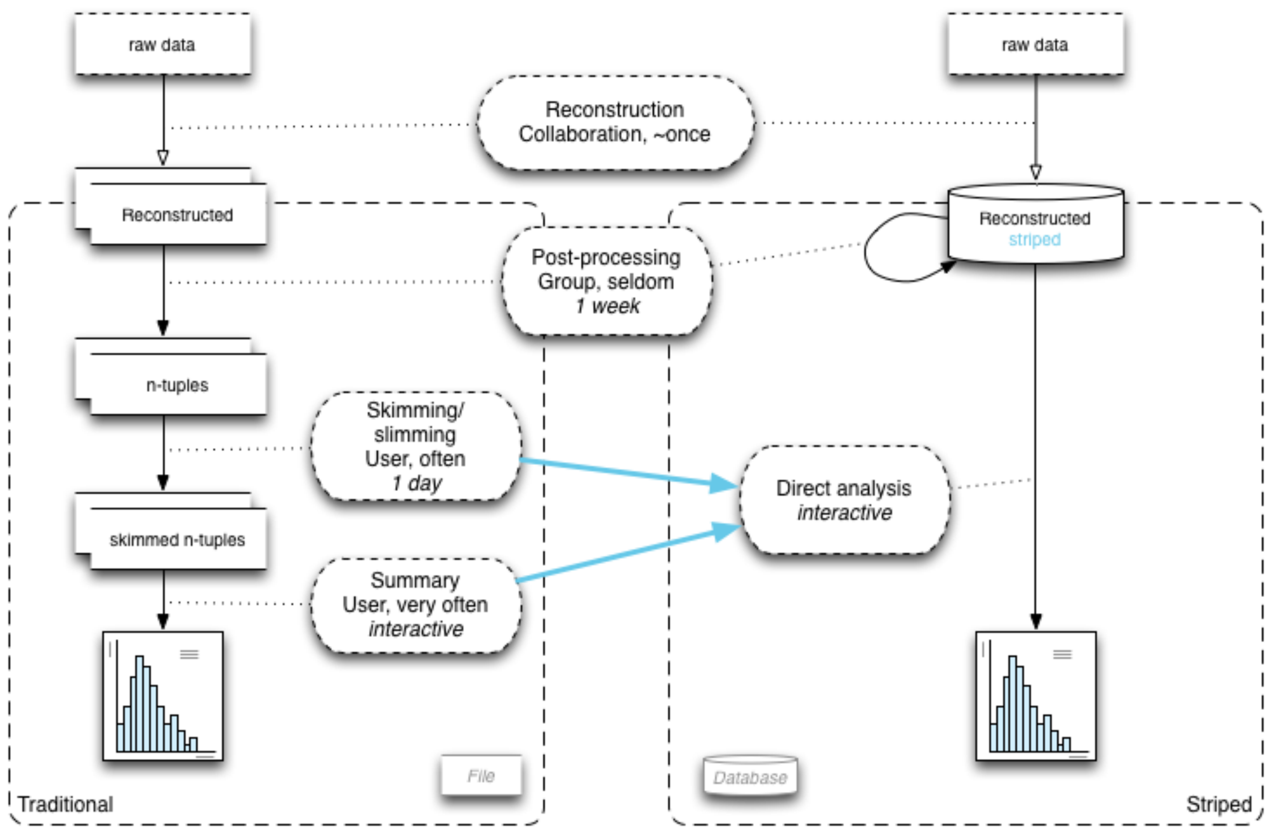
Striped Data Server for Scalable Parallel Data Analysis

Jin Chang (Caltech), Oliver Gutsche(FNAL), Igor Mandrichenko (FNAL), James Pivarski (Princeton University)

Goal: Reduce Time to Insight



Reduce analysis turn-around time

Programmatically, traditional analysis is iterative process, repeating:

