Analysis Framework for Reduced Data Warehouse

Authors: Franck Ravat, Jiefu Song, Olivier Teste
Team SIG, IRIT, Toulouse.

EDA-2015
Brussels, Belgium
Plan

1. Context

2. Multi-states analyses
   2.1 Data management
   2.2 Analysis processing

3. Implementation

4. Conclusion
• **General Context**  
  o Data is stored permanently in a **Multidimensional Data Warehouse (MDW)**  
  o New data are periodically added

• **Issues**  
  o Low analysis performance due to important and increasing data volume.  
  o Decreasing relevance of detailed information with age.

• **Objective**  
  OLAP analyses in MDW keeping only useful data over time:  
  • Propose a complete analysis process.  
  • Facilitate decision-makers’ tasks.
• Related Work

1. Modeling solutions for data reduction in MDW

- Partial (incomplete) solutions: either the fact or dimensional data updates
  [Chen et al. 2002], [Skyt et al. (2008)], [Kimball and Ross (2011)], [Golfarelli and Rizzi (2009)],
  [Iftikhar and Pedersen (2011)]

- Generalized data reduction process to the whole MDW schema [Atigui et al. (2014)]

2. Analysis framework compatible with reduced MDW

- No discussion about analysis-relative component
  [Chen et al. 2002], [Golfarelli and Rizzi (2009)], [Kimball and Ross (2011)], [Iftikhar and Pedersen (2011)]

- Limited discussion about querying solution in reduced MDW
  [Morzy and Wrembel (2004)], [Skyt et al. (2008)]

- Complete analysis support framework including
  - data management,
  - analysis processing
  - interactive data restitution.
Plan

1. Context

2. Multi-states analyses
   2.1 Data management
   2.2 Analysis processing

3. Implementation

4. Conclusion
Multi-states analysis

- Data management

Extended classical MDW model [Atigui et al. (2014)]:
- A MDW is composed of a set of star schemas;
- Each star schema, called state, is valid for a certain period of time.
Multi-states analysis

- Data management

Metamodel

<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
<th>IDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td>DGEOGRAPHY</td>
<td></td>
</tr>
<tr>
<td>d2</td>
<td>DTIMES</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td>DTHEME</td>
<td></td>
</tr>
<tr>
<td>d4</td>
<td>DGEOGRAPHY_E1</td>
<td>d1</td>
</tr>
<tr>
<td>d5</td>
<td>DTIMES_E1</td>
<td>d2</td>
</tr>
<tr>
<td>d6</td>
<td>DTHEME_E1</td>
<td>d3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IDF</th>
<th>NAME</th>
<th>IDD</th>
<th>NAME</th>
<th>DATES</th>
<th>DATEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>f2</td>
<td>FNEWS_E1</td>
<td>d5</td>
<td>DTIMES_E1</td>
<td>2010</td>
<td>2014</td>
</tr>
<tr>
<td>f2</td>
<td>FNEWS_E1</td>
<td>d4</td>
<td>DGEOGRAPHY_E1</td>
<td>2010</td>
<td>2014</td>
</tr>
<tr>
<td>f2</td>
<td>FNEWS_E1</td>
<td>d6</td>
<td>DTHEME_E1</td>
<td>2010</td>
<td>2014</td>
</tr>
</tbody>
</table>

Meta_Dimension

Meta_Hierarchy

Meta_Level

- pos
- typa

Meta_Attribute

- nameAtt

Meta_Element

- nameEle

Meta_Measure

Meta_Star

- dateS
- dateE

is derived from

derive

Meta_Fact

is derived from

derive

derive
Plan

1. Context

2. Multi-states analyses
   2.1 Data management
   2.2 Analysis processing

3. Implementation

4. Conclusion
Multi-states analysis

- Analysis Engine
  Analysis processing
Multi-states analysis

- Analysis Engine

Algebraic operator: Drilldown\textsuperscript{multi-states}

\textbf{Algorithm 1: Drilldown\textsuperscript{multi-states}} ([I]; MT_k, D_i; P_{\text{visf}})

Input: Set of temporal intervals I, displayed multidimensional table MT_k, displayed dimension D, parameter P. Output: new multidimensional table MT_{k+1}

1. Let $H_{\text{actual}}$ be the actually displayed hierarchy
2. Let $P_{\text{actual}}$ be the actually displayed parameter
3. If $P_{\text{actual}} \not\in H_{\text{actual}}$ then
   - Impossible operation
   Else
5. Find the subset of states $E_z \rightarrow \forall E_j \in E_z \mid I_{E_j} \in I \vee I_{MT_k}$ — Adaptation
6. Let $P_{\text{Drilldown}} = P$
7. Let $r = FALSE$
8. While $ALL_D \not\in H_{\text{actual}}$ $P_{\text{Drilldown}} \land r = FALSE$
9. If $\forall E_j \in E_z \mid P_{\text{Drilldown}} \in A_{E_j}^P$ then
   - $r = TRUE$
10. Else
11. $P_{\text{Drilldown}}$ increases one granularity level
12. End if
13. End While
14. If $r = FALSE$ then
15. Impossible operation
16. Else
17. For $E_z \in E_z$
18. Let $MT_{k+1}^{E_z}$ be the part of MT in states $E_z$
19. Translate $\text{Drilldown}(MT_{k}^{E_z}, D, P_{\text{Drilldown}})$ into query $Q$ — Translation
20. $MT_{k+1}^{E_z} = \text{Result of query } Q$
21. $MT_{k+1} = \bigcup MT_{k+1}$
22. End for
23. End if
24. End if

