EarthServer: a Summary of Achievements in Technology, Services, and Standards

Peter Baumann

Jacobs University, Bremen, Germany (p.baumann@jacobs-university.de)

Big Data in the Earth sciences, the Tera- to Exabyte archives, mostly are made up from coverage data, according to ISO and OGC defined as the digital representation of some space-time varying phenomenon. Common examples include 1-D sensor timeseries, 2-D remote sensing imagery, 3D x/y/t image timese
ries and x/y/z geology data, and 4-D x/y/z/t atmosphere and ocean data. Analytics on such data requires on-demand processing of sometimes significant complexity, such as getting the Fourier transform of satellite images. As network bandwidth limits prohibit transfer of such Big Data it is indispensable to devise protocols allowing clients to task flexible and fast processing on the server.

The transatlantic EarthServer initiative, running from 2011 through 2014, has united 11 partners to establish Big Earth Data Analytics. A key ingredient has been flexibility for users to ask whatever they want, not impeded and complicated by system internals. The EarthServer answer to this is to use high-level, standards-based query languages which unify data and metadata search in a simple, yet powerful way.

A second key ingredient is scalability. Without any doubt, scalability ultimately can only be achieved through parallelization. In the past, parallelizing cod
e has been done at compile time and usually with manual intervention. The EarthServer approach is to perform a samentic-based dynamic distribution of queries fragments based on networks optimization and further criteria.

The EarthServer platform is comprised by rasdaman, the pioneer and leading Array DBMS built for any-size multi-dimensional raster data being extended with support for irregular grids and general meshes; in-situ retrieval (evaluation of database queries on existing archive structures, avoiding data import and, hence, duplication); the aforementioned distributed query processing. Additionally, Web clients for multi-dimensional data visualization are being established. Client/server interfaces are strictly based on OGC and W3C standards, in particular the Web Coverage Processing Service (WCPS) which defines a high-level coverage query language.

Reviewers have attested EarthServer that "With no doubt the project has been shaping the Big Earth Data landscape through the standardization activities within OGC, ISO and beyond".

We present the project approach, its outcomes and impact on standardization and Big Data technology, and vistas for the future.