Exploring the Feasibility of In-Line Compression on HPC Mini-Apps:

To meet the demands of solving ever larger and more complicated problems, high-performance computing (HPC) systems have grown in size and complexity and are capable of solving previously intractable problems. However, as applications become more capable, the size of their data increases. These large data sets lead to significant transfer times, causing bottlenecks as applications attempt to work with them. These bottlenecks reduce the system's efficiency by reducing the amount of productive work done. Lossy compression is capable of reducing the size of data while introducing a small, user-controllable, amount of error into the data. In this study, we show the effects of using in-line compression within two HPC mini-apps to reduce the size of the data being work on within the application. We analyze the effects of using in-line compression on overall accuracy, storage, and throughput. Our analysis shows that, while significant improvements to the in-line lossy compression algorithm are needed, the algorithm can reduce the amount of storage required while introducing a small amount of inaccuracy and overhead, all of which are controlled by the user. We aim to further improve the algorithm by reducing the overhead introduced while enhancing its usability within HPC applications. Specifically, we aim to develop a standard MPI interface for the compressed data, enable the in-line compression to be used with OpenMP, and improve the algorithms API to handle dynamic data types better.

LA-UR-20-25813