

Budování sítě v datových centrech

Ing. Pavel Danihelka

Teamleader network administrator



Obsah

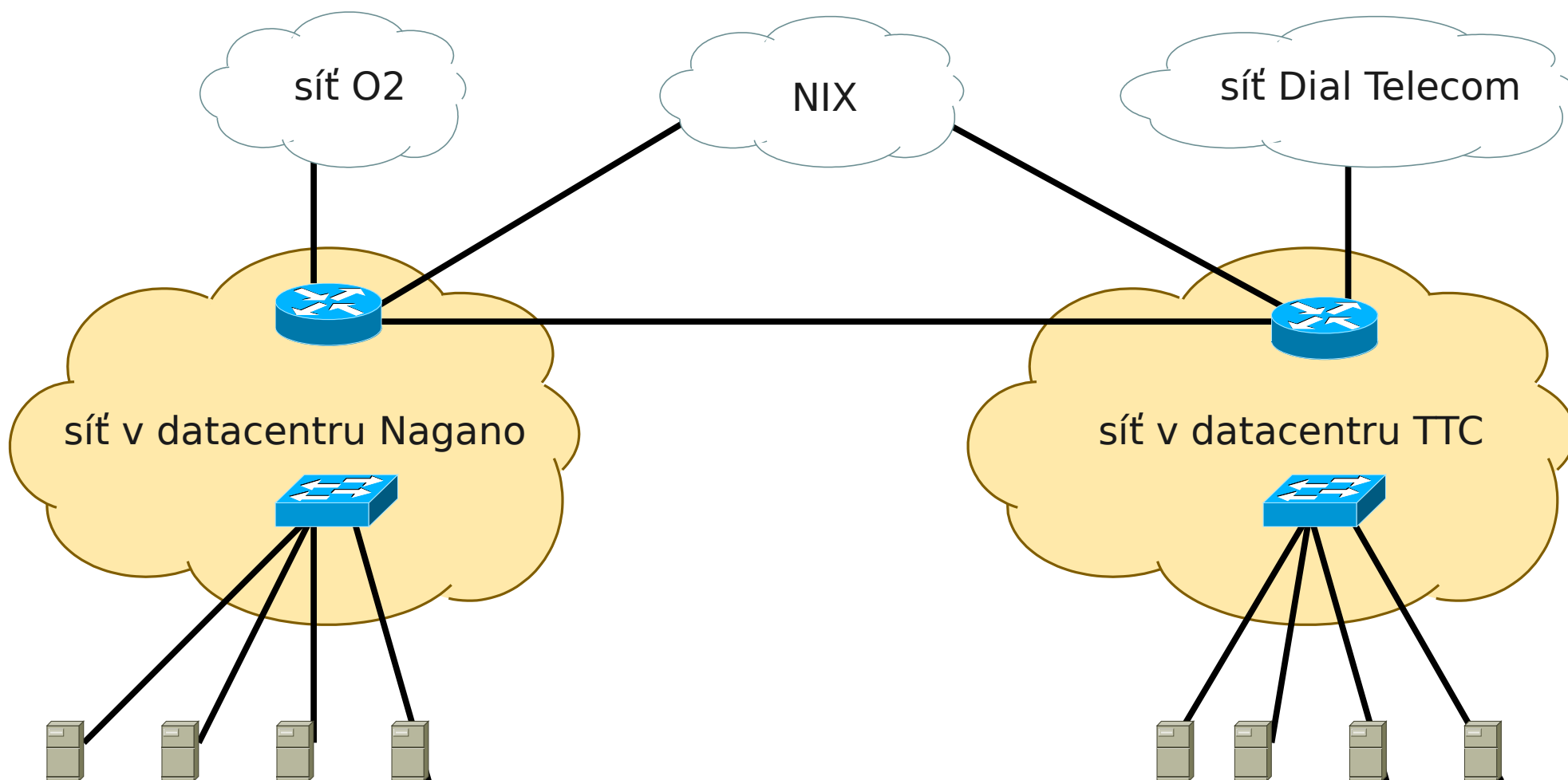
- **Seznam a jeho síť**
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Load balancing
- Návrh architektury sítě

Představení služeb Seznamu

- návštěvnost HP Seznamu přes 2M uživatelů denně (zdroj: Netmonitor)
- řada služeb v top 10 českého internetu
- široké portfolio služeb

