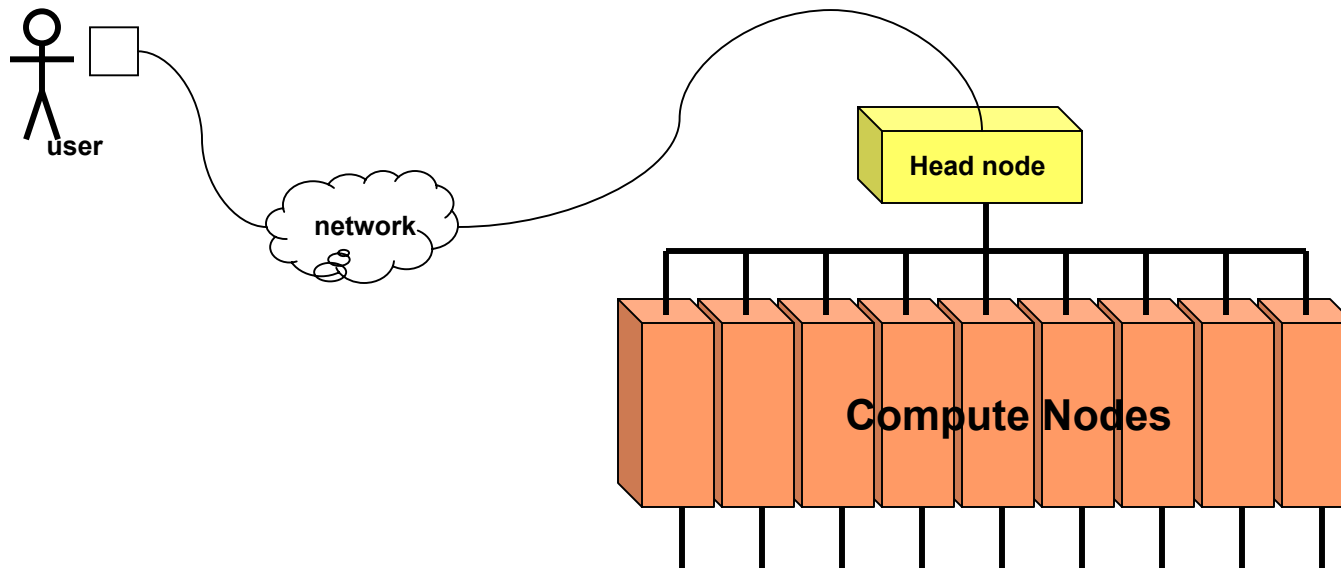


A typical Linux cluster consists of a group of compute nodes for executing parallel jobs and a head node to which users connect to build and launch their jobs.

Often the compute nodes are connected to the head node by one network and to the other compute nodes by a second network. The second network is often used solely for communication between parallel program tasks and in some implementations may be a special low-latency, high-bandwidth network such as Myrinet.

Logically a Linux cluster looks something like the following:



NC State has implemented a Linux cluster using IBM Blade Center hardware.

The following pages describe the Blade Center in more detail focusing on the features of the Blade Center architecture that may impact the performance of parallel applications running on the university Linux cluster.

