Big Data Analytics and Visualization for Ocean Science

Thomas Huang (1), Edward Armstrong (1), Frank Greguska (1), Joseph Jacob (1), Nga Quach (1), Lewis McGibbney (1), Vardis Tsontos (1), Brian Wilson (1), Shawn Smith (2), Mark Bourassa (2), Jocelyn Elya (2), Steve Worley (3), Tom Cram (3), Zaihua Ji (3), Chaowei Yang (4), Yongyao Jiang (4), and Yun Li (4)

(1) NASA Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA, (2) Center for Ocean-Atmospheric Prediction Studies, 2000 Levy Avenue, Building A, Suite 292, Tallahassee, FL 32306-2741, USA, (3) National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA, (4) George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

What if I tell you we have the solution to generate decades of regional time series that is hundreds of times faster than the current methods without having to download a single file? How about the ability to compute daily differences on-the-fly? What about the ability to perform on-the-fly match-up with associated in situ measurements? How about being able to discover relevant measurements, past anomalies, news, and publications? Finally, how about being able to subset swath and gridded data in great speed? In recent years, the NASA's Advanced Information System Technology (AIST) and Advancing Collaborative Connections for Earth System Science (ACCESS) programs have invested significantly in JPL to develop various Big Data solutions in the areas of analytics, anomaly detection, in situ match-up, quality screening, and search relevancy. OceanWorks is a 2-year integration effort to develop an integrate data analytic platform to improve ocean science through partnership with NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC). OceanWorks is a cloud-based big data analytic solution developed using the open source Apache Science Data Analytics Platform (SDAP).

With increasing global temperature, warming of the ocean, and melting ice sheets and glaciers, tangible impacts can be observed from changes in anomalous ocean temperature and circulation patterns, sea-level rise, increasing extreme weather events, and more intense tropical cyclones and storm surges affecting coastlines. These impacts also may involve drastic changes and shifts in marine ecosystems. NASA's Surface Water and Ocean Topography (SWOT) mission, scheduled to launch in April of 2021, is expected to generate over 23PB of data for a nominal 3-year mission. This will challenge all existing NASA Earth Science data archival/distribution paradigms. It will no longer be feasible for scientists to download and analyze such volumes of data. OceanWorks will enhance the physical ocean archive with fast, web-accessible search, discovery, data analysis, and in situ data match-up in order to minimize data movement and provide access, including subset, only to the relevant data.