# Project 0.1 External Hash

資工四 404410076 陳鐸元

# Arguments

- -m: memory limit in bytes
- -h: number of entries in the hash table
- -s: largest key size

#### Structure

- Input: read and parse all files
- hInsert: hash and store the record
  - · After hitting memory limit, write to a temporary file
- Output: iterate through each list in each entry
- Sort: use HW1's record sort to sort it
  - ../HW1/record\_sort -d \$'\n'\$'\n' -sd \$'\n' -r -n ../HW0/tcount.out -o ../HW0/tout

#### Benchmark

# Computer Hardware

- CPU: i7 6700HQ (4C8T)
- Memory: 16GB (2x8GB)
- Storage: 512Gb M.2 PCIE Gen 3 x4 SSD (Plextor M8pegn)

#### **Test Data**

1 GB news records

#### In-memory only

• 51s

```
100 2018/11/26 20:25:51
X130[danny@loc

ightharpoonup time ./tcount v1
                 ettoday1.rec
file ettoday0.rec
file ettoday1.rec
                  ettoday39 rec
    ettoday2.rec
                  ettoday5.rec
file ettoday3.rec
                        output
file ettoday4.rec
file ettoday5.rec
                      /HW0/tcount.out
       0m51.859s
real5
       0m45.404s
user
       0m2.961s
SYS
```

#### In-memory

• 4m7s

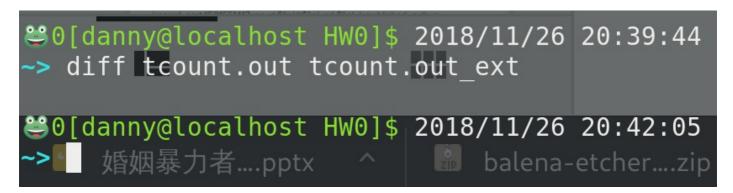
```
2018/11/24 23:38:24
-> ./tcounts_m_3000000000 ettoday0.rec ettoday2.rec ettoday4.rec ettoday1.rec ettoday3.rec ettoday5.rec && ../HW1/record_sort -d $'\n'$'\n' -sd $'\n' -r -n ../HW0/tcount.out -o ../HW0/tout && less tout
file ettoday0.rec
file ettoday2.rec
file ettoday4.rec
file ettoday1 rec
file ettoday3.rec
file ettoday5.rec
start yo output
file: __/HW0/tcount.out
thread_token: 0
thread_token: 1
thread token: 2
Finished reading file: ../HW0/tcount.out
Finished internal sort
start: 0, ccnt: 3, target: 3, final: ../HW0/tout
stream 0 Finished
stream 1 Finished
stream 2 Finished
Bye~
₩0[dannv@localhost HW0]$ 2018/11/24 23:44:3
```

# Limited-memory

```
≌0[danny@localhost HW0]$ 2018/11/24 23:44:31
  > ./tcount/ m 1700000000 ettoday0.rec ettoday2.rec ettoday4.rec ettoday1.rec ettoday3.rec ettoday5.re && ../HW1/record_sort -d $'\n'$'\n' -sd $'\n' -r -n ../HW0/tcount.out -o ../HW0/tout && less tout/
filerettodayo.recdanny danny 596449659 Nov 26 15:45
file ettoday2.recdanny danny
file ettödaya rec-channel mp.bin
3 -rw-rw-r--. 1 danny danny 1146792061 Nov 26 17:23
file ettoday1.rec
4 -rw-r---- 1 danny danny 1526743552 Sep 13 18:08 chromeos_10895.56.0_peppy_recofileeettoday3:reconnel_mp-v3.bin
5 -rw-rw-r--. 1 danny danny 613955837 Nov 26 11:57 chromeos_10895.56.0_peppy_reco
file ettoday5.rec
start√yō oūtput1 danny danny 8751432192 Nov 26 17:42 chromiumos image.img
file: 0.../HW0/tcount.out_hrome]$ 2018/11/26 17:44:41 thread_token: 0
thread token: 1
thread_token: 2
Finished reading file: ../HW0/tcount.out
Finished internal sort
start: 0, ccnt: 3, target: 3, final: ../HWO/tout
stream 0 Finished
stream 1 Finished /HWO/tcount.out
stream 2 Finished token: 0
Bye~

    @0[danny@localhost HW0]$ 2018/11/24 23:50:16
```

#### Validation



#### Code

https://github.com/Superdanby/Search-Engine/tree/master/Project%204