

Analýza softvérových nástrojov pre Big data

Abstrakt práce:

Predmetom skúmania práce sú softvérové nástroje pre prácu s Big data. Cieľom práce je analyzovať momentálne dostupné softvérové nástroje vhodné na spracovanie Big data a v jednom z týchto nástrojov vytvoriť vlastnú aplikáciu, ktorá bude analyzovať dostupné údaje.

V teoretickej časti práca obsahuje analýzu a zhodnotenie momentálnej situácie v oblasti ukladania a spracovania údajov. Z výsledkov analýzy vyplýva potreba hľadania nových a účinnejších softvérových nástrojov pre prácu s rýchlo pribúdajúcimi údajmi. Momentálne dostupné softvérové nástroje pre prácu s Big data sú rozdelené do základných kategórií a najpoužívanejšie z nich sú predstavené. Teoretická časť práce ďalej obsahuje podrobnejší popis systému Hadoop, ktorý je momentálne najdôležitejším systémom v oblasti spracovávania Big data.

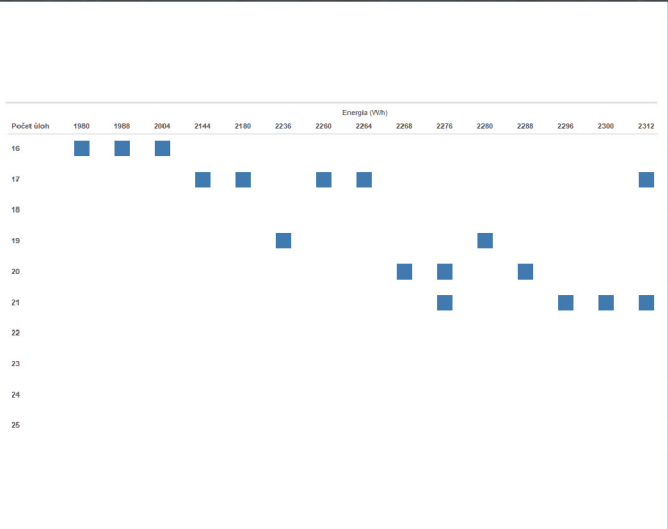
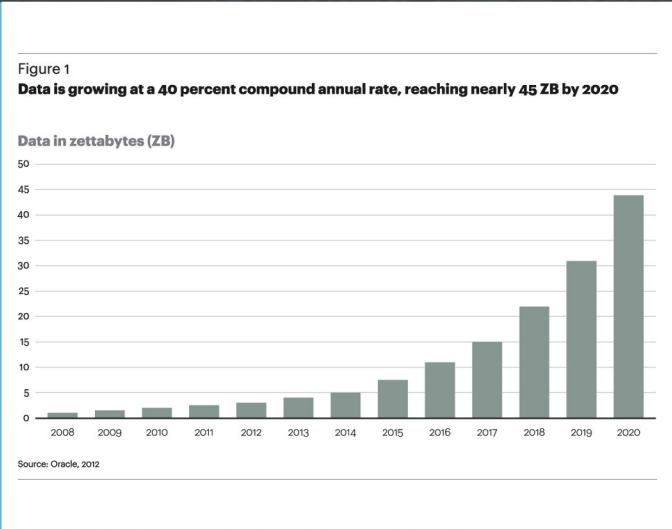
Praktická časť práce sa zaoberá analýzou, návrhom a implementáciou aplikácie pre hľadanie takého rozloženia a rozvrhovania úloh na vysokovýkonnom výpočtovom klastri, ktoré vedie k čo najnižšej spotrebe elektrickej energie za pomoci systému Hadoop. V závere praktickej časti sa nachádza zhrnutie výsledkov, ktoré vyprodukovala implementovaná aplikácia. Analyzované údaje pochádzajú z monitorovacích nástrojov vysokovýkonného výpočtového klastra, ktorý sa nachádza na pôde Univerzity Mateja Bela v Banskej Bystrici. Predpokladaným prínosom práce je zoznam momentálne dostupných softvérových nástrojov použiteľných pre prácu s Big data a takisto skripty a programy použiteľné v prostredí systému Hadoop, ktoré prispievajú k hľadaniu takého rozloženia a rozvrhovania úloh na vysokovýkonnom výpočtovom klastri, ktoré vyžaduje najnižšiu spotrebu elektrickej energie.

Kľúčové slová:

Big data. Hadoop. Vysokovýkonný výpočtový klaster. Spotreba energie. Analýza záznamov.

Komentár vedúceho práce:

„Ide o komplexnú prácu, ktorá obsahuje rešeršný, výskumný a realizačný aspekt, vrátane vyhodnotenia riešenia. Prínos práce je špecifikovaný vo vytvorenej metodike, ktorá obsahuje postupy, programy a skripty, umožňujúce použiť nástroje pre mnohorozmerné údaje pre spätnú rekonštrukciu rozvrhu, spolu so zodpovedajúcou spotrebou v jednominútových intervaloch. Prínosom sú tiež výstupy, ktoré je možné použiť na priamu analýzu, alebo ukazujú smer, ktorým sa treba uberať pri ďalších analýzách. Navrhnuté metódy je možné použiť pre komplexnejšie systémy a tiež pre spätné skúmanie veľkého časového úseku.“



Autor práce: Martin Trník
Vedúci práce: doc. Ing. Jarmila Škrinárová, PhD.