



**From: =?ISO-8859-2?Q?Ing=2E_Michal_Bukovsk=FD?=
<micHAL.bukovsky@firma.seznam.cz>**

To: undisclosed-recipients: ;

Subject: =?ISO-8859-2?Q?P=F8edn=E1=B9ka_o_emailu?='

Obsah

- **Architektura emailu**
 - **Příjem a odesílání pošty**
 - **Storage dat**
- **Ochrana proti Spamů**
- **Ochrana proti Virům**
- **Technologie**



Nějaká čísla na úvod...

Podle počtu zpráv

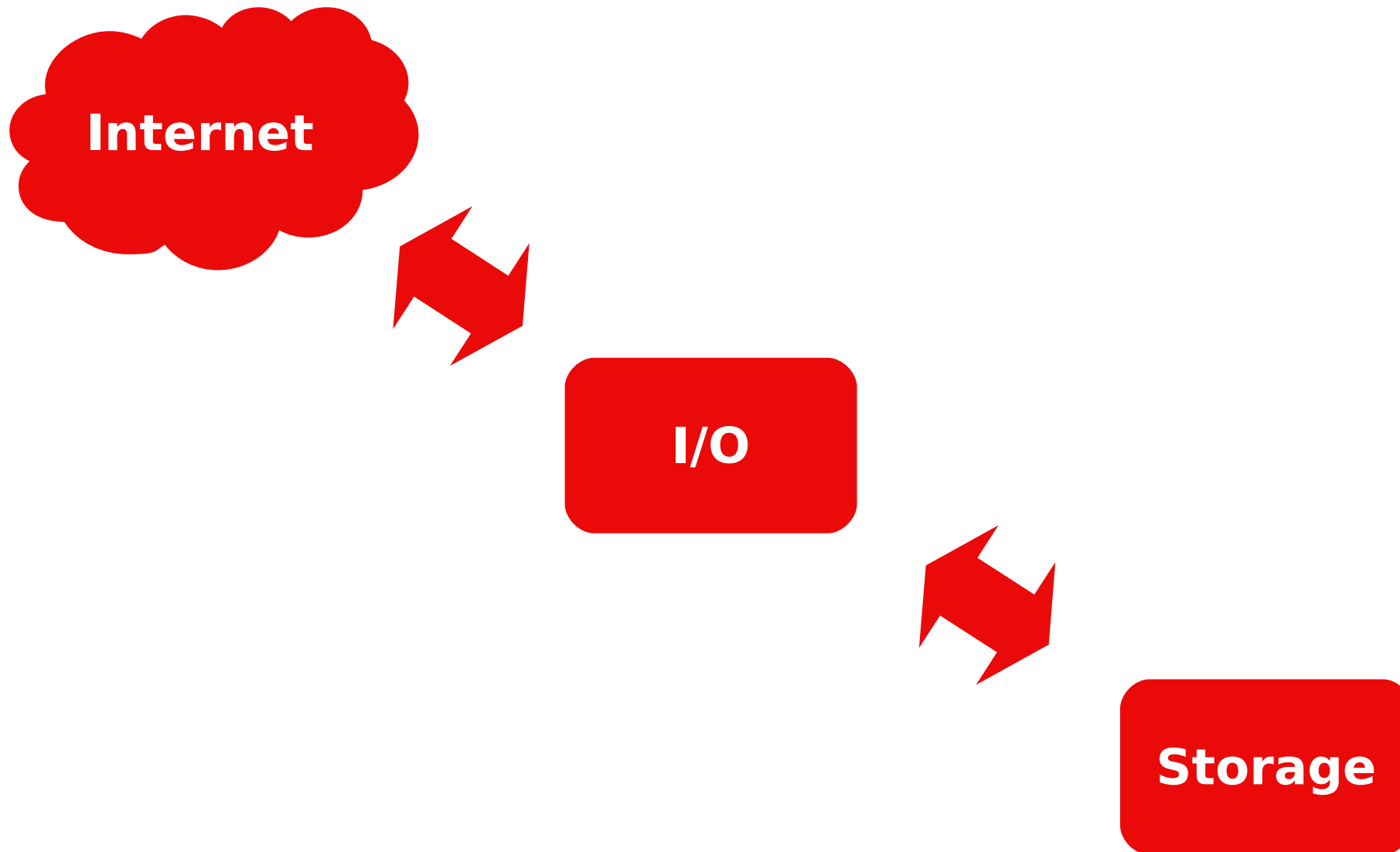
- **Přichází 85 000 con/min**
- **Přijímáme 15 000 msg/min**
- **Ukládáme 2 000 msg/min**
- **Odesíláme 2 500 msg/min**
- **Generujeme 300 bounce/min**
- **A to vše pro více jak 8 000 000 uživatelů**

Nějaká čísla na úvod...

