

Final Progress Report

DOE Award Number: DE-SC0004935

Name of the recipient: North Carolina State University

Project Title: Damsel - A Data Model Storage Library for Exascale Science

Name of the PI: Alok Choudhary, Northwestern University

Name of the NCSU PI: Nagiza F. Samatova

Project description:

The goal of this project is to enable exascale computational science applications to interact conveniently and efficiently with storage through abstractions that match their data models. We will accomplish this through three major activities: (1) identifying major data model motifs in computational science applications and developing representative benchmarks; (2) developing a data model storage library, called Damsel, that supports these motifs, provides efficient storage data layouts, incorporates optimizations to enable exascale operation, and is tolerant to failures; and (3) productizing Damsel and working with computational scientists to encourage adoption of this library by the scientific community.

Accomplishments Summary:

The following summarizes our accomplishments:

- Published eight papers in peer review conferences and journals (see the list below).
- Demonstrated improved compression rates and throughput for the following application codes: S3D, FLASH, GTS, XGC-1.
- Improved I/O throughput for standard I/O libraries by 15%-50% for check point and restart data for the following simulation codes: S3D, FLASH, and XGC-1.
- Presented our results at the following conferences: IPDPS'12, ICDM'11, ICDE'12, HPDC'12, SC'11, EuroPar'11.
- Produced two software libraries for open source release, available from <http://www.freescience.org/cs>:
 - ISABELA is an in-situ, embarrassingly parallel B-Spline based lossy compression technique for scientific floating point data with user-controlled accuracy bounds.
 - ISOBAR is a preconditioner-based, high-throughput lossless compression technique for hard-to-compress scientific datasets.
- Integrated the compression routines into the HDF5 and ADIOS parallel I/O libraries.
- Designed indexing and query processing over the compressed AMR data.
- Established collaborations with major AMR teams including Chombo (Dr. Phil Colella).

Next, we will briefly summarize R&D advancements and the impact on the applications of the DOE mission relevance. The details are available through the published papers.

ISOBAR Preconditioner for Effective & High-throughput Lossless Data Compression

Data reduction is a key to avoid the I/O bottleneck. Effective lossless compression ensures simulation fidelity, especially at checkpoints. Interleaving lossless compression and parallel I/O is a “secret sauce” for addressing the I/O bottleneck. NCSU’s ISOBAR code and theoretical performance model (ICDE'12, HPDC'12, EuroPar'11) offer a predictive, scalable and power-

efficient implementation of this strategy. Using various performance metrics about the system (e.g., network bandwidth and latency, disk read/write throughput) and the application (e.g., compression ratio and throughput, amount of data per core), ISOBAR's theoretical model accurately predicts the optimal balance among the placement of compression (on cores and/or compute/I/O/staging nodes), the data movement, and I/O. Using the pre-conditioner, ISOBAR dynamically identifies highly compressible bytes to process by a compression method, while asynchronously writing the remaining, uncompressible bytes to storage with parallel I/O library. This way it effectively hides the cost of compression and I/O synchronization. It operates within the data staging architecture, where various data transformations occur while data is *in transit* or *in situ*, from compute nodes to disk. ISOBAR exhibits both read and write performance gains proportional to the degree of data reduction, which ranges as high as 46% on scientific datasets, in addition to reducing the total amount of data that is being stored and accessed. Because ISOBAR applies a pre-conditioner to identify compressible bytes in small chunks, it avoids wasting CPU cycles trying to compress incompressible bytes in the data. By operating on a lower memory footprint and in an embarrassingly parallel manner, this method results in energy-efficiency, while simultaneously offering a high throughput, reduced data movement, and data reduction that collectively translate to a factor of 2-6 reduction in energy consumption.

ISOBAR showed superior performance when tested against state-of-the-art lossless compressors on data generated by XGC1 fusion particle codes, S3D combustion, GTS fusion core plasma and FLASH astrophysics DOE petascale simulation codes. Specifically, CS Chang, PPPL, noted: "Very nice work! Yes, we would like to have it incorporated into XGC1. Nearly 50% data reduction at almost 100MB/sec compression throughput and more than 380 MB/sec decompression are the right ISOBAR performance metrics for our ongoing work on in-memory code-coupling."

