Week 4 results

Compare graph-based placement and access frequency based placement

Parameter settings:

Trace file: 120 days file access trace

Edge adding window: 10s

Edge expire window: 1 day

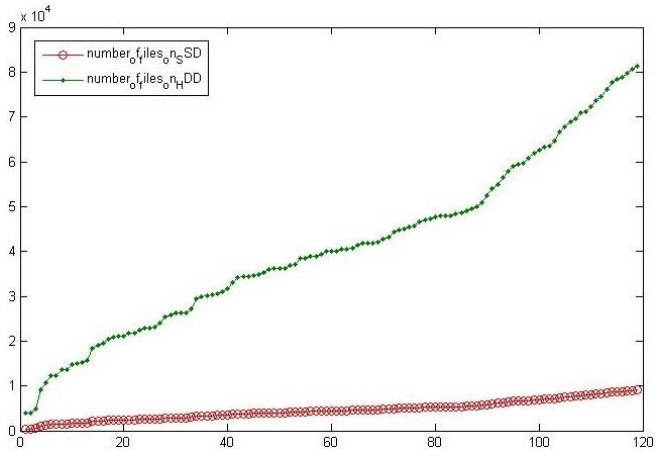
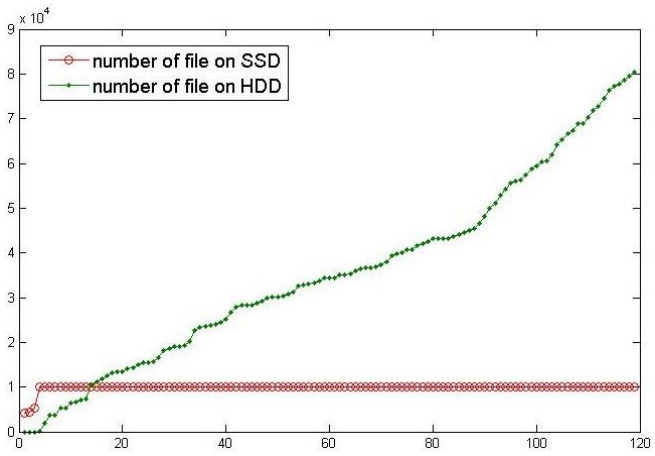
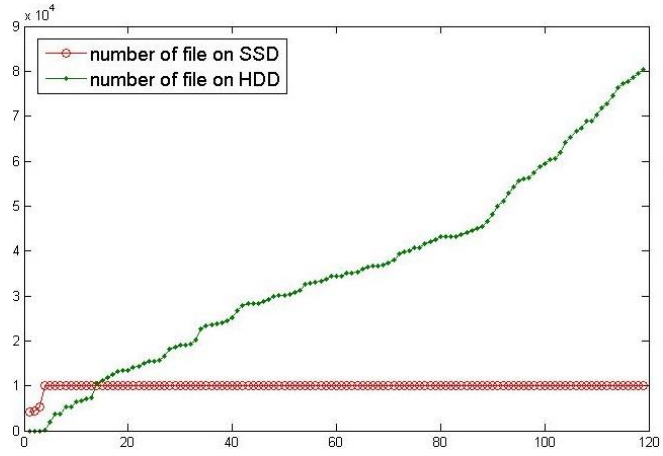
Update period: 1day

