A Case for Online Mixed Workload Processing

Jens Krüger, Christian Tinnefeld, Martin Grund, Alexander Zeier, Hasso Plattner
Introduction

- Data model acted as main guideline for DBMS development
- For a long time: few DBMSs for many different applications
- Recent DBMS discussions: heavily application oriented
  - Two examples:
    - Stonebraker et al– One size fits all
    - Vogels et al- VLDB Keynote 2007 / Amazon Dynamo
- DBMS for Enterprise Applications
  - Beginning of 1990s separation into OLTP and OLAP
  - But, we claim:

  \textit{OLTP-style workloads also require the ability to frequently compute OLAP-style aggregate queries}
1. Company estimates future demand for its products – 
   Demand Planning / Supply Chain Management
- Customers contact the company and place orders – 
  Sales Order Processing / Enterprise Resource Planning
- The availability of the requested products have to be checked – 
  Available-to-Promise / Supply Chain Management
- Customers fall behind on their payments – 
  Dunning / Financial Accounting
- The company wants to analyze its sales performance – 
  Sales Analytics / Enterprise Resource Planning
Available-to-Promise

**Temporary aggregate containing number of products to be produced**

**Subtracted by**

**Temporary aggregate containing number of products already promised to customers**

**Result set containing number of products that still can be promised to customers**

**Sales Order promised on 05.02.2008**

**Sales Order promised on 07.02.2008**

<table>
<thead>
<tr>
<th>05.02.2008</th>
<th>06.02.2008</th>
<th>07.02.2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>70</td>
<td>90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>05.02.2008</th>
<th>06.02.2008</th>
<th>07.02.2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>140</td>
<td>150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>05.02.2008</th>
<th>06.02.2008</th>
<th>07.02.2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>135</td>
<td>130</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>05.02.2008</th>
<th>06.02.2008</th>
<th>07.02.2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>
### Comparison of Application Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Demand Planning</th>
<th>Sales Order Processing</th>
<th>Available to Promise</th>
<th>Dunning</th>
<th>Sales Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Granularity of Data</strong></td>
<td>Transactional</td>
<td>Transactional</td>
<td>Transactional</td>
<td>Transactional</td>
<td>Pre-Aggregated</td>
</tr>
<tr>
<td><strong>Operations on Data</strong></td>
<td>Read &amp; Write</td>
<td>Read &amp; Write</td>
<td>Read &amp; Write</td>
<td>Read &amp; Write</td>
<td>Read-Only</td>
</tr>
<tr>
<td><strong>Preprocessing of Data</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Timeframe of Data</strong></td>
<td>Historical &amp; Recent</td>
<td>Recent Only</td>
<td>Historical &amp; Recent</td>
<td>Historical &amp; Recent</td>
<td>Historical &amp; Recent</td>
</tr>
<tr>
<td><strong>Update Cycles of Data</strong></td>
<td>Always Up-to-Date</td>
<td>Always Up-to-Date</td>
<td>Always Up-to-Date</td>
<td>Always Up-to-Date</td>
<td>Cyclic Updates</td>
</tr>
<tr>
<td><strong>Amount of Data per Query</strong></td>
<td>Large</td>
<td>Small</td>
<td>Large</td>
<td>Large</td>
<td>Large</td>
</tr>
<tr>
<td><strong>Complexity of Queries</strong></td>
<td>High</td>
<td>Standard</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Predictability of Queries</strong></td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Response Time of Queries</strong></td>
<td>Seconds</td>
<td>Seconds</td>
<td>Seconds</td>
<td>Seconds to Hours</td>
<td>Seconds to Hours</td>
</tr>
</tbody>
</table>

*OLTP Characteristics are colored light grey*  
*OLAP Characteristics are colored dark grey*
- Large amount of data is needed to perform transactional query
- Nothing new, mixed workloads are a well established topic
- But: here they are originated by a single application.

<table>
<thead>
<tr>
<th>Granularity of Data Operations on Data</th>
<th>Transactional Read &amp; Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing of Data</td>
<td>No</td>
</tr>
<tr>
<td>Timeframe of Data</td>
<td>Historical &amp; Recent</td>
</tr>
<tr>
<td>Update Cycles of Data</td>
<td>Always Up-to-Date</td>
</tr>
<tr>
<td>Amount of Data per Query</td>
<td>Large</td>
</tr>
<tr>
<td>Complexity of Queries</td>
<td>High</td>
</tr>
<tr>
<td>Predictability of Queries</td>
<td>Medium</td>
</tr>
<tr>
<td>Response Time of Queries</td>
<td>Seconds</td>
</tr>
</tbody>
</table>
Conclusion

- Not only Web 2.0 companies need application-specific data stores
- Mixed workloads are originated by a single application
- Huge potential for increasing performance and functionality of enterprise applications by supporting mixed workload

Future Work
- Need for a Mixed Workload Benchmark
  - TPC-E and TPC-H claim one part of the process separately
  - But, OLTP-style workloads also require the ability to frequently compute OLAP-style aggregate queries
- DBMS Draft for Mixed Workloads
  - Read-optimized, in-memory columnar store with transactional support
Backup Slides
Application Characteristics at Amazon

Diagram:

- **Object Size**
  - > 1MB: AC Set #1
  - < 1MB
    - Query Model
      - Primary Key Access
        - Strong Consistency
          - yes: AC Set #2
          - no
            - yes: AC Set #6
            - no
              - Relation-Driven Queries
                - yes: RDBMS Features
                - no
                  - Low Latency
                    - yes: AC Set #5
                    - no: AC Set #4
          - no
            - yes: AC Set #3
            - no: AC Set #7
      - Multi-Attribute Queries
        - yes
          - Strong Consistency
            - yes: AC Set #8
            - no: AC Set #9
        - no
          - RDBMS Features
            - yes
              - Low Latency
                - yes: AC Set #9
                - no: AC Set #8
            - no
              - Low Latency
                - yes: AC Set #8
                - no: AC Set #9
            - no
              - Low Latency
                - yes: AC Set #9
                - no: AC Set #8
            - no
              - Low Latency
                - yes: AC Set #9
                - no: AC Set #8
Sales Order Processing Data Logs

Load for all customers

Workload

Fiscal year

January, February, March, April, May, June, July, August, September, October, November, December