When Can I Connect a Node to the Spine?

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 when it is OK and Not OK to connect nodes directly to the spine.

This Visual HOWTO was Prepared by:
Brian Finley, IBM
Brian Forbes, Mellanox

When Can I Connect a Node to the Spine?

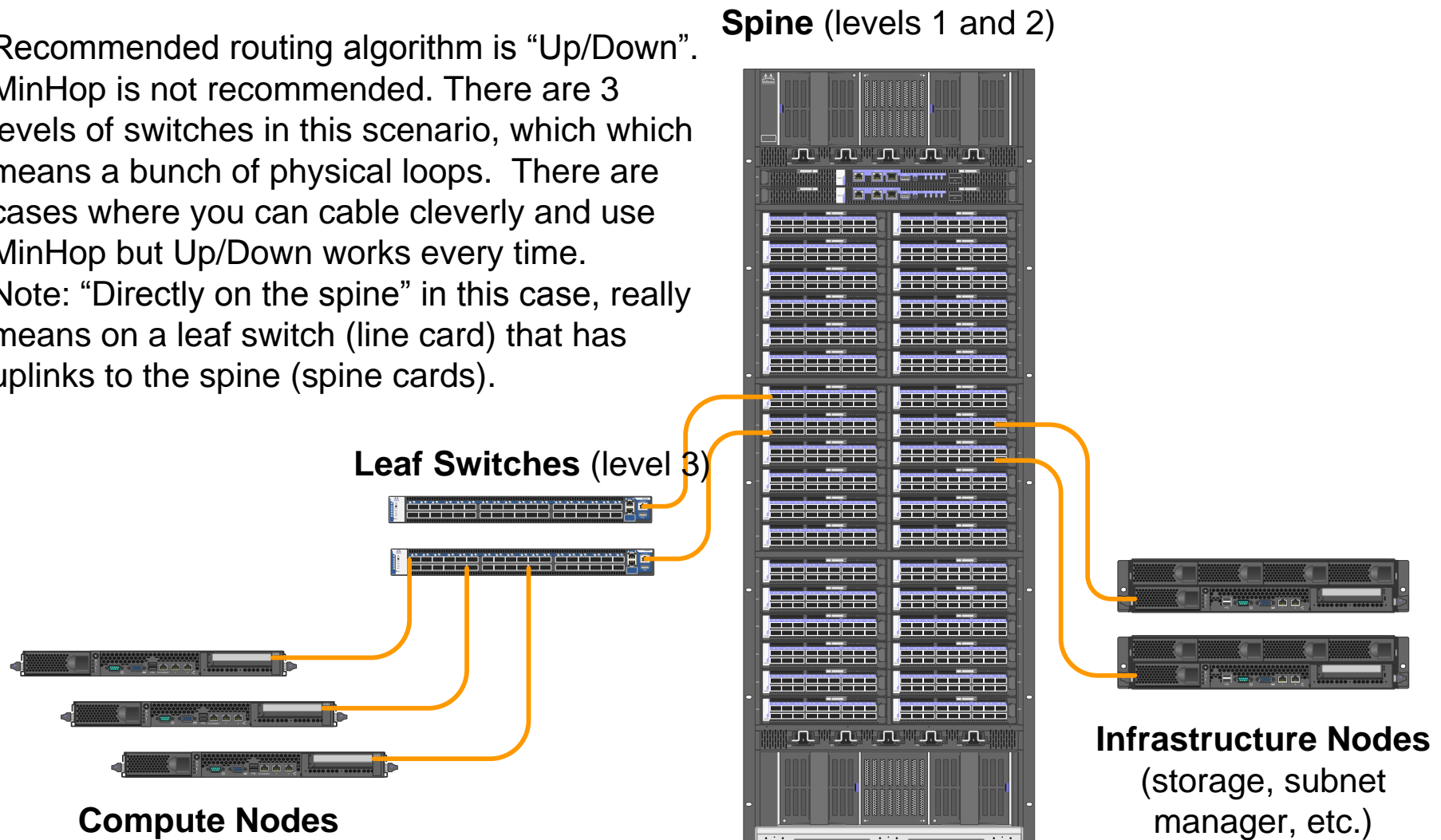**OK:** A single director class switch with no leaf switches.

- 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.

