

# Indexovanie veľkých dát v internete vecí: metódy a využitie\*

Miloš Bardáč

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií  
`xbardac@stuba.sk`

30. september 2023

## Abstrakt

V súčasnej dobe nepretržitého rastu objemu dát v oblasti internetu vecí je efektívne indexovanie neoddeliteľnou súčasťou. Z toho dôvodu sa tento článok zaoberá problematikou indexovania veľkých dát v IoT a rôznymi technikami, ktoré sa využívajú v IoT prostredí. Okrem toho sa článok venuje aj požiadavkám a výzvam spojeným s indexovaním objemných dát v IoT. Medzi tieto výzvy patrí napríklad bezpečnosť a ochrana súkromia, cena, či postupom času veľký nárast objemu dát. Tento článok poskytuje pohľad na dôležitosť efektívneho indexovania v kontexte Internetu vecí a taktiež pomáha čitateľom lepšie pochopiť techniky indexovania a výzvam v tomto odvetví. Taktiež vysvetľuje prečo je rýchly a spoľahlivý prístup k dátam kľúčovým aspektom využitia IoT dát pre rôzne aplikácie a inovácie.

## 1 Úvod

Vstupujeme do doby, kedy sa svet okolo nás stáva čoraz viac prepojeným a inteligentným. Internet vecí, skrátene IoT, môžeme chápať ako súbor sietí, v ktorých sú fyzicky prepojené dáta alebo zariadenia, čo umožňuje týmto zariadeniam zhromažďovať a zdieľať dáta. [1] Práve táto technológia zohráva kľúčovú úlohu v tomto pomerne rýchlom pokroku. V poslednom čase môžeme byť svedkami pomerne masívneho nárastu objemu dát v oblasti internetu vecí. S týmto nárastom je spojená potreba tieto dáta efektívne usporiadať a prehľadávať.

Preto sa v nasledujúcich kapitolách bude článok venovať rôznym aspektom tohto nevyhnutného procesu. Hneď v nasledujúcej kapitole (časť 2) bude čitateľ oboznámený s požiadavkami spojenými s touto problematikou. Nasledujúca kapitola (časť 3) je spojená so samotnými oblasťami využitia. V ďalšej časti (časť 4) sa článok zameriava na rôzne techniky indexovania objemných dát v oblasti Internetu vecí. Tie majú možnosť dosť výrazne zlepšiť rôzne odvetvia, či už je to priemysel, energetika či dokonca aj zdravotná starostlivosť. Jedná z posledných kapitol (časť 5) je venovaná problémom a výzvam, ktoré sa spájajú

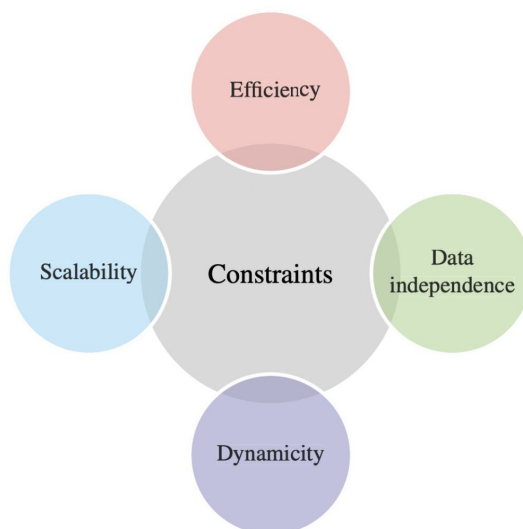
---

\*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2023/24, vedenie: Pavol Baťalík

s touto tematikou. Patrí sem otázka bezpečnosti dát, ochrany súkromia používateľov, či správa zariadení. V predposlednej časti (časť 6) ukončíme tento článok záverom a zhrnutím. Zároveň v tejto časti si čitateľ pripomenie kľúčové body, ktorým sa článok venuje a taktiež sa dozvie o víziách tejto problematiky do budúcnosti. Na úplnom konci, posledná kapitola (časť 7) bude obsahovať reakciu na témy z prednášok.

