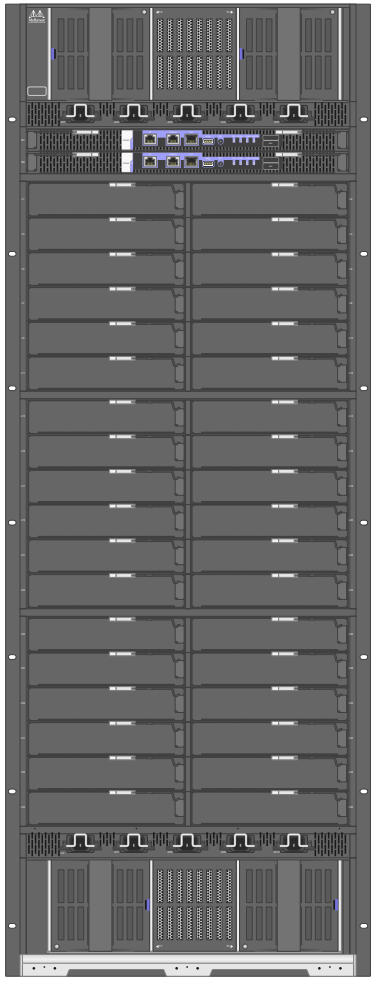
Common InfiniBand Diagrams and Implications *v5*

A Visual HOWTO

Prepared by:
Brian Finley, IBM
Brian Forbes, Mellanox



Where Can I Connect my Nodes?

A general rule of thumb with InfiniBand fabric architectures is to have all nodes at the same level (e.g.: all on a tier 2 leaf switch). While that will always work, there are other perfectly valid architectures that you can use.

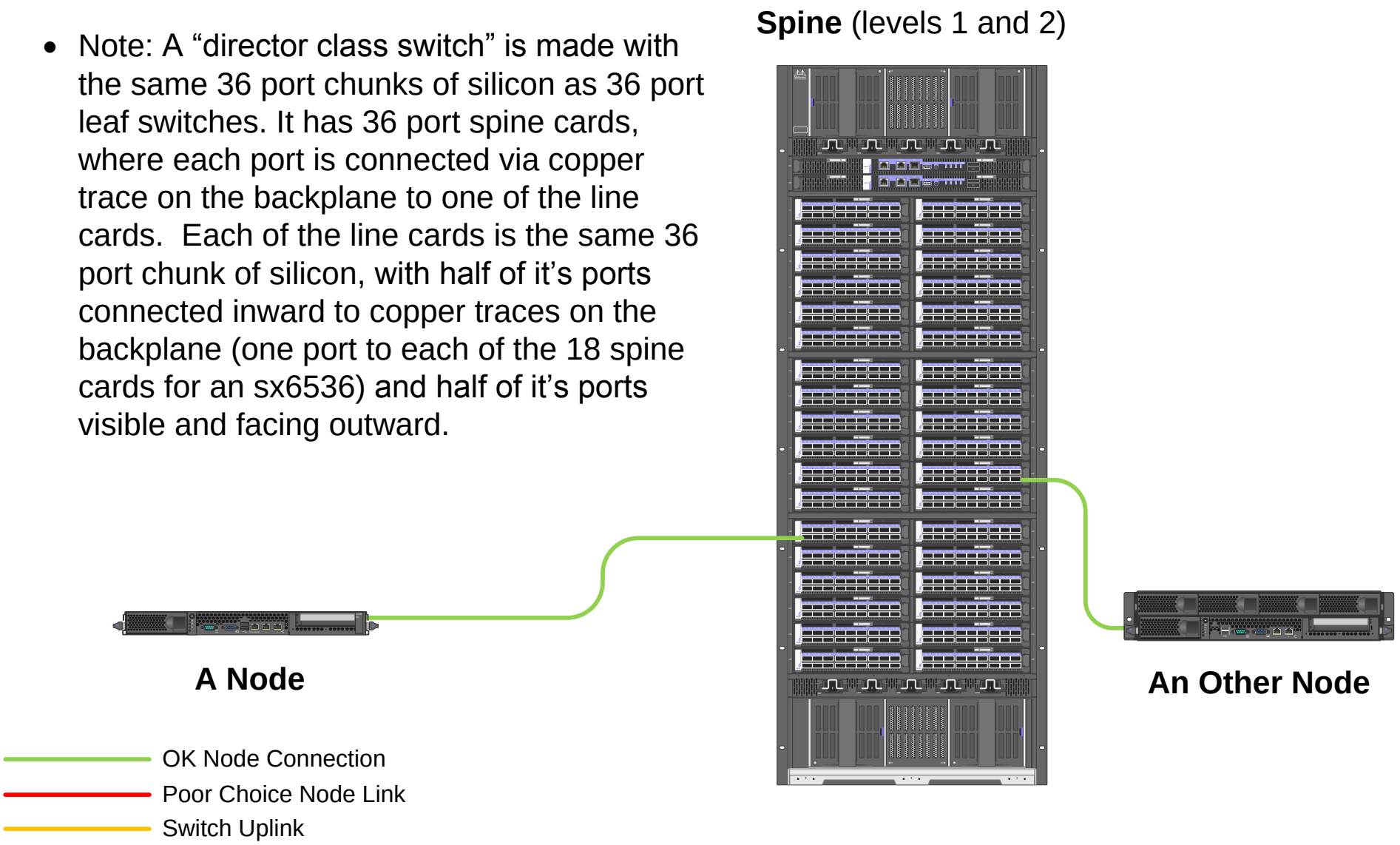
There are certain cases where connecting a node to the spine will work just fine, and others where it can dramatically complicate life for you and your client.

