jako řešení pro logy?

… vážně?

Jan Šimák

PR

aktuálně v 

devops team lead

dříve v Seznamu

architekt zastřešující projekt interního cloud řešení

project manager

částečně i product manager

Elasticsearch?

znáš a máš zkušenosti s ES?

jaké úlohy můžeš řešit díky ES?

v průměru, jaký je poměr mezi user data a indexem?

co běží v pozadí?

Otázka #1

Jak velký cluster provozuješ co do počtu nodes?

Otázka #2

Jak velký objem dat indexuješ během provozní špičky?

Question #3

Kolik dat držíš ve stavu “warm” (dostupných pro dotazování)?

Výhody vs. Nevýhody

+ dokáže zpracovat téměř cokoli
+ dynamická detekce datového typu and vytváření indexu -
+ silný DSL dotazovací jazyk
+ snadné rozběhnout a začít používat
- značné požadavky na zdroje
- celkem obtížné udržet řešení efektivní
- celkem výzva udržet řešení v chodu u velkých realizací
- cena

Cena na prvním místě

výpočetní zdroje (index/query)

- v průměru, jeden node je schopen **zpracovat 20k zpráv/sec**
- cena za dotazování závislá na typu a frekvenci dotazů

storage zdroje

- v průměru můžeme připočíst **60% z velikosti původních dat na index**

licencování (když základ nestačí)

- ne vše je možné řešit cestou open source/free software

čas engineerů

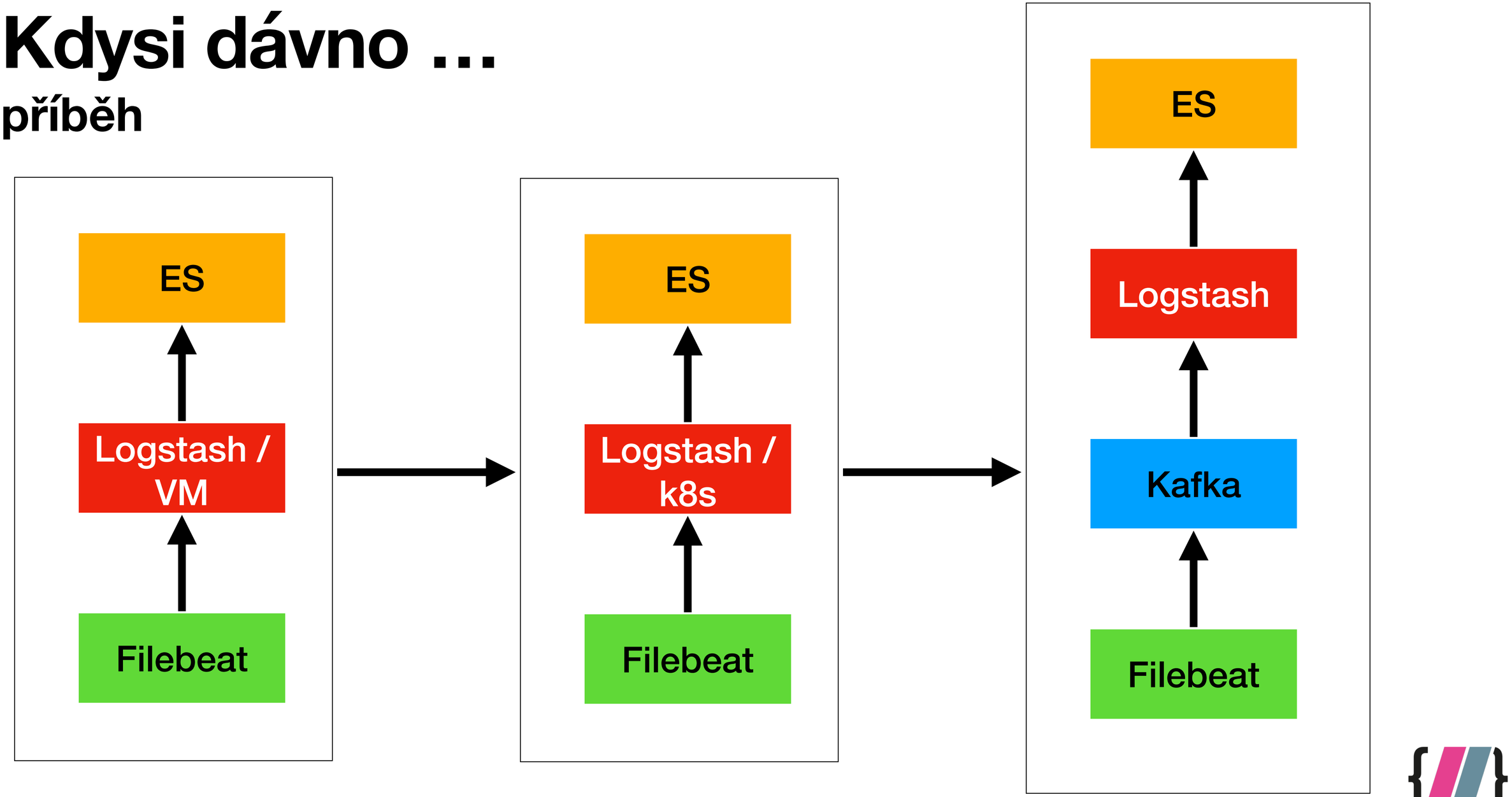
Kdysi dávno …
předmluva

cíle projektu

- jedno řešení pro všechny logy firmy
- čas zpracování blížící se reálnému času
- neexistující jednotná struktura logů
- vstup ve stovkách tisíc za sekundu (prostě hodně velké číslo)
- 100% dostupnost
- 0% ztráta dat

Kdysi dávno …

příběh



Kdysi dávno …
letíme na měsíc

vstup 400k/sec v průměru

~200TB dat dostupných pro okamžité dotazování

dostupnost řešení docela vysoká, ale …

ztráta dat docela nízká, ale …

data uchována v rozsahu dnů až týdnů podle typu vstupu

TLS všude a nasazen **RBAC**, aniž bychom utratili jedinou korunu

Kdysi dávno …

nohama zpátky na zemi

ohromné množství shardů na jeden node

index field explosion

data type collisions

zastaralá verze ES

index alias != data protection

Kdysi dávno
poučení

vždy následovat doporučení -> důkladně číst dokumentaci a brát ji vážně

skutečně chápat základní metriky a alertovat nad nimi

od začátku trvat na jednotné struktuře logů

neustále upgradovat ES s vysokou prioritou (např. log4shell)

omezení na počet fieldů per index je tvůj přítel

ES není Švýcarský nůž -> implementuj více tierů za použití různých produktů

měj vždy na paměti compatibility matrix

Co nějaká alternativy?

AWS OpenSearch

AWS fork originální verze Elasticsearch

přibaleno velké množství funkcionalit a stále zdarma

(cross-cluster replication, multitenacy kibana, document/field level security, ...)

můžeme provozovat on-prem nebo si zaplatit AWS aaS produkt

drží krok s novými verzemi ES

Co nějaká alternativy?

Grafana Loki

alternativní řešení pro logy od roku 2019?

zcela jiný přístup k indexovaní vstupu (data vs. metadata)

výrazně levnější v porovnání s řešením ES

project vznikl, aby řešil pouze zpracování logů

dobrý kandidát = základní kámen pro další tier

Díky!