## 2 Požiadavky

Veľké dáta sú obrovské množstvá štruktúrovaných alebo neštruktúrovaných údajov, ktoré musia byť spracované. Internet vecí reprezentuje sieť zariadení, ktoré sú pripojené k internetu a komunikujú medzi sebou. S príchodom Internetu vecí sa objem generovaných údajov výrazne zvýšil, čo znamená, že je potrebné zvýšiť efektivitu skladovania a správy týchto údajov. Požiadavky sú špecifikácie, ktoré určujú akým spôsobom by mal byť proces indexovania vykonaný a zároveň aby bol úspešný a čo najefektívnejší. Požiadavky taktiež stanovujú ako má byť daný systém indexovania navrhnutý a implementovaný aby najlepšie splňoval potreby daného IoT prostredia. Nedávne štúdie sa stále viac zameriavajú na optimalizáciu výkonnosti v rozsiahlych dátových sadách s čo najväčšou efektívnosťou. [2] Preto sa táto kapitola pokúsi čitateľovi priblížiť základné požiadavky, ktoré treba zvážiť pri indexovaní dát v kontexte internetu vecí. Medzi tieto požiadavky patrí škálovateľnosť, efektivita, dynamicnosť, či nezávislosť na dátach. [3] Každá z týchto kľúčových vlastností je samostatne vysvetlená v podkapitolách nižšie.



Obr. 1: Požiadavky pri indexovaní veľkých dát [3]

## 2.1 Škálovateľnosť

Výraz škálovateľnosť, rozšíriteľnosť, nám hovorí o schopnosti indexovacieho systému prispôbiť sa postupne rýchlejšie narastajúcemu objemu dát. V bežnom svete to znamená, že daný systém musí byť schopný spracovávať stále väčšie množstva nových dát bez výraznejšieho poklesu rýchlosti. [3] Na to je potrebný optimalizovaný algoritmus a dátové štruktúry, aby bol tento systém schopný udržať výkon aj pri väčšom zvýšení záťaže. Obzvlášť v oblasti IoT kde sa prevažne pracuje s obrovským množstvom dát je táto vlastnosť nevyhnutná.

## 2.2 Efektivita

Efektivita (efektívnosť) spočíva v schopnosti systému poskytovať najrýchlejšiu možnú odozvu na dotazy vyhľadávania. [4] V jednoduchosti musí systém od odoslania požiadavky užívateľom v čo najkratšom čase doručiť relevantný výsledok. Ďalej efektivita zahŕňa aj výber vhodných algoritmov, pretože výpočet podobnosti objektov môže byť značne náročnejší.

## 2.3 Dynamickosť

Dynamickosť (dynamika) znamená schopnosť za čo najmenší čas reagovať na zmeny dátach a databáze bez nutnosti reorganizácie, ktorá by zabrala príliš veľa času. To nám umožňuje mať nepretržitú indexáciu a stále aktuálne údaje. [3] V dnešnom svete Internetu vecí je tato špecifikácia z dôvodu nutnosti zmien a pridávania nových zariadení taktiež nevyhnutná.

## 2.4 Nezávislosť na dátach

Táto vlastnosť umožňuje indexovaciemu systému účinnosť aj v prípade, že sa na vstupe objavia rôzne typy dát ako aj ich rozloženie. Najmä v oblasti IoT kde sú mnoho krát rôzne formáty, štruktúry a značkovanie dát, musí byť schopný tieto údaje efektívne spracovať, bez ohľadu na ich charakteristiky. [2] To robí nezávislosť na dátach neoddeliteľnú súčasť Internetu vecí.

# 3 Oblasti využitia

Spektrum využitia Internetu vecí je veľmi široké. Dôvody používania sú rôzne, či už je to kvôli zvýšeniu efektivity, zníženiu nákladov alebo celkovému zlepšeniu kvality produktov, či služieb. Aplikácia indexovania big IoT dát je široká a zohráva kľúčovú úlohu v mnohých odvetviach.

## 3.1 Využitie v zdravotníctve

Aplikácia veľkých dát spojených s IoT v oblasti zdravotnej starostlivosti je novým a pomerne rýchlo sa rozvíjajúcim oblastným výskumom. Využitie zariadení a senzorov internetu vecí umožňuje zdravotnému personálu jednoducho zbierať a analyzovať veľké množstvo dát v reálnom čase, čo prispeje k lepším výsledkom pacientov, lepšej prevencii, zníženiu nákladov, či celkovému zlepšeniu kvality zdravotnej starostlivosti. [2] Stále využívanie internetu vecí a veľkých dát v oblasti zdravotníctva má vysoký potenciál prispieť k udržateľnejšiemu rozvoju.

