Multi-Dimensional, In-GPU-Memory Databases: Streaming Conditional Calculations in Big Data Sets

Peter Strohm  |  GTC15  |  San José  |  03/17/2015
2002
Founded in Freiburg, Germany

Today
Offices in Freiburg, Frankfurt, Düsseldorf, Paris, Boston

120
Global Business Partners

125
Countries with Jedox Users

25
Language Versions
Big Data in real-time: Twitter

- 500,000,000 tweets per day
- 6,000,000 localized tweets per day
- 60,000,000 data points per day
- > 4,000 tweets per minute
- 12 GB data volume per GPU
- < 1 second of computation time
Big Data in real-time: Twitter

Join the Jedox Session #S5481 about Big Data Analytics on GPU In-Memory databases at #gtc15 on Tuesday at 13:30!

California, USA

Twitter: Big Data in Real-time

Join the Jedox Session #S5481 about Big Data Analytics on GPU In-Memory databases at #gtc15 on Tuesday at 13:30!


<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Date</th>
<th>Hour</th>
<th>Minute</th>
<th>Second</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Language</th>
<th>TweetID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Join</td>
<td>2015-03-16</td>
<td>23</td>
<td>36</td>
<td>49</td>
<td>-121</td>
<td>37</td>
<td>en</td>
<td>577599223119548416</td>
</tr>
<tr>
<td>Jedox</td>
<td>2015-03-16</td>
<td>23</td>
<td>36</td>
<td>49</td>
<td>-121</td>
<td>37</td>
<td>en</td>
<td>577599223119548416</td>
</tr>
<tr>
<td>Session</td>
<td>2015-03-16</td>
<td>23</td>
<td>36</td>
<td>49</td>
<td>-121</td>
<td>37</td>
<td>en</td>
<td>577599223119548416</td>
</tr>
<tr>
<td>#S5481</td>
<td>2015-03-16</td>
<td>23</td>
<td>36</td>
<td>49</td>
<td>-121</td>
<td>37</td>
<td>en</td>
<td>577599223119548416</td>
</tr>
</tbody>
</table>

...
In-GPU-Memory OLAP-Database

1. All data in GPU memory

2. Hierarchical structure of consolidated elements

Business Driven Intelligence, Analytics & Performance Management.

www.jedox.com  @JedoxAG  @PSJedox  #gtc15  #s5481
In-GPU-Memory OLAP-Database

3 Calculation “on-the-fly”

4 Store only non-zero values

i In-GPU-Memory & “on-the-fly”
Jedox Social Analytics Workflow

Public Twitter Stream

Streaming via php in **Jedox**.Web

Adding Twitter data in **Jedox**.ETL

Update trending with **Jedox**.Web+ETL

Persistent Twitter Data in **Jedox**.OLAP

**Jedox**

ExcelPLUS

Mobile

Web

SA App

Business Driven Intelligence, Analytics & Performance Management.

www.jedox.com  -  @JedoxAG  -  @PSJedox  -  #gtc15  -  #s5481
Jedox Social Analytics with GPU

1. System: 2 x K40 with 12 GB GPURAM
   CPU with 128GB RAM

2. Data sets with > 500 million entries

3. Calculating “on-the-fly” on all data entries
Jedox Social Analytics Showcase
Jedox Social Analytics Showcase
Social Media Analytics: HeatMap

over 250,000 grid points

over 500 Million database entries
Social Media Analytics: Trending

1. Top 100 *Trending Factor*
   
   ![Image](image-url)

   

2. ~28,000 dictionary words
   
   > 80 Million DB entries
Social Media Analytics: Trending

1. Assigning *Trending Factor* to *virtual cells* (e.g. \([TF] = \frac{[B]}{[C]}\))

2. *Trending Factor* is *calculated on the fly*

3. Conditional Calculation: *IF-Rule*

   ```
   IF ([count] > 1000)
   THEN \([\text{count, currentDay}] / [\text{count, previousWeek}]\)
   ELSE 0
   ```
Conditional Rule with constants

\[ [TF] = \text{IF} \quad ([B] > 1000) \quad \text{THEN} \quad [C] \quad \text{ELSE} \quad 0 \]
Conditional Rule with constants

\[
[TF] = \text{IF} ([B] > 1000) \text{ THEN } [C] \text{ ELSE } 0
\]
Conditional Rule with constants

\[ [TF] = \text{IF } ([B] > 1000) \text{ THEN } [C] \text{ ELSE 0} \]
[TF] = IF

([B] > 1000) THEN [C] ELSE 0

Rule Processor

[T] = IF

([B] > 1000) THEN [C] ELSE 0

A,1
A,2
A,3
A,4
A,5

T

T
Conditional Rules on GPU

\[ [TF] = \text{IF } ([B] > 1000) \text{ THEN } [C] \text{ ELSE 0} \]
Conditional Rules on GPU

\[ [TF] = \text{IF} \quad \begin{cases} (B > 1000) & \text{THEN} \quad [C] \\ \text{ELSE} \quad 0 \end{cases} \]

