Adaptive Vectorized Query Planning

Luciano Arango
Lukas Maas (Advisor)
Things to Consider

Memory Bound

Full scans often preferable

Exponential growth in data
Predicate Ordering

Select * from table where $A > x$ and $B < y$

Predicate:

- $A > x$
- $B < y$

Selectivity:

- 33%
- 83%

8 pages*

11 pages*
Varying Selectivity

Select * from table where $A > x$ and $B < y$

Predicate: $A > x$, $B < y$

Selectivity: 33%, 50%

Query Planner: $A - B$
Varying Selectivity

Select * from table where $A > x$ and $B < y$

Predicate: $A > x$  $B < y$
Selectivity: 33%  50%

Query Planner: $B_1 A_1 A_2 B_2$
Clusters of Correlated Data

<table>
<thead>
<tr>
<th>Datetime</th>
<th>Country</th>
<th>Continent</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>03/27/2016 23:20:01 GMT</td>
<td>China</td>
<td>Asia</td>
<td>7.5</td>
</tr>
<tr>
<td>03/27/2016 23:19:59 GMT</td>
<td>China</td>
<td>Asia</td>
<td>3.2</td>
</tr>
<tr>
<td>...</td>
<td>Thailand</td>
<td>Asia</td>
<td>4.4</td>
</tr>
<tr>
<td>03/27/2016 15:13:10 GMT</td>
<td>USA</td>
<td>N. America</td>
<td>8.9</td>
</tr>
<tr>
<td>...</td>
<td>USA</td>
<td>N. America</td>
<td>7.2</td>
</tr>
</tbody>
</table>
Foundational Work

MonetDB/X100: Hyper-Pipelining Query Execution
Adaptive Predicate Ordering

Do{
    data = next_vector(vec_sz)
    stats = select(pred_ord, data, results)
    pred_ord, vec_sz = query_plan(stats)
}
while (data)
But….

- P. Bizarro, S. Babu, D. DeWitt, and J. Widom. [Content-Based Routing: Different Plans for Different Data](#).
- S. Babu and J. Widom. [StreaMon: An Adaptive Engine for Stream Query Processing](#).
- C. Olston, J. Jiang, and J. Widom. [Adaptive Filters for Continuous Queries over Distributed Data Streams](#).
Adaptive Query Planning

\[
\text{Do}\{
\text{data} = \text{next\_vector}(\text{vec\_sz})
\text{??} = \text{select}(\text{pred\_ord}, \text{data}, \text{results})
??, \text{vec\_sz} = \text{query\_plan}(??)
\}
\text{while (data)}
\]
Numa Aware

Entropy Measurements

Selectivity:
- 50%
- 50%

Qualifying Row:
<table>
<thead>
<tr>
<th>ID</th>
<th>1</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 4, 8</td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>20, 22, 27</td>
<td>15</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>45, 47, 49</td>
<td>20</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>47</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>58</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>67</td>
<td></td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

CPU Cache vs
Entropy continued

Selectivity: 50%  50%
Entropy: 4.32  3.3
Class ideas?