Femtocode: querying HEP data

Jim Pivarski

Princeton University – DIANA

April 17, 2017

Distributed server

Scaling up to full datasets

Exploratory data analysis requires human-scale time to completion: seconds at most. If a query server takes much longer than this, physicists will go back to skimming.

Scaling estimates for one query:

- Several dozen samples, totaling $\mathcal{O}(10\ \text{TB})$.
- Every query runs over all events; a single query rarely uses as much as 1% of the *columns*. (Popularity distribution is steep.)
- Worst cases observed in early implementation: 30 ms/MB.
- Implies 3000 core-sec for that query: 3 seconds for 1000 cores.

Scaling up to full collaborations

diana hep

What about multiple users?

- Most analyses have significantly overlapping needs. Evidence: home-grown skimming frameworks select the same 10% of CMS MiniAOD (Bacon, Pandas, Cms3, TreeMaker).
- File I/O is more expensive than processing: ∼40 ms/MB versus ∼2 ms/MB. Major gains if users *share* cache.
- 10% of 10 TB of samples is 1 TB, which easily fits in RAM on a cluster of 1000 cores (hard to fit in one user's machine).
- Short-lived queries are less likely to use resources at the same time, so shortening latency also reduces contention.

The parameters of the final system depend on the hardware allocated for it, but improving software can steepen the performance per price.

Distributed system layout