Geophysical Research Abstracts
Vol. 13, EGU2011-4486, 2011
EGU General Assembly 2011
© Author(s) 2011

Boosting Scalability of OGC Standards on Massive Data Sets Through Database Technology

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

Users of large-scale raster archives demand increasingly flexible functionality, ranging from simple navigation and download to complex ad-hoc aggregation, fusion, and data analysis. One often adopted response to this is to add more and more individual functions to the service; however, this is not only unwieldy for the user, but also hard to maintain. Another approach is to offer a flexible query language, following the tremendously successful example of SQL.

The rasdaman ("raster data manager") system [10] follows this generic approach. Its raster query language, rasql, allows SQL-style declarative expressions on n-dimensional raster data of unlimited size. OGC-based access to raster objects is accomplished through rasdaman's petascope component which adds geo semantics to the pure raster server. Currently supported are WMS 1.3, WCS 1.1, WCPS 1.0, and WPS 1.0; the new WCS 2.0 [1][3] is being finalized. All W*S requests are transformed into rasql statements processed by rasdaman.

Rasdaman employs a multi-parallel middleware architecture: Raster data are partitioned and stored in tables of the some conventional DBMS. Adapters exist for PostgreSQL, MySQL, Oracle, DB2, Informix, and flat file system. Hence, all vector, raster, and meta data end up in the same database, which tremendously eases administration.

Internally, rasdaman heavily relies on storage and query optimization for fast access. We mention only a few due to space constraints. A storage layout language integrated with rasql allows tuning each incoming object for best disk access performance by choosing partitioning, compression, and indexing scheme [2]. Query rewriting assesses each incoming query against 150 rules to rephrase it into a more efficient, but semantically equivalent version if possible [4]. Just-in-time compilation translates suitable parts of an incoming query into C code, compiles it, links it into the server, and executes it; this code is cached for reuse by similar future queries [6]. A generalization of image pyramids to n-dimensional pre-aggregates has shown very satisfying speedup [5].

This approach has proven very successful in terms of usability, performance, and scalability. The SQL-style query interface provides a high-level, flexible access interface making it straightforward to develop flexible client applications. Experience with rasdaman also has had considerable impact on the design of the OGC Web Coverage Processing Service (WCPS) standard which defines an XQuery-flavored raster retrieval language bridging the OGC WPS and OGC WCS standards. In the VAROS project, a commercial WPS client composes WCPS requests dynamically and sends them to rasdaman. On Earthlook [8], open-source JavaScript clients act in a similar way.

Among the relevant related work is SciDB; however, concepts and implementations are in their infancy, not yet ready for thorough assessment. PostGIS Raster conveys a nice SQL integration, but scalability has not been visibly addressed up to now. Further research work includes AQL, AML, TerraLib, and MonetDB. In a hands-on evaluation carried out by former US NIMA in 2003 where over 700 vendors had been invited rasdaman was evaluated as "most favourable tool" for image management. At the ACM PoDS 2007 conference Rona Machlin has dubbed rasdaman "the most comprehensively implemented [array database] system". Meantime rasdaman looks back on many years of stable use with dozen-Terabyte size objects. Recently, rasdaman has been invited to become part of the OSGeo Live DVD which showcases outstanding open-source geo tools.

In our talk, we will briefly discuss the rasdaman architecture and then present a live demo of ad-hoc remote WCS, WCPS, and WPS access to 1-D through 4-D remote sensing, oceanography, and climate data.

References

1. Baumann, P. (ed.): WCS 2.0 Core. OGC document 09-110r3

2. Baumann, P., Fezyabadi, S., Jucovschi, C.: Putting Pixels in Place: A Storage Layout Language for Scientific Data. Proc. SSTDM'10, December 14, 2010, Sydney, Australia

3. Baumann, P.: The OGC Web Coverage Processing Service (WCPS) Standard. Geoinformatica 14(4)2010, pp. 447ff, DOI 10.1007/s10707-009-0087-2

4. Furtado, P., Baumann, P.: Storage of Multidimensional Arrays based on Arbitrary Tiling. Proc. ICDE'99, March 23-26, 1999, Sydney, Australia, pp. 328-336

5. Garcia Gutierrez, A., Baumann, P.: Using Preaggregation To Speed Up Scaling Operations on Massive Spatio-Temporal Data. Proc. ER-2010, Vancouver, Canada, November 1, 2010, pp. 188 - 201

6. Jucovschi, C., Baumann, P., Stancu-Mara, S.: Speeding up Array Query Processing by Just-In-Time Compilation. Proc. SSTDM'08, Pisa, Italy, December 15, 2008, pp. 408-413

7. Machlin, R.: Index-based multidimensional array queries: safety and equivalence. Proc. PODS'2007, pp. 175-184

8. www.earthlook.org, last seen 10-jan-2011

9. www.opengeospatial.org/standards/wcs, last seen: 10-jan-2011

10. www.rasdaman.org, last seen: 10-jan-2011