

Welcome

Free Range Routing

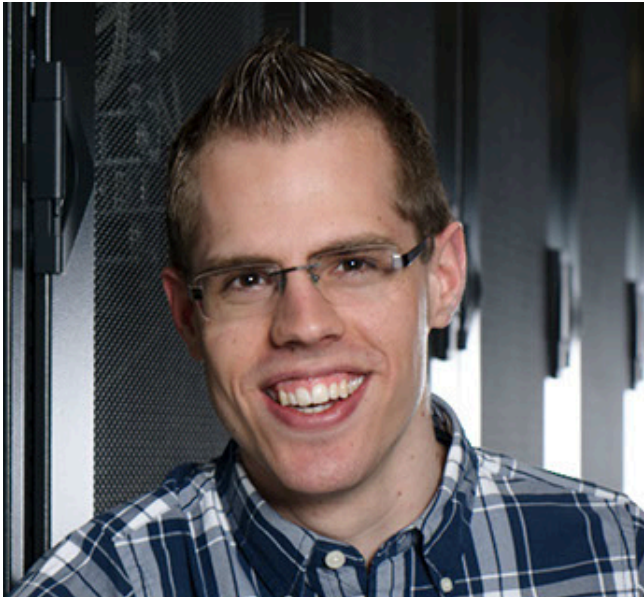
or how we ditched OSPF for BGP unnumbered (based on RFC5549)

2017-11-09, Gurtenpark (Berne)



cloudscale.ch

Your Speaker



Manuel Schweizer
@geitguet

- Network Engineer at cloudscale.ch AG
- Board Member at SwissIX Internet Exchange

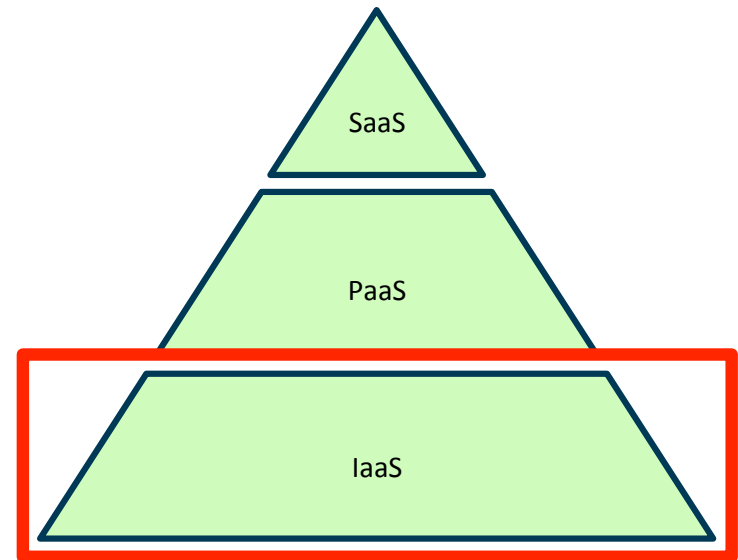
Dayjob



If you choose to, we can be your «someone else»

cloudscale.ch

- Founded in 2014
- **Swiss** IaaS Provider
- Linux Cloud Server (VMs)
- Focus on Simplicity



