Introduction

Applications and experiments in all areas of science are becoming increasingly complex and more demanding in terms of their computational and data requirements. Some applications generate data volumes reaching hundreds of terabytes and even petabytes. Analyzing, visualizing, and disseminating these large data sets has become a major challenge and data intensive computing is now considered as the "fourth paradigm" in scientific discovery after theoretical, experimental, and computational science.

As scientific applications become more data intensive, the technologies of handling "Big Data" have gathered great importance. This necessity has made that applications have seen an increasing adoption on cloud infrastructures. The computing models, system software, programming models, analysis frameworks, and other clouds services need to evolve and accommodate them to face the challenge of big data applications.

DataCloud 2014 provides the scientific community a dedicated forum for discussing new research, development, and deployment efforts in running data-intensive computing workloads on Cloud Computing infrastructures. It focuses on the use of cloud-based technologies to meet the new data intensive scientific challenges that are not well served by the current supercomputers, grids or compute-intensive clouds. The workshop features keynote talk by Narayan Desai and 5 technical paper presentations. The workshop is sponsored by IEEE, ACM, and SIGHPC.