

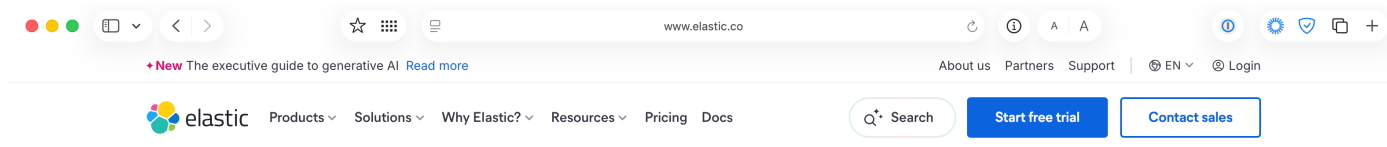
01 - Úvod do Elasticsearch

Elastic je společnost, která vytvořila nástroj Elasticsearch.

Elasticsearch je distribuovaný vyhledávací a analytický engine pro všechny druhy dat - textová, číselná, geografická, strukturovaná i nestrukturovaná.

Elastic nabízí také další nástroje a služby, které lze využívat samostatně, nebo společně - především jde o Kibana, Beats a Logstash. Tato kombinace nástrojů používá označení **Elastic Stack** (dříve ELK Stack).

Dále Elastic nabízí spravovanou službu **Elastic Cloud**, a v neposlední řadě také knihovny pro většinu běžně používaných programovacích jazyků.



The open source platform that powers search, observability, security, and more ...

Build with Elasticsearch

Start free trial

Explore Elastic

Elastic is a Leader in the 2025 Gartner® Magic Quadrant™ for Observability Platforms for the second year in a row!

[Access the report →](#)

Effortless Elasticsearch management — AutoOps is here, and it's free for Elastic Cloud customers.

[Read the blog →](#)

Automatically migrate your legacy SIEM with AI. Cut onboarding from days to minutes.

[Read the blog →](#)

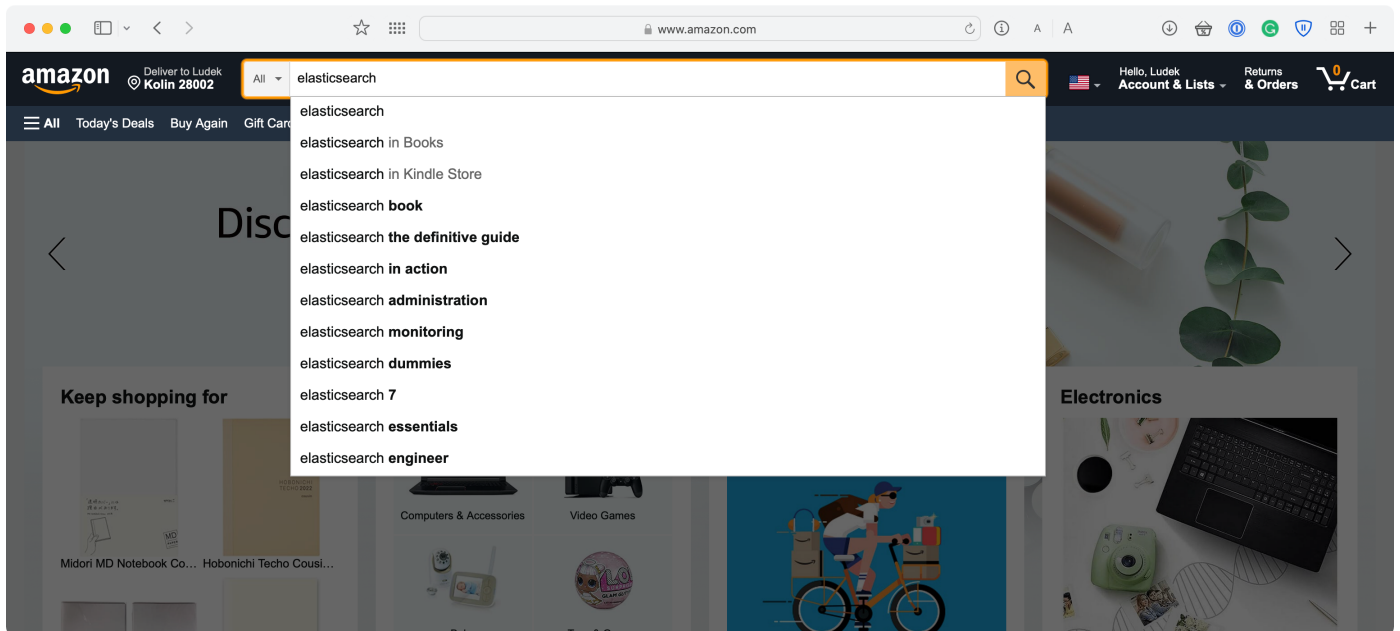
TRUSTED BY 50% OF THE FORTUNE 500 TO DRIVE INNOVATION