„For Developers Who Care“

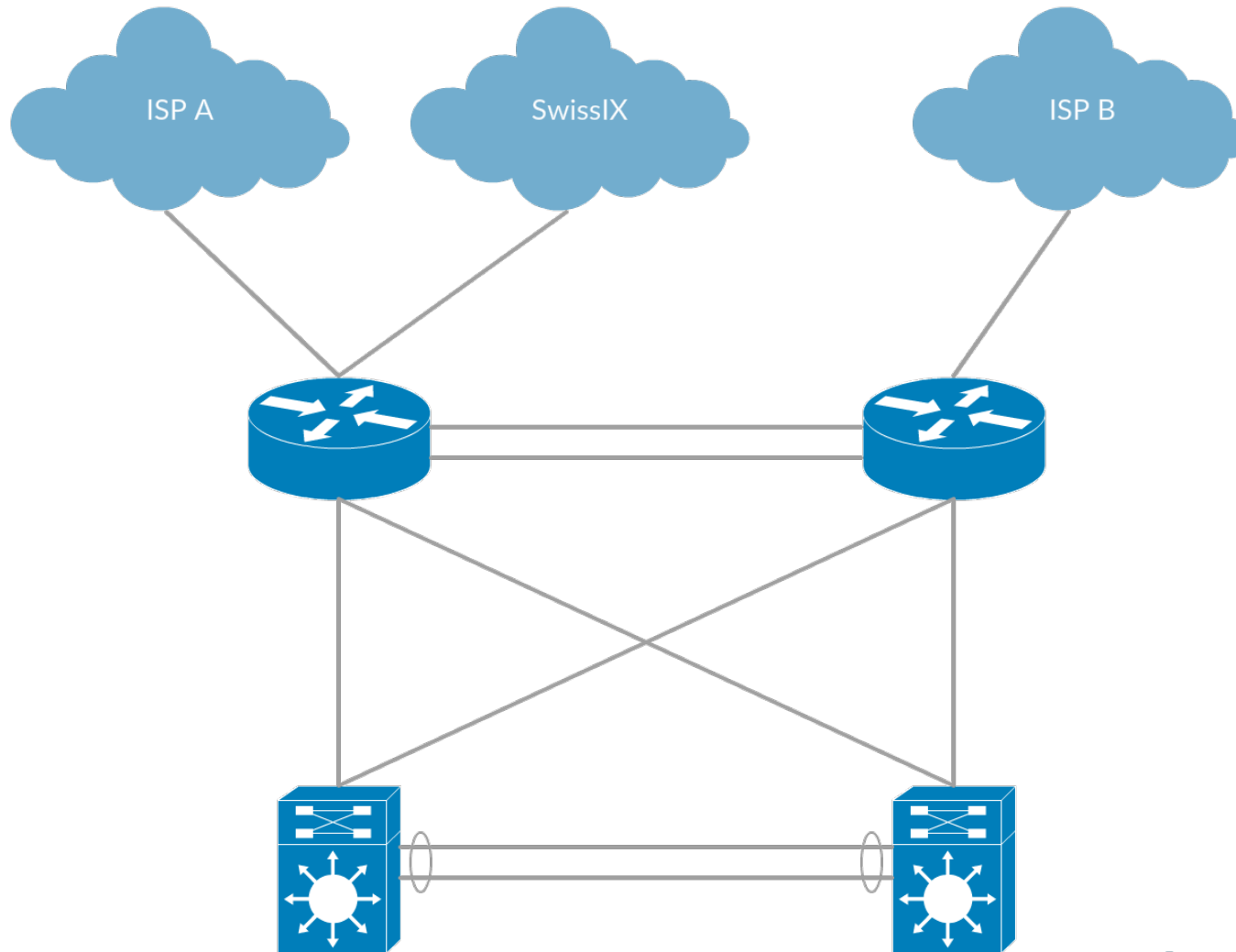
Agenda

- Initial and Target Situation
- Evaluation Phase
- Hardware
- Software
- Demo

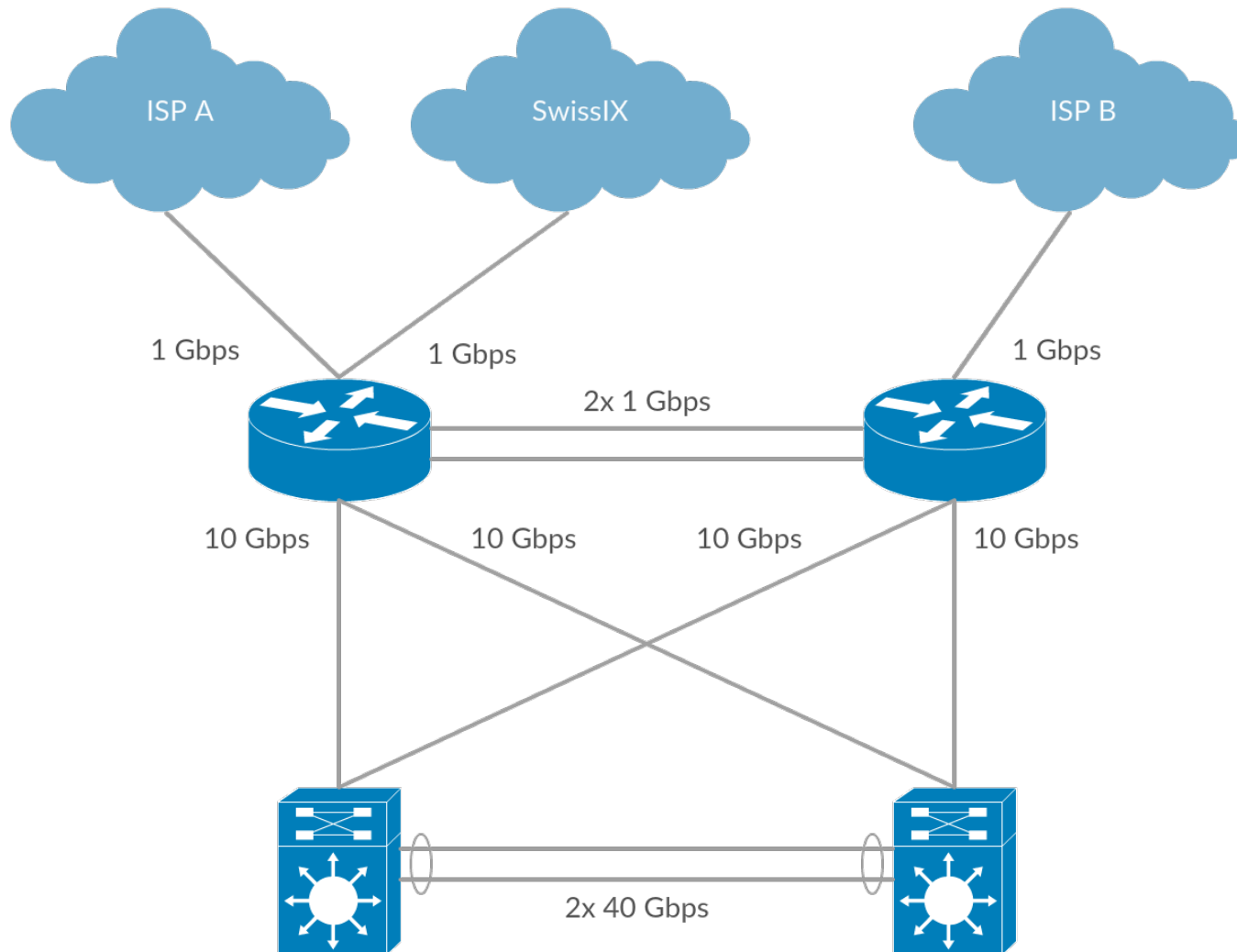
Agenda

- **Initial and Target Situation**
- Evaluation Phase
- Hardware
- Software
- Demo