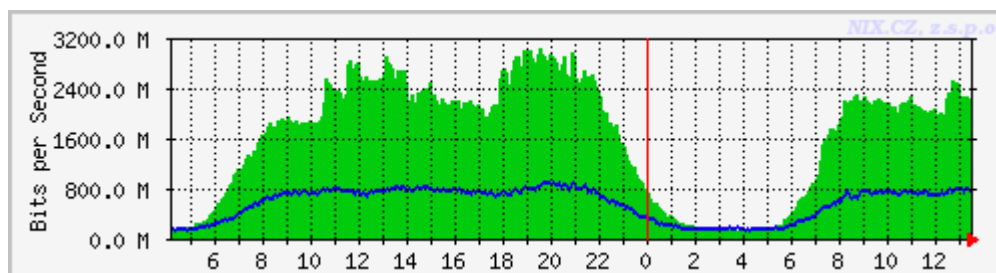
The logo for Seznam, featuring a stylized red 'S' followed by the word 'SEZNAM' in black uppercase letters.The logo for FIRMY.CZ, with 'FIRMY.CZ' in red uppercase letters and 'SEZNAM' in smaller black uppercase letters below it.The logo for LIDÉ, featuring a red circle with two white eyes to the left of the word 'LIDÉ' in blue and green uppercase letters.The logo for SUPER.CZ, with 'SUPER.CZ' in white uppercase letters on a red rectangular background.The logo for EMAIL SEZNAM, with 'EMAIL' in orange uppercase letters and 'SEZNAM' in smaller black uppercase letters below it.The logo for MAPY.CZ, with 'MAPY.CZ' in red uppercase letters.The logo for Novinky.cz, with 'Novinky.cz' in black uppercase letters, where the 'i' in 'Novinky' is red.The logo for Sport.cz, with 'Sport.cz' in black uppercase letters, where the 't' is blue.

Síťová infrastruktura Seznamu

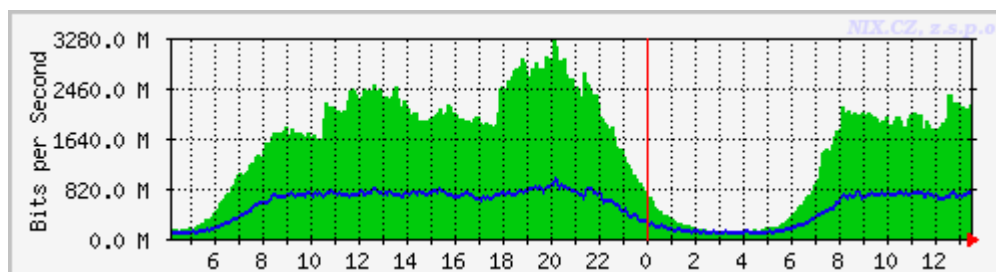


Síťový provoz Seznamu

- 2x 10Gb do NIX, 2x 10Gb tranzit (Dial Telecom a O2)
- linka do NIX2



- linka do NIX4



Požadavky na síť

- vysoká dostupnost
zhruba 99.999% tj. výpadek max 30s za měsíc
- redundance síťových prvků a datových cest
- rychlá konvergence sítě
řádově sekundy
- škálovatelnost
neustálý růst, kolem 1200 serverů
- velká síťová propustnost
odchozí provoz přes 6,5Gb/s

Obsah

- Seznam a jeho síť
- **Hardware**
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Load balancing
- Návrh architektury sítě

Hardware

- přepínače (switch)
- konvertory (transceiver)
- kabeláž
- směrovače (router)
- ukázka jednoduchého zapojení

Přepínače (switch)

- L2 prvek, rack mount
- zapojení koncových serverů
- možnost stohování switchů
- management, SNMP, VLAN



Konvertory (transceiver)

- GBIC, SFP
 - 1 GigE, metalické/optické
- XENPAK, X2, XFP, SFP+
 - 10 GigE, optické



Kabeláž

- metalická kabeláž
kategorie 6 - 1000BASE-T
- optická kabeláž
LC/SC konektory
singlemode/multimode

