CHEP Poster Session – Adelaide, Nov 2019

ATLAS Event Store and I/O developments in support for Production and Analysis in Run 3

During the long shutdown, ATLAS is preparing several fundamental changes to its offline event processing framework and analysis model. These include moving to multi-threaded reconstruction and simulation and reducing data duplication during derivation analysis by producing a combined mini-xAOD stream. These changes will allow ATLAS to take advantage of the higher luminosity at Run 3 without overstraining processing and storage capabilities. They also require significant changes to the underlying event store and the I/O framework to support them.

- The Run 2 I/O framework was overhauled to be thread-safe and minimize serial bottlenecks.
- For object navigation, new immutable references are deployed, which don't rely on storage container entry number so data can be merged in-memory.
- Filter decisions can be used to annotate combined output stream allowing for fast event selection on input.
- Compression algorithms and settings were optimized to allow efficient reading of event selections.

Thread safety and concurrency in the I/O framework

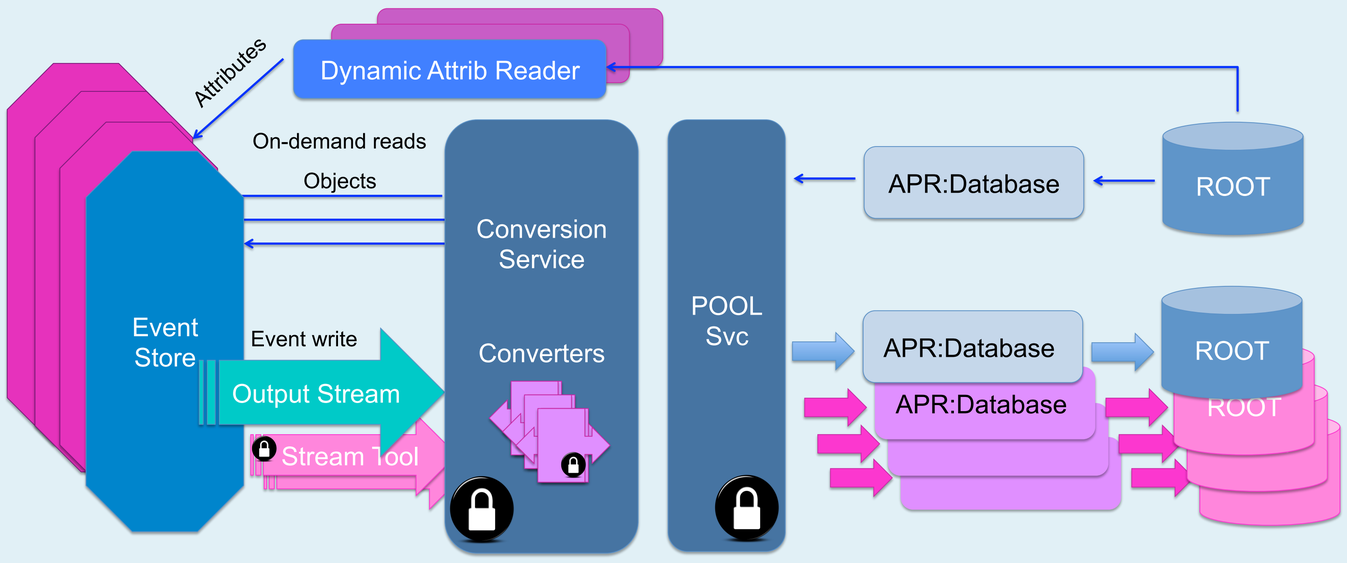

Fig 1. I/O Framework evolution from Single-Process (blue) to Multi-Threaded (blue and pink). Concurrent execution is made possible by creating additional copies of components that work in parallel. Multiple Events are processed in at the same time in dedicated Event Stores and then written out. Serialization is necessary when writing to the same file, but concurrent writing can be performed on different files.
Central services were made thread-safe with the use of thread locking mechanisms.