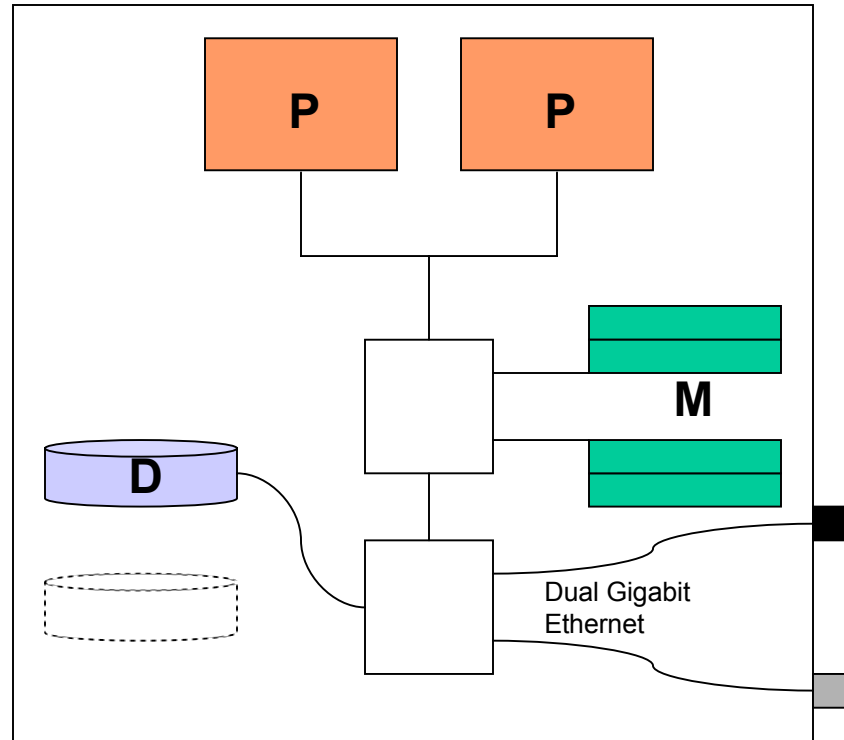
BladeCenter HS20



IBM server BladeCenter

Blade Center has two major components: blades, which are dual-processor computers built from server grade parts; and chassis which provide power, cooling, management, and connectivity for fourteen blades.

To the right is a cartoon of the major parts which make up a typical Blade Center blade.

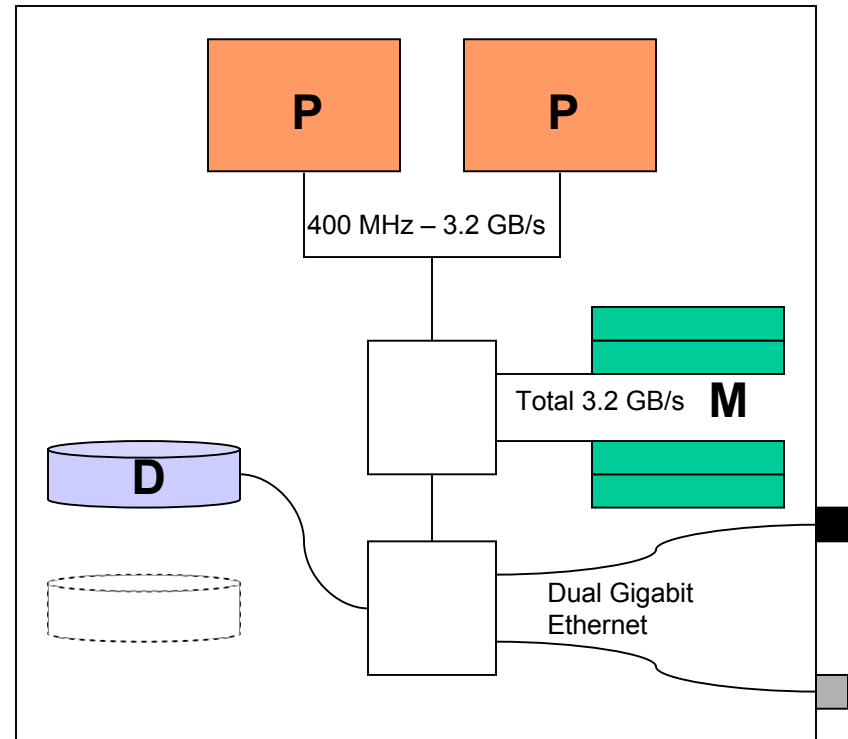


Blade

- Dual Processors
- 4 Memory Slots
- 2 Disk Drive Bays
- Dual Gigabit Ethernet
- All power, cooling, and connectivity supplied by chassis*