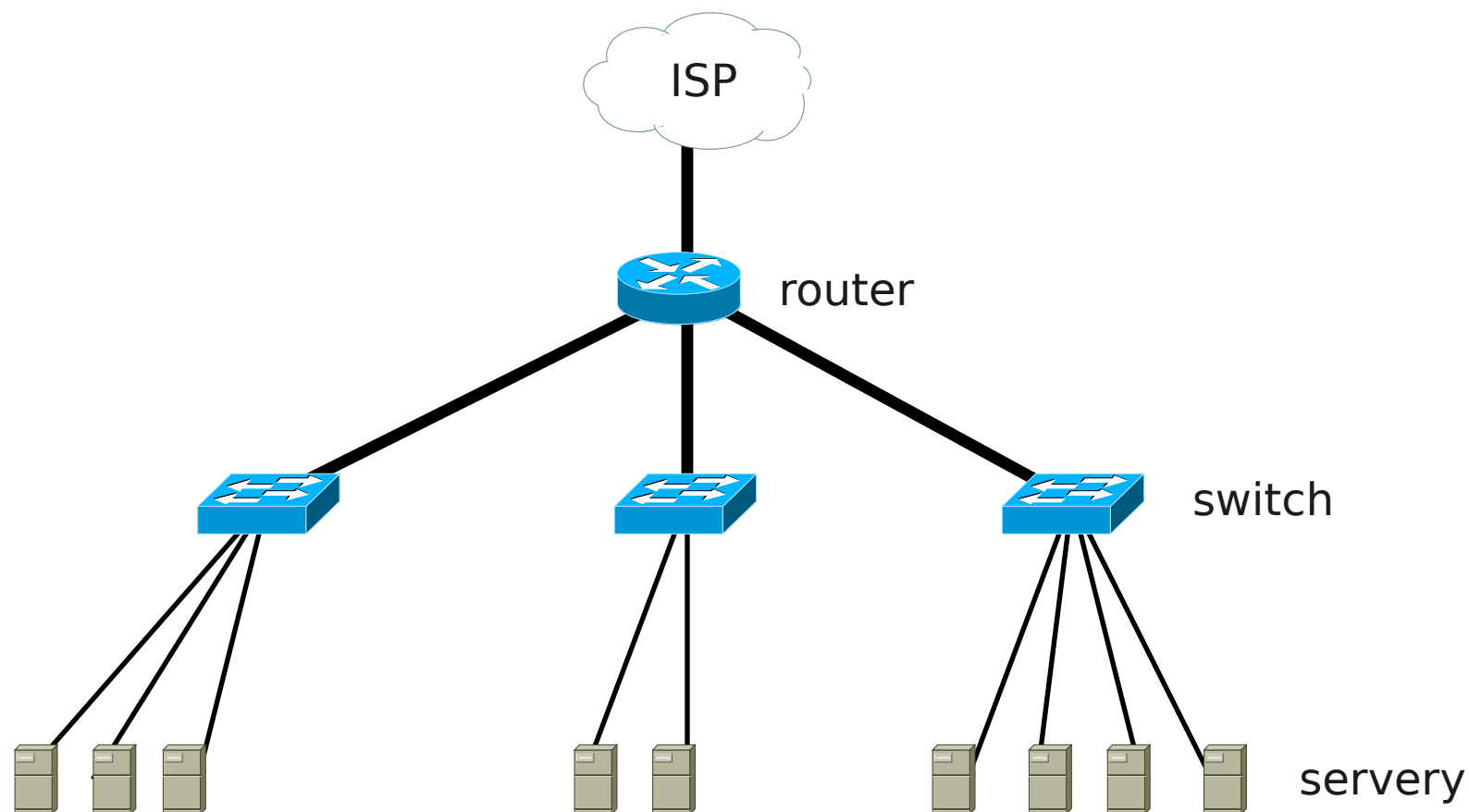


Směrovače (router)

- L3 prvek, redundantní zdroje, modulární design
- centrální bod sítě
- propojuje ke switchům a k providerovi
- směrování paketů



Ukázka jednoduchého zapojení



Obsah

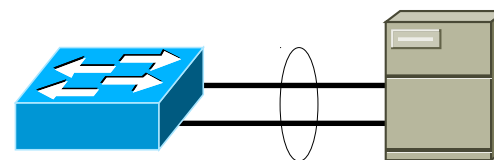
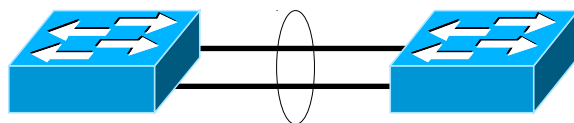
- Seznam a jeho síť
- Hardware
- **Škálovatelnost a propustnost**
- Zajištění vysoké dostupnosti
- Load balancing
- Návrh architektury sítě

