Největší uložiště malwaru (*minimálně*) v Evropě

Hitachi Content Platform v Gen Digital

11. 4. 2024

Rudolf Plíva

Kdo je Gen Digital?

Kdo je Gen Digital?

Gen je globální společnost působící ve více než 150 zemích světa, která podporuje digitální svobodu prostřednictvím značek, jako jsou Norton, Avast, LifeLock, Avira, AVG, ReputationDefender a CCleaner.

Přinášíme špičková technologická řešení v oblasti kybernetické bezpečnosti, soukromí a ochrany identity více než 500 milionům uživatelů, aby mohli svůj digitální život žít bezpečně, soukromě a sebevědomě dnes i po další generace.

Gen™

norton™ Avast™ LifeLock™ by norton Avira™

AVG™ REPUTATION DEFENDER™ by norton CCleaner™

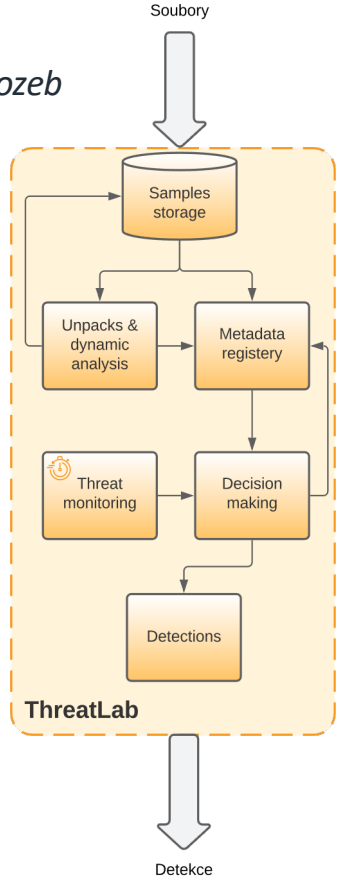
Jak funguje ThreatLab

ThreatLab = oddělení zabývající se vyhledáváním, analýzou a následnou detekcí různých druhů hrozeb

- Do ThreatLabu proudí denně velké množství souborů
  - *Kolem 4 - 5 mil souborů, z toho ~2 mil nových*
- Všechny musíme uložit a co nejdříve je zpřístupnit dalším systémům pro analýzu a další zpracování
- Průběžné vyhodnocování nasbíraných dat a jejich klasifikace
- V případě malwaru se vytvoří detekce a distribuuje se uživatelům

- Kromě nových souborů probíhá monitoring chování detekcí u uživatelů a řešení "*false positives*"

- Se soubory pracujeme jako s objekty - vždy zapíšeme a čteme celý soubor
- Soubory jsou pro nás neměnné, jejich identifikátorem je SHA256 hash



Požadavky

- Stabilita a výkonnost uložiště je pro nás klíčová
- Při jeho nedostupnosti se prakticky zastaví většina procesů v Threatlabu - nedostupný soubor = není co analyzovat
- V tu chvíli začíná relativně rychle klesat ochrana našich uživatelů a nejsme schopni například řešit ani hlášené "*false positives*"
- Dlouhodobější nedostupnost může mít vliv na důvěryhodnost firmy a ve výsledku i na tržby a akcie

Od Samby k objektovému uložišti

2008

Dva **Windows servery s diskovým polem**, sdílení přes SAMBA, metadata a umístění v databázi

2015

Šestnáct Windows serverů a pokukujeme po jiném řešení

2015

Zkoušíme CEPH, GlusterFS a zvažujeme i vlastní řešení

2016

Zvolena NAS technologie

Kapacitně dostatečné, ale výkonnostně nezvládá

2017

Výběrové řízení na novou objektovou storage - vlastní řešení postavené na CEPH nebo HCP.