Technológia IoT má významné uplatnenie najčastejšie v oblastiach zdravotnej starostlivosti, ako je monitorovanie UV žiarenia, zubná starostlivosť, či detekcia pádov. [5] Napriek tomu v spojitosti s touto témou vzniká mnoho problémov a výziev, ktorým treba čeliť. Či už sú to problémy s ochranou osobných údajov, bezpečnosti a mnoho ďalších, ktoré budú všetky spomenuté v jednej z posledných kapitol (časť . 5).

### 3.2 Využitie v priemysle

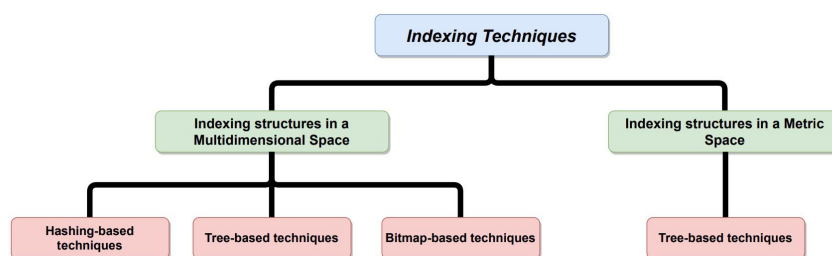
Veľké množstvo údajov má v priemyselnom odvetví mimoriadny význam, pretože slúži na zlepšenie a potenciálne urýchlenie výrobných procesov. Tieto rozsiahle zdroje údajov, často označované ako "priemyselné údaje IoT" (internet vecí), pochádzajú z rôznych snímačov, meracích prístrojov a automatizovaných systémov vo výrobných zariadeniach. Zahŕňajú informácie týkajúce sa veličín, ako sú teplota, tlak, vlhkosť, stav strojov, spotreba energie a množstvo ďalších parametrov. [6]

### 3.3 Využitie vo výskume

Výskumníci už aj v dnešnej dobe skúmajú potencionálne využívanie a dôsledky technológie internetu vecí pomocou veľkých objemov dát. Predpokladá sa [7]

## 4 Techniky indexovania veľkých dát v IoT

Techniky použité pri indexovaní údajov musia byť čo najefektívnejšie, čo vôbec neľahčuje príchod nových a objemnejších typov dát. Z dôvodu veľkého počtu techník indexovania objemných dát v internete vecí sú v tejto kapitole spomenuté iba tie najviac rozšírené.



Obr. 2: Globálna taxonómia techník indexovania [2]

### 4.1 Techniky založené na stromoch (Tree-Based)

Technika Tree-Based sa používa na indexovanie viacrozmerných údajov na základe rozdelenia priestoru na podpriestory alebo bunky. [2]

4.1.1 R\*-tree

4.1.2 BB-tree

4.1.3 Kd-tree

4.1.4 Quad-tree

4.1.5 Pyramid-tree

5 Výzvy

6 Záver a zhrnutie

7 Reakcia na témy z prednášok

## Literatúra

- [1] Lubhna Kaul and R.H Goudar. Internet of things and big data - challenges. In *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, pages 1–5, 2016.
- [2] Zineddine Kouahla, Ala-Eddine Benrazek, Mohamed Amine Ferrag, Brahim Farou, Hamid Seridi, Muhammet Kurulay, Adeel Anjum, and Alia Asheralieva. A survey on big iot data indexing: Potential solutions, recent advancements, and open issues. *Future Internet*, 14(1):19, Dec 2021.
- [3] Mohamed Amine Ferrag, Zineddine Kouahla, Hamid Seridi, and Muhammet Kurulay. Big iot data indexing: Architecture, techniques and open research challenges. In *2019 International Conference on Networking and Advanced Systems (ICNAS)*, pages 1–6, 2019.
- [4] Karima Khettabi, Zineddine Kouahla, Brahim Farou, Hamid Seridi, and Mohamed Amine Ferrag. Clustering and parallel indexing of big iot data in the fog-cloud computing level. *Transactions on Emerging Telecommunications Technologies*, 33, 2022.
- [5] Gurinder Singh, Vikas Garg, Pooja Tiwari, and Richa Goel. Emergence of iot and big data. 2020.
- [6] Xiaokang Zhou, Yiyong Hu, Wei Liang, Jianhua Ma, and Qun Jin. Variational lstm enhanced anomaly detection for industrial big data. *IEEE Transactions on Industrial Informatics*, 17:3469–3477, 2021.
- [7] Frederick J. Riggins and Samuel Fosso Wamba. Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. *2015 48th Hawaii International Conference on System Sciences*, pages 1531–1540, 2015.