EarthServer-2: Agile Datacube Analytics

Peter Baumann, Angelo Pio Rossi, Vlad Merticariu

{p.baumann,ap.rossi,v.merticariu}@jacobs-university.de

Jacobs University, Bremen

Responding to:
- Innovative Open Science methods, applications and prototypes supporting daily scientific working routines
- Repositories, networks and working platforms for open scientific communication, collaboration, exchange and access to knowledge

Abstract
Big Earth Data at your fingertips - this is the vision of EarthServer (www.earthserver.eu), an intercontinental initiative for unleashing the potential of Big Data through a disruptive paradigm shift in technology:

- from isolated silos of data with disparate functionality towards a single, uniform information space;
- from a difficult, artificial differentiation between data and metadata access to unified retrieval;
- from zillions of files towards few whatever-size datacubes;
- from limited functionality to the freedom of asking anything, anytime, any server in a peer network of data centers worldwide.

Researchers from Europe, the US and Australia have teamed up to materialize the datacube paradigm. On massive spatio-temporal sensor, image, simulation, and statistics non-experts and experts alike can "ask any question, any time" thereby enabling them to "build their own product on the go". EarthServer has established client and server technology for such spatio-temporal datacubes rigorously using open standards, which the project also is shaping actively. Based on the European array engine, rasdaman, and US virtual globe, NASA WorldWind, a service federation is up and running incorporating about 250 TB today and planning to cross the PB frontier in a few months. Ad-hoc data fusion between ECMWF (UK) and NCI (Australia) has already been demonstrated.

The underlying datacube technology actually is domain-agnostic, and it is expected that, based on the ISO and OGC datacube standards, Earth, Space, Life, Social science and engineering as well as business will benefit.
In phase 1, FP7-EarthServer has established open ad-hoc analytics on massive Earth Science data, based on and extending leading-edge Array Database technology, rasdaman. According to EU Commission and phase 1 reviewers, rasdaman will "significantly transform the way that scientists in different areas of Earth Science will be able to access and use data in a way that hitherto was not possible". Six Lighthouse Applications with over 200 TB have demonstrated this. Additionally, EarthServer "with no doubt has been shaping the Big Earth Data landscape through the standardization activities within OGC, ISO and beyond". Phase 2, H2020 EarthServer-2, has started with an even more ambitious goal: crossing the Petabyte frontier for 3-D and 4-D datacubes, including intercontinental fusion. Leading data centers (ECMWF, PML, MEEO/ESA, NCI Australia, and Jacobs University operate already standards-conformant water, air, weather, and planetary services; now a standards-based federation is being built. At the time of this writing, the largest service is exceeding 250 TB, and ECMWF is working on establishing standards-based access to their 90+ PB climate archive, the largest worldwide.

Jacobs University is driver, working group chair, and editor for the “Big Geo Data” standards suite in OGC and, transcending the Earth science domain, it has initiated an extension of ISO SQL with “Array SQL” which is considered by many a game changer in Big Datacube technology. In 2016, rasdaman has been listed under the 100 most promising Big Data technologies by US magazine CIO Review.

Figure: Intercontinental datacube mix and match in the EarthServer initiative (left) and collection of geo portals utilizing the EarthServer platform, rasdaman (right)