IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323207

HOW DATABASES NOSQL HELPS TEACHING DATABASES, GEOMETRY, AND REMOTE SENSING SIMULTANEOUSLY

Peter Baumann

Jacobs University | rasdaman GmbH
Bremen, Germany
baumann@rasdaman.com

ABSTRACT

Array Databases, offering actionable datacubes, are increasingly used for flexible,s callable Earth data services. Technically, Array Databases converge several concepts that are also taught – separately – in science ang engineering: Linear Algebra as in Analytical Geometry, query languages as in databases, and image processing as in remote sensing. This begs the question: can such a tool be used for such diverse teaching in a meaningful way, thereby also provoking possibly surprising new correlations with students? In following this idea, in this contribution we explore how Array Databases can serve for educational purposes in undergraduate education.

***Index Terms*—** Array Databases, teaching, rasdaman, WCPS

1. OVERVIEW

Datacubes represent a general concept for serving multi-dimensional arrays as they appear, for example, in the Earth sciences as sensor, image (timeseries), atmospheric, and general statistics data. Commonly, datacubes are seen today as a cornerstone for serving Earth data in an analysis-ready manner. Two standards exist for datacube query languages: ISO SQL/MDA (Multi-Dimensional Arrays) for domain-agnostic datacube querying [5], and OGC Web Coverage Processing Service (WCPS) for Earth datacubes aware of the concepts of space and time [3][2][6]. In all cases, the common underlying mathematical concept is that of 1D vectors, 2D matrices, and 3D+ tensors.

As these are common to many subjects taught in science and engineering, and we have used the rasdaman datacube technology occasionally in teaching situations the idea has emerged to combine several normally distinct course topics into a single teaching unit [4]. Several advantages we anticipated: The interactivity of the query language allows experimentation with direct visual feedback. From a database perspective, minds are broadened to experience NoSQL data models as well as declarative query languages beyond SQL. In Analytical Geometry, declarative queries allow visualizing fields interactively. Finally, in geo sciences, learning about spatio-temporal analytics and corresponding standards could be accomplished.

In this contribution, we present three learning scenarios taken from spatio-temporal Earth data analysis, Analytical Geometry, and NoSQL database querying, all utilizing ISO SQL and OGC WCPS, respectively.

2. THE RASDAMAN ARRAY DBMS

The rasdaman ("raster data manager") datacube engine is a domain-neutral Array DBMS whose query language adds multi-dimensional array processing to SQL. A geo frontend provides space/time semantics and support for regular and irregular grids through OGC coverage standards compliant APIs. The architecture of rasdaman is a full-stack implementation in fast C++, boosted by a collection of optimization, parallelization, and federation techniques. The rasdaman technology has pioneered Array DBMSs [1]and today represents the de-facto standard in datacube engines. A spectrum of third-party clients is available for flexible extraction, visualization, and ML, ranging from Leaflet over QGIS to R and python (Figure 1). Services using rasdaman have exceeded 2.5 PB and 1,000-fold parallelization in AWS, and data centers have been federated on a planetary scale.

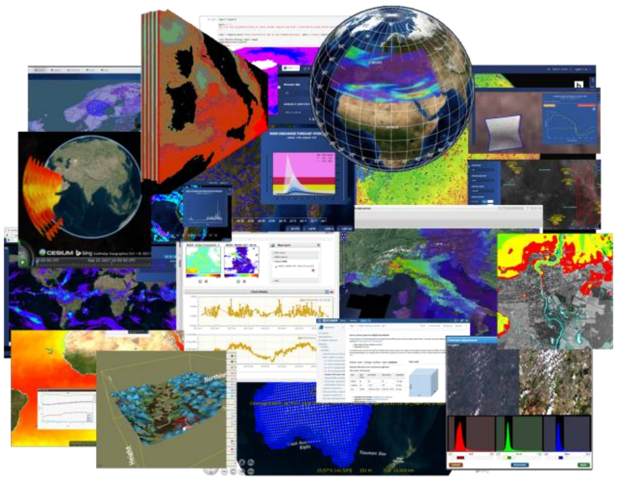

Figure 1 – Variety of third-party clients accessing datacubes, for matching skills

3. SPATIO-TEMPORAL EARTH DATA ANALYSIS

The OGC Web Coverage Processing Service (WCPS) Earth datacube analytics language allows high-level expression of complex spatio-temporal analysis on satellite imagery, image timeseries, climate timeseries, etc. WCPS abstracts away from algorithmic aspects and programming, allowing for a simplified formulat-ion of many common tasks. As processing is per-formed server side, experimentation on „Big Data“ can be done from small devices. Quota, as definable in rasdaman, protect students from sending overly resource-intensive queries.

Learning objectives:

- Formulation of spatial and temporal extraction as WCPS query
- Formulation of spatial and temporal analysis as WCPS query
- Interpretation of results plotted

Method: Introduce WCPS through the interactive demos provided by the EarthLook WCS demo portal [7]. Suggest problems (such as NDVI) of increasing complexity, and encourage students to define and solve own challenges, based on their Earth Science background. Datasets are avail-able through operation-al services providing Petabytes of Earth data.

