



EarthServer

Big Earth Data Analytics

Large Scientific Information Systems
 Research Group
 Head: Peter Baumann



Project Summary: EarthServer is establishing open access and ad-hoc analytics on extreme-size Earth Science data, based on and extending the rasdaman Array DBMS. Database query languages serve as c/s interface for "mix & match" access to multi-source, any-size, multi-dimensional spatio-temporal data and their metadata, across all Earth sciences – in short: *Big Earth Data Analytics*. These interfaces are strictly based on OGC (for geospatial information) and W3C XQuery (for metadata) standards. **Started** Sep 2011, runtime 3 years, 5.38m EUR budget (7m US\$).

rasdaman

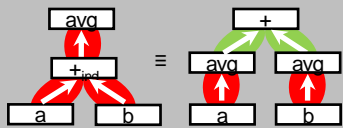


Rasdaman („raster data manager“) is a **multi-dimensional Array DBMS** [Baumann, VLDBJ 1994] adding n-D arrays as a new attribute type to relations. The rasdaman query engine extends SQL with n-D array operators, based on Array Algebra which also guides server-side optimizations. Arrays of unlimited size are partitioned („tiled“) and stored in a standard relational DBMS.

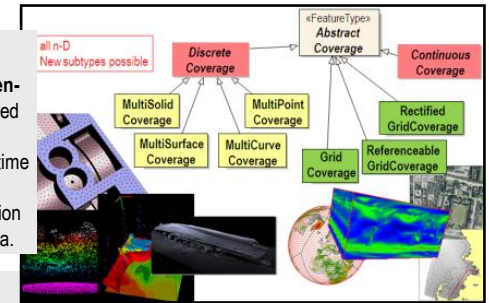
Incoming queries routinely undergo optimization, like: *tile streaming; adaptive compression; query rewriting (heuristics); physical operator clustering; just-in-time compilation; utilizing GPUs; dynamic preaggregation; tile caching; etc.* Benchmarks have shown that this achieves a speed-up typically by orders of magnitude.

Large-scale **sensor, image, model, & statistics** data can benefit from all classical database advantages, such as **information integration, flexible data retrieval, and scalability.**

```
select jpeg (img.green[x0:x1,y0:y1] > 130 )
from LandsatArchive as img
where avg_cells (img.red+img.green+img.blue) > 120
```



Coverages: A coverage [ISO 19123] is a „space-time varying phenomenon“. Coverage data are gathered in massive, rapidly growing amounts and variations, such as: 1-D sensor time series; 2-D maps; 3-D x/y/t satellite image time-series and x/y/z exploration data; 4-D x/y/z/t climate & ocean data.



WCS & WCPS

Coverages can be served via OGC **WCS** (Web Coverage Service). The WCS suite of standards establishes a modular set of extensions around a Core facilitating n-D spatio-temporal subsetting (trim & slice). One of these extensions is the OGC **WCPS** (Web Coverage Processing Service) Language, an n-D Array QL enriched with geo semantics and flavored along XQuery.

The following example [Baumann, Geoinformatica 2009] is "From MODIS scenes M1, M2, and M3, the absolute of the difference between red & nir, in HDF-EOS - but only those where nir exceeds 127 somewhere inside region R":

Rasdaman is WCS and WCPS reference implementation.

```
for $c in (M1, M2, M3),
   $r in (R)
where some($c.nir > 127 and $r)
return encode( abs( $c.red - $c.nir ), "hdf-eos" )
```

OGC

The Open GeoSpatial Consortium (OGC), in collaboration with ISO, develops standards for geo-spatial and location based services. Some coverage standards: P. Baumann (ed.): GML 3.2.1 Application Schema – Coverages, OGC 09-146r2 P. Baumann (ed.): WCS 2.0 Core, OGC 09-110r3 P. Baumann (ed.): WCPS Language, OGC 08-068r2

RTD Goals: +++ Extend support from arrays to coverages +++ databases & in-situ integration +++ data (WCPS) & metadata (XQuery) query integration +++ Federated query processing +++ Cloud parallelization +++ GIS integration +++ n-D visual clients, from mobile to immersive VR +++ Advance OGC standards

Lighthouse Applications: Six services are being established, together covering all Earth sciences, each a mix of at least 100 TB of in-situ and imported data.

Cryospheric Science
landcover

E-O-X

Airborne Science
high-altitude
long-endurance

NASA

Atmospheric Science
climate variables

MEEO
Metecological Environmental Earth Observation

Geology
geological models

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Oceanography
marine model
runs + in-situ

PML
PLYMOUTH MARINE LABORATORY

Planetary Science
Mars geology

JACOBS UNIVERSITY

Developer Partners:

Advisory Board: ESA | IEEE | OGC