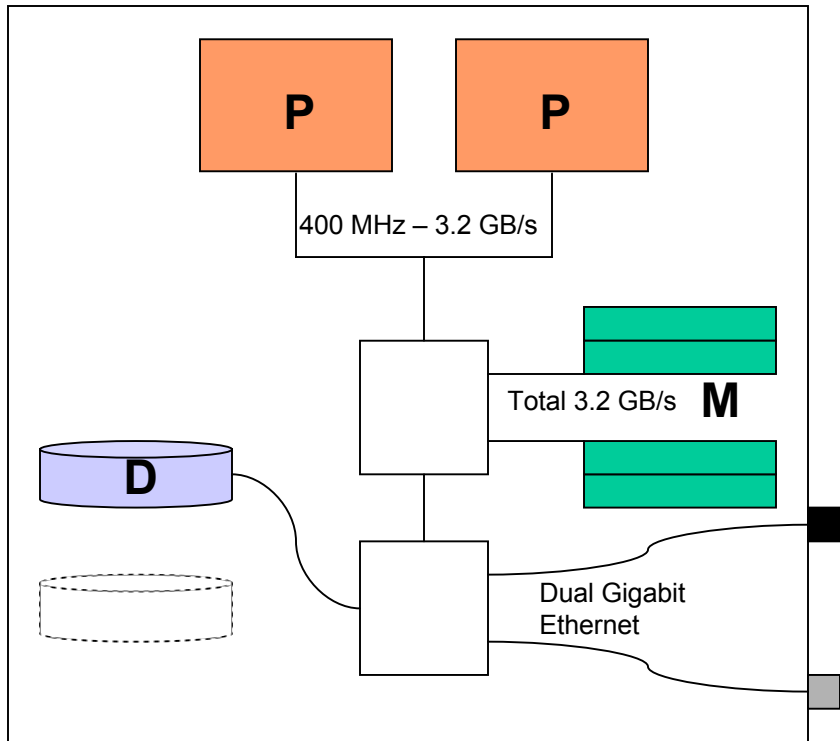
NC State began purchasing blades more than a year ago. As with other computer systems, processor speeds on blades have increased steadily during the year. To date the university Linux cluster contains blades with Xeon processors running at three different clock speeds.

Two metrics are shown for the various blades in the university cluster: the peak speed, taken to be the peak rate that the chip can perform floating point operations (expressed as Giga Floating Point Operations per Second – GFLOPS); and the peak memory bandwidth divided by the peak speed (expressed as Bytes/floating point operation).



2.4 GHz Dual Xeon Blade

- 3 GB memory
- Dual 40 GB disk
- Peak Speed – 4.8 GFLOPS per processor
- Peak Bytes/FLOP – 0.33



2.8 GHz Dual Xeon Blade

- 4 GB memory
- 40 GB disk
- Peak Speed – 5.6 GFLOPS per processor
- Peak Bytes/FLOP – 0.28

The initial NC State high performance computing (HPC) cluster had 64 of these 2.8 GHz dual-Xeon blades.

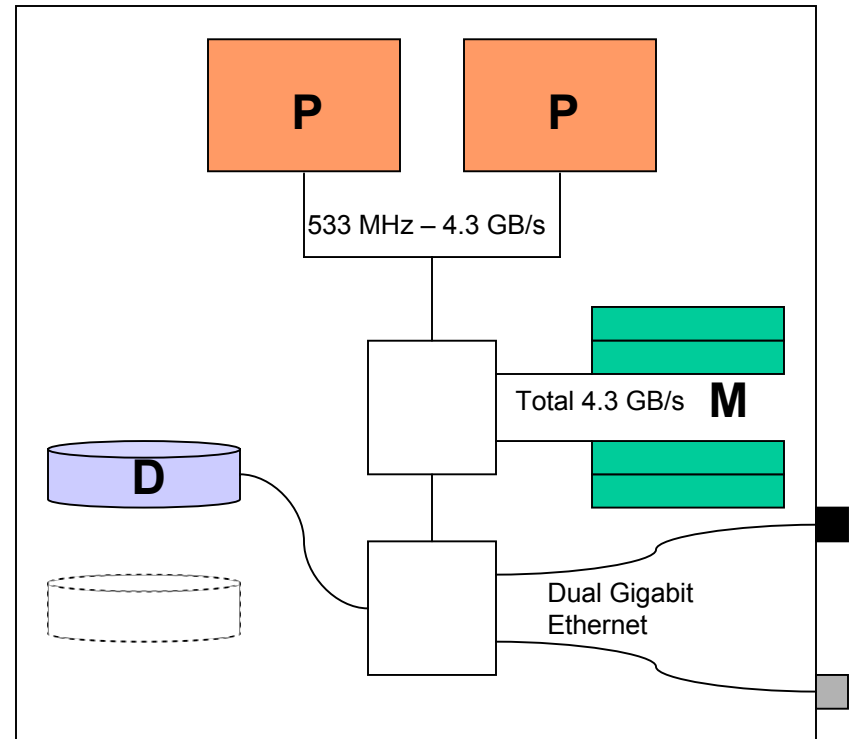
Two of the 2.4 GHz blades are available for code development and debugging.