Nedostatek interních zdrojů + naše specializace nejsou datová uložiště => HCP

2018

Nasazení HCP ve dvou datacentrech, odlišná konfigurace, ale totožná kapacita - 1.2PB v každém datacentru

2024

Rovnocenné konfigurace, kapacita každého datacentra 5.4PB

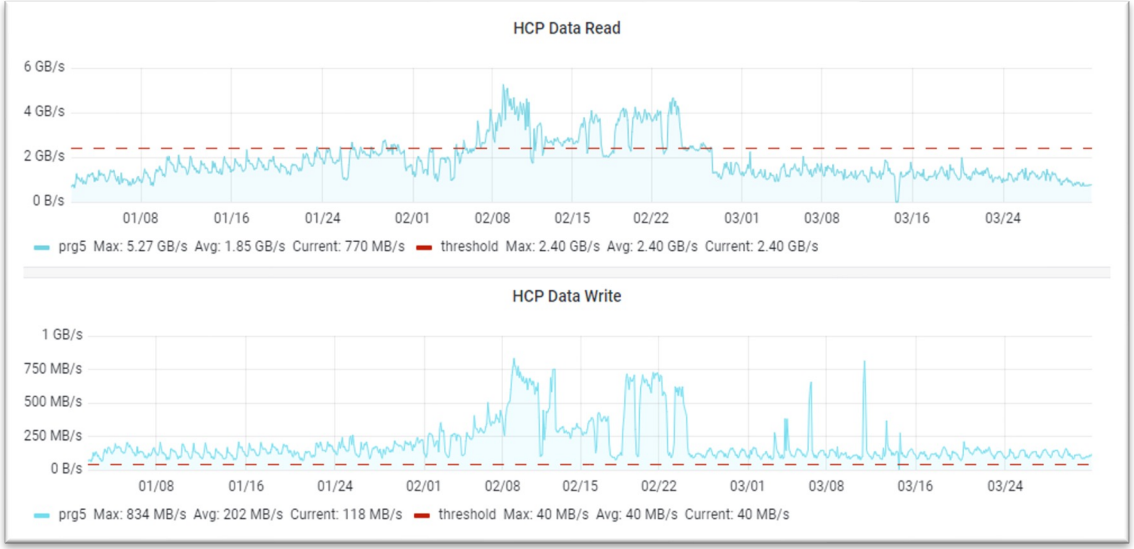
1990 — 2025

Nahlédnutí pod pokličku našeho uložiště

- Dvě rovnocenná datacentra – Praha a Brno – v režimu Active-Pasive z důvodu asynchronní replikace
- Největší namespace - vzorky pro Threatlab o velikosti ~4.36PB a s ~5 miliardami objektů

- Rychlosti
  - Čtení: MAX 5.31 GB/s, AVG 1.85 GB/s
  - Zápis: MAX 892 MB/s, AVG 202 MB/s

- Průměrné denní počty requestů
  - PUT ~5M
  - GET ~44M
  - HEAD ~42M
  - DELETE ~300k



Statistiky za posledních 90 dní (1.1. - 31.3.2024)

Co nám to přineslo

- Jednoduché API pro práci s objekty (+ podpora S3 API)
- Objekt, atributy, metadata - vše na jednom místě
- Vysoká dostupnost
- Dostatečný výkon
- Konzistence dat
- Snadné škálování
- Automatické rebalancování a rovnoměrné vytěžování všech storage nodů
- Asynchronní replikace dat
- Možnost pravidel pro automatické odstraňování objektů
- ...

Co nám ne úplně vyhovuje

- **Asynchronní replikace**
  - Metadata jsou v druhé lokalitě během pár minut, soubory o něco déle
  - V naší event-driven architektuře to znamená, že můžeme používat jen jedno aktivní datacentrum (= nevyužíváme na maximum, co máme)
  - Zvažovali jsme návrat k replikaci na straně aplikace, ale zatím přiliš komplikované :)

- **Řízení přístupu k objektům není možné na základě Active Directory**
  - Přístupy přes AD je možné využít pro webový portál HCP, ale ne pro (H)S3 API z důvodu omezení používaného S3 protokolu
  - Každá služba má svůj lokální účet na HCP - pro služby použitelné, pro uživatele už ne
  - Vyřešeno anonymním přístupem jen pro čtení, nově uživatelé směrováni do cloudu

Další vývoj

- Aktuálně přecházíme do cloudu
- Nové objektové uložiště poběží v hybridním režimu
  - Soubory z posledního roku uloženy na HCP
    - *Cache pro on-prem systémy*
  - Vše uloženo v cloudu

GenDigital.com

