

# CSP 509 Major Project

Angelic 2018CSM1005, Megha Mahobe 2018CSM1016

November 30, 2018

## Assumptions :

- Grid size = 400 x 400
- Alternate axis used for scaling along with respective median values of the bucket in grid files.
- No. of knn queries for each  $k[5, 20, 50, 100] = 100$ . All the queries are randomly generated.

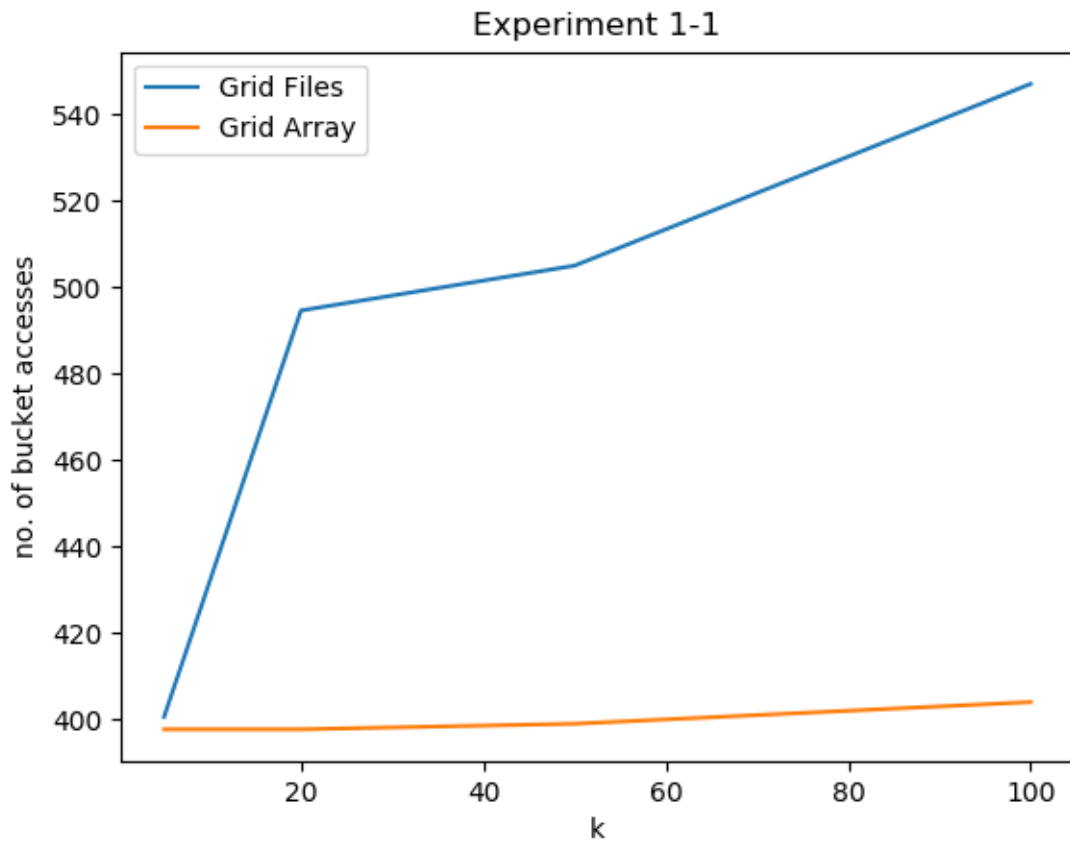


Figure 1: Dataset A, bucket capacity = 30

## Observations :

- The no. of bucket accesses for bucket capacity = 30 are greater than for capacity = 100 in both grid arrays and grid files because more buckets are generated when capacity is set low.

