Apache Science Data Analytics Platform (SDAP)

Thomas Huang

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA (thomas.huang@jpl.nasa.gov)

An Analytics Center Framework (ACF) is an environment that enables the confluence of resources for scientific investigation. It harmonizes data, tools and computational resources which subsequently enable the research community to focus on the investigation. The Earth science community is an innovative community. We produce many tools and solutions to improve how we do science. In computer science, a framework is a reusable, semi-complete application that can be specialized to produce custom applications [Johnson:88]. Over the past years, NASA has been developing an open source ACF in collaboration with Apache Software Foundation, called the Apache Science Data Analytics Platform (SDAP). It is a big data analytics platform designed for cloud-based data management, analytics, match up, and data discovery services. It is a community-support, extensible open source GIS platform. The motivation is empowering the Earth and Space Science Informatics community to develop a common big data solution for the cloud and on-premise cluster. This talk describes the Apache SDAP and lesson learned from developing and moving SDAP in production to support various NASA projects.