Sorting Exercise

Jonas Schenke
28.01.2019

Concept

1.Stream data in:

Read Data	Read Data	Read Data	Read Data	
	Quicksort	Quicksort	Quicksort	Quicksort
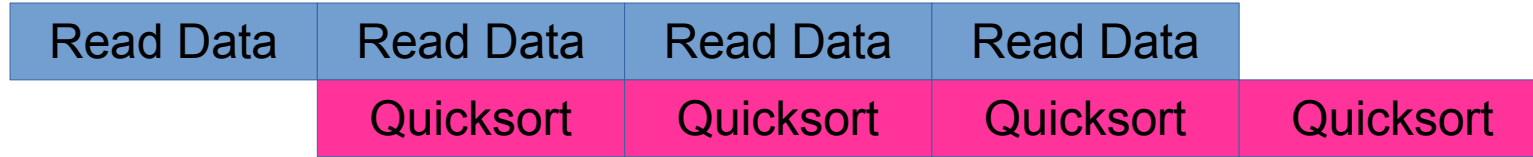
2.Synchronize threads

3.Stream data out:

Merge Lists	Merge Lists	Merge Lists	Merge Lists	
	Wirte Block	Wirte Block	Wirte Block	Wirte Block

4.Flush file

Current Implementation

- Implementation in C++

- Everything loaded to RAM (has to fit!)

- Each Quicksort executed in seperate thread

- Select smallest element from lists

    $\rightarrow$ create output blocks for streaming

- Single threaded merging

Configuration

- Block size ~ 610MiBi

- Input file size ~1.9GiBi (20 mio. Entries)

- Test System:
  - Ryzen 1700 (16*3.2GHz)
  - 16 GiBi RAM
  - 240GB Corsair MP500 NVMe SSD (read:~3GB/s)

    ==> ~10s runtime / 2 mio. entries/s

Further Notes

- Time for input * 2 = (ca.) time for output

- ~140% CPU usage
  - Input multithreaded
  - Output single threaded → improvable

- Potential speedup with parallel input sorting algorithm

- Merging algorithm improvable

- Code adaptable using template metaprogramming