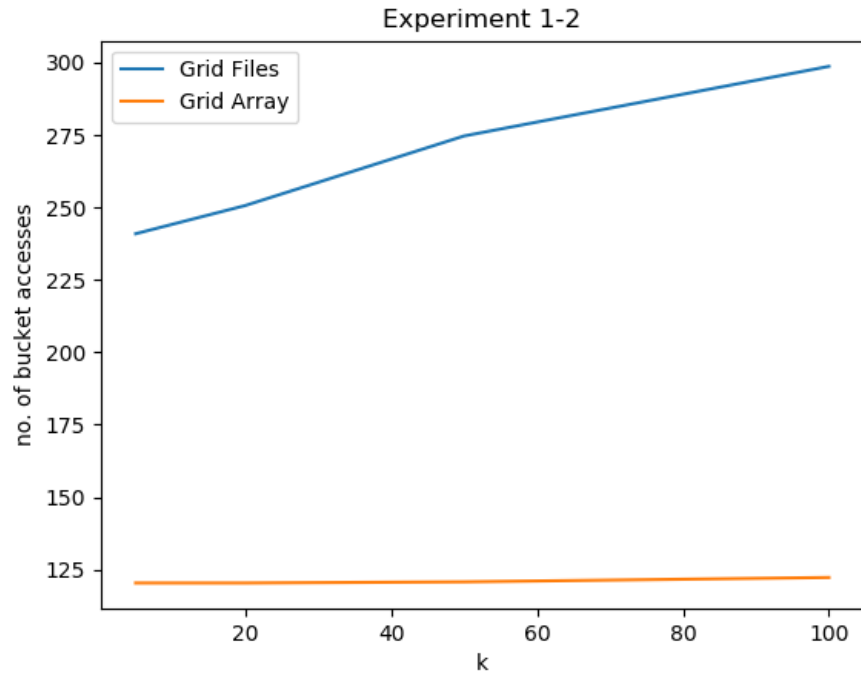


Figure 2: Dataset A, bucket capacity = 100

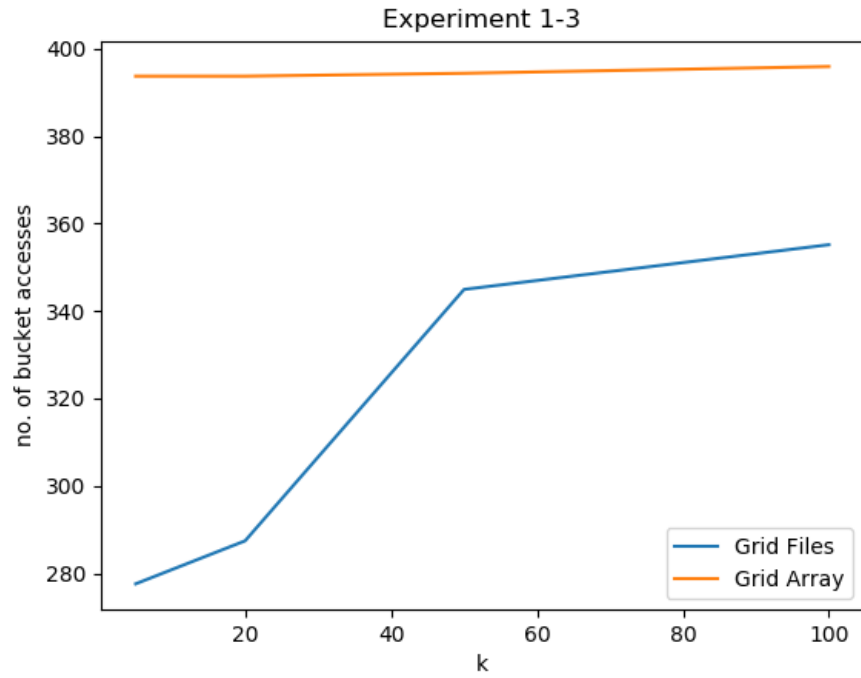


Figure 3: Dataset B, bucket capacity = 30

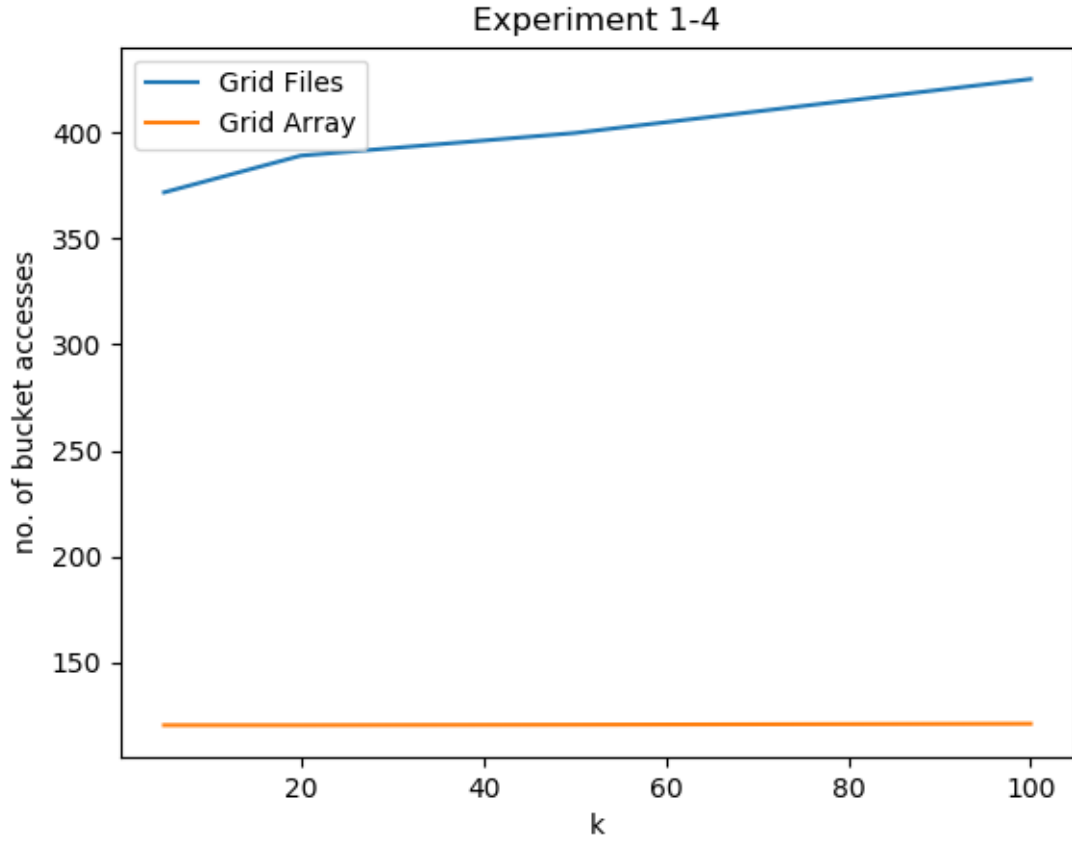


Figure 4: Dataset B, bucket capacity = 100

- The no. of bucket accesses for grid arrays don't vary much for any value of  $k$  because it is regular-sized grid structure.
- On the contrary, the trend for grid files is unpredictable. It can increase or decrease along with increasing  $k$ . In general, according to the trend analysis of the experiment, it increases steeply in some cases which again depends entirely on the data distribution in the grid cells.