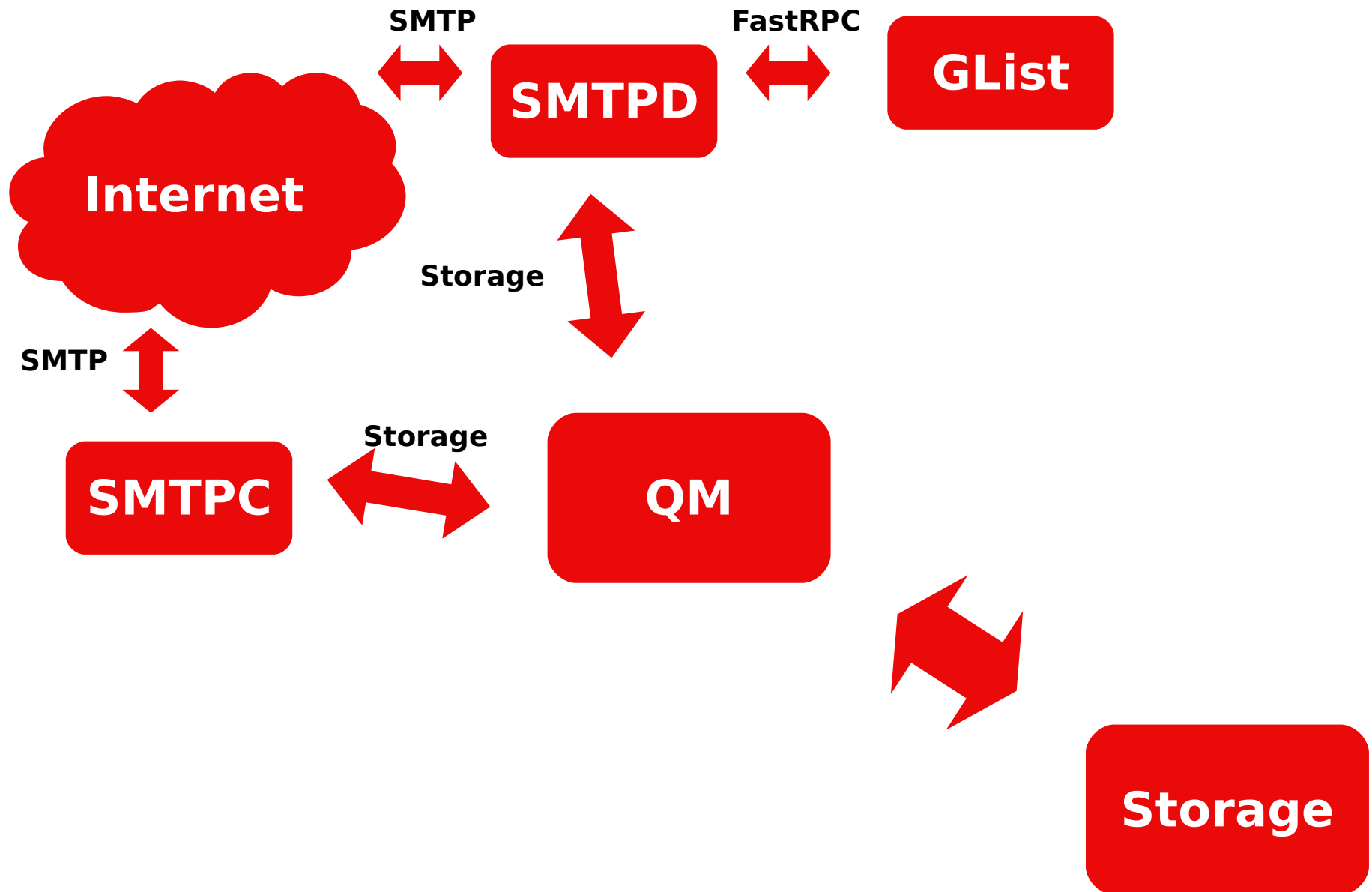
Z hlediska velikosti zpráv a hardware

- **Přes 3,2 peta byte**
- **Průměrná velikost zprávy je cca 400kB**
- **Více jak 1 řada racků v každé serverovně**
- **Což skoro 500 serverů** (více jak $\frac{1}{4}$ serverů seznamu)
- **Příkon přes 250 kWh** (spotřeba 1250 prům. domácností)
- **350 RO/s a 150 RW/s**