Škálovatelnost a propustnost

- 10 GigE po optice (IEEE 802.3ae)
- 1 GigE po optice (IEEE 802.3z) a metalice (IEEE 802.3ab)
- link aggregation (IEEE 802.3ad)
- dynamické směrování
OSPF, IS-IS

Link aggregation

- IEEE 802.3ad, Etherchannel (Cisco), Trunking (Sun), Bonding (Linux)
- L2 protokol, agregace až 8 linek
- rozvažování provozu na základě hashe
dst-ip, dst-mac, dst-port, src-dst-ip, src-dst-mac



OSPF (Open shortest path first)

- RFC 2740
- dynamický směrovací protokol
- používá Dijkstrův algoritmus k výpočtu nejkratší cesty
- používá se pro směrování ve vnitřní síti
- konvergence kolem 1s

OSPF - pokračování



Obsah

- Seznam a jeho síť
- Hardware
- Škálovatelnost a propustnost
- **Zajištění vysoké dostupnosti**
- Load balancing
- Návrh architektury sítě

Zajištění vysoké dostupnosti (High availability)

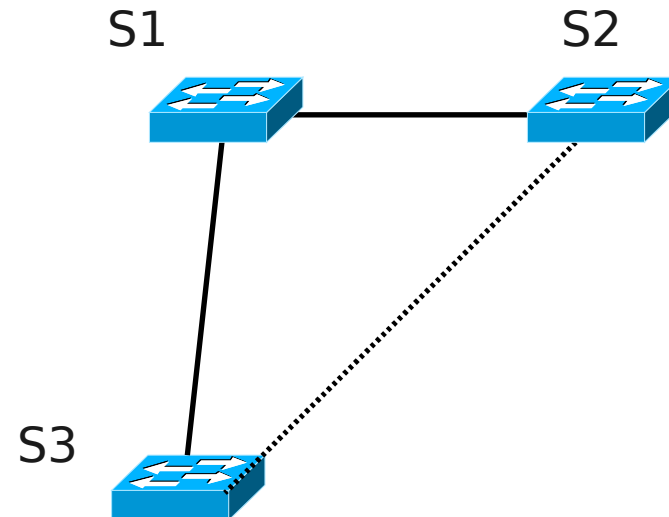
- VRRP
- STP (spanning tree protocol)

VRRP

- VRRP (Virtual router redundancy protocol)
RFC 3768
- IP adresa “sdílená” mezi více boxy
- komunikace přes multicast, priority
- konvergence cca 3s

STP (Spanning tree protocol)

- L2 protokol, BPDU rámce každé 2s
- Rapid-PVST+ (802.1w)
- stromová topologie, redundance datových cest
- root bridge, path cost
- strom nejkratších cest
- konvergence kolem 1s



Obsah

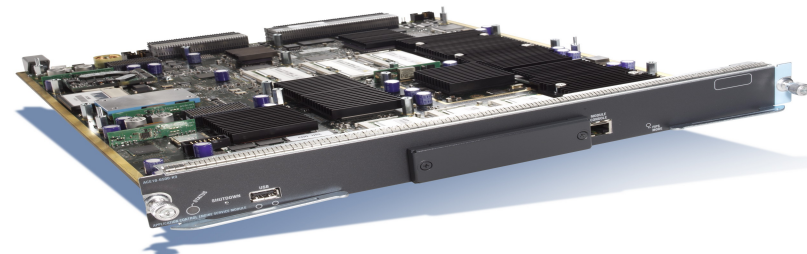
- Seznam a jeho síť
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- **Load balancing**
- Návrh architektury sítě