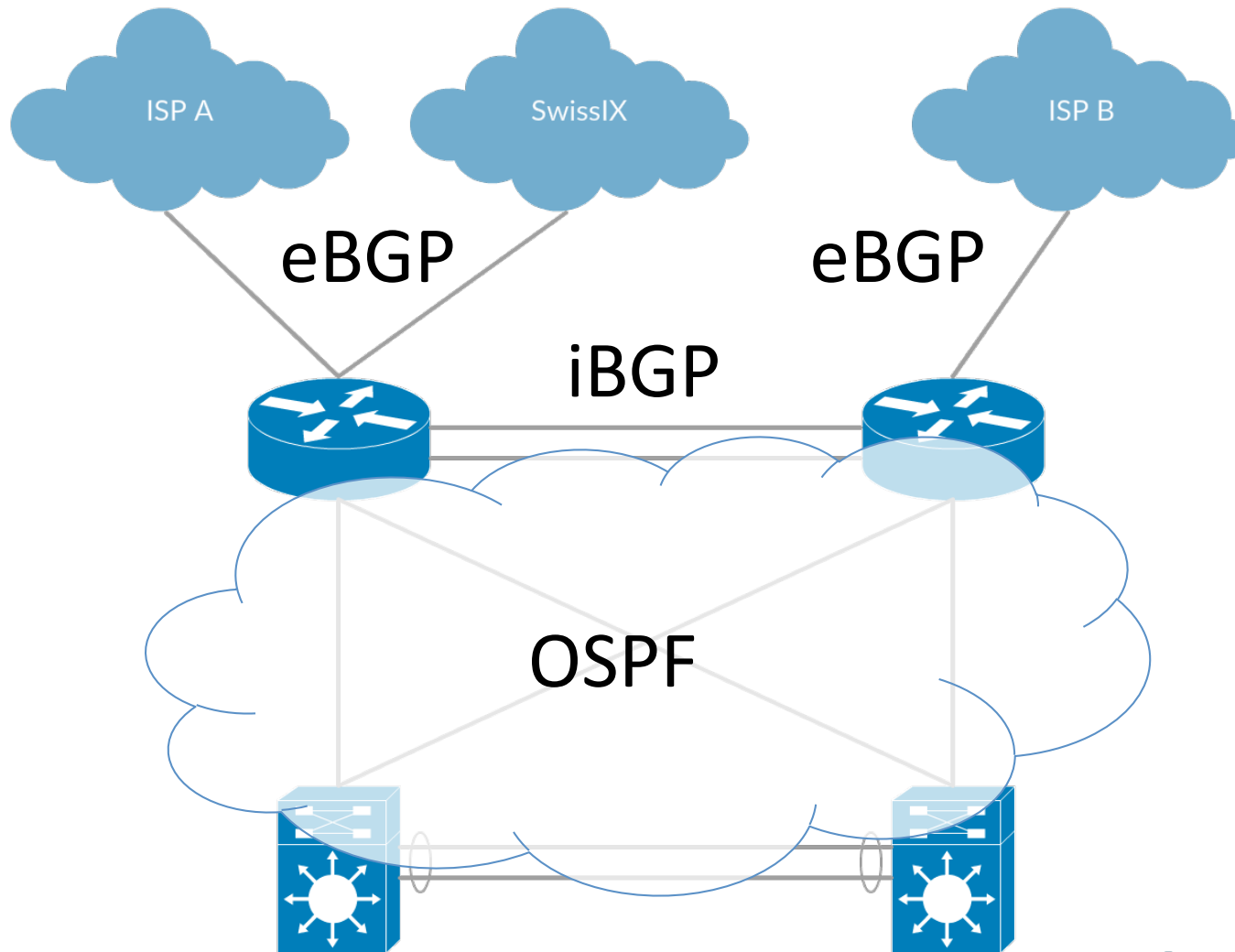
Initial Situation



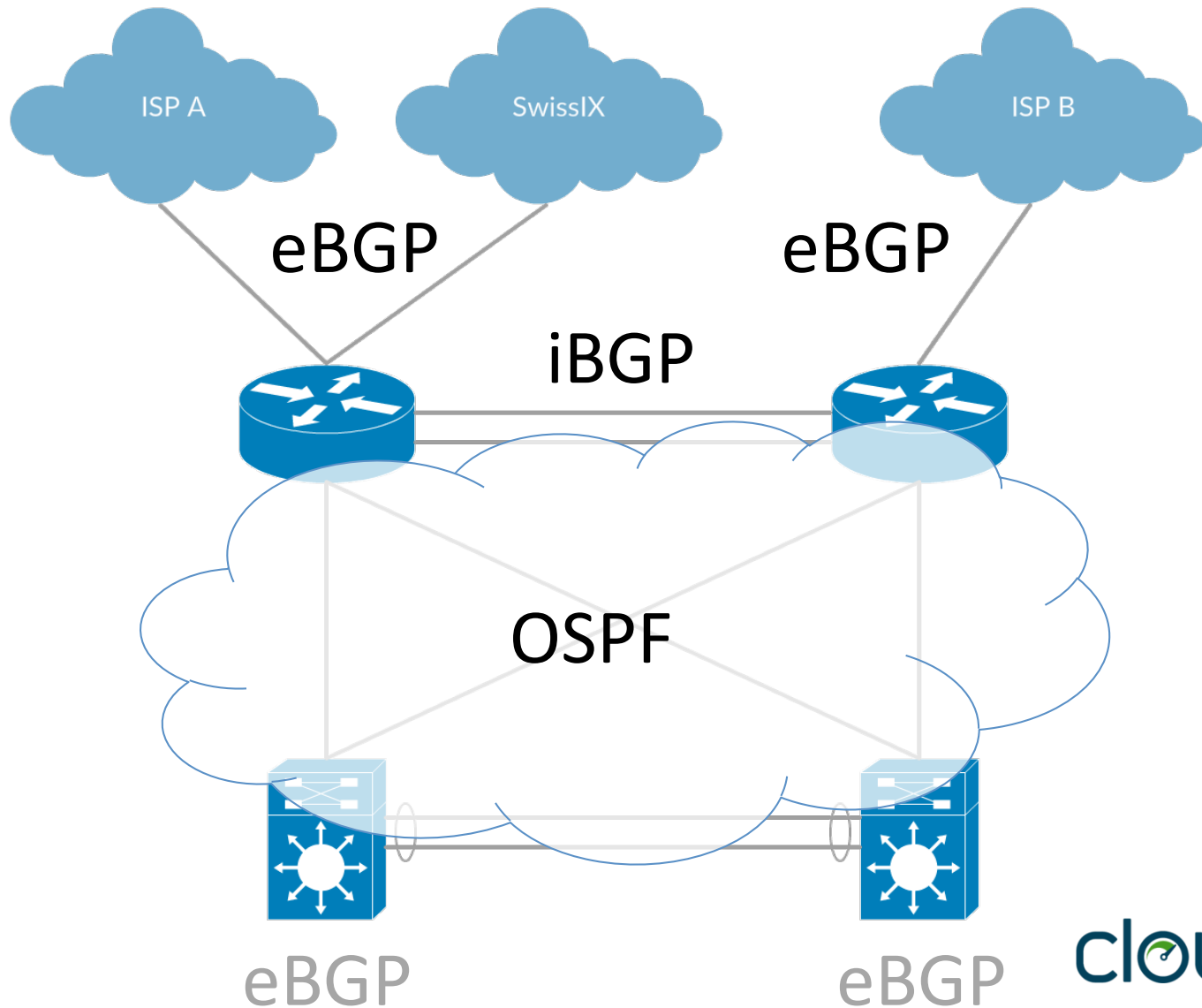
Initial Situation: Bandwidth



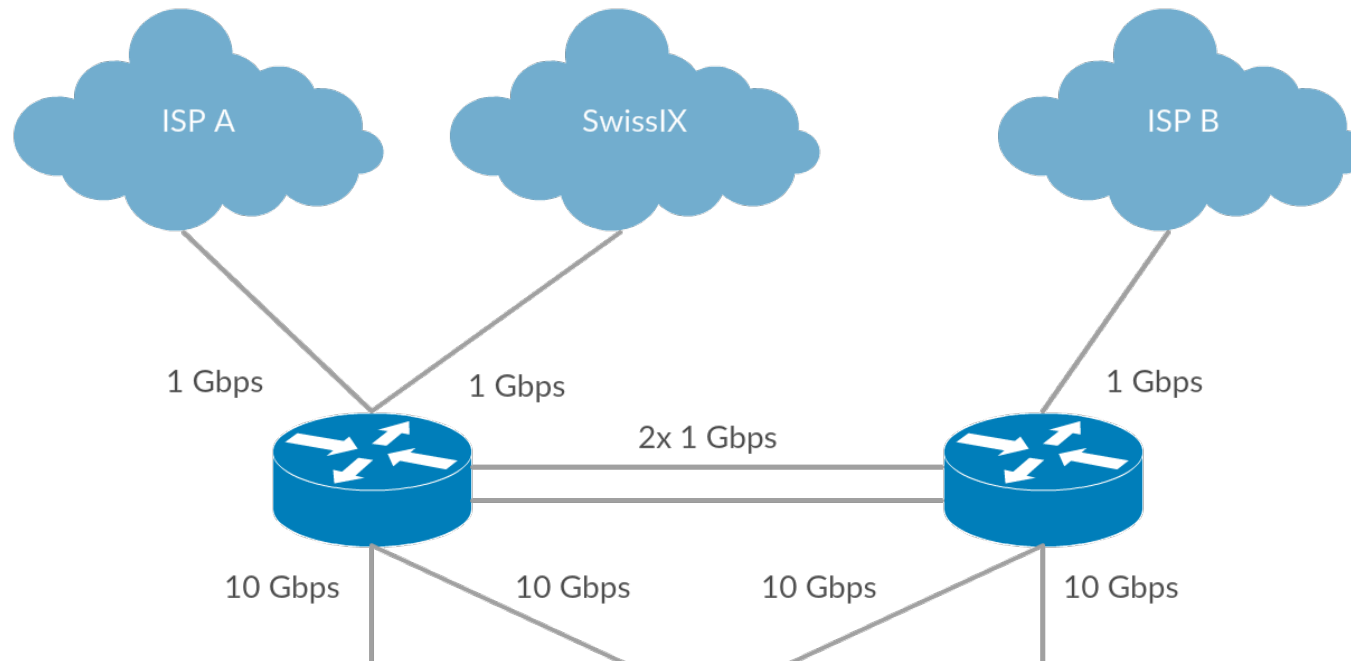
Initial Situation: Routing Protocols