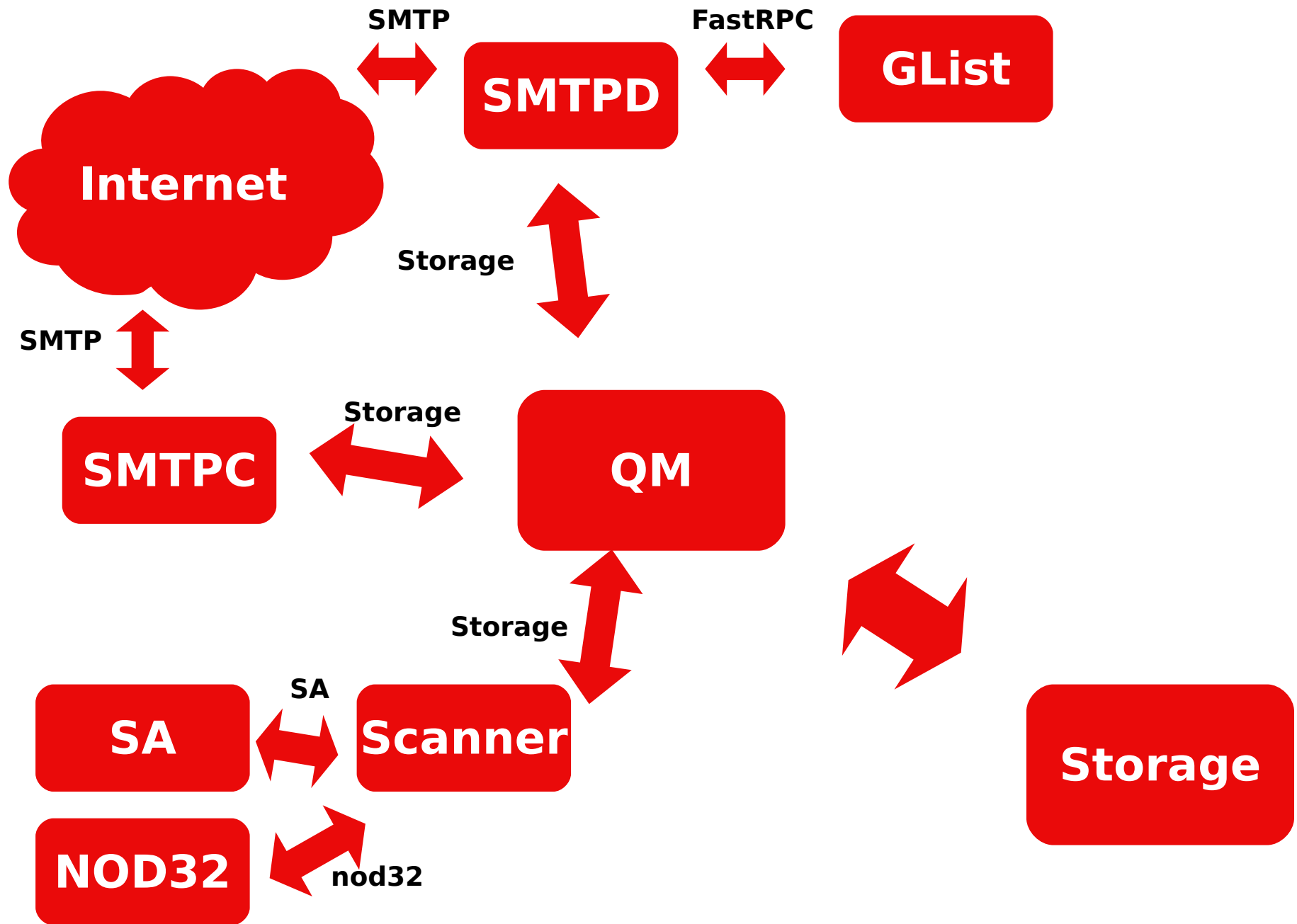
Architektura



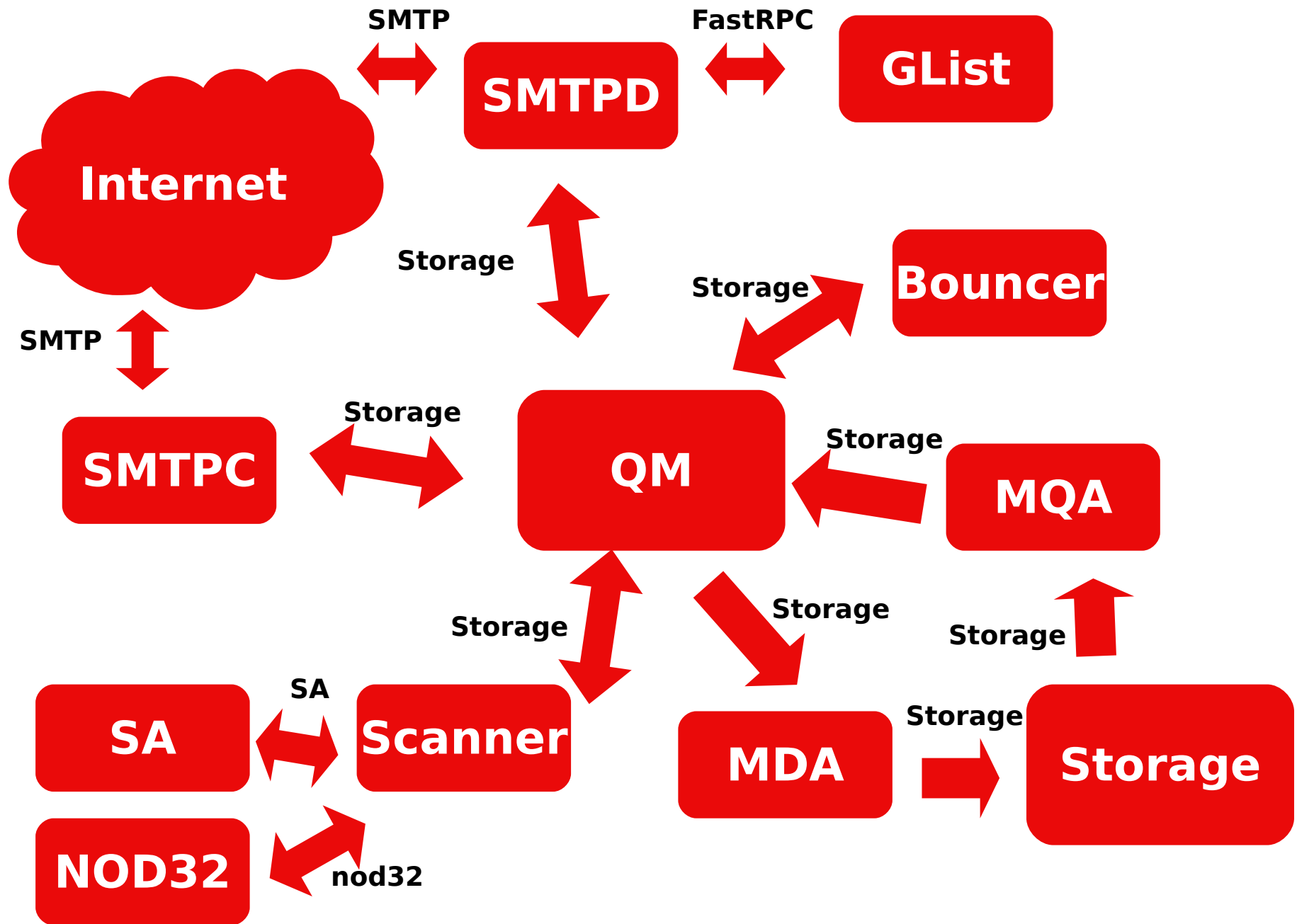
Architektura - I/O



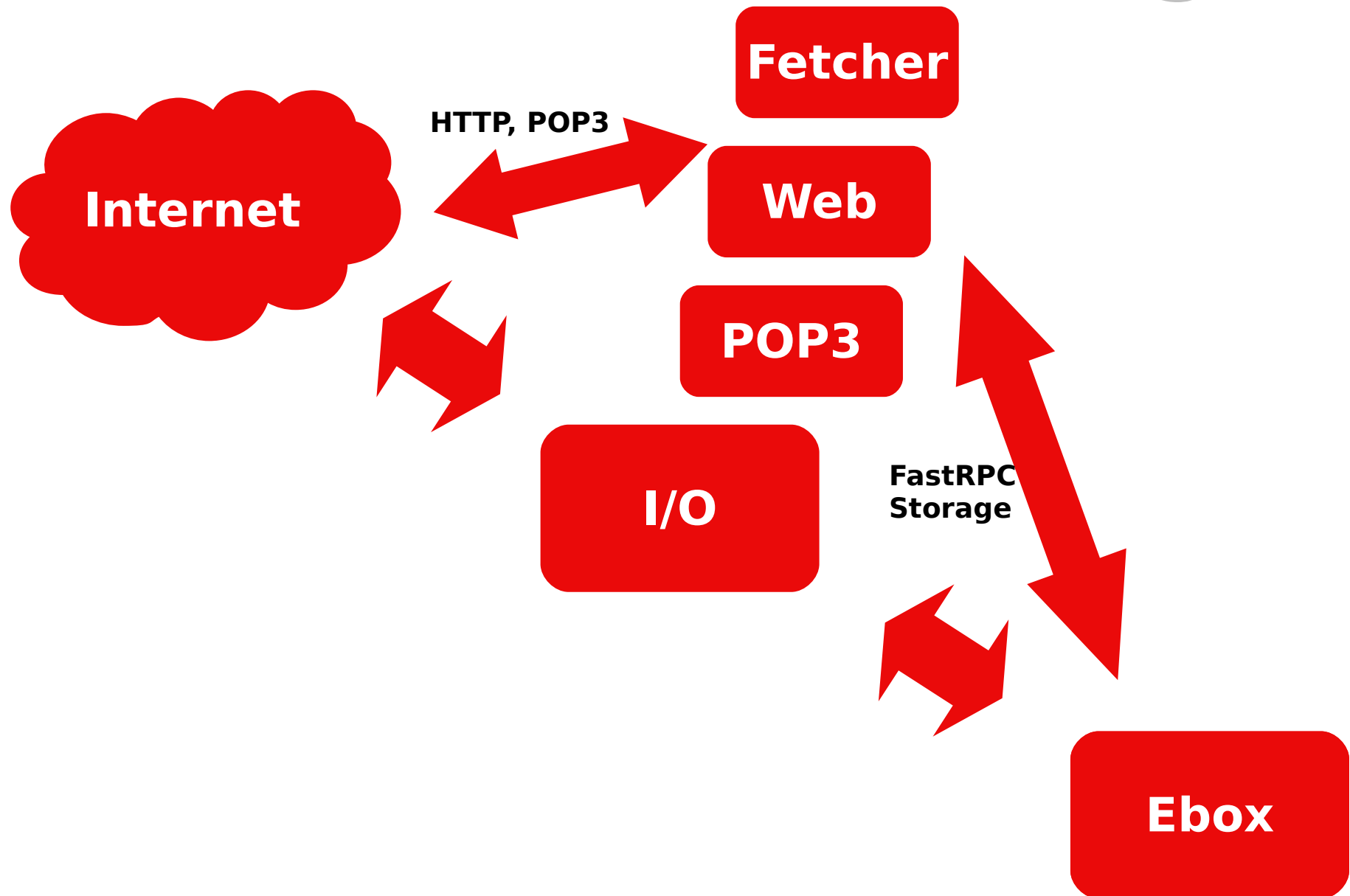
Architektura - I/O



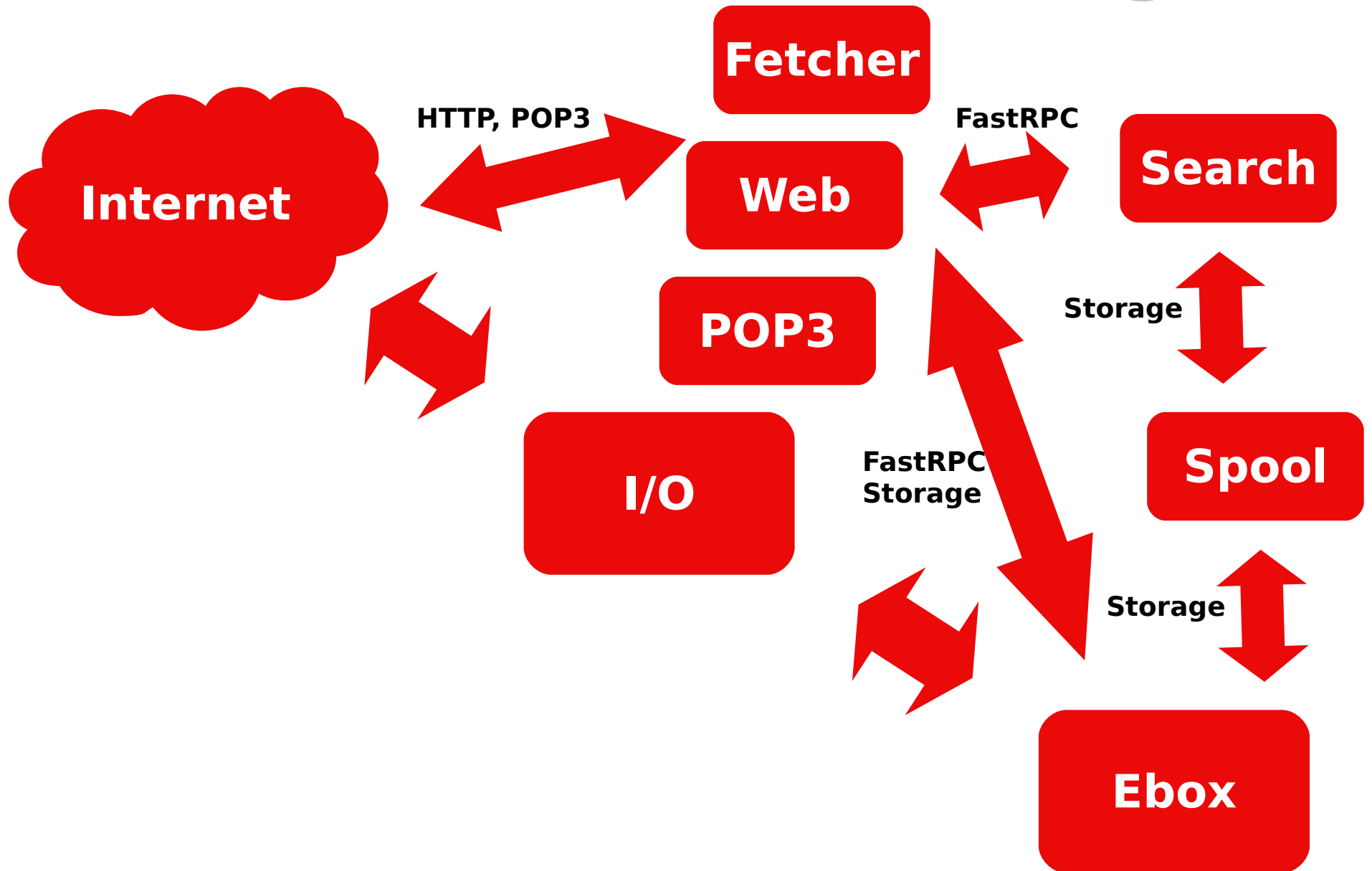
Architektura - I/O