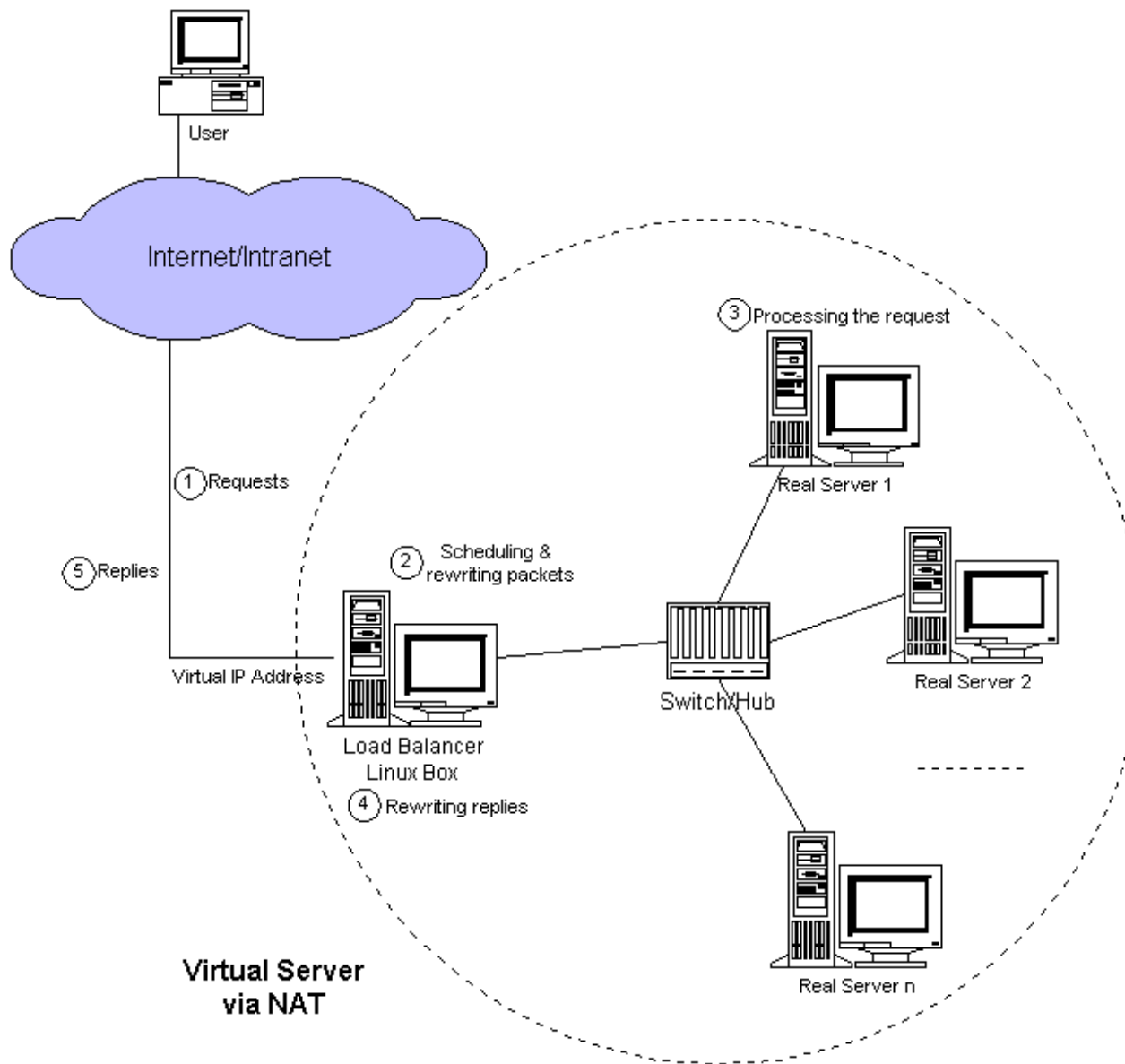
Load balancing

- motivace - rozkládání provozu na více fyzických serverů
- DNS round robin
 - výhody - jednoduché řešení
 - nevýhody - spotřeba IP adres, řešení výpadků strojů

Load balancery

- samostatné HW řešení
Cisco, Citrix, F5, Juniper, Nortel
- integrované HW řešení
Cisco ACE modul, CSM modul
- softwarové řešení
LVS - <http://www.linuxvirtualserver.org>



