

Static vs. Switched Networks

- Static networks: direct connections between compute nodes
- Switched networks: switch is used between compute nodes

Shopping for a 96-port switch



Cisco Nexus 93128TX Layer 3 Switch

by [Cisco](#)

[Be the first to review this item](#)

List Price: \$26,000.00

Price: **\$23,941.08** + \$99.90 shipping

You Save: **\$2,058.92 (8%)**

Note: Not eligible for Amazon Prime.

Usually ships within 6 to 10 days.

Ships from and sold by [Get724](#).

Estimated Delivery Date: Feb. 17 - 25 when you choose Expedited at checkout.

- Device Type: Layer 3 Switch
- Form Factor: Rack-mountable

What is inside a switch?

Crossbar and Multistage Connections

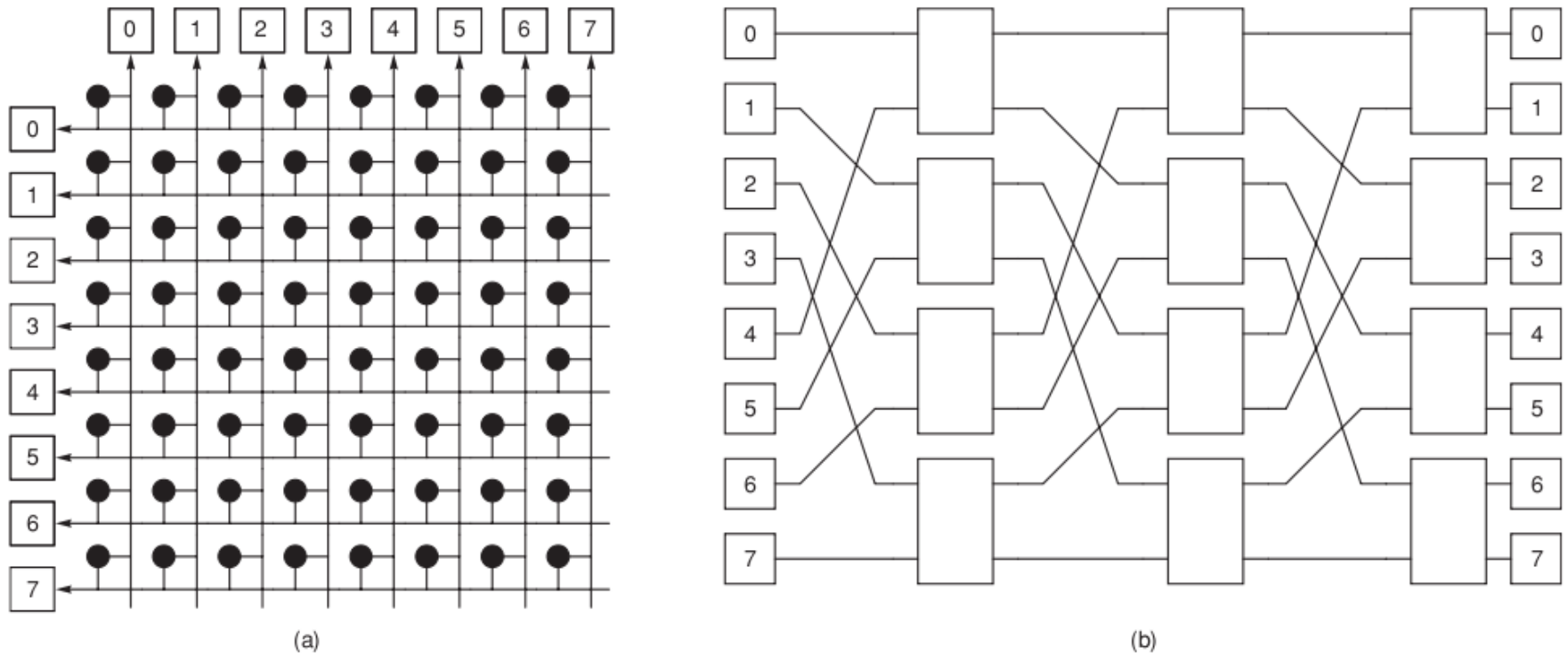


Figure E.11 Popular centralized switched networks: (a) the crossbar network requires N^2 crosspoint switches, shown as black dots; (b) the Omega, a MIN, requires $N/2 \log_2 N$ switches, shown as vertical rectangles. End node devices are shown as numbered squares (total of eight). Links are unidirectional—data enter at the left and exit out the top or right.

