#23164

FASTEST GPU-BASED OLAP AND DATA MINING: BIG DATA ANALYTICS ON DGX

Speaker: Roman Raevsky, Co-Founder & CEO, Polymatica
rar@polymatica.com
ABOUT POLYMATICA

OWN BI PLATFORM DEVELOPMENT SINCE 2010

OFFICES IN LONDON, MOSCOW, ZURICH

ENTERPRISE-SCALE RETAIL, FINANCE, TRANSPORTATION AND STATE CUSTOMERS

BASE ANALYTICAL PLATFORM FOR TAX AUTHORITY

WE ARE PROUD OF OUR MAIN PARTNERS

NVIDIA®

IBM®
WHAT IS ADVANCED ANALYTICS?

MAIN QUESTIONS

WHAT YOU COULD DO WITH THE DATA?

WHO CAN OPERATE WITH THE DATA?

HOW FAST AND FLEXIBLE ARE THE MOST POPULAR ANALYTICAL ACTIONS (DRILL-DOWN, ROTATE, GROUPING)?
WHY OLAP?

FLEXIBILITY

- MULTIDIMENSIONAL VIEW
- INSTEAD OF FLAT TABLES
- FOR USERS: VISUAL PLAYING WITH DATA

TIME-TO-SOLUTION:
- REACHING RESULTS MUCH FASTER
- NO NEED OF CONSULTING OR EXTRA PROGRAMMING SERVICES
WHY OLAP? CONCERNS

- OBSOLETE TECHNOLOGY
- SPEED DEGRADATION AS $N \times \log(N)$ ON CPU
- HUGE RESOURCE REQUIREMENTS
- PRE-AGGREGATED LEVELS AND COMPLEX FORMULAS
HOW TO FIX OLAP CONCERNS?

- **SPEED DEGRADATION AS N*LOG(N) ON CPU**
- **HUGE RESOURCE REQUIREMENTS**
- **PRE-AGGREGATED LEVELS AND COMPLEX FORMULAS**

- **SORTING SPEED AS O(N) ON GPU**
- **HYBRID CPU+GPU ARCHITECTURE**
- **EVERYTHING IS CALCULATED “ON-THE-FLY”**
NEW ANALYTICAL PARADIGM

- ABSOLUTE FREEDOM WITH DATA MANIPULATIONS
- RAW DATA VOLUMES
- REAL TIME PROCESSING

- OLAP-BASED
- HIDDEN MACHINE LEARNING ALGORITHMS
- EASY AND INTUITIVE INTERFACE FOR USERS
- PRODUCING INTERACTIVE VISUALIZATION AND PIXEL-PERFECT REPORTS
INTEGRATED OLAP + MACHINE LEARNING

“MAGIC BUTTONS”

NO SETTINGS, JUST DRAG & DROP

HYBRID CPU+GPU PROCESSING

NO NEED TO LEARN

ADDING MUCH MORE VALUE TO BUSINESS

“CLUSTERING” IS THE MOST POPULAR BUTTON
FROM 2D TO REAL LIFE PICTURE
WHY SPEED IS CRITICAL FOR USER

HUMAN ASPECT

- CONCENTRATION TIME DECREASED
- STRESS AND ROUTINE
- FUNDAMENTAL EDUCATION FALL

12 SEC IN 2000
8 SEC IN 2015

WHY SPEED IS CRITICAL FOR BUSINESS

BUSINESS ASPECT

- TIME-TO-DELIVER FOR THE NEW IDEAS
- COMPETITION FOR CUSTOMER ATTENTION
- THE SPEEDIER YOU ARE THE MORE VALUE YOU GAIN

USE CASE

- CAMPAIGN CONVERSION RATES GREW FROM 14% TO 35%
- TECHNOLOGY PENETRATION – FROM 3 BUSINESS ANALYSTS TO 40 IN 6 MONTHS
- ANY BUSINESS ANALYST GENERATES MORE PROFIT THAN THE WHOLE SAS DEPARTMENT
<table>
<thead>
<tr>
<th>1 BILLION-RECORDS OLAP CUBE</th>
<th>SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 YEAR</td>
<td>960’000’000</td>
</tr>
<tr>
<td>HISTORY OF RECEIPT LINES FROM 20 HYPERMARKETS</td>
<td>RECORD DATASET</td>
</tr>
<tr>
<td>2 YEAR</td>
<td>4,2 TB</td>
</tr>
<tr>
<td>HISTORY OF BANK TRANSACTION FOR 2 MILLION CARDS</td>
<td>INITIAL DATA (MS SQL DATABASE)</td>
</tr>
<tr>
<td>1 BILLING DAY</td>
<td>490 GB</td>
</tr>
<tr>
<td>OF A NATIONAL MOBILE OPERATOR</td>
<td>POLYMATICA OLAP-CUBE</td>
</tr>
<tr>
<td></td>
<td>NO DATA DETAIL LOSS</td>
</tr>
<tr>
<td></td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>DIMENSIONS, 22 FACTS</td>
</tr>
</tbody>
</table>
### TEST SCENARIO

The image shows a screenshot of an OLAP cube interface with the title "OLAP cube - BANK" at the top. The interface appears to be part of a data analysis tool, likely for financial or banking data analysis.

