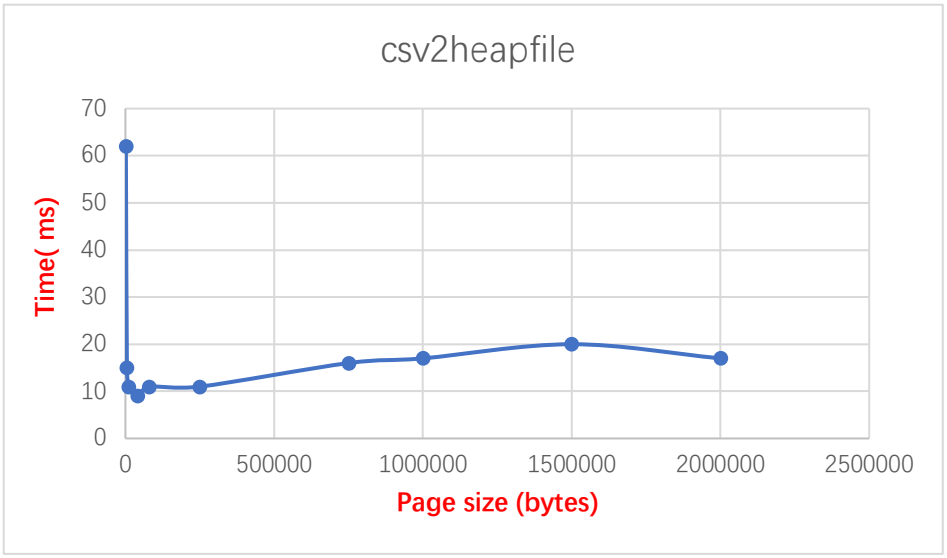
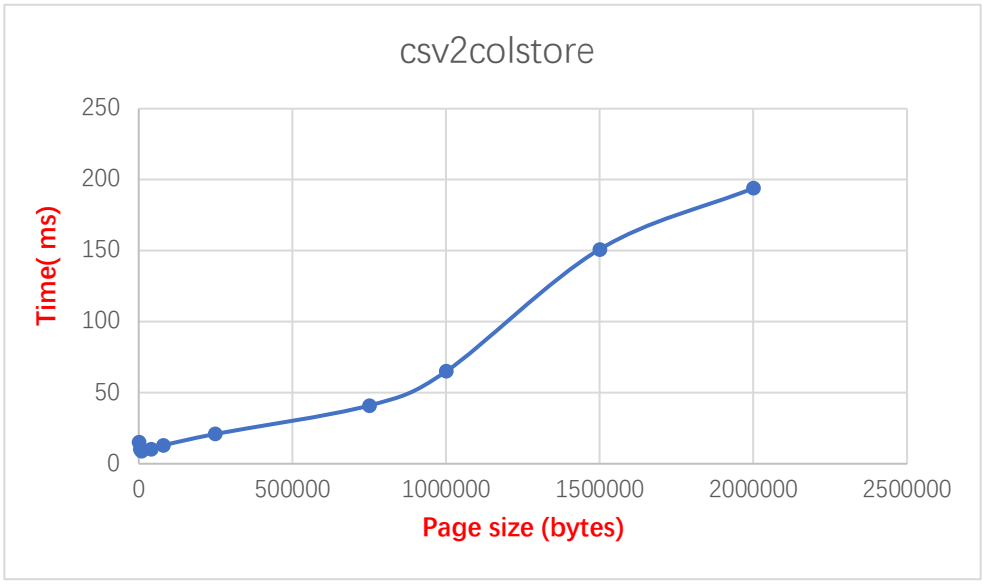
Report for part3: csv2colstore & select2 & select3

Total records: 1000 records
Script name: csv2colstore.sh & select2&3.sh

1)



