Monitoring DNS

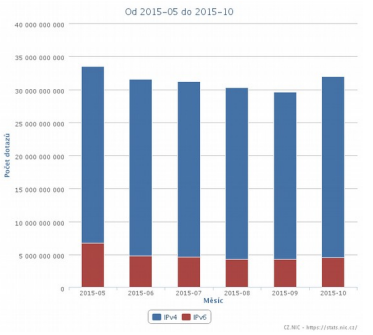
Pomocí BigData nástrojů

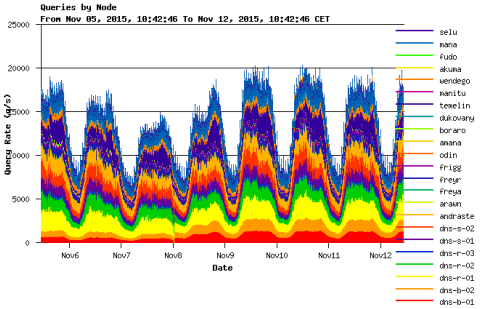
Petr Černohouz • petr.cernohouz@nic.cz • 13. 11. 2015

CZ.NIC | SPRÁVCE DOMÉNY CZ

Současný stav

- Více systémů pro sběr a vyhodnocení DNS dat
- Hrubá data například v IPv6 statistikách
  - Souhrnná data za celý měsíc
- Podrobná data v DSC
  - Sumarizace za daný časový úsek
  - Neumožňuje podrobnější analýzu dat





Několik čísel – průměrný den

- 1 000 000 000 dotazů
- 75 GB komprimovaných PCAP souborů
- Očištěná data exportovaná do CSV - 75 GB
- Uložení do PostgreSQL – 75 GB
  - Bez indexů
  - Indexy zabírají stejně místa jako data

Proč nestačí PCAP

- Ke potřeba celý soubor rozbalit a zpracovat
- Souborů je mnoho a/nebo jsou velké
- Zbytečně mnoho informací
- Nakonec stejně data „někam“ odsypeme
- PostgreSQL je jen dočasné řešení
  - Roční neindexovaná data – cca 27 TB

Čas pro BigData

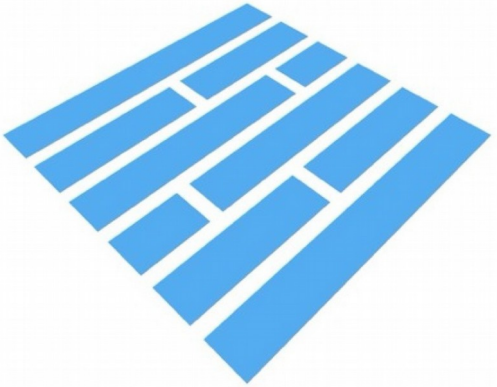
- Přerostli jsme konvenční řešení
- Potřeba získat výstupy v reálném čase
- Dobrá škálovatelnost
- Snadné napojování dalších komponent
- Můžeme využít i v dalších projektech

Zvolené technologie

- Apache™ Hadoop®
  - Snadné přidávání dalších uzlů
  - Kompletní ekosystém nástrojů
  - OpenSource
- Impala
  - SQL dotazování nad daty v hadoopu
  - Podpora mnoha formátů pro data

Apache Parquet

- Umožňuje ukládat data po sloupcích
- V rámci sloupce provádí kompresi dat
- U DNS dat je úspora místa kolem 80%
- Teoreticky „jen“ 5 TB dat za rok
  - 60 % dotazů je na QTYPE A
  - Přes 95 % odpovědí má RCODE NOERROR



Co s daty?

- Brzká detekce útoků
- Vizualizace patternů v tocích
- Odhalení problémových resolverů
- Přidání některých výstupů do stats.nic.cz
- Uvolnění datasetů jako OpenData

Plány do budoucna

- Vlastní sběrač pro serverová data
- Efektivní předávání nasbíraných dat do hadoopu
  - Apache Kafka ?
- UI pro práci s daty
- Spolupráce s bezpečnostním týmem
- Stále hledáme nejvhodnější nástroj

Dotazy?

Děkuji za pozornost

Petr Černohouz • petr.cernohouz@nic.cz

CZ.NIC | SPRÁVCE DOMÉNY CZ