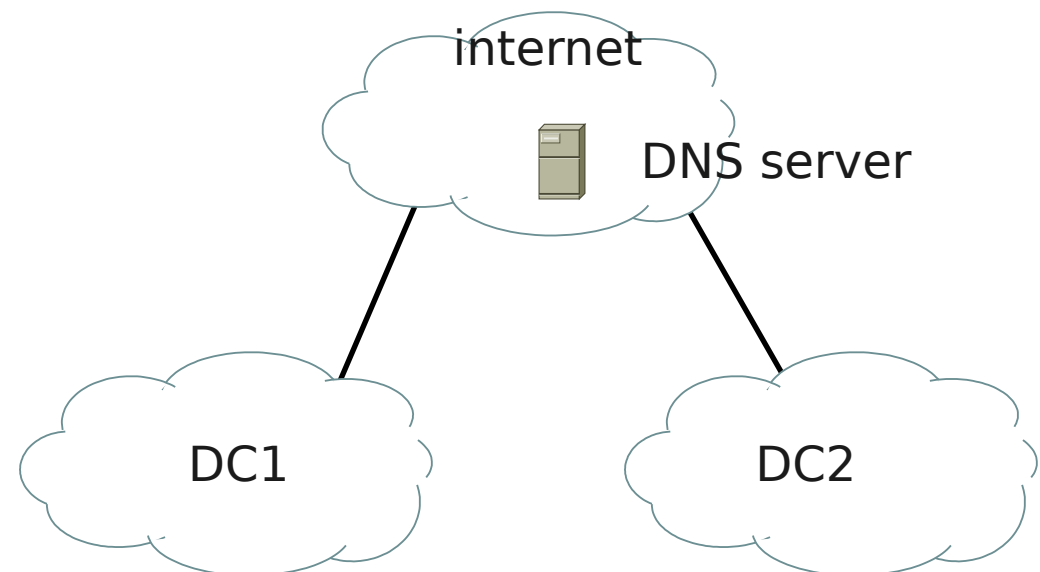


Load balancing - popis

- L4-L7 switching
- kontrola živých služeb
- full proxy, SSL offloading, HTTP caching
- rozkládání - round robin, least connections

Load balancing - více datacenter

- v každém DC jiné IP rozsahy
- funguje podobně jako DNS round robin
- kontrola živých služeb na DNS serveru



```
DNS  
;DC1  
seznam.cz A 77.75.72.3
```

```
;DC2  
seznam.cz A 77.75.76.3
```

Obsah

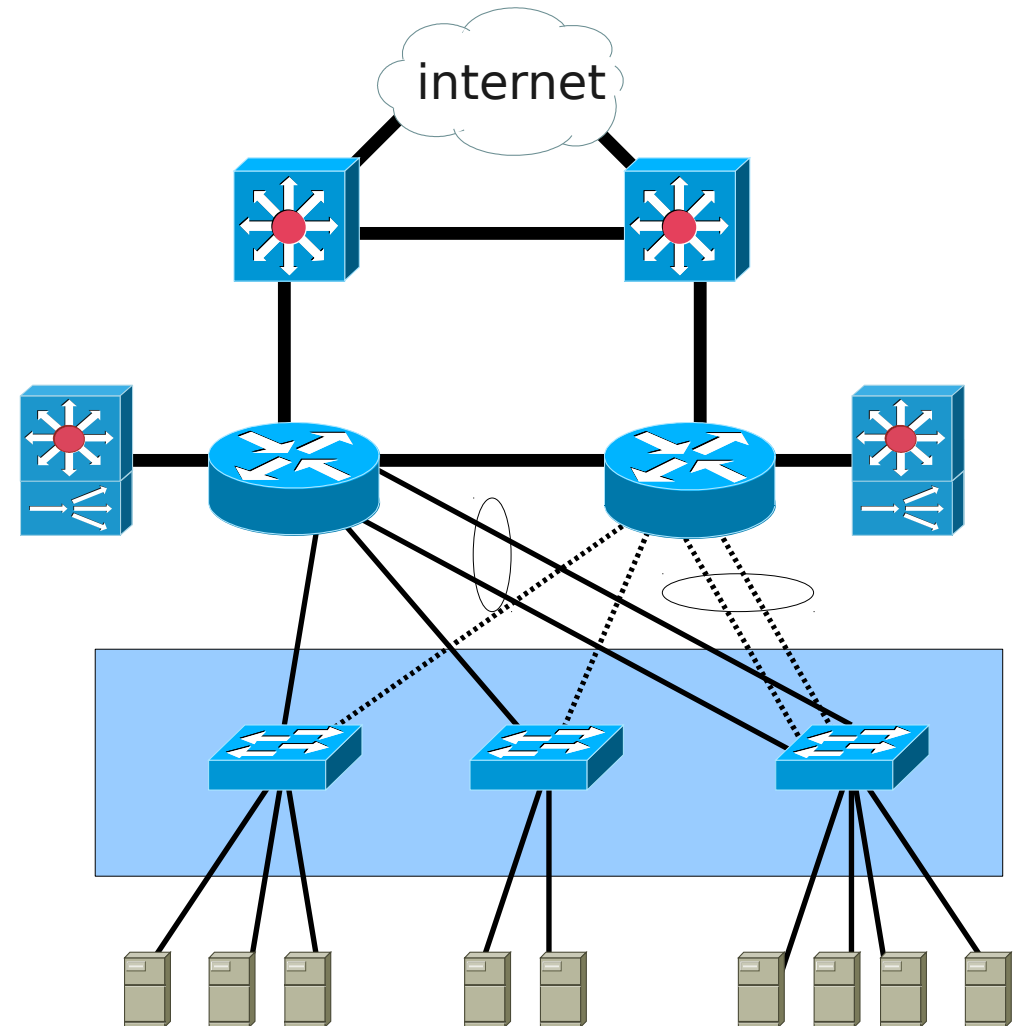
- Seznam a jeho síť
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Load balancing
- **Návrh architektury sítě**

