Background

- Array databases: we built rasdaman since 90s
  - Used in data centers, mapping agencies, etc
  - 100s of TB database sizes, 1,000+ cloud node parallelization
  - Dons life & space sciences, much earth science, including geo service standardization
- Me = computer scientist / database guy exposed to data centers and scientist users

Thoughts

- Can the HPC and Big Data communities learn from each other? How far down the HPC path do Big Data systems need to go? And, how should HPC change to meet its Big Data needs?
- DBers need to accept particularities of compute/data center environments
- HPCers need to learn value of flexible query interfaces
  - „we know the needs of our users“
- Ex Array Databases:
  - since long topic in HPC
  - But mostly for batch, not interactive, concurrent analysis

Thoughts

- What would the new software environments and stacks look like for HPC+Big Data systems?

- Different tools for differnt purposes
  - Possibly data coupled
- Ex: batch simulation, model coupling, user (query) frontends

Thoughts

- What are the workflow requirements, and will the tools and environments need to change for both domains?

- DB introduces way more dynamics than traditionally offered in HPC
  - Less predictable storage access, compute node workload
  - Users want variety of tools; maintenance!
  - Security issues
- Different workloads need to coexist (and collaborate), need for elastic orchestration

Bottom Line

- Accept each other's different thinking, it can ignite exciting new vistas