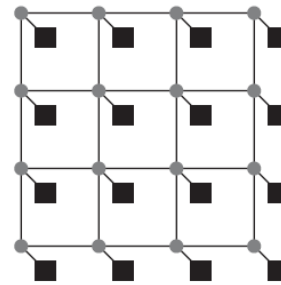
How would you build a cluster computer?

- Allowed components:
 - network cards
 - switches
 - cables
- Exercise: Propose how to connect 64 nodes with any resources, but only 16-way switches are available

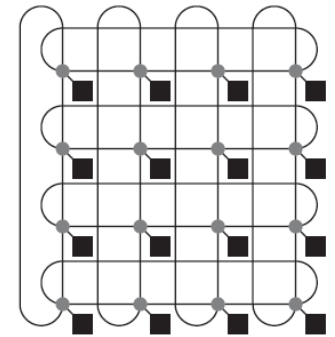


Some Network Topologies

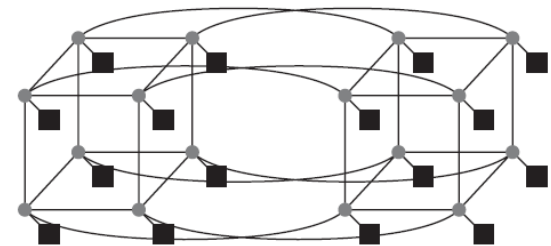
- Trade-off between cost and performance
- Match the topology with the application
- Ring
- Mesh/Torus
 - Good for applications using nearest-neighbor communication
 - Prevalent for proprietary interconnects
- Hypercube
- Fat tree (uses switches)
 - Popular for commodity clusters
- Dragonfly (switch attached to node)
 - Low diameter network



(a) 2D grid or mesh of 16 nodes



(b) 2D torus of 16 nodes

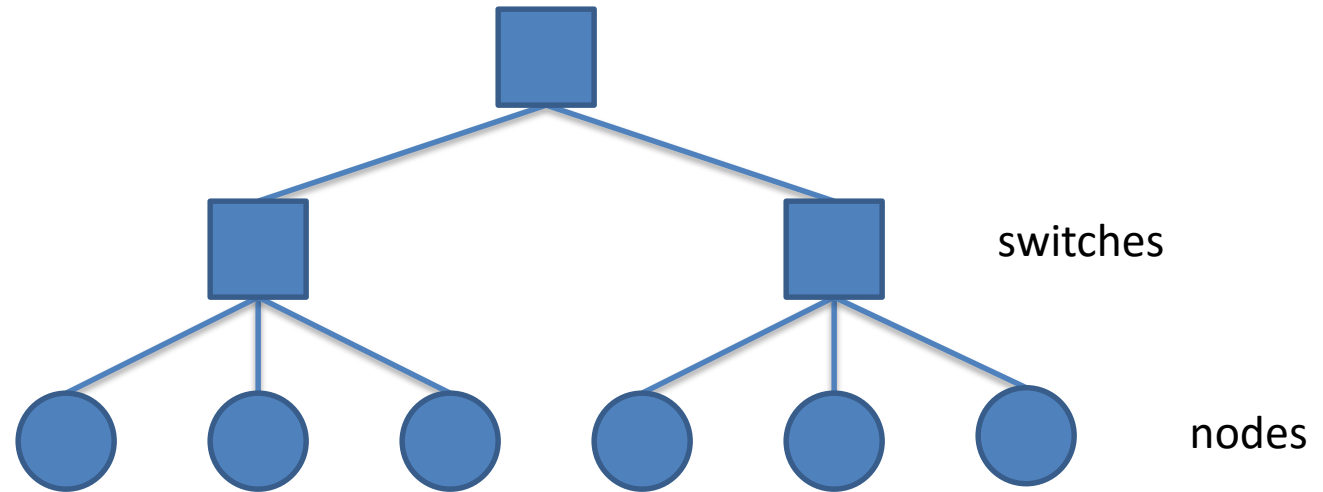


(c) Hypercube of 16 nodes ($16 = 2^4$ so $n = 4$)

