Industry 5.0 – technológie: bezpečný prenos, ukladanie a analýza údajov (6)

V tejto časti série sa zameriavame na ďalší technologický koncept, na ktorom je založené Industry 5.0 [1]. Týmto konceptom sú veľké dáta (Big Data), konkrétne rozoberieme bezpečný prenos, ukladanie a analýzu dát. Veľké dáta vyžadujú energeticky účinné a bezpečné technológie na prenos, ukladanie a analýzu s vlastnosťami ako sieťovanie senzorov/akčných členov, interoperabilitu medzi dátami a systémom, škálovateľnosť, viacúrovňovú kybernetickú bezpečnosť v IT infraštruktúre (od snímačov až po cloud), manažment veľkých dát, pôvod dát (traceability), spracovanie dát na strojové učenie, a samozrejme, výpočty na hrane siete a správnu distribúciu výpočtového výkonu. Dobre nastavená distribúcia výpočtového výkonu medzi hranou siete a cloudom napomáha analýze v reálnom čase, ktorá bola veľmi dôležitá pre Industry 4.0 aj pre Industry 5.0.

Veľké dáta v Industry 5.0

Tak ako pred pár desaťročiami bola hybnou silou ekonomiky ropa, dnes sú touto hybnou silou dáta. Preto sú veľké dáta dôležitým technologickým konceptom aj pre Industry 5.0, aj keď hlavné ciele sa rozšírili. Spracovanie dát pri Industry 4.0 bolo zamerané na technológie (technology-driven) a zvyšovanie zisku, pri Industry 5.0 sú to hodnoty (value-driven) [1] a snaha o udržateľný rozvoj, odolnosť, zameranie na človeka/operátora.

Veľké dáta môžu priniesť nové príležitosti a z toho dôvodu sa stali mantrou väčšiny odvetví. V publikácii [2] sa autori zameriavajú na udržateľné a ekonomické reportovanie s využitím veľkých dát, ktoré pokrývajú ekonomické údaje, objemy výroby a informácie o emisiách. Takéto využitie dát podporuje čistejšiu produkciu a zároveň ponúka viac informácií pre vývoj výnosov a zisku. Tieto veľké dáta prinášajú obchodné výhody celej spoločnosti, ak sú dátové dopyty a rozhrania vytvorené tak, aby boli interaktívne, intuitívne a používateľsky prívetivé. Množstvo informácií súvisiacich s prevádzkou, nákladmi, emisiami a dodávateľským reťazcom by sa enormne zvýšilo, ak by sa veľké dáta používali na tento cieľ v rôznych výrobných odvetviach. Je nevyhnutné nájsť relevantné korelácie medzi rôznymi atribútmi a dátami. Správny návrh algoritmu a programovanie sú kľúčom k maximálnemu využitiu veľkých dát. V [2] autori riešia environmentálne aspekty, riadenie nákladovej efektívnosti a návrh procesu zlepšenia pri výrobe papiera. Vďaka tomuto riešeniu výrobca vidí, koľko emisií bude mať objednávka zákazníka na konkrétnom výrobnom stroji s použitými surovinami, čo pomôže pri plánovaní výroby a možno tak znížiť náklady a emisie pri výrobe. Ďalšou výhodou je meranie emisií v reálnom čase. Primárnou myšlienkou väčšiny spoločností je zdôrazniť zákazníkom, že výroba je udržateľná a údaje prezentované zákazníkom a úradom sú spoľahlivé.

Vo všeobecnosti platí, že ak má fungovať interoperabilita medzi dátami a systémom, potrebujeme mať navrhnutú vhodnú architektúru takéhoto riešenia. Táto architektúra musí zahŕňať sieťovanie senzorov a akčných členov, vhodne navrhnutú distribúciu výpočtov medzi hranou siete a cloudom a zabezpečený prenos dát.

