Contribution ID: **678** Contribution code: **contribution ID 678** Type: **Oral**

ROOT Files Improved with Extreme Compression

Thursday, 2 December 2021 11:40 (20 minutes)

For the last 7 years Accelogic pioneered and perfected a radically new theory of numerical computing codenamed "Compressive Computing", which has an extremely profound impact on real-world computer science [1]. At the core of this new theory is the discovery of one of its fundamental theorems which states that, under very general conditions, the vast majority (typically between 70% and 80%) of the bits used in modern large-scale numerical computations are absolutely irrelevant for the accuracy of the end result. This theory of Compressive Computing provides mechanisms able to identify (with high intelligence and surgical accuracy) the number of bits (i.e., the precision) that can be used to represent numbers without affecting the substance of the end results, as they are computed and vary in real time. The bottom line outcome would be to provide a state-of-the-art compression algorithm that surpasses those currently available in the ROOT framework, with the purpose of enabling substantial economic and operational gains (including speedup) for High Energy and Nuclear Physics data storage/analysis. In our initial studies, a factor of nearly x4 (3.9) compression was achieved with RHIC/STAR data where ROOT compression managed only x1.4.

As a collaboration of experimental scientists, private industry, and the ROOT Team, our aim is to capitalize on the substantial success delivered by the initial effort and produce a robust technology properly packaged as an open-source tool that could be used by virtually every experiment around the world as means for improving data management and accessibility.

In this contribution, we will present our efforts integrating our concepts of "functionally lossless compression" within the ROOT framework implementation, with the purpose of producing a basic solution readily integrated into HENP applications. We will also present our progress applying this compression through realistic examples of analysis from both the STAR and CMS experiments.

Significance

Completion of this work has the potential to enable the ability to save significant amount of storage space (or bandwidth) while not affecting noticeably the physics result quality.

References

Speaker time zone

Compatible with America

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

Presenters: VAN BUREN, Gene (Brookhaven National Laboratory); LAURET, Jerome (Brookhaven National Laboratory); CALI, Ivan Amos (Massachusetts Inst. of Technology (US)); Dr GONZALEZ, Juan (Accelogic); CANAL, Philippe (Fermi National Accelerator Lab. (US)); Mr NUNEZ, Rafael; 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