However, note that Bytes/FLOP (B/F) for the 2.4 GHz blades is better (0.33) than for the 2.8 GHz blades (0.28).

Compute intensive applications may need B/F=8 to run unimpeded by memory bandwidth limitations.

By June 2004 the university Linux cluster included 32 of the 3.06 GHz dual-Xeon blades.

Note that the system bus speed increased on these blades from 400 to 533 MHz. The resulting memory bandwidth improvement produced an improved B/F (0.35).



3.06 GHz Dual Xeon Blade

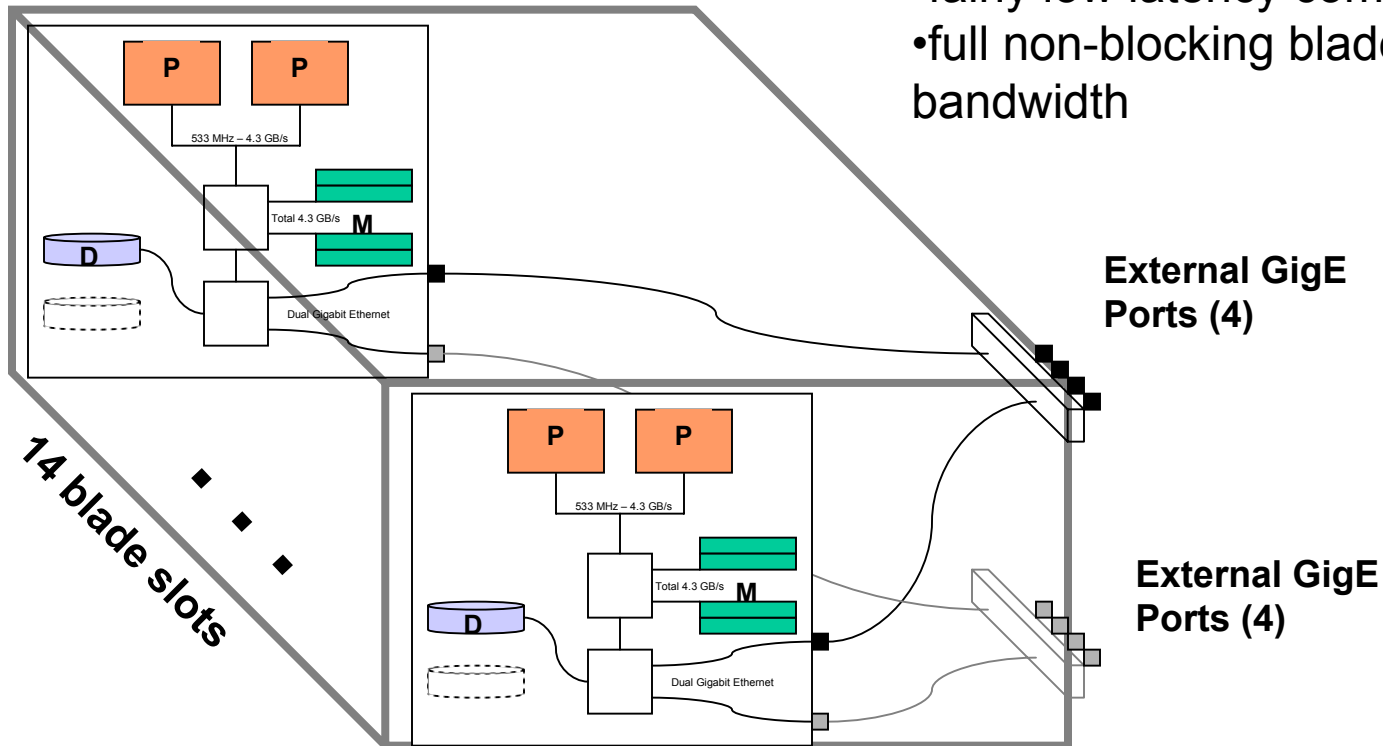
- 4 GB memory
- 40 GB disk
- Peak Speed – 6.12 GFLOPS per processor
- Peak Bytes/FLOP – 0.35

