

Hromadný sběr anonymizovaných dat pomocí Collectoru

Miroslav Slugeň - Deployment Manager
CSNOG, 11.6.2018

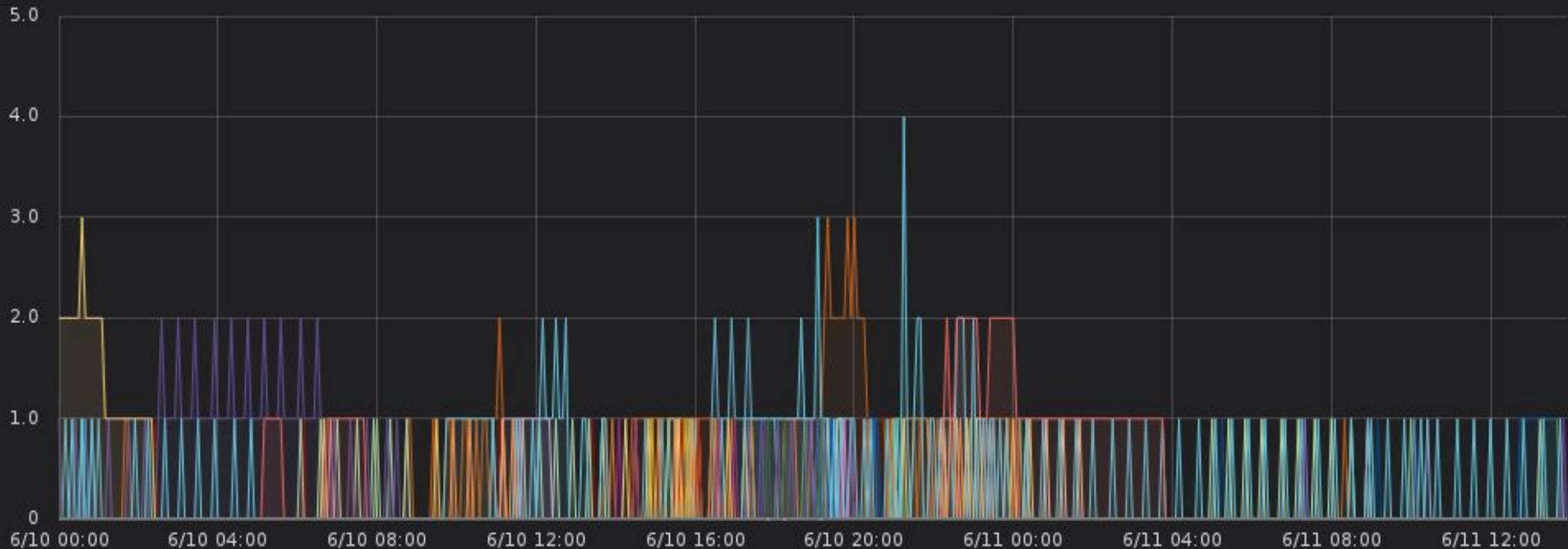
Co to je Collector

- Projekt zajišťující jednotný sběr a zpracování dat z různých zdrojů
- Sjednocení hlášení chyb a debug na různých aplikacích
- Hlášení statistických dat, na co se zákazníci dívají, verze aplikace, verze zařízení, úroveň signálu, rychlost internetu, teplota procesoru, atd...
- Rozdělení do různých FLOW (podle postupu kroků)
- Kompletně vlastní návrh
- Podporuje google analytics
- Určeno pro obrovské množství hlášených zpráv

