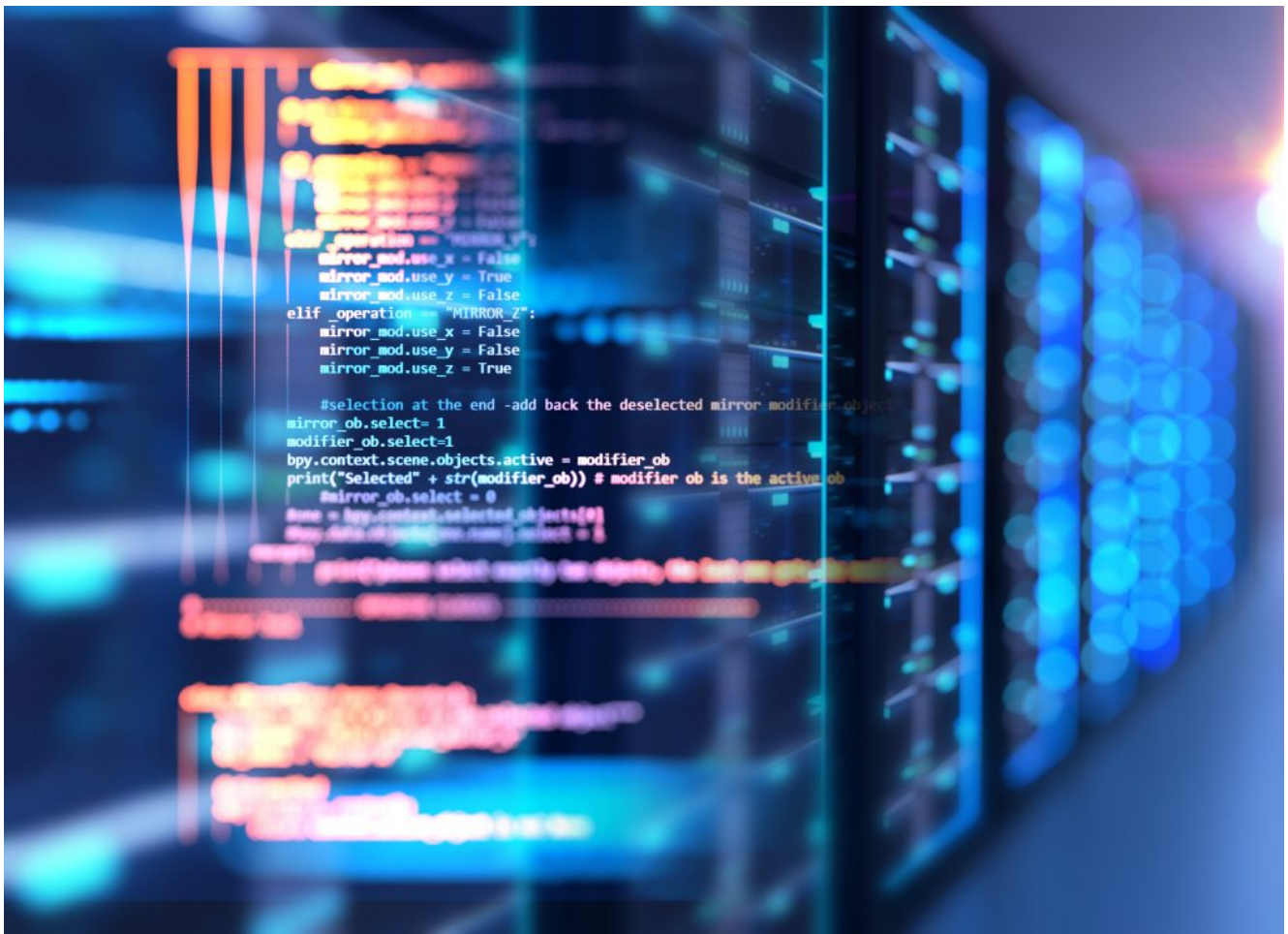


Začiatky BIG DATA

Od autonómnych áut po odhaľovanie podvodov – BIG DATA v T-Systems Slovakia



V roku 1663 sa obchodník s galantériou John Graunt rozhodol zaznamenávať štatistiky týkajúce sa verejného zdravia v Londýne. So svojimi spolupracovníkmi monitoroval epidémiu čierneho moru v meste a na základe získaných dát podnikol prvý pokus o výstražný systém na prevenciu možného rozšírenia nákazy. Okrem toho, že sa tak stal jedným z prvých demografov a epidemiológov, stal sa aj prvým dátovým analytikom.