Here are some visual examples of where it is OK to connect nodes in a fabric, including when you can connect nodes directly to the spine.

Where Can I Connect my Nodes?

Single director class switch with no leaf switches.

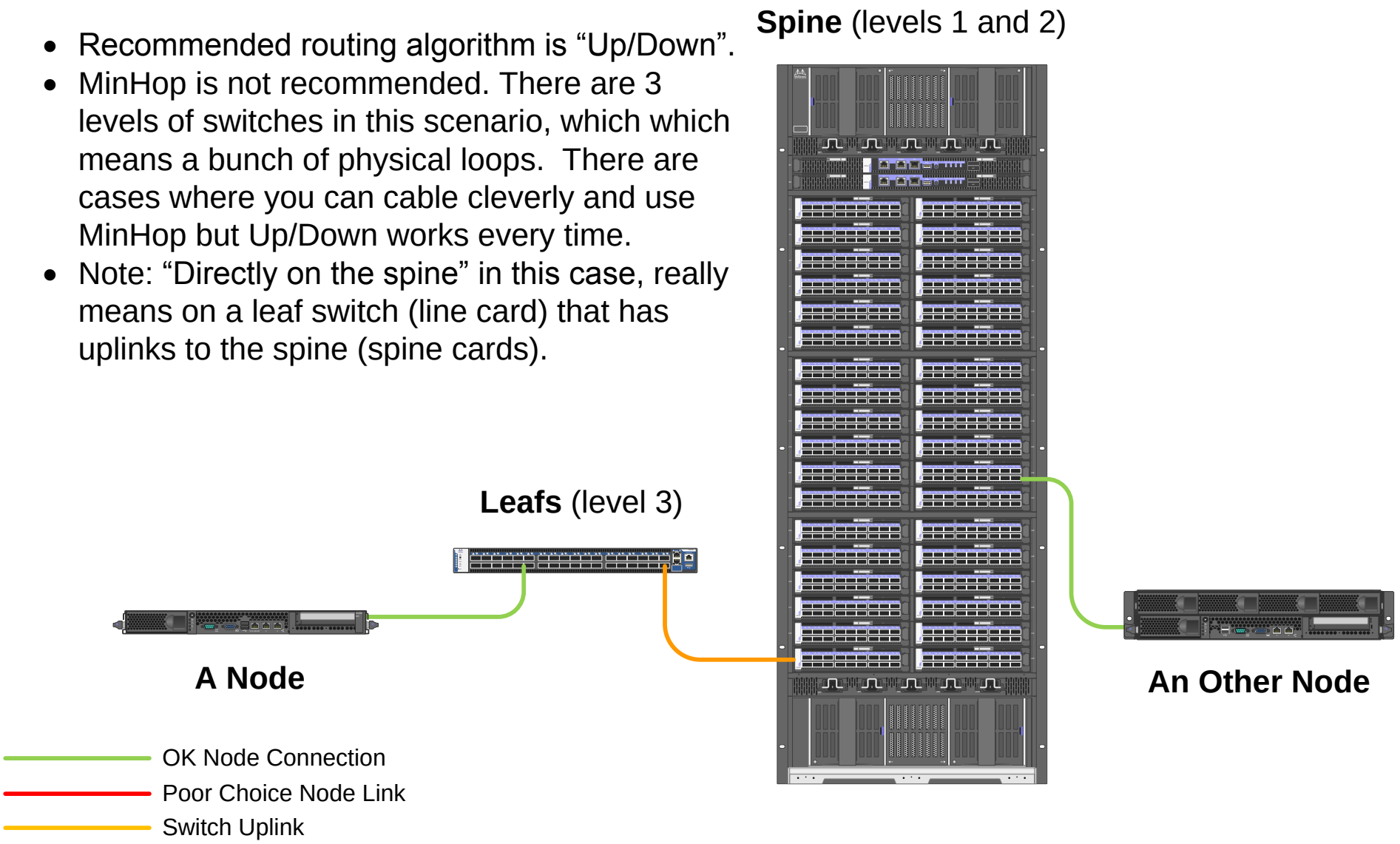
- Note: A “director class switch” is made with the same 36 port chunks of silicon as 36 port leaf switches. It has 36 port spine cards, where each port is connected via copper trace on the backplane to one of the line cards. Each of the line cards is the same 36 port chunk of silicon, with half of it’s ports connected inward to copper traces on the backplane (one port to each of the 18 spine cards for an sx6536) and half of it’s ports visible and facing outward.