Fat Tree



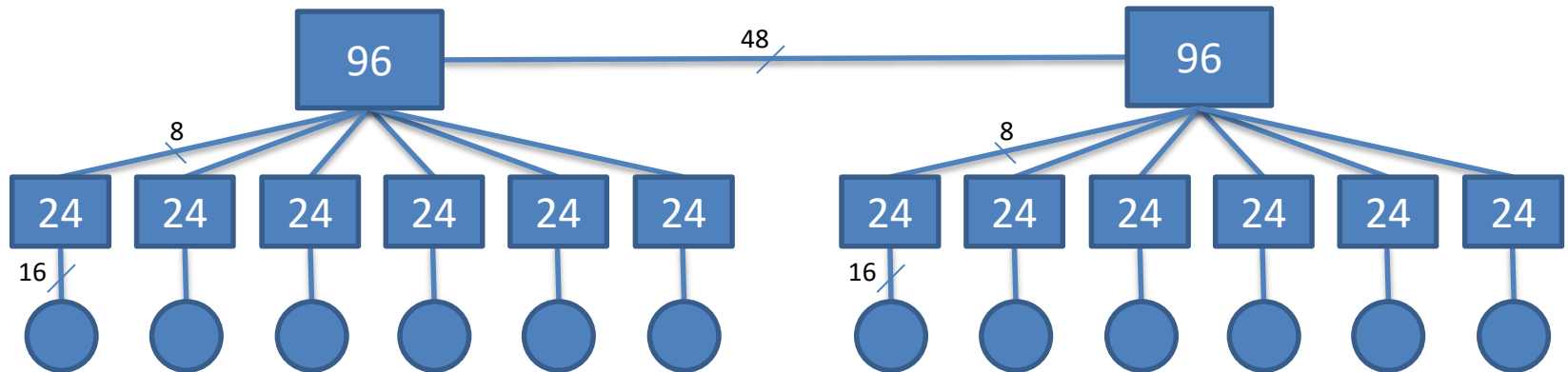
Trees and Fat Trees



- Tree (above figure)
- Fat Tree: additional links to increase bandwidth near the root of the tree

Fat Tree Challenge

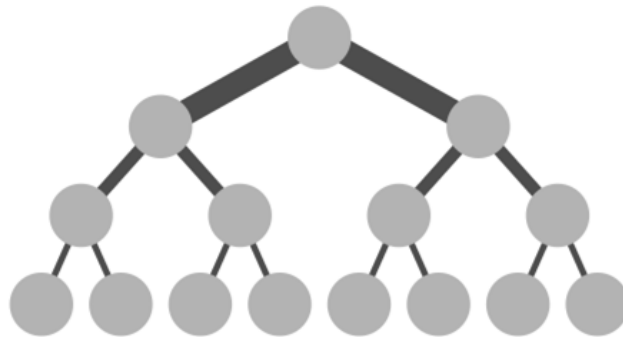
- Build a 192 node fat tree cluster with two 96-way switches and any number of 24-way switches.



