Data-Aware Compression for HPC using Machine Learning

- ▶ Author: Julius Plehn
- ▶ Type: Master's Thesis
- ▶ Date: 2022-05-09
- ▶ Reviewers: Jun.-Prof. Dr. Michael Kuhn, Prof. Dr. Thomas Ludwig
- ▶ Supervisors: Jun.-Prof. Dr. Michael Kuhn, Anna Fuchs, Jakob Lüttgau
- ▶ Download: PDF

While compression can provide significant storage and cost savings, its use within HPC applications is often only of secondary concern. This is in part due to the inflexibility of existing approaches where a single compression algorithm has to be used throughout the whole application, but also because insights into the behavior of the algorithms within the context of individual applications are missing. Several compression algorithms are available, with each one also having a unique set of options. These options have a direct influence on the achieved performance and compression results. Furthermore, the algorithms and options use for a given dataset are highly dependent on the characteristics of said dataset. This thesis explores how machine learning ca help identify fitting compression algorithms with corresponding options based on actual data structure encountered during I/O. T do so, a data collection and training pipeline is introduced. Inferencing is performed during regular application runs and shows promising results. Moreover, it provides valuable insights into the benefits of using certain compression algorithms and options fc specific data.