Bezpečnosť prenosu dát

Ako sme už spomenuli, bezpečnosť prenosu dát je veľmi dôležitá pre správnu interoperabilitu medzi systémom a dátami, preto by ani táto téma pri spracovaní údajov nemala byť odsúvaná bokom. Architektúra potrebná na správne spracovanie veľkých dát sa skladá z viacerých úrovni, preto je dôležité riešiť viacúrovňovú kybernetickú bezpečnosť z pohľadu IT infraštruktúry, siete, cloudu, hrany siete či fyzickej bezpečnosti. Spoločnosť nechce, aby dochádzalo k útokom z vonku či z vnútra. Každá spoločnosť má citlivé dáta a nechce, aby boli zneužité alebo odstránené. Takže je potrebné zabezpečenie dát a ich prenosu, čo sa vykonáva šifrovaním dát medzi prijímateľom a zariadením posielajúcim dáta pomocou protokolu HTTPS v RESTful API či OPC-UA v priemysle. Tieto dáta môžu následne ostať zašifrované aj na samotnom úložisku. Na bezpečný prenos dát má veľký vplyv správna implementácia hardvérových a softvérových prostriedkov. Zamyslieť sa treba nad správnym zapojením a nastavením smerovačov, prepínačov a hardvérových firewallov, ako aj nad softvérovým zabezpečením koncových zariadení pomocou antivírusových programov a firewallov. Prenos a zabezpečenie dát sa mnohokrát overuje pomocou rôznych simulovaných útokov pri asistencii etických hackerov.

Ďalšou dôležitou časťou je bezpečná škálovateľnosť, či už výpočtov, alebo ukladacieho priestoru, nakoľko rýchlosť pribúdania veľkých dát sa môže v podniku časom meniť. Pre obojstranný šifrovaný prenos dát je dôležité, aby sa výpočty vykonali v presne vymedzenom čase, rovnako je potrebné, aby v ukladacom priestore bol dostatok miesta. Z toho dôvodu je ďalšou dôležitou časťou bezpečnosti ukladanie a manažment dát.

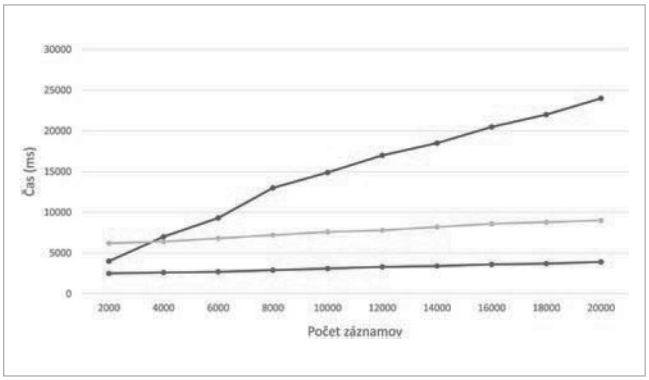
Bezpečné ukladanie a manažment dát

Na sledovanie stavu v priemysle sa neustále zbierali a archivovali dáta rôzneho charakteru. Mnohokrát nie je problém dáta vytvárať a ukladať, ale v prípade potreby je dôležité vedieť sa dopytovať na ne tak, aby sme rýchlo vedeli získať potrebné informácie, ktoré nás zaujímajú. Dopytovacie systémy sú dnes založené na jednej z viacerých základných technológií, ako sú SQL, NoSQL alebo dopyty na základe ontológií. Pri veľkých dátach sa používajú obe technológie: SQL pri štruktúrovaných dátach založených na relačných databázach a NoSQL pri neštruktúrovaných dátach založených na dokumentoch či grafoch.