Obecný návrh architektury sítě

- přístupová vrstva
konektivita pro servery, STP
- agregační vrstva
řešení přístupu, HA, load balancing, VRRP
- páteřní vrstva
škálovatelnost, HA, rychlá konvergence, propustnost

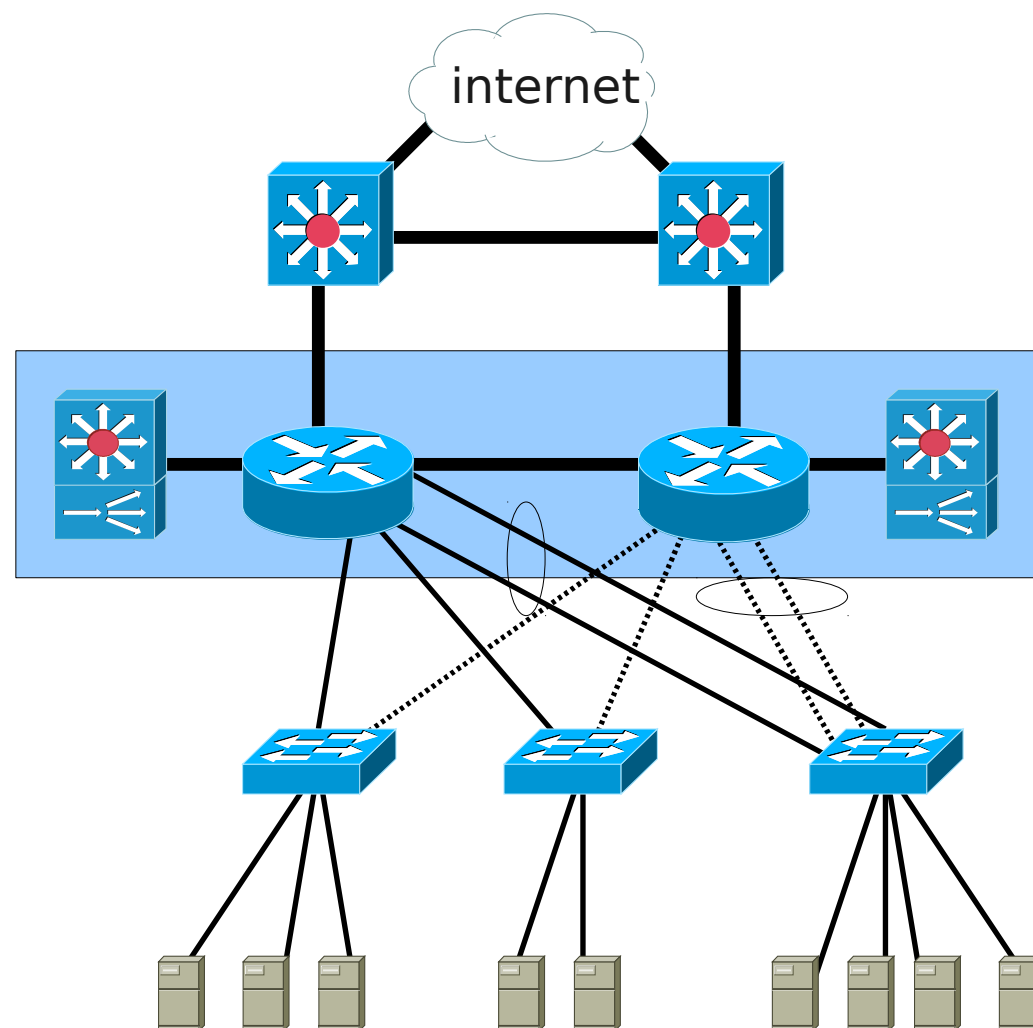
Návrh přístupové vrstvy

- rozdělení sítě do VLAN
- link aggregation propoje