Sample query:

```
for $ir in (INSPIRE_OI_IR)
return encode( (127*(1+
                ($ir.red-$ir.green)/($ir.red+$ir.green)
                )), "jpeg“ )
```

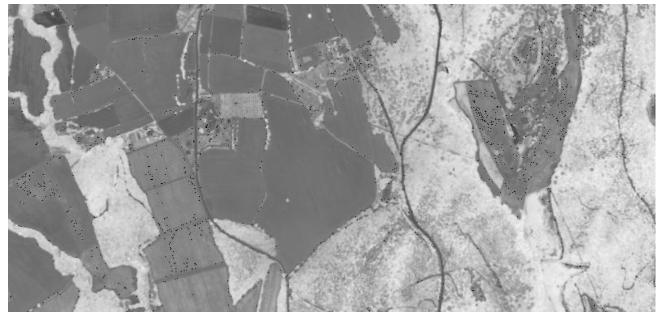

Figure 2 – Result of sample WCPS query

3. ANALYTICAL GEOMETRY

The query language allows expressing analytical geometry in a declarative manner, thereby enabling interactive visualization of functions and relations.

Learning objectives:

- Formulation of analytical geometry as a database query
- Interpretation of results plotted
- Application to spatio-temporal Earth data

Method: Presentation of various 2D functions and relations, with interactive coding and result analysis and interpretation.

Sample query:

```
select encode(
   127 * marray x in [1:500,1:500]
          values   (x[0]-250)*(x[0]-250)
                   + (x[1]-250)*(x[1]-250)
                   < 100*100,
  "png" )
```

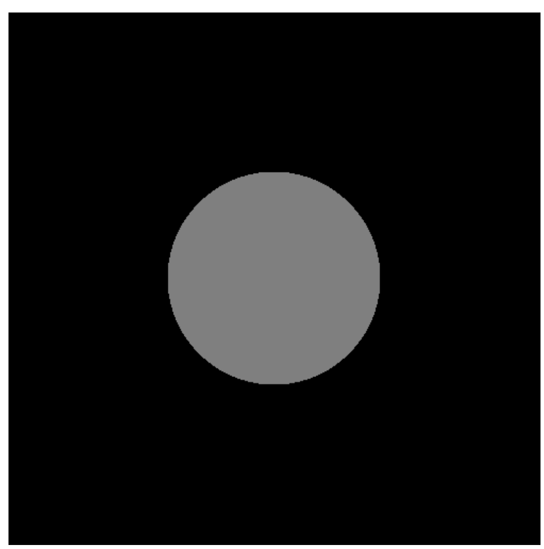

Figure 3 – Result of sample array query

4. LINEAR ALGEBRA AS DATABASE QUERY

Students normally get exposed to querying tables. NoSQL databases normally lack support for tensors and matrices as commonly needed in science and engineering, leaving the impression that databases mostly are useful in accounting departments, but not so in research and development departments. Array Databases close this gap. SQL/MDA corrects this impression by allowing server-side execution of Linear Algebra in a scalable manner.

Learning objectives:

- Arrays in science, engineering, business, and statistics
- Sample Big Array Data domain: spatio-temporal Earth data
- Declarative NoSQL querying
- How does physical data organization (here: tiling) affect performance?
- Exemplary array query optimization

Method: Interactive queries to highlight conceptual aspects and principles, but also showing physical aspects and optimizations of an array database engine.

Sample query (result is the scalar "3"):

```
select encode( count_cells(c[18,50,1:12].0 > 15),
               "csv")
from AverageChlorophyllScaled as c
```

5. CONCLUSION

Utilizing uncommon tools for teaching can have surprising benefits, and the surprise for students actually can stimulate curiosity. As databases since long are more than just passive data stores, but rather striving to support all of today's data structures – such as sets, hierarchies, graphs, and arrays –, and since databases are more relevant than ever in today's era of Big Data, there might be benefit in a playful inclusion of non-standard databases in undergraduate teaching.

ACKNOWLEDGEMENT

Research supported by EU H2020 EarthServer-1/-2, LandSupport, CopHub.AC, PARSEC, German BigDataCube, BigPicture, ORBiDANSe, DeepRain; and several more.

REFERENCES

[1] Baumann: Language Support for Raster Image Manipulation in Databases. Proc. Int. Workshop on Graphics Modeling, Visualization in Science & Technology, Darmstadt/Germany, 1992, pp. 236 – 245

[2] P. Baumann, D. Misev, V. Merticariu, B. Pham Huu: Datacubes: Towards Space / Time Analysis-Ready Data.. In: Service Oriented Mapping - Changing Paradigm in Map Production and Geoinformation Management, Springer 2018

[3] P. Baumann: The OGC Web Coverage Processing Service (WCPS) Standard. Geoinformatica, 14(4)2010, pp 447-479

[4] C. Bauzer Medeiros, P. Baumann, C. Jucovschi: Introducing multidisciplinary thinking in Computer Engineering: A new way of teaching database systems. Proc. IEEE Annual Global Engineering Education Conferencey (EDUCON'2010), 2010, Madrid, Spain

[5] ISO: 9075-15:2019 SQL/MDA (Multi-Dimensional Arrays)

[6] OGC: OGC Coverages / Datacube Standards. http://myogc.org/go/coveragesDWG

[7] N.n.: EarthLook – Big Earth Data Standards. http://www.earthlook.org