Sebastian Krey and Freja Nordsiek

Exercise Interconnects

Contents

Task 1: Cable planning (5 min) 1

Task 2: Switch planning (5 min) 1

Task 3: Storage server placement (3 min) 1

Task 4: Communication overhead calculation (5 min) 2

Optional Task 5: Fabric topolgies (10 min) 2

Task 1: Cable planning (5 min)

How should cable management be done, to ensure:

1. All nodes have all necessary connections (Ethernet, IPMI, Infiniband, power)
2. The airflow behind and in front of the nodes is as best as possible
3. The cables are as short as possible (Infiniband cables are expensive)
4. Replacing broken cables is as easy as possible.

Hints

- Consider placing of switches at different positions
- Consider front serviced nodes (network front, power back) vs. back serviced nodes (everything in the back)

Task 2: Switch planning (5 min)

How many 36 port switches are needed to connect 120 nodes to a fabric? Calculate this for a non-blocking fabric as well as for a blocking factor of 2 at the leaf switch level.

Task 3: Storage server placement (3 min)

How should the storage servers connected to the fabric for best performance, which requires even load distribution?

Task 4: Communication overhead calculation (5 min)

You have a 200GBit/s Infiniband HDR network available. How many MPI messages of size 8 Byte could be transfered theoretically?

How many messages per second are left over in different fabrics:

1. Infiniband with a message header of size 20 Byte.
2. RoCEv2 with a message header of size 66 Byte.

Optional Task 5: Fabric topolgies (10 min)

This is a difficult **additional** task which will support your understanding in the topic.

Why does even a non-blocking Fat Tree not guarantee 100% bandwidth for every communication pattern? Create a minimal example.

How can this problem be avoided?