Geophysical Research Abstracts
Vol. 20, EGU2018-17939, 2018
EGU General Assembly 2018

Array Databases: Useful Technology for Earth Datacubes?

Dimitar Misev
Jacobs University, Bremen, Germany (p.baumann@jacobs-university.de)

Datacubes as an emerging paradigm in the Earth sciences are promising because they establish a homogenized, simplified view for users and they moreover add elegant capabilities for temporal analysis. Various approaches for implementation are being proposed recently which vary substantially: Array databases like rasdaman, file processing frontends like OPeNDAP, programming language pack-ages like numpy, python scripting based tools like Ophidia, array libraries like TensorFlow, and business logic extending MapReduce, such as SciHadoop.

This concept is not new in other domains – business intelligence utilizes OLAP datacubes since long, med¬ical imagery involves 3-D x/y/z CAT scans becoming 4-D datacubes when adding time, astrophysics deals with huge amounts of gridded optical and radio data, etc.

In database research, the field of Array Databases has been coined so as to support massive multi-dim¬ensional arrays at the same service quality as, say, tabular data. This involves provision of high-level dec¬larative query languages on n-D arrays as well as architectural approaches towards efficient pro-cessing of such queries. While some researchers in the Earth sciences tend to consider only “pleasingly parallel” operations (such as differential indices) when it comes to array services, Array Databases go far beyond this aiming at general Linear Algebra support.

Hence, it is a relevant question whether such technology can be leveraged for Earth science datacubes as well. While such datacubes, being gridded data, have arrays at their heart, they certainly are more than just the bare arrays – they require deep exploitation of associated metadata, such as information providing the semantics of space and time axis, adequate support for different types of irregularity, et.

In our talk we will present Array Database principles and contrast them with other datacube approaches in terms of functionality, architecture, and performance. The contribution is based on our work in the Research Data Alliance (RDA) Array Database Assessment Working Group whose findings will be summarized. Further, we present results from the EarthServer initiative where Petascale services are being set up and federated using Array Database technology.