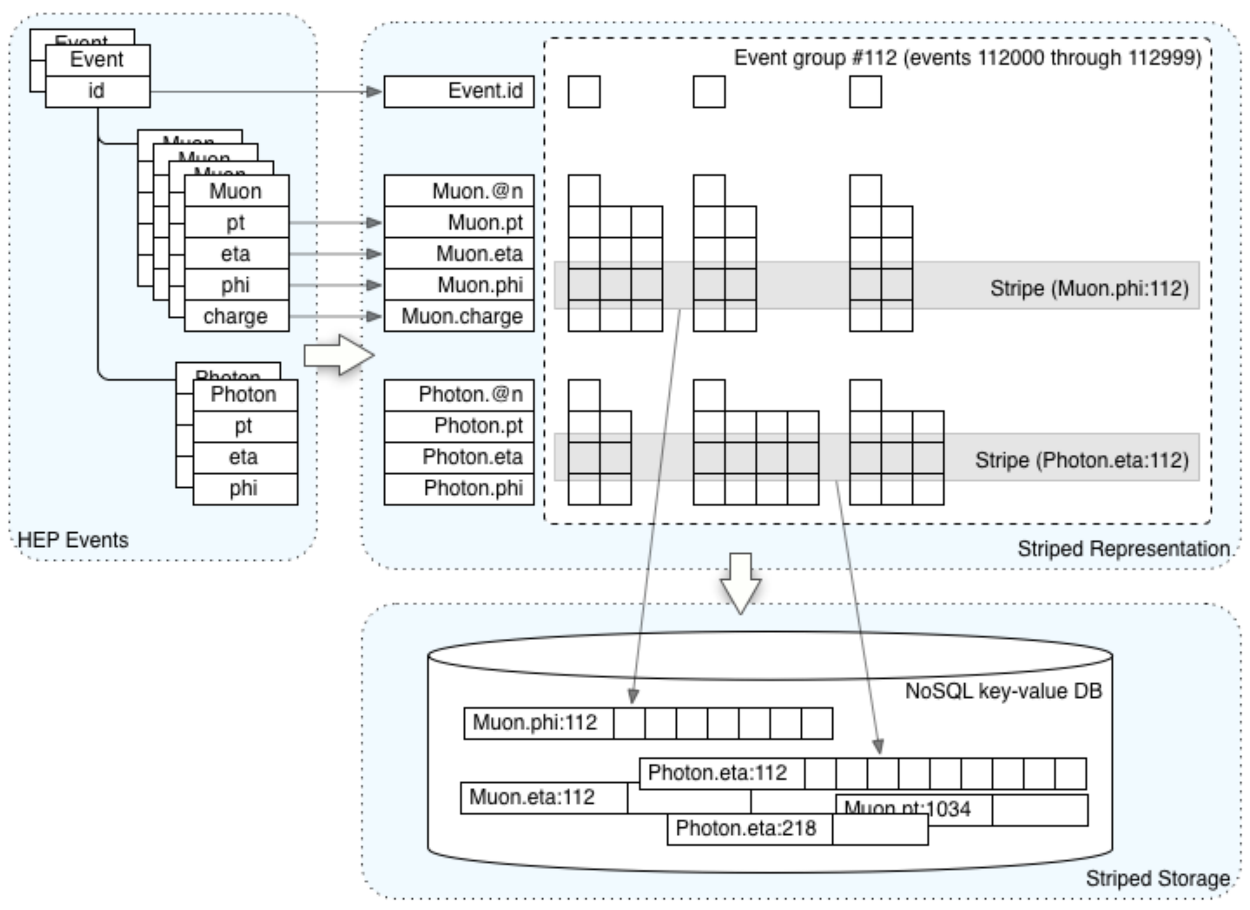
- Skimming (drop not interesting events, disk-to-disk)
- Slimming (drop unneeded attributes, disk-to-disk)
- Filtering (selectively read events into the memory)
- Pruning (selectively read attributes into the memory)

Provide access to the needed data and only to the needed data

- Direct, scalable, efficient

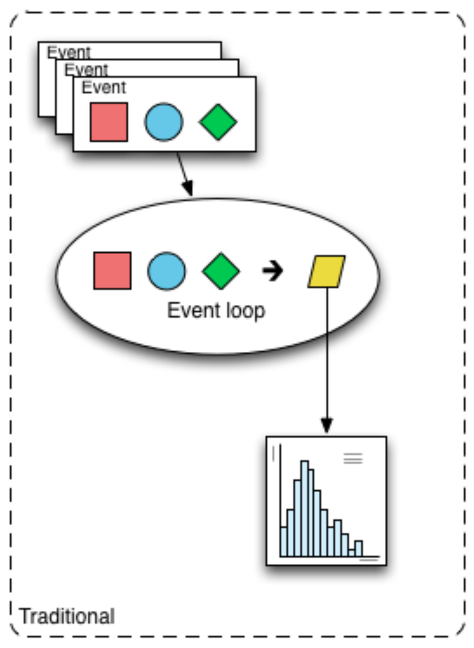
Eliminate the need to skim or slim data as a disk-to-disk operation

Innovation: Striped sata representation – move from file-based to database-based analysis paradigm



- Variation of *columnar representation*
- Columns are broken into *stripes* at the *event group*
  - 1K-10K events boundaries
- Data Stripe – one column of data for one event group
  - unit of representation efficiently stored in a key/value database
- Can be used for variety of non-HEP data