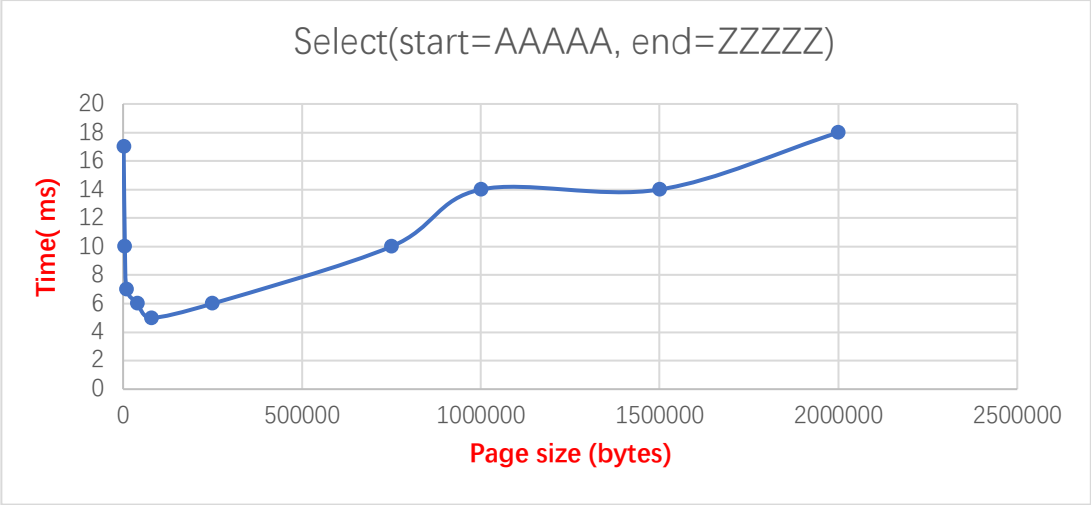


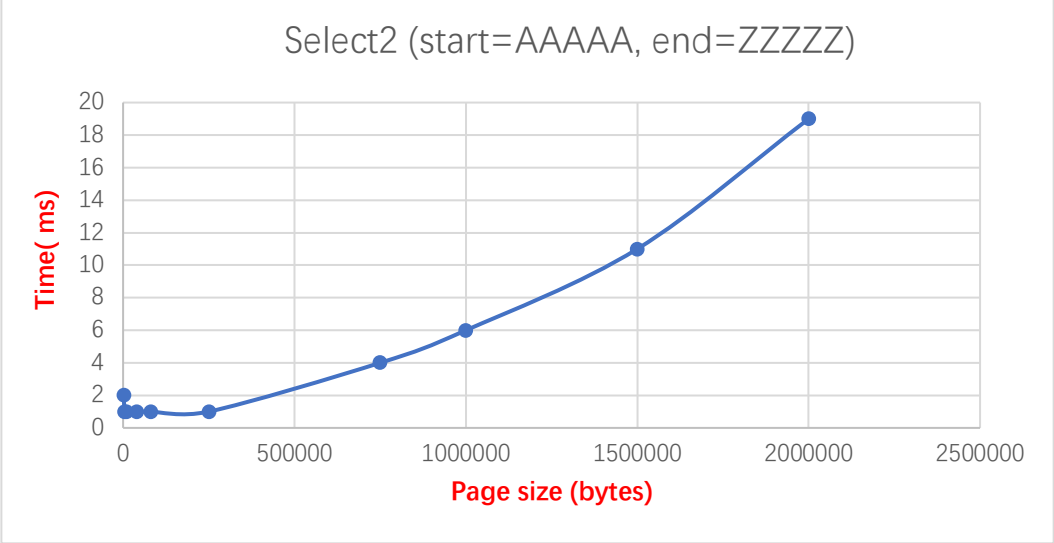
Comment:
It takes much longer time to create the column store heap file than creating row store file. And as the page size grows, the runtime will increase as well. The CSV file generated is stored as row-store format. Turning a row record takes extra time and memory to create column store, and the result shown that.

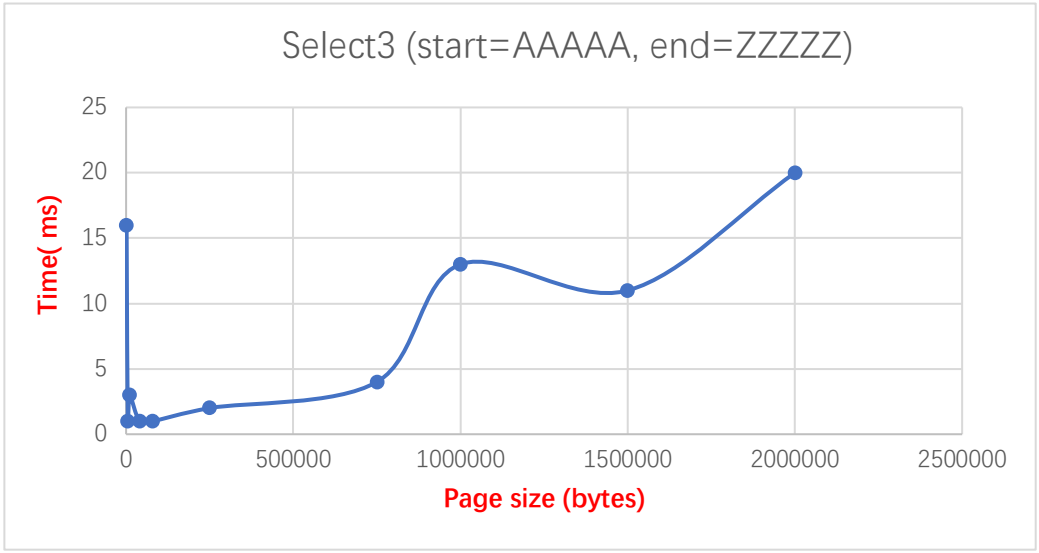
2) Select query vs Page size (select vs select 2 vs select3)