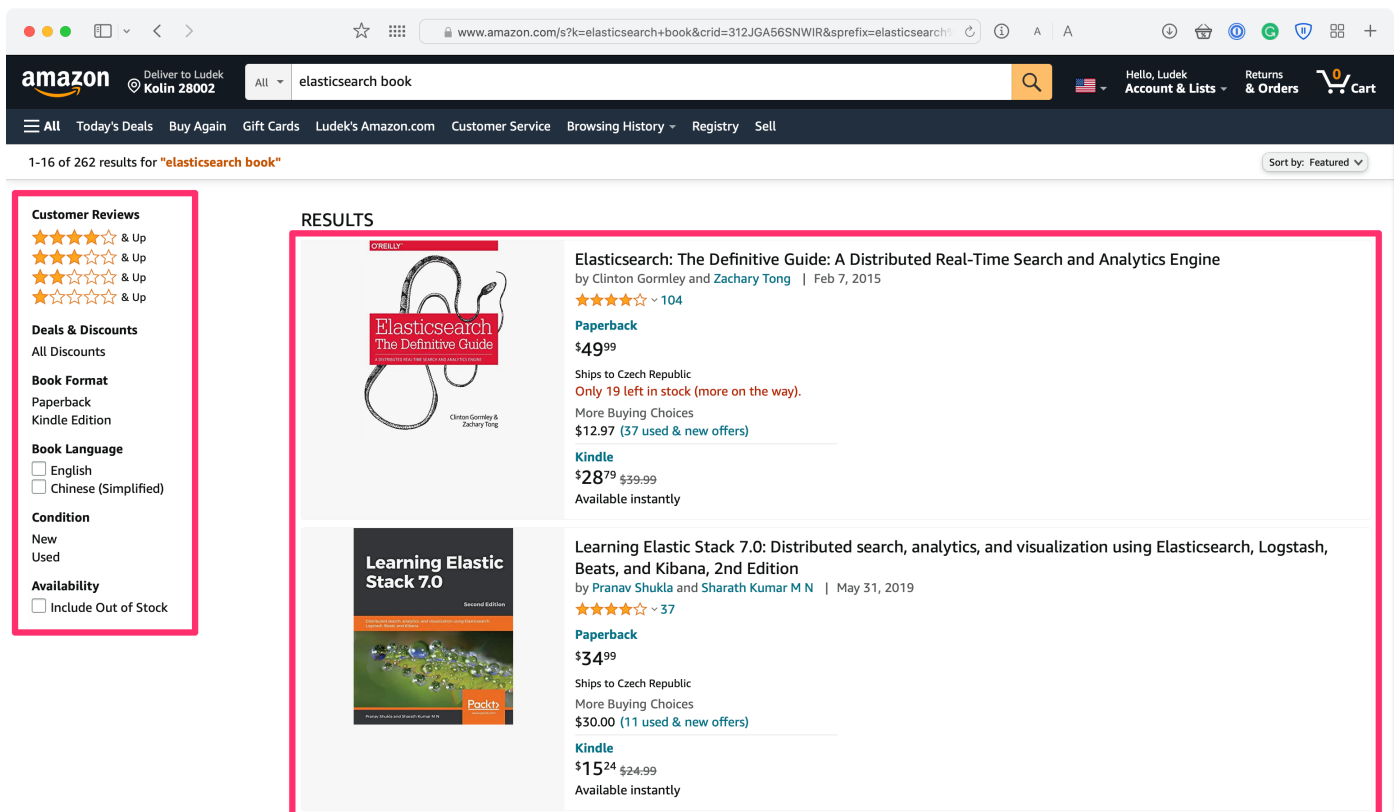
Co to je Elasticsearch?