Spine (levels 1 and 2)

Compute Nodes

Infrastructure Nodes
(storage, subnet manager, etc.)
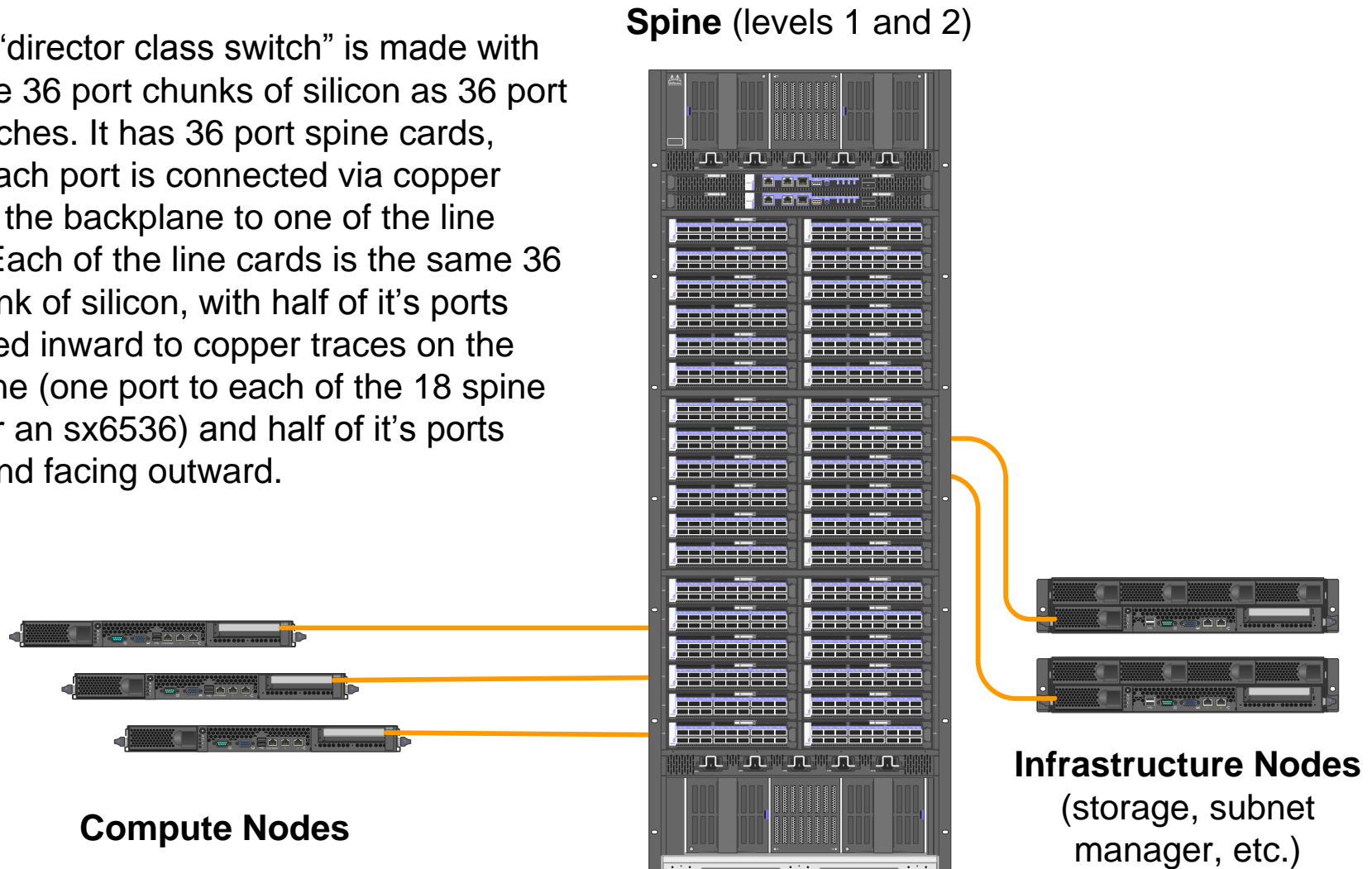
When Can I Connect a Node to the Spine?

