Managing Volatile Data within a Data Warehouse: It Can Be Done
Ed Schaefer
Senior Applications Developer
Enrollment & Marketing Research, Enrollment Management & Marketing
DePaul University

Association for Institutional Research, 2010 Annual Forum
Chicago, Illinois
June 1, 2010
Background...
Large, urban, private, Catholic

Nine colleges/schools

Six campuses; overseas programs
Enrollment Management & Marketing (EMM)

Admissions
Career and Money Management
Center for Access and Attainment
Enrollment & Marketing Research
Enrollment Systems & Operations

Financial Aid
Office of Institutional Planning and Research
Student Records
University Marketing Communications
University Marketing Strategy
EMM Data Warehouse

Admissions and Financial Aid
Research and Analysis
Office of Institutional Planning and Research
Enrollment & Marketing Research
What we know about DW...
Large, complex...

...and pricey technology
Time consuming to build...
DW Concepts...
Integrated
Time-variant
Subject oriented

Non-volatile

Bill Inmon
So, what's my issue with non-volatile...
Data changes are the T of ETL...
Types of slowly changing dimensions (SCD)

- Type 1: Truncate and reload; history rewritten daily
- Type 2: New rows written with data changes; full historical capture
- Type 3: Previous and new value columns; partial historical capture
When it comes to non-volatile...

Type 1  OK

Type 2  X
New business process impacts history

- Biology
- Mathematics
- Etc...

- Biology
- Mathematics
- Etc...

- Journalism
- Communication
- Etc...

- Journalism
- Communication
- Etc...
New logic impacts history
History is just wrong

Mr.
When it comes to non-volatile...

Type 1  OK  Type 2

New business process impacts history

- Biology
- Mathematics
- Journalism
- Communication
- Etc...

New logic impacts history

History is just wrong

Mr. ?
EMM looks for DW partner in Information Services (IS)
EMM DW needs:

- Wish to utilize university DW...but not maintain

Information Services (IS) DW:

☑ Maintain university DW
EMM DW needs:

- Maintain near daily captures of cleaned data

Information Services (IS) DW:

☑ Daily loads
EMM DW needs:

- Supply history for YTD and trends

Information Services (IS) DW:

✓ Type 2 SCDs
EMM DW needs:

- Must be able to modify history when needed

Information Services (IS) DW:

❌ Non-volatile
Our solution...
Data from DW-Q may feed into the transformation process.

**REPORTING EXTRACTS**
- Copy
- Restructure data but data values remain unaltered

**Data Warehouse**
- DW Data Quality (DW-Q)
- DW Enterprise (DW-E)
- DW Transformed (DW-T)
- DW Census (DW-C)

**Quality Feedback Process**
- Apply OIPR, EMR, and FA transformation logic

**E/I: Errors and Inconsistencies**

**EMM Reporting Systems include:**
- Admission Summary
- Enrollment update report
- HCRC data extract
- EMM report suite
- OIPR Fact File & cubes

Reporting systems vary between departments but data from the DW does not.

Further census data checks and, if required, transformations.
Example of how it works...
D_PERSON in DW has GENDER
D_PERSON in DW has GENDER

?
EMM builds and maintains D_PERSON_AUX with R_GENDER
Integrated: PERSON_KEY in D_PERSON_AUX

<table>
<thead>
<tr>
<th>D_PERSON</th>
<th>D_PERSON_AUX</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Female Icon" /></td>
<td><img src="image" alt="Female Icon" /></td>
</tr>
<tr>
<td>?</td>
<td><img src="image" alt="Male Icon" /></td>
</tr>
<tr>
<td><img src="image" alt="Male Icon" /></td>
<td><img src="image" alt="Male Icon" /></td>
</tr>
<tr>
<td>?</td>
<td><img src="image" alt="Female Icon" /></td>
</tr>
<tr>
<td><img src="image" alt="Female Icon" /></td>
<td><img src="image" alt="Female Icon" /></td>
</tr>
</tbody>
</table>
Since EMM controls
D_PERSON_AUX...

<table>
<thead>
<tr>
<th>D_PERSON</th>
<th>D_PERSON_AUX</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Female" /></td>
<td><img src="image" alt="Female" /></td>
</tr>
<tr>
<td><img src="image" alt="Unknown" /></td>
<td><img src="image" alt="Male" /></td>
</tr>
<tr>
<td><img src="image" alt="Male" /></td>
<td><img src="image" alt="Female" /></td>
</tr>
<tr>
<td><img src="image" alt="Female" /></td>
<td><img src="image" alt="Male" /></td>
</tr>
</tbody>
</table>
Groundrules...

- UPDATE non-key fields only
- INSERT cannot split a record
- DELETE with extreme caution
- Only IS can execute DELETE
Groundrules...

- UPDATE non-key fields only
- INSERT cannot split a record
- DELETE with extreme caution
- Only IS can execute DELETE

Likelihood...

- Somewhat likely
- Very unlikely
- Extremely unlikely
Thank you!

For further questions or comments:

Ed Schaefer  
eschaefe@depaul.edu