Elasticsearch

Elasticsearch je distribuovaný, RESTful vyhledávací a analytický engine navržený pro horizontální škálovatelnost, maximální spolehlivost a snadnou správu. Je schopen prohledávat a analyzovat obrovské objemy dat v reálném čase.

Zatímco klasické relační databáze (SQL) jsou skvělé pro transakce, Elasticsearch exceluje v **full-textovém vyhledávání** a komplexní analýze logů.

Základní koncepty

Abychom pochopili, jak Elasticsearch funguje, musíme znát jeho vnitřní strukturu:

- **Index:** Logický prostor pro ukládání dat (odpovídá „databázi“ v SQL). Např. index `logs-2024-05`.
- **Document:** Základní jednotka informací uložená ve formátu JSON (odpovídá „řádku“ v SQL).
- **Field:** Konkrétní datové pole v dokumentu (odpovídá „sloupci“ v SQL).
- **Inverted Index (Invertovaný index):** Klíčová datová struktura, která umožňuje bleskové vyhledávání. Místo prohledávání dokumentů hledá slova v seznamu a okamžitě ví, ve kterých dokumentech se nacházejí.

Architektura a škálování

Elasticsearch je navržen jako distribuovaný systém, což znamená, že data jsou rozdělena mezi více serverů (**Nodes**), které tvoří **Cluster**.

- **Shards (Střepy):** Index je rozdělen na menší části (shardy). To umožňuje rozdělit zátěž mezi více uzlů.
- **Replicas (Repliky):** Každý shard může mít svou kopii. Pokud jeden server selže, data jsou stále dostupná z repliky na jiném serveru. Tím je zajištěna vysoká dostupnost (High Availability).

Hlavní vlastnosti

- **Near Real-Time (NRT):** Od okamžiku indexace dokumentu do chvíle, kdy je vyhledatelný, uplyne obvykle méně než jedna sekunda.
- **REST API:** Veškerá komunikace s Elasticsearch probíhá pomocí standardních HTTP metod (GET, POST, PUT, DELETE), což usnadňuje integraci s jakýmkoliv programovacím jazykem.
- **Schéma-less:** Dokumenty lze do indexu posílat bez předem definovaného schématu. Elasticsearch se pokusí datové typy (číslo, text, datum) odhadnout automaticky.
- **Agregace:** Umožňuje provádět složité statistické výpočty nad daty (např. „vypočítej průměrnou latenci webu za poslední hodinu seskupenou podle zemí“).

Použití v praxi

Oblast	Příklad použití
Vyhledávání	Full-textové vyhledávání v e-shopech (např. našeptávač, filtry).
Log Management	Centrální úložiště pro ELK Stack (analýza systémových logů).
Bezpečnost	Detekce hrozeb v reálném čase (SIEM).
Business Intelligence	Sledování prodejů a chování uživatelů v reálném čase.

Srovnání: SQL vs. Elasticsearch

Pojem	Relační DB (SQL)	Elasticsearch
Organizace	Databáze	Index
Struktura	Tabulka	Index (skupina dokumentů)
Záznam	Řádek	Dokument (JSON)
Schéma	Pevně dané (Strict)	Dynamické / Flexibilní
Silná stránka	Integrita, vztahy (JOIN)	Rychlost vyhledávání, analýza

> **Důležité:** Elasticsearch není náhradou za primární SQL databázi. Nehodí se pro komplexní relace mezi daty (JOINy jsou výpočetně drahé) a nemá klasické transakce (ACID). Nejlépe funguje jako sekundární vrstva pro rychlé vyhledávání.

— **Viz také:** ELK Stack, Kibana, JSON, Full-text vyhledávání