

Spracovávanie dát v reálnom čase pomocou in-memory databáz

Motivácia k výberu témy bakalárskej práce

Vedúci práce: Ing. Miron Kuzma PhD.

Tomáš Kekeňák

Matematike som sa venoval počas celej základnej aj strednej školy. Zúčastňoval som sa rôznych súťaží, seminárov, aj táborov. Vždy ma ale fascinovalo, ako moderné technológie menia náš svet a aké problémy sa dajú vyriešiť pomocou informačných technológií. To ma postupne priviedlo k štúdiu informatiky.

Dlho som rozmýšľal, akú tému si mám vybrať na bakalársku prácu. Nakoľko s tém na burze ma neoslovila žiadna, rozhodol som sa, že si vymyslím novú. Cez môjho otca som sa dostal k novým zaujímavým technológiám od firmy SAP, ktoré vzbudili moju pozornosť na prvý pohľad. Je to najmä ich nová revolučná in-memory relačná databáza SAP HANA. Odlišuje sa od bežných databáz jednak tým, že všetky dáta sú v operačnej pamäti počítača a tým, že dáta sú uložené v stĺpcovom a nie riadkovom formáte, čo prispieva masívnemu zrýchleniu agregáčnych dopytov. SAP HANA nie je len databáza, ale celá platforma, ktorá prináša so sebou množstvo zaujímavých nových funkcionalít. Jednou z nich je SAP HANA Smart Data Streaming, ktorá slúži na spracovávanie streamovaných dát. Práve táto funkcionalita ma zaujala.

V poslednom čase internet vecí naberá na popularite, a čím ďalej, tým viac senzorov sa inštaluje pomaly do všetkých vecí, ktoré poznáme. To prináša ale aj množstvo nových problémov, lebo počet dát na spracovanie rastie obrovským tempom, a postupne nebudeme vedieť spracovávať tie dáta štandardnými metódami. Rastie aj sieťová prevádzka, a veľké firmy ako Google, alebo Facebook monitorujú naše správanie na webe, čo zase prináša zo sebou obrovské množstvo dát. Pre veľa aplikácií je žiadané, aby tieto dáta boli spracované v reálnom čase, napr. vzdialené monitorovanie prevádzky, monitorovanie sieťových útokov, kde by sme chceli hneď identifikovať, keď nastane problém. Toto nás vedie k tomu, aby sme dáta spracovali prúdovo, teda hneď ako ich dostaneme, ešte predtým, ako ich uložíme v databáze, aby sme vedeli zakročiť v čas a vyhnúť sa väčším škodám. Práve tejto oblasti, rýchlemu spracovávaní dát, ich analýze a generovaniu notifikácií v prípade problémov v reálnom čase sa budem venovať vo svojej práci.