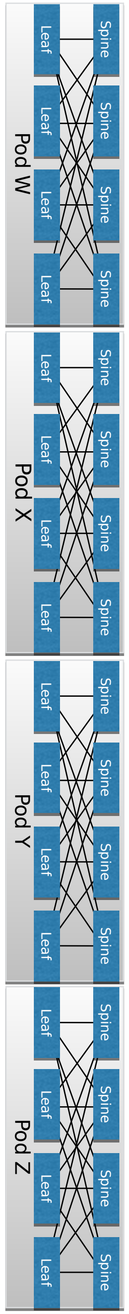
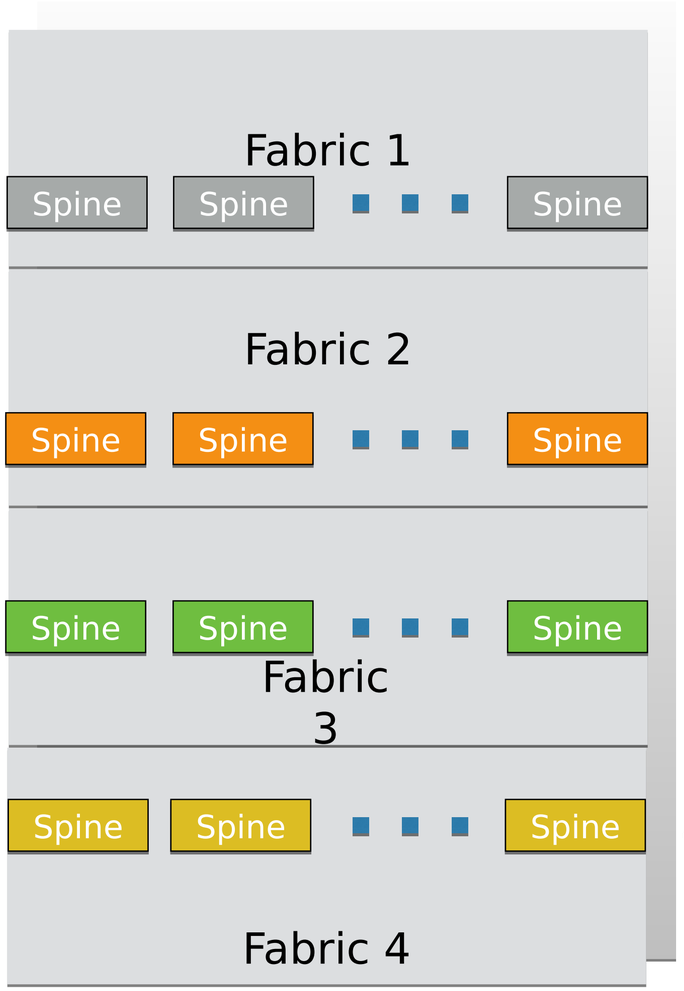
Data Center Routing Challenges - LinkedIn

