

Apache Kafka

Apache Kafka je open-source platforma pro distribuovaný přenos a zpracování datových proudů (streams). Na rozdíl od tradičních systémů pro zasílání zpráv (Message Brokers) funguje Kafka spíše jako distribuovaný transakční protokol (log), který umožňuje data ukládat a zpětně procházet.

1. Základní koncepty

Architektura Kafky stojí na několika klíčových prvcích:

- **Producer (Producent):** Aplikace, která posílá (publikuje) data do Kafky.
- **Consumer (Konzument):** Aplikace, která data čte a zpracovává.
- **Topic (Téma):** Kategorie nebo název kanálu, do kterého jsou data ukládána. Témata jsou rozdělena na **Partitions** (partice) pro zajištění paralelismu a škálovatelnosti.
- **Broker:** Server v clusteru Kafky, který ukládá data a vyřizuje požadavky klientů.

2. Proč je Kafka unikátní?

Na rozdíl od starších systémů (jako RabbitMQ), kde se zpráva po přečtení smaže, Kafka funguje jinak:

- **Persistence:** Zprávy jsou ukládány na disk a uchovávány po nastavenou dobu (např. 7 dní) nebo do dosažení určité velikosti. To umožňuje konzumentům číst data i zpětně.
- **Vysoká propustnost:** Kafka dokáže zpracovat miliony zpráv za sekundu díky efektivnímu zápisu na disk a minimální režii.
- **Škálovatelnost:** Přidáním dalších brokerů do clusteru lze prakticky neomezeně zvyšovat kapacitu systému.
- **Záruka pořadí:** V rámci jedné partice je zaručeno, že zprávy budou přečteny přesně v tom pořadí, v jakém byly zapsány.

3. Ekosystém Kafky

Kafka není jen „potrubí“ na data, ale celý ekosystém nástrojů:

- **Kafka Connect:** Nástroj pro snadné propojení Kafky s databázemi (např. SQL), indexy (Elasticsearch) nebo cloudovými úložišti bez psaní kódu.
- **Kafka Streams:** Knihovna pro budování aplikací, které transformují, agregují a analyzují data přímo „za letu“ uvnitř Kafky.
- **ksqlDB:** Databáze pro streamy, která umožňuje psát dotazy nad živými daty pomocí jazyka podobného SQL.

4. Využití v praxi

- **Sledování aktivit (Activity Tracking):** Logování každého kliknutí na webu pro real-time analýzu.
- **Metriky a monitoring:** Shromažďování dat z tisíců serverů pro sledování jejich stavu.
- **Log Aggregation:** Centrální sběr logů z mnoha mikroslužeb.
- **Finanční transakce:** Zpracování plateb a detekce podvodů v reálném čase.

5. Porovnání: Kafka vs. Tradiční Message Broker

Vlastnost	Apache Kafka	RabbitMQ / ActiveMQ
Model	Pull (konzument si říká o data)	Push (server posílá data)
Ukládání dat	Persistentní log (zůstávají v systému)	Dočasná fronta (smažou se po doručení)
Škálovatelnost	Velmi vysoká (distribuovaná)	Omezená (hůře se škáluje horizontálně)
Použití	Streamování, Big Data, Analýza	Komunikace mezi službami, úkoly na pozadí

Související články:

- [Big Data a 5V](#)
- [Architektura mikroslužeb](#)
- [Apache Spark: Zpracování v paměti](#)

Tagy: *it big-data kafka streaming backend architecture infrastructure*

From: <https://serviceit.cz/> - IT ENCYKLOPEDIE

Permanent link: https://serviceit.cz/doku.php?id=it:sw:apache_kafka

Last update: 2026/01/02 20:00