Initial Situation: Routing Protocols

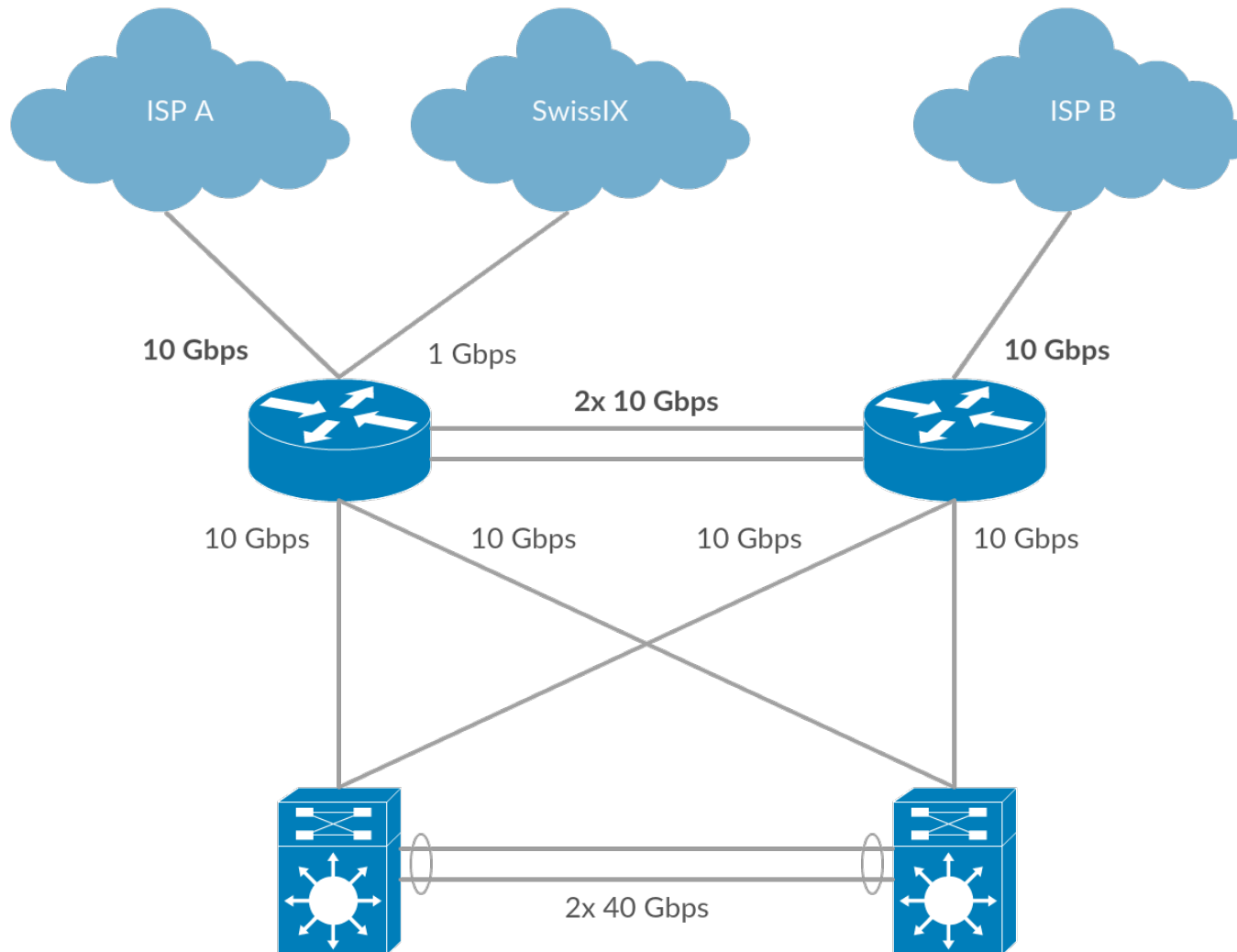


Initial Situation: Summary

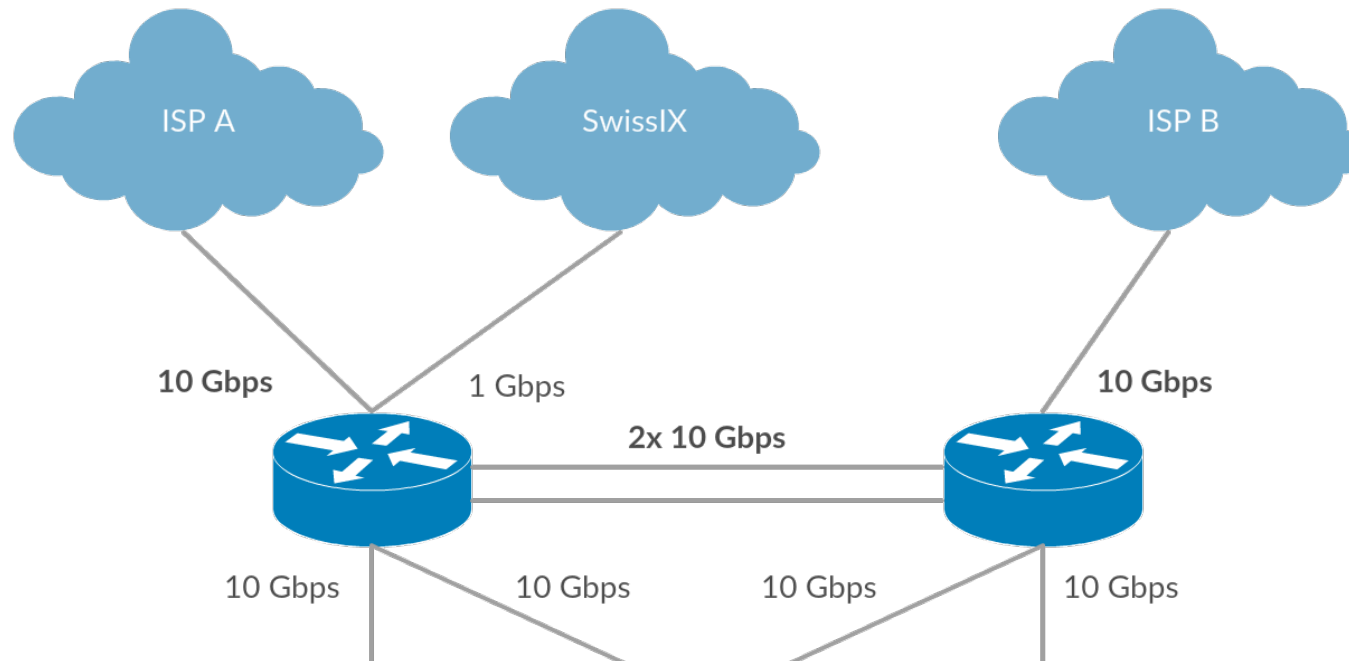


- 2x 1 Gbps IP Transit
- 2x 10 Gbps Interfaces (only!)
- IGP: OSPF (and BGP)
- EGP: BGP