Chassis

- Redundant Power supplies
- Redundant Blowers
- Management module
- Dual GigE switches
- Space for 14 blades

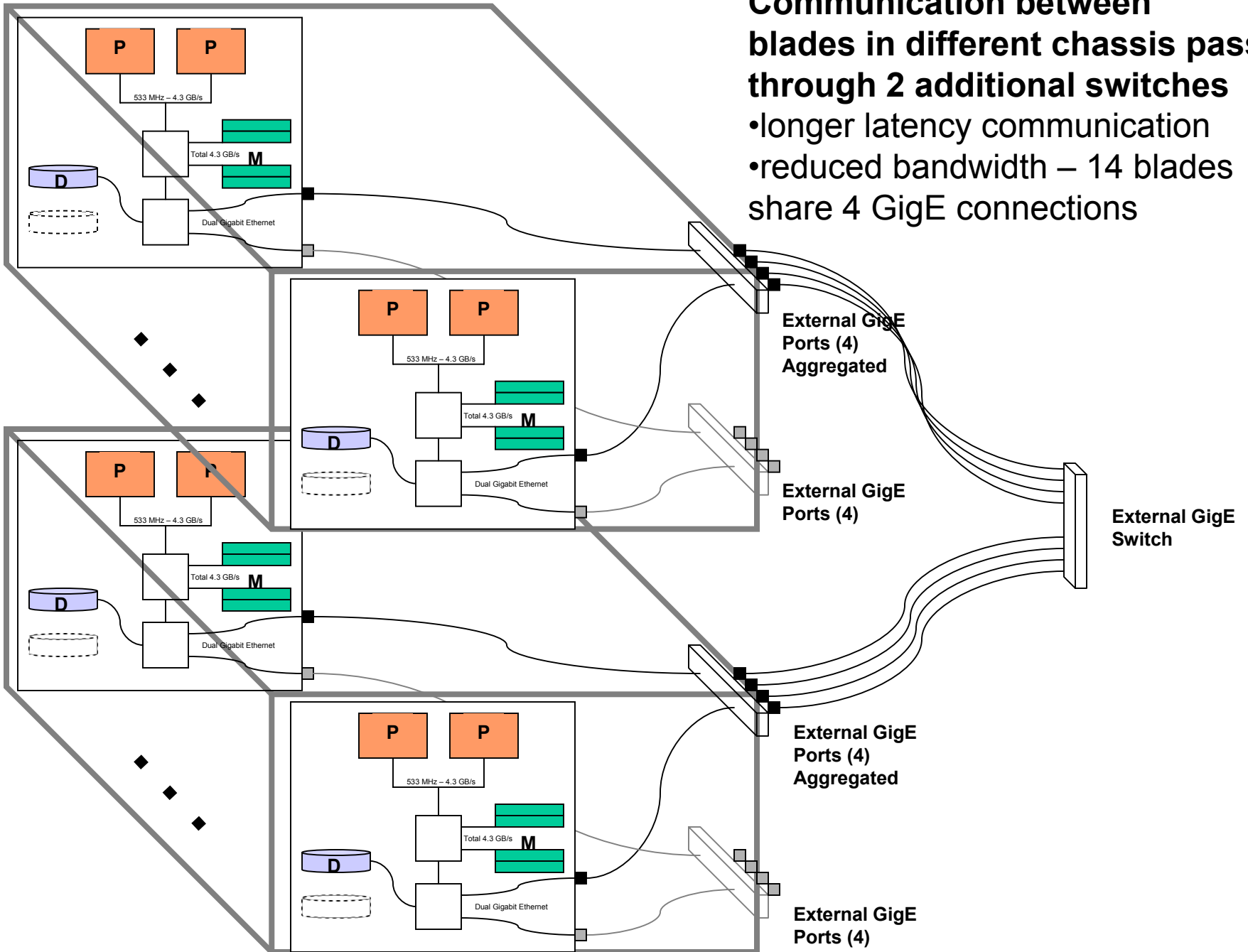
Communication between blades in same chassis uses internal GigE switch