Veľmi dôležitou časťou bezpečného ukladania dát je vysledovateľnosť pôvodu dát (traceability) a distribuovaný súborový systém na bezproblémovú prácu s dátami. Distribuované súborové systémy sú napríklad HDFS (Hadoop Distributed File System) alebo novší IPFS (InterPlanetary File System), čo je internetový peer-to-peer

protokol na priamu distribúciu dát bez použitia serverov. Problém distribuovaných systémov je, že pôvod dát je ťažko dohľadateľný. Preto sa tu naskytá možnosť využiť distribuovanú decentralizovanú databázu uchovávajúcu neustále sa rozširujúci počet záznamov, ktoré sú chránené proti neoprávnenému zásahu z vonkajšej strany aj zo strany samotných uzlov peer-to-peer siete. Takýmto systémom je blockchain, ktorý uchováva transakcie vykonané používateľmi. Kombinácia s kryptografiou umožňuje zaistiť atomicitu operácií a zabrániť neoprávneným transakciám a dokonalo rieši pôvod dát. Problémom v tomto prípade môže byť výpočtový výkon potrebný pri vykonaní každej transakcie, ale pri postupnom zvyšovaní výpočtového výkonu zariadení sa môže pri zachovávaní pôvodu dát tento systém ujať. V [3] autori predstavili spôsob, ako vhodne doplniť súborový systém IPFS o distribuovanú databázu blockchain, tzv. BlockIPFS. Vďaka tomuto spojeniu môže byť pôvod všetkých dát dohľadateľný aj v distribuovaných systémoch.

Vďaka jedinečným vlastnostiam blockchainu, ako je kybernetická bezpečnosť a sledovateľnosť, má veľký aplikačný potenciál v manažmente veľkých dát, dokáže efektívne vyriešiť mnohé problémy tradičnej správy súborov. Blockchain je však obmedzený svojimi vlastnými nedostatkami, ako je malý úložný priestor a pomalý čas synchronizácie, tiež ho nemožno priamo aplikovať na veľké dátové pole. Práve tieto nevýhody sa riešia napr. v [4, 5], aby sa veľké dáta mohli kombinovať s technológiou blockchain. Na demonštráciu výkonnosti blockchainu sa autori v [4] rozhodli spraviť test (obr. 10), kde modrá čiara reprezentuje záznam s veľkosťou 50 kB uložený do databázy blockchain, červená čiara záznam s veľkosťou 10 kB uložený do tej istej databázy, žltá čiara záznam s veľkosťou 10 kB uložený do databázy blockchain a zvyšných 40 kB do lokálnej databázy. To poukazuje na to, že kým nemáme dostatočný výpočtový výkon na ukladanie všetkých dát pomocou blockchainu, vieme vyriešiť zabezpečenie dôležitejších dát, na základe ktorých sa bude môcť zistiť pôvod aj tých dát, ktoré sú v inom type databáz.



Obr. 10 Demonštrácia výkonnosti technológie blockchain [4]

Samozrejme existujú aj iné technológie, ktoré by dokázali vyriešiť manažment, bezpečnosť a trasovanie pôvodu veľkých dát, ako napríklad metadata či tokenizácia, ale práve nové a bezpečné distribuované súborové systémy (IPFS) a databázy (blockchain) sú z pohľadu bezpečnosti v tejto dobe vhodnými technológiami.

Analýza dát

Aj keď máme nástroje na distribuovanú analýzu veľkých dát pomocou strojového učenia ako Apache Spark, HDFS, Cassandra a podobne, často nevieme odhadnúť, ktoré dáta sú validné na vypracovanie predikčného modelu. Dátoví inžinieri často riešia tento problém takým spôsobom, že modely strojového učenia predtrénujú pomocou menšej vybranej vzorky dát.

