

Návrh a implementace škálovatelné místní infrastruktury pro shromažďování, zpracování a ukládání dat ve velkém měřítku v aplikacích strojového učení

Jan Troják

České vysoké učení technické v Praze

Fakulta informačních technologií



Úložiště

Pipeline

Úložiště

- Malé soubory, velké množství
- Paralelní zápis
- Paralelní čtení
- Write-Once-Read-Many, no update
- Plochá adresovatelnost
- Vlastní metadata
- Výpadku tolerantní (RAID 01)
- Škálovatelné (jednoduše rozšiřitelné)
- Vysoká dostupnost

Úložiště

- Malé soubory, velké množství
- Paralelní zápis
- Paralelní čtení
- Write-Once-Read-Many, no update
- Plochá adresovatelnost
- Vlastní metadata
- Výpadku tolerantní (RAID 01)
- Škálovatelné (jednoduše rozšiřitelné)
- Vysoká dostupnost

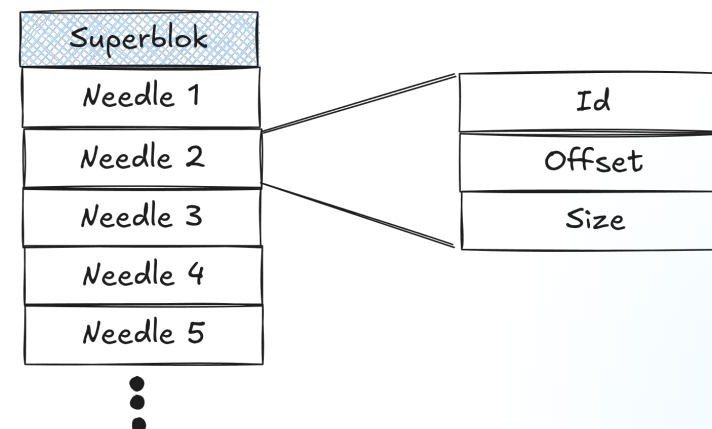
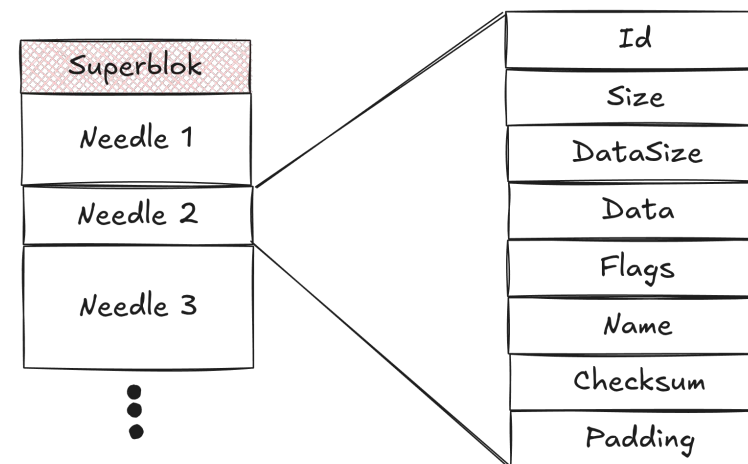
- POSIX
 - Pomalé vyhledávání
 - Nepotřebná metadata
 - Synchronizace a zámky
- Složitost a těžkopádnost univerzálních řešení
 - Snaha o vytvoření univerzálního řešení
 - Složitá implementace, složitý tuning
 - Často různá architektonická omezení

Volba SeaweedFS

Implementace Facebook Haystack designu.

Optimalizace pro malé soubory

- Volumes (~30 GiB)
 - Velké soubory jako kontejner pro hodně malých
 - Append only
 - Obsahují jen nezbytně nutné metadata pro soubory
 - Jednoduché pro replikaci
 - Write-enabled, read-only
 - Neumožňuje jednoduché mazání dat
- Index
 - Pole adres ve volume
 - Uložen v RAM
 - Drží otevřený FD na volume



| Úložiště

| Pipeline

Pipeline

- Idempotence a obnova po selhání
- Trvalé ukládání a indexace výsledků, postupů
- Kombinace vysoké propustnosti a nízké latence
- Horizontální škálovatelnost
- Autoscaling
- Vysoká dostupnost

