools – COB Utilities

These commands interact with the shelf manager and COB IPMC

- **cob_dump** shelfmanager_ip[/slot[/bay[/rce]]]
  Print voltage, current, power, and temperatures.

- **cob_cold_data_reset** shelfmanager_ip[/slot[/bay[/rce]]]
  Send a reset to an RCE and re-read the cold data store.
  Equivalent to removing the board from the shelf.

- **cob_rce_reset** shelfmanager_ip[/slot[/bay[/rce]]]
  Send a soft reset to one or more RCEs.

SDK Tools - Administration

Creating an RCE SD Card

https://confluence.slac.stanford.edu/display/RPTUSER/Creating+an+RCE+SD+Card

Upgrading RCE Core Software

https://confluence.slac.stanford.edu/display/RPTUSER/Upgrading+RCE+Core+Software

Updating RCE User Software

- Secure copy (scp)
- Can make use of SDK update utility
- User responsible for Arch Linux package updates

Diskless RCE Node Using NFS

Host Configuration:

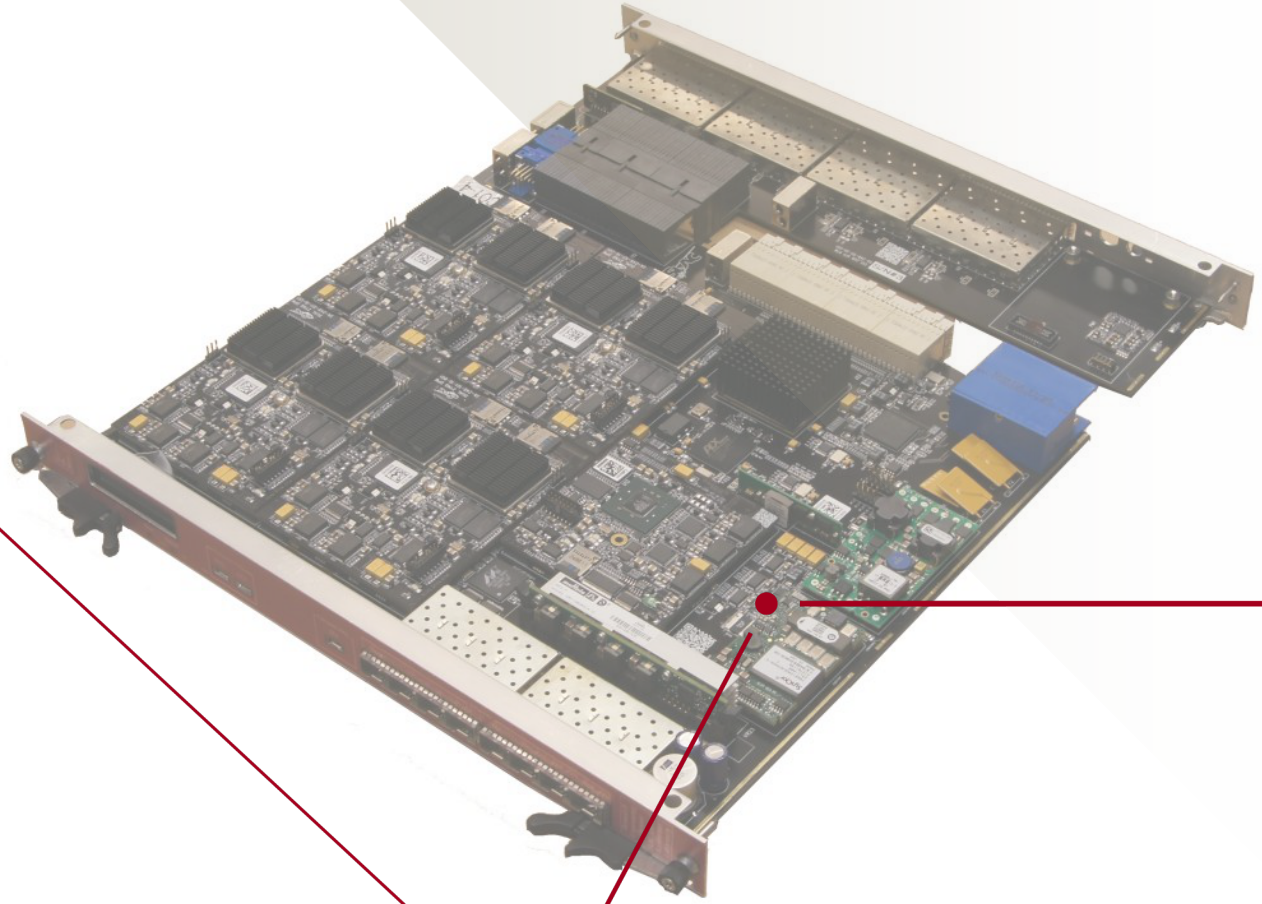
- NFS server setup
- DHCP setup
  - DHCP server can push configuration to bootloader
- Prep Arch Linux for NFS root filesystem
  - Read/write permissions to host disk
  - SSH keys

RCE Bootloader Configuration:

- RCE and NFS server IP addresses
- Linux kernel, device tree, and firmware bitfile
- NFS rootfs path

https://confluence.slac.stanford.edu/display/RPTUSER/RCE+Diskless+Node+(NFS)

Backup Slides



Network Topology: Development vs. Production

- The development environment is different than the production environment
  - It is very reasonable for the networks to be fundamentally different
- The development environment is dynamic
  - The production system might include many COBs in many Shelves
    - Development probably starts with a single RCE
    - As development continues RCEs and COBs may be added
- The production environment is stable
  - The network in the two cases can and probably should be different
- In production, there may be restrictions on how the you are allowed to configure the network
  - In Development there is usually more freedom to do whatever you need
- Example: DHCP
  - In a developments systems, COBs may be moving in and out of the Shelf
    - Might prefer the DHCP server outside the Shelf (The host machine may be more stable)
  - In production, COBs should be stable
    - Having DHCP inside the Shelf scales better than DHCP outside the Shelf
- Example: Should a single external network connection be 1G or 10G Ethernet?
  - It's not all about bandwidth!
  - 1G is cheaper and easier to find external networking hardware for
  - 10G doesn't require the Cluster Interconnect to perform rate matching between the internal Shelf network and the External connection