

Performance of distributed data structures on MBES data

Gandhi Kalidasan Vishwa Barathy¹, Ray Suprio², Church Ian³

¹ Master of Computer Science student, University of New Brunswick, Canada

² Associate Professor, Faculty of Computer Science, University of New Brunswick, Canada

³ Associate Professor, Faculty of Engineering, University of New Brunswick, Canada
vgandhi@unb.ca

The amount of data collected from hydrographic surveys continues to increase and is expensive to store and transform; This leads to the objective of finding alternative data structures to store the data. Here we report on the latest distributed data structures and files, such as xarray and Zarr, which use the principles of distributed computing and chunking to perform I/O operations faster on the data and store in a compressed format. We analyzed the performance of the distributed data structures with several extensive hydrographic surveys, and benchmarked the parameters of the data structures with respect to the duration for I/O operations, memory, and storage consumption. The metrics of the benchmark results are provided, along with the other key findings of the latest distributed data structures.