Real-Time Quantification Filters for Multidimensional Databases

Peter Strohm, Jedox AG
Jedox: In-Memory OLAP Database

2002
Founded in Freiburg, Germany

Today
- 100+ Employees
- Offices in Freiburg, Frankfurt, Düsseldorf, Paris
- 100+ Business partners globally

Jedox Suite
Version 5.1
Business Intelligence, Analytics & Performance Management
Excel-, Web-, Mobile-Client
GPU Acceleration

Jedox OLAP Server
Jedox ETL
Jedox SAP Connector
ERP, CRM, SCM
RDB, DWH
SAP/R3
SAP BI/BW
ODBO
XMLA
3rd Party Tools

GPU Accelerator
What is an OLAP-Database?

1. Multidimensional Cube
2. Hierarchical Structure
3. Consolidated Elements
4. Elements as “dimension path, value” pairs
In-Memory OLAP-Database

All data in main memory

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Q1</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Q2</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>All regions</td>
<td>27</td>
<td>14</td>
<td>14</td>
<td>55</td>
<td>29</td>
<td>20</td>
<td>10</td>
<td>59</td>
<td>20</td>
<td>18</td>
<td>15</td>
<td>53</td>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Europe</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>18</td>
<td>10</td>
<td>11</td>
<td>0</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>15</td>
<td>32</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>France</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>11</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Italy</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UK</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>North America</td>
<td>20</td>
<td>9</td>
<td>8</td>
<td>37</td>
<td>19</td>
<td>9</td>
<td>10</td>
<td>38</td>
<td>7</td>
<td>14</td>
<td>0</td>
<td>21</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>USA</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Canada</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mexico</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Deviation
Actual Budget

All data in main memory
In-Memory OLAP-Database

1. All data in main memory
2. Store only base elements
In-GPU-Memory OLAP-Database

1. All data in **GPU** memory
2. Store only base elements
3. Store only non-zero values
4. Calculate consolidated elements “on the fly”
5. Save Memory, be up to date
What is a Quantification Filter?

1. ANY and ALL Quantifier on one dimension
2. Conditional Filter

Example: Time period with any element > 10
What is a Quantification Filter?

1. ANY and ALL Quantifier
2. Conditional Filter

Ex. Time period with any element > 10

Ex. Region with all elements < 10
What is a Quantification Filter?

1. ANY and ALL Quantifier
2. Conditional Filter
Ex. Time period with any element > 10
Ex. Region with all elements < 10

Self-Service Business Intelligence, Analytics & Performance Management.
www.jedox.com - @JedoxAG - @PSJedox - #GTC14 - #S4395
Quantification Filter: Challenges

Pre-processing, e.g. Aggregation, Rules, etc.

Condition for one dimension, e.g. value > 5
## Quantification Filter: Challenges 2

| Is AnyProcessor | Is Zero In Result? | | FALSE | All & & 0 included | Any & & 0 excluded |
|-----------------|--------------------|---|-------------------|---------------------|
| TRUE            | TRUE               | TRUE | Any & & 0 included | FALSE               |
|                 |                    | satisfied | Any & & 0 excluded |
|                 |                    | Not satisfied | satisfied |
|                 |                    | Counter != sliceCellCount | flag > 0 |
| FALSE           | FALSE              | FALSE | All & & 0 included | All & & 0 excluded |
|                 |                    | Not satisfied | Not Satisfied |
|                 |                    | flag == 0 | Counter == sliceCellCount |
Quantification Filter: Algorithm

Preprocessed Cells

Any/All Processor

- Check cell condition?
  - No: Count
  - Yes: Put cell into hash table
    - Insert zeros
      - Check
      - Post-processing

Result Cells

Multi-GPU
Quantification: Algorithm

1. Using hash table
2. Skipping already checked elements
3. Avoiding atomics and locks
Wikipedia Page Stats Example

1. Starting point: Big Data (743GB reduced to 2GB)
2. Getting the data into the cube
3. Getting amazing speed-up with GPU

See also: www.saphana.com, www.wikipedia.org, blog.gbrueckl.at
Wikipedia Page Stats cube

1. Pages (1,2 Million)
2. Languages (~360)
3. Hours (24), Date (~360)
4. Projects (16)
5. Measures (~4)

Cube has ~2,48 Trillion possible cells (about 276 Million filled)
Wikipedia Example: Superbowl