Possible solution

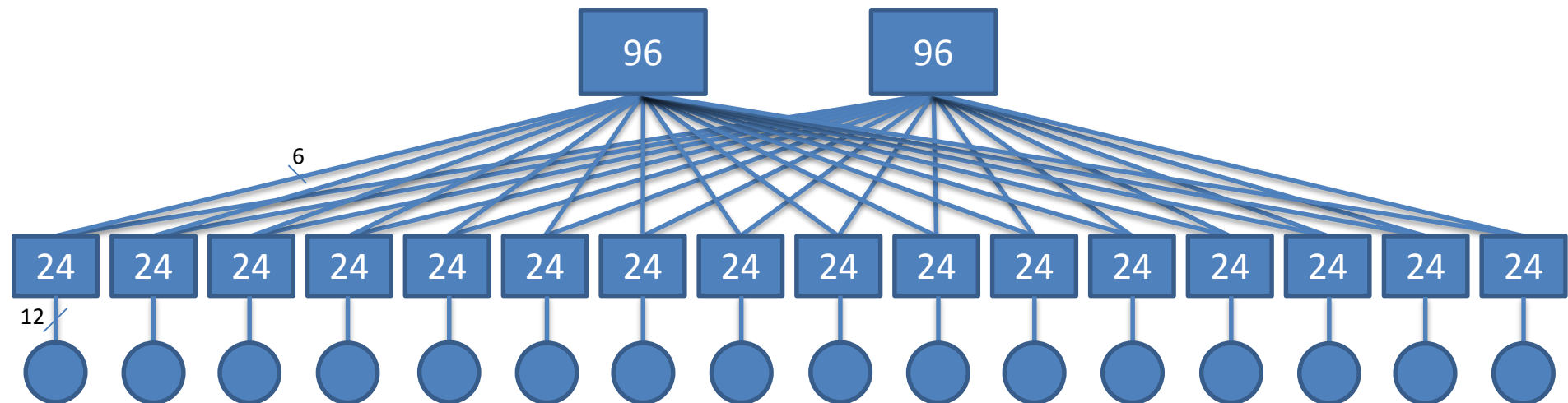
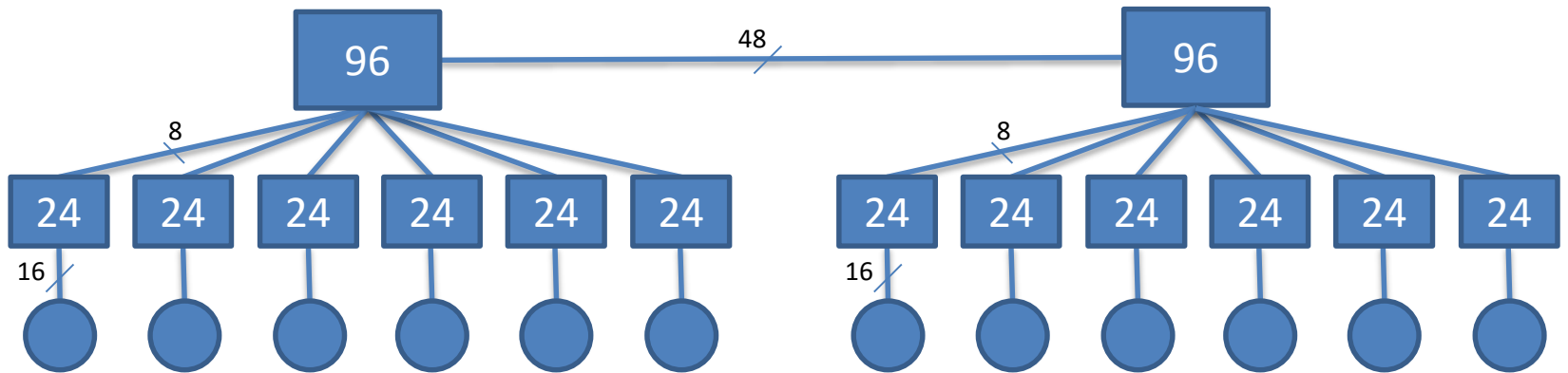
Network Properties

- Diameter = max number of hops between two nodes
- Bisection width = minimum number of links cut to divide the network into two halves
- Bisection bandwidth = bandwidth of above
- Full bisection bandwidth = one half of nodes can communicate simultaneously with other half of nodes



Bisection Bandwidth

- Suppose half the nodes send to the other half of the nodes (each node sends or receives). What is the minimum bandwidth for any choice of which nodes send or receive?
- **Bisection bandwidth**: sum of the bandwidths of the minimal number of links that are cut when splitting the system into two parts
- Suppose that half the nodes can inject data into the network at a rate of B bytes/sec. The network has **full** bisection bandwidth if the bisection bandwidth is B .



What is the bisection bandwidth in these two examples?

Network Properties

Evaluation category	Bus	Ring	2D mesh	2D torus	Hypercube	Fat tree	Fully connected
Performance							
BW _{Bisection} in # links	1	2	8	16	32	32	1024
Max (ave.) hop count	1 (1)	32 (16)	14 (7)	8 (4)	6 (3)	11 (9)	1 (1)
Cost							
I/O ports per switch	NA	3	5	5	7	4	64
Number of switches	NA	64	64	64	64	192	64
Number of net. links	1	64	112	128	192	320	2016
Total number of links	1	128	176	192	256	384	2080

Figure E.15 Performance and cost of several network topologies for 64 nodes. The bus is the standard reference at unit network link cost and bisection bandwidth. Values are given in terms of bidirectional links and ports. Hop count includes a switch and its output link, but not the injection link at end nodes. Except for the bus, values are given for the number of network links and total number of links, including injection/reception links between end node devices and the network.

One port per node; nodes attached to switches.
Hennessy and Patterson, 2007.

Summary of Interconnects

- Shared media (did not talk about this)
 - Used in small shared memory machines
 - e.g., bus shared by up to 4 cores
- Centralized switched networks (indirect/dynamic)
 - Used in distributed memory and shared memory machines
 - Network fabric consists of switches and links
 - e.g., crossbar, multistage network, fat tree
- Distributed networks (direct/static)
 - Used in distributed memory machines
 - Usually, a switch is attached to the node
 - e.g., mesh/grid/array, torus, hypercube