Where Can I Connect my Nodes?

Single director class switch with one layer of leaf switches.

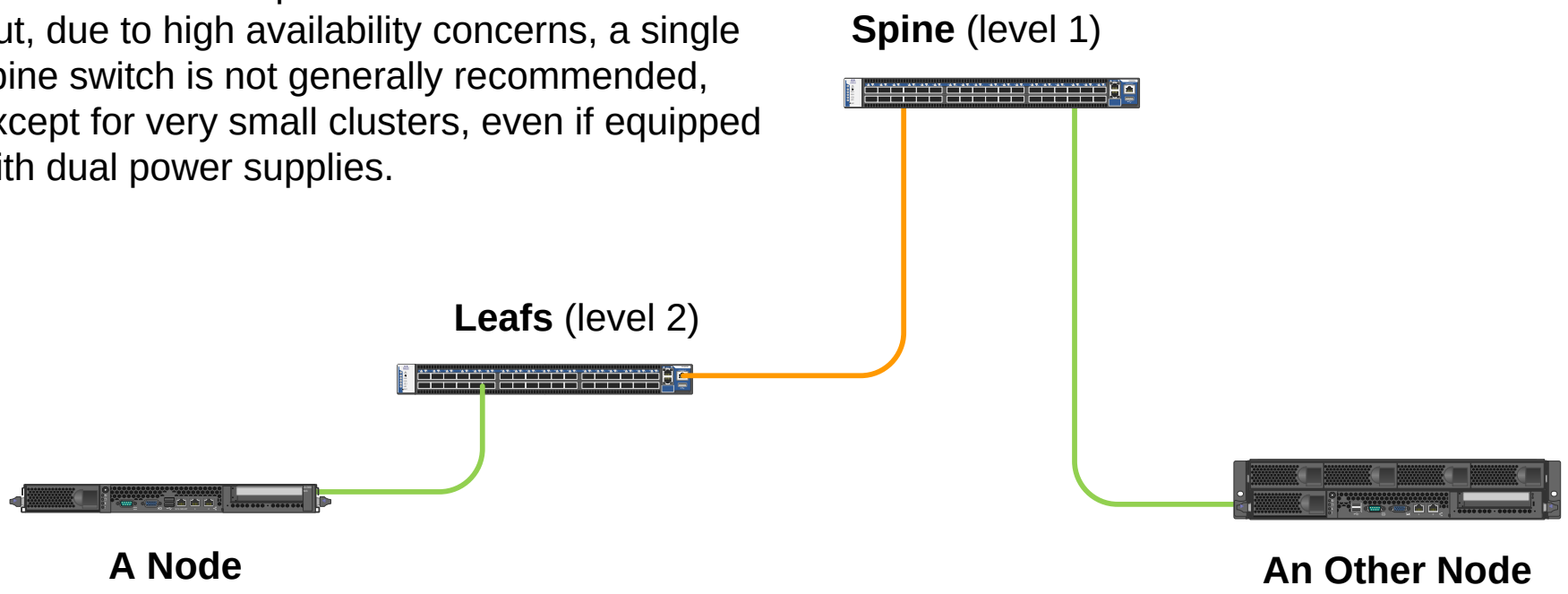
- Recommended routing algorithm is “Up/Down”.
- MinHop is not recommended. There are 3 levels of switches in this scenario, which which means a bunch of physical loops. There are cases where you can cable cleverly and use MinHop but Up/Down works every time.
- Note: “Directly on the spine” in this case, really means on a leaf switch (line card) that has uplinks to the spine (spine cards).