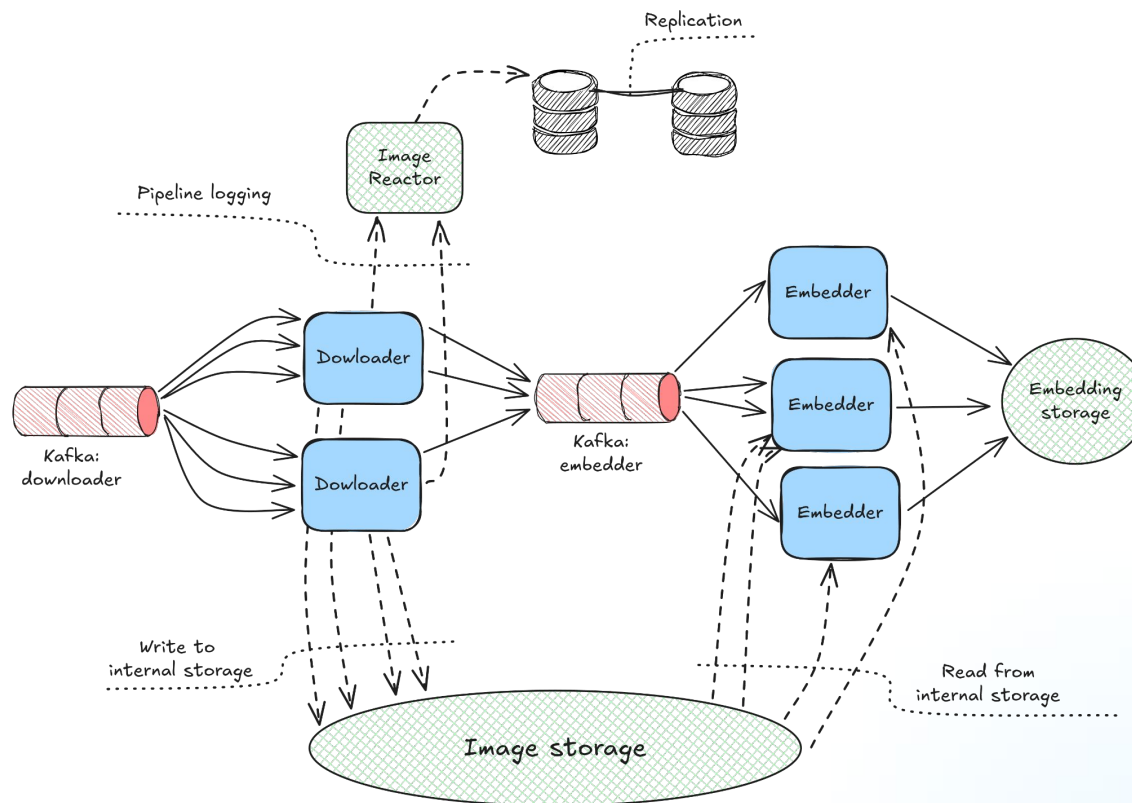
Pipeline

- Idempotence a obnova po selhání
- Trvalé ukládání a indexace výsledků
- Kombinace vysoké propustnosti a nízké latence
- Horizontální škálovatelnost
- Autoscaling
- Vysoká dostupnost

- Online pipelines
 - Vysoká dostupnost
 - Nízká latence
 - Důraz na výkonnost
 - Založený na požadavcích
- Offline pipeline
 - Vysoká propustnost
 - Důraz na ne-chybovost
 - Dávkový vstup
- Streaming pipeline
 - Kombinace online a offline
 - Nízká latence
 - Vysoká propustnost
 - Nutnost plánovat práci

Streaming pipeline

- **Kafka**
 - Perzistentně drží práci
 - HA, Parallel write, Parallel read
- **Downloader**
 - Stahuje obrázky z internetu
 - Transformuje data
 - Nahrává a registruje stažené soubory
- **Embedder**
 - Schránka pro ML model (embedding, annotations...)
- **Keda**
 - Kubernetes native job, service scheduler
 - Umí škálovat na 0
 - Podporuje joby a services
 - Pokročilé možnosti plánování



Děkuji za pozornost

Jan Troják | FIT ČVUT | 2026

Otázka

- `enable.auto.commit: true (5s)`
- Co se stane s rozdělanou prací při restartu podu?

- Očekávaný restart
 - Všechna přijatá práce je dokončena, až poté se pod ukončí
- Neočekávaný restart
 - Progress se ztratí. Administrátor musí udělat ruční zásah pro opravu
- Řešení - záleží na nasazení (produceru)
 - `enable.auto.commit: false`
 - Opakovaná práce není problém
 - Job fetcher
 - Použití message queue s podporou retry (rabbitMQ)

