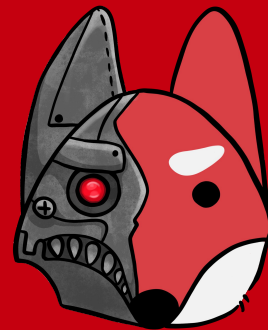


Elasticsearch jako primární databáze

Jakub Kluvánek
jakub.klivanek@biddingfox.com
@kluvi, @biddingfox_devs

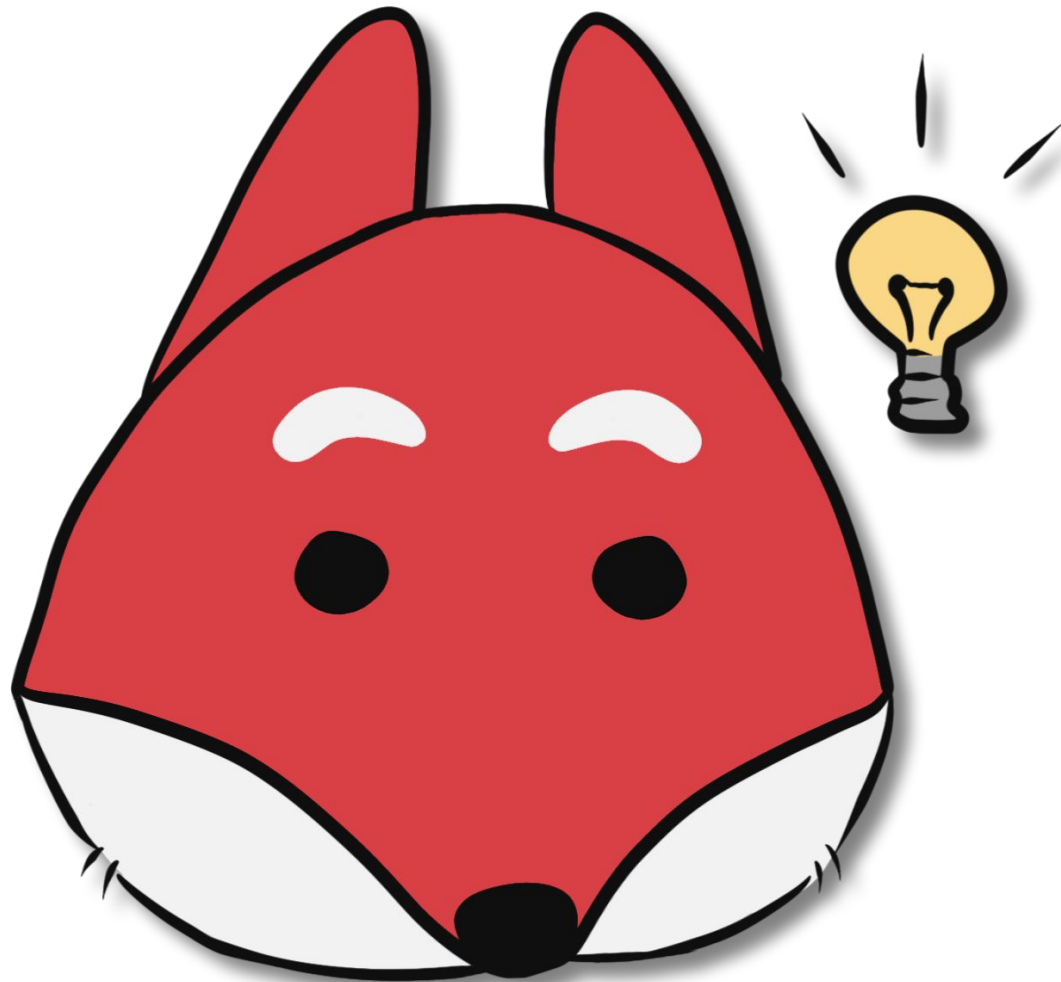


Doba před-Elasticová

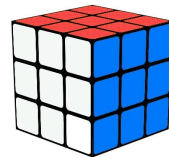
- MariaDB
- 600 GB+ dat
- relace

Doba před-Elasticová

- velké indexy
- data nešlo mazat



Elasticsearch v



- WTF
- NRT
- základy

Výhody

- škálovatelnost
- replikace
- rychlost

Výhody

- rolling restarts
- agregace
- snapshots

GUIs

- Kibana
- Cerebro



Přesun dat

- schizofrenní aplikace
- nárůst objemu

Mapping

- správný návrh indexů a mappingů
- nelze měnit
- scripty na update

Ztráta dat

- index-time (queues)
- shard-fail
- duplicity

