



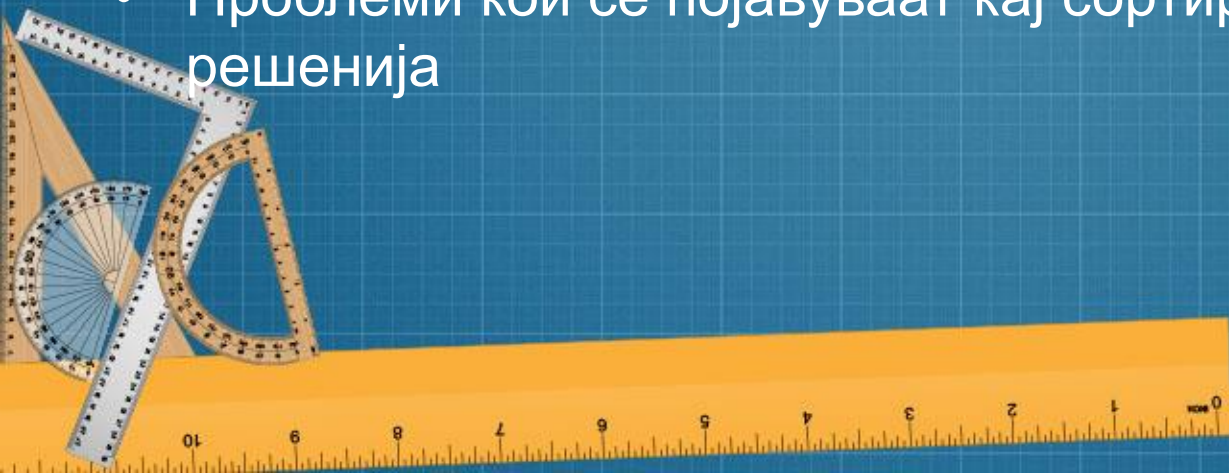
АЛГОРИТМИ ЗА СОРТИРАЊЕ НА МАСИВНИ БАЗИ НА ПОДАТОЦИ

Дејан Каракашов
dejan.karakashov@gmail.com
SkyOptical LLC, Dubai

АЛГОРИТМИ И ПРИМЕНА 10.12.2016 год, ПМФ-Скопје

Масивни бази на податоци

- Што се масивни бази на податоци?
- Карактеристики на масивни бази на податоци
- Можности на масивните бази на податоци
- Технологии со кои се развиваат системи за менаџирање на масивни бази на податоци
- Структури на податоци со кои се претставуваат масивните бази на податоци
- Анализа на сортирање кај масивни бази на податоци
- Проблеми кои се појавуваат кај сортирањето и можни решенија



Што се масивни бази и нивни карактеристики

- Големи и комплексни множества на податоци
- 3v's model of Big Data
 - ❑ data volume(обем на податоци)
 - ❑ data velocity(брзина на генерирање на податоци)
 - ❑ data variety(спектар на податоци)





Можности на масивните бази на податоци

- Параметарско пребарување или насочена навигација (faceted search at scale)
- Multimedia search
- opinion mining and sentiment analysis
- Збогатување на базата на податоци (Database Enrichment)
- Прелиминарна анализа
- Оперативна анализа



Технологии

- Паралелно процесирање и дистрибуирани архитектури
- Флексибилни податочни модели(NoSql,NewSql, Search-based platforms)
- Кеширање и процесирање во меморијата(in-memory processing)

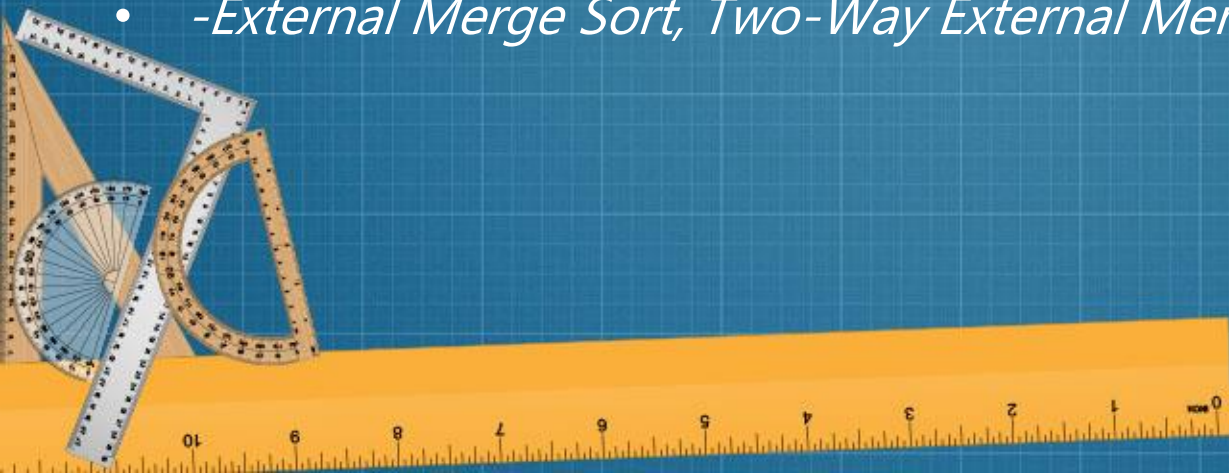
Структури на податоци за претставување на масивни бази на податоци