Po vytvorení modelu sa model skúša v rôznych nasadeniach a postupne sa dotrénuje podľa potreby situácie. Dotrénovanie predtrénovaných modelov je tzv. transfer learning (prenosové učenie), ktoré sa môže vykonávať viacerými spôsobmi na hrane siete aj v cloude. V súčasnosti je veľmi rozšírený aj trend federated learning (federované učenie) pomocou zariadení na hrane siete. Tento trend vznikol z toho dôvodu, že mnohokrát sa dáta kvôli rôznym reguláciám o súkromí nemôžu len tak zdieľať a posielať medzi dátovými centrami v rámci cloudu. Federated learning funguje na princípe, že každé zariadenie na hrane siete si vytvorí lokálne modely strojového učenia. Tieto lokálne modely, ktoré sú natrénované na hrane siete, sa posielajú do cloudu, kde sa jednotlivé podobné klasifikačné vlastnosti viacerých modelov agregujú. V praxi to znamená, že nemusíme posielať veľa dát do datacentra, ale prakticky si vytvoríme jeden veľký model zoskupovaním viacerých parciálnych modelov natrénovaných viacerými zariadeniami na hrane siete. Tieto zariadenia môžu byť potom využité aj na analýzu dát v reálnom čase, či už ide o deskriptívnu, predikčnú alebo preskriptívnu analýzu. Je viacero postupov, ako k tomu pristúpiť. Jeden zo spôsobov je pripojiť zariadenia na hrane siete k systémom senzorov, z ktorých bude toto zariadenie na základe modelu predikovať alebo klasifikovať dáta, prípadne dáta agregovať a posielať do cloudu, na deskriptívnu analýzu v reálnom čase [6].

Záver

V tejto časti série sme predstavili koncept veľké dáta z pohľadu bezpečného prenosu, ukladania a analýzy ako jeden z podporných konceptov pri realizácii riešení Industry 5.0. V nasledujúcej časti série priblížime ďalšie technológie podporujúce Industry 5.0 tak, ako sú opísané v dokumente, ktorý bol vydaný EK [7]. Konkrétne pôjde o umelú inteligenciu v Industry 5.0.

Poďakovanie
Táto publikácia vznikla vďaka podpore grantu APVV – ENISaC – Edge-eNabled Intelligent Sensing and Computing (APVV-20-0247).

Referencie

[1] Zolotová, Iveta – Kajáti, Erik – Pomšár, Ladislav: Industry 5.0 – koncept, technológie, ciele (1). In: ATP Journal, 2021, roč. 28, č. 11, s. 42 – 43.

[2] Hämäläinen, Esa – Inkinen, Tommi: How to Generate Economic and Sustainability Reports from Big Data? Qualifications of Process Industry. In: Processes, 2017, vol. 5. DOI 10.3390/pr5040064.

[3] Nyaletey, Emmanuel – Parizi, Reza M. – Zhang, Qi – Choo, Kim-Kwang Raymond: BlockIPFS – Blockchain-enabled Interplanetary File System for Forensic and Trusted Data Traceability. In: IEEE International Conference on Blockchain, 2019. DOI 10.1109/Blockchain.2019.00012.

[4] Chen, Jian – Lv, Zhihan – Song, Houbing: Design of personnel big data management system based on blockchain. In: Future Generation Computer System – The International Journal of Escience, 2019, č. 101, s. 1122 – 1129. DOI 10.1016/j.future.2019.07.037.

[5] Karafiloski, Elena – Mishev, Anastas – Karadzinov: Blockchain Solutions for Big Data Challenges A Literature Review. In: 17th IEEE International Conference on Smart Technologies, 2017.

[6] Cao, Hung – Wachowicz, Monica – Cha, Sangwhan: Developing an edge computing platform for real-time descriptive analytics. In: 2017 IEEE International Conference on Big Data, 2017.

[7] European Commission: Industry 5.0 – Towards a sustainable, human-centric and resilient European industry. Directorate-General for Research and Innovation, 01/2021. DOI 10.2777/308407.

doc. Ing. Peter Papcun, PhD.
Ing. Kristián Mičko
Ing. Erik Kajáti, PhD.

Technická univerzita v Košiciach FEI
Katedra kybernetiky a umelej inteligencie
Centrum inteligentných kybernetických systémov
http://ics.fei.tuke.sk