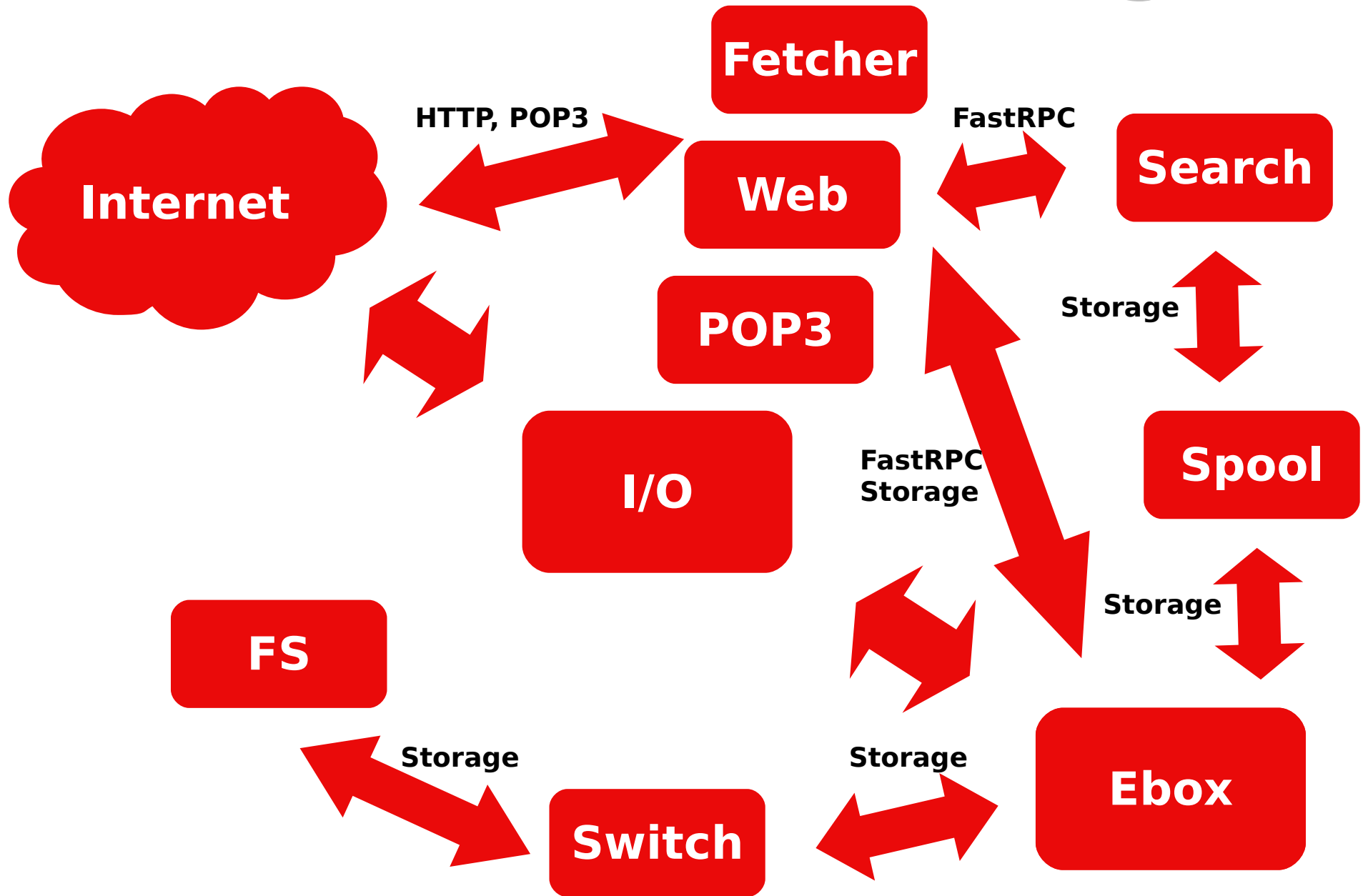
Architektura - Storage



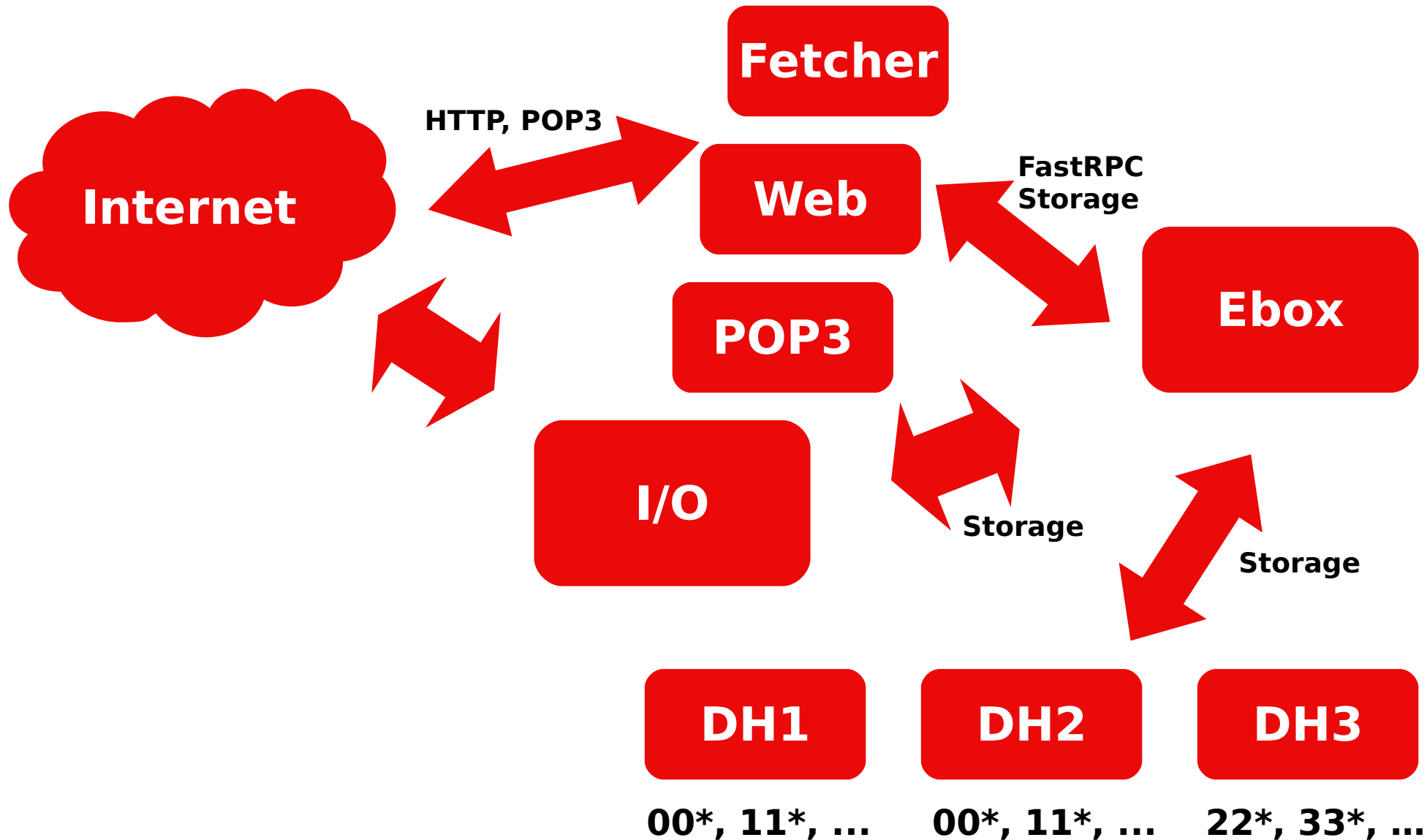
Architektura - Storage



Architektura - Storage



Architektura - Deduplikace



Princip hashování:

EMAIL → 0011223344556677 → DH1