Object indexing in the storage layer

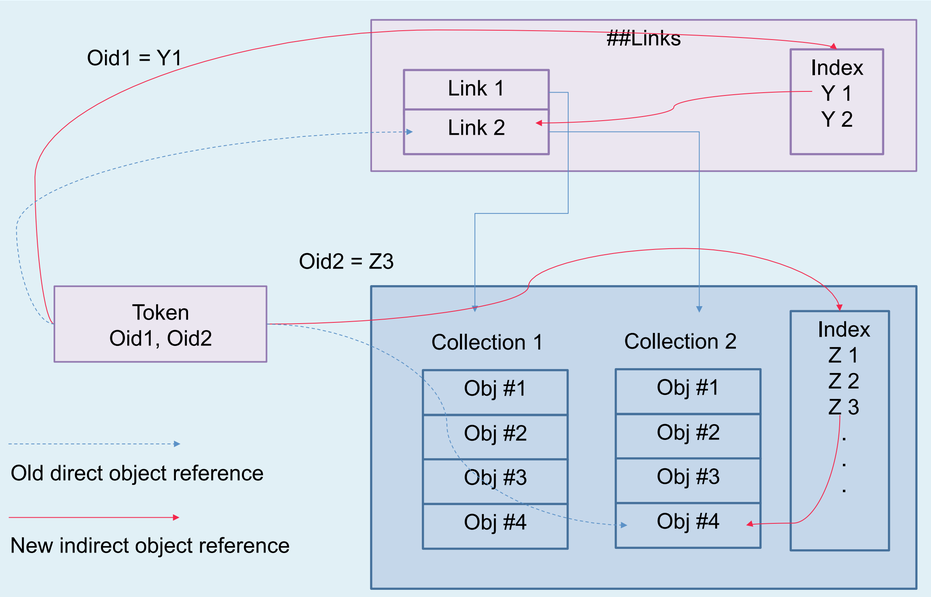

Fig 2. Persistent object reference (Token) received an additional level of indirection in the form of in-file index (TTreeIndex). The index can be automatically updated when merging file, thus preserving navigational references even though row numbers change.

Combined output stream with Event tagging

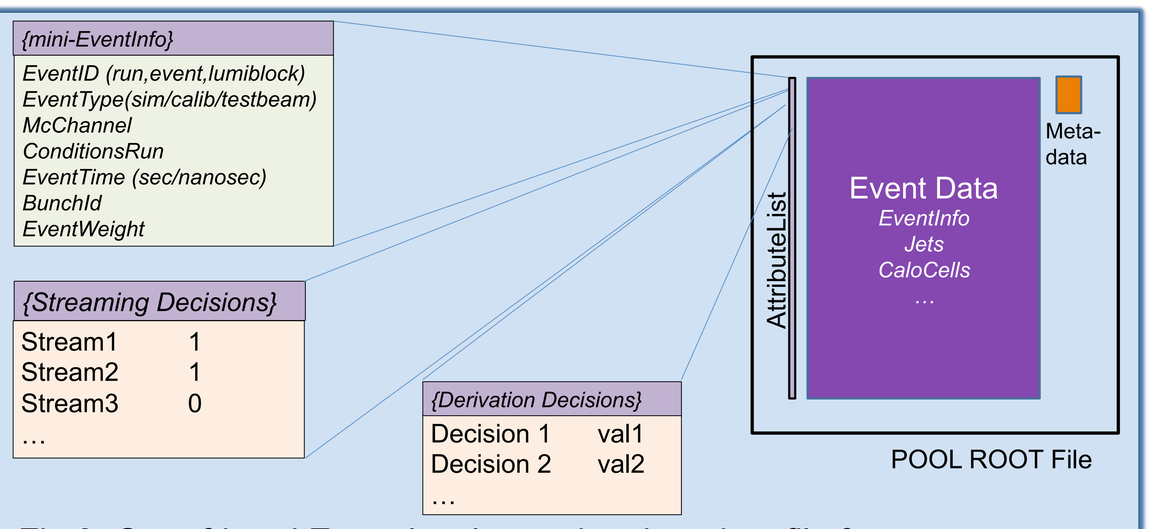

Fig 3. Out-of-band Event-level metadata in a data file for fast Event selection on input