Elasticsearch je dokumentové úložiště, NoSQL databáze nebo také vyhledávací a analytický engine. Byl primárně navržen pro **fulltextové vyhledávání**, nyní je na něj však navázána řada dalších nástrojů, takže bývá často využíván i jako **analytický nástroj**.

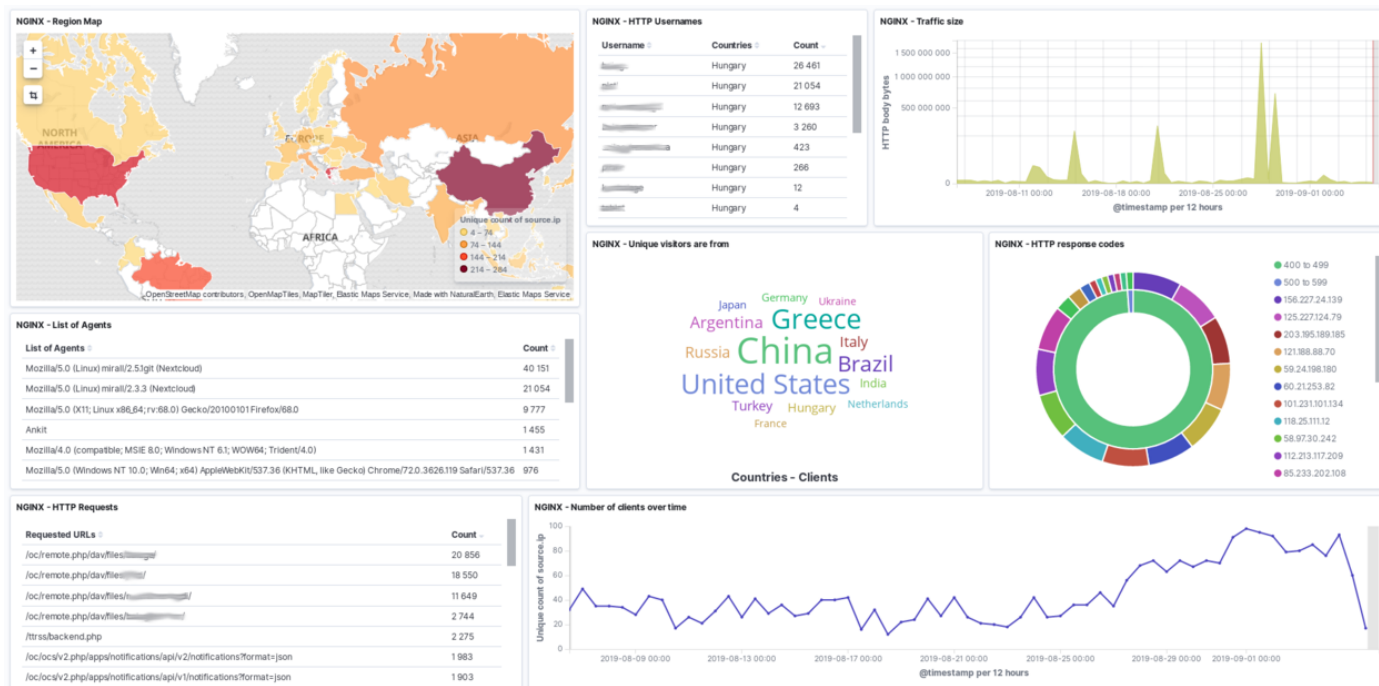
Mezi jeho hlavní funkcionality při implementaci vyhledávání patří práce s přirozeným jazykem, tedy zohlednění tvarosloví, nebo například překlepů:



Elasticsearch lze také využívat jako distribuované úložiště, které dokáže pracovat s opravdu velkým množstvím dat. Zároveň zůstává velmi rychlý, lze jej tak využít například k **výpisu dat v kombinaci s použitím filtrů**:



Elasticsearch je také **analytický engine**, který nám umožňuje zjistit přehled o veškerých uložených datech. V kombinaci s Beats a Kibana může být Elasticsearch využit pro ukládání a analýzu logů a metrik:



Základní vlastnosti Elasticsearch

Elasticsearch je snadno **horizontálně škálovatelný**. Pokud vám nedostačuje výkon nebo kapacita serveru, jednoduše spustíte další a umožníte mu připojit se do existujícího clusteru. Elasticsearch si sám řídí komunikaci mezi jednotlivými servery, rozložení dat napříč clusterem. Díky tomu dokáže pojmout opravdu velké množství dat a přitom v nich stále rychle vyhledávat. Současně s tím pracuje s replikami dat, při výpadku nebo pádu některého z propojených serverů se dokáže s touto událostí vypořádat a data obnovit z jejich kopie.