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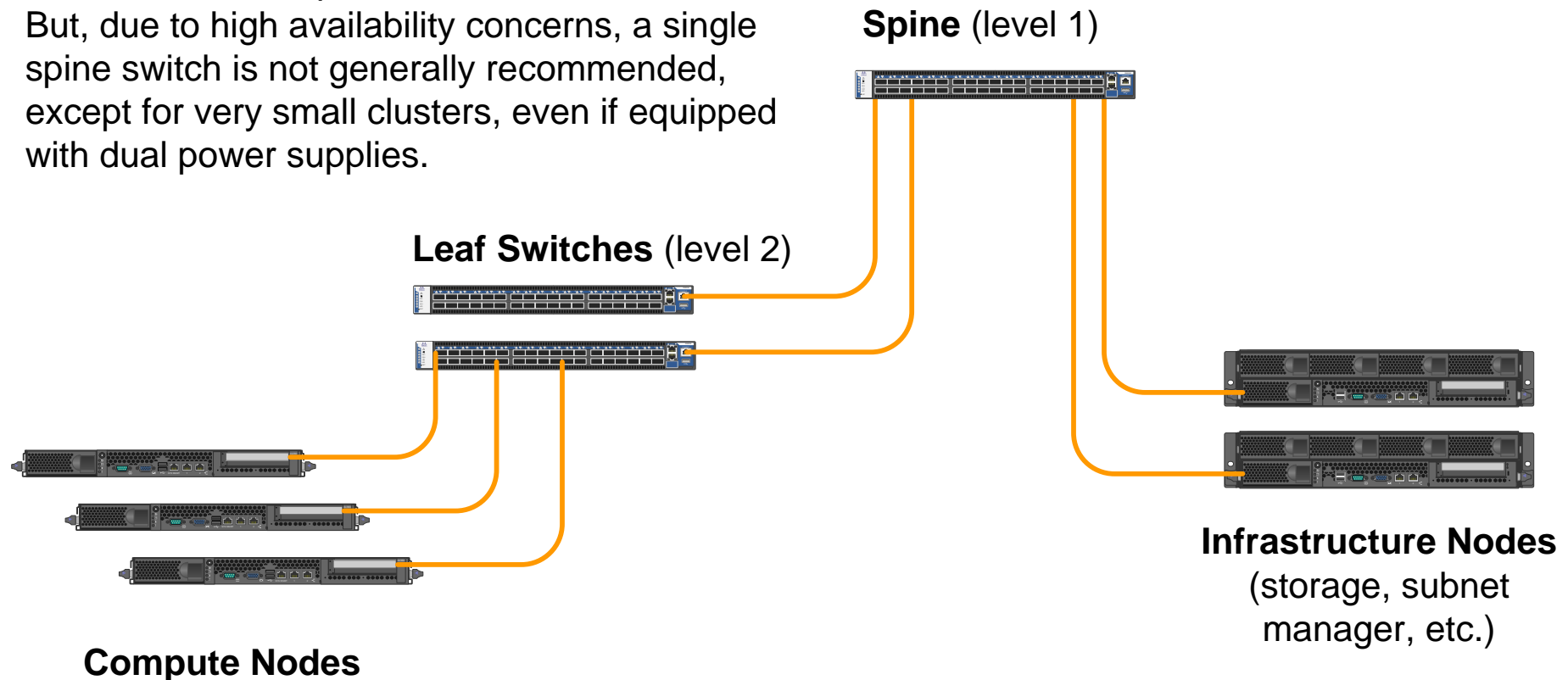
- Recommended routing algorithm is "Up/Down".
- MinHop is not recommended. There are 3 levels of switches in this scenario, which which means a bunch of physical loops. There are cases where you can cable cleverly and use MinHop but Up/Down works every time.
- Note: "Directly on the spine" in this case,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

Spine (levels 1 and 2)

Leaf Switches (level 3)

Compute Nodes

Infrastructure Nodes
(storage, subnet manager, etc.)

When Can I Connect a Node to the Spine?

OK: A single 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.
- 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.

Spine (level 1)

Leaf Switches (level 2)

Infrastructure Nodes (storage, subnet manager, etc.)

Compute Nodes

When Can I Connect a Node to the Spine?

OK: A 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.
- 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.

Spine (level 1)

Leaf Switches (level 2)

Infrastructure Nodes
(storage, subnet manager, etc.)

Compute Nodes