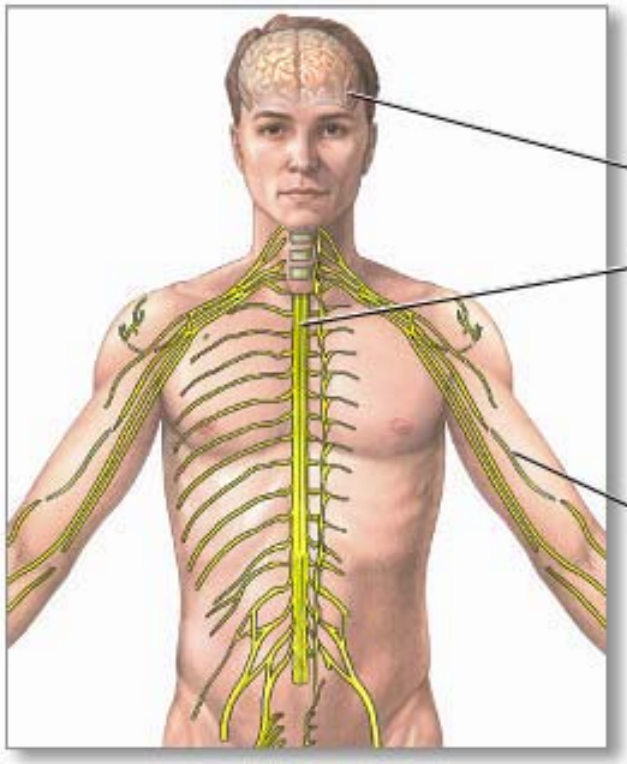
Red Storm Management System (RSMS)

The Central Nervous System of the Cray XT3 and RedStorm Computer Systems

Steve Sjoquist

Mark Swan

CUG 2005

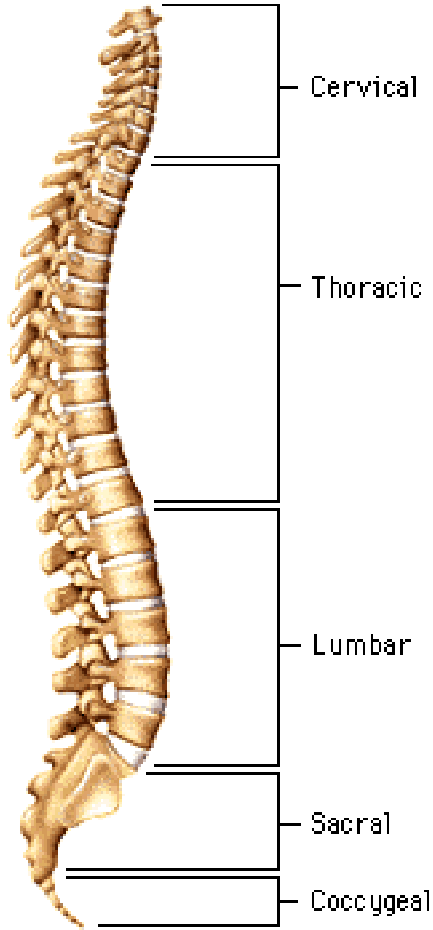

Central nervous system
Brain
Spinal cord
Peripheral nervous system
Peripheral nerve
A.D.A.M.

What's in a name?

- The entire system is called 's0'
- Cabinets are named 'cX-Y'
- Cages are named 'cX-YcC'
- Modules are named 'cX-YcCsS'
- Nodes are named 'cX-YcCsSnN'
- SeaStars are named 'cX-YcCsSsS'
- Sections are named 'tA-B'

Event-based System

- Indirect request/response protocol
- Must subscribe to an event to get it
- Event sizes vary greatly
- One request can cause a single response or thousands of responses
- Unsolicited events