Target Situation: Bandwidth



Target Situation: Summary

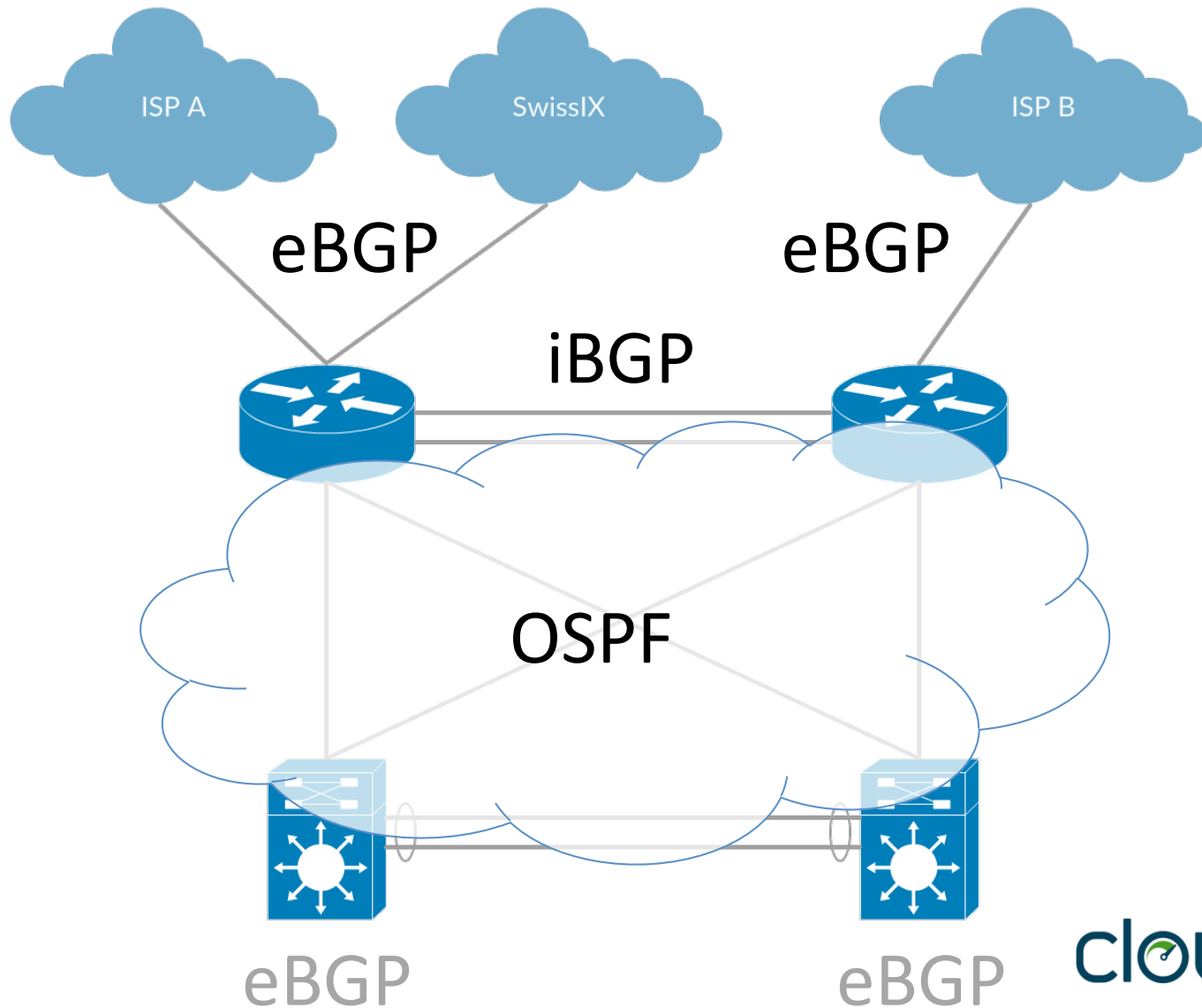


- 2x 10 Gbps IP Transit
- 6-8x 10 Gbps Interfaces
- Reduce Complexity!
- Price...

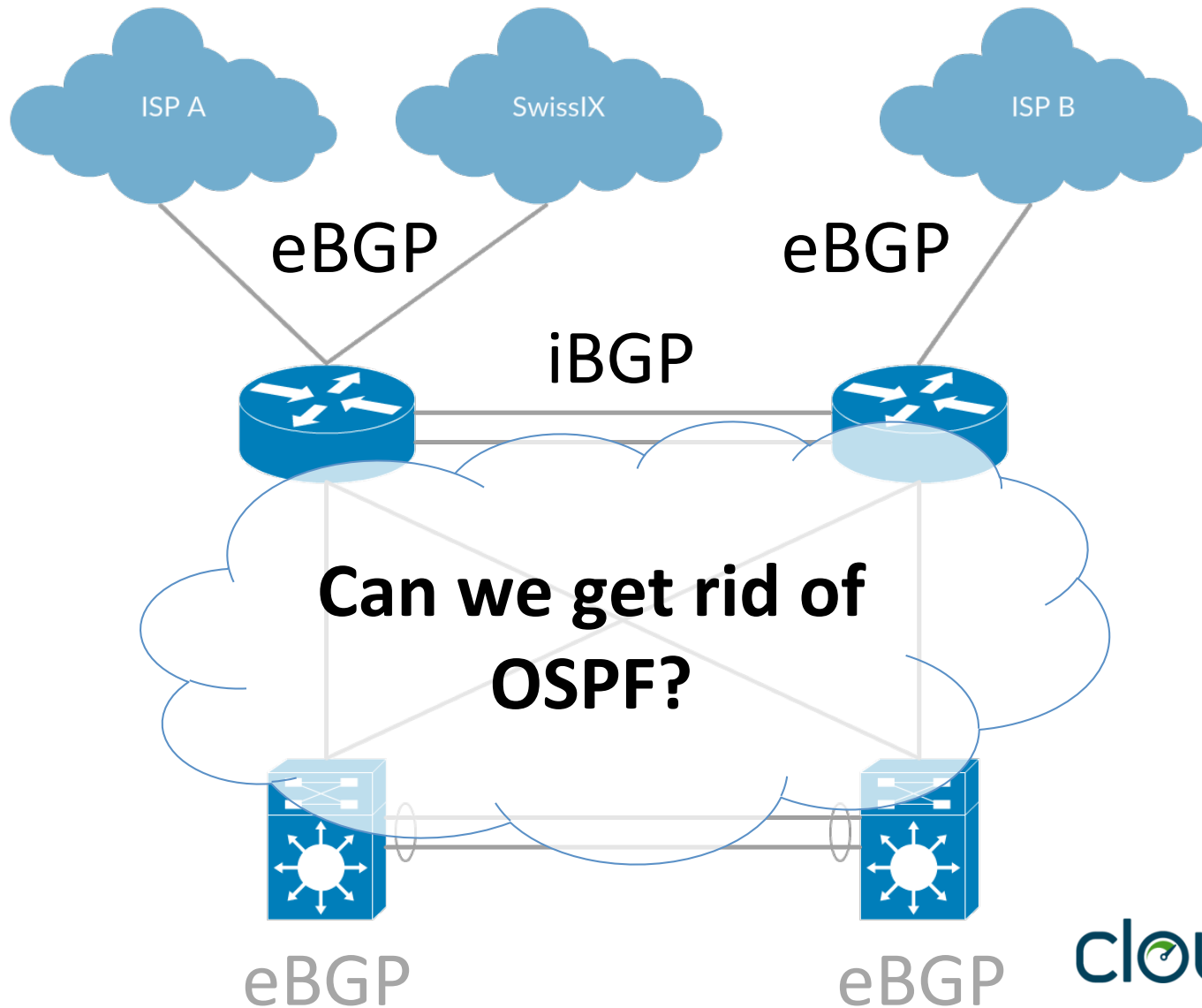
Agenda

- Initial and Target Situation
- **Evaluation Phase**
- Hardware
- Software
- Demo

Evaluation Phase: Reduce Complexity!



Evaluation Phase: Reduce Complexity!



Evaluation Phase: RFC5549

<https://tools.ietf.org/html/rfc5549>

In (very) short:

„[...] this document only concerns itself with the advertisement of IPv4 NLRI (Network Layer Reachability Information) [...] with an IPv6 Next Hop.“

Evaluation Phase: RFC5549

- Use of existing IPv6 link-local address
- You are running dual-stack, are you?
- Next hop: Loopback IP address

Evaluation Phase: Price...

- Commercial routers with 8x 10 Gbps:
 - Starting at CHF 10-15k (hardware only!)
 - + support contract
 - + license fees
- Experience so far:
 - TAC, oh boy
 - Blackbox (bugs => workarounds?)

