SciDB Progress

Kian-Tat Lim
SLAC National Accelerator Lab

for the SciDB Team

XLDB3, 29 Aug 2009
History

XLDB1
Science/Academia Meeting
XLDB2
Features

Array data model, attribute-based storage

Parallel, scalable, elastic, extensible

Versioning, provenance, uncertainty, ...

Attribute-based storage gives advantages of columnar databases. Versioning, provenance, uncertainty, and other features yet to come, but planned for in design.
Array Data Model

Ragged N-dimensional arrays.
Each cell contains a tuple of attribute/value pairs.
Attributes can contain nested N-dimensional arrays (to one level of nesting currently).
Why Arrays?

Familiar interface

Ordering

Physical relationships

Differences from Oracle: can distribute arrays
Differences from Monet column arrays: multidimensional
Short-Term Goals

- Develop array algebra
- Build prototype for demo at VLDB
Team

Implementation

Design

Steering
# Team: Implementation

<table>
<thead>
<tr>
<th>Philippe Cudre-Mauroux</th>
<th>Roman Simakov</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmitry Gorchakov</td>
<td>Artyom Smirnov</td>
</tr>
<tr>
<td>Hideaki Kimura</td>
<td>Emad Soroush</td>
</tr>
<tr>
<td>K-T Lim</td>
<td>Pavel Velikhov</td>
</tr>
<tr>
<td>Jennie Rogers</td>
<td>Daniel Wang</td>
</tr>
</tbody>
</table>
## Team: Design

<table>
<thead>
<tr>
<th>Magda Balazinska</th>
<th>Ugur Cetintemel</th>
<th>Sam Madden</th>
<th>Pavel Velikhov</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacek Becla</td>
<td>Philippe Cudre-Mauroux</td>
<td>Dave Maier</td>
<td>Daniel Wang</td>
</tr>
<tr>
<td>Paul Brown</td>
<td>David DeWitt</td>
<td>Jignesh Patel</td>
<td>Stan Zdonik</td>
</tr>
<tr>
<td>Mike Carey</td>
<td>K-T Lim</td>
<td>Mike Stonebraker</td>
<td>others...</td>
</tr>
</tbody>
</table>

---

SLAC

XLDB3 Workshop
August 28–29, 2009
## Team: Steering

<table>
<thead>
<tr>
<th>Jacek Becla</th>
<th>K-T Lim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ugur Cetintemel</td>
<td>Andy Palmer</td>
</tr>
<tr>
<td>David DeWitt</td>
<td>Mike Stonebraker</td>
</tr>
<tr>
<td>Bobbi Heath</td>
<td>Stan Zdonik</td>
</tr>
</tbody>
</table>
Architecture

Client API

Executor

Operators

Storage

Executor

Operators

Storage

Executor

Operators

Storage
Each input cell (e.g. the gray one) can contribute to multiple output cells (red).
Each output cell receives data from multiple input cells.
Future

Rebuild with clean interfaces

Open design and source

Integrate with other projects