Context Multi-states analysis Implementation Conclusion
Multi-states analysis

- Analysis Engine
  Example: $MT_0$ - Number of published news by Continent from 2000 to 2015

  Drilldown_{multi-states} ([2000, 2015], $MT_0$, DGeography, Ville)

  Drilldown_{multi-states} ([2000, 2015], $MT_0$, DGeography, Country)

  Drilldown ($MT_{E1}$, DGeography, Country)

  Drilldown ($MT_{E2}$, DGeography, Country)

  Query over $E2$

  Query over $E1$

  Results from 2000 to 2010

  Results from 2010 to 2015

  Results from 2000 to 2015
Plan

1. Context

2. Multi-states analyses
   2.1 Data management
   2.2 Analysis processing

3. Implementation

4. Conclusion
• Architecture of analysis framework
Implementation

• Interactive restitution

Graphical interface

Context
Multi-states analysis
Implementation
Conclusion
Implementation

• Interactive restitution
  Example:
  published news in the world by month from 2000 to 2014

Drilldown\textsuperscript{multi-states}

published news by month and by continents from 2000 to 2014

\begin{verbatim}
SELECT SL.FACT_NAME AS FACT, 
  SL.DIMENSION_NAME AS DIML, 
  SC.DIMENSION_NAME AS DIMC 
FROM META_STAR SL, 
  META_FACT F, 
  META_DIMENSION DL, 
  META_DIMENSION DC, 
  META_STAR SC 
WHERE SL.NAME = F.NAME 
  AND SC.NAME = SL.NAME 
  AND F.NAME_PARENT = 'FNEWS' 
  AND DL.NAME = SL.DIMENSION_NAME 
  AND DL.NAME = 'DGRAPHIE' 
  AND DC.NAME = 'DTEMP' 
  AND SL.DATES <= TO_DATE ('30-12-2014', 'DD-MM-YYYY') 
  AND SC.DATEE >= TO_DATE ('01-01-2000', 'DD-MM-YYYY')
FROM FNEWS_E1 FAIT, DGEOGRAPHY_E1 DIML, DTIMES_E1 DIMC, DTIMES_E1 DMT 
WHERE FAIT.ID_DGEOGRAPHY_E1 = DIML.ID_DGEOGRAPHY_E1 
  AND FAIT.ID_DTIMES_E1 = DMT.ID_DTIMES_E1 
  AND TO_DATE (DIML.NUMMONT, 'MM-RRRR') 
  BETWEEN TO_DATE ('01-01-2000', 'DD/MM/RRRR') 
  AND TO_DATE ('30-12-2014', 'DD/MM/RRRR') 
  AND FAIT.ID_DTIMES_E1 = DIMC.ID_DTIMES_E1 
GROUP BY DIML.COUNTRY, DIMC.NUMMONT

(UNION)

(SELECT SUM(NBN) AS FNEWS, DIML.COUNTRY, DIMC.NUMMONT 
FROM FNEWS_E1 FAIT, DGEOGRAPHY_E1 DIML, DTIMES_E1 DIMC, DTIMES_E1 DMT 
WHERE FAIT.ID_DGEOGRAPHY_E1 = DIML.ID_DGEOGRAPHY_E1 
  AND FAIT.ID_DTIMES_E1 = DMT.ID_DTIMES_E1 
  AND TO_DATE (DIML.NUMMONT, 'MM-RRRR') 
  BETWEEN TO_DATE ('01-01-2000', 'DD/MM/RRRR') 
  AND TO_DATE ('30-12-2014', 'DD/MM/RRRR') 
  AND FAIT.ID_DTIMES_E1 = DIMC.ID_DTIMES_E1 
GROUP BY DIML.COUNTRY, DIMC.NUMMONT)
\end{verbatim}
Implementation

- Interactive restitution

  Example:
  - Published news in the world by month from 2000 to 2014
  - Published news by month and by continents from 2000 to 2014

Drilldown^multi-states

---

Context  Multi-states analysis  Implementation  Conclusion
Plan

1. Context

2. Multi-states analyses
   2.1 Modeling solution
   2.2 Analysis processing

3. Implementation

4. Conclusion
A MDW model composed of a set of states over time.
- Decreases data volume
- Facilitate decision-makers’ tasks

A generic multi-states analysis framework
- Graphical interface: interaction with decision-makers
- Analysis engine: algebraic operators and its execution algorithms
- Data management: meta-model and its instances

Implementation: graphical multi-states analysis tool
- Automatic generation of queries
- Transparency of data reduction

Future work
- Influence of data reduction over pre-aggregated data
- Other analysis operators