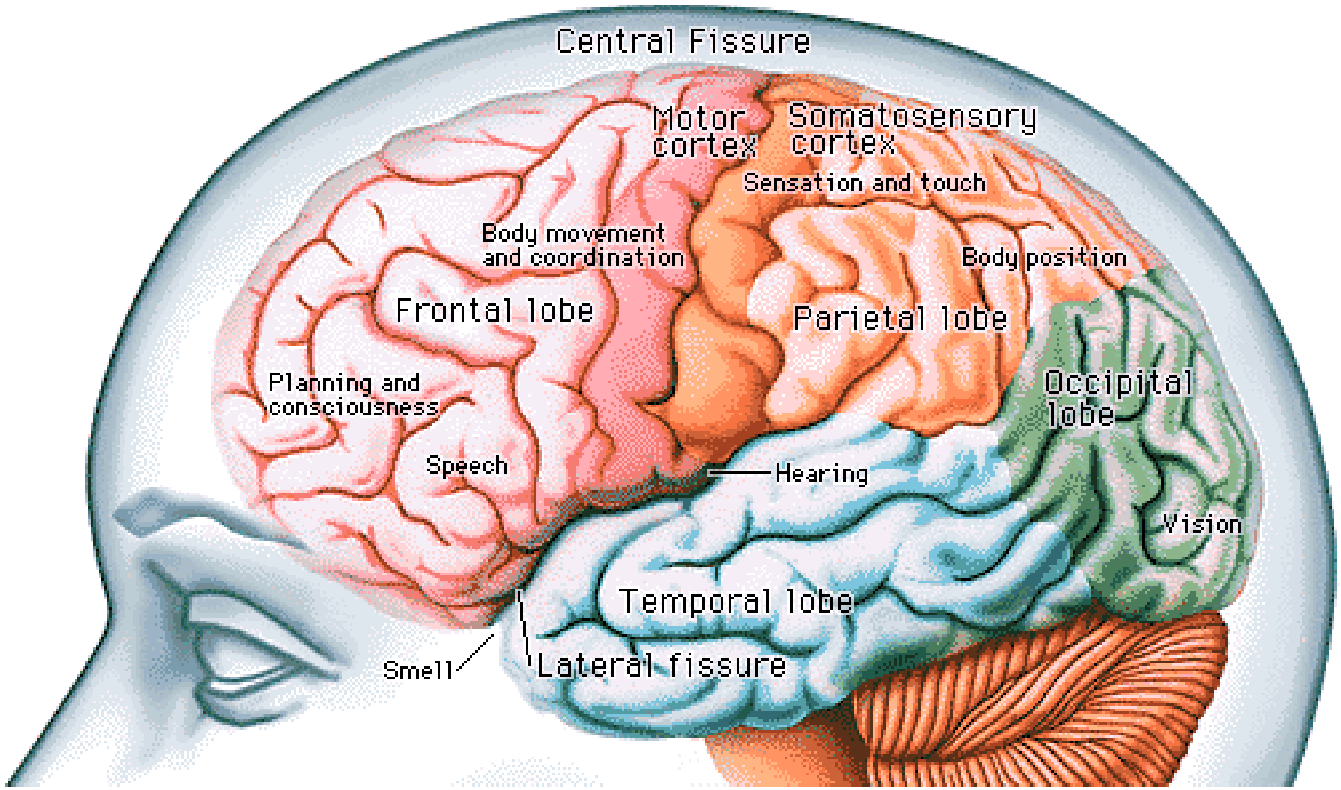


Event Router

- Manages hierarchical tree of connections to other Event Routers on L1s and L0s
- Manages connections to Managers, Special Agents and User Interfaces
- Delivers events
- Honors subscriptions
- Logs all events to /opt/craylog/eventlog

Event Router (continued)

- Manages localization and globalization of events
  - Some events go everywhere
  - Some events go all the way to the SMW
  - Some events only go from L0 to L1
  - Some events only go from L1 to SMW
  - Some events don't go anywhere
- Acts somewhat like a "ethernet router" when determining if an event *needs* to go down a path. These are called targeted events.