Performance

- “it depends”
- počty replik
- hardware
- load-balancing
- proxy

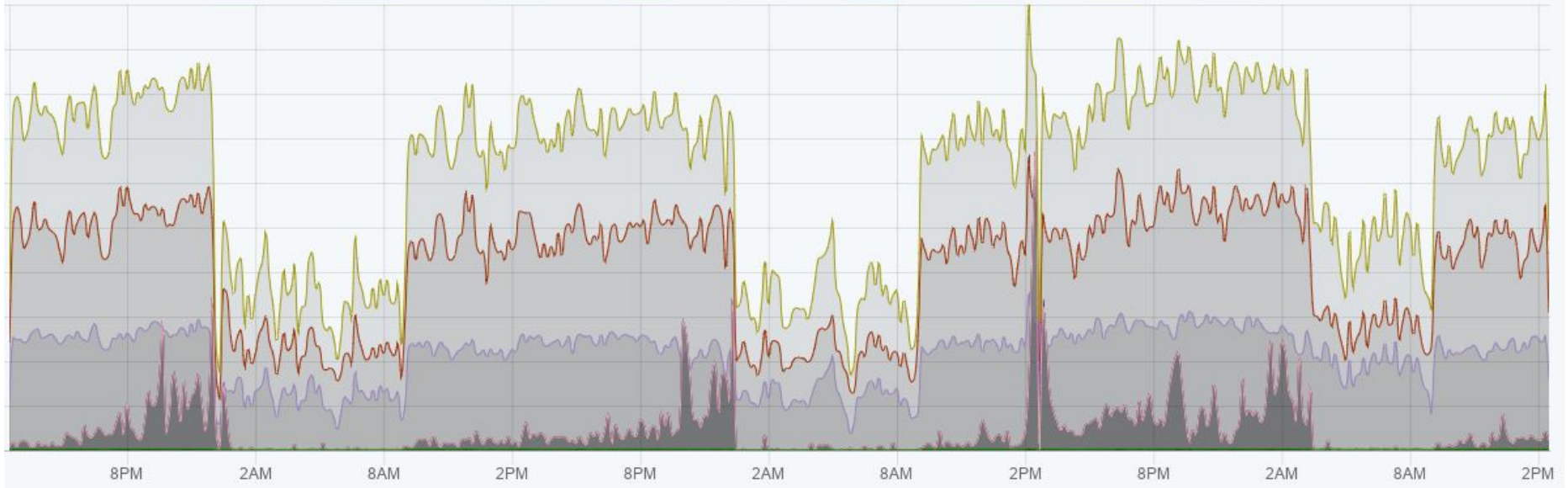
Update scripty

- stored / inline
- limity
- versioning
- konflikty

Write queues

- RabbitMQ
- bulk requests
- stats

requests operations conflicts errors index update conflict error avg



Lišáček Elasticsearch

- 2 oddělené clustery - produkce / archiv
- 3 nodes, 62 indices, 562 shards, 3 mld docs, 3.4TB
HW: 72 CPU, 384 GB RAM, 5.4 TB SSD
- 1 node, 201 indices, 473 shards, 29 mld docs, 8.6 TB
HW: 16 CPU, 64 GB RAM, 11 TB HDD

K čemu?

- aktuální / historická data k produktům
- machine learning
- validace

Otázky?





Jakub Kluváněk
jakub.klivanek@biddingfox.com
@kluvi, @biddingfox_devs