

# ALL AT ONCE CLONE DETECTOR

PA2577 | Alexander Jarnstrom | aljr21@student.bth.se

## I. HOW TO RUN

I don't like to make things easier for me so i decided to use C++ for this with FTXUI<sup>1</sup> and MangoCXX<sup>2</sup>, it was a fun experience but a terrible choice. Making the resulting graphs lovely to look at but not very precise.

The way to run it is a bit unique as well, there is a make command to run it with the corresponding services:

```
make monitor
```

This will run the following script:

```
cljDetector: corpusVolume
@docker compose -f all-at-once.yaml up -d
clear

monitor: cljDetector
@docker run \
  -eDBHOST=dbstorage \
  -eOUTPATH=/out/out.txt \
  --network big_data_analytics_clones \
  --volume ./out:/out\
  -it monitor-tool:latest
```

<sup>1</sup><https://github.com/ArthurSonzogni/FTXUI>

<sup>2</sup><https://github.com/mongodb/mongo-cxx-driver>

but you can't do it just with `docker compose`. Why? You may ask, the monitoring tool need access to a pseudo-tty and stdout, using the `-it` flags. These only come with `docker run` and `docker exec`, `docker compose` does have a feature to toggle these on but it looks horrendous.

## II. QUESTIONS

The time to generate chunk is fluctuating with some major hiccups in between, Figure 1. This simply shows the difference in processing time for larger files. The results make sense thanks to the fact it's just walking through the files separating them and then pushing it onto the database, meaning the most congested part is probably the database. The process reached a total of 18,211,863 chunks over 9,600 files. Giving us an average of 1,897 chunks per file, which is a valid excuse to stall sometimes.

The candidates are all loaded and added to the database in one sweep shown in the gap between database events, Figure 2.

The number of candidates processed per second slowly fell off, giving us the graph in Figure 3, meaning the time per processing went up for every iteration. The statistics show this as well, where the first one took 1ns while the last one took 2754ns giving us an exponential growth. This can be seen in Figure 3 where each row

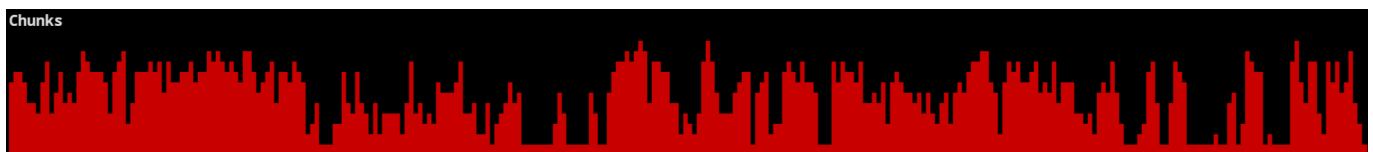


Figure 1: The number of processed chunks every second.

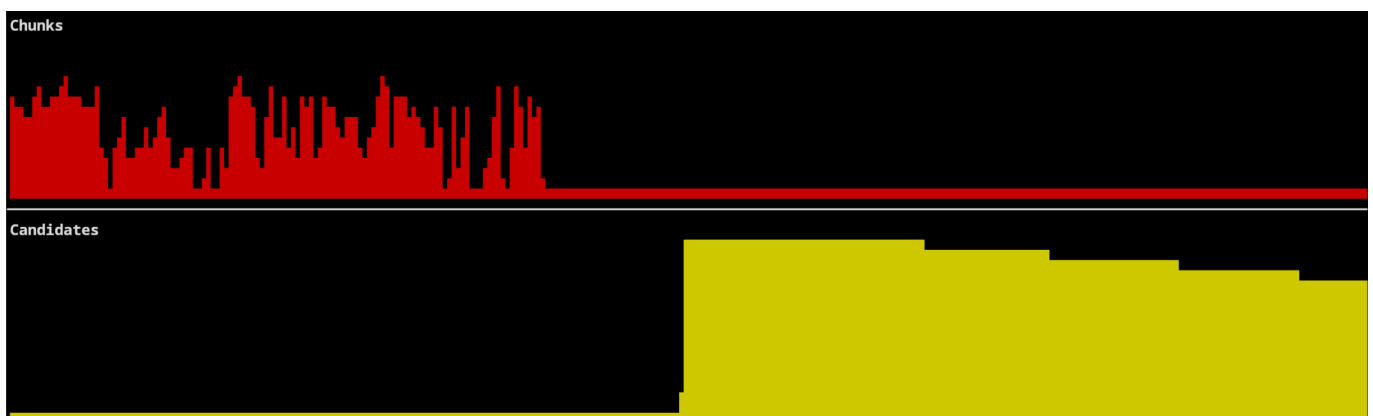


Figure 2: Visualization of the gap between chunk loading and candidates getting added.

get slightly longer for each step. This makes sense, the first one will just be placed, the second one will attempt to stretch over the ones which overlap, and so on. This means every candidate has to be compared to every other candidate giving us this growth in time.

