Geophysical Research Abstracts
Vol. 18, EGU2016-14982, 2016
EGU General Assembly 2016

In-database processing of a large collection of remote sensing data: applications and implementation

Vladimir Kikhtenko (1,2), Elena Mamash (1), Dmitri Chubarov (1), and Polina Voronina (1)

(1) Institute of Computational Technologies SB RAS, Novosibirsk, Russian Federation, (2) Novosibirsk State University, Novosibirsk, Russian Federation

Large archives of remote sensing data are now available to scientists, yet the need to work with individual satellite scenes or product files constrains studies that span a wide temporal range or spatial extent. The resources (storage capacity, computing power and network bandwidth) required for such studies are often beyond the capabilities of individual geoscientists. This problem has been tackled before in remote sensing research and inspired several information systems. Some of them such as NASA Giovanni [1] and Google Earth Engine have already proved their utility for science.

Analysis tasks involving large volumes of numerical data are not unique to Earth Sciences. Recent advances in data science are enabled by the development of in-database processing engines that bring processing closer to storage, use declarative query languages to facilitate parallel scalability and provide high-level abstraction of the whole dataset.

We build on the idea of bridging the gap between file archives containing remote sensing data and databases by integrating files into relational database as foreign data sources and performing analytical processing inside the database engine. Thereby higher level query language can efficiently address problems of arbitrary size: from accessing the data associated with a specific pixel or a grid cell to complex aggregation over spatial or temporal extents over a large number of individual data files.

This approach was implemented using PostgreSQL for a Siberian regional archive of satellite data products holding hundreds of terabytes of measurements from multiple sensors and missions taken over a decade-long span. While preserving the original storage layout and therefore compatibility with existing applications the in-database processing engine provides a toolkit for provisioning remote sensing data in scientific workflows and applications. The use of SQL - a widely used higher level declarative query language - simplifies interoperability between desktop GIS, web applications and geographic web services and interactive scientific applications (MATLAB, IPython). The system is also automatically ingesting direct readout data from meteorological and research satellites in near-real time with distributed acquisition workflows managed by Taverna workflow engine [2].

The system has demonstrated its utility in performing non-trivial analytic processing such as the computation of the Robust Satellite Technique (RST) indices [3]. It had been useful in different tasks such as studying urban heat islands, analyzing patterns in the distribution of wildfire occurrences, detecting phenomena related to seismic and earthquake activity.

Initial experience has highlighted several limitations of the proposed approach yet it has demonstrated ability to facilitate the use of large archives of remote sensing data by geoscientists.

1. J.G. Acker, G. Leptoukh, Online analysis enhances use of NASA Earth science data. EOS Trans. AGU, 2007, 88(2), P. 14–17.
2. D. Hull, K. Wolsfencroft, R. Stevens, C. Goble, M.R. Pocock, P. Li and T. Oinn, Taverna: a tool for building and running workflows of services. Nucleic Acids Research. 2006. V. 34. P. W729–W732.
3. V. Tramutoli, G. Di Bello, N. Pergola, S. Piscitelli, Robust satellite techniques for remote sensing of seismically active areas // Annals of Geophysics. 2001. no. 44(2). P. 295–312.