Regardt, O., Rönnbäck, L., Bergholtz, M., Johannesson, P., Wohed, P.

Anchor Modeling
Agile Modeling Technique using the Sixth Normal Form for Structurally and Temporally Evolving Data

“You can never step into the same river twice.”
Entity Relationship Modeling

Sixth Normal Form

Temporal Databases

```
select pCU.GEN_Gender, pCU.CUHAC_CustomerHairColor, 
COUNT(pCU.CUHAC_CustomerHairColor) as Customers 
from pCU_Customer('1985-11-09') pCU 
where pCU.CUDOB_CustomerDateOfBirth < '1980-01-01' 
group by pCU.GEN_Gender, pCU.CUHAC_CustomerHairColor
```
Attributes and ties come in four flavors, historized or static combined with knotted
**ANCHORS**

```sql
SELECT top 5 * from CU_Customer
```

<table>
<thead>
<tr>
<th>CU_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

**KNOTS**

```sql
SELECT top 5 * from GEN_Gender
```

<table>
<thead>
<tr>
<th>GEN_ID</th>
<th>GEN_Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
</tr>
</tbody>
</table>

**ATTRIBUTES**

```sql
SELECT top 5 * from CUDOB_CustomerDateOfBirth
```

<table>
<thead>
<tr>
<th>CU_ID</th>
<th>CUDOB_CustomerDateOfBirth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1905-03-02</td>
</tr>
<tr>
<td>2</td>
<td>1905-07-02</td>
</tr>
<tr>
<td>3</td>
<td>1908-09-14</td>
</tr>
<tr>
<td>4</td>
<td>1910-02-03</td>
</tr>
<tr>
<td>5</td>
<td>1912-04-01</td>
</tr>
</tbody>
</table>

**TIES**

```sql
SELECT top 5 * from CUHH_Customer_Household
```

<table>
<thead>
<tr>
<th>CU_ID</th>
<th>HH_ID</th>
<th>HOW_ID</th>
<th>CUHH_FromDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2009-02-13</td>
</tr>
<tr>
<td>1</td>
<td>895</td>
<td>0</td>
<td>2009-09-21</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2006-10-17</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2002-08-20</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1993-08-29</td>
</tr>
</tbody>
</table>
All previous versions of the schema are present and were never modified, allowing extensions to be made "online".
**Latest View**

- Joins all attributes and finds the attribute row with the latest FromDate if historized.

**Point-in-Time Function**

- Joins all attributes and finds the attribute row with the latest FromDate earlier or on the given timepoint if historized.

```sql
select
    pCU.GEN_Gender,
    pCU.CUHAC_CustomerHairColor,
    COUNT(pCU.CUHAC_CustomerHairColor) as Customers
from
    pCU_Customer('1985-11-09') pCU
where
    pCU.CUDOB_CustomerDateOfBirth < '1980-01-01'
    group by
    pCU.GEN_Gender,
    pCU.CUHAC_CustomerHairColor
```
The query execution plan shows that only two tables are touched (the anchor and the selected attribute) despite of the fact that several others are joined into the view we are using.
The query optimizer will remove table $T$ from the execution plan of a query if the following two conditions are fulfilled:

i. no column from $T$ is explicitly selected

ii. the number of rows in the returned data set is not affected by the join with $T$

Support
Microsoft SQL Server
Oracle
IBM DB2
PostgreSQL
MariaDB (fork of MySQL)
Teradata (partial)
Pseudo loading code given "wide" source data:

- Check if there already is an associated surrogate key for each natural key
- For unknown individuals
  - Create and associate surrogate keys
  - Directly insert data into all relevant tables
    (most tables including the anchor)
- For known individuals
  - If this is a delta file, directly insert data into all relevant tables
  - If this is not a delta file, check if the value in the source differs from the latest value in the destination and insert if the data is new
    (few tables excluding the anchor)
Ease of Modeling
Simple concepts and notation
Historization by design
Iterative and incremental development
Reduced translation logic

Simplified Maintenance
Ease of temporal querying
Absence of null values
Reusability and automation
Asynchronous arrival of data

High Performance
High run-time performance
Efficient storage
Parallelized physical media access

www.anchormodeling.com