SciDB: A Science-Oriented Database System

Motivation and Interface

Existing RDBMSs Missing Key Features for Scientific Applications

- Array Oriented Data Model
- Append-only storage w/ support for versions, lineage, time travel
- In-Situ Data
- First-class support for UDFs
- Massively Parallel Computations
- Simple continuous model of uncertainty

SciDB goal is to provide these features in a community supported, open-source project

Array-oriented Data Model

- Multi-dimensional arrays
  - Integer-indexed dimensions
  - Cells contain scalars, UDFs, or arrays
  - Coordinate systems map from user-defined types to integer indices in enhanced arrays
  - Ragged arrays allow each row/column to have a different dimensionality

Query Language and Operators

- Array Specific Operators
  - Aggregate, Apply, Compose, Filter, Join, Lookup, Project, Regrid, Subsample
- Postgres-style User-defined Functions
  - E.g., to cook imagery or perform customized search
- C++ Language Bindings

Architecture

System Architecture

- Native language bindings
  - ORM-like model
- Shared nothing cluster parallelism
  - 10’s–1000’s of nodes
- Queries refer to arrays as if not distributed
- Query planner converts queries into plans for each node
- Run by node’s local executor/storage manager
- Runtime supervisor coordinates execution

Storage Manager

- Array-Oriented Storage Manager
  - Splits large arrays into series of multidimensional chunks
  - Stores dense-packed array chunks on disk
  - Chunks may overlap
  - Supports sparse and unbounded arrays
- Multi-node support
  - Arrays partitioned over cluster nodes
- Compression
  - Support for various types of compression (e.g., Run Length encoding, Null suppression, Huffman encoding)
  - Array attributes can be compressed individually using different schemes

Executor

- Basic primitive: Array-Oriented UDFS
  - Consumes and produce arrays
  - Built in ops internally written as UDFs
- Parallel execution – “Spit / Slurp”
  - Splits large arrays into overlapping chunks
  - Redistributions chunks over cluster nodes
  - Runs UDFs in parallel over cluster nodes
  - Flexible partitioning over dimensions or attributes
- Materialization strategies
  - Currently intermediate arrays stored in files
  - Future versions will pipeline execution

Demonstration Application: LSST

- Telescopic images of the sky
  - 55 Physes of raw imagery
- “Raw” Imagery “cooked” to find “observations” of astronomical objects (e.g., stars, galaxies)

Demonstration

Demo Query Workflow

- Filter query
  - Find observations in a particular region with flux above a threshold

- Regrid/cook query
  - Compare observations in a region to observations when region is re-cooked with a different algorithm

Conversion to XML & Distribution

Query plan run on all nodes, results merged to produce final answer at client

Coming Soon...

- Query Optimizer
  - Exploit partitionable, commutative operations
- Query execution on compressed data
- Shared scans
- Provenance and Versioning
  - Time travel for data, meta-data and UDFs
- Ability to re-derive any computation
- Uncertainty

SciDB is supported by NSF Grants OCS-0454423, 05-0704624, 06-0463022, gifts from Microsoft Research, and Prof. Balazinska's Microsoft Research New Faculty Fellowship.