



Contribution ID: 115

Type: Oral

Precision Cascade: A novel algorithm for multi-precision extreme compression

Tuesday 25 October 2022 17:00 (20 minutes)

Lossy compression algorithms are incredibly useful due to powerful compression results. However, lossy compression has historically presented a trade-off between the retained precision and the resulting size of data compressed with a lossy algorithm. Previously, we introduced BLAST, a state-of-the-art compression algorithm developed by Accellogic. We presented results that demonstrated BLAST can achieve a compression factor that undeniably surpasses compression algorithms currently available in the ROOT framework. However, the leading concern of utilizing the lossy compression technique is the delayed realization that more precision is necessary. This precision may have been irretrievably lost in an effort to decrease storage size. Thus, there is immense value in retaining higher precision data in reserve. Though, in the era of exabyte computing, it becomes extremely inefficient and costly to duplicate data stored at different compressive precision values. A tiered cascade of stored precision optimizes data storage and resolves these fundamental concerns.

Accellogic has developed a game-changing compression technique, known as “Precision Cascade”, which enables higher precision to be stored separately without duplicating information. With this novel method, varying levels of precision can be retrieved, potentially minimizing live storage space. Preliminary results from STAR and CMS demonstrate that multiple layers of precision can be stored and retrieved without significant penalty to the compression ratios and (de)compression speeds, when compared to the single-precision BLAST baseline.

In this contribution, we will present the integration of Accellogic’s “Precision Cascade” into the ROOT framework, with the principal purpose of enabling high-energy physics experiments to leverage this state-of-the-art algorithm with minimal friction. We also present our progress in exploring storage reduction and speed performance with this new compression tool in realistic examples from both STAR and CMS experiments and feel we are ready to deliver the compression algorithm to the wider community.

Significance

Lossy compression algorithms are incredibly useful due to powerful compression results and represents the next evolution in the HEP/NP IO workflows for extreme space saving. Precision cascade is a novel technique that targets the primary concern of lossy compression (that too much precision is lost), by enabling higher precision to be stored and retrieved later on, without duplicating data.

Track 1 encompassing Computing Technology for Physics Research and architecture seem suited for our abstract.

References

- G. V. Buren, J. Lauret, J. Gonzales, R. Nunez, P. Canal, A. Naumann – “Extreme compression for Large Scale Data store” – CHEP 2019 proceedings, EPJ Web of Conferences 245, 06024 (2020), doi 10.1051/epjconf/202024506024
- P. Canal, G.V. Buren, J. Lauret, I.A. Cali, J. Gonzales, P. Canal, R. Nunez, Y. Ying – “ROOT Files Improved with Extreme Compression”, ACAT 2021 proceedings, accepted for publication.

Experiment context, if any

STAR, CMS and the ROOT team

Primary authors: VAN BUREN, Gene; Dr LAURET, Jerome (Brookhaven National Laboratory); Dr GONZALEZ, Juan (Accelogic, Inc.); CANAL, Philippe (Fermi National Accelerator Lab. (US)); YING, Yueyang (Massachusetts Inst. of Technology (US))

Presenter: YING, Yueyang (Massachusetts Inst. of Technology (US))

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research