The clones shown in Figure 4 doesn't seem like a lot, and that's mostly thanks to the another less then optimal choice, show the whole timeline for all. This means, to fit the whole graph in such a small frame, an average over segments was made. And a clone is registered every one to two seconds in the beginning before ramping up, means they are heavily averaged out by the hundreds of zeros around them, showing us this cute peak with the highest jump of 208 clones. Though we got the average number of rows per clone, which ended up on 43.

### III. FUTURE IMPROVEMENTS

---

The monitoring tool need plenty of fixes, the first result in every graph is an overflow creating a vertical bar when the program runs. Using virtual classes for the statistics would have been a way cleaner solution and save many branch misses. Using C++ with FTXUI was a fun experience, should one do it on an assignment with a deadline one day earlier than you thought, probably not. Improve planning and structure would make it simpler to complete the task.



Figure 3: The number of remaining candidates per second.



Figure 4: Number of found clones per second, so zoomed out it merged with the zeros

## IV. APPENDIX

---

The assignment mentions a 100 odd lined cut out of the statistics, it saves the stats every two seconds to a file, these are already the odd lined statistics due to the monitor updateing every second. This means there's a total of 7300 lines of statistics, *all on odd lines*. And the first part is just the chunk part, which leads us to this, 50 line from the beginning and 50 line from where the candidates processing started.

There are several shortend names used in the statistics:

- `f[:]`: file category.
- `cu[:]`: chunk category.
- `ca[:]`: candidate category.
- `cl[:]`: clone category.
- `to`: total count.
- `c`: change from last.
- `ti`: average time per object in nanoseconds.
- `cf`: deprecated, used for file chunk average.
- `in`: instances per clone, changed to average row count per clone.

### IV.I. FIRST 50

```
3: f:[to:4300] cu:[to:0 c:0 ti:0 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
5: f:[to:8600] cu:[to:0 c:0 ti:0 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
7: f:[to:9600] cu:[to:11200 c:7000 ti:142 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
9: f:[to:9600] cu:[to:35700 c:11300 ti:88 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
11: f:[to:9600] cu:[to:42800 c:0 ti:140 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
13: f:[to:9600] cu:[to:42800 c:0 ti:140 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
15: f:[to:9600] cu:[to:49600 c:6800 ti:735 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
17: f:[to:9600] cu:[to:70800 c:16200 ti:61 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
19: f:[to:9600] cu:[to:92600 c:17100 ti:58 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
21: f:[to:9600] cu:[to:92600 c:0 ti:58 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
23: f:[to:9600] cu:[to:103000 c:10400 ti:384 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
25: f:[to:9600] cu:[to:113600 c:6800 ti:147 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
27: f:[to:9600] cu:[to:136100 c:13100 ti:76 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
29: f:[to:9600] cu:[to:162600 c:15600 ti:64 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
31: f:[to:9600] cu:[to:179500 c:4000 ti:249 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
33: f:[to:9600] cu:[to:201000 c:11000 ti:90 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
35: f:[to:9600] cu:[to:222800 c:9700 ti:103 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
37: f:[to:9600] cu:[to:242200 c:8000 ti:125 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
39: f:[to:9600] cu:[to:268200 c:12200 ti:81 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
41: f:[to:9600] cu:[to:290300 c:11900 ti:83 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
43: f:[to:9600] cu:[to:311600 c:12500 ti:79 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
45: f:[to:9600] cu:[to:334200 c:12400 ti:80 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
47: f:[to:9600] cu:[to:352300 c:11300 ti:88 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
49: f:[to:9600] cu:[to:365400 c:8400 ti:118 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
51: f:[to:9600] cu:[to:382300 c:6800 ti:146 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
53: f:[to:9600] cu:[to:408400 c:14400 ti:69 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
55: f:[to:9600] cu:[to:420800 c:5500 ti:181 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
57: f:[to:9600] cu:[to:429500 c:8700 ti:229 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
59: f:[to:9600] cu:[to:449600 c:11400 ti:87 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
61: f:[to:9600] cu:[to:467000 c:7100 ti:140 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
63: f:[to:9600] cu:[to:478400 c:6700 ti:149 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
65: f:[to:9600] cu:[to:494100 c:7800 ti:128 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
67: f:[to:9600] cu:[to:503500 c:0 ti:106 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
69: f:[to:9600] cu:[to:517700 c:4500 ti:222 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
71: f:[to:9600] cu:[to:522100 c:0 ti:227 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
73: f:[to:9600] cu:[to:524100 c:2000 ti:1498 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
75: f:[to:9600] cu:[to:542500 c:13100 ti:76 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
77: f:[to:9600] cu:[to:556600 c:5300 ti:188 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
79: f:[to:9600] cu:[to:562000 c:700 ti:1428 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
81: f:[to:9600] cu:[to:581600 c:11300 ti:88 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
83: f:[to:9600] cu:[to:603800 c:15000 ti:66 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
85: f:[to:9600] cu:[to:626400 c:10200 ti:98 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
87: f:[to:9600] cu:[to:657200 c:16900 ti:59 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
89: f:[to:9600] cu:[to:684400 c:14700 ti:67 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]
```

