Cloud-based data-proximate visualization and analysis

Ward Fisher

Unidata/UCAR, Unidata, Boulder, United States (wfisher@ucar.edu)

The rise in cloud computing, coupled with the growth of "Big Data", has lead to a migration away from local scientific data storage. The increasing size of remote scientific data sets increase, however, makes it difficult for scientists to subject them to large-scale analysis and visualization. These large datasets can take an inordinate amount of time to download; subsetting is a potential solution, but subsetting services are not yet ubiquitous. Data providers may also pay steep prices, as many cloud providers meter data based on how much data leaves their cloud service.

The solution to this problem is a deceptively simple one; move data analysis and visualization tools to the cloud, so that scientists may perform data-proximate analysis and visualization. This results in increased transfer speeds, while egress costs are lowered or completely eliminated. The challenge now becomes creating tools which are cloud-ready.

The solution to this challenge is provided by Application Streaming. This technology allows a program to run entirely on a remote virtual machine while still allowing for interactivity and dynamic visualizations. When coupled with containerization technology such as Docker, we are able to easily deploy legacy analysis and visualization software to the cloud whilst retaining access via a desktop, netbook, a smartphone, or the next generation of hardware, whatever it may be.

Unidata has harnessed Application Streaming to provide a cloud-capable version of our visualization software, the Integrated Data Viewer (IDV). This work will examine the challenges associated with adapting the IDV to an application streaming platform, and include a brief discussion of the underlying technologies involved.