Evaluation Phase: Free Range Routing

- Supports BGP unnumbered (RFC5549)
- Runs on x86 hardware and Linux
- Proven basis for Big Switch, Cumulus etc.

Evaluation Phase: Free Range Routing

- Supports BGP unnumbered

Reduce Complexity

- Runs on x86 hardware and

Low Price

- Proven basis for Big Switch, Cumulus

Stability

FRR – About the Project

- FOSS (Free and Open Source Software)
- Open Community Model
- Linux Foundation Project (since 04/2017)
- Version 3.0.2 released 2 days ago
- Fork of Quagga

FRR – What's Different?

- Methodical vetting of submissions
- Extensive automated testing of contributions
- Git pull requests
- Github centered development
- Elected maintainers & steering committee
- Common assets held in trust by the Linux Foundation

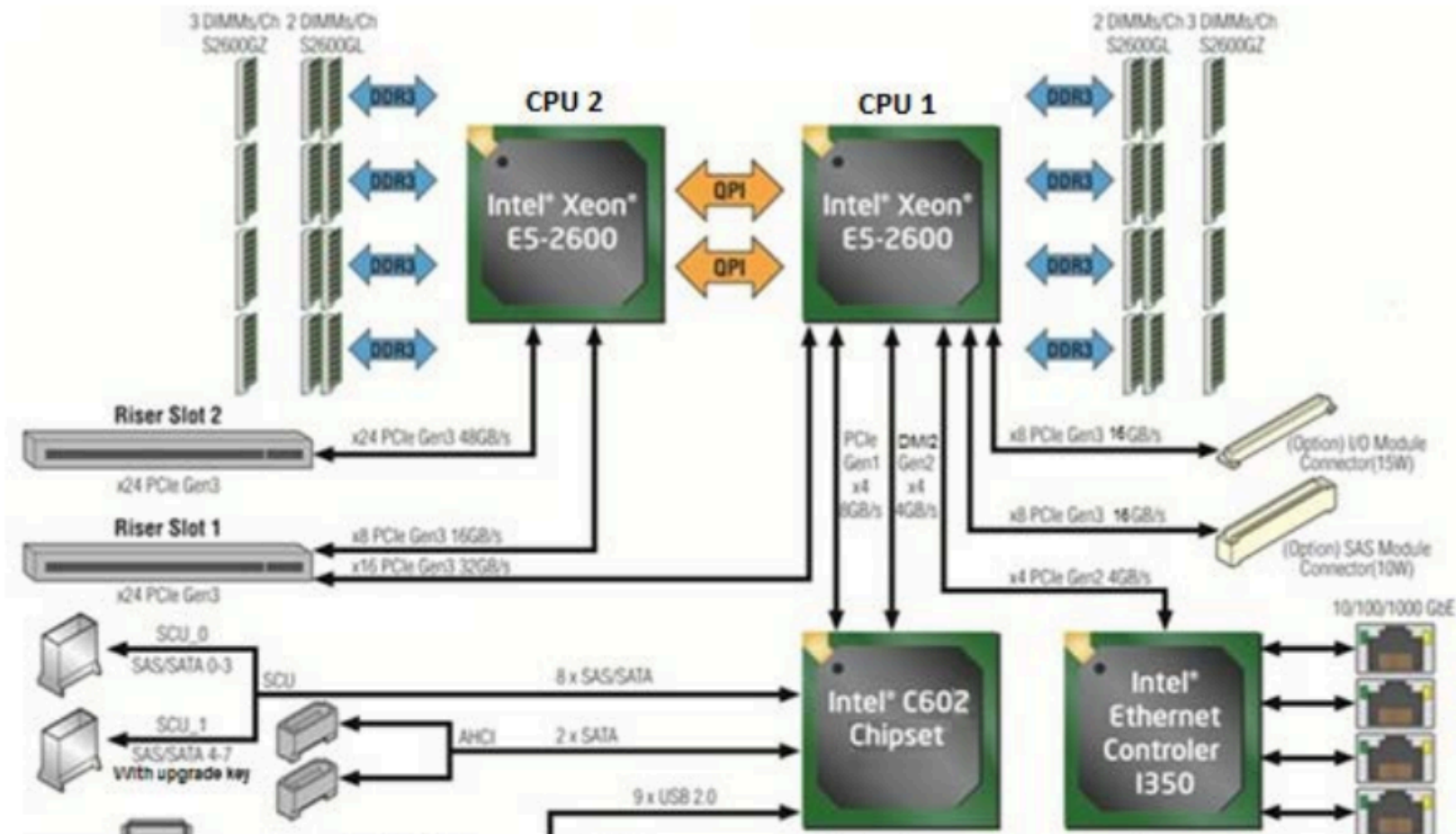
FRR – Links

- Website:
<https://frrouting.org>
- Github:
<https://github.com/FRRouting/frr/>
- Issue Tracker:
<https://github.com/FRRouting/frr/issues>
- Continuous Integration:
<https://ci1.netdef.org/browse/FRR>

Agenda

- Initial and Target Situation
- Evaluation Phase
- **Hardware**
- Software
- Demo