Total file count: 90345

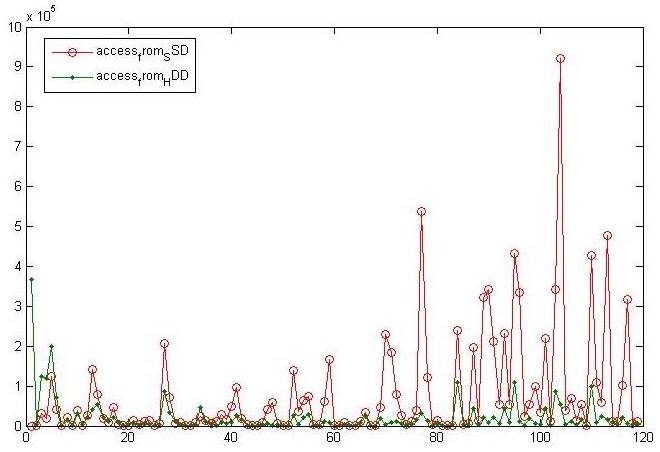
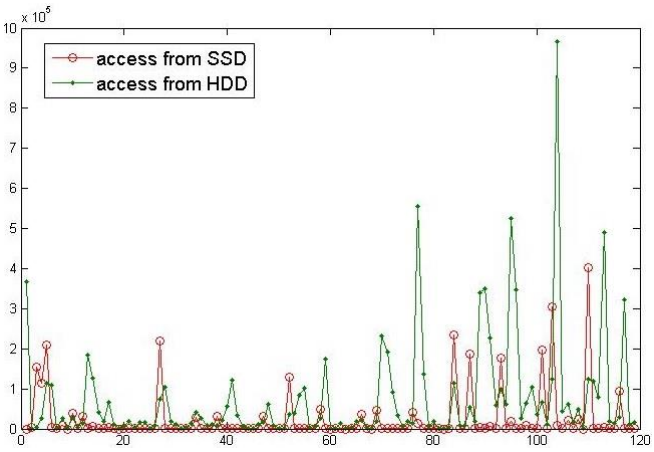
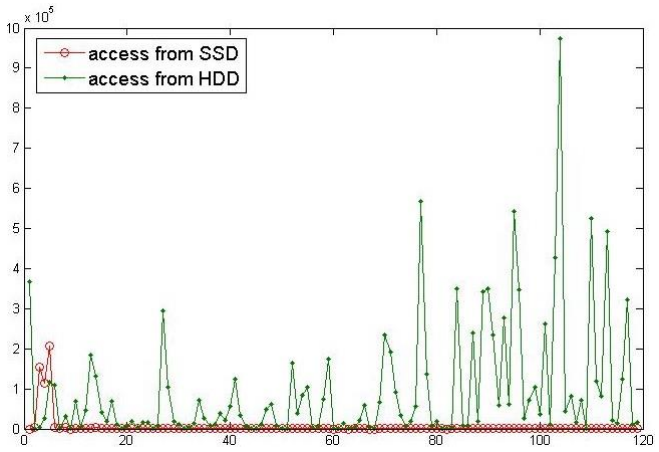
Using fixed ssd capacity ratio or fixed ssd capacity to control the data placement instead of using threshold.

Results comparison:

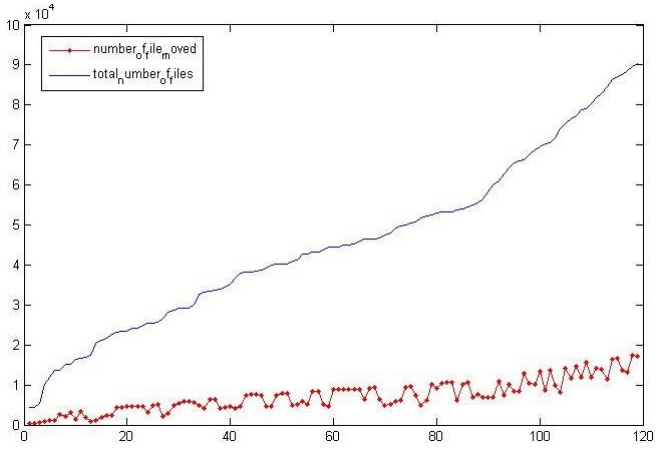
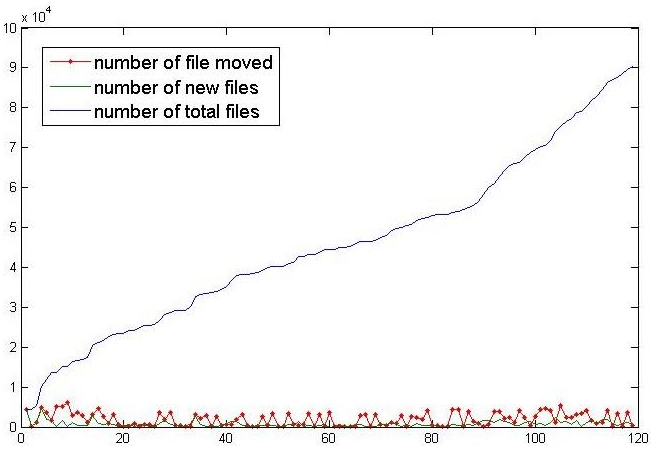
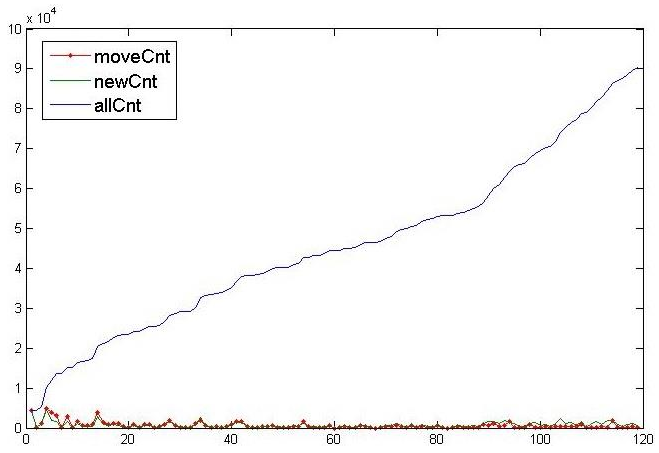
1. Placement (Number of files on SSD/HDD):

