



JACOBS
UNIVERSITY

Boosting Scalability of OGC Standards on Massive Data Sets Through Database Technology

EGU 2011

Vienna, 2011-apr-05

Peter Baumann

Jacobs University Bremen | rasdaman GmbH



Introduction



- Data providers are offering geo data since long
 - Usually simple ftp or similar
- In future: ad-hoc geospatial **analytics**
 - Disaster relief collaboration, science, ...
- This demo is about **operational complexity** using OGC standards
 - Specifically, on raster-type coverages
 - OGC Web Coverage Service (WCS) standard and Web Coverage Processing Service (WCPS) standard and EO-WCS (time permitting)





WCPS By Example

- "From MODIS scenes **M1**, **M2**, and **M3**, the absolute of the difference between **red** and **nir**, in HDF-EOS"

```
for c in ( M1, M2, M3 )
return
    encode (
        abs ( c.red - c.nir ),
        "hdf"
    )
```

(hdf_A,
hdf_B,
hdf_C)



WCPS By Example

- "From MODIS scenes **M1**, **M2**, and **M3**, the absolute of the difference between **red** and **nir**, in HDF-EOS"
 - ...but only those where nir exceeds 127 somewhere

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

(hdf_A,
hdf_C)



WCPS By Example

- "From MODIS scenes **M1**, **M2**, and **M3**, the absolute of the difference between **red** and **nir**, in HDF-EOS"
 - ...but only those where nir exceeds 127 somewhere
 - ...inside region R

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

(hdf_A)



rasdaman

www.rasdaman.org



JACOBS
UNIVERSITY

- „Raster data manager“ =
C/S **Array DBMS** for massive n-D raster data

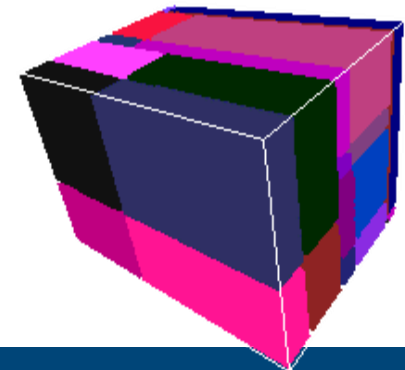
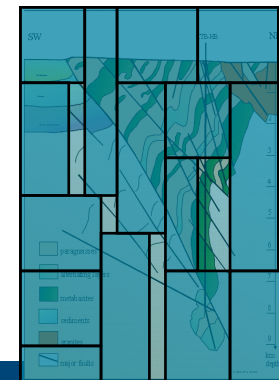
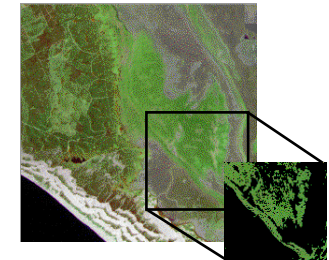
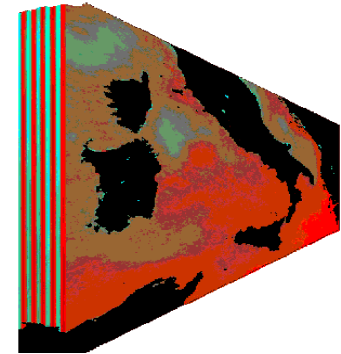
- storage & query optimization
- In operational use on dozen-TB objects

- rasql = declarative **array query language**

```
select img.green[x0:x1,y0:y1] > 130  
from LandsatArchive as img
```

- n-D array → set of n-D **tiles**

- tiles stored inside (!) DBMS in BLOBs
- arbitrary tiling (layout language)