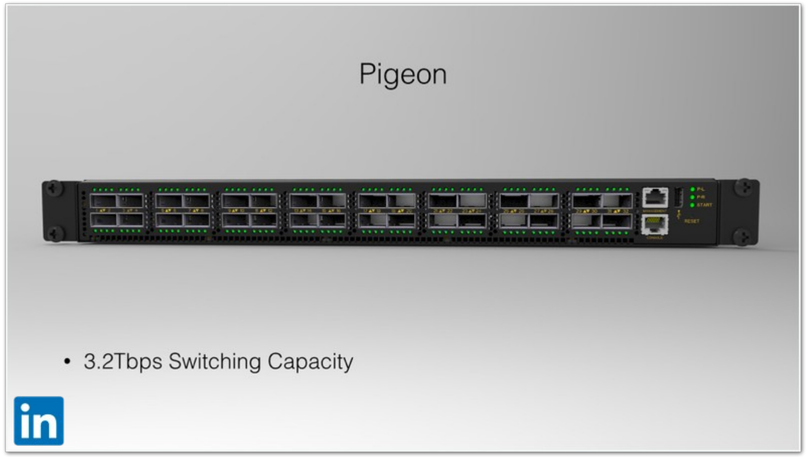
Russ White

Shawn Zandi

Single SKU Data Center





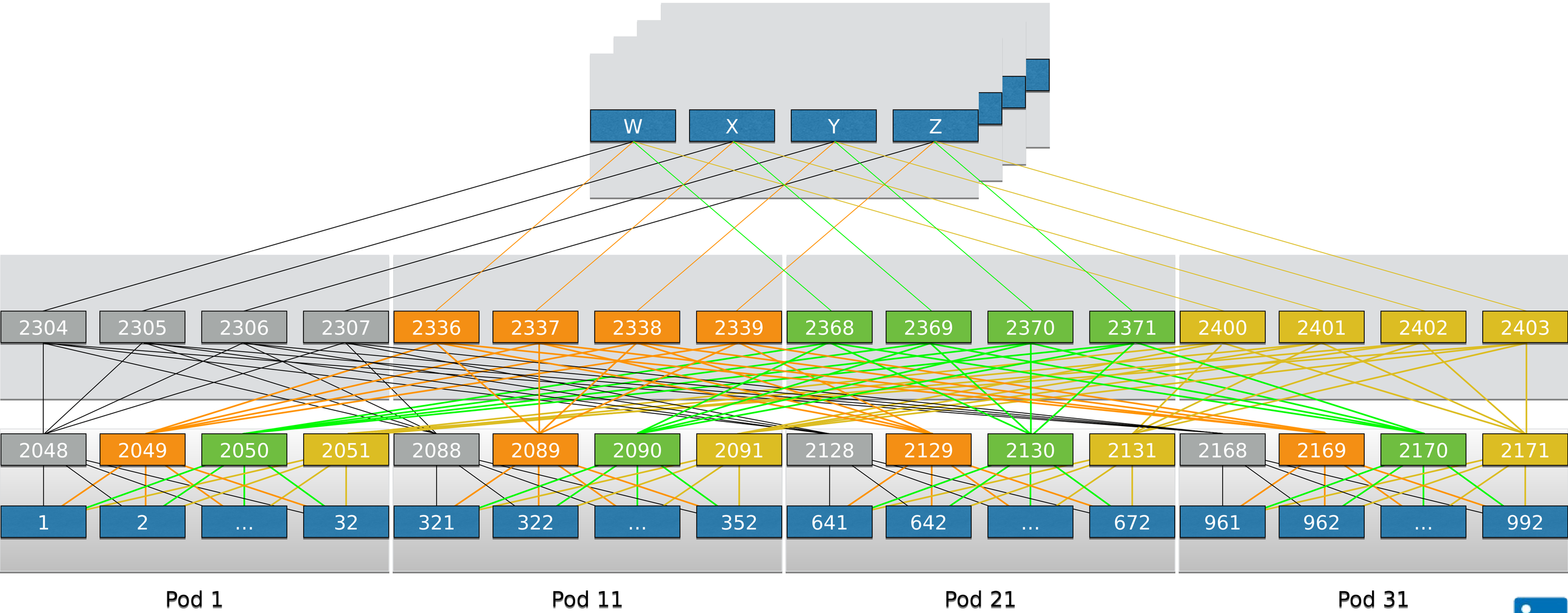


4,096 x100G ports
Non-Blocking
Scale-out

Complexity within Chassis

- Chassis: Robust-yet-Fragile
  - Complex due to NSR, ISSU, feature-sets, etc.
  - Larger fault domain, Failover/Fail-back
  - Indeterministic boot up process and long upgrade procedures
- Moved complexity from big boxes to pizza boxes, where we can easily manage and control!
- Better control and visibility to internals by removing black-box abstraction!
- Same Switch SKU on ToR, Leaf and Spine (Entire DC)
- Single chipset uniform IO design (same bandwidth, latency and buffering)
  - True 5-Staged Clos Topology! with deterministic latency
- Dedicated control plane, OAM and CPU for each ASIC

Control Plane Complexity at Scale



Control Plane Requirements

Fast, simple distributed control plane
No tags, bells, or whistles (no hacks, no policy)
Auto discover neighbors and build RIB
Minimal (to zero) configuration
Must use TLVs for future, backward compatible, extensibility
Must carry MPLS labels (per node/interface)

