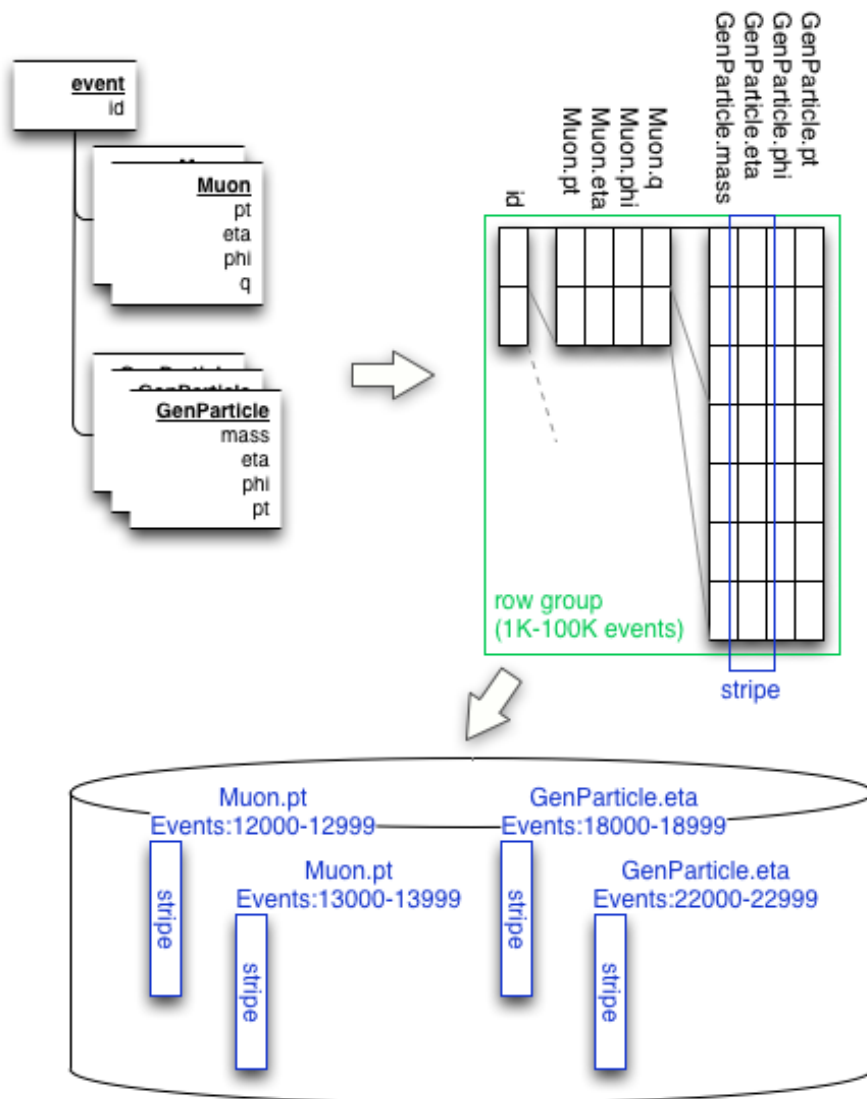# STRIPED EVENT STORAGE: UPDATE AND DEMO

Igor Mandrichenko

March 8, 2017

Striped Data Representation

- Unit of DB storage is a Stripe
- Stripe is a numpy array representing single leaf for a range of events
- Different leaves Stripes are split along the same event boundaries
  - Multiple stripes from same event range can be combined

Advantages of striped representation

- Only needed columns/events have to be downloaded

- Downloaded stripe is immediately useful via numpy without additional decoding

- Some calculations can be performed using numpy
  - E.g.:   Muon.p = Muon.pt * cosh(Muon.eta)

- Stripes are cacheable
  - Same repeating analysis processes
  - Different analysis processes using same pieces of data

Demo Structure
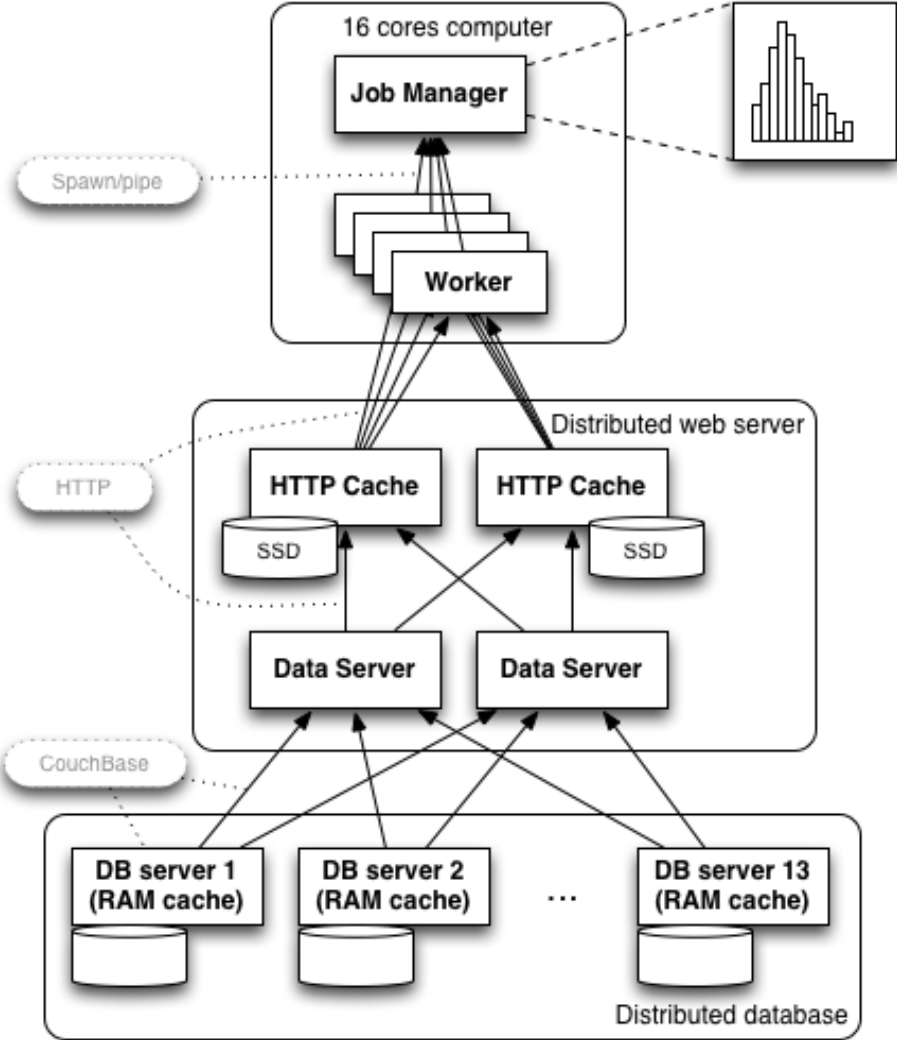
Update

- All the datasets stored in the database
- 1.95 TB
- Most datasets stored with 1000 events/event group (stripe)
  - Not most efficient
  - The most efficient seems to be 10,000 events/group
- Some are stored at 20,000 events/group
- http://dbdata0vm:9090/striped/app/datasets (takes time to load!)

Demo