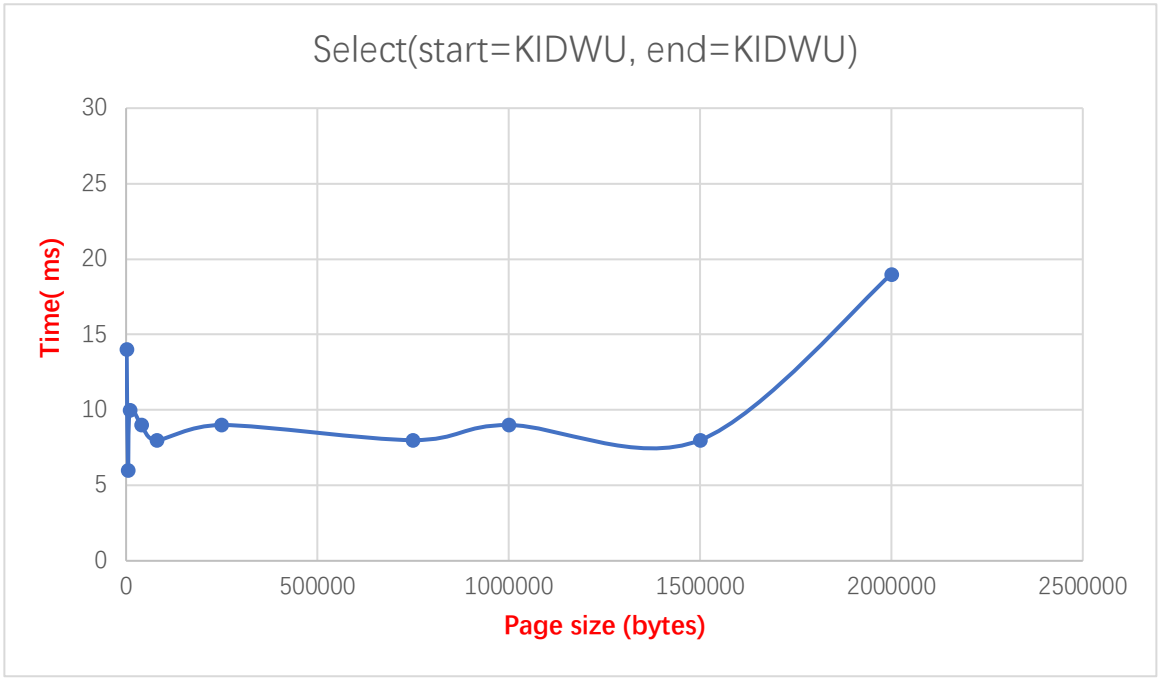
Performace graph without printing the query result.
(Query result will be printed in the program)

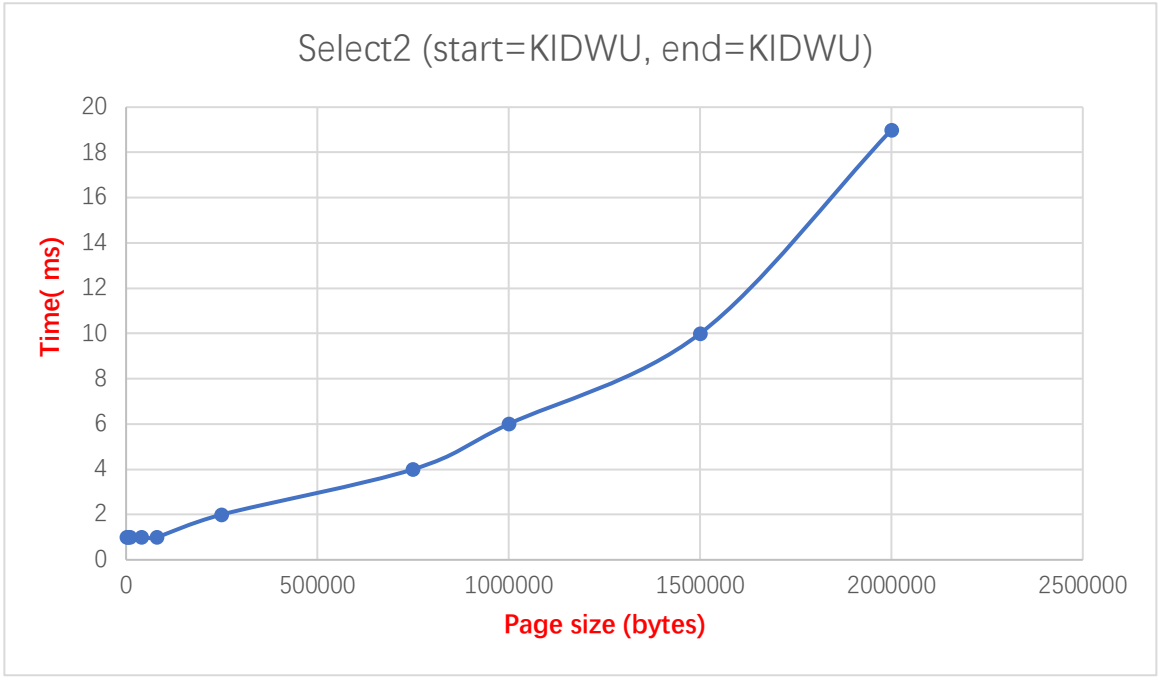
Select2 is the fastest to finish the query since we are looking at the attribute value and column store is built for storing attribute values inside a single heapfile. We can just iterate through the file that has way less records than row-store heap file. And select3 has a better result than select as well, since going through two heapfiles still has way less records to check that doing the select query on single row-store heapfile.