Elasticsearch pracuje **real-time** (respektive near real-time při vyhledávání). Jakmile jsou data do Elasticsearch uložena, je okamžitě možné v nich vyhledávat, nebo nad nimi spouštět analytické dotazy. Reálně zde však drobná prodleva vzniká, ve výchozím nastavení je to maximálně 1s.

Rychlost je zajištěna mimo jiné tím, že pracuje s **denormalizovanými daty**. Pokud tedy již máte data v relační databázi, je třeba navrhnout způsob, jakým je do Elasticsearch přenášet a udržovat aktuální.

Elasticsearch je psaný v **Javě**, lze jej tedy provozovat pod libovolným běžně používaným operačním systémem. Konfiguruje a provozuje se podobně, jako jakákoliv jiná Java aplikace. Na jednu stranu je relativně náročný na hardwarové zdroje, na druhou stranu je však velmi rychlý. Knihovna, kterou Elasticsearch využívá pro implementaci vyhledávání je **Apache Lucene**.

Při dotazování se Elasticsearch máme k dispozici řadu vyhledávacích (**queries**) a analytických (**aggregations**) dotazů. Kromě vyhledávání konkrétních záznamů nám tak dává rychle odpověď na otázky typu "Kolik zákazníků si koupilo telefon Samsung?" nebo "Jak se vyvíjí průměrná cena položky objednávky v čase?".

S Elasticsearch se komunikuje primárně pomocí **REST API** a data se předávají ve formátu **JSON** (QueryDSL). Alternativně je možné s Elasticsearch komunikovat i pomocí SQL nebo JDBC/ODBC klientů (vyžadují platnou licenci). Plnou funkcionalitu však získáme pouze použitím REST API. Při psaní dotazů v nástroji Kibana využijeme také další syntax - KQL (Kibana Query Language), případně ES|QL.

Omezení Elasticsearch

Klady Elasticsearch jsou na druhé straně vyváženy omezeními, které z vlastností Elasticsearch vyplývají. Před využitím Elasticsearch pro danou úlohu je tedy třeba pečlivě zvážit, zda je Elasticsearch optimální technologií v dané situaci.

- Omezené možnosti pro uložení relací - při přenosu dat z RDBMS je nutné je transformovat
- Změna schéma (např. změna pole na jiný datový typ) je náročnou operací vyžadující reindexaci dat
- Elasticsearch není ACID - nepodporuje transakce, tak jak je známe z ostatních RDBMS, může docházet k nekonzistenci dat, některé výpočty (např. relevance) mohou být nepřesné (ale používá optimistic concurrency control na úrovni dokumentu)
- Omezené možnosti Kibany ve srovnání s nástroji typu Tableau (na druhou stranu je však funkcionality pravidelně rozšiřována)

Základní pojmy

Document

Dokument je základní jednotka, s kterou umí Elasticsearch pracovat. Pokud chceme nějakou informaci uložit do Elasticsearch, nebo ji později získat, pracujeme právě s dokumentem. Příkladem dokumentu může být objednávka, produkt nebo záznam v logu. Ve světě relačních databází dokument odpovídá záznamu (řádku) v tabulce. Dokumenty jsou vyjádřeny v formátu JSON. Každý dokument má jednoznačný identifikátor (`_id`), který můžeme definovat při vytváření, nebo nechat Elasticsearch vygenerovat nový.

Field

Field (pole) je část dokumentu, která má určitý datový typ. Elasticsearch podporuje běžné datové typy, jako celé číslo, desetinné číslo, textový řetězec nebo datum, podporuje i další datové typy, jako pole nebo vnořený dokument. Zjednodušeně řečeno - co jde napsat do JSON dokumentu, to se do Elasticsearch uloží.

Index

Index je sada dokumentů stejné struktury. Jako příklad lze uvést index s produkty nebo index s objednávkami. Index je identifikován názvem (malými písmeny). Ve světě relačních databází je ekvivalentem tabulky. Indexů můžeme vytvořit libovolné množství.