Where Can I Connect my Nodes?

Single 36 port switch as spine, with one layer of leaf switches.

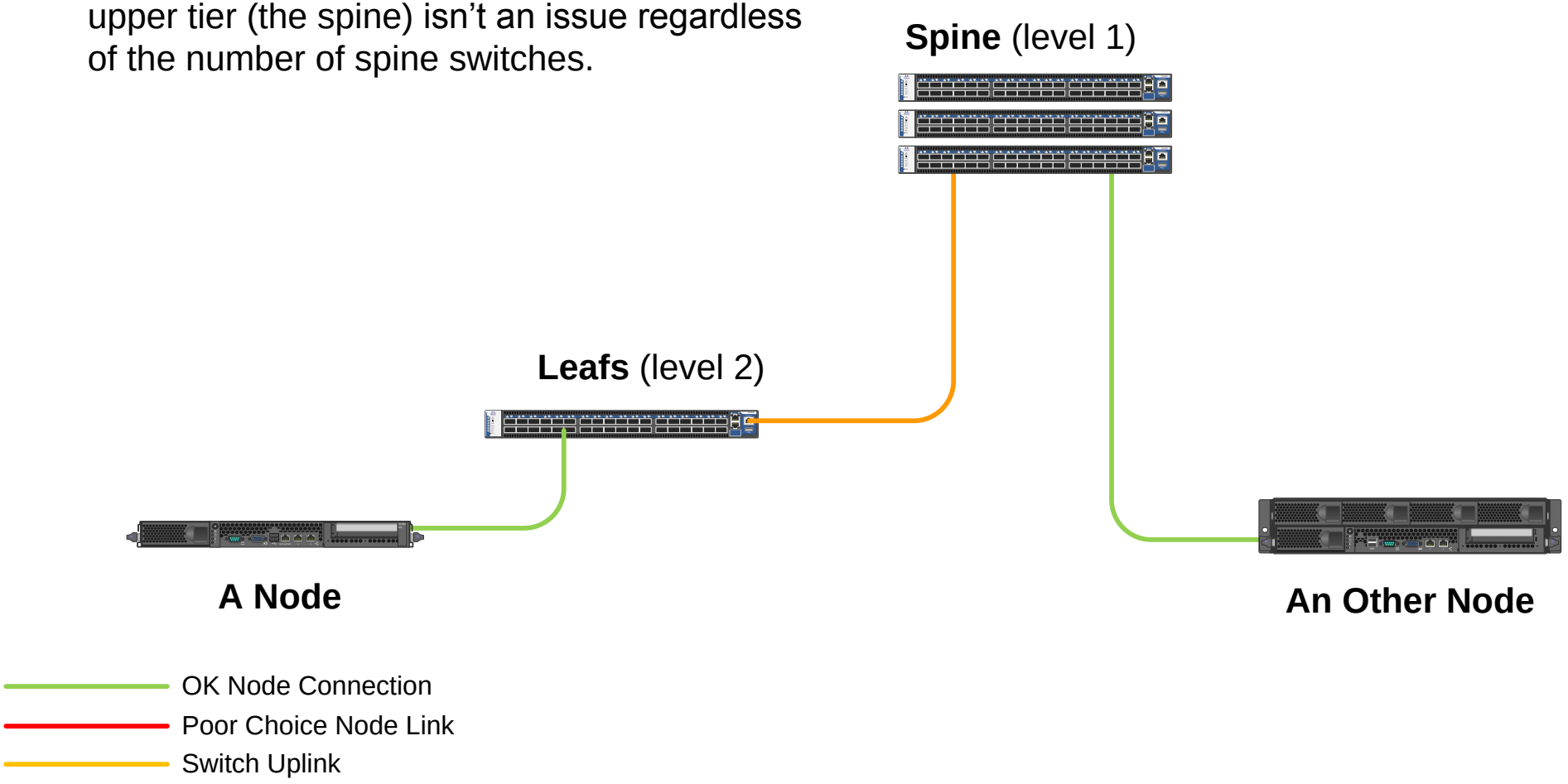
- Recommended protocols: MinHop or Up/Down
- You can use MinHop on any fat tree that only has 2 levels, therefore connecting nodes to the upper tier (the spine) isn't an issue regardless of the number of spine switches.
- But, due to high availability concerns, a single spine switch is not generally recommended, except for very small clusters, even if equipped with dual power supplies.



