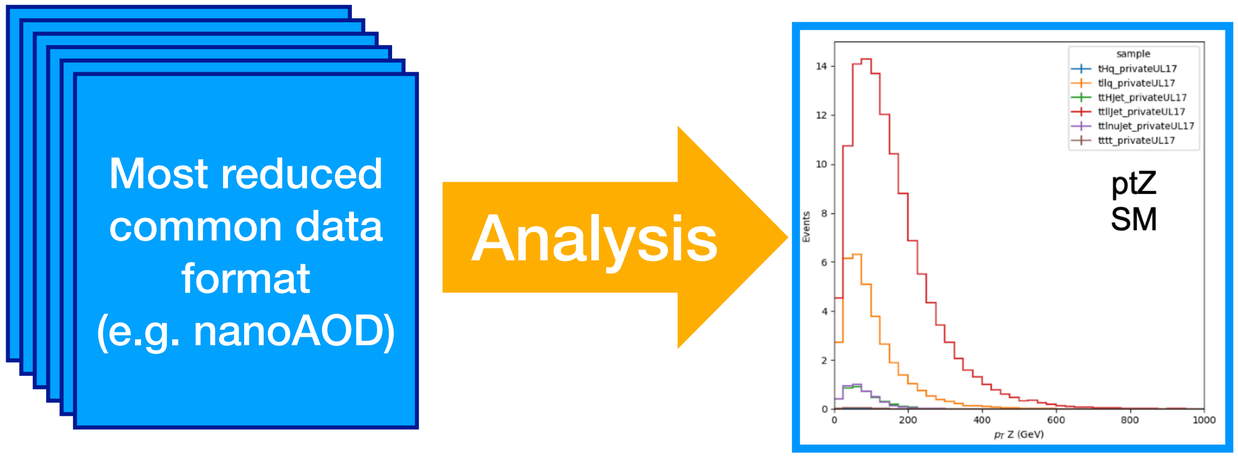
Analysis Cyberinfrastructure: Challenges and Opportunities

contribution to Snowmass 2021

K. Lannon, P. Brenner, M. Hildreth, K. Hurtado Anampa, A. Malta Rodrigues, K. Mohrman, D. Thain, B. Tovar



Cyberinfrastructure

||

Software + Hardware

Takeaway #1: Analysis Can be Done Fast…

…but usually isn't

- Given expected data volumes from LHC (from now to HL-LHC), analysis should take 10-60 min
- Common experience is usually that it takes "overnight"
- To get past this, we need
  - ~1000 CPU cores and sufficient I/O bandwidth (multiple data servers)
  - Robust, resilient software

Take Away #2: Many Layers of Software Stack

Mind the PieCaken!

- Ideal software stack: a layer cake
- Real software stack: a PieCaken

PieCakens are harder to bake. Require better integrated tools (full stack monitoring!), better documentation, and expert help with integration and operations.

Takeaway #3: You Can't Count on Users Tuning or Cooperating

Need automatic decent performance and coordination

- You can't expect users to tune their parameters for days to achieve a 10 min run time!
- What seem like small changes to the users (e.g. just change a command line argument) can have huge impacts in terms of analysis resource (CPU, memory, etc.) needs
- There is enough information available that analysis software can start from user guesses and adaptively find optimal parameters. (See here.)
- When you add more users, the system needs to be designed to coordinate among users to preserve desired performance. Coordination likely needs to come at a higher level than the batch system.