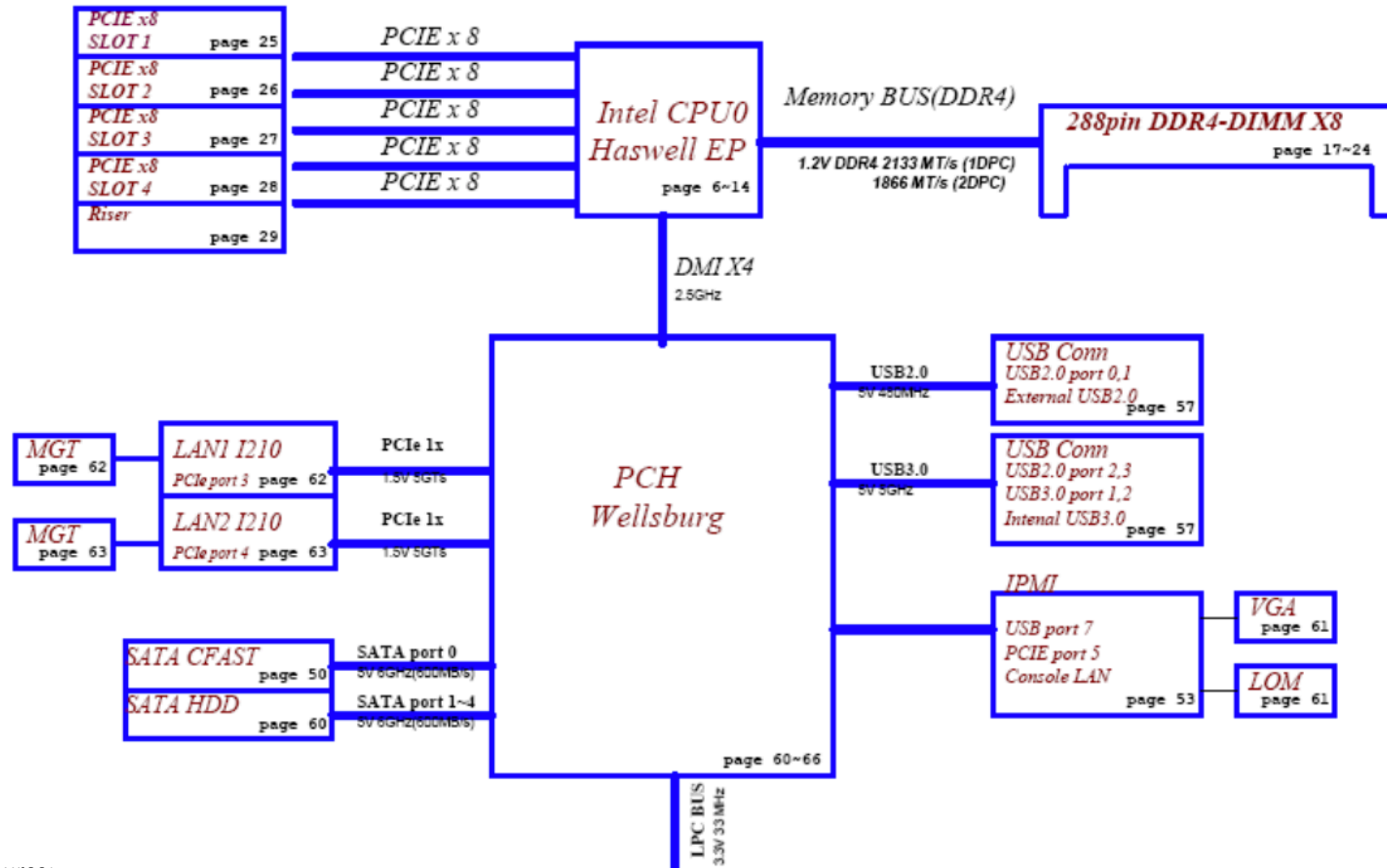
Intel 1U Server Hardware



Source:

https://www.intel.com/content/dam/support/us/en/documents/motherboards/server/sb/s2600gzgl_tps_r2_4.pdf - Page 11

NCA-5510 Block Diagram



Source:

<http://www.lannerinc.com/download-center/User-Manuals/x86-Network-Appliances/?download=1840> - Page 14

Hardware: Lanner NCA-5510



- Dual PSU
- Hot swappable fans
- 4x front-facing PCIe x8

Source:

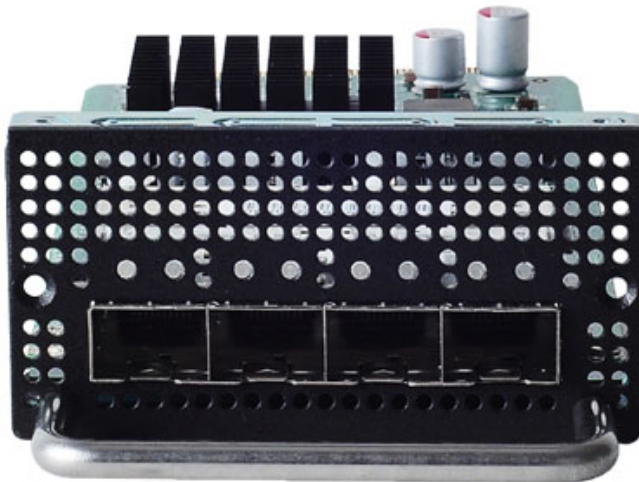
<http://www.lannerinc.com/network-appliances/x86-rackmount-network-appliances/?view=article&id=1667:nca-5510>

Hardware: „Linecards“



Source:

<http://www.lannerinc.com/support/download-center/brochures?download=1086>



- 4-8x 1 Gbps
- 2-4x 10 Gbps
- 2x 40 Gbps
- 2x 100 Gbps (new!)

Source:

<https://www.landitec.com/products/x86-network-appliance-hardware/ncs2-ixm405a-detail>

Hardware: The Real Deal



- 1x 1 Gbps Mgmt NIC
- Serial Console
- 8x 10 Gbps SFP+
- 4x 1 Gbps Base-T
- IPMI (LOM)

Agenda

- Initial and Target Situation
- Evaluation Phase
- Hardware
- **Software**
- Demo

Software Setup

- Ubuntu 16.04 LTS
- Xenial HWE Kernel (4.10) – for VRF Support
- FRR 3.x

Additional Packages:

- ifupdown2, iproute2, vrf, mgmt-vrf
- ptmd, lldpd, snmp, hsflowd

Concerns

Question – Answer Game

Security Concerns (1)

Question

You cannot honestly run Linux in the core?

Answer

Our cloud infrastructure depends on Linux.

Most of the commercial vendors use Linux as a basis for their solution.

Security Concerns (2)

Question

But how about security patches?

Answer

What's the release cycle of your current vendor?
Include updating your routers in your scheduled maintenance windows.

Security Concerns (3)

Question

You are using a firewall then, right?

Answer

Firewall = „latency generator“

Services (SSH, SNMP, sFlow) run in Mgmt-VRF only.

Performance Concerns

Question

But how about line-rate forwarding?

Answer

Current CPUs can easily handle ~ 100 Gbps.

In our tests: 20 Gbps = 0.5 CPU cores (out of 10!)

Performance Concerns

Question

Can FRR handle a BGP full table?

Answer

From enabling the BGP session to fully converged in *less than 20 seconds*.

„1206398 RIB entries, using 156 MiB of memory“

Configuration Concerns

Statement

FRR is not for me, I need a CLI.

Answer

vtysh, Cisco-like syntax.

vtysh -c „command“ instead of expect scripts.

Simple transition to config management with Puppet, Ansible etc.

Monitoring Concerns

Statement

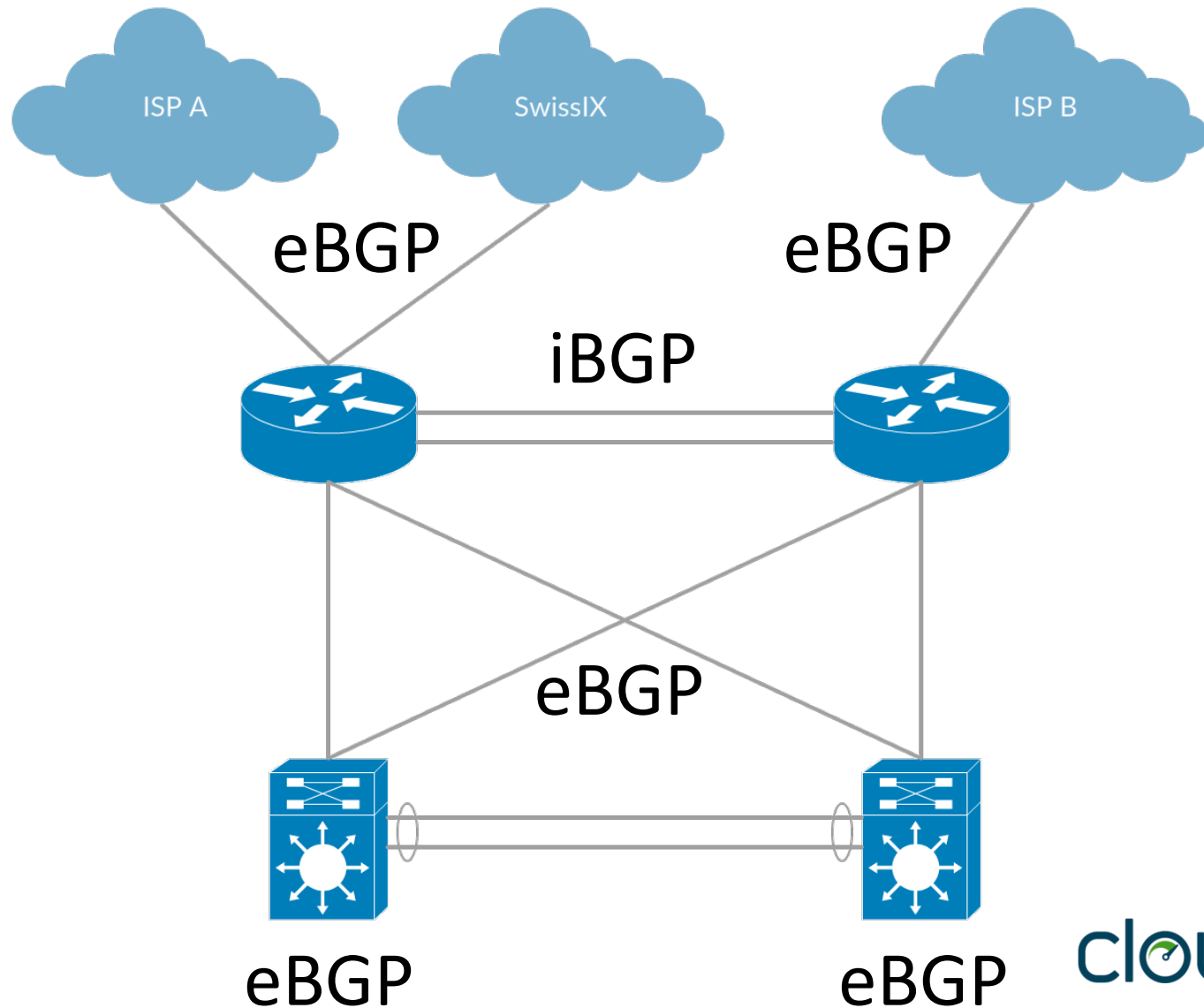
But I need SNMP!