Spam - SMTPD

- **Velký traffic nemusí být vždy nevýhoda**
- **GeoIP - IP rasismus**
- **Autoblocker - automatický IP/User blacklist**
- **SPF - csas.cz**
- **GrayList - 80% úspěšnost**
- **Blacklisty a Whitelisty**
- **Delivery test**

Spam - Scanner

- **Domain Keys**
- **SpamAssassin**
- **Bayes - SA, Bogofilter**
- **IP/URL DNS Blacklisty**
- **DCC Blacklisty**
- **Sbírání URL**



Spam - SMTPC

- Nejen spameři jsou na blacklistech
- CAPTCHA na webmailu
- Více odchozích IP adres

Viry

- **Nod32 - scanner**
- **Aktualizace - opět velký traffic výhodou**
- **V dnešní době menší problém než spam**
- **<http://www.virovyradar.cz>**



Technologie

- **C++ a STL, Python**
- **MySQL**
- **Proprietární databáze metadat**
- **FastRPC, HTTPStorage**
- **MetaServer**
- **Memcache**
- **Apache, lighttpd, mod_python, Teng**

Odkazy a otázky

- <http://vyvojari.seznam.cz>
- <http://seznam.sblog.cz>
- <http://opensource.seznam.cz>



<micHAL.bukovsky@firma.seznam.cz>