

Scientific Data Management at Exascale

Dr. John K. Wu, Lawrence Berkeley National Laboratory, USA

Summary

Major recent scientific discoveries including Higgs boson and gravitational waves are results of massive data analyses. For example, the discovery of Higgs boson, also known as the god particle, because it gives particles their mass, was based on a large physics experiment that produces data at 600 terabytes per second, while the discoveries involving gravitational waves not only process a large amount of data but also integrate information from multiple astronomical catalogs. In this talk, we will describe some of the key data management technologies involved in such analysis operations and show examples we have developed. These tools include I/O libraries, indexing techniques, a data reduction package, and an automatic parallelization framework. In addition, we also briefly examine the flurry of new computing and storage technologies and speculate on the ever more exciting research opportunities.