- Synopsys структури на податоци
 - брзо процесирање
 - брз пренос на податоци
 - помала цена
 - подобри перформанси
- Multi-way search trees



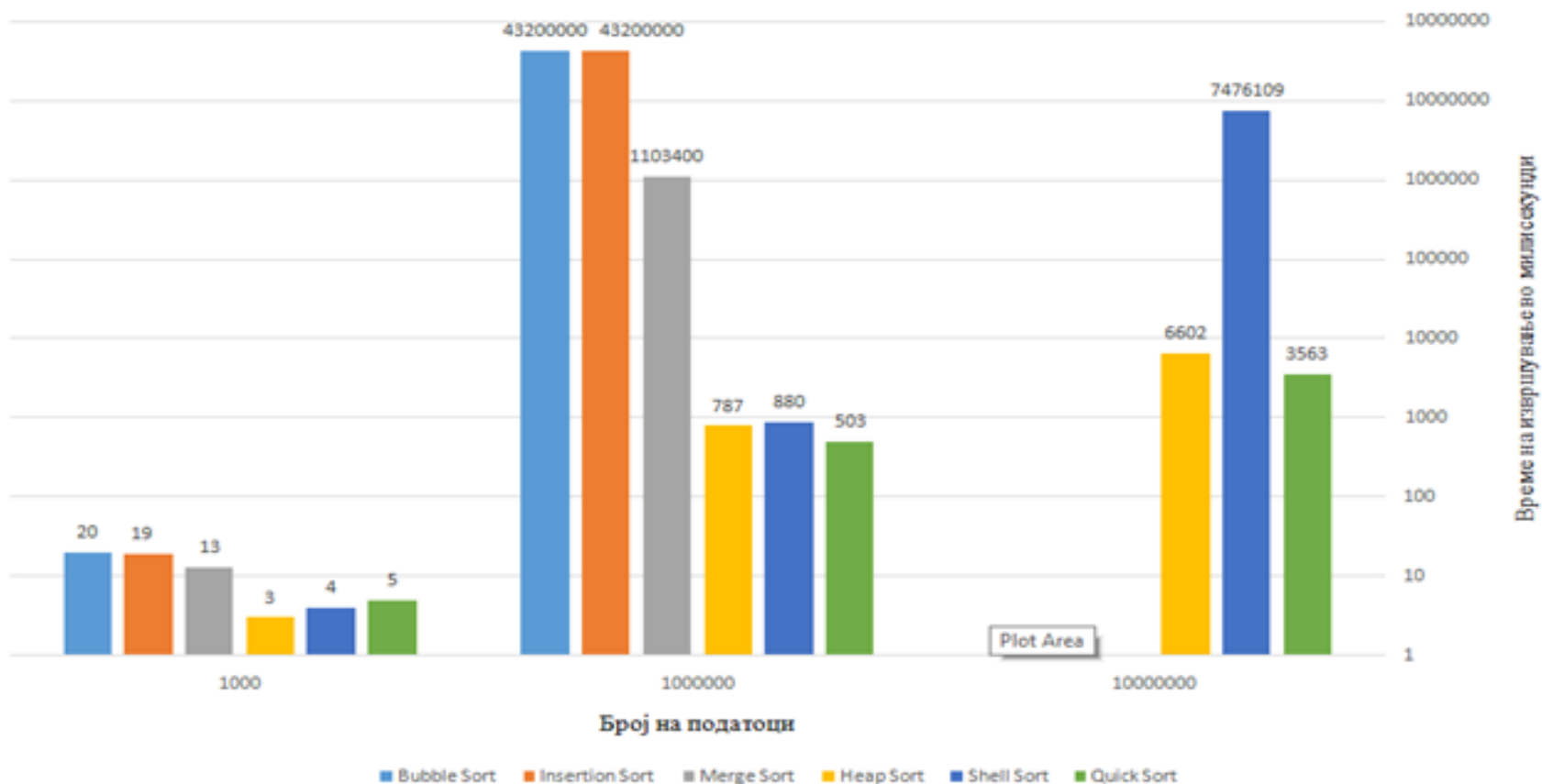
Сортирање кај масивни бази на податоци

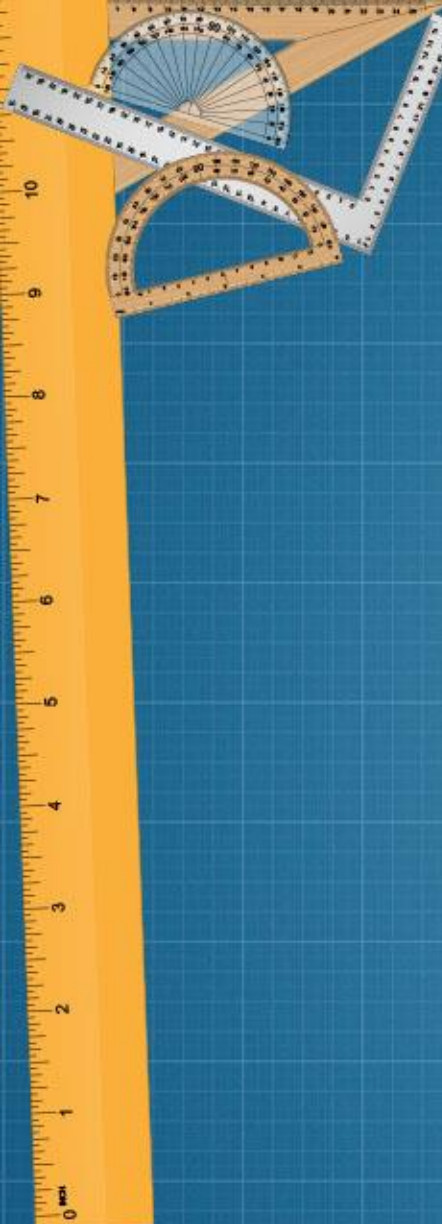
- Алгоритми за внатрешно сортирање (Insertion Sort, Bubble Sort, Merge Sort, Quick Sort, Heap Sort, Shell Sort)
- *Сортирање во меморијата*
- *Неадекватни алгоритми за масивни бази на податоци*
- *Потребни се модификации*
- Алгоритми за надворешно сортирање
- *-sort-merge метод на сортирање*
- *-External Merge Sort, Two-Way External Merge Sort...*



Анализа на сортирањето

Споредба на алгоритми за сортирање изразена во милисекунди





Проблеми кои се појавуваат при сортирањето на масивни бази на податоци

