

High Energy Physics Center for Computational Excellence

Storage OPTimization: Outlook

Peter van Gemmeren (ANL)
for HEP-CCE

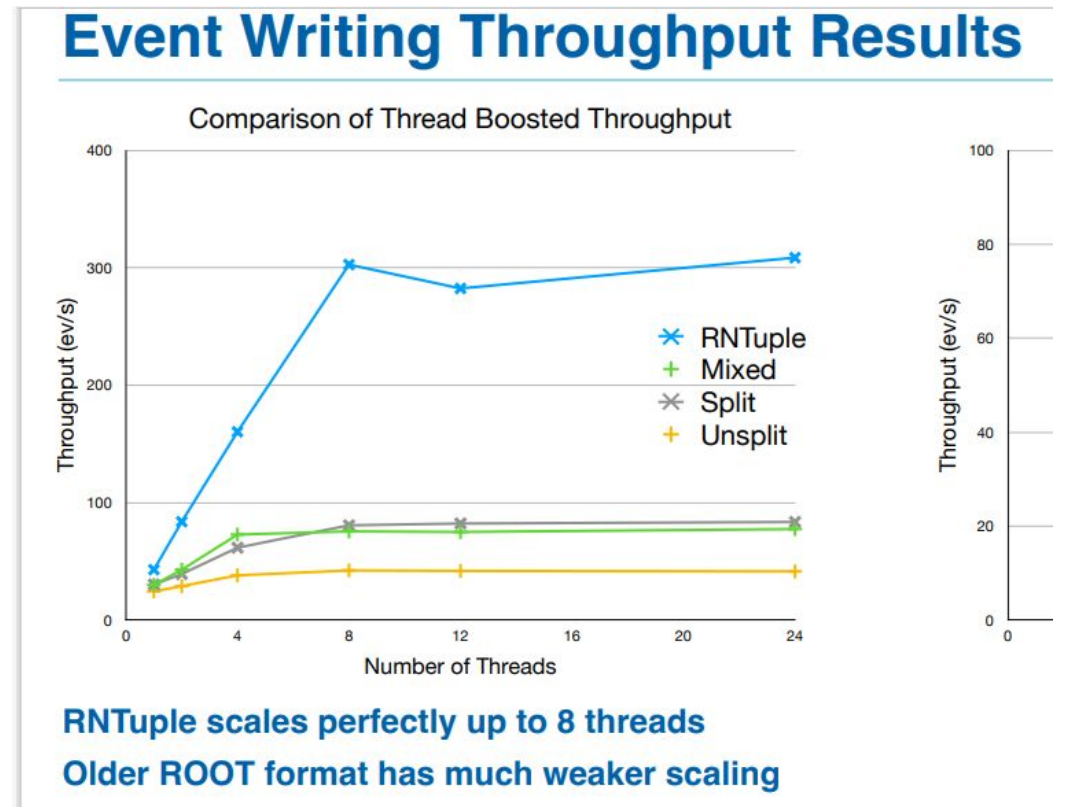
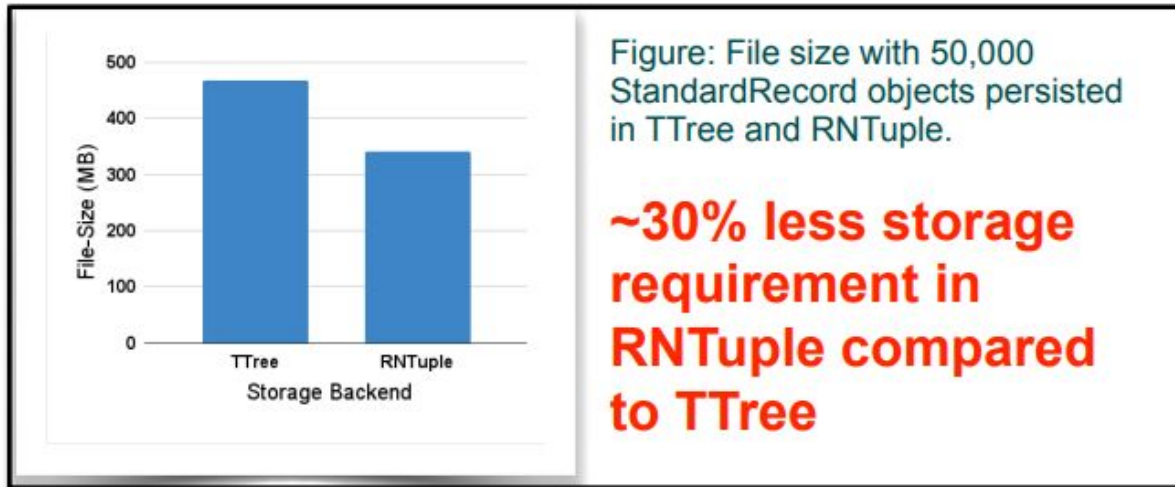
HEP CCE AHM
July 24, 2024

Questions (and Answers)

What have been the two most significant accomplishments of Year 1 so far?

1. RNTuple API Review and Progress
Presented by Chris on Monday

2. Development for CAF data in RNTuple
Presented by Amit on Monday



Questions (and Answers)

What are the focus areas of the work in your technical area right now?

1. Continue RNTuple
2. Adding object store
 - Using references to avoid duplication. Using RNTuple with DAOS (because backend exists)
3. And data delivery
 - xrootd, HDF5, new SLAC effort

Questions (and Answers)

What are the two or three main priorities for the next 18 months (till the end of Year 2)? Do they match our stakeholders' priorities/requirements/use cases? How much effort would it take to deliver on these priorities? What other CCE areas should work with you on these?

- 1. RNTuple completing review, finding common solution to performance optimization,**
 - Requested/supported by ATLAS, CMS, DUNE,
 - 1 - 1 1/2 FTE
- 2. Object store,**
 - Supported by ATLAS, CMS,
 - 1 FTE, collaborate with PAW
- 3. Data delivery,**
 - Primarily targeted towards DUNE as well as ATLAS/CMS,
 - >1/2 FTE (for work planning), collaborate with PAW, SML

Questions (and Answers)

Based on what you have learned about the IRI/HPDF program from Debbie and Lavanya's talks, are any of CCE's priorities aligned with IRI's goals (1&2)? How could we benefit from IRI work(3)? Could CCE work help IRI progress(4)? Any suggestions for the Task Force idea(5-6)?

1. **Common data access interfaces for experiments**
 - To storage
 - Importing/Exporting data, including API, policies and permissions
 - Desired, common monitoring tools for site-specific problems.
2. **Same workflow interfaces**
3. **Dialog on whether Monarch/Tiered system can be matched with Hubs/Spokes.**
 - E.g. Tier1 (smaller Tiers?) as Spokes, advantages (better connection to Hubs), disadvantages?
4. **Providing use-cases and real-examples (make them not go off the deep end).**
5. **Find out about meetings, suggest infrequent/annual IRI/CCE meetings**
6. **CCE should include representatives for all its work groups in the task force.**

Questions (and Answers)

How would you use AI/ML methods to achieve your goals? Do you have in-house AI/ML expertise, or would you need to collaborate with some AI/ML specialists outside your area?

- **Data layout and caching optimization via AI/ML (e.g. in RNTuple, maybe object store, data access)**
- **Using AI/ML to safe-guard intelligent/lossy compression algorithms (compression)**
 - **Help identify quantities of interest that need to be preserved**
-
- **These activities are beyond SOP's original scope of work and if we would do them we would need outside AI/ML experts to engage.**

Thanks

- To everyone for all their work and especially
- The LBNL crew (plus Samantha) for organizing a great workshop