Multi-level Layout Optimization for Fast Spatio-temporal Queries

The size and scope of cutting-edge scientific simulations are growing much faster than the I/O subsystems of their runtime environments, not only making I/O the primary bottleneck, but also consuming space that pushes the storage capacities of many computing facilities. These problems are exacerbated by the need to perform data-intensive analytics applications, such as querying the dataset by variable and spatio-temporal constraints, for what current database technologies commonly build query indices of size greater than that of the raw data. To help solve these problems, we have been developing a parallel query-processing engine that can handle both range queries and queries with spatio-temporal constraints, on B-spline compressed data with user-controlled accuracy. Our method adapts to widening gaps between computation and I/O performance by querying on compressed metadata separated into bins by variable values, utilizing Hilbert space-filling curves to optimize for spatial constraints and aggregating data access to improve locality of per-bin stored data, reducing the false positive rate and latency-bound I/O operations (such as seek) substantially. We show our method to be efficient with respect to storage, computation, and I/O compared to existing database technologies optimized for query processing on scientific data [ICDM'12, ICPP'12, IPDPS'12, CCPE'12, SC'11].

References:

1. [IPDPS'12] Gong, Z., Lakshminarasimhan, S., Jenkins, J., Kolla, H., Ethier S., Chen, J., Ross, R., Klasky, S., and Samatova, N.F., "Multi-level layout optimization for efficient spatio-temporal queries on ISABELA-compressed data." *The 26th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2012)*, May 21-25, 2012, Shanghai, China.

2. [ICPP'12] Z. Gong, T. Rogers, J. Jenkins, H. Kolla, S. Ethier, J. Chen, R. Ross, S. Klasky and N. F. Samatova, "*MLOC: Multi-level Layout Optimization Framework for Compressed Scientific Data Exploration with Heterogeneous Access Patterns*," Proc.41st International Conference on Parallel Processing (ICPP), Pittsburgh, PA. September 10-13, 2012.
3. [ICDM'12] Jin, Y., Lakshminarasimhan, S., Shah, N., Gong, Z., Chang, C.S., Chen, J., Ethier S., Kolla, H., Ku, S., Klasky, S., Latham, R., Ross, R., Schuchardt, K., and Samatova, N.F., ``S-preconditioner for multi-fold data reduction with guaranteed user-controlled accuracy." *IEEE International Conference on Data Mining (ICDM 2011)*, December 11-14, 2011, Vancouver, Canada.
4. [CCPE'12] Lakshminarasimhan, S., Shah, N., Ethier, S., Ku, S-H., Chang, C.S., Klasky, S., Latham, R., Ross, R., and Samatova, N.F., ``ISABELA: Effective in-situ compression of spatio-temporal data." *Concurrency and Computation: Practice and Experience Journal*, 2012 (Accepted).
5. [HPDC'12] Schendel, E.R., Pendse, S.V., Jenkins, J., Boyuka II, D.A., Gong, Z., Lakshminarasimhan, S., Liu, Q., Klasky, S., Ross, R., and Samatova, N.F., ``ISOBAR hybrid compression-I/O interleaving for large-scale parallel I/O optimization." *The 21st International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC 2012)*, Delft, the Netherlands, June 20-22, 2012.
6. [ICDE'12] Schendel, E.R., Jin, Y., Shah, N., Chen, J., Chang, C.S., Ku, S-H., Ethier, S., Klasky, S., Latham, R., Ross, R., and Samatova, N.F., ``ISOBAR preconditioner for effective and high-throughput lossless data compression." *The 28th IEEE International Conference on Data Engineering (ICDE 2012)*, April 1--5, 2012, Washington DC.
7. [EuroPar'11] S. Lakshminarasimhan, N. Shah, S. Ethier, S. Klasky, R. Latham, R. Ross and N. F. Samatova, Compressing the Incompressible with ISABELA: In-situ Reduction of Spatio-Temporal Data, 17th International European Conference on Parallel and Distributed Computing [Euro-Par 2011], **Distinguished Paper Award**.
8. [SC'11] S. Lakshminarasimhan, J. Jenkins, I. Arkatkar¹, Z. Gong, H. Kolla, S-H Ku, S. Ethier, J. Chen, CS Chang, S. Klasky, R. Latham, R. Ross and N. F. Samatova, ISABELA-QA: Query-driven Data Analytics over ISABELA-compressed Scientific Data, Supercomputing Conference [SC' 2011].