Tento index (v rámci Elasticsearch) nemá nic společného s indexem používaným v relačních databázích.

Cluster

Cluster je označení pro jeden nebo více serverů (nodů), které jsou spolu propojeny a pracují se stejnými daty. Cluster je identifikován jednoznačným názvem. Ve výchozím stavu je zapnutá security, tedy komunikace v rámci clusteru je šifrovaná.

Node

Node je spuštěná instance Elasticsearch, která běží na jednom serveru. Pokud si spustíte Elasticsearch lokálně na svém počítači, bude celý cluster tvořen jediným nodem. V produkčním prostředí se zpravidla používá nodů více, je však možné Elasticsearch používat i s jediným nodem, pouze se připravíte o HA. Nody jsou identifikovány pomocí vygenerovaného UUID, které je případně možné přenastavit.

Shard

Data v indexu jsou dělena do shardů, přičemž shard je interně tvořen Lucene indexem. To je ta nejmenší jednotka, v které probíhá ukládání dat nebo jejich vyhledávání, přičemž dokáže využít právě jedno vlákno procesoru. Každý shard může být uložen na jiném serveru. Díky tomu je možné vytvořit index větší, než je kapacita jednoho serveru a zároveň lze díky tomu paralelizovat vyhledávání.

Shardy dělíme na primární (**primary**) a jejich repliky (**replica shards**), což jsou kopie primárních shardů. Slouží k záloze dat a díky tomu, že se repliky automaticky umístí na různé nody jsou data ochráněna před ztrátou v případě výpadku některého nodu.

Počet replik i shardů se definuje při vytváření indexů, pokud však není některý z nich uveden, použije se defaultní nastavení (1 shard a 1 replika). Počet replik lze snadno měnit i po vytvoření indexu. Je technicky možné měnit také počet shardů, jde ale o náročnou operaci a v praxi bývá zpravidla lepší vytvořit nový index.

Porovnání pojmů Elasticsearch a relační databáze

Pro představu uvádím následující tabulku přiřazující výše uvedené pojmy k pojmům známým ze světa relačních databází:

Elasticsearch	Relační Databáze
Cluster	Databáze
Index	Tabulka
Document	Řádek v tabulce
Field	Sloupec v tabulce

Oficiální dokumentace

Elasticsearch má kvalitní dokumentaci, která je dostupná na adrese:

- <https://www.elastic.co/docs/solutions/search>

Dokumentace je obsáhlá a je v ní i řada konceptů detailně vysvětlena. Stejně, jako má svou dokumentaci Elasticsearch, mají ji i další části Elastic Stacku (Kibana, Logstash, nebo Beats). Jako rozcestník k dílčím dokumentacím můžete využít následující odkaz:


- <https://www.elastic.co/docs>

www.elastic.co/docs

New

The executive guide to generative AI [Read more](#)

[About us](#) [Partners](#) [Support](#) | [Login](#)

 [Products](#) [Solutions](#) [Why Elastic?](#) [Resources](#) [Pricing](#) [Docs](#)

Search

Start free trial

Contact sales

Docs

[Release notes](#) [Troubleshoot](#) [Reference](#)


Elastic Docs

Welcome to the docs that cover all changes in Elastic Stack 9.0.0 and later, including Elastic Stack 9.2.0 and Elastic Cloud Serverless. For easy reference, changes in 9.1.0 are marked inline. For details, check [Understanding versioning and availability](#).

Elastic Fundamentals

Upgrade versions

[View previous docs versions](#) →




Elasticsearch

Build powerful search and RAG applications using Elasticsearch's vector database, AI toolkit, and advanced retrieval capabilities.

[9.0+ \(latest 9.2.0\) docs](#)

[Previous versions](#)




Observability

Resolve problems with open, flexible, and unified observability powered by advanced machine learning and analytics.

[9.0+ \(latest 9.2.0\) docs](#)

[Previous versions](#)



Security

Detect, investigate, and respond to threats with AI-driven security analytics to protect your organization at scale.

[9.0+ \(latest 9.2.0\) docs](#)

[Previous versions](#)

Looking for more?

We'll show you how to solve your toughest challenges with Search AI. Join us at an Elastic{ON} event near you!

Register now