Managers

- State Manager
- Power Manager
- Diagnostic Manager
- Boot Manager
- Routing Manager
- Warning Manager
- Flash Manager

State Manager

- Maintains state of all components in the system
- Maintains locks to prevent overlapping operations
- Provides state change notifications to daemons on the mainframe
- Use '*xtcli status …'* command to see state information

Power Manager

- Manages sequencing of component power down and power up
- Use '*xtcli power …*' command

Diagnostic Manager

- Manages the running of *cpuburn*, *memtest* and *seacheck* diagnostics
- Logs summary information in /opt/craylog/diaglogs subdirectories
- Use '*xtcli diag …'* command

Boot Manager

- Moves operating system data into boot node memory
- Coordinates with *boot node daemon* for movement of operating system data into all other node memories across high speed network
- Use '*xtcli boot …*' command

Routing Manager

- Computes and asserts routing tables
- Initializes links
- Routes around disabled components when possible
- Use '*rtr …*' command

Warning Manager

- Displays information about components that require attention due to environmental conditions
- Use '*wm …*' command

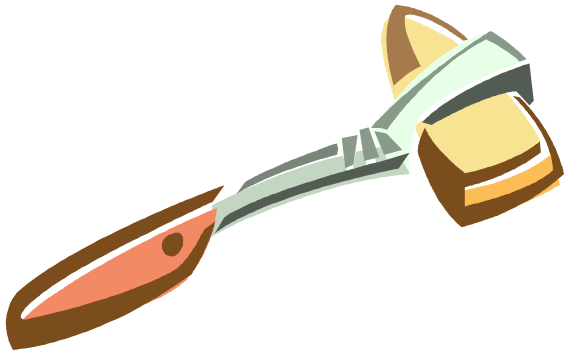
Flash Manager

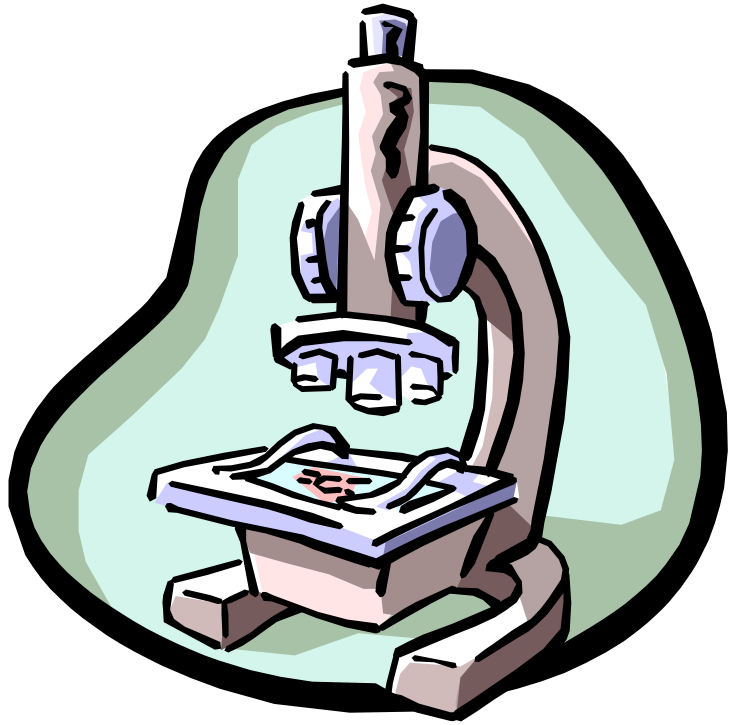
- Flashes L1s and L0s with new images stored in /opt/tftpboot
- Can wipe out L1 and L0 flash memory to cause a network boot
- Use '*fm …*' command

PETROGLYPHS TO PETAFLOPS
CUG 2005

Billy, how are you feeling today?

- Running on the L1 is a system daemon that monitors and reports cabinet air flow pressure, air temperature and heartbeats from L0s.
- Running on the L0 is a system daemon that monitors and reports component voltages, temperatures and other health indications as well as heartbeats from the SeaStar and Opterons.

