



## **Python Winding Itself Around Datacubes: How to Access Massive Multi-Dimensional Arrays in a Pythonic Way**

Vlad Merticariu, Dimitar Misev, and Peter Baumann  
Jacobs University, Bremen, Germany (v.merticariu@jacobs-university.de)

While python has developed into the lingua franca in Data Science there is often a paradigm break when accessing specialized tools. In particular for one of the core data categories in science and engineering, massive multi-dimensional arrays, out-of-memory solutions typically employ their own, different models.

We discuss this situation on the example of the scalable open-source array engine, rasdaman ("raster data manager") which offers access to and processing of Petascale multi-dimensional arrays through an SQL-style array query language, rasql. Such queries are executed in the server on a storage engine utilizing adaptive array partitioning and based on a processing engine implementing a "tile streaming" paradigm to allow processing of arrays massively larger than server RAM. The rasdaman QL has acted as blueprint for forthcoming ISO Array SQL and the Open Geospatial Consortium (OGC) geo analytics language, Web Coverage Processing Service, adopted in 2008. Not surprisingly, rasdaman is OGC and INSPIRE Reference Implementation for their "Big Earth Data" standards suite.

Recently, rasdaman has been augmented with a python interface which allows to transparently interact with the database (credits go to Siddharth Shukla's Master Thesis at Jacobs University). Programmers do not need to know the rasdaman query language, as the operators are silently transformed, through lazy evaluation, into queries. Arrays delivered are likewise automatically transformed into their python representation.

In the talk, the rasdaman concept will be illustrated with the help of large-scale real-life examples of operational satellite image and weather data services, and sample python code.