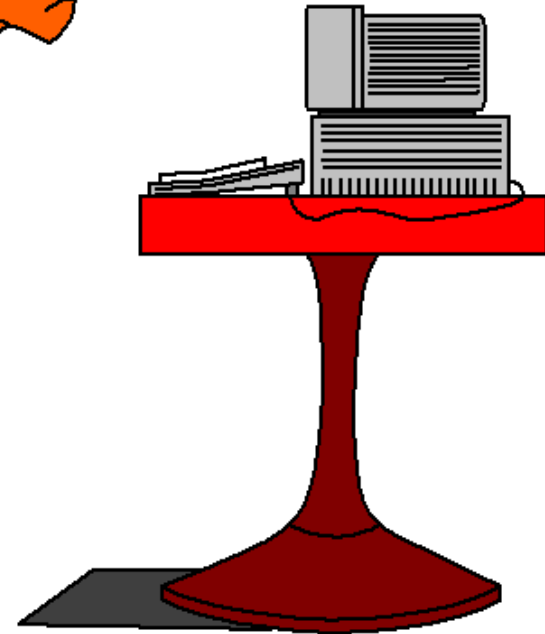


Demo



JACOBS
UNIVERSITY

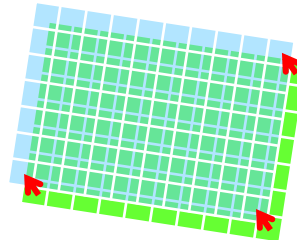
- <http://www.eoxserver.org/doc/en/users/demonstration.html>
- www.earthlook.org
- [rasdaman](#)



Query Optimization – Ex. 2

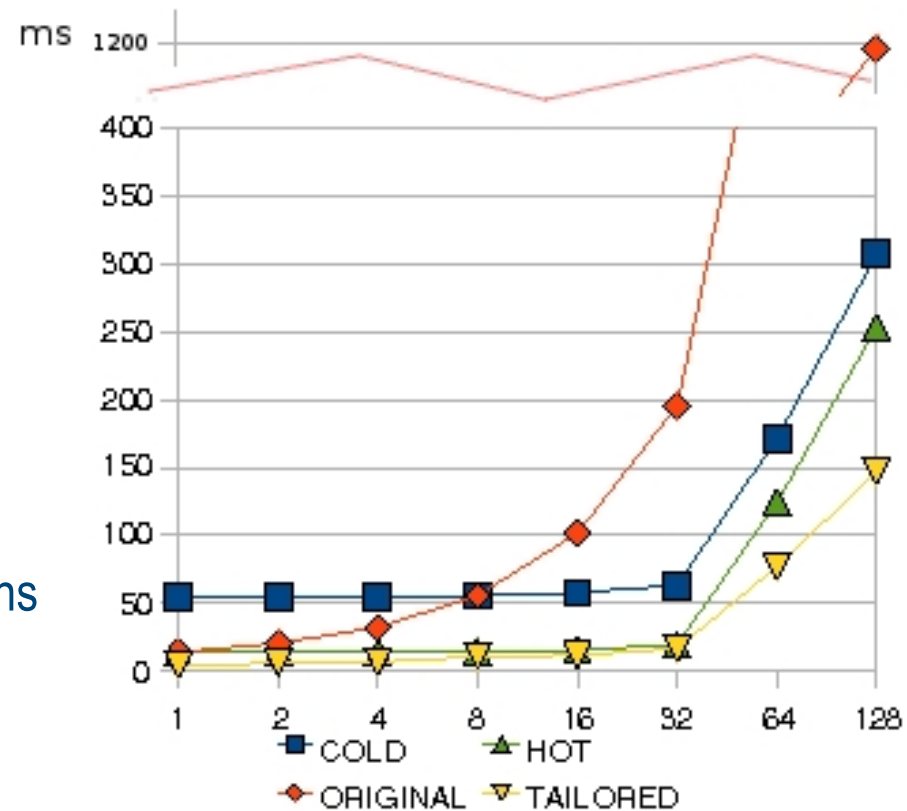


- Observation:
interpreted mode
slows down



```
select x*x*...*x  
from float_matrix x
```

- Approach:
 - cluster suitable operations
 - compile & dynamically bind
- Benefit:
 - Speed up complex, repeated operations
- Variation:
 - compile code for GPU