Control Plane

BGP

- Heavy weight; lots of features and “stuff” that are not needed
- Modifications to support single IP configuration required
- Does not supply full topology view
- Proven scaling

IS-IS

- Not proven to scale in this environment
- Light weight
- Most requirements for zero configuration are already met
- Provides full topology view

Build New

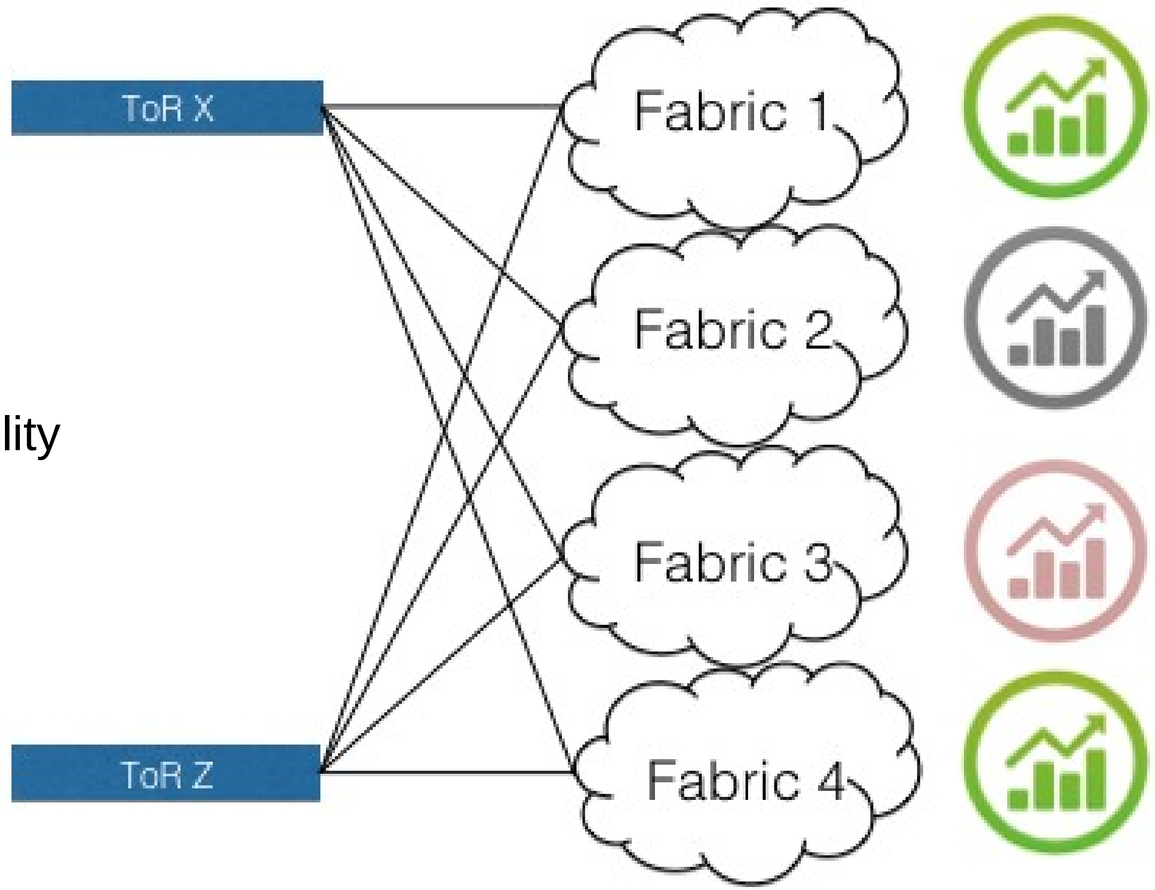
- A lot of work
- But could use bits and pieces from other places

Forwarding Challenges

- ECMP is blind
- End to end path selection is required for some applications.
- Application / Operator cannot easily enforce a path...

ECMP: Blind Forwarding

Reachability ≠ Availability



Other challenges

- Auto-Configuration is important. Protocols should negotiate and come up without any manual configuration...
- Provisioning can be simplified (lack of standardization)
- Turning on a network requires another network (out of band)

(To hardware vendors) BMC in every switch is a MUST!