Where Can I Connect my Nodes?

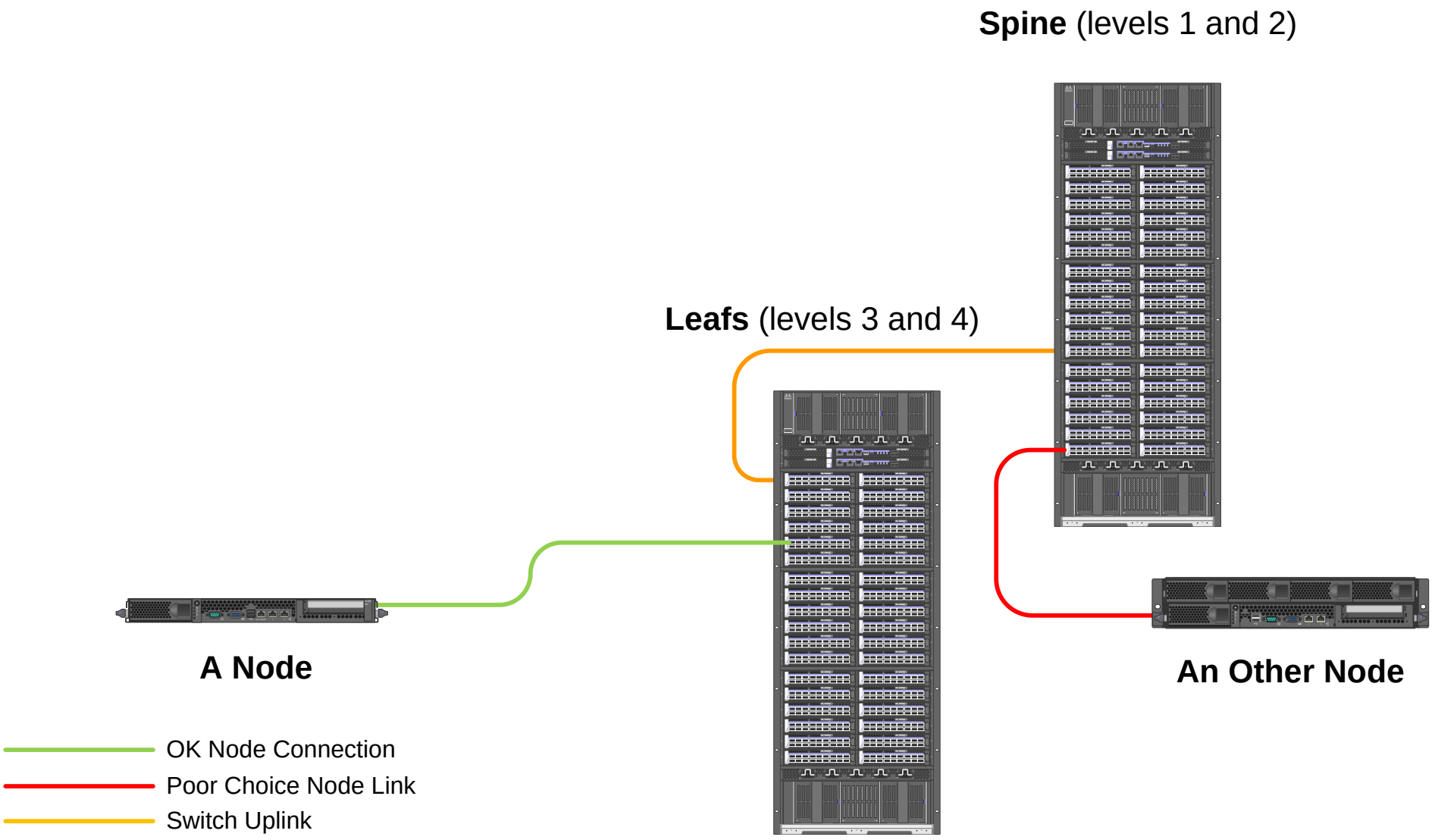
Multiple 36 port switch based spine, with one layer of leaf switches.