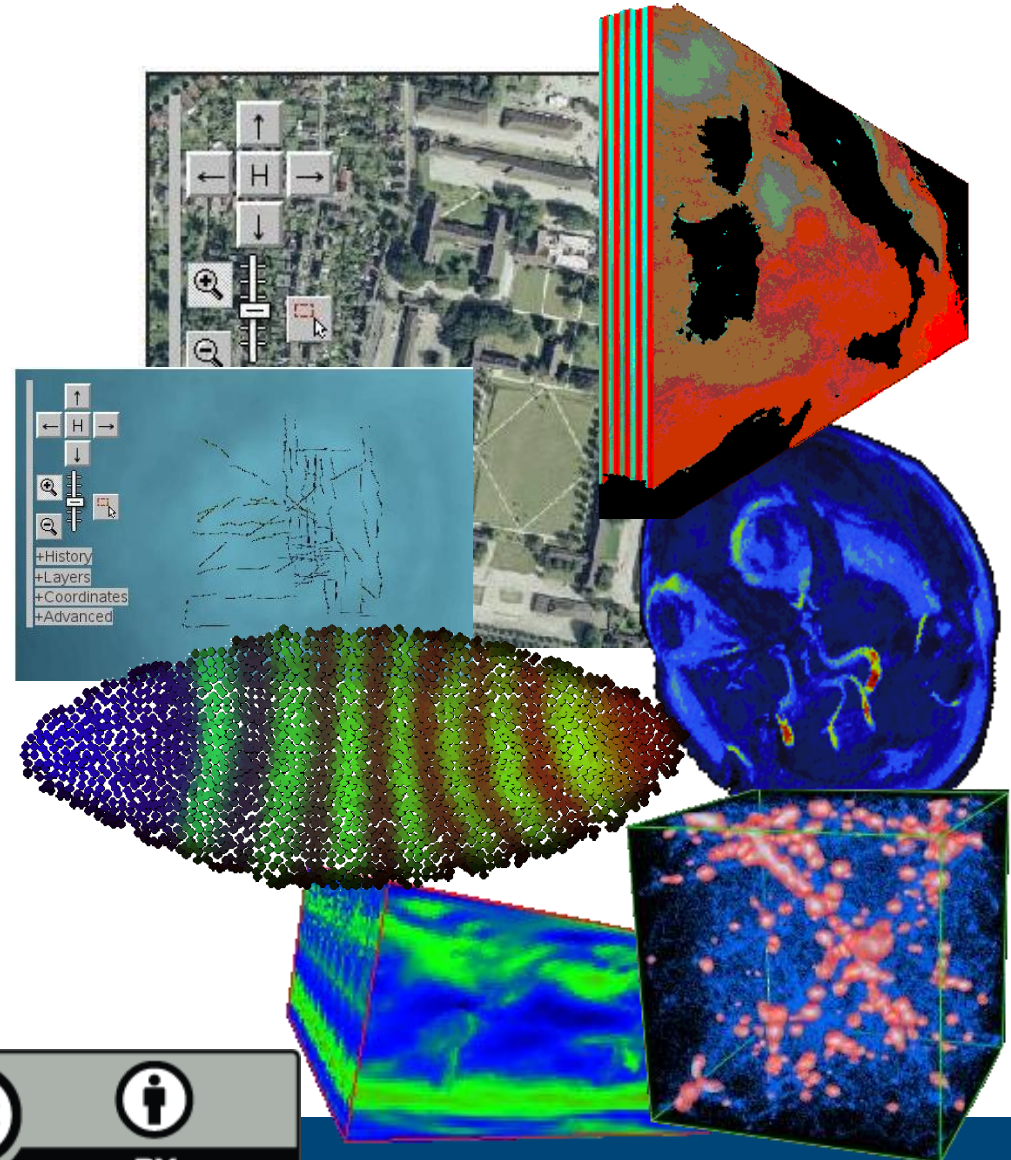


Optimization Techniques & Application Domains Studied



JACOBS
UNIVERSITY

- Adaptive tiling
- Adaptive compression
- Multi-dimensional indexing
- Distributed query processing
- Query rewriting
- Pre-aggregation
- Physical operator clustering
- Transparent tape integration
- Just-in-time compilation
- GPU processing
- Tile caching
- ...



Outlook: Research Directions



JACOBS
UNIVERSITY

- NASA considers using WCPS standard for ground/space interface
 - Satellite = ad-hoc analytics server
- **EarthServer** project: Scalable Earth Science Service Environment
 - 100 TB online analytics, incl. distributed fusion
 - 11 partners, ESA + NASA; start in Sep 2011
- Emerging research area:
Array Databases
 - workshop last week in Uppsala:
www.rasdaman.com/ArrayDatabases_Workshop
 - **New issues** like: imaging/database coupling



Conclusion

- **Array Databases** can give high-performance support on massive sets
 - Highly effective optimization techniques: adaptive tiling, query rewriting, JIT compilation, ...
- Still lots of research issues, but perspective of substantially improved **analytics & collaboration** support is clear
 - Both at acquisition time and for „long-tail science“
 - Extension from raster services to full coverage services
 - Server-side dynamic optimization helps!