Graph based	Access frequency based	Access frequency based (Accumulated)
		

2. Performance (Number of access from SSD/HDD):

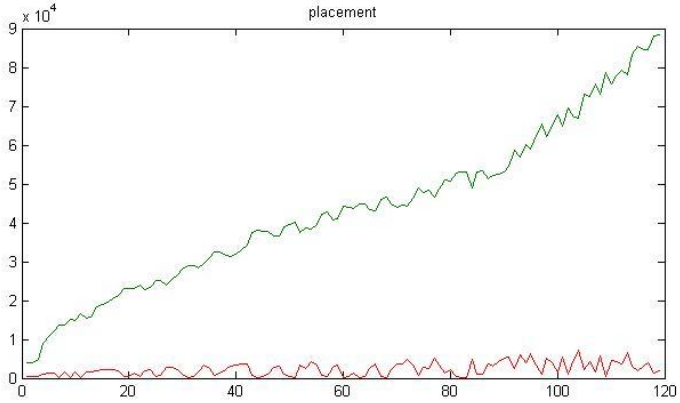
Graph based	Access frequency based	Access frequency based (Accumulated)
		

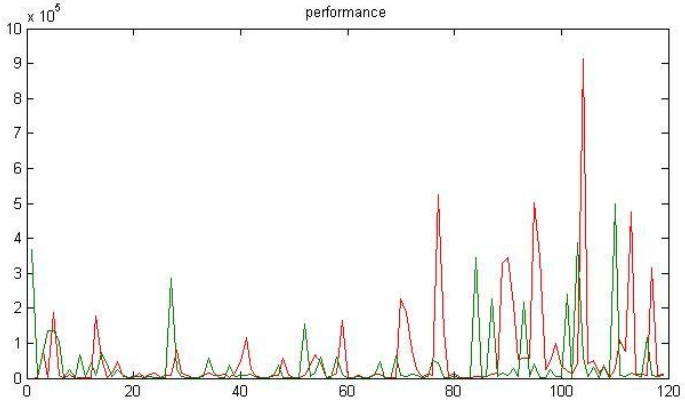
3. Overhead (Number of file movements):

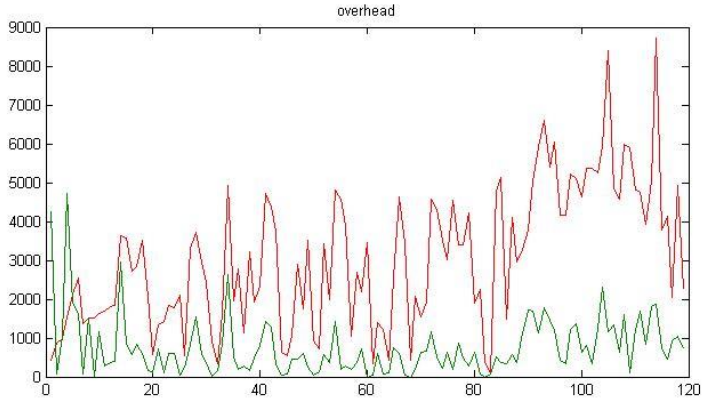
Graph based	Access frequency based	Access frequency based (Accumulated)
		

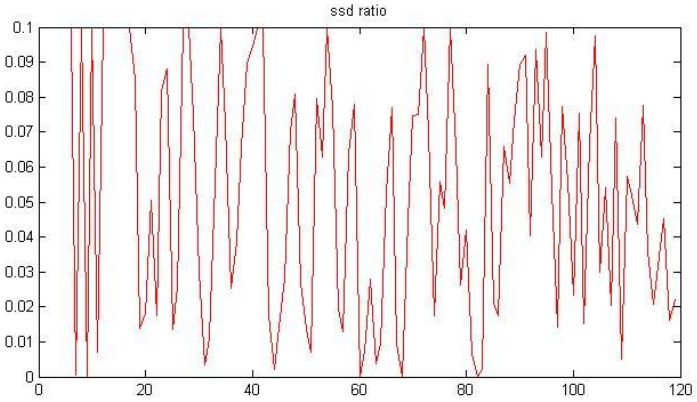
Extended Results

1. Graph mixed with high weight or high access frequency
   when only use high ssd/hdd ratio, there might be not enough files in SSD. Sometimes, files in SSD may die out.