- redundantní cesty



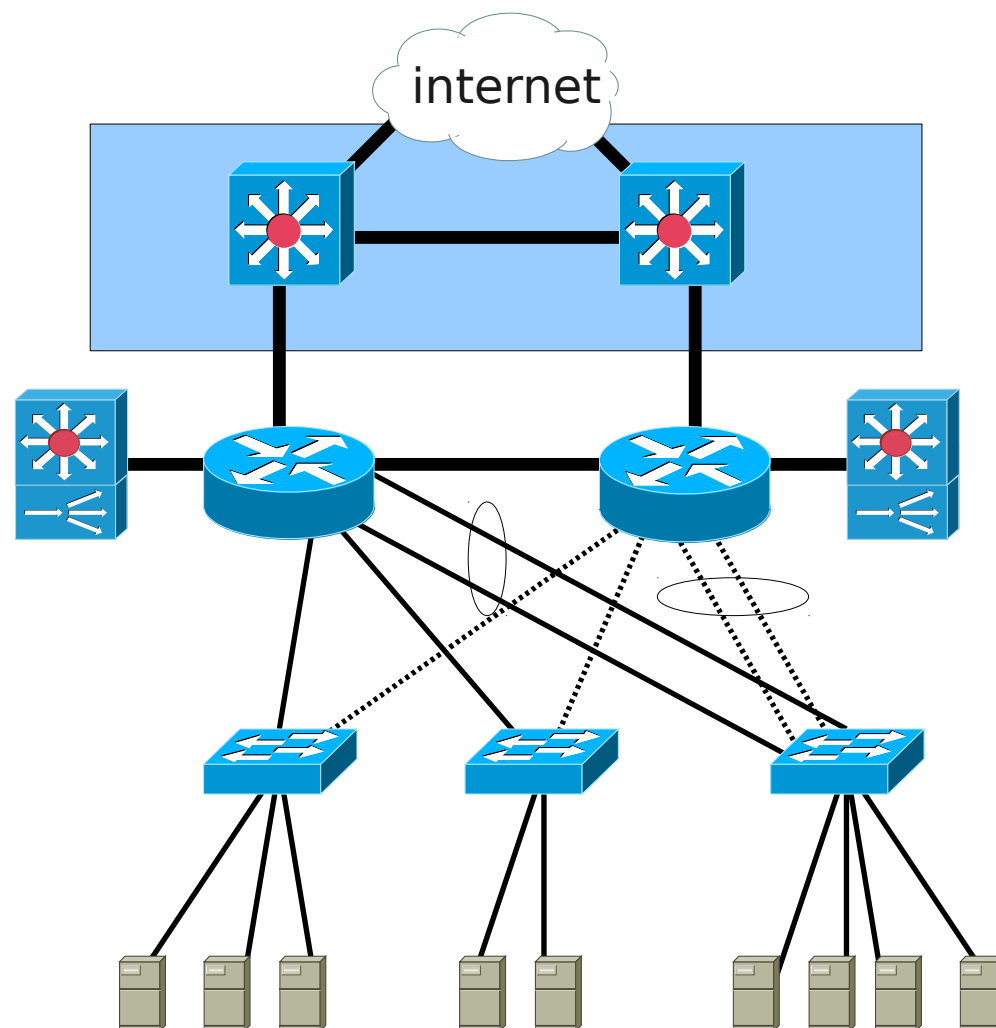
Návrh agregační vrstvy

- VRRP
- L2, Rapid-PVST+
- load balancing, firewall, SSL



Návrh páteřní vrstvy

- ne vždy je potřeba
- 10 GigE
- vše na L3
- směrování přes OSPF



Dotazy?

pavel.danihelka@firma.seznam.cz

Odkazy

- Seznam blog
seznam.sblog.cz
- Vývojáři
vyvojari.seznam.cz