Raster Databases

Peter Baumann
Jacobs University Bremen
Campus Ring 1, 28759 Bremen, Germany
+49-421-200-3178
p.baumann@jacobs-university.de

1. Arrays as New First-Class Database Citizens

Since the launch of Google Earth at the latest it is clear that online services for multi-Terabyte satellite imagery are becoming integral part of our Internet experience. Actually, 2-D imagery is but the tip of the iceberg - the general concept of multi-dimensional spatio-temporal raster data covers 1-D sensor time series, 2-D imagery, 3-D image time series (x/y/t) and exploration data (x/y/z), 4-D climate models (x/y/z/t), and many more.

The common data abstraction behind all these is that of a multi-dimensional array of some extent and cell ("pixel", "voxel") type. In the sequel, we will use the terms "raster" and "array" synonymously.

Today's inexpensive high-capacity storage media indeed allow to give online access to such data. Still, neither Google nor most data providers store their raster data in databases. Relational technology doesn't support arrays as first-class citizens. Mappings of data cubes to relations, which have been studied extensively in OLAP, fail on arrays in general, let alone for the dense population. Hence, standard database technology can only offer blobs (binary large objects) which, however, offer no nearly adequate functionality (such as spatiotemporal subsetting) and disastrous performance when implementing operations on top of them. Consequently, databases are well accepted in scientific communities for metadata management, and ignored for the bulk of data: the arrays themselves.

Domain-specific solutions, therefore, prevail as of today; recently, a trend towards integrated data services using off-the-shelf database technology can be observed. To some extent driven by geo services, an emerging trend is to not just offer navigation and extraction facilities, but advanced server-side analysis and processing. This poses new challenges on data management and, in particular, the design of open, interoperable services.

VLDB '07, September 23-28, 2007, Vienna, Austria.

2. Goal of the Tutorial

The tutorial will introduce to the emerging field of flexible, large-scale multi-dimensional raster services, ranging from simple, fast navigation to ad-hoc online analysis and mining.

The conceptual model presented grounds on the rasdaman ("raster data manager") array algebra. A safe, declarative query language, rasql, is derived from the algebra. The implementation architecture is based on partitioning arrays into sub-arrays, called tiles, which are stored as relational blobs. Query evaluation is tile-based which allows to avoid materialization of large objects in server main memory. Based on the array algebra, query optimization rules can be deduced which have been shown to boost performance sometimes by orders of magnitude. Storage management also involves optimization, such as finding suitable tiling layouts, but moreover addresses issues like extension from disks to tertiary storage, i.e., tape archives.

In the tutorial, high emphasis will be devoted to raster database support for geo and life sciences in particular; to this end, real-life use cases will be presented stemming from our project work carried out. Specifically for geo sciences our discussion is based on our standardization work in the Open GeoSpatial Consortium (OGC, www.opengeospatial.org).

Current status is that array databases have proven feasible in theory and industrial practice. However, many research issues exist (and continuously arise), making array databases a promising research field. A goal of this tutorial is to point out results achieved as well as research avenues.

3. Presenter's Bio

Peter Baumann is Professor of Computer Science at Jacobs University Bremen. His core research interest is large-scale multidimensional raster services and their application in geo, life science, Grid, and e-learning. He has published more than 60 book chapters and journal / conference articles, holds international patents, and has received international innovation awards. He is member of the Open Geospatial Consortium (OGC) and further bodies where raster services are being addressed.

See www.faculty.jacobs-university.de/pbaumann for more information.