ATLAS Run3 uses a single combined derivation output stream in order to avoid event duplication, observed among Run2 separate streams, and save disk space. Every event in the common stream is marked as belonging to any of the derivation streams. Additional values relevant to event filtering can also be stored.
Keeping the information outside the main Event body in a simple, easily readable ROOT format, permits the Main Event Loop to efficiently read only the events from the desired derivation stream, with additional selection on stored attributes.

Compression optimization for efficient selective reading

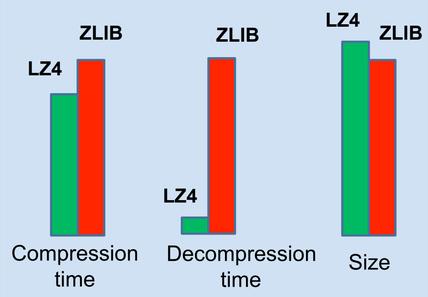

Fig 4. Comparison between LZ4 and ZLIB compression algorithms for ATLAS Event data files

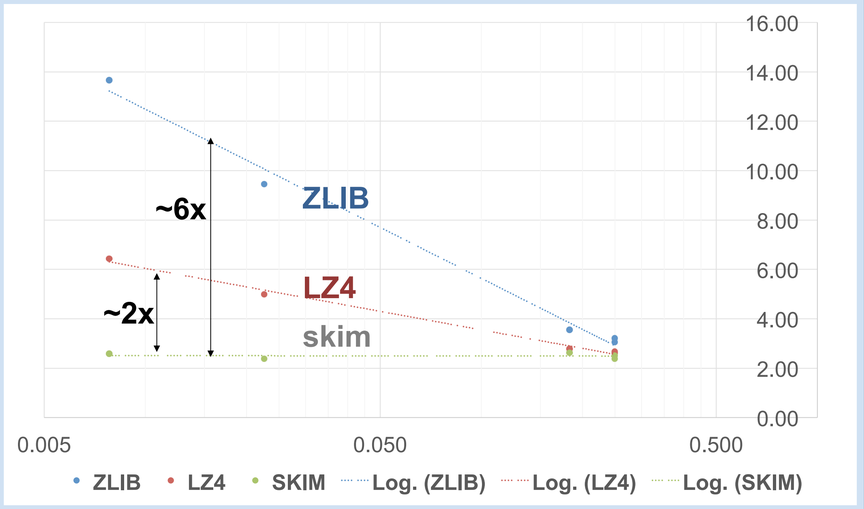

Fig 5. Effects of different compression algorithms on selective reading speed compared to 100% read (skim)

Selective reading of objects from ROOT files introduces inefficiencies due to reading and decompressing entire buffers, when only a fraction of that data had been requested. LZ4, with it's fast decompression, makes it possible to read data at **40%-100% of the rate of pre-selected samples,** while retaining most of the size reduction advantages of ZLIB compression.

Summary

ATLAS Computing is undergoing important changes to meet the challenges of Run3 data handling and processing. The offline software framework ATHENA, and in particular its I/O components, has evolved to support these changes. The framework foundations have achieved stability in the multi-threaded environment, which provides the ground for adaptation and testing of physics algorithms and other framework components. Data storage format was modified to achieve better balance between disk space and performance. Object referencing was made more robust. The work on the I/O layer continues to increase concurrency on writing and increase performance.

M.Nowak[1], P. van Gemmeren[2] and J.Cranshaw[2], for the ATLAS Collaboration
[1]Brookhaven National Laboratory, Upton, New York 11973, USA, [2]Argonne National Laboratory, Argonne, Illinois 60439, USA