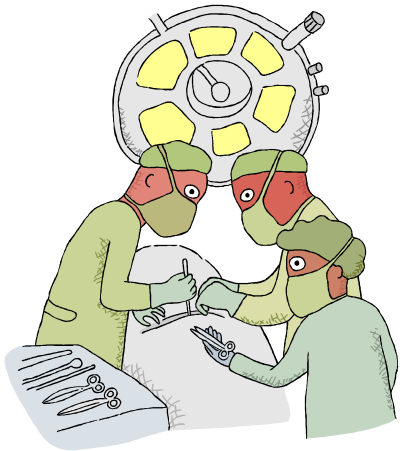


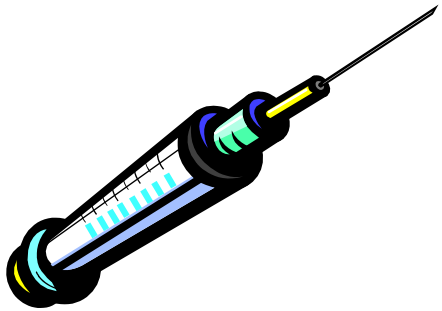



PETROGLYPHS TO PETAFLOPS
CUG 2005

Hey Jim, have you ever seen anything like this before?

- “xtconsumer” will show you all (or just a selected few) of the events that are flying around the system (when watching the SMW) or just on a specific component (when watching a L1 or L0).
- “wm” will show you any health problems
- “xthb” is an intrusive, active SeaStar and Opteron heartbeat checker.



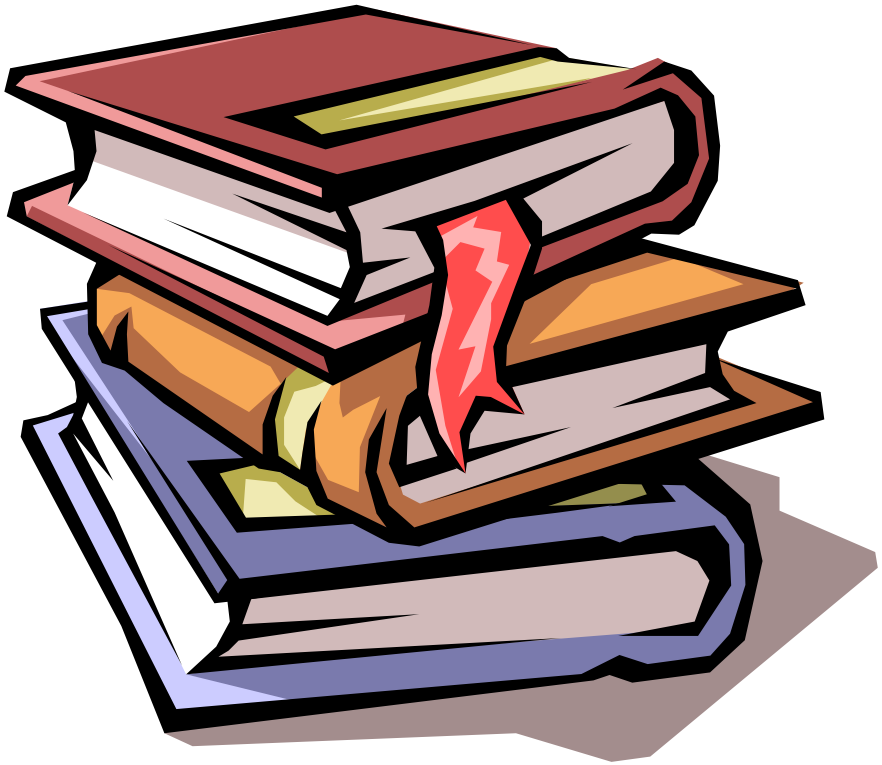



PETROGLYPHS TO PETAFLOPS
CUG 2005

This is going to hurt me more than it is going to hurt you.

- We can disable nodes that are causing problems so jobs don't get scheduled on them.
- We can route around interconnect faults.
- In an upcoming release, we'll be able to remove, repair and replace faulty modules.

Where to learn more…



- http://docs.cray.com