Rule Processor

\[
\begin{array}{c}
\text{>}
\hline
1 \\
1 \\
0 \\
1 \\
0 \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
T \\
A,1 \\
A,2 \\
A,3 \\
A,6 \\
A,8 \\
\vdots \\
\vdots \\
\end{array}
\]
Conditional Rules on GPU

\[
[TF] = \text{IF} \ (B > 1000) \ \text{THEN} \ [C] \ \text{ELSE} \ 0
\]
Conditional Rules on GPU

GPU MEMORY

SPACE FOR CALCULATIONS
Conditional Rule on GPU

\[ [TF] = IF ([B] > 1000) \]  
\[ \text{THEN} [C] \]  
\[ \text{ELSE} 1 \]
Streaming in Conditional Rules

\[ [TF] = \text{IF} ([B] > 1000) \text{ THEN } [C] \text{ ELSE } [D] \]

Rule Processor

Matching Processor

Processor

Part 1

Part 2

Constant (1000)

Part 1

Part 2

Matching Processor

Total

Business Driven Intelligence, Analytics & Performance Management.

www.jedox.com - @JedoxAG - @PSJedox - #gtc15 - #s5481
Streaming in Conditional Rules

\[ [TF] = \text{IF} \quad ([B] > 1000) \quad \text{THEN} \quad [C] \quad \text{ELSE} \quad [D] \]

Rule Processor

Matching Processor

Part 1

> Processor

Part 2

Constant (1000)

Part 1

Matching Processor

Part 2

Total

TRUE

FALSE
Streaming in Conditional Rules

\[ TF = \text{IF} \ (B > 1000) \ \text{THEN} \ [C] \ \text{ELSE} \ [D] \]
Streaming in Conditional Rules

\[
[TF] = \text{IF} \quad ([B] > 1000) \quad \text{THEN} \quad [C] \quad \text{ELSE} \quad [D]
\]

Rule Processor

Matching Processor

Matching Processor

Constant (1000)

Part 2

Matching Processor

Part 2

Matching Processor

Part 2
Streaming in Conditional Rules

\[ [TF] = \text{IF} \quad ([B] > 1000) \quad \text{THEN} \quad [C] \quad \text{ELSE} \quad [D] \]
Streaming in Conditional Rules

\[ [TF] = \text{IF} \quad ([B] > 1000) \quad \text{THEN} \quad [C] \quad \text{ELSE} \quad [D] \]

1. Rule Processor
2. Matching Processor
3. Constant (1000)
4. Matching Processor
5. Part 2
6. True
7. Part 2
8. Matching Processor
9. False
10. Total

Business Driven Intelligence, Analytics & Performance Management.
www.jedox.com - @JedoxAG - @PSJedox - #gtc15 - #s5481
Streaming in Conditional Rules

$$[TF] = \text{IF } ([B] > 1000) \text{ THEN } [C] \text{ ELSE } [D]$$

Rule Processor

- Processor
  - > Processor
    - Matching Processor
      - Matching Processor
        - Constant (1000)
        - Matching Processor
      - Constant (1000)
    - Matching Processor
      - Matching Processor
      - Matching Processor

Part 2
Sub-Total
1
Streaming in Conditional Rules

\[ TF = \text{IF} \ (B > 1000) \ \text{THEN} \ [C] \ \text{ELSE} \ [D] \]

1. Rule Processor
2. Matching Processor
   - Processor
   - Constant (1000)
   - Matching Processor
   - Matching Processor
Streaming in Conditional Rules

\[ TF = \text{IF} \quad ([B] > 1000) \quad \text{THEN} \quad [C] \quad \text{ELSE} \quad [D] \]

Rule Processor

- \text{Matching Processor}
- \text{Constant (1000)}
- \text{Matching Processor}

complete result set
Features: Performance with GPU

1. Top 100
2. Heat Map
3. Trending Factor

![In-GPU-Memory database & calculations](image)

- Speed-up up to 50x and more!
Jedox Social Analytics Showcase

Jedox Social Analytics App download at: www.jedox.com/en/jedox-social-analytics

GPU- & Social Analytics App-Team:
Alexander Haberstroh
Steffen Wittmer
Leo Mehlig