The RNTuple Storage Format

Jonas Hahnfeld, for the RNTuple development team
jonas.hahnfeld@cern.ch

GEANT4 Collaboration Meeting – October 7, 2024

Context: Geant4 Analysis Tools

- `G4AnalysisManager`: store histograms and n-tuples
  - Supports CSV, XML, HDF5, ROOT (TTree)
  - Integrated in GEANT4, no need to link external software

Context: Geant4 Analysis Tools

- `G4AnalysisManager`: store histograms and n-tuples
  - Supports CSV, XML, HDF5, ROOT (TTree)
  - Integrated in GEANT4, no need to link external software

- n-tuples support columns of ...
  - integer, `float`, `double`, and `string` types,
  - `vectors` of these types

Context: Geant4 Analysis Tools

- `G4AnalysisManager`: store histograms and n-tuples
  - Supports CSV, XML, HDF5, ROOT (TTree)
  - Integrated in GEANT4, no need to link external software

- n-tuples support columns of ...
  - integer, `float`, `double`, and `string` types,
  - `vectors` of these types

- For more details, see contribution by Ivana and Guy at CHEP 2023

ROOT TTree

- Binary columnar data format since more than 25 years
  - Optimized for analysis reading of only a few columns
  - Transparently compressed to save storage space

ROOT TTree

- Binary columnar data format since more than 25 years
  - Optimized for analysis reading of only a few columns
  - Transparently compressed to save storage space
- Used in production by many HEP experiments
  - For example at the LHC, storing more than 2 exabytes

RNTuple

- Designated successor to TTree, to be used in HL-LHC
  - Expecting a 10x increase of data volume

- Designated successor to TTree, to be used in HL-LHC
  - Expecting a 10x increase of data volume
- Modern design, based on decades of experience

RNTuple

- Designated successor to TTree, to be used in HL-LHC
  - Expecting a 10x increase of data volume
- Modern design, based on decades of experience
- Smaller files, better performance, systematic use of checksums

RNTuple: Smaller Files

- Simulated 1 million 18 GeV π^+ events with ATLTileCalTB
  - 4 scalar `double` fields: `ELeak`, `Ecal`, `EdepSum`, `SdepSum`
  - 2 `std::vector<double>` fields: `Edep`, `Sdep`
  - 2 fields describing primary particle: `PDGID` (`int32_t`) and `EBeam` (`float`)

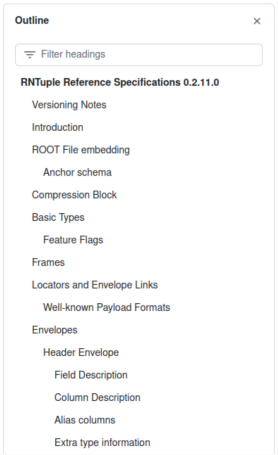
Compression	TTree	RNTuple	
`zlib` (101)	391M	367M	-6.3 %
`zstd` (505)	361M	334M	-7.3 %
	-7.8 %	-8.9 %	-14.7 %

RNTuple Specification

- Written document on the RNTuple binary format
  - Maintained as part of the ROOT repository

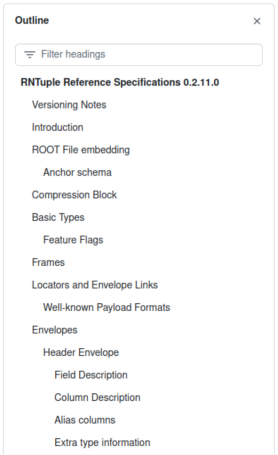
RNTuple Specification

- Written document on the RNTuple binary format
  - Maintained as part of the ROOT repository



RNTuple Specification

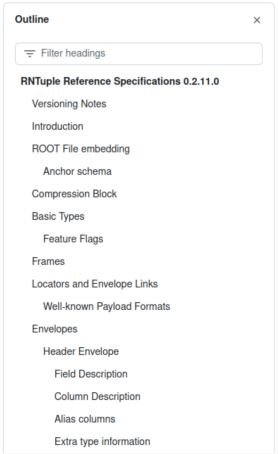
- Written document on the RNTuple binary format
  - Maintained as part of the ROOT repository
- Version 1.0 will be declared end of this year
  - ROOT promises to keep backwards compatibility



RNTuple Specification

- Written document on the RNTuple binary format
  - Maintained as part of the ROOT repository
- Version 1.0 will be declared end of this year
  - ROOT promises to keep backwards compatibility

$\Rightarrow$ Compliant implementation usable for a long time



RNTuple in Geant4 Analysis Tools?

- Medium- to long-term: development will move to RNTuple

RNTuple in Geant4 Analysis Tools?

- Medium- to long-term: development will move to RNTuple
- To be clear: TTree reading will continue to be supported in ROOT
  - Data written by `G4AnalysisManager` will remain usable

RNTuple in Geant4 Analysis Tools?

- Medium- to long-term: development will move to RNTuple
- To be clear: TTree reading will continue to be supported in ROOT
  - Data written by `G4AnalysisManager` will remain usable
- Advantages of RNTuple for GEANT4:
  - Written specification, no reverse-engineering needed
  - Embedding into ROOT `TFile` can be reused
  - Smaller files? (to be tested per application)
  - "Moving with the trend" as other software adopts RNTuple

RNTuple in Geant4 Analysis Tools?

- Currently no resources in the ROOT team to provide a second implementation
  - Open to discuss requirements to link RNTuple library into GEANT4

RNTuple in Geant4 Analysis Tools?

- Currently no resources in the ROOT team to provide a second implementation
  - Open to discuss requirements to link RNTuple library into GEANT4
- RNTuple specification simplifies external implementations
  - Successfully demonstrated with RNTuple support in `uproot`

RNTuple in Geant4 Analysis Tools?

- Currently no resources in the ROOT team to provide a second implementation
  - Open to discuss requirements to link RNTuple library into GEANT4
- RNTuple specification simplifies external implementations
  - Successfully demonstrated with RNTuple support in `uproot`
- Of course available to help and discuss in case of problems