

DNP17-2017-000167

Abstract for an Invited Paper
for the DNP17 Meeting of
the American Physical Society

Bridging the Particle Physics and Big Data Worlds

JAMES PIVARSKI, Princeton University

For decades, particle physicists have developed custom software because the scale and complexity of our problems were unique. In recent years, however, the "big data" industry has begun to tackle similar problems, and has developed some novel solutions. Incorporating scientific Python libraries, Spark, TensorFlow, and machine learning tools into the physics software stack can improve abstraction, reliability, and in some cases performance. Perhaps more importantly, it can free physicists to concentrate on domain-specific problems. Building bridges isn't always easy, however. Physics software and open-source software from industry differ in many incidental ways and a few fundamental ways. I will show work from the DIANA-HEP project to streamline data flow from ROOT to Numpy and Spark, to incorporate ideas of functional programming into histogram aggregation, and to develop real-time, query-style manipulations of particle data.