- Recommended protocols: MinHop or Up/Down
- You can use MinHop on any fat tree that only has 2 levels, therefore connecting nodes to the upper tier (the spine) isn't an issue regardless of the number of spine switches.



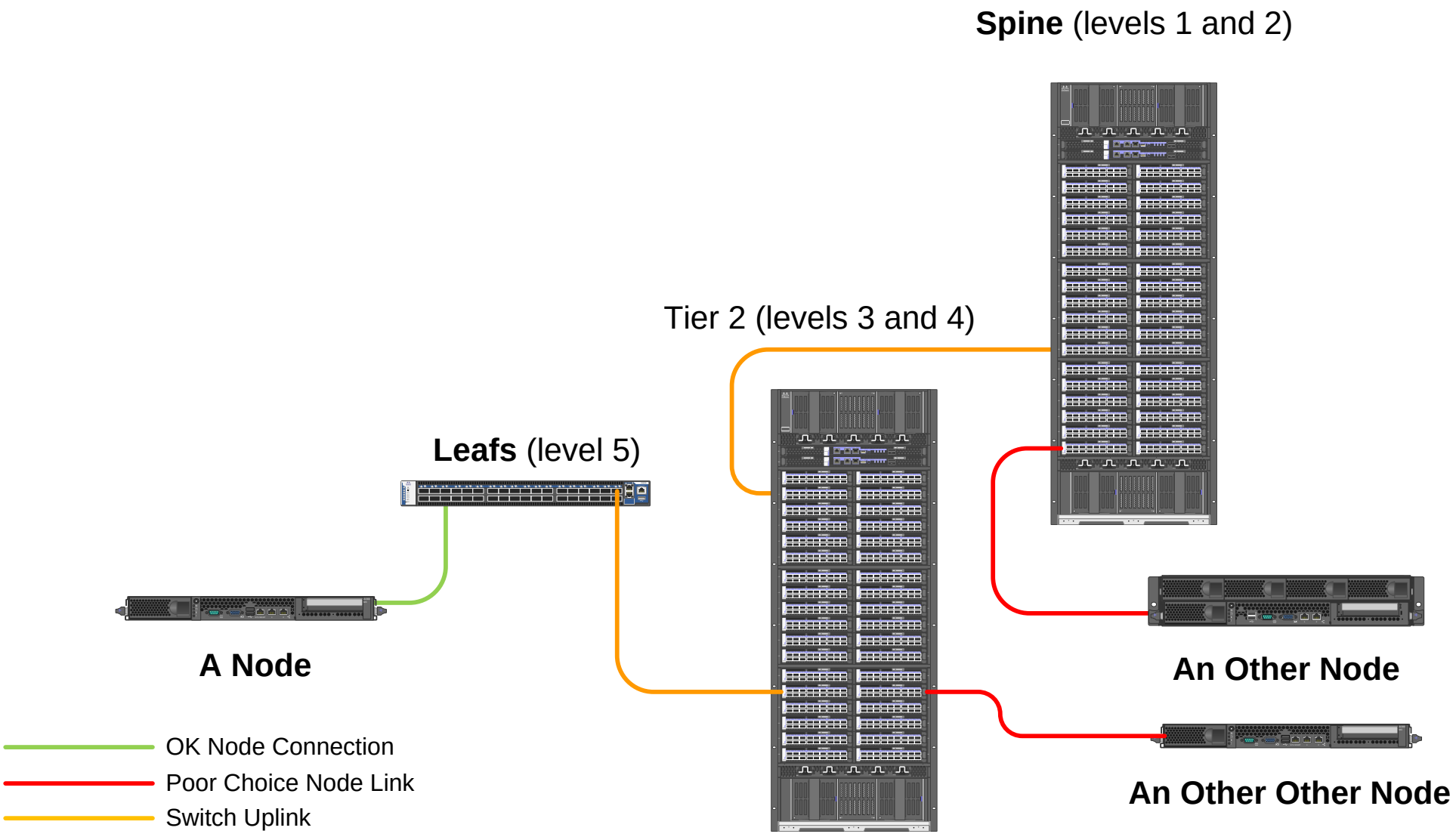
Where Can I Connect my Nodes?

Single director class switch as spine, other director class switches as leafs.