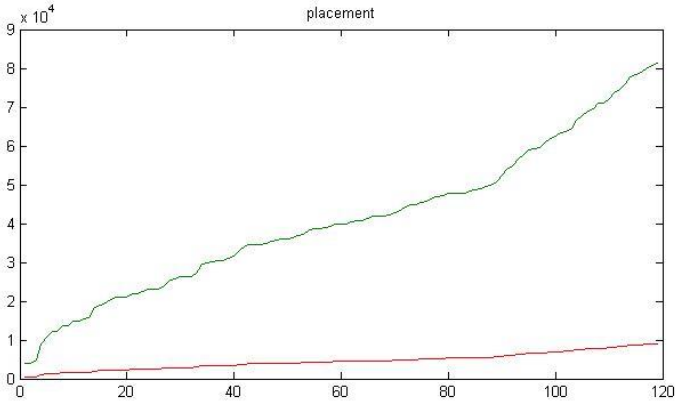


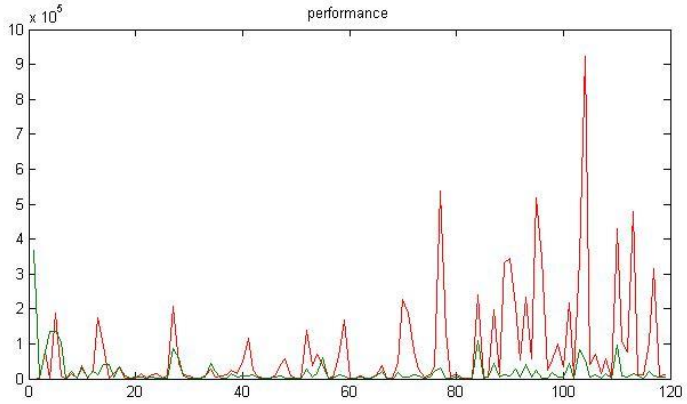


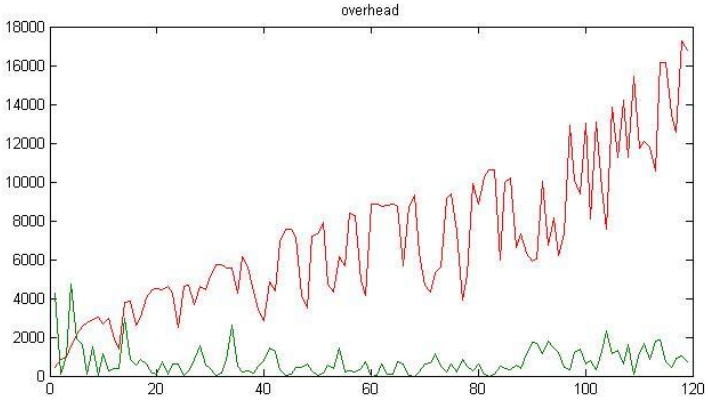


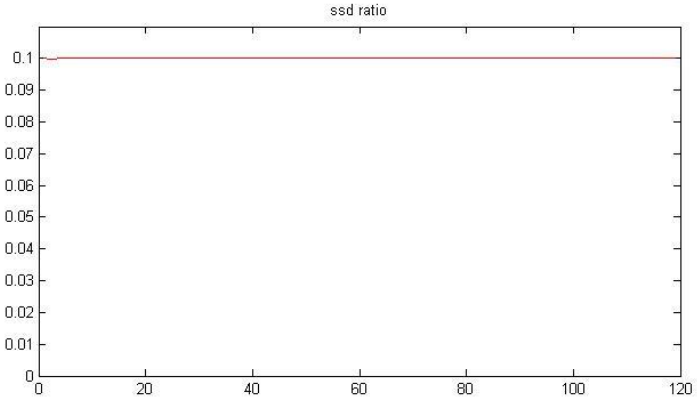


   a. High weight:









b. High access frequency: no help, worse

2. Optimal file movement

   "If we know 20th day's access frequency for each file already, how many files do we need to move on 19th day."