AVS: The Array Versioning System

Philippe Cudré-Mauroux
eXascale Infolab - U. of Fribourg - Switzerland

joint work w/ Adam Seering and Sam Madden - CSAIL - MIT

October 6, 2010
XLDB
N-dimensional array objects

Frequent insertions of new versions
- New observations
- New simulations

Arbitrary queries over version history
Problems w/ Versioning Systems

- Current Versioning Systems (e.g., SVN, GIT) are
  - Slow for large / numerous binary files
  - Catastrophic for spatial queries over series of versions
    - No spatial chunking
    - No co-location of versions
AVS Algorithms

- Given a workload of frequent / typical queries:
  - Optimal co-location of (N+1) dimensional chunks on disk
  - Optimal materialization / $\Delta$-compression of versions to minimize
    - Storage space
    - Query execution time
The Array Versioning System

Current Status

- Under development at MIT / UNIFR
- Supports arbitrary N-dimensional data and (N+1)-dimensional queries
- Supports version trees
- On-going Integration w/ SciDB

→ Orders of magnitude faster than svn / git on some workloads
AVS: The Array Versioning System

Philippe Cudré-Mauroux
exascale Infolab - U. of Fribourg - Switzerland

pcm@unifr.ch

October 6, 2010
XLDB