Množstvo dát, s ktorými John Graunt v 17. storočí pracoval,

bolo na tú dobu ohromujúce. S nástupom 20. storočia sa však množstvo zaznamenaných dát zvýšilo nepredstaviteľným spôsobom – 90 % objemu všetkých dát na svete vzniklo v priebehu posledných pár rokov. Každé dva roky sa tento objem zdvojnásobí. V roku 2018 sa na celom svete denne generovalo 2,5 trilióna bajtov dát. Tieto dáta sú štruktúrované, neštruktúrované či sčasti štruktúrované, preto štatistiky a tradičné relačné databázy na ich spracovávanie už dávno nestačia. V roku 2004 vydal Jefferz Dean a Sanjaz Ghemawat z firmy Google článok o programovacom modeli MapReduce s názvom MAPREDUCE: SIMPLIFIED DATA PROCESSING ON LARGE CLUSTERS, ktorý sa zaoberal prístupom Googlu k zberu a analýze dát z internetových stránok na optimalizáciu vyhľadávania. Google používal MapReduce spolu s Google File System (skratka GFS). O rok neskôr vznikol v rámci open source komunity Apache projekt Hadoop, ktorý je zbierkou open source softvérov a programov, využívajúcich zdroje mnohých počítačov na spracúvanie obrovských množstiev dát v reálnom čase. Jadrom Hadoopu je HDFS (Hadoop Distributed Filesystem) a MapReduce. MapReduce využíva aj firma MapR, ktorá integruje Hadoop do vlastných riešení big data.

Využitie big data

Big data priniesli revolúciu do takmer všetkých odvetví priemyslu a majú vplyv aj na súčasnú kultúru. Sú výsledkom informačnej doby a okrem technológií menia aj ľudské správanie a pohľad na svet. Príklady využitia big data:

- Big data sa používajú v oblasti zdravotníctva v epidemiológii a pri testovaní alternatívnych spôsobov liečby.
- NASA používa big data na prieskum vesmíru.

- Big data sa používajú v oblasti kyberbezpečnosti na zamedzenie kyberzločinu.
- Vo finančnom sektore big data pomáhajú analyzovať riziká investícií a predpovedať vývoj burzy.

Potenciál využitia veľkého množstva dát je nesmierny, uvedomujú si to aj firmy na celom svete. Preto sa v súvislosti s dátami často používa spojenie „Data is the new oil“ (dáta sú novou ropou).

Big data v Košiciach

Na neustále sa zvyšujúci dopyt po riešeniach big data v Európe odpovedá od roku 2015 aj naša firma. Tím ICT inžinierov, administrátorov a architektov zozbiera potrebné vedomosti a know-how a vybuďoval prvé POC (proof-of-concept) klastre, na ktorých si zákazníci mohli vyskúšať aplikovať riešenia big data pre svoj biznis. V priebehu nasledujúcich mesiacov a rokov sme získali viacero zákazníkov, pre ktorých sa stali big data skutočným prínosom a využívajú ich v produkčnom prostredí. Na poskytovanie služieb využívame produkty najväčších hráčov na trhu, ako sú Cloudera, Hortonworks, MapR a pod. Naš tím poskytuje kompletný balík služieb – technický koncept a dizajn infraštruktúry, vybudovanie klastrov (softvér, hardvér, sieť a bezpečnosť), administráciu, optimalizáciu, konzultačné služby a automatizáciu.

Súčasnú a budúce projekty big data u nás zahŕňajú:

- Machine learning na vývoj autonómnych áut

- Monitorovanie stavu súčiastok a pneumatík pomocou senzorov a predikcia ich výmeny
- Predpovedanie meškania vlakov
- Prevenciu podvodov
- Zber a analýzu dát z IP kamier
- Anonymizáciu dát a iné

Väčšina klastrov Hadoop sa štandardne nachádza vo veľkých dátových centrách. Naša spoločnosť sa podieľa aj na netradičnom projekte, ktorým je koncept mobilného dátového centra. To pozostáva z niekoľkých mobilných a centrálnych klastrov, ktoré sú spojené do logickej infraštruktúry. Princíp je takýto: Dáta sa bezdrôtovo prenášajú z testovacích vozidiel do lokálneho mobilného klastra. V rámci takzvaného edge computingu sa predbežné výpočty vykonávajú priamo na mieste a na ďalšie spracovanie už putujú len výsledky zbavené zbytočného balastu. Hlavné výpočty prebiehajú v logickom klastri podľa princípu „prenášaj algoritmus, nie dáta“. Lokálne klastre sú budované s dôrazom na výkon, prevenciu dátových strát a aj s ohľadom na klimatické podmienky danej lokality. Kým všetky mobilné klastre sú prenášateľné, tie najmenšie sú také malé, že sa vojdú do jedného transportného vozidla.

Na záver už možno len dodať, že big data v IT sektore rozhodne nekončia. Implementácia big data veľkými spoločnosťami vzrástla zo 17 % v roku 2015 na 59 % v roku 2018, čím dosiahla

kumulovanú ročnú mieru rastu 36 %. Big data sa tak stali jednou z najrýchlejšie sa rozvíjajúcich technologických oblastí súčasnosti.

Autor: [Viktória Šusterová, T-Systems Slovakia](#)