Jak funguje

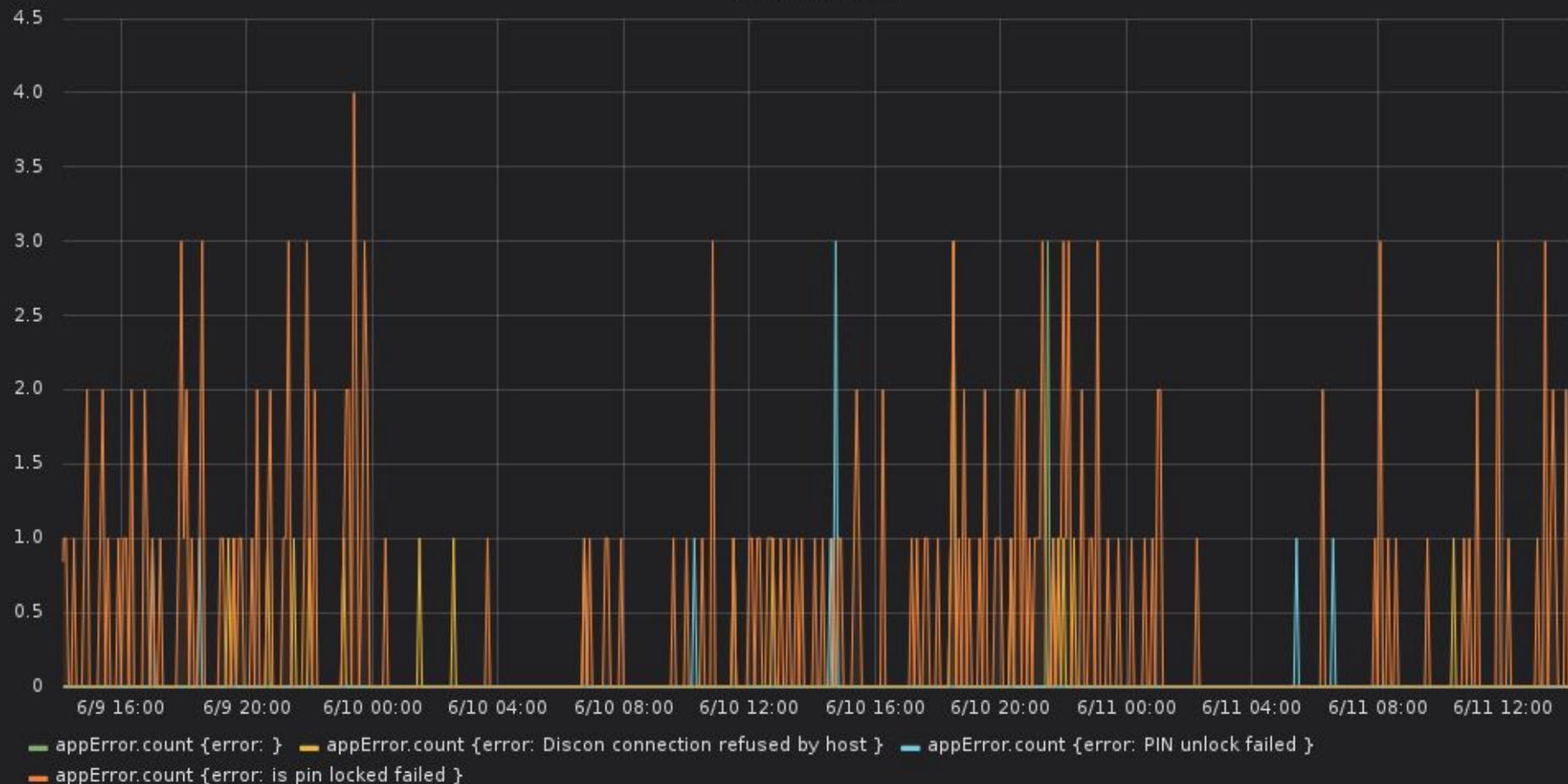
- Klient generuje data a spojuje je do zpráv
- Klient generuje šifrovanou zprávu pomocí jeho klíče
- Klient odešle zprávu do collectoru
- Collector přijme zprávu a začne její asynchronní zpracování
- Collector ověří autorizaci a poté začne zpracovávat data pomocí příslušných pravidel každé FLOW
- Collector smaže zprávu, která je již zpracovaná