- Недоволно меморија за да се извршат толкав број на инструкции
- Несоодветна компјутерска конфигурација
- Сортирање според повеќе колони
- Неоптимални алгоритми за сортирање



Можни решенија

- Алгоритми за надворешно сортирање
- Компресирање на датотеките што треба да се сортираат
- Индексирање на клучевите

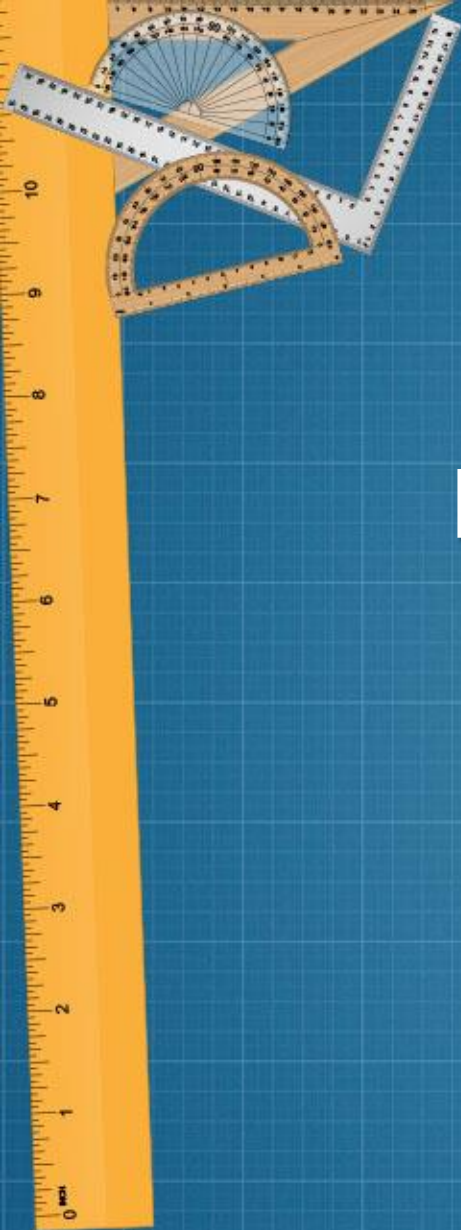


Заклучок

Денес сме длабоко навлезени во ерата на масивните бази на податоци.

Потенцијалот за зголемување на профитабилноста и успехот кај претпријатијата постојано се зголемува поради поголемата достапност на масивните бази на податоци.

Сортирањето е еден од главните предизвици, па затоа имаме постојан развој на нови технологии, структури и алгоритми за сортирање на масивни бази на податоци.



Благодарам на вниманието

Прашања?