91: f:[to:9600] cu:[to:712500 c:15600 ti:64 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]  
93: f:[to:9600] cu:[to:728700 c:8200 ti:121 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]  
95: f:[to:9600] cu:[to:748100 c:7900 ti:126 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]  
97: f:[to:9600] cu:[to:767200 c:7200 ti:138 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]  
99: f:[to:9600] cu:[to:771200 c:0 ti:249 cf:0] ca:[to:0 ti:0] cl:[to:0 c:0 ti:4294911728 in:0]

## IV.II. SECOND 50

3029: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:586086 ti:5165] cl:[to:0 c:0 ti:4294911728 in:0]  
3031: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:585895 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3033: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:585705 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3035: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:585477 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3037: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:585287 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3039: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:585097 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3041: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584907 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3043: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584780 ti:1] cl:[to:0 c:0 ti:4294911728 in:0]  
3045: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584740 ti:1] cl:[to:2 c:1 ti:998817 in:2]  
3047: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584644 ti:1] cl:[to:3 c:1 ti:1998972 in:2]  
3049: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584454 ti:1] cl:[to:3 c:0 ti:1998972 in:2]  
3051: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584351 ti:1] cl:[to:3 c:0 ti:1998972 in:2]  
3053: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584191 ti:1] cl:[to:4 c:0 ti:4999001 in:2]  
3055: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:584001 ti:1] cl:[to:4 c:0 ti:4999001 in:2]  
3057: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583811 ti:1] cl:[to:4 c:0 ti:4999001 in:2]  
3059: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583653 ti:1] cl:[to:5 c:1 ti:6998490 in:2]  
3061: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583561 ti:1] cl:[to:6 c:1 ti:1999899 in:2]  
3063: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583500 ti:1] cl:[to:7 c:1 ti:1999873 in:2]  
3065: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583456 ti:1] cl:[to:9 c:1 ti:999828 in:2]  
3067: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583393 ti:1] cl:[to:10 c:1 ti:2000089 in:2]  
3069: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583240 ti:1] cl:[to:11 c:0 ti:999133 in:2]  
3071: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583115 ti:1] cl:[to:11 c:0 ti:999133 in:2]  
3073: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:583020 ti:1] cl:[to:11 c:0 ti:999133 in:2]  
3075: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:582925 ti:1] cl:[to:11 c:0 ti:999133 in:2]  
3077: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:582829 ti:1] cl:[to:12 c:1 ti:8998781 in:2]  
3079: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:582723 ti:1] cl:[to:13 c:1 ti:1998919 in:2]  
3081: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:582533 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3083: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:582343 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3085: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:582117 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3087: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581958 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3089: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581844 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3091: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581749 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3093: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581664 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3095: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581557 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3097: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581462 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3099: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581367 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3101: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581253 ti:1] cl:[to:13 c:0 ti:1998919 in:2]  
3103: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581129 ti:1] cl:[to:14 c:0 ti:22989698 in:2]  
3105: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:581034 ti:1] cl:[to:14 c:0 ti:22989698 in:2]  
3107: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580941 ti:1] cl:[to:15 c:1 ti:4996495 in:2]  
3109: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580815 ti:1] cl:[to:16 c:0 ti:999763 in:2]  
3111: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580720 ti:1] cl:[to:16 c:0 ti:999763 in:2]  
3113: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580606 ti:1] cl:[to:16 c:0 ti:999763 in:2]  
3115: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580511 ti:1] cl:[to:16 c:0 ti:999763 in:2]  
3117: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580418 ti:1] cl:[to:16 c:0 ti:999763 in:2]  
3119: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580302 ti:1] cl:[to:18 c:1 ti:999935 in:2]  
3121: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580156 ti:1] cl:[to:19 c:0 ti:1000120 in:2]  
3123: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580063 ti:1] cl:[to:20 c:0 ti:1998023 in:2]  
3125: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:580023 ti:1] cl:[to:21 c:0 ti:1999329 in:2]  
3127: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:579970 ti:1] cl:[to:23 c:1 ti:999104 in:2]  
3129: f:[to:9600] cu:[to:18211863 c:0 ti:2160 cf:0] ca:[to:579923 ti:1] cl:[to:25 c:1 ti:998958 in:2]