Number of people watching by channel



- appWatch.count {content: DorcelHD}
- appWatch.count {content: DorcelTVHD}
- appWatch.count {content: Eurosport}
- appWatch.count {content: HBO}
- appWatch.count {content: HBO2}
- appWatch.count {content: Nickelodeon}
- appWatch.count {content: Seznam}
- appWatch.count {content: Sport2}
- appWatch.count {content: amc}
- appWatch.count {content: barrandov}
- appWatch.count {content: brazzers}
- appWatch.count {content: ct1}
- appWatch.count {content: ct2}
- appWatch.count {content: ct24}
- appWatch.count {content: ct4sport}
- appWatch.count {content: ctart}
- appWatch.count {content: ctdecko}
- appWatch.count {content: eroxHD}
- appWatch.count {content: eroxxxHD}
- appWatch.count {content: fanda}
- appWatch.count {content: filmpro}
- appWatch.count {content: hustler}
- appWatch.count {content: kinobarrandov}
- appWatch.count {content: kinosvet}
- appWatch.count {content: leo}
- appWatch.count {content: markizaint}
- appWatch.count {content: nova}
- appWatch.count {content: novacinema}
- appWatch.count {content: orko}
- appWatch.count {content: orf1}
- appWatch.count {content: prima krimi}
- appWatch.count {content: prima max}

Errors over time



Ukázka zprávy - MESSAGE

```
<message> = {  
  "type": 0 = for log messages, 1 = for join messages,  
  "source": {  
    "app": application_name,  
    "client": client_id,  
  },  
  "index": unique_message_id,  
  "data": [  
    {  
      "flow": <flow_id>,  
      "time": time_of_the_event,  
      "data": {  
        key1: value1,  
        key2: value2,  
        ...  
      }  
    }  
  ]  
}
```

APP INFO

- Informace které se nemění
- Posílají se vždy po startu aplikace jen jednou
- Verze aplikace
- Jaké je to zařízení
- Verze firmware
- Verze aplikace
- Hlavně pro Android Mobile

APP ERROR

- Informace o chybách na zařízení
- FLOW DEVICE - unikátní ID zařízení
- Popis chyby
- Celý call stack
- Úroveň chyby

APP WATCH

- Vytvoří se session po startu aplikace s unikátním identifikátorem
- Session sjednocuje zprávy
- Posílají se informace o programu, kanálu, akce (play,pause,seek)
- Pozice seeku, informace o tom jestli jde o reklamu, pořad
- Slouží hlavně pro recommendation engine

STREAM ERROR

- Všechny chyby související se streamem
- Výpadek
- Zpožděné dodání
- Poškozená data
- Restart streamu
- Pozice streamu

STORES

- Typy úložišť
- Influx DB - hlavní databáze
- Redis - hlavně pro spojená data, dočasná data
- Lze použít jakékoliv úložiště včetně souboru
- Všechny stores jsou konfigurovány dynamicky

Podpora pro různé výrazy a podmínky

- * foo - ((3 + 5) * (10 / bar))
- * foo < 43
- * foo > 41
- * foo <= 42
- * foo >= 42
- * foo == 42
- * foo != 43
- * foo < 41 AND bar < 70 OR lorem < 89
- * foo > 41 AND foo < 43 OR lorem > ipsum
- * (foo < 41 OR bar > 68) AND (lorem * 2 < ipsum)
- * dolor LIKE '%bar'
- * dolor LIKE 'foo%'
- * dolor LIKE '%ob%'
- * foo IN (1, 2, 3, 42, 58)
- * dolor IN ('foo', 'bar', 'fooba', 'foobar')

Co připravujeme

- Matematické operace na daty
- Kubernetes (orchestrátor dockerů)
- Hlášení pro partnery
- Doporučování pořadů a VOD pomocí sesbíraných dat
- Schopnost systému přesně určit sledovanost jednotlivých reklam/přadů
- Propojení s Grafanou
- Přesná data o použití timeshiftu
- Propojení hlášení chyb s podporou

Děkuji za pozornost!

Miroslav Slugeň - Deployment Manager