- fairly low latency communication
- full non-blocking blade to blade bandwidth

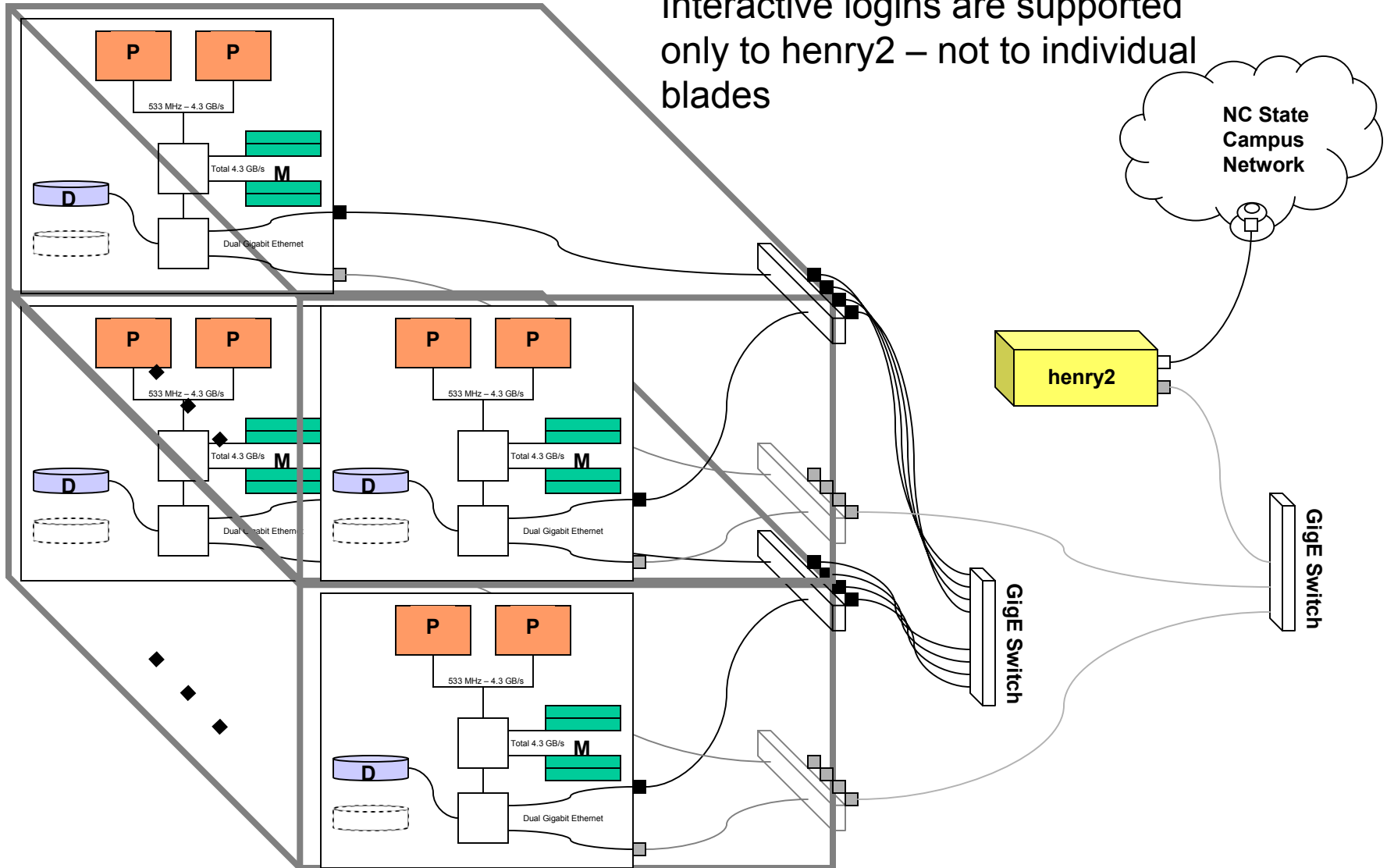


Communication between blades in different chassis pass through 2 additional switches

- longer latency communication
- reduced bandwidth – 14 blades share 4 GigE connections



User access to blades is through a queuing system (LSF) controlled by a Linux head node (henry2)
Interactive logins are supported only to henry2 – not to individual blades



Storage available from cluster

• Shared network attached storage available from blades

- /home
- /usr/local (read only)
- /share
- /share3

• Network attached storage available from henry2

- /ncsu/volume1
- /ncsu/volume2

