Science storms the cloud

Chelle Gentemann, Chris Holdgraf, Ryan Abernathey, Daniel Crichton, James Colliander, Edward Kearns, Yuvi Panda, and Richard Signell
(cgentemann@faralloninstitute.org)

The core tools of science (data, software, and computers) are undergoing a rapid and historic evolution, changing what questions scientists ask and how they find answers. Earth science data are being transformed into new formats optimized for cloud storage that enable rapid analysis of multi-petabyte datasets. Datasets are moving from archive centers to vast cloud data storage, adjacent to massive server farms. Open source cloud-based data science platforms, accessed through a web-browser window, are enabling advanced, collaborative, interdisciplinary science to be performed wherever scientists can connect to the internet. Specialized software and hardware for machine learning and artificial intelligence (AI/ML) are being integrated into data science platforms, making them more accessible to average scientists. Increasing amounts of data and computational power in the cloud are unlocking new approaches for data-driven discovery. For the first time, it is truly feasible for scientists to bring their analysis to the data without specialized cloud computing knowledge. Practically, for scientists, the effect of these changes is to vastly shrink the amount of time spent acquiring and processing data, freeing up more time for science. This shift in paradigm is lowering the threshold for entry, expanding the science community, and increasing opportunities for collaboration, while promoting scientific innovation, transparency, and reproducibility. These changes are increasing the speed of science, broadening the possibilities of what questions science can answer, and expanding participation in science.