<table>
<thead>
<tr>
<th>Wikipedia Example</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Super_Bowl</td>
<td>946.783,00</td>
</tr>
<tr>
<td>Frank_Ocean</td>
<td>919.531,00</td>
</tr>
<tr>
<td>Tunguska_event</td>
<td>909.623,00</td>
</tr>
<tr>
<td>Gangnam_Style</td>
<td>897.481,00</td>
</tr>
<tr>
<td>Martin_Luther_King,_Jr.</td>
<td>893.704,00</td>
</tr>
<tr>
<td>Baltimore_Ravens</td>
<td>809.333,00</td>
</tr>
<tr>
<td>Joe_Flacco</td>
<td>777.894,00</td>
</tr>
<tr>
<td>Mardi_Gras</td>
<td>768.077,00</td>
</tr>
<tr>
<td>Mumford_%26_Sons</td>
<td>757.956,00</td>
</tr>
<tr>
<td>List_of_Super_Bowl_champions</td>
<td>747.618,00</td>
</tr>
<tr>
<td>2013_in_UFC</td>
<td>724.661,00</td>
</tr>
<tr>
<td>George_Washington</td>
<td>702.503,00</td>
</tr>
<tr>
<td>Michael_Oher</td>
<td>692.753,00</td>
</tr>
<tr>
<td>Chinese_zodiac</td>
<td>633.067,00</td>
</tr>
<tr>
<td>Mohandas_Karamchand_Gandhi</td>
<td>622.810,00</td>
</tr>
<tr>
<td>Roman_numerals</td>
<td>551.791,00</td>
</tr>
<tr>
<td>List_of_Downton_Abbey_episodes</td>
<td>541.991,00</td>
</tr>
<tr>
<td>Beasts_of_the_Southern_Wild</td>
<td>541.338,00</td>
</tr>
<tr>
<td>Alabama_Shakes</td>
<td>540.304,00</td>
</tr>
<tr>
<td>San_Francisco_49ers</td>
<td>538.824,00</td>
</tr>
</tbody>
</table>

**Top 50 - Peak in February 2013 - Superbowl**
Wikipedia: Peak + QFilter-ALL

1. Top 50 Peaks in 02/13
2. ALL QFilter < 0.7
3. Correlations

Top Elements Superbowl – Peak in February && ALL other months < 0.7

- Baltimore_Ravens
- Joe_Flacco
- Michael_Oher
- Roman_numerals
- San_Francisco_49ers
- Flag_of_the_United_States

Correlations

Top 50 Peaks in 02/13

ALL QFilter < 0.7

Correlations
Top 50 Peaks in 02/13

ALL QFilter < 0.7

Correlations

Wikipedia: P+QfALL Performance

QFilter ALL < 0.7 on Pages

- GPU (2xK40): 46.700ms, 558ms
- CPU (Xeon E5-2643): 1.161ms, 73.745ms

Natural languages (110,965,726 cells)

en (63,303,959 cells)

GPU: 83x, 0ms
CPU: 63x, 10.000ms, 20.000ms, 30.000ms, 40.000ms, 50.000ms, 60.000ms, 70.000ms, 80.000ms
Wikipedia: What’s new in June?

1. ALL elements > 4 June compared to Jan-May
2. ALL QFilter > 0.5 June compared to Jul-Dec
3. No Peak but steady interest
Wikipedia: WNij Performance

1. ALL elements > 4 June compared to Jan-May
2. ALL QFilter > 0.5 June compared to Jul-Dec
3. No Peak but steady interest

Graph:
- GPU (2xK40) vs CPU (Xeon E5-2643)
  - en: 46.700ms (63.303.959 cells)
  - Natural languages: 1.161ms (110.965.726 cells)
  - What's new: 1.466ms (117.484.560 cells)

Assessment:
- GPU: 83x faster
- CPU: 63x faster
- What's new: 70x faster
Wikipedia: What’s new (event)?

1. Aggregation + DFilter
2. On a daily base
3. Even more data

The chart shows the popularity of various names over a period from 2013-03-08 to 2013-04-01. The names include Francisco (papa), Jorge Bergoglio, Jorge Mario Bergoglio, Pope Francis, Franziskus (Papst), and Papa Francesco.
Future works

1. Multi-Node-GPU performance
2. Fast massive & continuous insertion
3. New OLAP Features
Visit us in the exhibit hall!

Visit **Jedox** at booth 1030!

Download at [www.jedox.com](http://www.jedox.com)
Tweet to [@JedoxAG](https://twitter.com/JedoxAG)
Mail to peter.strohm@jedox.com


Thanks to:

Alex Haberstroh, Jedox AG
Tobias Lauer, Jedox AG
Steffen Wittmer, Jedox AG