Stripe is a numpy data array, immediately consumable by numpy



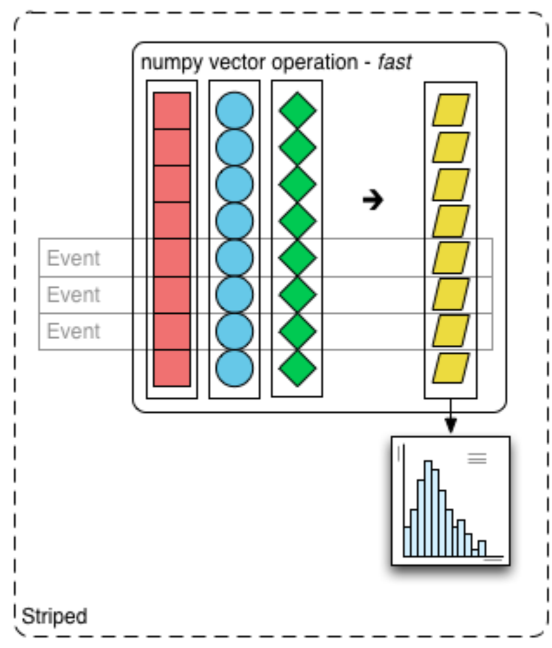
Traditional analysis:

Calculate missing quantities for each event, one event at a time in the event loop

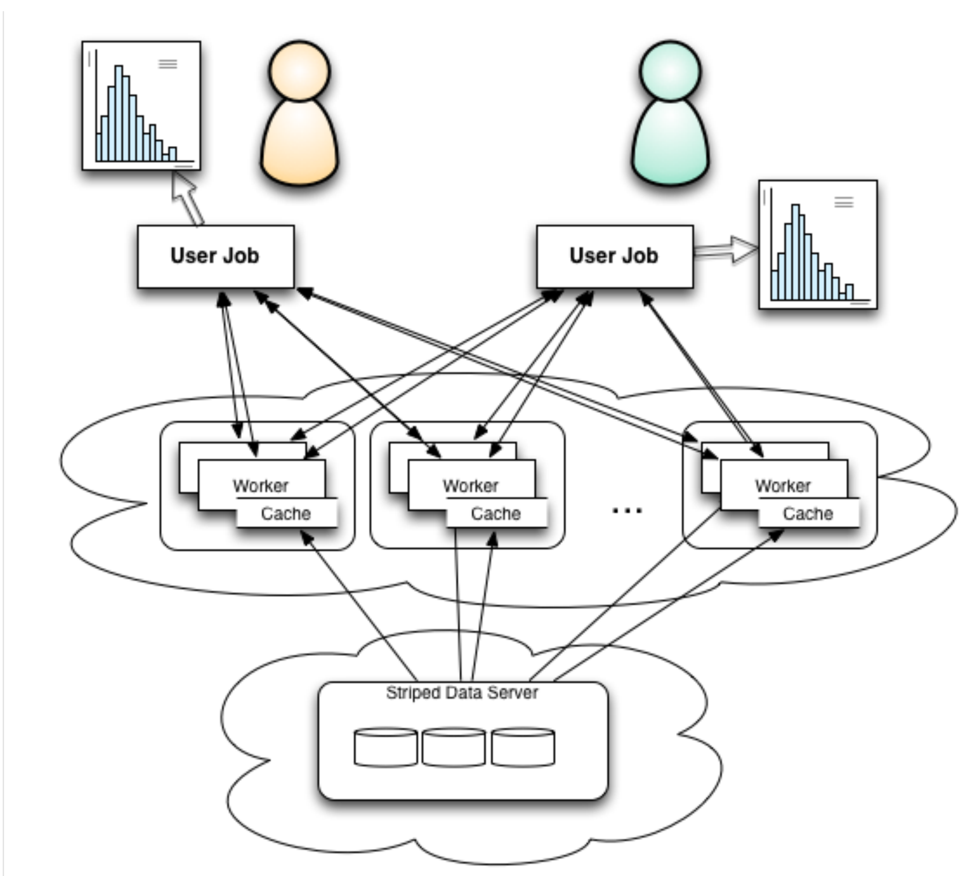
Striped analysis:

Move as much calculations as possible from the event loop to numpy vector operations over stripes

- Vector operations are much more efficient
- Can leverage power of GPU



Implementation: cloud-friendly client/worker/database architecture



Striped Data Server

- Distributed, scalable, redundant no-SQL key/value storage
- Web service with simple REST interface, web cache

Computing Component/Workers

- Worker is a single-threaded stateless process with its private data cache
- Cloud-ready: can be deployed elastically, using Docker containers

User code - Python, Jupyter notebook compatible

~1M events/second performance on 30 core demo 2-server cluster, CMS dark matter search dataset

- 1 worker per core, 13 node database no-SQL DB cluster

This project is funded by FNAL-LDRD-2016-032

Fermilab U.S. DEPARTMENT OF ENERGY