#### Dimensions and Facts

- **Dimensions**:
  - **Date**
  - **Year**
  - **Quarter**
  - **Month**
  - **Week**
  - **Weekday**
  - **Hour**

- **Facts**:
  - **Σ Amount**

#### Data Table

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Facts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Total</td>
</tr>
<tr>
<td>Year</td>
<td>Σ Amount</td>
</tr>
<tr>
<td>Quarter</td>
<td>Total</td>
</tr>
<tr>
<td>Month</td>
<td>69,002,278,136.11</td>
</tr>
<tr>
<td>Week</td>
<td></td>
</tr>
<tr>
<td>Weekday</td>
<td></td>
</tr>
<tr>
<td>Hour</td>
<td></td>
</tr>
</tbody>
</table>

The total amount is displayed as 69,002,278,136.11.
## Test on 1 Billion Records – Intel Platform

**2*Intel Xeon E5-2667 3.2 GHz, 512 GB RAM, 1*Tesla P100**

<table>
<thead>
<tr>
<th>Operation</th>
<th>CPU+GPU Acceleration, Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Dim to the Left</td>
<td>1.21</td>
</tr>
<tr>
<td>2 Dims to the Left</td>
<td>1.34</td>
</tr>
<tr>
<td>3 Dims to the Left</td>
<td>1.62</td>
</tr>
<tr>
<td>Apply Profile</td>
<td>1.40</td>
</tr>
</tbody>
</table>

![Graph showing CPU+GPU acceleration times for different operations]
TEST ON 1 BILLION RECORDS – IBM POWER PLATFORM*

2*POWER8 2.86 GHZ (20-CORES), 512 GB RAM, 4*TESLA P100 (“MINSKY”)

<table>
<thead>
<tr>
<th>MACHINE</th>
<th>CPU+GPU ACCELERATION, TIMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DIM TO THE LEFT</td>
<td>2.0</td>
</tr>
<tr>
<td>2 DIMS TO THE LEFT</td>
<td>1.58</td>
</tr>
<tr>
<td>3 DIMS TO THE LEFT</td>
<td>1.59</td>
</tr>
<tr>
<td>APPLY PROFILE</td>
<td>1.06</td>
</tr>
</tbody>
</table>

*Draft testing of non-optimized code (cross-platform compilation)
**TEST ON 1 BILLION RECORDS – NVIDIA DGX-1**

2*INTEL XEON, 512 TB RAM, 8*TESLA P100

<table>
<thead>
<tr>
<th>OPERATION</th>
<th>CPU+GPU ACCELERATION, TIMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DIM TO THE LEFT</td>
<td>6.33</td>
</tr>
<tr>
<td>2 DIMS TO THE LEFT</td>
<td>3.11</td>
</tr>
<tr>
<td>3 DIMS TO THE LEFT</td>
<td>2.85</td>
</tr>
<tr>
<td>APPLY PROFILE</td>
<td>1.26</td>
</tr>
</tbody>
</table>

![Bar Chart](chart.png)

NVIDIA DGX is kindly provided by: [SCAN® Business](www.scan.co.uk/business/deep-learning)  
Email: aiservices@scan.co.uk  
Stand G.06
ALL PLATFORMS

1 DIM TO THE LEFT
2 DIMS TO THE LEFT
3 DIMS TO THE LEFT
APPLY PROFILE

INTEL PLATFORM
IBM POWER PLATFORM
NVIDIA DGX-1
ANALYTICAL PLATFORM OF THE FUTURE

THE NEAR FUTURE

SCENARIOS – ROBOTIZATION
OF ANALYTICAL WORK

CUBE EDITIONS – BUDGETING, MODELING
AND PLANNING ALL-IN-ONE

SEMANTIC AND AI INTEGRATION – SEARCH
FOR BUSINESS MEANING
Thank you!

Roman Raevsky, Co-Founder & CEO, Polymatica

Polymatica system demo: Booth E.25
rar@polymatica.com
www.polymatica.com