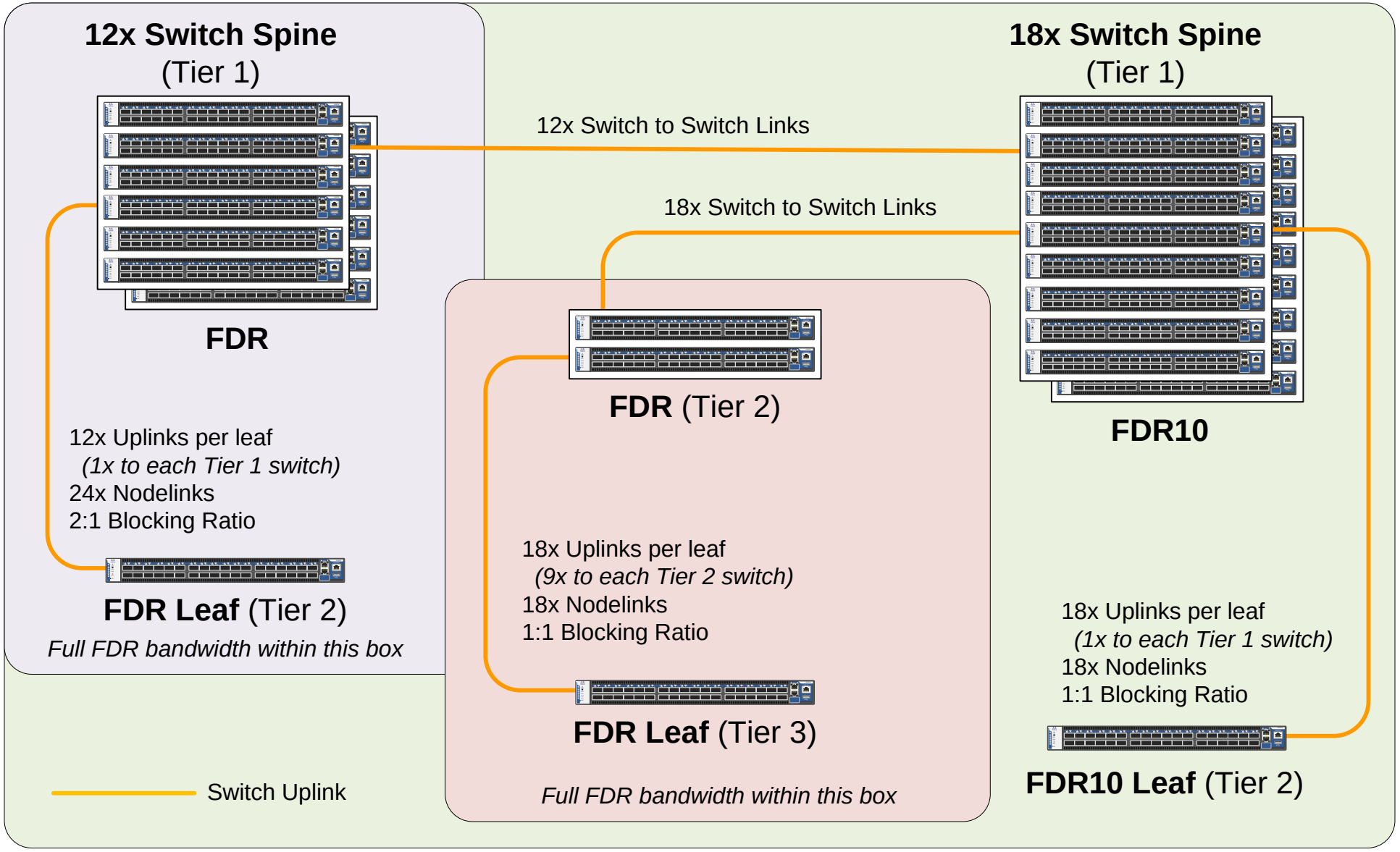
Where Can I Connect my Nodes?

Single director class switch as spine, other director class switches as a second tier, with 36 port switches as leafs.



A Valid Multi-Speed, Multi-Blocking Ratio Fabric

- All 12x FDR and all 18x FDR10 are roots
- UpDn Protocol



For more information, contact:

Brian Finley
IBM, Executive IT Specialist
ATS Cloud and Technical Computing
Mobile: +1 469.667.2110
bfinley@us.ibm.com

Brian Forbes
Mellanox, Sr. Solutions Architect
Mobile: +1 415.706.2755
brianf@mellanox.com