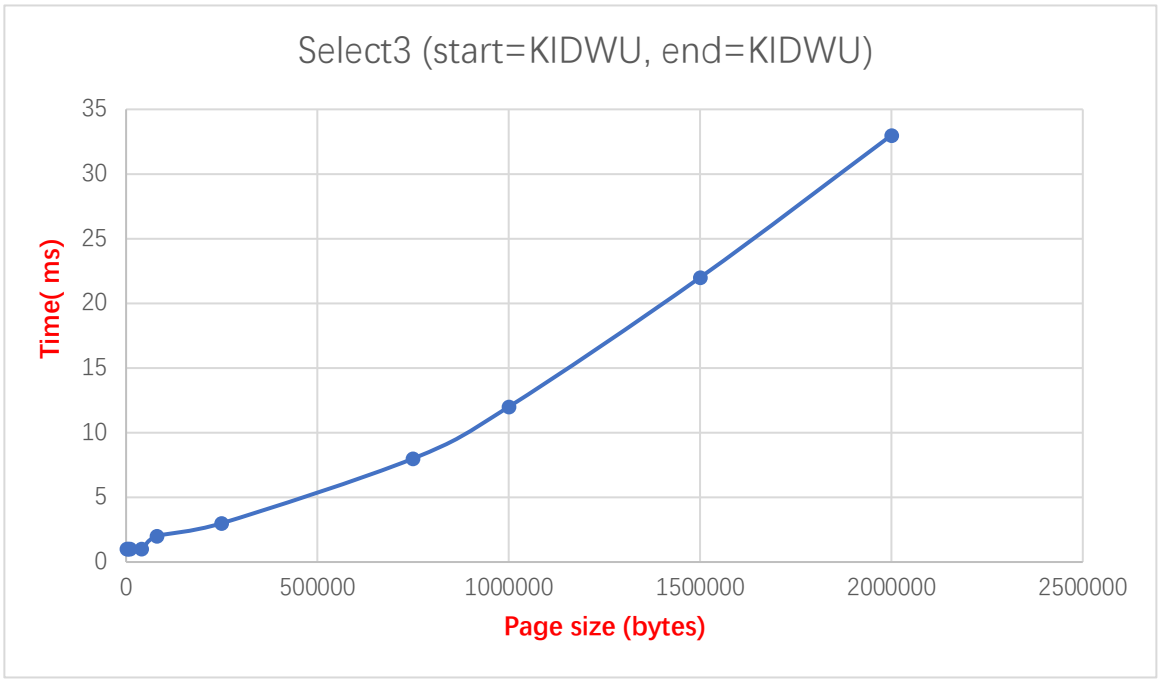





Select3 (start=AAAAA, end=ZZZZZ)
Time(ms)
25
20
15
10
5
0
0
500000
1000000
1500000
2000000
2500000
Page size (bytes)


Select(start=KIDWU, end=KIDWU)
Time(ms)
30
25
20
15
10
5
0
0
500000
1000000
1500000
2000000
2500000
Page size (bytes)


Select2 (start=KIDWU, end=KIDWU)
Time(ms)
20
18
16
14
12
10
8
6
4
2
0
0
500000
1000000
1500000
2000000
2500000
Page size (bytes)


Select3 (start=KIDWU, end=KIDWU)
Time(ms)
35
30
25
20
15
10
5
0
0
500000
1000000
1500000
2000000
2500000
Page size (bytes)