Answer

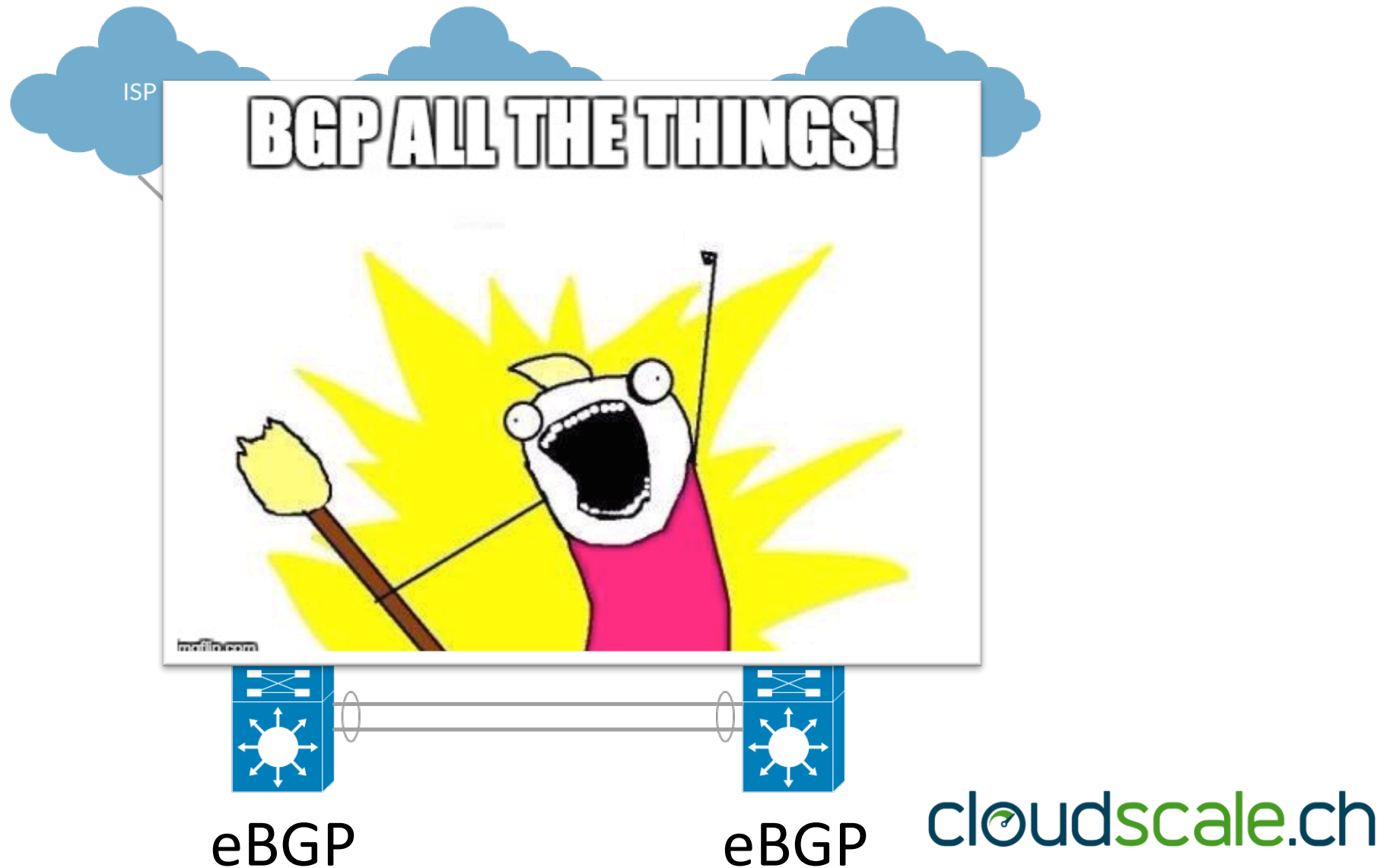
Available as a package.

Also: Use Zabbix, Icinga2, ... directly on your routers.

Software Setup: Routing Protocols



Software Setup: Routing Protocols



Agenda

- Initial and Target Situation
- Evaluation Phase
- Hardware
- Software
- **Demo**

Demo Setup (iBGP)



- BGP unnumbered, no OSPF
- Advertise loopback IPs through iBGP

Demo Config (iBGP)

```
int lo
  ip address 203.0.113.1/32
  ipv6 address 2001:db8::1/128

router-id 203.0.113.1

int slp1
  no ipv6 nd suppress-ra
  ipv6 nd ra-interval 10

int slp2
  no ipv6 nd suppress-ra
  ipv6 nd ra-interval 10

router bgp 65001
  no bgp default ipv4-unicast
  bgp bestpath as-path multipath-relax
  bgp bestpath compare-routerid
  neighbor PG-IBGP peer-group
  neighbor PG-IBGP remote-as internal
  neighbor PG-IBGP description iBGP Peer Group
  neighbor PG-IBGP capability extended-nexthop
  neighbor slp1 interface peer-group PG-IBGP
  neighbor slp2 interface peer-group PG-IBGP
```

```
addr ipv4 uni
  network 203.0.113.1/32
  neighbor PG-IBGP activate
  neighbor PG-IBGP next-hop-self
  neighbor PG-IBGP send-community
  neighbor PG-IBGP soft-reconfig inbound

addr ipv6 uni
  network 2001:db8::1/128
  neighbor PG-IBGP activate
  neighbor PG-IBGP next-hop-self
  neighbor PG-IBGP send-community
  neighbor PG-IBGP soft-reconfig inbound
```

Demo

Showtime!

cloudscale.ch

Questions



We are hiring...

We are looking for a
Senior Linux System Engineer

(Ubuntu, Debian, OpenStack, Ceph,
Ansible, Python, ...)

Get in touch: jobs@cloudscale.ch

 cloudscale.ch

Thank you!

I am looking forward to your feedback:
manuel.schweizer@cloudscale.ch

 cloudscale.ch